

Through the crisis  
UK SMEs performance during the 'credit crunch'

Meng Ma



THE UNIVERSITY  
*of* EDINBURGH

Degree of Doctor of Philosophy  
in Management Science and Business Economics

## • Contents

List of Figures.....	I
List of tables.....	III
Acknowledgement.....	V
Abstract .....	1
1. Introduction.....	3
1.1 Motivation .....	8
1.2 Aims and research questions .....	9
1.3 Contributions.....	11
1.4 Objective of Research.....	12
1.5 Underlying Philosophy.....	14
1.6 Structure of this Thesis.....	15
2. Literature and Regulation Reviews.....	18
2.1 Introduction.....	18
2.2 Background.....	18
2.2.1The ‘Credit Crunch’ on SMEs.....	18
2.3.1 Influential Regulations.....	22
2.4 Credit risk modelling.....	33
2.5 Literature review for panel data model .....	38
2.6 Literature review for GAM .....	41
2.7 Summary.....	43
3. Methodology .....	44
3.1 Introduction.....	44
3.2 Data transformation .....	45
3.2.1 Coarse classification .....	46
3.2.2 Weights of Evidence .....	47
3.3 Variable Selection.....	49
3.4 Relevant Models.....	50
3.4.1 Logistic Regression .....	51

3.4.2 Panel Data Models.....	53
3.4.3 Residual check .....	63
3.4.4 Generalised additive model.....	66
3.5 Seperation Measures.....	75
3.5.1 AUROC, Gini and KS-statistics.....	75
3.5.2 H measure .....	78
3.5.3 Akaike Information Criterion (AIC) .....	79
3.6 Summary.....	80
4 Data description .....	81
4.1 SMEs Definition .....	81
4.2 Sample .....	82
4.3 Segments .....	84
4.4 'Bad' rates.....	85
4.5 Macroeconomic Variables (MVs) .....	87
4.6 Variable Selection .....	89
4.7 Explained variables' description.....	92
4.7.1 SMEs performance according to SIC, region and account lateness. ....	93
4.7.2 The missing category .....	100
4.7.3 Other variables' trends.....	103
4.8 Replacing missing values .....	106
4.8.1 Non-start-up SMEs .....	107
4.8.2 Start-up SMEs .....	114
4.9 Conclusion .....	119
5. Results .....	121
5.1 Logistic regression .....	122
5.2 Panel data.....	126
5.2.1 Introduction:.....	126
5.2.2 Adding year dummies.....	128
5.2.3 Adding Macroeconomic Variables.....	128

5.2.4 Results for 'start-up' SMEs .....	135
5.2.5 Results for 'non-start-up' SMEs.....	136
5.3 Residual check .....	141
5.4 GAM results .....	143
5.4.1 GAM with WoE Data.....	144
5.4.2 GAM using original format of continuous variables.....	149
6. Conclusion .....	189
6.1 Research questions answered.....	189
6.2 Limitations and further research.....	192
Appendix:.....	194
Bibliography.....	198

## List of Figures

Figure 3.1 ROC (Souce: Li, 2014).....	76
Figure 4.1 UK SMEs 'bad' rate during the 'credit crunch' .....	86
Figure 4.2 UK Macroeconomic data from 2007 to 2011 .....	88
Figure 4.3 Start-up SMEs categorical frequency according to 1992 SIC code.....	94
Figure 4.4 Non-start-up SMEs categorical frequency according to 1992 SIC code .....	94
Figure 4.5. Start-up SMEs 'bad' rate according to 1992 SIC code .....	95
Figure 4.6 'bad' rate for non-start-ups according to 1992 SIC code.....	96
Figure 4.7 'start-up' SMEs statistics according to region .....	97
Figure 4.8 'Non-start-up' SMEs statistics according to region .....	97
Figure 4.9 SMEs frequency of lateness of account through 'credit crunch'.....	98
Figure 4.10 Frequency and 'bad' rate of Non-start-up SMEs according to number of current directors .....	101
Figure 4.11 (left) Start-ups 'Bad' Rate according to Proportion of Current Directors to Previous Directors in the Last Year.....	102
Figure 4.12 (right) Non-start-ups 'Bad' Rate according to Proportion of Current Directors to Previous Directors in the Last Year.....	<b>Error! Bookmark not defined.</b>
Figure 4.13 Start-up SMEs frequency and 'bad' rate according to Oldest Age of Current Directors .....	104
Figure 4.14 Non-start-up SMEs 'bad' rate according to the percentage change in shareholders' funds.....	105
Figure 4.15 No. of 'Current' Directors: 'bad' rate of missing V.S. MA of observed value (2007-2010).....	107
Figure 4.16 No. of 'Current' Directors: 'bad' rate of missing V.S. MA of observed value ....	108
Figure 4.17 Current Directors Proportion: 'bad' rate of missing V.S. MA of observed value	109
Figure 4.18 No. of Previous Searches: 'bad' rate of missing V.S. MA of observed value .....	109
Figure 4.19 Time since last Derog: 'bad' rate of missing V.S. MA of observed value.....	110
Figure 4.20 Lateness of Account: 'bad' rate of missing V.S. MA of observed value.....	111
Figure 4.21 Figure 20 Time since Last Annual Return: 'bad' rate of missing V.S. MA of observed value .....	112
Figure 4.22 Pct. of Total Fixed Assets: 'bad' rate of missing V.S. MA of observed value .....	112
Figure 4.23 Pct. Change in Total Assets: 'bad' rate of missing V.S. MA of observed value ..	113
Figure 4.24 missing frequency for Proportion of current Director .....	114
Figure 4.25 Oldest Age of Current Directors: 'bad' rate of missing V.S. MA of observed value (2007-2010) .....	114

---

Figure 4.26 Oldest Age of Current Directors: 'bad' rate of missing V.S. MA of observed value .....	115
Figure 4.27 No. of Previous Searches: 'bad' rate of missing V.S. MA of observed value .....	116
Figure 4-28 Time since Last Derog. Data: 'bad' rate of missing V.S. MA of observed value .....	116
Figure 4.29 Lateness of Accounts: 'bad' rate of missing V.S. MA of observed value .....	117
Figure 4.30 Time since last Annual Return: 'bad' rate of missing V.S. MA of observed value .....	117
Figure 4.31 Prop. of Current Directors: 'bad' rate of missing V.S. MA of observed value ...	118
Figure 4.32 Total Assets: 'bad' rate of missing V.S. MA of observed value.....	118
Figure 5.1 histogram plotting for estimated residual V.S. normal distribution (2007) (left)	142
Figure 5.2 histogram plotting for estimated random effects V.S normal distribution. (2007) (right).....	142
Figure 5.3 QQ plot for estimated residual V.S. normal distribution (2007) (left) .....	142
Figure 5.4 QQ plot for estimated random effects V.S normal distribution. (2007) (right) ...	142
Figure 5.5 additive effects for Number of 'Current' Directors: .....	154
Figure 5.6 additive effects for Proportion of Current Directors to Previous Directors in the Last Year .....	156
Figure 5.7 additive effects for PP Worst (Company DBT - Industry DBT) in the Last 12 Months .....	158
Figure 5.8 Total Value of Judgements in the Last 12 Months .....	160
Figure 5.9 additive effect for Number of Previous Searches (last 12 months) .....	161
Figure 5.10 additive effects for Time since last derogatory data item (months).....	163
Figure 5.11 additive effects for Lateness of Accounts .....	164
Figure 5.12 additive effects for Time since Last Annual Return.....	166
Figure 5.13 additive effects for Total Fixed Assets as a Percentage of Total Assets.....	169
Figure 5.14 additive effects for Percentage Change in Total Assets .....	170
Figure 5.15 additive effects for Proportion of Current Directors to Previous Directors in the Last Year .....	179
Figure 5.16 additive effects for Oldest Age of Current Directors/Proprietors supplied (Years) .....	180
Figure 5.17 additive effects for Number of Previous Searches (last 12m).....	182
Figure 5.18 additive effects for Lateness of Accounts .....	183
Figure 5.19 additive effects for Time since Last Annual Return.....	185

---

## List of tables

Table 3.1 Clasification table .....	76
Table 4.1 Segment size across 'credit crunch' .....	85
Table 4.2 selecting significant explanatory variables .....	90
Table 5.1 logistic regression results for start-ups .....	123
Table 5.2 logistic regression results for non-start-ups .....	125
Table 5.3 MVs sign for both segments when adding one MV to the model.....	130
Table 5.4 AIC value when adding one averaged MV into panel model .....	133
Table 5.5 start-up SMEs random effect panel data models parameter estimation.....	139
Table 5.6 non-start-up SMEs random effect panel data model parameter estimation.....	140
Table 5.7 model fittings of training sample.....	141
Table 5.8 model fittings of holdout sample .....	141
Table 5.9 Summary of insignificant smoothers for both segments at level 5% .....	144
Table 5.10 GAM via WoE parameter estimation for non-start-up SMEs Note: 1. ***refers to significant at level 1%; 2. **refers to significant at level 5%; 3. *refers to significant at level 10% .....	148
Table 5.11 GAM model separation measures for training sample .....	149
Table 5.12 GAM model separation measures for holdout sample .....	149
Table 5.13 Parameter estimation: GAM with original values for non-start-up SMES .....	153
Table 5.14 correlation between lateness of account and annual reports .....	168
Table 5.15 independent variables' trend summary for non-start-ups.....	174
Table 5.16 non-start-ups separation measure: GAM & logistic regression based on data's original format.....	176
Table 5.17 Parameter estimation: GAM with original values for start-up SMES.....	178
Table 5.18 start-ups separation measure: GAM & logistic regression based on data's original format.....	186

---

To my son Tinghe Gu, my husband Xin Gu, my Mum and Dad



---

## Acknowledgement

I want to thanks my supervisor Jake and Galina. Thank you for enlightening my mind and guide me to the knowledge of credit risk.

Dear Tinghe, your name is picked up from *Song of Chu* which is a very romantic traditional Chinese poem, however, you seems only cares about how to make noise and destroy all the cups. But mummy always loves you.

My friends, you stayed with me and you are always in my heart!

---

## **Abstract**

The influence of ‘credit crunch’ on Small and Medium sized Enterprises (SMEs) has been of concern to the government, regulators, banks, the enterprises and the public. Using a large dataset of UK SMEs’ records covering the early period of the ‘credit crunch’, the influence of the ‘credit crunch’ on SMEs have been studied. It uses cross-sectional method, panel data models and GAM to provide a detailed examination of SMEs performance. Both newly established and matured SMEs, segmented by age, are considered separately. The data contains 79 variables which covered obligors’ general condition, financial information, directors’ portfolio and other relevant credit histories.

The ‘credit crunch’ is a typical ‘black swan’ phenomenon. As such there is a need to examine whether the stepwise logistic model, the industries prime modelling tool, could deal with the sudden change in SMEs credit risk. Whilst it may be capable of modelling the situation alternatives models may be more appropriate. It provides a benchmark for comparison to other models and shows how well the industry’s standard model performs. Given cross-sectional models only provide aggregative level single time period analysis, panel models are used to study SMEs performance through the crisis period. To overcome the pro-cyclic feature of logistic model, macroeconomic variables were added to panel data model. This allows examination of how economic conditions influence SMEs during ‘credit crunch’. The use of panel data model leads to a discussion of fixed and random effects estimation and the use of

---

explanatory macroeconomic variables. The panel data model provides a detailed analyse of SMEs' behaviour during the crisis period.

Under parametric models, especially logistic regression, data is usually transformed to allow for the non-linear correlation between independent variable and dependent variable. However, this brings difficulty in understanding influence of each independent variable's marginal effects. Another way of dealing with this is to add non-parametric effects. In this study, Generalized Additive Models (GAM) allows for non-parametric effects. A natural extension of logistic regression is a GAM model with logistic link function. In order to use the data in their original state an alternative method of processing missing values is proposed, which avoids data transformation, such as the use of weights of evidence (WoE). GAM with original data could derive a direct marginal trend and plot how explanatory variables influence SMEs' 'bad' rate. Significant non-parametric effects are found for both 'start-ups' and 'non-start-ups'. Using GAM models results in higher prediction accuracy and improves model transparency by deriving explanatory variables' marginal effects.

---

## 1. Introduction

Credit risk is defined as *'the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms'* by the Basel Committee and Banking supervision, (BCBS 1999, p. 1). Before the statistical method proposed by Fisher (Fisher, 1936) was recognized to classify customers in 1950, an obligor's credit risk was estimated by a specialist's experience. The early issue of credit cards in 1966, such as Barclay cards and BankAmericard, proved the power of credit scoring in its early applications, (Thomas, 2009). Since then statistical models have become widely used to evaluate customers' credit risk. With the development of financial markets and the fast growing number of loans, credit risk models have become more salient in assessing credit risk for both retail and SMEs. After 2007 the global financial crisis, known as 2007 or 'credit crunch, credit risk models have seen a period of increased the attention. The global financial crisis challenges the existing credit rating system and illustrates the importance of risk management.

This crisis is regarded as the most significant economic disaster since the 'great depression'. Started in 1929, the 'great depression' decreased the US real GDP by 26.5 %. By comparing the 'great depression' with the 2007's global financial crisis, a world bank report produced by Brahmhatt and Silva (2009) says that the recent recession has more impact on financial market which uniquely caused the 'credit crunch', Its impact is especially significant for the banking sector. Almost all the major economies suffered from negative GDP growth rate in 2009, including the UK,

---

France, Germany, Japan and the US. Given the importance of financial industry in the UK, the impact of the crisis is even more severe than elsewhere. The ‘credit crunch’ has caused a negative GDP growth rate in both 2008 and 2009. In the latter year, the UK experienced 4.3% of GDP decline, which is the biggest economic constrain since 1980 (the World Bank, 2015). In such a downturn in the business cycle, obligors’ performance has become even worse due to supervisors’ preferences for prudential strategy, leading to banks’ decreasing lending and higher counterparty default risk. According to Moody’s reports, UK firms’ annual default rate jumped from 0.9% in 2007 to the extremely high level of 13.1% in 2009 (Moody’s, 2011).

Among banks’ borrowers, Small and Medium sized Enterprises (SMEs) are an interesting and unique group. During the latest global financial crisis, SMEs lending has attracted scrutiny due to their importance in the UK’s economy, high default risk and media reports on SMEs struggling to get financial support. According to BIS’s data in 2010, more than 99% of enterprises in the UK are SMEs, which employ 56.7% of the workforce in the job market (BIS, 2010; Ma and Lin, 2012). In general, SMEs account for 90% of businesses and employ a major part of the labour force in Europe. Therefore, SMEs are often regarded as the ‘engine of European economy’ (Verheugen, 2005). However, this ‘engine’ seemed to lack ‘petrol’ during the ‘credit crunch’. The amount of monthly lending to SMEs in UK dropped from 991m to 566m between 2008 and 2010 (British Banking Association, 2010). Such a big decline has caused national concern about the SME default rate. Ma and Lin (2012) summarised

---

the media's concerns and the SMEs' prospects of financial difficulties during the 'credit crunch'. SMEs in a wide range of industries reported a lack of financial support. More than 2800 SMEs went through bankruptcy during February 2009 in the UK according to Panorama (Panorama, 2009). In the researcher's data, the UK SME 'bad' rates were 6.9%, 11.8%, 16.1% and 11.9% from 2007 to 2010. There was a noticeable increase in the 'bad' rate in 2008, however, 2009 was the most difficult year for SME survival. Compared to 2007, which can be regarded as the normal economic condition, the UK SME 'bad' rate increased by almost ten per cent in 2009. Although, the 'bad' rate reduced in 2010, it was still significantly higher than that of 2007. Yet, similar to other SME financial issues, there were surprisingly little research focusing on SME risk modelling during this period.

In fact, SME credit risk is a less discussed topic compared to other counterparty risks, such as credit cards and listed large operations. Altman and Sabato (2007, p. 333) even states that '*The only study that the authors are aware of that focused on modelling credit risk specifically for SMEs is a fairly distant article by Edmister (1972).*' Although there is research covering SMEs credit risk, for example Dietsch and Petey (2002) estimate French SMEs Variance at Risk (VaR) and their marginal risk in 2002 and Behr and Guttler (2007) employed logit model on a German SMEs data set in 2007. Yet SMEs credit risk has not received as much discussion compared to credit risk modelling of consumer or large corporations. The lack of research could be due to various reasons. Amongst them lack of data is a key issue. Basel II

---

introduced Internal Rating-Based (IRB), SMEs default modelling becomes essential for financial institutions. Larger banks want to establish their own rating system in order to decrease their capital requirement. As a special type of obligor, SMEs' credit scoring becomes more widely used and discussed since then. According to Basel II, SMEs can be treated as retail if their total exposures have not exceeded €1 million (BCBS, 2004). Therefore, the risk weight of SMEs is reduced from 100% to 75% (Lin, 2009). This reduced the capital requirement of SME loans and made financial support more accessible for SMEs. SMEs seemed to lose the advantages of lower associated capital requirement when the Basel III accord came into force. Basel III is proposed by the Basel Committee in order to enhance supervision on banks and guarantee that banks have sufficient capital during any kind of economic conditions, and especially during severe recessions, see Ma (2007). In doing so, banks are required to have a countercyclical capital and maintain their leverage ratio above 3%. The increase of capital requirement could significantly increase the cost of lending for smaller financial institutions, which provide a large number of SMEs with financial support. Therefore, these actions are believed harmful for SME lending (Tolley, 2010). However, it is difficult to summarise the overall effect of Basel III since it has not been fully implemented. Therefore, no accurate data could be collected to estimate its impact (ACCA, 2011).

The changing regulations, Basel Committees' special treatment of SMEs and the 'credit crunch' have encouraged more scholars to carry out SME relevant research in

---

recent years. However, several disadvantages appear in most SME research. As discussed previously, data accessibility is always the primary concern in this field. Compared to other obligors, SME information is more opaque, which causes information asymmetry for their credit suppliers. Another issue with SME data is the low default rate of loans, which is a common phenomenon for credit scoring in normal economic conditions. That data limitation restricts research in SMEs credit risk modelling. For example, there are only 120 defaults out of 2,010 US firms' data in Altman and Sabato (2007); meanwhile, the date set in other research is even smaller, such as Lin (2009) who was only able to identify 28 bankruptcies out of 450 listed SMEs in the UK. The second problem is that the widely used single stage model cannot account for the impact of business cycle, such as what happened during the 'credit crunch'. Is the widely used logistic model robust under these circumstances? How are SMEs affected by the business cycle? Due to the limitation of data, those fundamental questions remain unresolved for SMEs. The third issue is to explore the marginal influence of variables on SMEs' 'bad' rate. A common belief is that variables influence obligors' credit risk in a non-linear way. Therefore, original variables have often been transformed before they are entered into credit scoring models. Hence, independent variables' influence cannot be interpreted by their coefficients. The non-linear nature in probability of default (PD) models makes it hard to demonstrate independent variables' contribution towards firms' credit risk. Hence, the writer is motivated to carry on this research and make contributions to this field.



---

## 1.1 Motivation

Despite the importance of SMEs, there are much fewer researches in the field of SME credit risk compared to large corporations' credit ratings and credit card analysis. However, more concerns have been raised regarding lending to SMEs due to the 'credit crunch'. Therefore, this research should be interesting for an academic audience since it adds to the discussion of SME performance during the 'credit crunch', explores new methodology of SME credit risk modelling, improves forecasting accuracy and explains how independent variables influence SME performance. At the same time, the practitioners may be more interested in standard models' performance during the 'credit crunch' and the influence of business cycles.

In addition, lending to SMEs is a big social issue during the 'credit crunch'. There are countless media reports about SME performance during this period. Hence the rise in my interest in SME credit risk modelling, and I would like to make a contribution to improve the prediction of SMEs 'bad' rate. There are lots of entrepreneurs whose businesses provide unique products unlike larger firms' goods which are more standardised. The lack of research into SME credit risk drives financial institutions away from those obligors. I would therefore like to carry on research and build more reliable credit scoring models to separate 'good' SMEs from 'bad'. Therefore, I can help find reliable entrepreneurs who should receive financial support and build a better business. Secondly, financial institutions face a difficult situation on SME lending. On one hand the authorities want to expand the power of SMEs to boost the

---

economy after the financial crisis, yet they also require banks to carry on a prudential lending policy to reduce their expected losses. I want to study the features of ‘bad’ SMEs to help financial institutions choose ‘good’ firms to support. In this way, the financial institutions’ sustainability will be enhanced.

In summary, my interest in understanding SME performance is due to the intense media interest and would like to improve the ‘bad’ rate prediction to help maintain financial sustainability.

## **1.2 Aims and research questions**

This research explores the SME ‘bad’ rate, especially focusing on their performance during the past ‘credit crunch’. Based on the existing literature, there are three questions being proposed by the author:

1. Could the widely used logistic model provide accurate prediction of SMEs PD?

Logistic models are industrial standard methods in credit scoring. During the past ‘credit crunch’, the high default rate of banks’ portfolios raised concern about the existing rating systems. This question aims to explore whether the standard model could separate ‘good’ from ‘bad’ with a high prediction accuracy during the ‘credit crunch’. Logistic regression models are also benchmark models to compare other models to, such as panel data and generalized additive models.

---

2. How should modellers employ multi-period models to SME PD modelling?

Further, how could analyst reflect macroeconomic changes during business cycle?

The aim is to expand SME PD modelling from single time period methods into multi-period models. The author analysed panel data models and selected suitable model for SME modelling. To reflect the influence of macroeconomic conditions, Macroeconomic Variables (MVs) are involved in SME modelling. The author provides analysis concerning MV format, influence and effects of improving model performance.

3. Is there any non-linear effect in SME performance, especially during ‘credit crunch’? If any, can credit risk prediction be improved by involving those effects? Furthermore, could independent variables’ marginal trend be derived by those non-linear effects and showing their trend during the ‘credit crunch’?

The objective is to consider non-parametric effects to improve PD model performance for credit scoring models since linear assumptions are challenged by the ‘credit crunch’. Since the ‘credit crunch’ can be regarded as a ‘black swan’ event, the commonly used assumptions are questioned for this period. Non-parametric models are ideal for such circumstances and may provide more accurate analysis without imposing too many constraints. The variables’ trends are not revealed explicitly when using dummy variables or weights of evidence (WoE). For example, significant dummies may vary from time to time. Using dummies could not fully illustrate

---

variables' trends. WoE converts continuous variables to categorical variables and re-orders the categories of the variable. This research provides an alternative way of processing original data and involving non-parametric effects. The derived trend of each independent variable would clearly demonstrate their trend during the 'credit crunch'.

### **1.3 Contributions**

Most researches in the section of credit risk focus on large corporates. This research extends the understanding of SME credit risk in various ways. First of all, this research is based on a very large UK SME dataset which contains the majority of, if not all, UK SMEs which borrow from financial institutions. Therefore, this research provides a rich landscape of UK SME behaviour during the 'credit crunch'. Secondly, a standard industry model is built which could answer the question of whether current SME credit scoring models are reliable during a crisis time. Thirdly, by applying panel modelling the researcher extends SME credit scoring from commonly used single time period into multi-period modelling. Although, a few researchers have used panel data model, such as Valles (2006), the writer pointed out several disadvantages which could lead to model risk and significant bias. It contributes to the literature on credit risk by comparing panel models to cross-sectional ones, and for the former by exploring the suitability of Random Effects (RE) versus Fixed Effects (FE). Panel models allow inclusion of Macroeconomic Variables (MVs) which is the fourth contribution of the paper: it is now possible to develop business cycle models, such as

---

‘credit crunch’ into SME credit modelling. Last but not least, the writer presents the marginal distributions of factors that influence SME credit risk. The marginal distributions are derived from a semi-parametric model—Generalised Additive Models (GAM). The non-parametric smoothers in GAM overcome linear assumptions commonly used in credit scoring models. Therefore, coarse classification, which changes continuous variable into categorical format, is not the best solution to this issue. In order to use continuous variables in their original formats and derive their marginal trends, this research proposes an alternative way of processing missing values which substitute the missing values with observed values with similar performance.

## **1.4 Objective of Research**

The aim of this study is to improve credit scoring models for SMEs, especially focusing on models used in crises such as the ‘credit crunch’ which occurred between 2007 and 2010. Scoring models are designed for various reasons, yet the most common one is that financial institutions use such models to distinguish a ‘good’ customer from a ‘bad’ one. A ‘good’ scoring model will be able to separate those two groups and hence may achieve a lower default rate for financial institutions. Given the size of the data the research should be able to search for the best performing model to separate ‘good’ SMEs from ‘bad’ ones during this time period. Due to the economic condition between 2007 and 2010, this research is particularly interested in

---

investigating the nature of the change that saw a dramatic rise in the default rate during this period. In detail, the research objectives are:

1. SME performance during the ‘credit crunch’ challenged the current scoring models. Review literatures and background to find gaps in SME credit scoring. Then, in order to support literature gaps and highlight modelling issues, summarise UK SME statistic features during the ‘credit crunch’.
2. To explore whether credit scoring models fail to separate ‘good’ from ‘bad’ during the ‘credit crunch’. This paper aims to join the discussion that when a ‘black swan’ event occurs, such as that which happened during the latest global financial crisis, whether standard model could sufficiently separate ‘good’ from ‘bad’.
3. Seeking for models which solve the disadvantages of current SME scoring models. Firstly, current models could not reflect how economic changes influence SME credit risk. Hence, any new model should be able to solve this issue whether this inclusion could improve separation accuracy. Secondly, it is doubtful if normal distribution fit SME performance during the ‘credit crunch’. Hence, this research will explore marginal trend of independent variables and how it could improve the model during a crisis period. Thirdly, current models either assume independent variables are linearly correlated with dependent variables or translating the independent variables first, such as the use of dummies or weights

---

of evidence. Hence, any new model should be able to present independent variables trends and show how those trends alter during a changing economy.

## **1.5 Underlying Philosophy**

The research design follows a deductive research framework. A review of the literature on SME credit scoring and other risk models is firstly carried out. The limitations of the existing SME credit scoring models raised the issue of whether there were better models to use. Obviously the introduction of macroeconomic data into the models using panel data models seemed appropriate. When exploring non-linearity, then GAM are the most appropriate models. A traditional frequentist approach is taken given the size of the data and research questions to be addressed. The data covers the ‘credit crunch’ and contains the vast majority of behaviours exhibited by UK SMEs which borrow from financial institutions. Given the data it is possible to test whether alternative models could improve the performance of the modelling of SME default and so shed light on SME behaviour during the ‘credit crunch’.

In this work it is assumed that the variables considered will act as determinants of performance of SMEs. This may be a simplification but leads naturally to assumptions of an objective world in which, perhaps, the difficulty arises from our observation of the truth not that the truth is a social construction. In that sense the underpinning philosophy is positivistic or post-positivistic.

---

## 1.6 Structure of this Thesis

The rest of this thesis is constructed in the following way: Chapter two provides a review of the background of SME credit risk and current literature. There are a series of regulations which control the banks' risk management. The most important policies are the Basel Accords. The development of the Basel Accords changed the perception of SME credit risk and SME access to finance. Therefore, the development of the Basel Accords are reviewed, along with their influence on SME credit risk. As a newly introduced regulation, IFRS9 is also briefly reviewed as models used in this research could be used for IFRS9 modelling. Subsequently, this chapter also reviews the commonly used credit scoring models in retail banking relevant to research into SME performance. Literature review for models used in this research will be presented later in relative chapters.

The third chapter explains the methodology employed in current research. The methods of transformation of original data are reviewed. To solve the collinearity issue among independent variables, variable selection is an essential step before building models. Then, the models used in this research are explored, which include logistic models, panel data models and generalised additive model (GAM). Furthermore, an alternative method of processing missing values is proposed to avoid transformation of data by WoE. With this method, variables' marginal effects are explored by GAM.



---

The fourth chapter describes the employed dataset in detail. It demonstrates SME crisis performance according to various factors, such as their industry classification, region, accounting information and other features. Subsequently the focus switches to missing values and other variables trends which support the assumptions that SMEs have to be segmented and non-parametric effects have significant impacts on SMEs credit scoring. In addition, these initial analyses shows annual differences of SME performance during the ‘credit crunch’. As an alternative method is proposed to process the missing value for GAM, the last section discusses how a continuous variables’ missing value is imputed by a matching observed value.

The following three chapters give a thorough demonstration of results and findings of this research. The fifth chapter is the result of variable selection and logistic models which is the benchmark model. It answers the research question whether the current industry standard model can accurately separate ‘good’ SMEs from ‘bad’ during the ‘credit crunch’. Panel data model results are provided in the sixth chapter. The panel data model builds a time series effect into SME modelling. As panel data is a multi-period model, it can control the annual difference by time dummies or the economic condition by adding MVs. Further, panel data results are compared with logistic regression to show the benefits of involving time effects. Additionally, this chapter also discusses how different MVs influence SME performance during the ‘credit crunch’. The seventh chapter presents the results from GAM. Firstly, the residuals derived from the panel data model is not normally distributed as had been assumed.

---

Therefore, a semi-parametric model is suggested to provide a better fit for the trend of SMEs. Then, the GAM model is applied using WoE transformation. This model tests which variables have significant non-parametric effects and how the non-parametric effects may improve the power of separation. The last section of this chapter provides results received from GAM with continuous variables' original format. By doing so, the impact of continuous variables' marginal effects are explored and the variables' influence on SME performance is investigated.

The last chapter summarises the thesis, provides the implication of the research and points out its limitations. This research explores UK SME performance by standard models and innovative models. It covers different aspects of SME credit risk during the 'credit crunch' and improves SME PD modelling in various ways. However, due to the limitations of data, the author is not able to answer questions such as how financial institutions' strategy influences SMEs.

---

## **2. Literature and Regulation Reviews**

### **2.1 Introduction**

This chapter discusses the background of this research and provides a literature review of previous research in credit risk modelling. As this research particularly focuses on the ‘credit crunch’ time period it is essential to identify any special circumstances. During the ‘credit crunch’ the economy undergoes significant changes, which lead supervisors to design new regulations to control financial institutions’ credit risk management. Therefore, the background section also reviews regulation changes which influence credit risk modelling. Well-developed credit risk models are reviewed in the credit risk modelling section. Literature about models used in this research will be provided in the methodology chapter.

### **2.2 Background**

#### **2.2.1 The ‘Credit Crunch’ on SMEs**

The ‘credit crunch’ defined by the Council of Economic Advisors (1991) as “a situation in which the supply of credit is restricted below the range usually identified with prevailing market interest rates and the profitability of investment projects” (quoted by Pazarbasioglu, 1997; Costa and Margani, 1999; Wehinger, 2014). A similar definition is given by Bernanke and Lown (1991), quoted by Holton, Lawless and McCann (2013), “a significant leftward shift in the supply for bank loans, holding constant both the safe real interest rate and the quality of the borrower.” During a ‘credit crunch’ the lenders

---

usually follow a very prudential strategy to reduce their lending associated with other phenomena such as a rise of interest rate, higher rejection rate, more arrangement fees, more collateral required and other issues. As this research focuses on a crisis period the ‘credit crunch’ influence has to be considered. This section reviews the influence of the ‘credit crunch’ on SMEs.

Wehinger’s OECD (Organisation for Economic Co-operation and Development) paper provides a review of SME financial difficulties during the financial crisis (Wehinger, 2014). The lending survey of ECB (European Central Bank) is used by Wehinger to provide evidence in the Eurozone which says one out of four SMEs in the Euro area have suffered from a shortage of credit. Besides the 2007 global financial crisis, sovereign debt issues caused a Euro area crisis starting in 2009. SMEs in the Eurozone suffered from significant credit constraints and banks’ lending decreased as the economy became weaker. Clearly, prudential policies were followed by banks to decrease the risk level they were taking (Holton, Lawless and McCann, 2013).

In a BIS UK survey of SME Finance (UKSMEF), Fraser (2010; 2012) investigates 2500 UK SMEs and concludes that there is a significant credit rotation among SME financing. Larger amounts of loans have been issued to SMEs with ‘lower’ risk while more risky SMEs suffer from much higher rejection rate. Fraser criticises banks for a lack of investigation before making decisions to reject SMEs. Armstrong, Davis, Liadze and Rienzo (2013) provides a further analysis of the UKSMEF by considering more firm features and their results support that UK SMEs suffered from credit shortage. The

---

economic downturn and credit constraints resulted in extremely high default rate amongst SMEs. The daily closure rate of SMEs reached as high as 100 and their stories of struggling to survive were told by a media which raised even more public concern (Howes, 2008; Panorama, 2009; Ma and Lin , 2010).

Series of recessions have been recorded in history. Hill and Thomas (2010) used UK macroeconomic data from 1700 and analysed different recessions that the UK has passed through. Although the last financial crisis, the ‘credit crunch’, was influenced more by the global market and the government provided more frequent policy changes to maintain money supply, all past recessions have a great level of similarity. Therefore, researching SME score card models during the last recession would not only help to understand the past event but could also enhance credit risk systems during economic downturn in the future.

Research concerning SME credit risk during the ‘credit crunch’ also covers other aspects. Nassr and Wehinger (2015) suggests that more effect should be made by other sources of finance besides traditional banking to enhance SME finance. However, trade credit has a negative influence with firms’ performance (Yazdanfar and Ohman, 2014). Bank lending has always been the main channel to finance SMEs (Armstrong, Davis, Liadze and Rienzo, 2013). And banks lending decisions are based on applicants’ score received from scorecards. For SME applications, besides the quantitative credit risk analysis, some may argue that SMEs have negotiation power gained through relationship lending. Compared to building score cards, relationship lending mainly uses ‘soft’ information to judge SME

---

credit risk. That information is not collected by accountant reports or credit bureaus data, but from qualitative methods such as interviews, entrepreneurial or other personal connections (Torre, Peria and Schmukler, 2010). However, relationship lending is not a solution to increase SME credit accessibility during crises. This has been shown by both the 1990s crisis and the last globe crisis. Fraser (2012) summarises that the relationship between SMEs and their credit suppliers are less reliable during times of crisis, with high rejection rates, especially from small banks, forcing SMEs to change their bank to gain credit. In addition, for such a large data, it is almost impossible to collect all those SMEs soft information during the ‘credit crunch’. Therefore, this research focus on quantitative data only.

In summary, during the latest global financial crisis, banks reduced their supply of credit which causes credit constraints for their obligators. If firms had insufficient financial support, the economy would be further damaged. Therefore, improving SME credit risk models becomes even more important during a downturn-economy since it would help banks to separate ‘good’ firms from ‘bad’. ‘Good’ firms would likely get credit and be more able to survive. Therefore, the economy could be maintained in a healthier condition.

SMEs ‘credit crunch’ is not the only problem banking systems suffered from during the latest global financial crisis. As the global financial crisis is closely related to the aggregative risk levels in mortgage markets and other credit risk issues, several regulators and supervisors have updated their regulations to guide or force banks to improve their

---

risk management frameworks. These regulations are aimed to incorporate business cycles into the framework and improve transparency of risk management. Among them, the most influential regulations are Basel III and IFRS9. The following section reviews the regulation background of this research and explains why this particular period is interesting to investigate.

### **2.3.1 Influential Regulations**

#### **i) Basel frameworks**

In 1973 USA came off the gold standard which led to extreme volatility of the US dollar. The well-known Bretton Woods system had fallen apart and banks experienced high exposure in foreign exchange, especially in G10 countries. Hence, the Basel Committee was established in order to maintain banks' capital levels under economic or financial crisis. To provide consistent supervision, the Basel Committee proposed their framework based on careful trading off among *simplicity, comparability and risk sensitivity* (BIS, 2015). The Basel Committee is part of the Bank for International Settlement which is based in a Switzerland city, Basel. Although the Basel Committee does not have any force to apply their standards, the participating central banks and authorities have accepted standards given by the committee and supervising financial institutions in their countries to conform to the adopted Basel framework (BIS, 2015). The latest update on European adoption is provided by the European Banking Authorities (EBA) to further develop the Internal Rating-Based (IRB) approach, (EBA, 2015). Then, the Bank of England (BoE) provided a more detailed framework

---

to further clarify definitions and standards used in the UK (BoE, 2015). These regulations provided guidelines for financial institution to improve their credit risk scoring models, increase the quality of their portfolio and reduce their required capital. All aspects of financial institutions are now influenced by those regulations. Here, only brief review of regulation developments is given to provide the regulations monitoring financial institutions to improve their credit scoring models. Further, different regulation gives different treatment of SME lending. In addition, the changing regulations also influence the accessibility of SME finance and raises new concerns about SME lending.

### ***Basel I***

The first milestone of the Basel Committee is the release of Basel Accord I in 1988 which is also known as the *Basel Capital Accord*. This accord is revolutionary due to its simplicity. It gives a single standard to calculate capital requirement for financial institutions (Lin, 2009). Assets are separated into four categories: risk-free assets, such as sovereign debts of OECD countries; other public claims and securities which have low default rate; retail loans which are secured by properties and other financial instruments with high default risk. Different weights are given according to their associated credit risks, namely: zero weight, 20 per cent, 50 per cent and 100 per cent weight correspondingly. The *Basel Capital Accord* also requires Tier one and total capital (Tier one plus Tier two capital) ratio no smaller than 4 per cent and 8 per cent correspondingly (BCBS, 1988).



---

The simple, required capital calculation that Basel I proposed would be beneficial for supervisors, yet several disadvantages appear as the economic condition changes and financial markets become more and more sophisticated. Capital requirement of Basel I ignored business cycle and operational risk. For instance, risk appears in various formats besides credit risk, which may lead banks to incur losses and so requires capital to help financial institutions to survive. Furthermore, Basel I could lead to poor estimates of capital requirement since it ignores the differences in credit risk for obligors that fall into the same category. For instance, SMEs are assumed to have the same level of risk worldwide as long as their countries adopt the Basel I.

### ***Basel II***

To overcome problems raised by Basel I, the committee proposed the *Revised Capital Framework* (BCBS, 2004). It took six years for the committee to reach an agreement and finally published the Basel II proposals in 2004. The most influential change is that banks are encouraged to using rating systems to price their risk and to establish their required capital. This replaced the uniform capital requirement calculation given by Basel I. According to the Basel Committee, these changes were made to encourage ‘financial innovations’ (BCBS, 2015).

There are two approaches of rating systems: the standard approach and the Internal Ratings-Based (IRB) approach. If financial institutions undertake the standard approach, their capital requirement would remain rather similar to those of Basel I. However, the IRB approach can

---

significantly reduce the capital requirements. The more comprehensive and robust their rating systems are, the more reduction they would receive on their capital requirements.

The SME default modelling becomes a much more appealing topic for both industrialist and academic due to the requirement of IRB approach. According to Basel II, SMEs can be treated as retail if their total exposures have not exceeded €1 million (Ma, 2012). Therefore, the risk weight of SMEs is reduced from 100% to 75% (Lin, 2009).

Besides the introduction of the IRB approach, Basel II classified risks into three categories: credit risk, market risk and operational risk. It also requires the following risk components to demonstrate banks' credit risk (BIS, 2006 paragraph 474):

1. Probability of default (PD): Banks need to use at least five years data to calculate their portfolio's estimated PD. Basel II define PD as the maximum value of the estimated PD and 0.03%. This rule applies to both standard and Advanced-IRB (A-IRB) approach.
2. Lost Given Default (LGD): LGD refers to the percentage of EaD when obligors default. If banks use the A-IRB approach, their LGD would be estimated by their own approach. Otherwise, they should follow policies made by their supervisors to calculate LGD. Under either circumstance, LGD should not be lower than their portfolio's weighted average loss.

- 
3. Exposure at Default (EaD): EaD measures the gross net of potential losses banks will be exposed to if the liability could not be fulfilled. Similar to LGD, the adoption of A-IRB approach makes banks more flexible to decide their own EaD.
  4. Maturity (M): M is defined as the maximum value between one year and the remaining life of liabilities in years.

Using the relevant risk components, the required capital under Basel II is calculated as following:

$$\text{Capital} = \text{PD} \times \text{LGD} \times \text{EaD}$$

This equation provides a clear link between required capital and credit risk components which are PD, LGD and EaD. The introduction of Basel II made the banking industry more enthusiastic to build credit risk models to estimate risk components for all of their obligors. Among the risk components, PD estimation is the most fundamental one which directly influences the portfolio's quality. With the improvement of PD prediction, financial institutions will more accurately separate 'good' firms from 'bad' ones. This helps financial institutions to build a healthier portfolio with lower PD. This research focuses on the SME 'bad' rate and the proposed methods that aim to improve PD prediction. Therefore, this research would be interested for practitioners to improve their IRB approach. As the Basel III proposals will be fully implemented in 2019, Basel II is still the most widely used

---

international banking regulation at present. The IRB-approach proposed by Basel II has guided banks to adopt innovative financial models to enhance their credit risk management. However, the IRB-approach does not require financial institutions to adjust their models and the corresponding required capital according to different business cycles. This pro-cycle feature of Basel II becomes an obvious disadvantage especially after the financial crisis. Further, different adoption of the IRB-approach might have provide financial institutions a high level of flexibility, which causes comparison difficulty and raises concern of incorrect implementation. Therefore, new accord and detailed guide of IRB-approach's implementation becomes in demand.

### ***Basel III***

Reacting to the global financial crisis and solving new issues appearing in the banking sector, the Basel Committee updates their regulation framework and publishes Basel III in 2010. The principles of the new accord are to prepare financial institutions for incidental events such as a financial crisis, to keep their risk level disclosed sufficiently, to prevent them from risking dangerously high leverage ratios, to correct the pro-cyclical feature of Basel II and to guarantee that they have sufficient capital during recessions. In doing so, the new Basel accord addresses the following issues (BCBS, 2015):

1. The minimum common equity capital ratio has been increased to 4.5%.
2. Financial institutions are required to maintain a capital conservation buffer above 2.5% and the capital would be used to deal with financial distress if the

---

institutions' capital ratio falls below 7% and helps the institution to recover from distress.

3. The minimum Tier I capital ratio will be raised to 6%.
4. Banks are required to have a counter-cyclical capital buffer between 0 to 2.5%.  
The new countercyclical buffer is introduced to correct the pro-cycle feature of Basel II. Financial institutions are required to keep the countercyclical capital buffer above 2.5% when they are issuing excess credit. This buffer is aimed to cover exposures that institutions may face if the bubbles caused by aggregate credit supply break and cause financial distress.
5. A leverage ratio is introduced and set as 3%. This ratio is defined as the ratio between capital and total assets.
6. Proposed the *global liquidity standard and supervisory monitoring* framework. It requires banks to maintain a sufficiently high liquidity coverage ratio to pass supervisors' stress test. Another required ratio is the *net stable funding ratio* which demonstrates banks' ability to cover mismatch issues.
7. Other changes including extensive financial instruments risk management, financial process monitoring, counterparty and central counterparty risk management and off-balance sheet risk exposures.

Basel III is a reaction to the financial crisis which could be partly blamed on aggregated credit issued. It gives a closer monitoring of the institutions' loan

---

performances and sets higher capital ratio, that would make it more and more difficult to access credit for enterprises. SMEs, which have never enjoyed the extra credit and in no sense caused the crisis, will be the sector potentially affected the most by Basel III according to the Association of Chartered Certified Accountants (ACCA, 2010). SME loans usually have a higher default rate than that of consumer and large firms (Cardone-Riportella, Trujillo-Ponce and Briozzo, 2013), and so the increased capital requirement would further reduce banks' lending to SMEs (Samon Hills, 2010). The Institution of International Finance (IIF) points out that Basel III could significantly reduce both short term and long term financial support for SMEs (IIF, 2010). The ACCA also mentioned that Basel III's influence on SMEs is difficult to predict before the full Basel Accord III becomes active, yet it is obvious that the SME credit scoring model should be improved to involve business cycles and to become more transparent to supervisors to meet the requirement of the new Basel Accord. In addition, the Bank of England (BoE) also requires a higher transparency of models employed by financial institution which allowing supervisors or stakeholders to check the impairment of Basel Accords. Semi-parametric models which proposed in this research will improve model transparency and clearly present how portfolios are affected by different factors.

**ii) IFRS9**

The international accounting standards board (IASB) proposed the new international financial reporting standards, IFRS9, in 2014. They will become active on 1<sup>st</sup> January 2018. Drafted in the financial crisis, the IFRS9 introduces a forward-looking

---

accounting standard and requires financial institutions to evaluate their expected losses on a much longer time period, including the past, present and future. As the panel data model proposed in this research analysed SME performance on a longer time horizon, it is also suitable for financial institutions to use under the IFRS9 standards. However, as the default definition required in IFRS9 is different from flags provided in the employed dataset, the researcher could not directly build an IFRS9 model.

In addition, IFRS9 requires the Lifetime Expect Loss (LEL). Therefore, obligor's PD should be forward looking and instead of one PD estimated under the IRB, IFRS9 model should provide obligor's PD through its entire life. Macroeconomic influence should be considered for lifetime PD due to the forward-looking features of IFRS9. Therefore, Through-The-Cycle (TTC) models should be used as TTC models is that they consider the economic cycle (Altman and Rijken, 2006). Panel data models analysed in this thesis is used with macroeconomic variables and could be regarded as a TTC model. TTC models will results in more smooth credit risk forecast and there is a trade-off between the estimation stability and the forecasting accuracy (Kiff, Kisser, and Schumacher, 2013).

When calculating LEL, the PD, EAD and LGD should all be provided in a time series format. That means although default may only occur once under the current standards, under IFRS9 definition an obligor could have multiple default through life time. And LEL is calculated by adding the obligor's possible lose at each upcoming month. This is the major difference between IFRS9 and Basel calculation. This concept is used to

---

require timely update of losses. In summary, IFRS9 challenges PIT models and require the consideration of business cycle. The panel data model used in this research could overcome the disadvantage of PIT models.

In conclusion, since the introduction of Basel I, BCBS have proposed standards for credit risk frameworks and made revolutionary changes to the banking industry. Especially after the release of Basel II, credit risk models become essential for financial institutions and credit scoring models have been developed to meet supervisors' requirements. Basel III addressed the importance of the business cycle, however, current SME credit scoring models are usually single response-level models, which could not reflect the business cycle. Hence, developing SME credit scoring models to involve the business cycle would not only could improve the model fitting, but also suits the principle of the Basel III accord. In addition, the Bank of England (BoE) requires financial institutions to provide their long-run averaged PD in their IRB-approach with consideration of different business cycles (BoE, 2015). A panel data model is ideal to tackle this issue. Hence, the panel data model with MVs proposed in this thesis becomes ideal to meet this requirement since it gives a close analysis of how to incorporate economic conditions into SME credit scoring. In addition, the semi-parametric models not only could introduce non-parametric effects and improve model fitting, but they could also give more transparency to supervisors and stakeholders.



---

The researcher has reviewed the regulations in credit risk management, which provide the regulatory background of SME credit risk management. Quantitative credit risk modelling becomes essential to financial institutions according to those regulations since the introduction of Basel II. The new regulations, such as IFRS9 and Basel III, has influence the credit risk modellers to pay more attention to the business cycles in SME modelling and take account of multi-period credit risk models. The improvement of credit scoring models will significantly influence the financial institutions as efficient scorecards could reduce default rate of their portfolio, decrease capital requirements, enhance their credit risk management and win regulators' approval.

The regulatory requirements are especially challenging for SME credit risk modelling due to certain unique features of SMEs, such as information opacity and the lack of robust data. Therefore, the researcher is very fortunate to carry on this research by employing a very large dataset of UK SME data. The methods proposed in this research will improve SME credit risk modelling by considering business cycle and explaining other default features of SMEs. Therefore, this research provides alternative methods for financial institutions to update their SME credit risk model to suit regulation requirements.

---

## 2.4 Credit risk modelling

Due to the different types of obligors and features of loans, there are two main streams of credit risk models. One type is represented by market based models, which are used for large corporations' credit risk. The other type is represented by scorecard models. As this research focuses on SME loans which have a high number of applications and each loan size is relatively small, especially due to the special treatment of SMEs in Basel II, more SMEs are treated as retail obligors and scorecard models are widely used for the SME segment. The majority of techniques used for scorecard modelling are operational research algorithms and statistic models. Previously, Lin (2007) and Ma (2009) have discussed these models in detail. This section will give a brief review of commonly used techniques.

Multivariate discrimination analysis (MDA) is one of the first models used to predict company default rates. Back in 1966, discrimination analysis was proposed by Beaver to predict company failure. As this was the earliest research that addressed the importance of financial ratios in failure prediction, it is usually regarded as the original, accounting-based credit risk model (Beaver, 1966). A major disadvantage of Beaver's research is that it only covers Moody's rated large US companies. Altman developed the traditional ratio analysis by using an MDA with 22 different ratios (Altman, 1968). This bankruptcy prediction model is the well-known 'Z-score' model. In Altman's research, firms of different sizes were selected, however, only a small fraction of them were small firms. One model is built for firms with different

---

size. Edminster (1972) applied similar analysis to an SME sample. This study is usually treated as the beginning of SME credit risk modelling. Edminster concluded that for SMEs credit risk modelling was distinct from that of larger firms. Unlike larger corporations which only demanded one financial statement, it required three (or more) continuous records to predict SME credit risk. Despite the early contributions MDA made in credit scoring field, MDA lost its popularity when new models were introduced. The reasons are that MDA assumed data follow a normal distribution and the same variance-covariance metrics is used for both 'good' and 'bad' firms; besides, MDA could not directly provide default probability (Lin, 2007).

Santomero and Vinso (1977) are commonly regarded as the beginning of using logit link functions in credit risk analysis (Ohlson, 1980). The logistic distribution fits the log odd ratio used in credit scoring. Instant of assuming a multinomial distribution, theoretically, logistic regression only assumes the probability is linked with a logit function. In the meantime, besides the link function, logistic regression retains the variables in a linear additive form (Keasey, McGuinness and Short, 1990). Its simple form can lead to clear implementations which are in favour of supervisors and stakeholders. Therefore, logistic regression has dominated scorecard building in recent years (Laitinen and Laitinen, 2000).

Besides statistical models, machine learning and other operational research algorithms are another type of technology often used in this area. Pioneers are Von Stein and Ziegler (1984), Anderson and Rosenfeld (1988), Kim and Scott (1991). These models

---

are shown to have better prediction accuracy by Coats and Fant (1993), Mahlhotra and Malhotra (2003) and Podding (1994). However, their disadvantages are also clear: firstly, the algorithm could take time to converge especially if the group difference was small and in some cases it may not converge; secondly, it could easily result in overfitting; thirdly, the procedure could not provide a clear explanation of decisions. The black box nature of machine learning models raises concern from both supervisors and regulators. In addition, the relevant customers question financial institutions' refusal of loan applications since there is not enough transparency of loan issuing.

A more recent development in credit scoring is survival models. Widely used in medical research, survival models can predict the probability of an event's occurrence at a certain time. Hence, when using this type of model in credit scoring, not only the PD of the obligors could be given, but also the failure time could be predicted. This feature is very attractive especially for profitability score card models (Crook and Belotti, 2010). Narain (1992) used survival analysis to build credit scoring models and clarifies the default time as well (Narain, 1992). This method has been further developed by Banasik, Crook and Thomas (1999), Stepanova and Thomas (2002), and Andreeva, Ansell and Crook (2005, 2007). Ju, Jeon and Sohn (2015) applied survival models to 4566 Korean SMEs with their monthly data. However, SME data received for this research is collected on an annual basis and only covers four years. Due to the feature of SME data, most SMEs update their account on a less frequent basis, which

---

is discussed in more details in the data description chapter. Therefore, current SME data contains large amounts of observation but uses a short observation time period. Besides SMEs dropping out due to their default, there can be other reasons for censoring. For example a firm may voluntarily close due to nonfinancial reasons, such as illness of entrepreneurs. The incomplete information of censored SMEs causes difficulty in applying survival analysis, which usually treats all observations that dropped out of sample as censored data.

Another stream to be mentioned is that of market models, which are well-known for setting listed companies' credit migration matrices. Those models are developed from the Merton model, in which the default events are replicated by a European call option on a firm's stock (Merton, 1973). The idea is that when a firm's liability is due, the listed company will default if their asset value falls below that of their liabilities, otherwise it would be able to survive. Besides their wide use in corporate credit risk analysis, several studies have used them in the retail side, Rosch and schedule (2004). In addition, Lin (2007) and Ma (2009) have applied Merton type models to SMEs. The Merton model can provide perfect theoretical explanations of default, yet it has several major disadvantages: first, it assumes that default can only appear at the maturity of debt; second, company debt is a single zero coupon bond; third, the market must follow the efficient market theory that all information has been included in the share price; fourth, as the Brownian motion follows the multi-dimension normal distribution, company equity is assumed to have a normal density; fifth, it assumes

---

any bankruptcy decision will be made as soon as the firm misses one payment. An empirical extension of the Merton model was developed by Kealhofer, McQuown and Vasicek which is commonly known as the KMV model. The detail of the original model is not fully published due to commercial reasons, however, its mechanism has been clearly explained by many scholars. KMV's extension overcame several disadvantages of the Merton model by providing numerical estimation of more complex capital structures than those assumed by the Merton model. Other models, such as credit risk plus model established by Credit Suisse, are also used to provide credit ratings for large corporations. Merton type models are applied by Lin (2007) to a small sample of UK listed SMEs. After a comprehensive comparison of Merton model and scoring-card models, Lin finds that the Merton model is not suitable for longer time period predictions and high default rate portfolios (Lin, 2007). Merton type models are usually developed by efficient market theory under which it is assumed that the listed companies asset value could be fully represented by their share prices. For SMEs, only a proportion of them have been listed on the stock market whose shares are usually traded so infrequently that efficient market theory is not always applicable. Therefore Merton type models face restrictions when applied to a large SME portfolio. In summary, this section provides a general review of different streams of credit risk models. For retail loans, the most widely used model for PD estimation is logistic regression, which is a statistic model. Logistic models could provide direct estimation of obligor's PD, give clear explanation of rejections and do not have high demand on data and hardware. This research seeks to develop the

---

statistic models in several different ways. Firstly, logistic models' performance is tested by SME crisis data. Secondly, the logistic model is extended to a panel model which considers the shifts of macroeconomic economic conditions. Thirdly, extend logistic models by adding non-parametric smoothers which captures irregular trends of SMEs during the 'credit crunch'.

## **2.5 Literature review for panel data model**

Panel data is a fundamental method in applied econometrics; its early application goes back to the 1960's. Both Mundlak and, then, Hoch used panel models to analyse a firm's production function (Mundlak, 1961; Hoch, 1962). Since then, there is an intensive development both in panel data's theoretical framework and its applications.

For credit risk-relevant research, panel data models are also the basis methodology for a wide range of macroeconomic topics such as sovereign risk and corporate bond default modelling. Back in 1977, Feder and Just (1977) applied binary response logit model to study country debt. More recently, panel data has been used to explore European government bonds by Lemmen and Goodhart (Lemmen and Goodhart, 1999). Their data covers the period 1987-1996 while a time dummy is used for the 1994's crisis. Das and Ghosh use panel data analysis to emphasise that macroeconomic significantly influences banks' risk level in emerging markets (Das and Ghosh, 2007). More recent research are focused on understanding the influence of past global financial crisis. Hassan and Wu prove that sovereign rating changes

---

interact with the volatility of GDP growth rate in a negative pattern by using instrumental variable methods. Combined with the theory that lower volatility of GDP growth rate would improve real GDP made by the Ramey and Ramey in 1995, their research shows the sovereign ratings have a significant indirect influence on GDP growth rate (Hassan and Wu, 2015). A large group of researchers have used panel data models to analyse whether there is a credit reduction or credit rotation during the last global financial crisis. A German companies' survey data is analysed by panel data models to develop indicators of credit supply constraints (Rottmann and Wollmersha, 2013). GMM is used in risk research by Sangalli to show that the inventory decumulation becomes a less preferred internal finance method for Italian manufacturing firms compared to what happened in previous recessions, although bank lending is believed to be significantly constrained during recession times (Sangalli, 2013). GMM is also applied by Tajik, Aliakbari, Ghalia and Kaffash to test the relationship between housing prices and non-performing loans, especially during the last financial crisis, in the US market at the bank level (Tajik, Aliakbari, Ghalia and Kaffash, 2015).

However, in the area of credit scoring, there is less application. Panel data models with probit link functions are used by Valles to evaluate the reliability of a through-the-cycle model. Focusing on corporate loans in an emerging market; the influence of GDP growth rate and unemployment rate are tested in this research (Valles, 2006). Later, Crook and Bellotti adopt panel models to consumer loans. Based on monthly



---

updated consumer loans, they mainly focus on survival analysis with Cox's proportional hazard function. Not only is a wider range of macroeconomic variables analysed, but also they consider the interactive effect of macroeconomic variables with obligors (Crook and Bellotti, 2010).

When focusing on SMEs, the use of panel data models becomes even less. Although lots of authors use data collected overtime, such as Altman et. al. (2007) and Gama et. al. (2012), the model they use is still cross sectional models without consideration of the time series effect. Probit panel data model is used by Fidrmuc and Hainz (2009) for a Slovakia SMEs sample. Their data is collected from one commercial bank and contains 667 SMEs' records from 2000-2005. Their research only covers a few industry sectors due to the small sample size they used. Another disadvantage of their research is that time-varying-only variables, such as MVs, have not been included. Valles (2006) is another example of using the probit panel data model whose research is based on Argentina's SMEs sample, containing around 6000 SMEs records from 2000-2004 during which a major crisis occurred in Argentina. According to Valles, there are two reasons that draw back the use of panel data model: One, is that collection data in panel format is more expensive; secondly, there is a misunderstanding that only current information affects firms' performance significantly; panel data model values historical information as well. Valles points out that historical information also plays an important role in default prediction since the panel data model actually provides a better prediction.

---

This research uses logit panel data model which is an extension of the widely used logistic regression. Comparing to probit model, which assumes predicted variable follows normal a distribution, logit model assumes that the underlying distribution is the cumulative logistic distribution which has been well adopted as the most appropriate one for credit scoring. Second, the UK SMEs is a large segment of obligor whose performance during the crisis needs further exploration. This research provides a guide on how to build scoring models by considering the significant macroeconomic swifts.

## **2.6 Literature review for GAM**

Another model employed in this research is Generalized Additive Model (GAM). McCullagh and Nelder introduced additive models which add non-parametric smoothers besides linear components. In this way, an additive model avoids assuming a linear structure between independent variables and a dependent variable, and allows to capture information that could not be addressed in traditional regression models (McCullagh and Nelder, 1989). GAM is a development of additive models which gives more flexibility on link functions. In various application fields, GAM suits the empirical distribution due to its advantage of using alternative link functions (Hastie, 2001).

Due to the flexibility and simplicity of GAM, it has been used in a wide range of research, such as Animal Ecology (Pereira and Itami, 1991), electricity recovery

---

ability after hurricane hits (Han, Guikema, Quiring, 2009), patients readmission (Demir, 2014), and others. The non-parametric feature of GAM allows it to provide an unrestricted estimation comparing to other more specified regression models in most applications where sample size is sufficiently large. However, the most attractive feature of GAM, argued by Demir, is that GAM provides more transparency about variables' influence, indicating which type of observations are worth investigating further (Demir, 2014).

In credit risk management, GAM is not a commonly used method yet. To the best of the authors' knowledge, Burkhard and Giorgi's research on Swiss residential mortgage's default density should be the first one using GAM in a credit risk area (Burkhard and Girogi, 2004). They find that GAM is comparable to reduced form models such as *CreditRisk<sup>+</sup>* and GAM captures independent variables' non-linear trend to improve credit risk management. Later, Berg applies GAM in 2007 to predict Norwegian firms. By comparison, GAM shows a distinct improvement in default prediction (Berg, 2007). However, Berg's priority is to improve prediction accuracy by obtaining the non-linear effects. Hence, this research fails to demonstrate an independent variables' trend. Recently, GAM with Gamma distribution is adopted by Tong, Mues, Brown and Thomas to estimate credit cards' EAD which has floating exposures, especially around the default time point. Their research shows that when little is known about a credit parameter's property, the credit parameter is a credit

---

card's EAD in this case, the GAM model out-performs due to their flexibility (Tong, Mues, Brown and Thomas, 2016).

For SMEs credit risk, GAM has rarely been used. It means, previously in SMEs credit risk, statistic models are mainly regressions which could ignore the non-linearity feature of independent variables and could not demonstrate their trend directly.

## **2.7 Summary**

Although SMEs are an important fraction of the economy, their credit risk rating systems have been challenged by the latest global financial crisis. To highlight issues raised since the last financial crisis, new regulations, such as Basel III and IFRS9, require financial institutions to include the business cycle into SME modelling and leads to a preference for multi-period credit risk models. These regulatory requirements are especially challenging for SMEs since information opacity allows limited analysis of SME credit risk. As the most fundamental risk parameter, SME PD estimation is still dominated by logistic regression. Therefore, this research would be especially interesting since it not only employs a huge dataset of UK SME but also the proposed models can reflect the influence of business cycles and explain features of defaulted SMEs. These alternative methods, such as panel data models and GAM, are ideal for financial institutions to use to meet new regulatory requirements and improve their portfolio performance.

---

## 3. Methodology

### 3.1 Introduction

The ‘credit crunch’ highlighted problems in current credit scoring systems which call for the development of SMEs credit risk methods to consider the changing business cycle and SMEs performance features in extreme situations. This chapter explains methodologies employed and how to explore UK SMEs credit risk during the financial crisis.

The researcher first explains the data preparation methods. The data describes a range of customer features which originally contains 79 independent variables. Several issues remain with the original dataset such as high collinearity and a high percentage of missing values. This chapter discusses methods of cleaning the original data for credit risk models which contain two steps: Firstly, variables are coarsely classified into categorical variables; then, each category is represented by its weight of evidence (Anderson, 2007). Obviously, even with a large set of data, there are dangers of over-fitting, hence, there needs to be an appropriate approach to select the ‘best’ variables for prediction from the original variable set. Therefore, variable selection is an essential step before building models which is shown in section 3.5.

In order to capture the features of SMEs during the ‘credit crunch’, a series of models are used in this research. Firstly, the logistic model is the benchmark, the second model is the panel data models and then, the generalised additive model (GAM) is

---

employed. Panel data allows consideration of the influence of business cycles (Valles, 2006). To do this, the researcher explores macroeconomic variables (MVs) that can be used in modelling the panel data. To explore the non-linearity in regressors, the research proposes an alternative method of processing missing values (Berg, 2007). After employing this approach, generalised additive models used this method in credit scoring SMEs.

This chapter is organised as the following: section 3.2 and 3.3 shows data transformation procedure and how to select variables from original dataset respectively; section 3.4 reviews employed models including logistic regression, panel data models and generalized additive models. In section 3.4, issues raised by using panel data models and generalized additive models are also discussed such as adding of macroeconomic variables and the use data's original format.

## **3.2 Data transformation**

This research is based on an empirical dataset which contains 'dirty' original data. Original data could not be used directly as firstly, missing values is such a common phenomenon in SME's data that one cannot ignore this category. Secondly, one cannot simply assume independent variables are related with dependent variables in a linear pattern. Third, variables have other formats besides numerical form. For example, although accounting relevant variables are usually presented in numerical form, firms' qualitative information, such as legal form, are recorded in a character

---

format. This section presents how to prepare original data to build credit risk models. Coarse classification is a commonly used method to transfer all variables into numerical form. Combined with Weights of Evidence, transferred data avoids issues mentioned above. This section explains the data perpetration process in detail.

### **3.2.1 Coarse classification**

Explanatory variables cover a wide range of data which could be nominal, ordinal or ratio. Therefore, some variables could have a very large range, such as industry classification due to SMEs diversity. On the contrary, the nominal or ordinal type of variables, such as legal form, tend to fall into a very limited range. Another feature of SMEs data is the sizeable number of missing values. For certain variables, up to 90% of the data is missing.

To solve those issue, variables are usually classified into categories in credit scoring. Classification covers two stages, which are coarse classification and fine classification:

1. Initially, variables are divided into small intervals. Each interval's performance is calculated. Categorical performance is usually presented by Weight of Evidence which will be discussed in detail in the following section.
2. Intervals with similar performance are combined together. Usually, though, the missing category is kept separated. Finally, around ten categories are chosen for each variable.

---

Although this method is fairly straightforward, given the large number of explanatory variables, 79, this procedure becomes cumbersome. To allow comparison of differences, the same classification is used for four years.

### **3.2.2 Weights of Evidence**

Coarse classification transfers all variables to categorical form and gives each category numerical value by its order. For variables with missing values, coarse classification classifies the missing value as one category. Therefore, coarse classification solves two issues of original data: it converts all variables into numerical format; instead of being absent, missing values is also treated as a separate category; thirdly, it copes with outliers. Then, the remaining issue is that those classified categories are not necessarily linearly correlated with dependent variables, especially considering there are missing categories among them. Often, dummy variables will then be formed to solve the linearity issue. However, dummy variables are not the appropriate method for panel data models. For panel data, the same set of variables would be used throughout time, however, significant dummies can vary from year to year. As this research aims at involving time variate differences of UK SMEs performance, using dummy variables could lead to two possible results: either, not enough dummy variables are selected to reflect the time variations; or, a large number of dummies, which may be significant in different years, are considered which would make the estimation matrix fulfilled by zeros. However, if the estimation matrix is fulfilled by zeros, the time difference matrix would also be dominated by



---

zero. This will significantly reduce the efficiency of panel data estimators (Cameron and Trivedi, 2010). Therefore, dummy variables are not the preferred method for this research.

Another way of avoiding the linearity assumption is to re-evaluate each category by its Weights of Evidence (WoE). Unlike dummy variables which only keeps the significant category of independent variables, WoE enters the entire information of the independent variable. Therefore, it not only reduces the number of entered variables but also maintains more information about SMEs changes during the financial crisis.

WoE is calculated by the following formula:

$$\text{WoE} = \ln\left(\frac{N_i}{P_i}\right) - \ln\left(\frac{\sum N_i}{\sum P_i}\right)$$

where  $P_i$  represents the number of ‘good’ customers in category I,  $N_i$  is the number of ‘bad’. WoE shows the difference between each category’s good-bad ratio and that of the whole sample. A higher value of the WoE means the category has a higher percentage of ‘good’ customers comparing to the whole sample. Hence, WoE decreases in a monotonous pattern with a corresponding category’s ‘bad’ rate. And, if the dependent variable focusses on ‘good’ applicants, then all the independent variables should be positively correlated with dependent variables.

---

### 3.3 Variable Selection

The initial data contains 79 variables which describe customers' general features, firms' board information, financial ratios, payment types and previous records. However, not every recorded variable has significant influence in identifying SMEs default events. In addition, if all of these variables were added into the model, one cannot avoid high collinearity amongst those regressors. One result that might arise is the estimated negative correlation between dependent variables and independent variables. For example, a logistic regression, which involves all the 79 variables, modelling the probability of being good based on 2007's non-start-up SMEs data, results in a negative coefficient for 10 of the variables. On the contrary, the use of WoE should only lead to positive coefficients.

Several reasons could cause correlation among significant independent variables, such as containing similar information and being dominated by one large category. These problems may be controlled by using coarse classification since different variables could be classified in different manners. For example, relevant categories may be combined with other categories, which could significantly reduce, not eliminate, collinearity. Hence, it is necessary to remove correlated variables.

To select significant independent variables and to solve the collinearity issue, this research chooses stepwise selection logistic regression, instead of forward or backward selection, to reduce the size of independent variables. In a stepwise

---

procedure, variables which have not been entered in previous steps will be entered if they have a high chi-square value. Meanwhile, existing variables with lower chi-square values would be removed. There are three stepwise selection methods considering the entering order of variables: forward, backward and no ordering which is the case of stepwise selection. In either a forward or backward selection procedure, variables are entered into the model in a chosen order. However, in this research, there is no preference of independent variables, this research selects variables without considering its entering order. When newly added variables cannot achieve a 'better' model, the stepwise procedure ends. Stepwise logistic models are preferred for several reasons. First of all, stepwise logistic regression will only select significant independent variables. Hence, it can reduce the set of independent variables efficiently. Second, variables are selected into model regardless of its order. Third, correlations within independent variables would cause unexpected coefficient signs; those variables would also be ruled out to further reduce collinearity (Littell, 1996).

### **3.4 Relevant Models**

The 'credit crunch' challenges the credit risk systems, especially on the rating of SMEs. It remains untold when 'black swan' event occurs, whether the current system is robust, how the macroeconomic shifts influence SMEs 'bad' rate, and how independent variables influence SMEs 'bad' rate. Different models need to be built to answer those questions. Those models are analysed in the following chapter which are: logistic regression, panel data models and GAM. The last section explains

---

separation measures used to demonstrate a models' ability to separate 'good' from 'bad'.

### 3.4.1 Logistic Regression

As discussed in the literature review, logistic regression is a classic credit scoring model and is the present industry standard. Logistic regression dominates the industry partially due to its simplicity. It is easy to understand, robust in a variety of empirical circumstances, gives a clear answer to rejections, and imposes lower requirements in computation. Since the default event is a binary event, discrete logistic regression is used here as a benchmark model to judge whether the introduced methodologies can improve PD estimation.

If  $y_i$  is the dependent variable taking 1 ('good') and 0 ('bad') then:

$$Pr(y_i = 1|X_i) = \frac{\exp(BX_i)}{1 + \exp(BX_i)} \quad (1)$$

or

$$Pr(y_i = 1|X_i) = \frac{1}{1 + \exp(BX_i)^{-1}} \quad (2)$$

where  $X_i = \{1, x_1, x_2, \dots, x_j\}$  is the  $j + 1$  dimensional vector represent  $j$  explanatory variable,  $B = \{\beta_0, \beta_1, \beta_2, \dots, \beta_j\}$  represents  $j + 1$  dimensional coefficients vector where  $\beta_0$  is the intercept.

---

Then, the linear combination of explanatory variables could be expressed by conditional default probability in the following form:

$$BX_i = \log \frac{Pr(y_i = 1|X_i)}{1 \pm Pr(y_i = 1|X_i)} \quad (3)$$

For logistic regression models, the obligor's probability of default can be predicted by the logistic transformation of selected variables. This transformation is chosen as it can limit our prediction within [0, 1] interval which fits the nature of default probability.

In this research, SMEs probability of default is estimated by the firms' general information, financial ratios, previous credit history and other relevant information. The model is used to explain the relationship between selected variables and the dependent variable. Coefficient  $\beta_i$  is the coefficient of  $x_i$  shows how much the dependent variable will be changed with a logistic transformation for per unit change of explanatory variable  $x_i$  given the collected information.

The logistic regression was initially applied to annually use a selected set of independent variables. This model provides the initial estimation of UK SMEs performance during the 'credit crunch' and provides the benchmark for further modelling.

---

### 3.4.2 Panel Data Models

Due to the impact of the ‘credit crunch’, the supervisors, governments, and financial institutions have become keen on developing more accurate credit risk models to control financial institutions’ exposure to default loss. Several disadvantages of the current risk system raise major concerns about its accuracy in separate ‘good’ and ‘bad’ customers, especially during economic downturns. First of all, the logistic regression, which dominates the financial industry, is a cross-sectional analysis which treats all data as if it’s at the same time and does not address the time series effects, except by regular recalibrations. The dynamics of credit risk caused by rapid economic change would not be considered in such a model. Therefore, as the aim of this research is to analyse SMEs performance during the crisis period, it is desirable to build a model to assist the analysis’s timely changes. Panel data is a natural development of cross-sectional analysis since it can involve both cross-sectional effects and in time series effects into the estimation. Panel data considers both the differences within each group and the differences between groups. Therefore, the data efficiency is significantly improved comparing to simple cross-sectional analysis. In order to add time effects, each obligor’s performance is tracked through ‘credit crunch’ which allows the consideration of macroeconomic changes from time to time. This improvement corrects the pro-cycle feature of the standard logistic model. Further, when using forecasted macroeconomic variables, the panel model could provide the future performance of loans under different economic conditions.

---

Therefore, panel data models can not only answer how the macroeconomic shifts during the ‘credit crunch’ influence the SMEs performance, but this type of model is also ideal to use to meet new regulation requirements, such as Basel III and IFRS9.

**i) Logit panel data model**

The panel data model allows the modelling of both cross-sectional and time series effects. There are various link functions that could be chosen to satisfy the data structure and specific circumstances of the research. The linear panel data model is the basic panel data model, while if the link function such as logit, probit, and hazard function are used, this model is usually classified as a nonlinear panel data model. As mentioned, credit scoring models are aimed at binary response dependent variables. This feature leads to the preference of logit panel data models which has a binary response and employs logistic distribution to identify default events during the ‘credit crunch’.

The logit panel data model is used to represent panel models which are based on logistic distribution with discrete responses (Wooldridge, 2010). Since panel data model considers both cross-sectional and time-series effects, the error structure is much more complex than regression models. As mentioned previously, modelling at default and denoting “good” as one, then the logit panel data model could be formed as follows:

$$\Pr(y_{it} = 1 | X_{it} = x_{it}) = \frac{\exp(\beta x_{it} + \alpha_i + \delta_t + u_{it})}{1 + \exp(\beta x_{it} + \alpha_i + \delta_t + u_{it})} \quad (4)$$

---

here  $x_{it}$  represents explanatory variables of observation  $i$  at time  $t$ ,  $\beta$  is the coefficient, and  $\alpha_i$  is the time-invariant variable showing the specification effect, the time effect  $\delta_t$  does not yield throughout time for the same obligor and residual is denoted as  $u_{it}$ . This expression is very much like the logistic regression model except this one uses double subscription and two more coefficients: time-only related effect and firm-only effect. This is the two-way panel data model as both the firm specific and time only effects are considered. However, as this research plans to explain the time only effects by macroeconomic variables (MVs), only the firm specific effect is involved. When only one effect is used, the model becomes a one way model. In addition, two way model will bring difficulty in estimation. The one way model with firm specific effect is denoted as:

$$\Pr(y_{it} = 1 | X_{it} = x_{it}) = \frac{\exp(\beta x_{it} + c_i)}{1 + \exp(\beta x_{it} + c_i)} \quad (5)$$

The impact of time-series introduces a firm-specific error. It is an important issue in panel data modelling whether to treat the unobserved effect as a fixed coefficient (FE) or as a random variable (RE). This section will discuss the logit panel data model and the selection of estimators. The discussion between fixed and random effects in the nonlinear model is believed more important than in a linear model since this choice will future influence the analysis of data and lead to significant differences in estimation results (Baltagi, 2005).



---

There is a long and continuous discussion about the choice between FE model and RE model among economists. Firstly, the choice between FE and RE depending on meaning of the error term in the content of the research and will reflect whether the researcher believes the firm-specific effect is constant. If FE is used, it assumes that firms have an unobserved effect that does not change along with economic conditions and other aspects that change over time. The modelling background leads to the choice of FE rather than RE. However, a constant firm-specific effect is a very strong assumption which is problematic in this research as it focusses on the firms' changes during the financial crisis. The firms' unobserved effects could be explained as their relationship with banks, their supply chains, and other unobserved information which influences the default event. During the financial crisis, unobserved information would change dramatically. In conclusion, the content of credit risk does not lead to the preference of FE.

The following content of this section further explains the researcher's choice between FE and RE. The analysis is based on maximum likelihood methods which are commonly used to solve logit panel data models (Wooldridge, 2010). Chamberlain's conditional maximum likelihood (CML) estimation is a widely used solution for the estimation of FE, while RE uses marginal likelihood.

**ii) The incidental parameter problem and CML for FE**

In panel data models, when a new observation is added to the sample, an additional coefficient  $\alpha_i$  is needed as well. Therefore, FE estimators will solve an unmanageable

---

set of coefficient. This is called the incidental parameter problem (Neyman and Scott, 1948). To avoid this problem, conditional likelihood is usually used for FE models (Lancaster, 2000). However, the algorithm of conditional likelihood would delete observations with no change to its dependent variable. The reason for deletion is explained as following. For customer  $i$  ( $i = 1, 2, \dots, n$ ), conditional probability is:

$$\Pr\left(y_i \mid \sum_{t=1}^{T_i} y_{it} = k_{1i}\right) = \frac{e^{\sum_{t=1}^{T_i} y_{it} x_{it} \beta}}{\sum_{d_i \in S_i} e^{\sum_{t=1}^{T_i} y_{it} x_{it} \beta}} =: f_i(T_i, k_{1i}) \quad (6)$$

In the above function,  $\sum_{t=1}^{T_i} d_{it} = k_{1i}$ , and  $k_{1i}$  is the sum of dependent variables which are binary in this research. So  $k_{1i}$  is also the total number of ones. Define  $k_{2i}$  as a total number of zeroes and  $S_i$  contain combinations of  $k_{1i}$  and  $k_{2i}$ . Therefore, the numerator depends on combinations of different values of independent variables through time. This problem is especially important for a short time period (Heckman, 1987; Hamerle and Ronning, 1995). Then conditional likelihood becomes:

$$\ln L = \sum_{i=1}^n \left\{ \sum_{t=1}^{T_i} y_{it} x_{it} \beta - \log f_i(T_i, k_{1i}) \right\} \quad (7)$$

For observations that do not change state, the conditional likelihood loses its efficiency as  $S_i$  has only one element and  $f_i(T_i, k_{1i})$  will always equal to one. Hence, observations without change of default status are removed out during estimation.

---

However, the majority of SMEs remain in the ‘non-default’ (Good) state even during an economic downturn such as the past ‘credit crunch’. Therefore, if FE is used, only SMEs that fail during the period will be used for estimation. It will not only lead to the neglect of data but will lead to significant estimation bias. For example if FE is used to estimate the start-up’s segment, which has a higher default rate, FE will remove 170,354 observations and only use the remaining 33.9% for estimation, and variables lose their significance at 5% level in the FE model.

In addition as FE weighs more on the within group difference, which is one customer’s change through time, FE is not suitable for data which has small or insignificant changes. Since credit risk data usually use WoE or dummy variable to transfer raw data and the dependent variable, credit risk data lacks within group differences, especially when majority of the portfolio remains as ‘good’. The low change within each group causes a significant loss of data and a bias data selection when FE is employed.

Hence, observations without change of default status are removed out during estimation. However, the majority of SMEs remain in the ‘non-default’ (Good) state even during an economic downturn such as the past ‘credit crunch’. Therefore, if one of the models used is FE, only SMEs that fail during the period will be used for estimation. It will not only lead to the neglect of data but will lead to significant estimation bias. In the start-up’s segment, which has a higher default rate, FE will remove 170,354 observations and only use the remaining 33.9% for estimation, and

---

variables lose their significance at 5% level in the FE model. Therefore, FE is not suitable to build scorecards. Hausman's specification test (Hausman, 1978) is not used here to choose between RE and FE since FE and RE. The reasons are as following. Hausman assumes FE estimator is the consistent estimator, which usually is true, while RE is more efficient. If the difference of those two estimators is not statistically different, RE consists with the consistent estimator FE. Therefore, RE is the preferred estimator since it has a higher efficiency in analysing data. Additionally, if the Hausman test is used, one accepts that losing efficiency will not affect FE's consistency and will not cause a statistically significant bias. However, when it comes to the empirical selection of the logit model, the problem is more complex as FE is not always the consistent estimator. When FE is the biased estimator, the Hausman test could not be used to make a decision. Marginal likelihood for RE

Compared to FE's disadvantages, RE is more suitable for this research. In the past 'credit crunch', an unobserved random firm-specific effect is assumed; SMEs both with 'good' and 'bad' stature are considered in estimation and a small variation of explanatory variables does not lead to a strong decrease of significance. As both FE and RE estimators are consistent at  $\sqrt{N}$  level where  $N$  is the size of data used in estimation. As FE only obtained a fraction of data comparing to the RE estimator, they are not comparable and using the Hausman test to check RE's consistency is not necessary and can be misleading given the size (Cameron and Trivedi, 2009). For the above reasons, the RE model is chosen for analysis and a marginal likelihood is

---

applied. Further, in RE models,  $\alpha_i$  is normally distributed. The marginal likelihood method integrates  $\alpha_i$  according to its distribution, therefore incidental problems are solved and data efficiency is improved significantly (Lancaster, 2000).

As  $\alpha_i$  is assumed to replicate unobserved firm-specific effects in our model. Since we analyse SME behaviours through the financial crisis, during which period macroeconomic policies and each lender is experiencing significant changes, it is reasonable to assume the SMEs relationships with their bank vary over time. That means relationships are random, independent and unobserved effects. This research employs a population-averaged method for RE logit panel models.

Here, assuming  $c_i \sim N(0, \sigma_c^2)$ , then

$$\Pr(y_{i1}, y_{i2}, \dots, y_{in} | x_{i1}, x_{i2}, \dots, x_{in}) = \int_{-\infty}^{\infty} \frac{e^{-c_i^2 / 2\sigma_c^2}}{\sqrt{2\pi\sigma_c}} \left\{ \prod_{t=i}^{n_i} F(y_{it}, x_{it}\beta + c_i) \right\} dc_i \quad (8)$$

where

$$F(y, z) = \begin{cases} \frac{1}{1 + e^{-z}}, & z \neq 0 \\ \frac{1}{1 + e^z}, & z = 0 \end{cases} \quad (9)$$

The panel level likelihood  $l_i$  follows

$$l_i = \int_{-\infty}^{\infty} \frac{e^{-c_i^2} / \sqrt{2\pi\sigma_c^2}}{\sqrt{2\pi\sigma_c^2}} \left\{ \prod_{t=i}^{n_i} F(y_{it}, x_{it}\beta + c_i) \right\} dc_i =: \int_{-\infty}^{\infty} g(y_{it}, x_{it}, c_i) dc_i$$

(10)

There is no closed form for the above function and quadratic form is usually used to estimate the above function. Therefore, RE avoids the incidental parameter problem by integrating the random effects with normal distribution.

#### **v) Adding Dummy and Macroeconomic Variables**

When using panel data models, there may exist a temporary dependence in time-series effects. It would lead to misleading results if the temporary dependence is inappropriately treated by ignorance (Mintrom, 1997). This issue has been discussed by Beck (1998) and adding time dummies has been found effective to overcome it. Considering in the observed time period, default rates as well as statistics of explanatory variables shift sharply from year to year. Therefore, year dummies should be added to control the sizable upwards and downwards business cycle from the crunch. However, the dummy variable has two main disadvantages. First, dummy variables cannot be explained. Hence, it would reduce the transparency of model and raise concerns from the supervisor, regulator and customer. Second, it is even more difficult to estimate dummies outside of the sample period. Hence, it is problematic to do stress testing or forecasting for the upcoming time period.

Macroeconomic variables (MVs) is the ideal alternative for dummies. Using MVs to replicate market movements can provide an understanding of how macroeconomic

---

conditions influence SMEs performance during the ‘credit crunch’ crisis. Aimed to explain the time variation of SMEs performance during the ‘credit crunch’, selected MVs should be reported in a consistent pattern during this period and most relative to SMEs performance in the UK. When selecting MVs, the other issue is the complex correlation structure amongst them and their interactions with firm-specific variables. To avoid noise caused by interaction, and bearing in mind that our aim is only to control for market movements, only major MVs are considered according to Figlewski’s *et. al.* (2012) framework, which described the direction of the economy, general economic condition, and the financial market movements.

The influence of MVs usually is believed to be latent, yet there is no clear rule for lag choice. As this research is about a very particular time, the guide about the choice of lag length is even more limited. In my dataset, SMEs suffered their highest default rate in 2009. In the same year, UK’s GDP growth rate reached its lowest record as well. Unlike relevant research, consumer and large corporation loans usually employ monthly update data; SMEs data is much less continuous on the time horizon. According to Figlewski’s framework, using the cumulative effect could contain all the past information and provide the most accurate market replication. As the MVs influence on SMEs is limited, this research joins the discussion by testing the influence of MVs with different lags. Those results provide insight of MVs influence and support for other research.

---

Similar to Figlewski et al's (2012) work, decalent factor  $\delta$  equals to 0.88. Considering the length of employed data, the time horizon is set as two years. That means one-year lagged MVs and two-year lagged MVs are added with giving higher weight to the most recent data:

$$MV_t = \frac{\sum_{p=0}^2 \delta^p MV_{t-p}}{\sum_{p=0}^2 \delta^p} \quad (11)$$

where  $MV_t = (mv_1, mv_2 \dots mv_s)_t$  and represents the MVs value at time  $t$ , and the decay factor  $\delta$  is set as 0.88. MVs at three different lags are considered, therefore  $p = 0, 1, 2$ , which is the current year's MV, MVs with one year's lag and MVs with two years' lag.

### 3.4.3 Residual check

In panel data models and the majority of other statistic models, normal distribution is widely used in credit risk analysis. Although logit panel model uses a logit link function, the following analysis would show that after transfer, normality is still assumed in estimation. However, the tail distribution of financially constrained events has been discussed for decades. The normal assumption for this research is particularly questionable since more vulnerable performances has been observed in SMEs during the observed period. The macroeconomic condition changes in such a dramatic pattern that SMEs' annual performance varies as well (Smallbone, Deakins, Battisti and Kitching, 2012). Hence, it is necessary to test whether the normal



---

assumption is valid for this research. The residual of panel models is explored and the estimated distribution of residuals compared to the normal distribution.

The residuals from the logit panel are obtained by the estimation as follows. Logit panel model has the following form:

$$\Pr(y_{it} = 1|X_{it} = x_{it}) = \frac{\exp(\beta x_{it} + c_i)}{1 + \exp(\beta x_{it} + c_i)} \quad (12)$$

where  $c_i$  is the time-invariant unobserved effect and set  $u_{it}$  as the residual. Denote estimated outcome as  $y_{it}'$  then

$$y_{it}' = \frac{\exp(\beta' x_{it} + c_i')}{1 + \exp(\beta' x_{it} + c_i')} \quad (13)$$

residual is

$$u_{it}' = y_{it} - y_{it}' \quad (14)$$

here  $u_{it}'$  follows logistic distribution while  $\alpha_i'$  is normally distributed, then

$$\ln\left(\frac{u_{it}'}{1 - u_{it}'}\right) \sim N(0, 1) \quad (15)$$

and

---


$$c'_i = \ln\left(\frac{y_{it} - u'_{it}}{1 - (y_{it} - u'_{it})}\right) - \beta' x_{it} \sim N(0, 1) \quad (16)$$

Therefore,

$$c'_i \sim N(0, 1) \quad (17)$$

And the following notation is used:

$$e_{it} = \ln(\mu'_{it}/(1 - \mu'_{it})) \quad (18)$$

$$rc_{it} = \ln((y_{it} - \mu'_{it})/(1 - (y_{it} - \mu'_{it}))) - \beta' x_{it} \quad (19)$$

By employing the above functions, residuals of logit panel models are expressed by empirical data and estimated statistics. The empirical distribution of residuals could be explored by those functions. In addition, the above models assume WoE is linearly correlated with the dependent variable and this assumption is widely used in credit scoring. If so, according to the assumptions of logit panel data model, those residuals should be normally distributed. However, considering the non-linear nature of risk event and the special time period this research focusses on, the standard parametric method may not be able to capture all the data features. The comparison of residuals and normal distribution will be discussed which provides the initial evidence of non-linearity. Graphical and descriptive statistics will be employed to give an initial understanding of the residuals.

---

### 3.4.4 Generalised additive model

#### i) Introduction

The residual check provides the initial evidence of the irregularity of SMEs' credit risk during a recession period that previous methods could not fully capture. However, a 'black swan' event as the 'credit crunch' is; too little information is known to choose the alternative parametric method. Hence, non-parametric estimations are used to fit the unexpected performance, provide a clear demonstration of SMEs performance during the 'credit crunch', and highlight the non-parametric nature of SMEs' credit risk during the recession period.

Therefore, it is worthwhile to employ a non-parametric approach to improve the default prediction. Smoothers become an interesting alternative for several reasons. First, a smoother is a non-parametric estimator used to fit the real trend of independent variables without assuming a prior structure of the trend, such as linearity in regression models. This non-parametric feature makes it a good tool to estimate irregular performance and provide a better model fitting. Secondly, as smoothers could plot the hidden trends clearly, it improves the transparency of credit risk models. The high transparency of such models not only suits the supervisors' perspective but also identifies the unknown trend of each independent variable during recession times. Therefore, this technique is beneficial for both improving prediction accuracy and increasing model transparency.

---

Among models which involve smoothers, the research chooses generalised additive models (GAM) since GAM is efficient on large datasets and could explore the association between explanatory variable and dependent variable.

## ii) GAM

For reasons discussed in the last section, GAM is employed to address the non-parametric nature of SMEs credit risk. The logistic link function is chosen since its interval falls between zero and one, and had been proven as the most appropriate one since Ohlson introduced it in 1980. Its form is:

$$\eta(y) = \frac{\exp(y)}{1 + \exp(y)} \quad (20)$$

Here  $y = s_0 + \sum_1^j g_i(X_i)$  and  $g_i(X_i)$  are estimated functions that could be either parametric or non-parametric, which means each variables' influence is determined by a parametric linear effect besides a non-parametric part. Yet, the non-parametric part can be insignificant. If variables do not exhibit a significant non-parametric effect,  $g_i(X_i)$  is equal to a linear parametric parameter. Meanwhile, for variables which have a significant smoothing component  $g_i(X_i)$  will refer to an additive function which is comprised of both a linear parametric part and a non-parametric part. If assuming the estimation function of  $X_i$  is linear:

$$g_i(X_i) = \beta_i X_i \quad (21)$$

---

The nature of GAM is to perform a smoother of explanatory variables towards a dependent variable without assuming a given pattern. Therefore, no distribution or function of  $g_i(X_i)$  is pre-determined on this step,  $g_i(X_i)$  is derived only as a smoothing function, usually by a backward fitting method. The second step of the algorithm is to separate  $g_i(X_i)$  into two parts: a functional linear part  $f_i(X_i)$ , where  $f_i(X_i) = \beta_i X_i$ ; as well as a non-parametric part  $s_i(X_i)$ . Then, adding the linear part and non-parametric part together, the influence of the explanatory variable is complete. In practice, not every variable enjoys significant effect from both parts.

The non-parametric effect will then be estimated by kernel functions. Assuming the distribution of the random effect  $\alpha_i$  is unknown and apply kernel distribution function:

$$\hat{f}_h(\alpha_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\alpha - \alpha_i}{h}\right) \quad (22)$$

Then obtain:

$$\Pr'(y_{i1}, \dots, y_{in_i} | x_{i1}, \dots, x_{in_i}) = \int_{-\infty}^{\infty} \hat{F}_h(\alpha_i) \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \alpha_i) \right\} d\alpha_i \quad (23)$$

Where  $\hat{F}_h(\alpha_i)$  is estimated by the kernel function. Previously, the residual check provides the initial evidence whether the model has fully captured the performance of SMEs. Yet, further analysis is required to demonstrate whether those ignored effects

---

could be explained by non-parametric smoothers. Hence, GAM is firstly applied to the same set of data used by logistic regression and the panel model to test whether the non-parametric effects are significant. Using the significant level 95%, variables with a significant non-parametric effect for three years or more are selected.

The data used here has been transferred by WoE for consistency and comparison. Although it helps to answer the question whether non-parametric effects are significant, GAM's advantage of increasing transparency could not be addressed. Therefore, the next section seeks to use data to demonstrate a variables' trend directly.

### **iii) GAM with original data**

As mentioned previously, GAM not only brings a non-parametric effect to increase model accuracy, it also improves model transparency which allows for the direct explanation of how independent variables are influencing SMEs credit risk. Mostly used parametric statistic models, such as survival and logistic regression, cannot directly incorporate original data since one cannot assume all the variables are linearly correlated with performance. GAM, however, can overcome this issue as non-parametric effects can be added as supplementary to the linear effect results from regression models. However, as our data is transferred with WoE, we have to combine the non-parametric movements with the trend of WoE to explain variables' influence. As powerful as WoE is in practice, its results only indicate the influence of transferred data and reduces GAM's explanatory power significantly. If one wants to dig a

---

particular variable's effect, original data is always the ideal choice. Next, we will discuss how to solve this issue.

However, the biggest challenge left for incorporating original value without WoE transformation is the existence of missing value in real data. WoE groups missing values and gives this absent group a real value according to this group's performance.

An alternative method should be introduced to process the missing group to avoid transferring original data.

### ***(1) Handling the missing values***

One difficulty of using original data is the large volume of missing value. This section provide an alternative method of proceeding missing value. In the missing value theories, the most traditional and widely used two methods are: deletion and single imputation (Baraldi and Enders, 2009). This research chooses a single imputation to solve the missing value issue. Due to the high frequency of missing value in SMEs data, the deletion will shrink the sample size and cause bias of the remaining sample since the missing group has proven informative in scorecard modelling. Although one cannot investigate the exact reason to why those have been failed to be reported, the fact that those variables are missing is informative enough in estimating the corresponding obligors' performance.

The single imputation is used because of the assumption that missing value does not occur at random. In the handling of the missing value, it is necessary to identify the mechanism of missing in the employed data. Generally speaking, there are three types

---

of missing values: 1) missing completely at random; 2) missing at random; 3) missing not at random (Enders, 2010). If the missing value appears completely at random, random numbering is a good method for fulfilling the gap. However, although it is almost impossible to answer why an SME fails to report some information to their credit supplier, missing values do not occur randomly in most cases and the missing value is either directly or indirectly correlated with the 'bad' performance of SMEs. In either way, the missing category is informative and could be used effectively for SMEs performance analysis. For example, if the SMEs total asset is missing, it is unlikely that such an important accounting ratio has been forgotten to be reported at random. It is more likely that the figure is not easy to be reported correctly or reporting the figure is negative for the firms' credit access. However, due to the large set of data, the real reason for each missing value is almost impossible to be recovered. Hence, this research assumes that the missing category is informative and brings the same information into scorecards.

The alternative method is proposed to fit the special feature of SMEs missing data in credit scoring. Traditionally used methods, such as dummy variable or WoE, treat the missing value as one category which has been proven successful in handling the lost information. Dummy variable is not appropriate since it could mix with the original value. To substitute the missing data with an observed value. However, the disadvantage of WoE is also clear. It will re-order the data and makes implementation of independent variable less direct.



---

This research chooses to match a missing category with observed values. If a missing category and a known value share the same performance, they bring the same information for estimation and the missing category could be fulfilled by this value without bringing bias to analysis. Further, a missing category's moving average (MA), which leads to a continuous performance measure of a dependent variable, is chosen to capture the performance of non-missing values. MA enjoys several advantages. First, MA keeps the original format of aimed variables without changing continuous variables into one category. It will lead to loss of information and cause bias if all variables are treated categorically. A value's moving average is calculated by taking the average of neighbourhood value's performance. Order the aimed variable's value by an increasing order, then the moving average of the  $i$ -th value  $a_j$  is defined as the average value of values fall into its neighbourhood, however, this research, the moving average of value's 'good' rate as  $g_i$ , is used :

$$MA_i = \frac{g_{i-n} \cdots + g_i + \cdots + g_{i+n}}{n} \quad (24)$$

Here  $MA_i$  represents the moving average of value  $i$ ,  $n$  is the number of value considered in the neighbourhood of value  $i$ ,  $n = 2000$ , And  $g_i$  is the 'good' rate for value  $i$

$$g_i = \frac{\text{number of 'good' customers which have value } a_i}{\text{total number of customers which have value } a_i} \quad (25)$$

---

In order to find the missing category's matching value, calculate the 'good' rate of missing category  $a_M$ :

$$MG = \frac{\text{number of 'good' customers with missing value}}{\text{total number of customers with missing value}} \quad (26)$$

Here  $MG$  is the 'good' rate of missing category.

If a given value  $j$  is found to have the same performance as the missing category:

$$MG = MA_j \quad (27)$$

Then this value provides the same information as the missing category during estimation. Now input the missing value with  $a_j$

Else if

$$MG \neq MA_i \text{ for } i = 1, 2, \dots, N \quad (28)$$

And  $N$  is the total number of observed values of the analyzed variable.

Then the missing category performs different from the rest of the data and one cannot explain their performance by any collected information. This variable should either been kept in WoE for accuracy or estimated  $a_j$

$$\text{if } MG - MA_j = \min |MG - MA_i| \text{ for } i = 1, 2, \dots, N \quad (29)$$

Although one may argue the missing group should be kept distinct, the aim of this method is to explore a variable's influence. Approximation allows for keeping

---

original data and remaining in more variation. Using this method, the missing category is imputed by an observed value if they have similar performance, which is measured by the corresponding 'good' rate. This method is proposed to fit the feature of SMEs credit scoring particularly. Single imputation is used due to the mechanism and volume size of missing value. It also allows one to keep the original value derive the clear marginal trend of explanatory variables. In addition, as an observed value is used to represent the missing category, it provides the insight that the missing category is close to which observed value. Therefore, this method is proposed to increase the transparency of analysis instant of bring more noise.

## **(2) Standardization**

Another disadvantage of using original data is that the variety of independent variables cause the significant scale difference among them. It will not only weaken our models' explanatory power as coefficients values could not directly explain the independent variables influence, but it brings estimation bias as well. Furthermore, as diverse as SMEs are, to some obligors' records, it can be lying far from the main body of data. Those outliers cause another source of estimation bias.

Both outliers and scale differences in original data can be solved by using standardisation. All of the selected variables have passed normality checks into their original form. Therefore, using mean and variance, all the original values can be transferred into the same scale while not changing the order of it. The transformation form is:

---

$$u = \frac{x - \mu}{\sigma}$$

(30)

where  $\mu$  is the variable mean,  $\sigma$  is the standard deviation,  $x$  is the original value and  $u$  is the standard variable. To remove outliers, 90 percent quantile is used and values greater than 1.67 or lower than -1.67 are removed. With the help of quantile values, outliers can be found and removed while keeping the majority of data used.

### **3.5 Separation Measures**

Although one would expect different models could be used for different questions, for example how the economic condition impact SMEs, it is always interesting to know which model can better separate SMEs with ‘good’ performance from ‘bad’ ones. To do this, separation measures are used as criteria. This section explains measures used in this research which are Area Under Receiver Operating Characteristic Curve (AUROC), Kolmogorov–Smirnov (KS) statistics, Gini coefficient and H measure.

#### **3.5.1 AUROC, Gini and KS-statistics**

Defining sensitivity as the ratio between classified true ‘bad’ and all observed ‘bad’; specificity as the ratio between classified true ‘good’ and all observed ‘good’. Receiver Operating Characteristic curve (ROC curve) plots the probability of sensitivity to 1-specificity, (Ma, 2007). The following shows the classification table:

Classified	Observed		Total
	Bad	good	
Bad	A	b	a+b
Good	C	d	c+d
Total	a+c	b+d	a+b+c+d

Table 3.1 Classification table

In Table 3.1, a, b, c and d representing True Positives (TP), False Negatives (FN), False Positives (FP) and True Negatives (TN) and. Using notations given by the above Table 3.1, define sensitivity equals to  $\frac{a}{a+c}$  which is the ratio of correctly classified 'true' bad out of total observed 'bad'. While recording specificity as  $\frac{d}{b+d}$  which is the percentage of 'truly' identified good out of total observed good. Then, 1-specificity becomes  $\frac{b}{b+d}$  which represents the ratio between misclassified 'true' good and total number of observed good. It means the ROC curve plots the cost of misclassify good against successfully identify bad, see Figure 3.1 below.

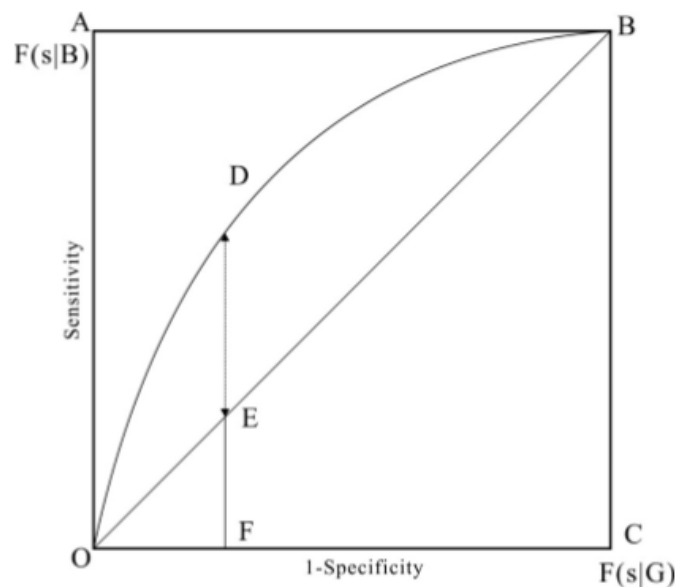


Figure 3.1 ROC (Source: Li, 2014)

---

If a random guess is made, the ROC curve becomes the diagonal line OB. Meanwhile if all good and bad are correctly identified, the ROC curve follows the line OAB. By calculating Area Under the Receiver Operator Curve (AUROC), all possible cut offs are considered. In Figure 3.1, AUROC refers to the area under the curve ODB, starting at O following the curve D to B. As the ODB curve gets closer to the edge OAB, the better separation ability the corresponding model will have (Li, 2009).

Gini coefficient is another measure, which has been widely used in credit scoring, especially in the America. It is defined by Gini (1909) and refers to the proportion of area ODB and OAB. Gini could be calculated as follows:

$$\text{Gini} = \frac{A_{ODE}}{A_{OAB}} = \frac{0.5 - A_{OABD}}{0.5} = \frac{0.5 - (1 - A_{ODEC})}{0.5} = \frac{A_{ODEC} - 0.5}{0.5} = 2A_{ODEC} - 1 \quad (31)$$

which provides the relationship between Gini and AUROC as:

$$\text{Gini} = 2\text{AUROC} - 1 \quad (32)$$

Kolmogorov–Smirnov (KS) statistic is usually used in statistic theory to calculate the distance between two distributions. As credit scoring tries to separate the distribution between ‘good’ and ‘bad’, KS statistic could also be used as a model separation measure. KS statistic in credit scoring content usually refers to the maximum distance between 1-specificity,  $F(s|G)$ , and sensitivity,  $F(s|B)$ , which could be defined as:

---


$$KS = \max_s \|F(s|G) - F(s|B)\| = \max_s \|DF - OF\| = \max_s \|DF - EF\| = \max_s \|DE\|$$

(33)

as  $OF = EF$ . It means the KS statistic represents the biggest vertical separation between the curve and the diagonal (Li, 2009).

### 3.5.2 H measure

Hand pointed out that different misclassification density will add different weight on False Positives (FP) and True Negatives (TN) in previous separation measures, (Anagnostopoulos, Hands and Adams, 2012). This issue will lead incoherent separation measures results, (Ma, 2007). Therefore, Hand introduced H measure. Denote  $\tilde{c}_0$  and  $\tilde{c}_1$  be the cost of misclassification class of False Positives (FP) and True Negatives (TN) and the cost classification  $\tilde{C}$  is defined as:

$$\tilde{C} = \frac{\tilde{c}_0}{\tilde{c}_0 + \tilde{c}_1}$$

(34)

Hand considers the misclassification with a given threshold and a cost proportion  $\tilde{C}$ :

$$\tilde{Q}_{\tilde{C}}(t, \tilde{C}) \triangleq \tilde{c}_0 \pi_0 (1 - F_0(t)) + \tilde{c}_1 \pi_1 (1 - \tilde{F}_0(t))$$

(35)

And the H-measure could be defined by calculating the expected loss as

$$\tilde{L}_{\tilde{C}} \triangleq \int_0^1 \tilde{Q}_{\tilde{C}}(\tilde{T}_{\tilde{C}}(\tilde{C}); \tilde{C}) \tilde{W}_{\tilde{C}}(\tilde{C}) d(\tilde{C})$$

(36)

---

$\tilde{T}$  is the threshold choice with the cost proportions and  $\widetilde{W}_{\tilde{c}}$  is the distribution for cost proportion.

Although Hand criticizes AUCROC would bring incoherent results, yet Flach, Hernandez-Orallo and Ferri (2011) argues that if all the possible thresholds are considered AUROC's results become coherent. This research use the default threshold which is 0.5. Since, AUCROC has been used by both academic and industry, it is easier to compare the results of the current research with others if AUROC is used.

### **3.5.3 Akaike Information Criterion (AIC)**

Akaike Information Criterion was introduced by Akaike (1973) which is calculated by the following equation:

$$AIC = 2k - 2 \ln L$$

(37)

where  $L$  represents the model's estimated maximum likelihood and  $k$  is the number of total estimated parameters (Akaike, 1973). AIC is positively correlated with the number of parameters included in the model, yet negatively correlated with the model's fitting. A lower AIC value means one could achieve a better model fitting with a less complex model, which contain fewer parameters. Hence, model with a lower AIC value is preferred.



---

## 3.6 Summary

Chapter three provides a theoretical foundation for data transformation, variable selection, and models used in this thesis. Three models discussed in this chapter are: 1) logistic regression which is the widely used model in industry and benchmark models for others; 2) panel data models for which the discussion contains the choice of estimators is presented in detail and so too is the use of macroeconomic variables; 3) GAM which prefers variables in its original format to improve model clarity and an alternative method of handling missing values is provided.

---

## 4 Data description

### 4.1 SMEs Definition

Small and Medium sized Enterprises (SMEs) exist globally, yet there is no clear definition of SMEs which is universally accepted. For example, different definitions of SMEs are adopted even among different government departments in U.S. (U.S. international trade commission, 2010). However, although the threshold varies, SMEs definitions usually consider the following aspects: the total number of employees, total assets or total turnover. Detailed definition of SMEs may have significant difference across the world, for example by U.S. Small Business Administrations, European Commission, China's Regulations on the Standards for Classification of Small and Medium-sized Enterprises as well as within other organisations.

Within the European Union, the SMEs definition is given by the European Commission stating that enterprises should be regarded as SMEs if they have no more than 250 employees and satisfy one of the following criteria (The European Commission, 2015):

- Total turnover less than €50 million
- Balance sheet total less than €43 million

---

For simplicity, currency difference is ignored in this research and total turnover less than 50 million GBP is used as threshold. Firms which could not satisfy the above criteria are removed from the sample.

## **4.2 Sample**

Following the removal of firms not meeting the European Union criteria for SMEs, a stratified random sample is selected to represent the original population and with a 'bad' flag chosen as strata definition. In my data, the 'good' and 'bad' state is given by a flag variable called 'myflage'. For preference of my data supplier, the definition of 'bad' statue could not be fully illustrated here. However, it is clear that not all SMEs labelled as 'bad' had missed three payment. Therefore, the 'bad' flag in the data set is not a default flag but a financial distressed indicator. If a SME is assigned with a 'bad' flag, it means the corresponding firm's credit risk has been raised significantly and its ability to replay is of deep concern. However, a more detailed definition of 'bad' flag could not be fully provided due to constraints from the supplier of the data. During the period of study, SMEs are clearly under pressure and their credit risk becomes much higher than in normal time periods. Given the heightened number of 'bad' outcomes a random selection could select enough 'bads' for model estimation.

To form an unbiased sample, the preferred method would be to randomly select a fraction of the original population. However, if a ten per cent sample was randomly

---

chosen each year, the selected sample could not form a consistency record over time. Random choice cannot guarantee that the same firms are selected each year which would lose both the time series effect and the panel data feature. To avoid aggregating the unbalanced pattern on both end of data and tracking the influence of ‘credit crunch’ on SMEs from 2007 to 2010, firms were randomly selected at 2007 and their performance monitored over a period of time.

In summary the data selection process is as following:

1. Remove any SME which contradicts the European Commission’s definition of SMEs;
2. Ten per cent of SMEs were randomly selected from the original data in 2007 to form our training sample;
3. Those SMEs are selected through the observed period.

An out of sample hold out sample is used in this research: 10 per cent of 2010’s SMEs were selected as holdout samples to test the robustness of the models. The aforementioned strategy resulted in a discreet time unbalanced panel data describing the 2007’s obligors’ performance from 2007 to 2010, during which the total number of observations decreased from period to period as the SMEs labelled as ‘bad’ flag drop out of the sample. As this research separated the sample according to SMEs firm age, the sample size and ‘bad’ rate will be presented in later sections.

---

## 4.3 Segments

When having such a large SMEs sample, segmentation can be used to reflect strata within the population. Newly established SMEs face different risks and issues compared to matured ones. For example, there is either none or too little accounting history for newly established SMEs when they are applying for financial support. Meanwhile, every mature SME has accounting records, but these records could suffer from a lack of frequent updating. The differential behaviour between newly established and mature SMEs in terms of risk means that it is better to divide SMEs into the two segments and analyse them separately.

Here whether a firm has been established for three years or not is chosen as the threshold:

- Start-ups: newly established, existing for no more than three years;
- Non-start-ups: more matured SMEs which have been in existence for more than three years.

This threshold is chosen mainly for distinct performance among the new SMEs and matured ones. Although other factors, such as region and SIC can also lead to segment difference, there is clearly a difference between newly founded SMEs and mature SMEs. Any specific split is arbitrary but historically people have often used three years as the threshold between young and old businesses. A detailed comparison will be provided in the following sections.

---

The size of the resulting samples are summarised as follows:

Size of the training sample				
	2007	2008	2009	2010
Non-start-ups	111021	103072	94039	83697
Start-ups	94742	72034	51350	39773
total	205763	175106	145389	123470

*Table 0.1 Segment size across 'credit crunch'*

For both segments, the data set is an unbalanced sample with a decreasing sample size. In start-ups in particular, there are a significant number of firms which drop out each year, causing our panel to become unbalanced. SMEs in the state 'bad' disappear in the following year. However, SMEs leave the databases for a series of other reasons than just being 'bad'. Therefore, there is censoring in the data.

#### **4.4 'Bad' rates**

Due to commercial sensitivity deemed by the data supplier, the default definition used in this research could not be fully disclosed. However, it could be stated that the 'good' or 'bad' flag used does not follow a simple 90 days past due threshold, as specified in Basel Accords (BCBS, 2004). Therefore, the dependent variable, which is the 'good'/'bad' flag, is used to refer to accounts, which show significant sign of being likely to default. Default is then used as a shorthand the term.

Figure 4.1 highlights the 'bad' rate of the training sample and that of two segments.

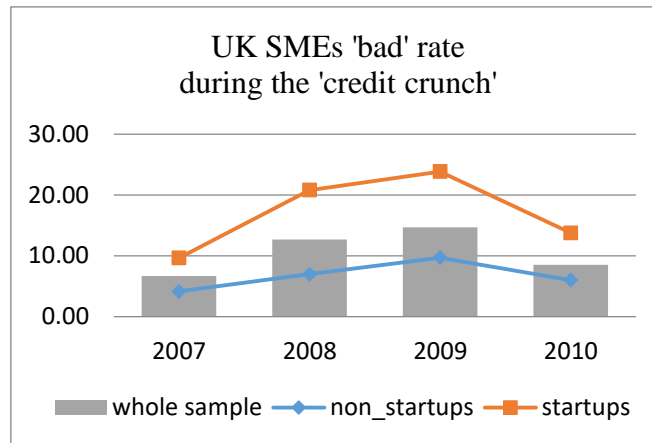


Figure 0.1 UK SMEs 'bad' rate during the 'credit crunch'

'Bad' rate of the whole sample is presented by bins in Figure 0.1; the orange and the blue lines show that of 'start-ups' and 'non-start-ups' respectively. The previous 'credit crunch' has deeply impacted SMEs as their 'bad' rate significantly rose in 2008 and 2009. A clear difference is exhibited between 'start-ups' and 'non-start-ups.' Start-ups present a much higher 'bad' rate compared to non-start-ups over time. A sharp increase of 'bad' rate was evident for start-up SMEs in 2008, the situation deteriorated in 2009 and subsequently recovered to a near normal situation in 2010. In 2009, its 'bad' rate was almost twice as high as that in the normal period. For non-start-ups, although the 'bad' rate line lies below start-ups, its 'bad' rate's trend performs in quite a similar manner to that of start-ups.

In 2007, the 'financial crisis' started in the U.S. and its negative effects soon spread to the rest of the world and especially in developed countries. In 2009, the UK's economy experienced a deep recession which led to a huge amount of SMEs becoming financial distressed. However, the UK SME's macroeconomic environment was not able to recover immediately in 2010 since the Euro Zone also suffered from a

---

crisis across a number of states including Greece, Iceland and Portugal. However, the Euro crisis was less influential in the UK as its major economic figures showed a significant recovery. The trend of 'bad' rate shown above summarises how UK SMEs struggle to survive when the macroeconomic conditions change dramatically.

#### **4.5 Macroeconomic Variables (MVs)**

During the 'credit crunch,' the macroeconomic condition noticeably shifted and there was intensive discussion of its influence on the SMEs performance. This research uses panel data models which not only fit the discrete time feature of SMEs data, but also consider multiple time periods. Therefore, MVs are introduced into SMEs credit risk modelling and their influence over time will be carefully discussed. As the observation period lasts for four years, selected macro variables (MV) should be reported in a consistent pattern during this period as well as being most relative to SMEs performance in the UK. Considering the short time period this research observes and the lack of update in SMEs data, only a small set of MVs are tested in this research. Additionally, MVs have a complex correlation structure among themselves and interact with firm specific variables.

To avoid noise caused by interaction and bearing in mind that the aim of using MVs is solely to control the market movements during a short period, only the most commonly used MVs are selected in this research. They are GDP growth rate (which is labelled as 'gdp3'), CPI growth rate (which is labelled as 'cpir'), inflation rate



(which is labelled as ‘inflation’), unemployment rate (which is labelled as ‘unemployment’), FTSE all share index (which is labelled as ‘ftsall index’) and FTSE 100 index (which is labelled as ‘fts100 index’). Most MVs data comes from World Bank except financial market data, which are FTSE all, share index and FTSE 100 index. The latter data comes from Datastream. Whilst only a small number of MVs are employed in this research, those MVs chosen accord to Figlewski et al’s (2012) framework. Since they are able to demonstrate important aspects of economy, such as direction of economy, general economic condition and the financial market movements.

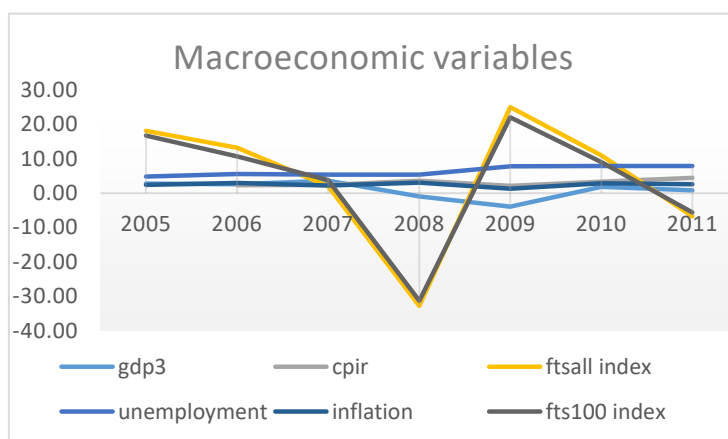


Figure 0.2 UK Macroeconomic data from 2007 to 2011

Above figure shows dynamics of the MVs from 2007 to 2011. GDP growth rate shows the direction of economy which marks 2009 as the worst economic condition year during the crisis. Although general economic condition MVs, which are unemployment rate, CPI and inflation, seemed to remain flat, it needs to be borne in mind that a fraction of MVs’ change could bring significant hazard on investment,

---

business activities and financial accessibility. Among three types of MV variables financial market related variables showed the most notable change.

## 4.6 Variable Selection

The data contains 79 variables describe firms' general features, directors' information, financial statements, payment type and previous records. Firms' general information contains descriptive variables, such as firms' legal form, location and industry classifications. Directors' information shows the size of board, the mobility of board members, age of directors and such. Financial statements of the firm include percentage of shareholders' fund, time since last annual return, lateness of account and other information. Previous records covers firms' searches for funding, derogatory records and others aspects.

There are various reasons for the interaction between the variables. Firstly, variables may provide very similar information and so will have high collinearity, for instance, *percentage change in DBT from current to one month previous* and *percentage change in DBT from current to three month previous*. Secondly, some of the categorical variables may have large categories which are also problematic. A large proportion of SMEs have one director, therefore *No. Of 'Current' Directors* is dominated by one large category. This will also influence other variables, such as *Number of Directors Holding Shares*. Another reason of correlation is the high

percentage of missing value. If the SME is missing *Total Number of Judgements* it will not have a record for *Total Value of Judgements*.

As mentioned in the Methodology Chapter, this research employs stepwise logistic regression for variable selection. In addition, owing to the use of WoE, the selected variables with negative coefficients are removed from the model, since they imply estimation bias. The WoE is calculated depending on variables' categorical 'good' rate in this research. While modelling on 'good' status of obligors, all variables should only present positive parameters. After deletion, stepwise logistic regression is performed again to select the significant variables. This process is repeated until the selected variables all present positive coefficients. This section provides the results of the stepwise selection and shows how many variables have been selected for both segments. Table below shows the process of variable selection with significant level 95%:

Year	'stat-up'		'non-start-up'	
	Times	No. of removed variables	Times	No. of removed variables
2007	12	18	7	25
2008	12	25	5	28
2009	9	23	4	38
2010	9	22	6	24

*Table 0.2 selecting significant explanatory variables*

Table Table 0.2, the columns headed 'Times' show how many rounds of logistic regressions were carried out to remove all the explanatory variables with negative

---

coefficients in the responding model. Columns headed as ‘No. of removed variables’ show how many variables, in total, have been removed before selected variables have all positive coefficients. For example, for 2007’s ‘start-ups’, after removing 18 variables, on the 12th round all selected explanatory variables showed a positive correlation with the dependent variable. Therefore, the ‘times’ column for ‘start-ups’ in 2007 has value 12 and ‘No. of removed variables’ column has the value 18. Notice that different sets of variables are selected as time varies as well as different segments.

From Table 0.2, it is evident that significant variables vary from segments with different sets of variables selected for ‘start-ups’ and ‘non-start-ups.’ The segmentation difference is supported by different number of selected variables. However, the same set of variables should be used over time for each segment. Only by analysing the same set of variables through the ‘credit crunch’ was it possible to capture the time series effect and demonstrate SMEs performance through this period.

To obtain the same set of independent variables, this research chooses variables which had been selected over three years or four years by stepwise logistic regression. The reason for this choice was that it not only results in a manageable size of dependent variables but also reduces the annual variation. If variables remaining in the models for one year or more are selected, there would be a total of 38 variables selected for ‘start-ups’ modelling among which only seven variables are selected for four years.

To select the most influential variables and further control for collinearity, the variables which appear less than three years are removed from analysis. The reason

---

for keeping variables being insignificant in one year is that those variables capture annual difference.

Additionally, although subsidiary indicator variables – *company is subsidiary* and *parent company* were selected for both segments, they have information value lower than 0.01 over four years. The reason is that one large category dominated this variable: more than 90% of observation belonging to a single category. Therefore, those two variables were also dropped from the model. In the end, 15 variables were selected for ‘non-start-ups’ and that number of ‘start-ups’ 12.

#### **4.7 Explained variables’ description**

This section provides detailed SMEs description according to various features, such as their industry classification, region, accounting information and other features. With these descriptions, this section gives an initial analysis of SMEs performance and these initial findings are useful to support the assumptions which have been made in methodology part. For example, the missing category’s feature supports the assumption that this category has a relatively consistent performance and single imputation is suitable in this case.

As previously mentioned, this research employs a huge UK sample, which not only contains 2.1 million SMEs recorded in 2007, but also covers a variety of SMEs attributes. There as many as 79 different explanatory variables which cover the following aspects of a SME:

- 
1. General information: such as their legal form, location information, 1992's Section of Industry Classification (SIC), No. of employees, age of company and so on.
  2. Directors' information: No. of directors in total and other general director's information, management ability not included.
  3. Previous relevant credit history: such as DBT, judgement and previous searches information
  4. Accounting information: all the commonly used financial ratios.

Firstly, the description gives SMEs categorical performance according to industry classification, location and lateness of account. The discussion then focuses on the missing category's performance. Its characteristics support the treatment of missing value in this research. The third section shows other variables trends which demonstrate the non-linearity feature of independent variables. This data character supports the use of WoE and the use of GAM.

#### **4.7.1 SMEs performance according to SIC, region and account lateness.**

##### **i) Cross industry performance**

Industry classification is an important feature that leads to categorical difference in SMEs performance. This research uses UK 1992 Standard Industrial Classification (SIC) Codes to divide SMEs into industry categories. This standard is the most widely used one during the observed period. With the help of coarse classification and

considering general industry divisions, more than nine thousand of SIC is classified into 12 category for start-ups and 14 for non-start-ups. Figure 0.3 and Figure 0.4 present SMEs categorical frequency according to SIC code. Years are labelled from 2007 to 2010 as ‘APR07’ to ‘APR10’:

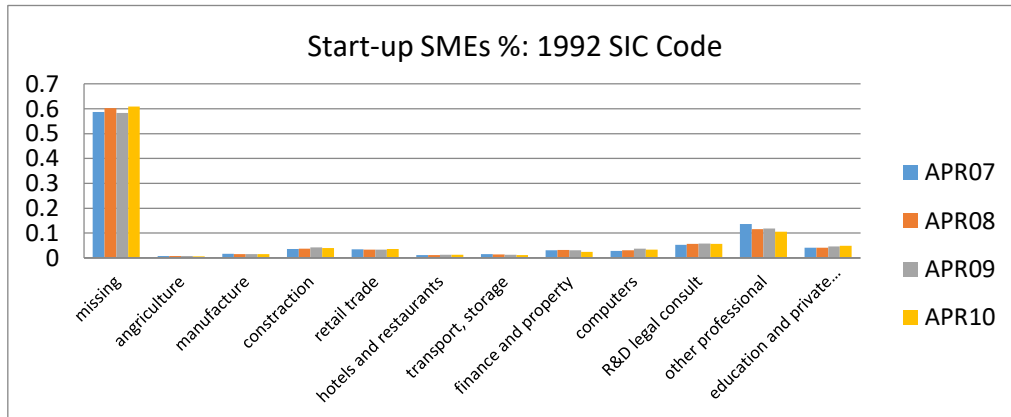


Figure 0.3 Start-up SMEs categorical frequency according to 1992 SIC code

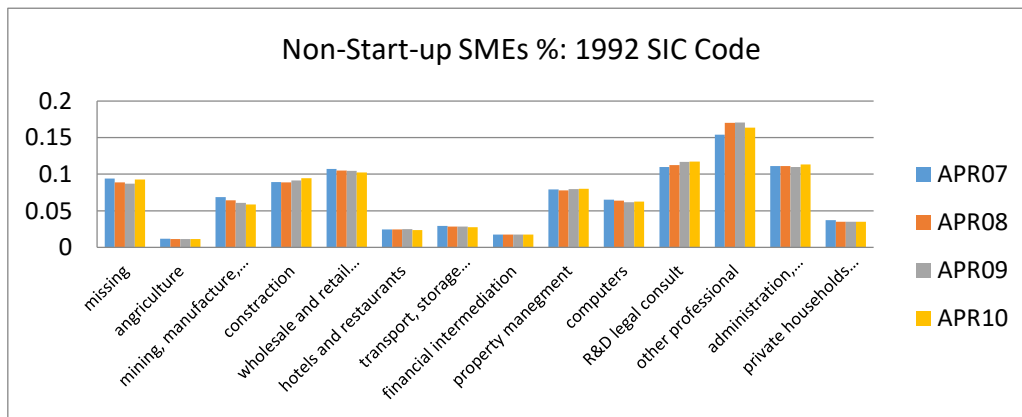


Figure 0.4 Non-start-up SMEs categorical frequency according to 1992 SIC code

The above two figures show the frequency of start-ups and non-start-ups by industry sectors. The most obvious difference between ‘start-ups’ and ‘non-start-ups’ is the size of the missing category which dominates ‘start-ups’ but have a far lower frequency for ‘non-start-ups’. Failing to report industry classification is a constant

feature for a SME since, for both segments, the frequency of the missing category remains stable over time. The following discusses SMEs' 'bad' rate according to SIC:

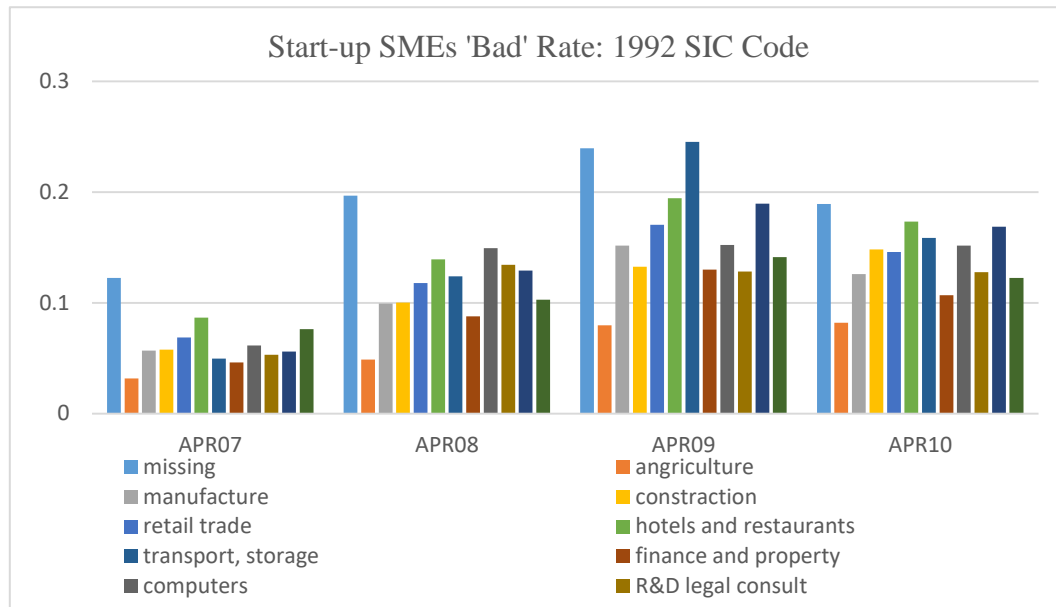


Figure 0.5. Start-up SMEs 'bad' rate according to 1992 SIC code

The 'start-ups' performance according to their industry classification is distributed in Figure 0.5. The missing category of 'start-ups' exhibits high 'bad' rate compared to other categories, especially in 2008. However, when other categories 'bad' rate increased to surprisingly high levels in 2009, this disadvantage became less noticeable. Professional firms, which are classified as *other professional* here, had the highest default rate in 2009. Thus, those SMEs offering professional service seemed to be more prone to suffer financial distress during the period of the crisis than in normal times.



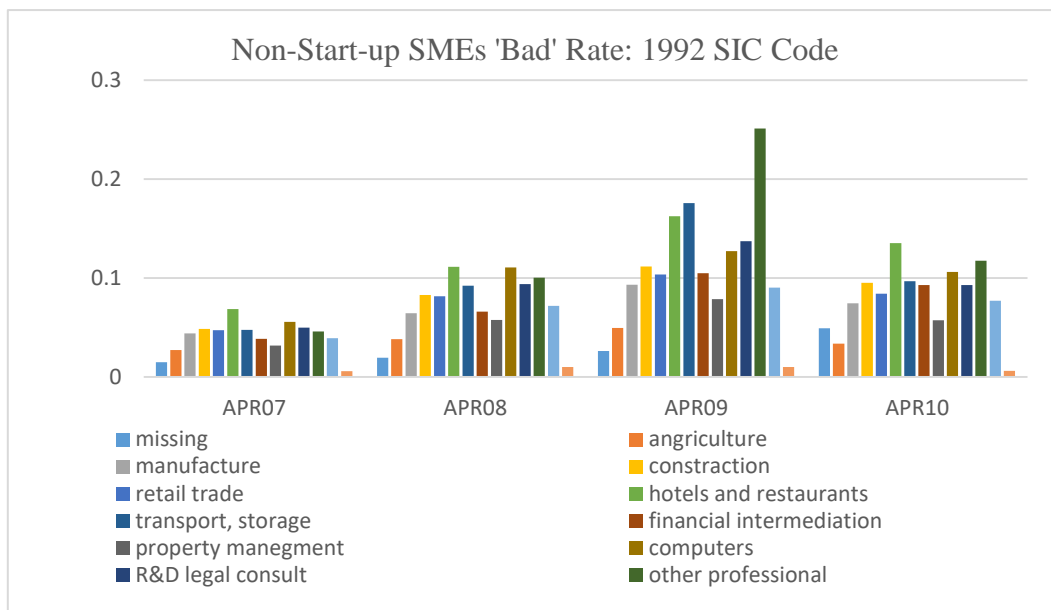


Figure 0.6 'bad' rate for non-start-ups according to 1992 SIC code

In comparison to 'start-ups,' the missing category had a very low default rate for the 'non-start-ups' segment. There was a sharp increase of default for *other professionals*. However, the time variation was far more noticeable for 'non-start-ups' which was particularly evident for *hotel and restaurants, transportation and storage and other professionals*. This suggests that the service sector was hit hardest by the 'crisis.'

## ii) Regional performance

SMEs' locations were spread across the UK and their business behaviour varies accordingly. The regional policy, economic condition and financial institutions accessibility would influence SMEs performance. For instance, the London area has the largest population, has a more highly developed commercial community and also a greater number of financial institutions. On the contrary, a firm in the Scottish Highlands would face a completely different business environment. During the 'credit crunch' the financial system faced a major challenge. London firms therefore faced a

greater risk owing to the large number of financial firms in this area. The following two tables summarise the regional frequency (which is labelled as ‘percent’) and ‘bad’ rate (which is labelled as ‘brate’):

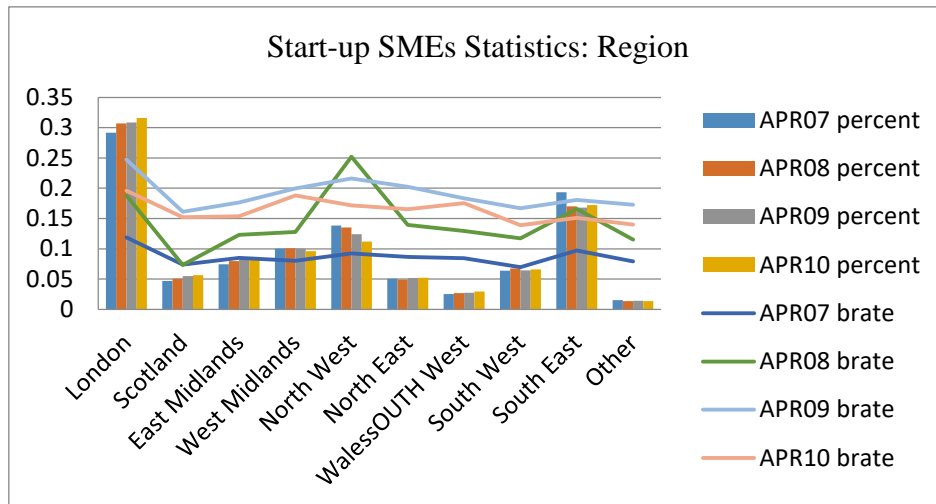


Figure 0.7 'start-up' SMEs statistics according to region

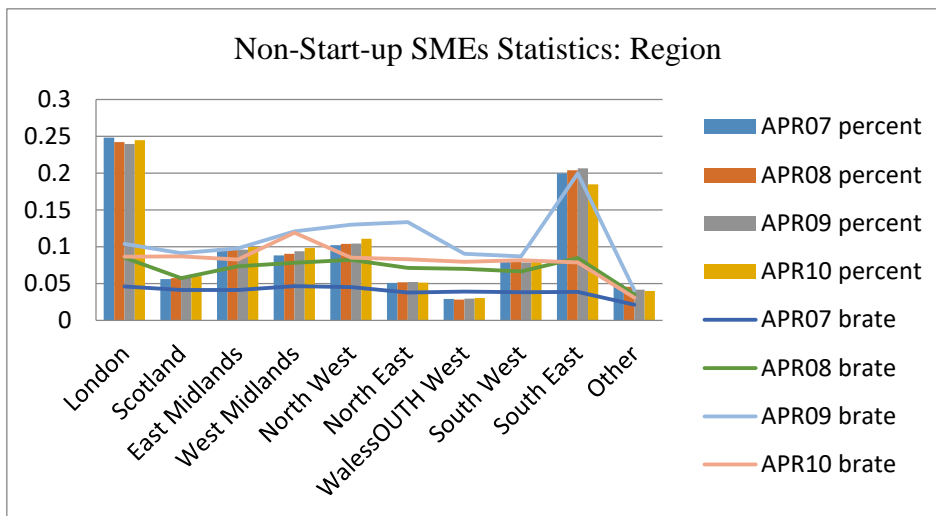


Figure 0.8 'Non-start-up' SMEs statistics according to region

The other category refers to firms which could not be classified into the nine categories. The London region has the largest number of SMEs for both segments. There is no significant regional frequency difference between the two segments. For

annual performance, in 2008 there was a significant regional variation for ‘start-ups’ while for ‘non-start-ups’ regional difference is not so obvious in that year. For categorical performance, London based ‘start-ups’ suffered a higher ‘bad’ rate over the four years, while for ‘non-start-ups’ one noticeable category was the south east based where the ‘bad’ rate sharply increased in 2009.

**iii) Lateness of account**

Figure 0.9 shows the percentage of lateness of accounts by the time of lateness through the four years.

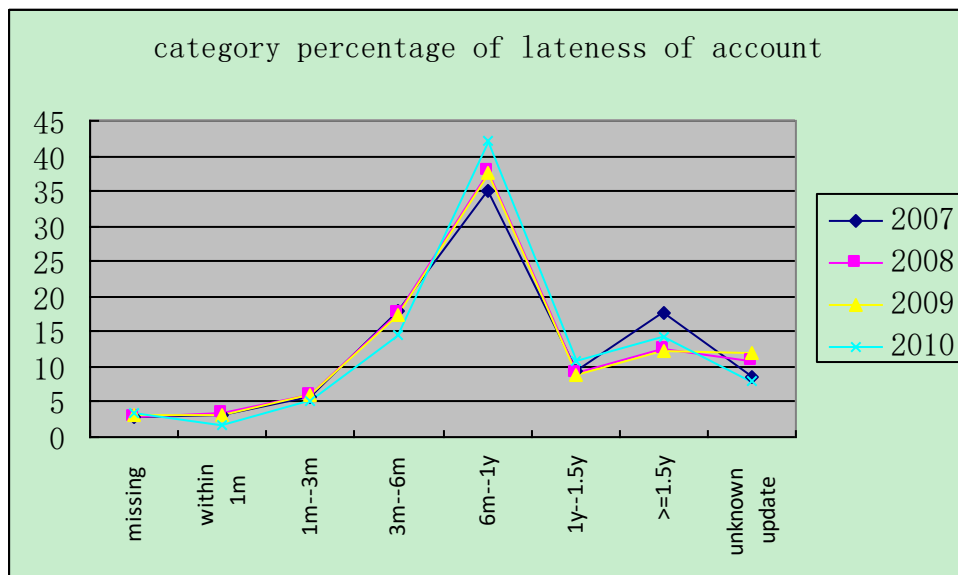


Figure 0.9 SMEs frequency of lateness of account through 'credit crunch'

The majority of SMEs take longer than three months to update their accounts which is a distinctive behaviour when compared to corporations and consumers data. For the retail consumers, their detailed account information could be updated by daily transition. Other information such as change of address is usually modified to the bank on time as well. The liquidity of corporations stock leads to the frequent

---

adjustment for their assets market price. The discrete time feature of SMEs accounting information challenges the application of various types of credit models, but may be explored by panel data modelling.

In this section three signature variable statistics are discussed, which are SIC, region and its lateness of account, to provide the general understanding of SMEs performance during the ‘credit crunch.’ There are two important common points when analysing SIC and region:

1. Segmentation difference. The first concern would be the difference between ‘start-ups’ and ‘non-start-ups,’ which proves the necessity of segmentation.
2. Annual variations. There is evidence that single level models, such as logistic regression faces great challenges during the ‘credit crunch.’ Additionally, the changes ensure the success of the panel data model application since the estimation of panel data model is based on time variation.

SMEs credit scoring models are limited by the data availability owing to their information opacity. Although the missing values categories have been mentioned, more analysis needs to be undertaken to support the assumption made since the treatment of missing category is an important part of this research. Hence, the following section focuses on the discussion of missing category.

---

#### 4.7.2 The missing category

Missing data is a feature in SMEs credit scoring, therefore simply deleting SMEs with missing information can cause a significant loss of data and lead to estimation bias. Given that frequently missing value category provides negative behavioural information. In the data more than 90% SMEs do not have a specific figure of *number of employees*. Instability in SMEs and movement of employees could both mean that the number of employees is difficult to measure. In addition, variables such as *capital employed* could be missing due to unclarified capital ownership between the SMEs and their owners. For example, where the owner of a SMEs is classified as private house holder, it may difficult to clarify whether the value of the property should be counted as an asset of the firm or not.

Furthermore, the missing category does not necessarily perform below the average. If an enterprise misses *Time since last derogatory data item (months)*, it is more likely a result of no previous derogatory data available, which means no previous concerns have occurred. Also, if the bank decides to accept obligors with missing information, those obligors may have counterbalancing advantages which caused them to be accepted.

This section exhibits a series of variables' missing category which usually not only have a large volume but also a relatively stable performance over time. Although it is almost impossible to collect exact reasons for the missing values for each variable and the 'bad' rate varies, yet the missing category can still be very informative even if it is

treated as a group. Therefore, this supports the assumption that the missing values for variables do not occur randomly, and it provides information useful for credit risk modelling.

**i) The number of 'current' directors**

Figure 0.10 shows the 'bad' rate of 'non-start-up' SMEs for a different number of current directors. This variable shows the size of the firm's board and is only significant for 'non-start-ups.' The statistics show the missing category has a stable frequency over time and missing values represent a group with lower 'bad' rate.

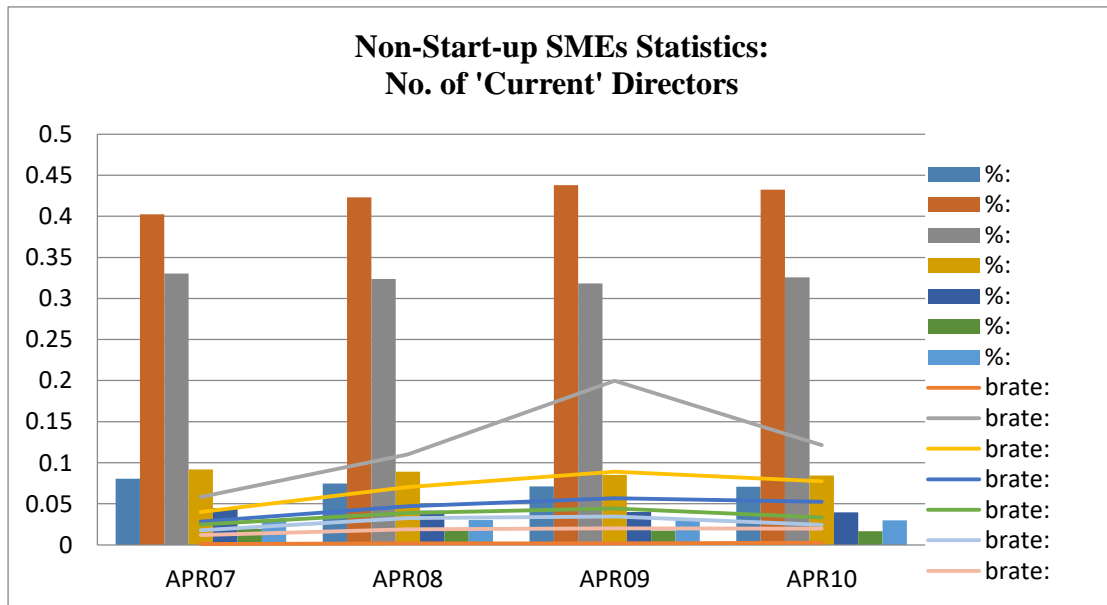


Figure 0.10 Frequency and 'bad' rate of Non-start-up SMEs according to number of current directors

Each segment is explained by the legend in Figure 0.10. For example, the light blue column represents missing category's frequency (which is labelled as '%') while its 'bad' rate is presented by the dark red line (which is labelled as 'brate'). The size of the missing category is stable over four year, meaning this category is not affected by the crisis. There is clearly a group of firms who fail to provide this information.

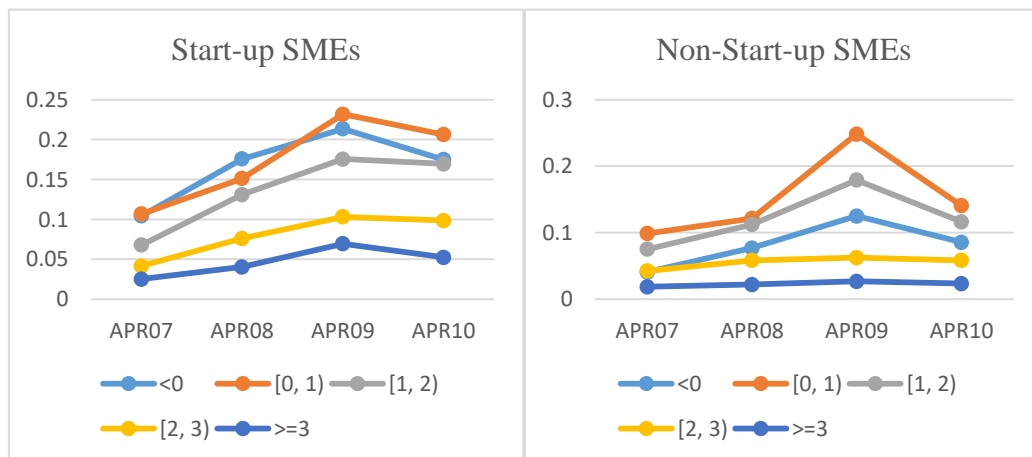
Furthermore, this group of firms has the lowest ‘bad’ rate over four years. It is strong evidence to support the assumption that the missing value does not occur at random and firms that fall into this category have similar performances.

This variable also addresses the benefits of a large sample set. The smallest category, SMEs with five directors, has a percentage lower than two per cent of the training sample. However, it still has more than a thousand observations each year which ensures the statistic results are reliable.

**ii) 'Bad' Rate of SMEs according to Proportion of Current Directors to Previous Directors in the Last Year**

The *proportion of current directors to previous directors in the last year* shows the change in the firms’ board. As mentioned before, the size of the current board is not significant for ‘start-ups,’ although the change is significant for both segments. Therefore, the changing of a board is a more important issue for ‘start-ups.’

Figure 0.11 and **Error! Reference source not found.** present SMEs categorical performance according to this variable:



---

Figure 0.11 (left) Start-ups 'Bad' Rate according to Proportion of Current Directors to Previous Directors in the Last Year

The missing category is represented by a negative value. By comparing these two charts, a noticeable performance difference could be found between the two segments. The worst performing category for both segments is the one falling into interval  $[0, 1)$ . SMEs falling into this category decreased their number of directors compared to previous year. For this variable, the missing category is the one with a negative value, and the missing category's performance is significantly different for the separate segments. 'Start-up' SMEs' missing category has a high 'bad' rate, whereas for 'non-start-up' SMEs the missing category has a medium 'bad' rate. Therefore, the missing category is not always the worst performing category and could be very informative in credit risk modelling.

#### **4.7.3 Other variables' trends**

The selected variables' trend in this section demonstrates that independent variables are not necessarily linearly correlated with dependent variable. This feature supports the use of WoE and GAM. In addition, these variables also present annual differences and segments variation for SMEs performance during the 'credit crunch.'

##### **i) Oldest age of current directors**

Age is related to directors' knowledge, experience and their risk talking preference. This variable is significant only for start-up SMEs, which means the directors'



experience is more influential for ‘start-ups.’ Figure 0.12 presents the categorical performance of ‘start-ups’ according to this variable:

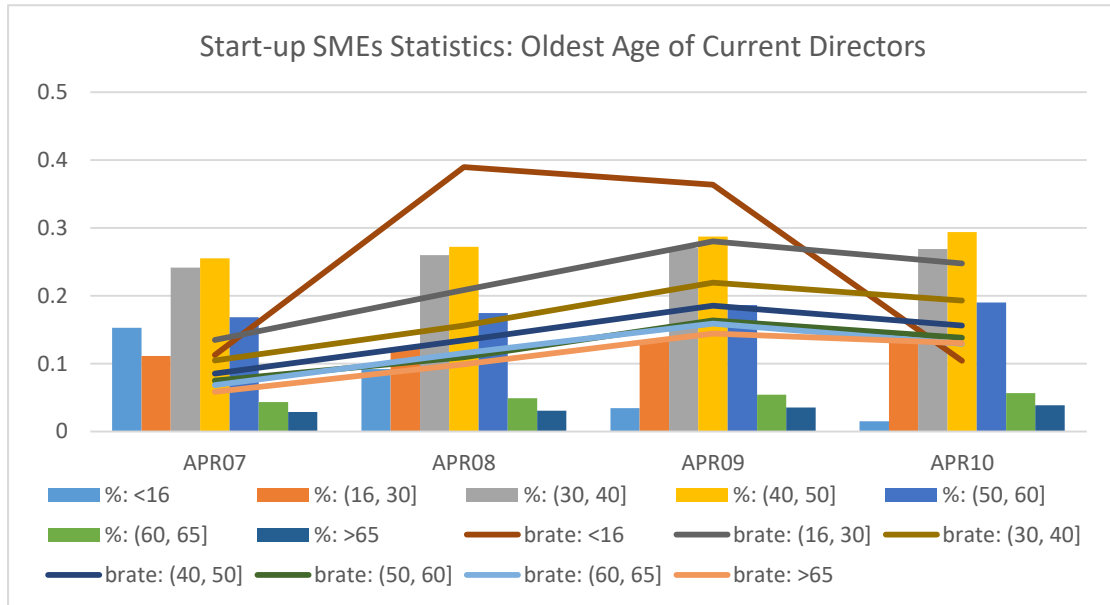


Figure 0.12 Start-up SMEs frequency and 'bad' rate according to Oldest Age of Current Directors

The category with youngest directors’ age has the most dramatic change through the ‘credit crunch.’ It starts with a higher level of ‘bad’ rate at the beginning of the financial crisis although not the highest one, then their ‘bad’ rate sharply increases in 2008. This means it is very risky if the ‘start up’ enterprise has very young directors and these enterprises may suffer most during the crisis.

## ii) Percentage change in shareholders’ funds

The *percentage change in shareholders’ funds* is significant only for ‘non-start-ups.’

Figure 0.13 presents ‘non-start-ups’ categorical performance:

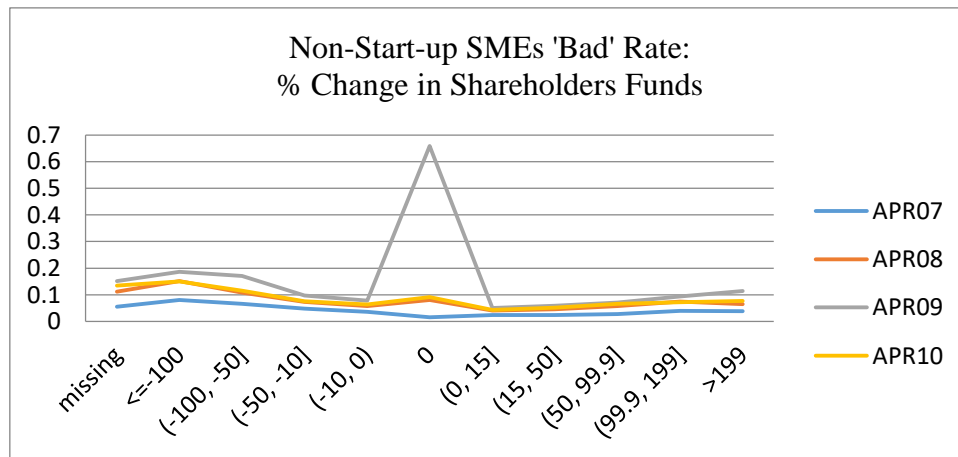


Figure 0.13 Non-start-up SMEs 'bad' rate according to the percentage change in shareholders' funds

There is a sudden 'bad' rate increase for 'non-start-ups' with the category 'none or small change' in shareholders' funds in 2009. In 2007 this category showed the lowest 'bad' rate while as the financial crisis occurred the 'bad' rate of this category increased in 2008 and hit the highest 'bad' rate in 2009. In 2010 the pattern was similar to 2008. This category is strongly influenced by the financial crisis which implies that during crisis any change in shareholders' funds could positively influence the SMEs performance regardless of the direction of change. If the shareholders' funds do not react towards the crisis, the risk of 'bad' performance can become extremely high. This variables' performance highlights the non-linearity feature of SMEs.

Section 4.7 provides SMEs performance according to selected variables. Firstly, the sample 'bad' rate is given to provide coherent evidence of the macroeconomic shocks which appeared through the 'credit crunch.' Then, SMEs location and SIC statistic summaries show how changeable SMEs' performances are and so the models employed have to be robust for such a large sample. Then the discrete time feature of

---

SMEs data is highlighted by *lateness of account*, which leads to the preference of using panel data models. Other approaches are presented later to address other features of SMEs data such as non-linearity and large number of missing value. The non-linearity nature supports the use of WoE and GAM. The missing values' feature confirms that missing value does not occur at random and missing values tend to be informative in building credit risk models.

## **4.8 Replacing missing values**

As demonstrated in the previous section, missing value is a feature of SMEs data. Missing values constitute a large proportion of SMEs data, also they aid PD prediction. The previous section discussed categorical variables or continuous variables categorical performance which leads to the use of WoE. The transform of WoE is suitable for logistic model and panel data model since it solves the missing value issue and overcomes the linearity limitation for most statistical models. However, as GAM involves non-linear component, this research explores an alternative methodology to retain continuous variables in their original format. Hence, it is necessary to fill the gaps caused by missing value. The detailed imputation method has been explained in the Methodology Chapter and this section discusses the imputation in more detail.

As discussed in the Methodology Chapter, the missing category will be replaced with the observed value by matching their performances. To do this for each variables a

moving average (MA) of the ‘good’ rate of non-missing values is plotted and compared to the default rate of the missing category which will be a horizontal line. The points where two lines either intersect or the distance between lines are minimised is taken as the imputation value. When more than one point exists, an average value is usually used. However, the situation varies for different variables. The selection of imputation is discussed in detail as each variable’s performance varies and the imputation value could explain

### 4.8.1 Non-start-up SMEs

For non-start-up SMEs the variables’ MA curves present a similar pattern through the ‘credit crunch.’ although the missing category’s performance switches around the MA curve. For example the *number of current directors* shows that there is a consistent pattern for this variable through time although the exact crossing point could be different.

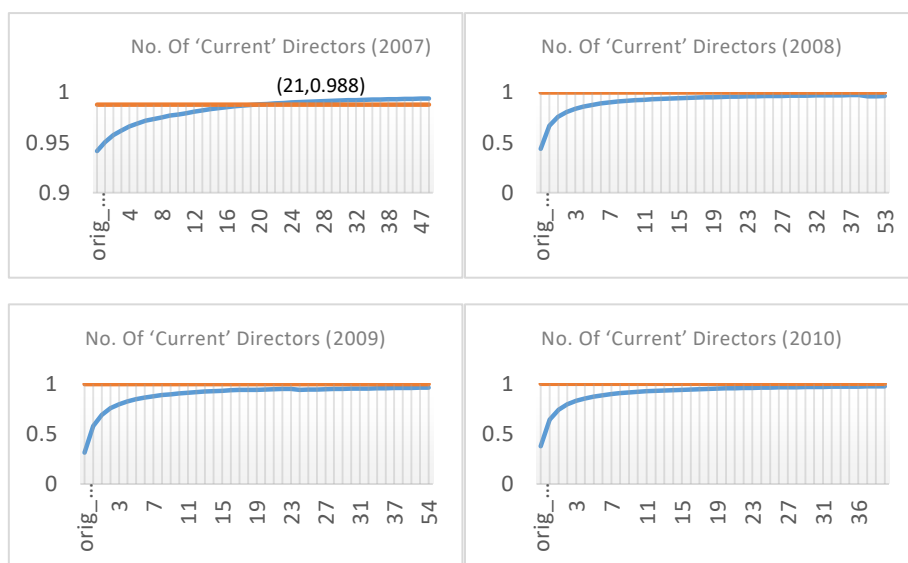


Figure 0.14 No. of 'Current' Directors: 'bad' rate of missing V.S. MA of observed value (2007-2010)

Figure 0.14 presents four years' data for 'No. of current directors'. The flat orange line represents 'bad' rate of missing category while the moving average of observed values are plotted by the blue curve. This notification is used in the rest of this section. Over time, the MA curve approaches the missing category's performance at its right hand tail. In 2007, those two lines actually intersected with each other, but not for the other three years. The chosen point minimises the distance between those two lines to impute the missing category value. All of these lay at the right hand side tail. Hence, the missing category always has similar performance with firms who have a large number of directors. The selected value is given in Figure 0.15 below for 2007 as illustration.

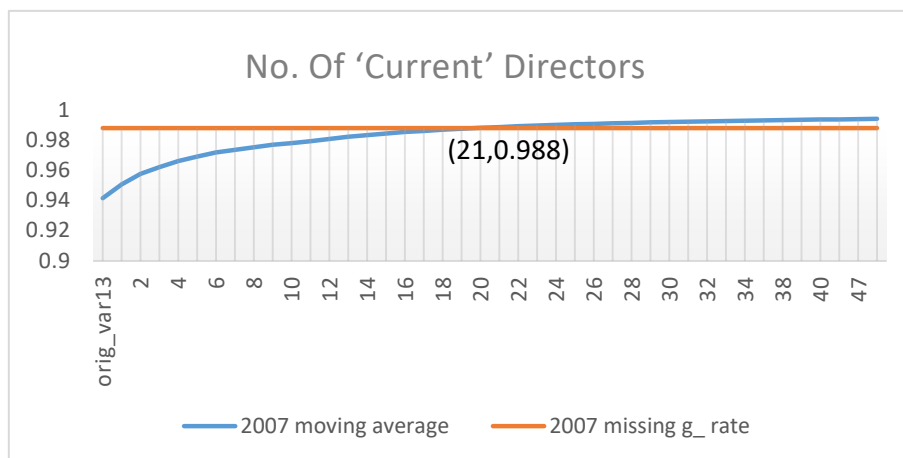


Figure 0.15 No. of 'Current' Directors: 'bad' rate of missing V.S. MA of observed value

*No. of Current Directors* describes the current size of the board. There is only one clear crossing point for this variable in 2007 which is ideal to be used as the filling value. . The curve and the line stays very close to the end. The missing category behaves similarly to the medium size of directors' board.

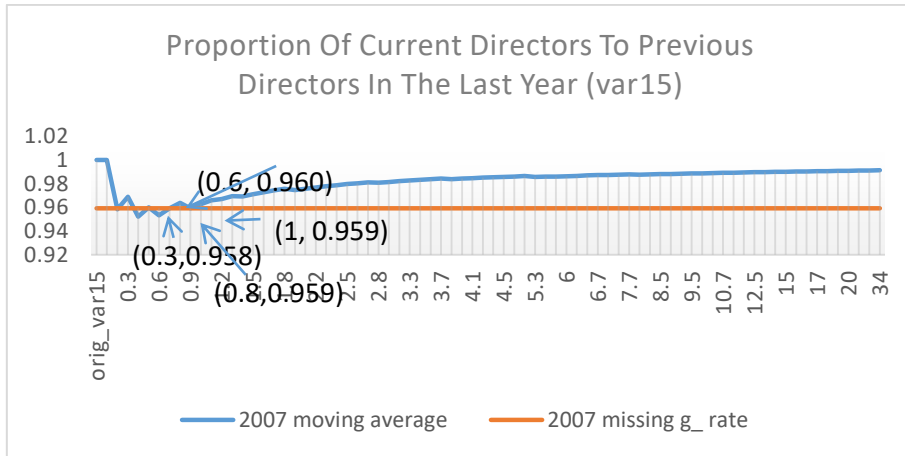


Figure 0.16 Current Directors Proportion: 'bad' rate of missing V.S. MA of observed value

The *Proportion of Current Directors to Previous Directors in the last year* describes the mobility of SMEs director boards. Figure 0.16 shows the MA which decreases sharply at the beginning, then hovers around the missing category's performance before monotonously increasing. Multiple crossing points occur at the beginning. It means the group of missing value has a similar performance to the early part of the curve. To avoid causing more noise, the average of the crossing points is used to match the missing category.

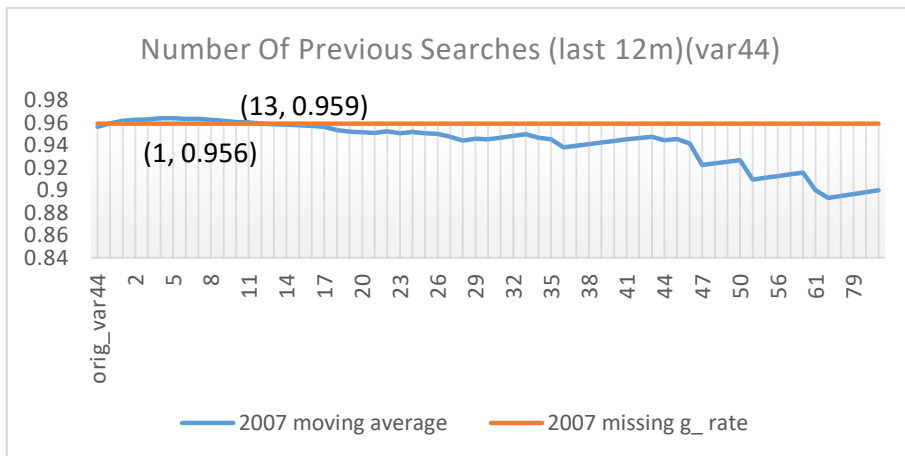


Figure 0.17 No. of Previous Searches: 'bad' rate of missing V.S. MA of observed value

The MA's curve is rather flat when *No. of Previous Searches* are less than 19, with this part of the curve close to the value of missing category. More than one crossing points are observed in this part. The average of the two crossing points is taken as the imputation for the missing value. In summary, the missing category performs very similar to non-start-up SMEs which have not been applied for funding too often in the past.

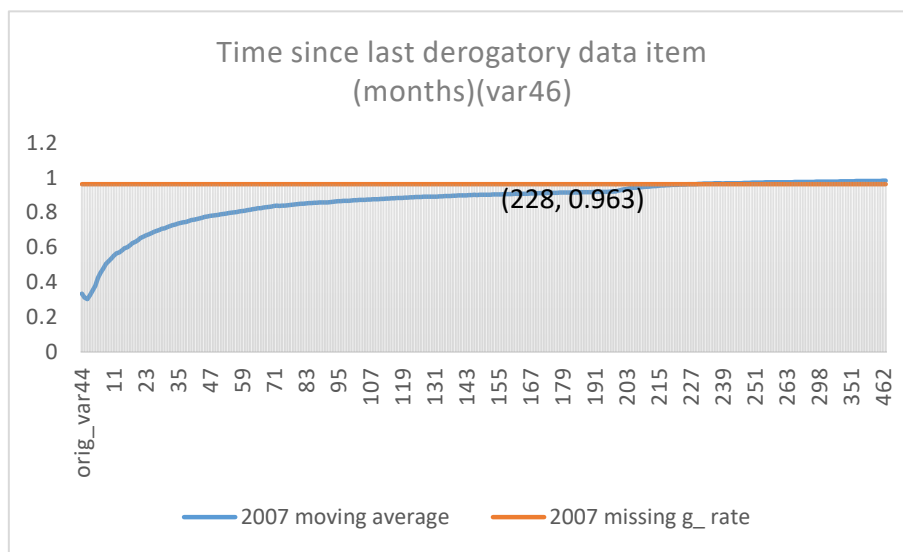


Figure 0.18 Time since last Derog: 'bad' rate of missing V.S. MA of observed value

Only one crossing point occurs for *Time since last derogatory data item*. MA's curve has an exponential-like pattern: it increases quickly at the beginning and then has an almost zero accelerate rate at the end. The missing value's performance matches with firms whose derogatory data is recorded long ago.

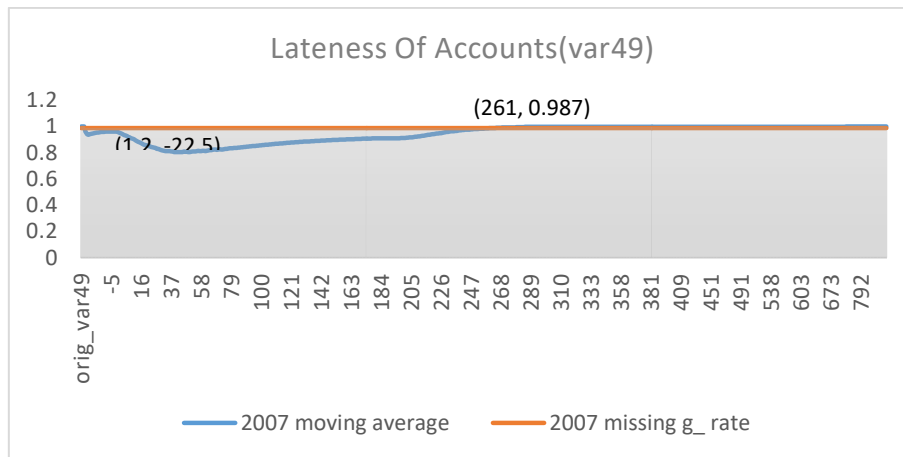


Figure 0.19 Lateness of Account: 'bad' rate of missing V.S. MA of observed value

The MA of *Lateness of Accounts* crosses with the missing category's performance twice. They start with very similar performances, then they are separated apart, crossing again in the middle and staying very close until the end. Therefore, the second crossing point is chosen to fulfil the missing for the following reasons:

1. The data is sparse around the first crossing point. Meanwhile, considerably more observations occur at the second crossing point;
2. The first crossing point is too close to an outlier on the left hand side, while the chosen point falls into the middle of the horizontal axis;

In summary, it is assumed that the missing category's performance is very similar with firms with less frequent of data updates.



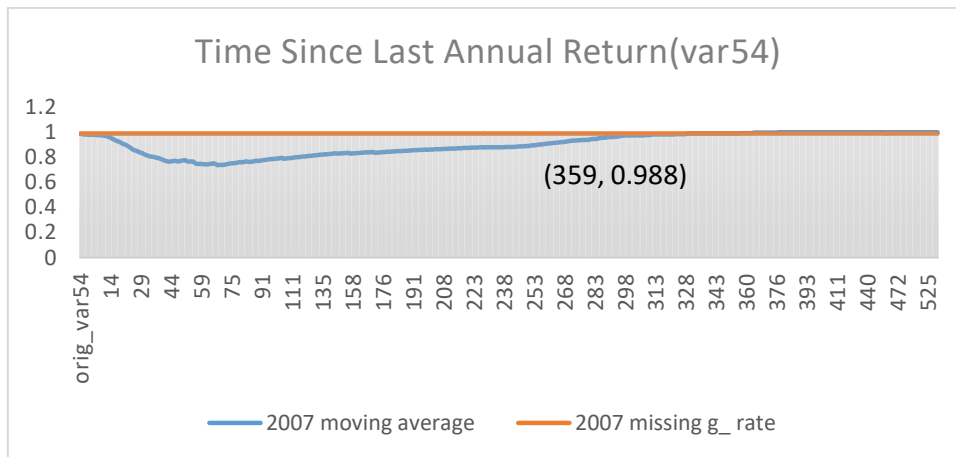


Figure 0.20 Figure 20 Time since Last Annual Return: 'bad' rate of missing V.S. MA of observed value

In both tails, the variables' MA stays very close to the missing category's performance. However, the crossing point occurs only at the right hand tail. Therefore, this research assumes the missing category's performance is similar with firms whose last annual return is reported long ago.

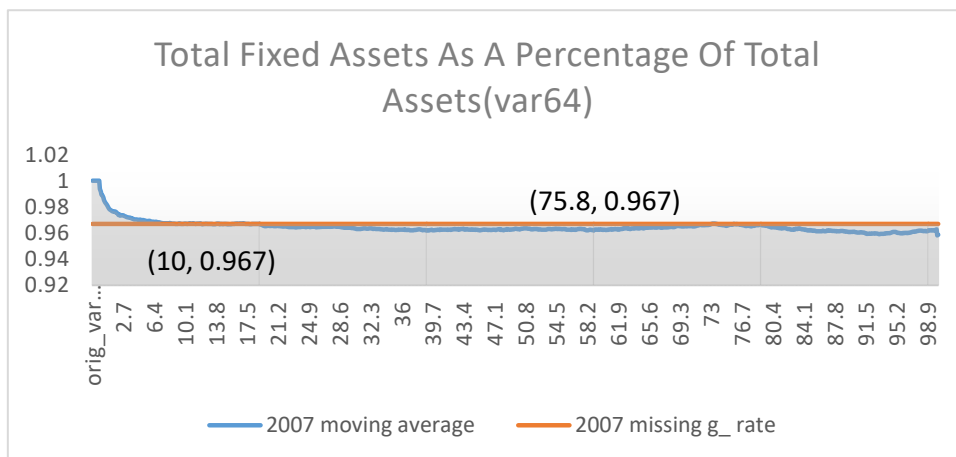


Figure 0.21 Pct. of Total Fixed Assets: 'bad' rate of missing V.S. MA of observed value

The MA decreases fast for firms with a low percentage of the fixed assets. After the MA's curve cross the missing category at the first crossing point, the two lines remain very close. Another crossing point occurs at round the tail. The average of the two crossing point is chosen as the imputation value since the MA remains relatively flat

between those two crossing points. In summary, the percentage of total fixed asset is more informative for firms with a lower proportion of fixed assets, yet the missing category performance is similar to a higher proportion of fixed assets.

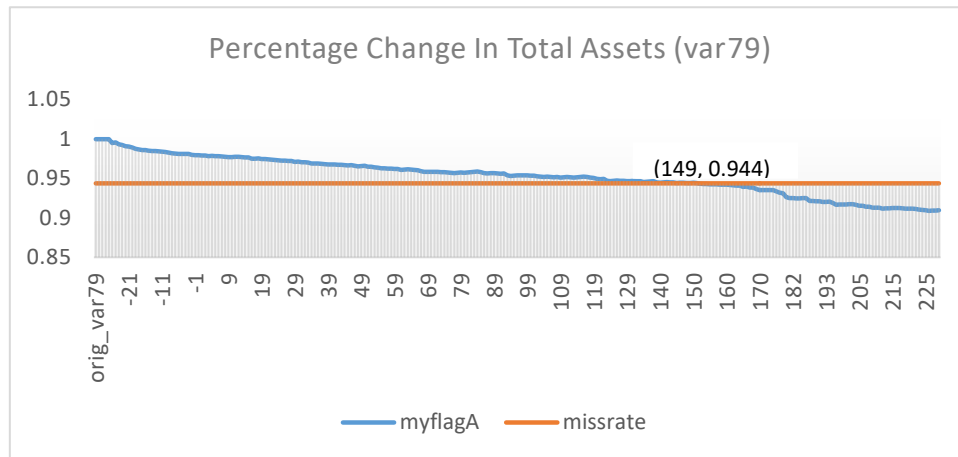


Figure 0.22 Pct. Change in Total Assets: 'bad' rate of missing V.S. MA of observed value

The MA curve decreases in an almost monotonous pattern except around the crossing point where the decreasing rate of the curve rate is far lower. This clear crossing point, which occurs at the right hand side is chosen as the imputation value for the missing category. Hence, firms with a missing value are assumed to have a similar performance to those whose total assets change is high.

In summary, in non-start-up SMEs' the missing category usually crosses the observed values' MA. Hence, the missing category could be replaced by the observed value with little bias. For the same variable, the exact imputed value may not be the same but variation is very limited even during the 'credit crunch.'

## 4.8.2 Start-up SMEs

For the ‘start-ups’ segment, the missing category is larger and the missing category’s annual variation is also more noticeable than that of ‘non-start-ups.’ The following two variables are used to demonstrate those features: the former one has high missing frequency and the later one has clear annual variation.

	2007	2008	2009	2010
Proportion of Current Directors to Previous Directors in the Last Year	81.61	93.66	94.17	

Figure 0.23 missing frequency for Proportion of current Director

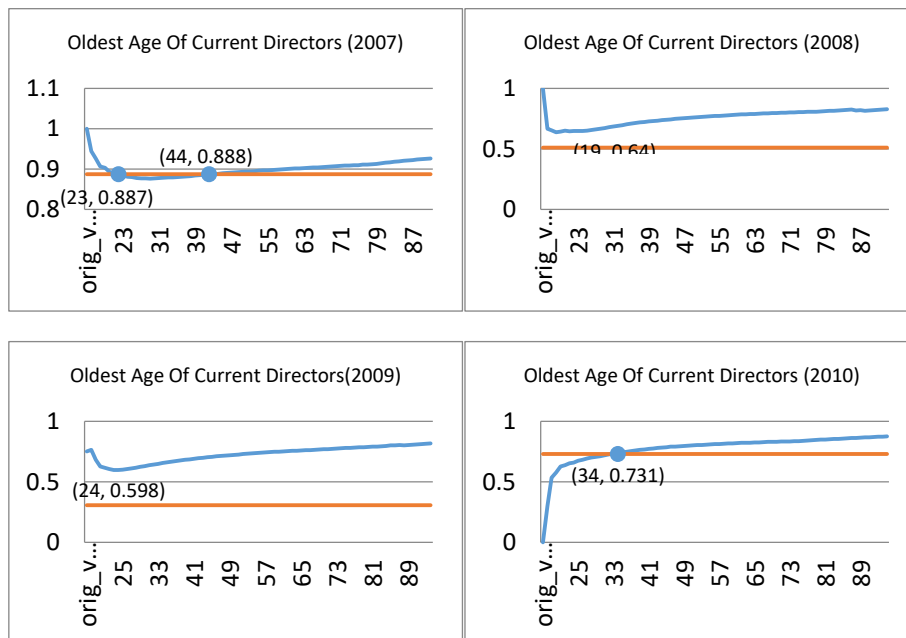


Figure 0.24 Oldest Age of Current Directors: 'bad' rate of missing V.S. MA of observed value (2007-2010)

*Proportion of Current Directors to Previous Directors in the Last Year* has a particularly large missing category. For the *Oldest Age of Current Directors*, in 2007 the missing category crossed the MA at two points. The MA curve was then higher than the missing category default rate for 2008 and 2009 with no actual crossing

existing. In these cases points with a minimum distance between the lines and MA curve are chosen as an imputed value in those two years. In summary, the missing category's performance became closer to firms with very young directors in 2008 and 2009. In 2010, MA shifted direction from younger current directors in 2010 and the crossing occurs at 34 which is close to that of 2007. The following part explains the selection of the crossing point for each variable in this segment.

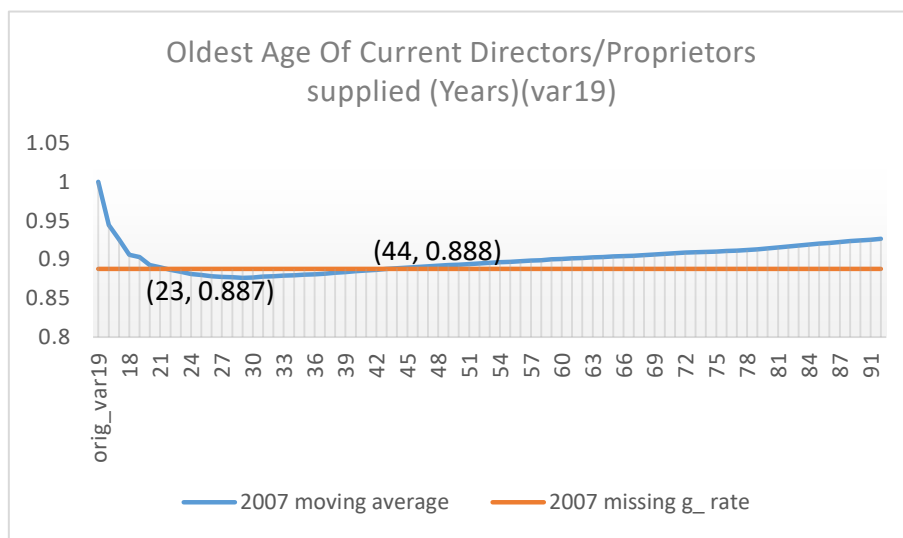


Figure 0.25 Oldest Age of Current Directors: 'bad' rate of missing V.S. MA of observed value

The MA curve of *Oldest Age of Current Directors/Proprietors Supplied (Years)* crosses the missing category's performance twice. The crossing point is close to the flat bottom of the curve of the moving average. Therefore, the average of those two crossing points is used as imputed value. The two crossing points for the oldest age of current directors is 23 and 44. After taking the average, imputed value is 33.5 which means the missing category's oldest director can be estimated as middle age.

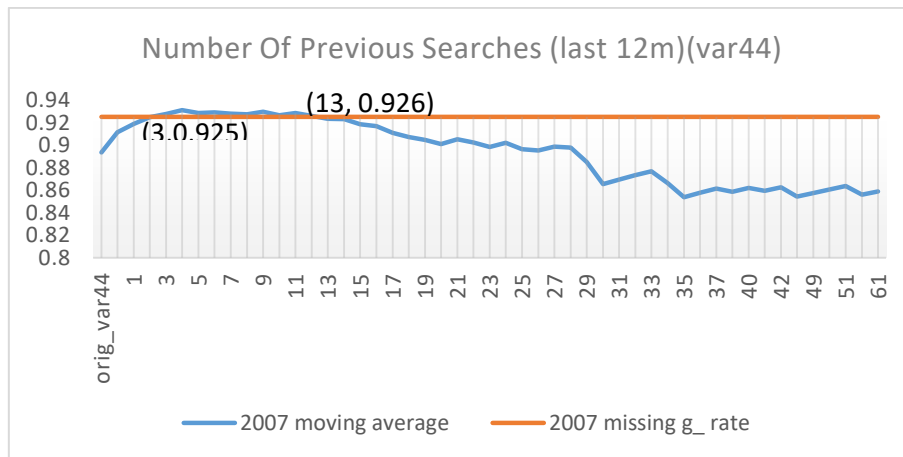


Figure 0.26 No. of Previous Searches: 'bad' rate of missing V.S. MA of observed value

The MA curve of *Number of Previous Searches (Last 12m)* also has a quadratic alike form with two crossing points and the missing category's performance is estimated as eight previous searches in the last 12 months.

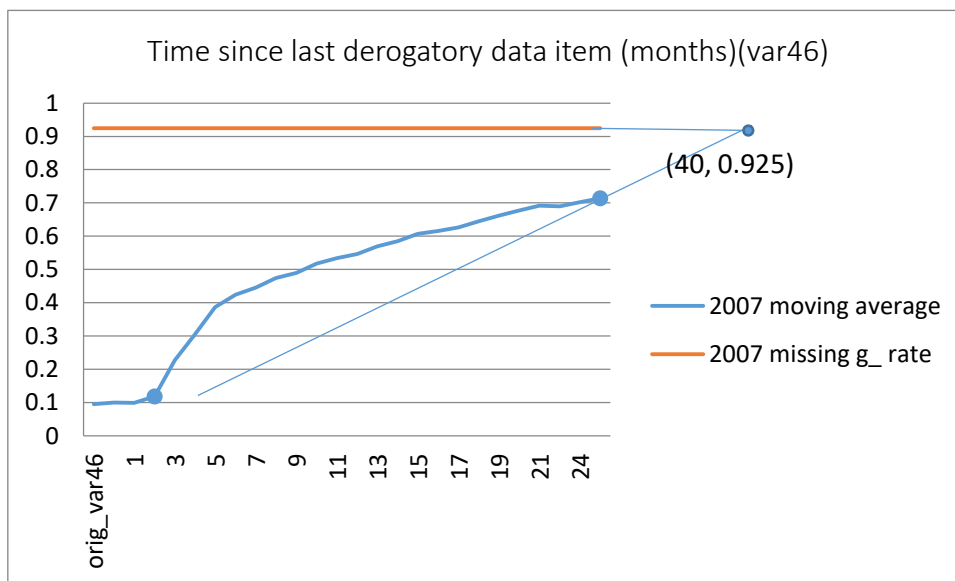


Figure 0.27 Time since Last Derog. Data: 'bad' rate of missing V.S. MA of observed value

MA of *Time since Last Derogatory Data Item (Months)* increases almost monotonously, while the missing category shows a distinct performance and is always above the MA curve. An approximated crossing point would be perceived as an

outlier. To avoid using an outlier, the maximum value is used as approximation to impute the missing value. It means the performance of missing category is mimicked by the longest last derogatory data.

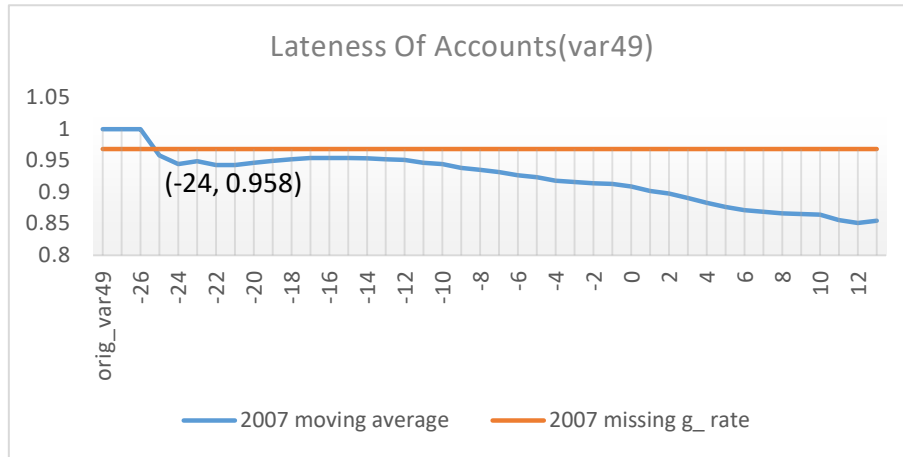


Figure 0.28 Lateness of Accounts: 'bad' rate of missing V.S. MA of observed value

For *Lateness of Accounts*, the only one crossing point is taken as the imputed value. Negative value of *Lateness of Accounts*, where it matches the missing category's performance, means that the firm's account is not available for the corresponding months.

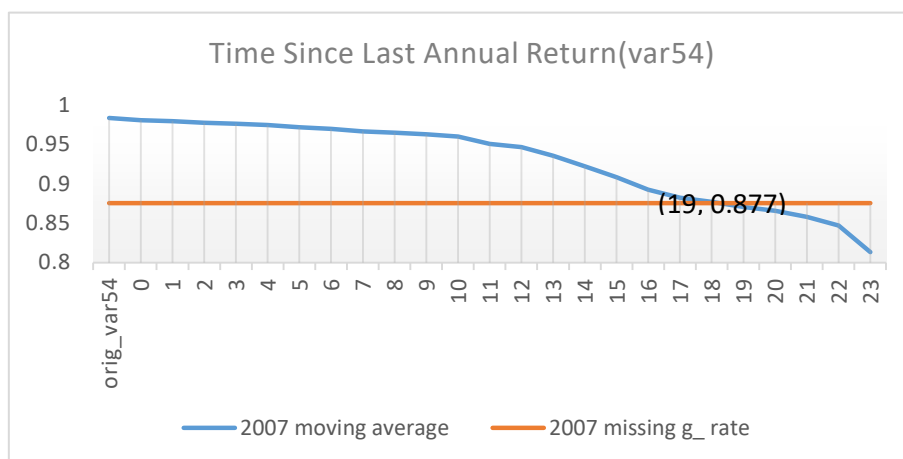


Figure 0.29 Time since last Annual Return: 'bad' rate of missing V.S. MA of observed value

One clear crossing point is found to fulfil the missing category for *Time since Last Annual Return*. Therefore, firms missing *Time since Last Annual Return* have a similar performance to those which reported their last annual return long ago.

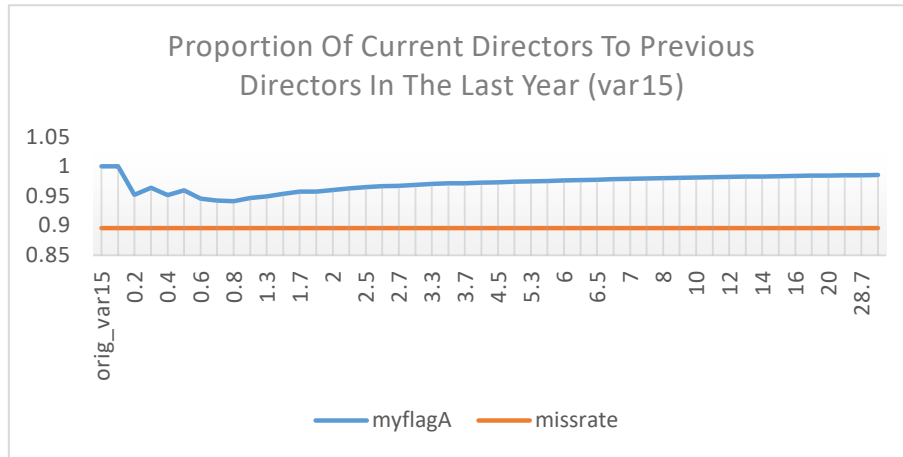


Figure 0.30 Prop. of Current Directors: 'bad' rate of missing V.S. MA of observed value

No real crossing point occurs for *Proportion of Current Directors to Previous Directors in the Last Year*. The point where the line and the MA curve are the smallest distance apart is chosen to approximate the performance of missing category.

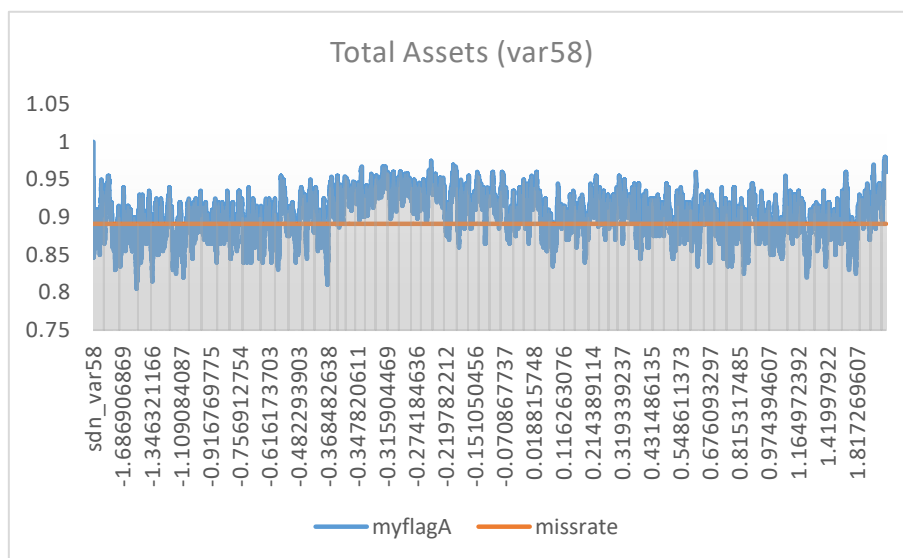


Figure 0.31 Total Assets: 'bad' rate of missing V.S. MA of observed value

---

The original value of *Total Assets* covers a very large range and 79.33% of *Total Assets* values are missing value. Owing to the volatility of the moving average and the size of missing category, WoE is used for *Total Assets* in order to avoid more noise.

In summary, the missing category's performance of start-up SMEs could not be easily replaced by observed values and the missing category's performance is less stable compared to 'non-start-ups.' Approximation is used when no exact crossing exists. However, as more approximations are used in this segment, there is a potential loss of information which could cause reduction in the predicted accuracy when using variables in their original format.

## **4.9 Conclusion**

This chapter provides the initial statistics to describe SMEs performance during the 'credit crunch.' The SMEs performance changes have been presented according to industry, region and other factors during the financial crisis. These statistics have clearly shown that 'start-ups' and 'non-start-ups' have distinct performances throughout time. By dividing them into two segments, this research identifies the segmentation difference and provides more accurate credit risk analysis. The *lateness of account* has clearly demonstrated that SMEs data are sparse and could only be fitted by discrete time models. By comparing different variables' missing categories, the author also pointed out that missing value could not be deleted since not only a



---

majority of SMEs suffer from failing to provide some information, but also the missing value itself is a very informative category.

The quality of employed data guarantees the observations are well supported by the empirical evidence and valid for a wide range of UK SMEs. Previous research, for example Altman and Sabato only analysed 2010 firms in their research (Altman and Sabato, 2007), used small sample sets, and therefore their conclusions are questionable owing to potential selection bias and model accuracy also being challenged by low frequency in analysed categories.

This chapter also provides the variable selection process used in this research. The subjective variable selection process controls the collinearity among independent variables and helps to build the most efficient forecasting PD model. The last section of this chapter discussed the process of substituting missing value with observed values. Therefore, this chapter completes data preparation and the following chapter will present results and findings.

---

## 5. Results

This Chapter gives a thorough presentation of results and findings of this research. Several pieces of software have been used in this research to achieve the best results of employed models. Logistic models and GAM are implemented in SAS, while STATA is used for panel data models. STATA is better designed for logit panel data model with unbalanced data. Separation measures are produced in R with the H measure package which summarise AUROC, Gini, KS-statistic and H measure. This package is developed by Anagnostopoulos and Hand (2012). It has three sections which shows results from the logistic model, panel model and GAM correspondingly. Logistic regression is the benchmark model and shows how well the current industry standard model can respond to the ‘credit crunch’. With the help of panel data model, the author involves time series effect into SMEs modelling. The second section shows detailed results of panel data modelling and how the improvement is achieved by the addition of time series effects using micro-economic variables (MVs). Additionally, it also explains how the MVs influence SMEs performance during the ‘credit crunch’. The last model used is GAM which address the non-linear behaviour of SME’s performance. After imputing missing values, original data values are used in GAM and independent variables trends are fully explored to increase the model transparency.

## 5.1 Logistic regression

As previously mentioned, variables which are significant for three years or more are selected for modelling, and the same set of variables are used for each segment in different models to obtain a comparison between cross-sectional models over four years: non-start-up segment has 15 variables and that number for start-up SMEs' is 12. Among those selected variables, 9 variables are commonly used for both segments. Coefficients and its significance is shown for two segments. Table 5.1 presents 'start-ups' results:

start-ups				
	2007	2008	2009	2010
Intercept	2.15***	1.32***	1.08***	1.78***
	25409.15	11686.90	5255.39	9104.01
Legal Form	1.82***	1.85***	1.90***	2.42***
	713.07	1094.13	1128.12	740.56
1992 SIC Code	1.67***	0.22***	-0.01	0.22***
	867.59	72.50	0.05	11.23
Region	0.60***	0.59***	0.59***	0.76***
	78.66	338.06	86.68	67.92
Proportion Of Current Directors To Previous Directors In The Last Year	0.74***	0.17**	0.47***	0.37**
	159.58	4.57	34.94	5.51
Oldest Age Of Current Directors/Proprietors supplied (Years)	0.71***	0.51***	0.55***	-0.07
	151.59	617.95	517.10	0.84
Number Of Directors Holding Shares	0.33***	0.14***	0.20***	0.23***
	32.69	42.52	46.93	10.18
Total Value Of Judgements In The Last 12 Months	0.59***	0.66***	0.73***	0.47***
	88.53	51.40	139.43	57.40
Number Of Previous Searches (last 12m)	1.21***	0.69***	0.60***	0.71***
	337.56	435.25	227.53	140.06
Time since last derogatory data item (months)	0.66***	0.62***	0.55***	0.64***
	1832.81	1720.41	2309.14	1752.58
Lateness Of Accounts	1.69***	0.83***	0.38***	0.37***
	3798.06	1615.47	678.10	322.21
Time Since Last Annual Return	0.72***	0.67***	0.49***	0.49***
	766.05	2233.70	1251.83	563.51
Total Assets	0.08*	0.17***	0.25***	0.46***
	3.55	62.30	219.50	220.67

---

Table 5.1 logistic regression results for start-ups

Different significant level is represented by stars following each coefficient: \*90%, \*\*95%, \*\*\*99.9%

The coefficient of each variable is listed first and its significance is marked by stars and chi-square statistic is given below the coefficient.

Using a significance level of 95%, only three variables have insignificant coefficients: *total assets* in 2007, *SIC* in 2009 and *oldest age of current directors* in 2010. The chi-square statistic for *SIC* code starts with a very high absolute value in 2007, then it drops significantly in 2008 and loses significance in 2009, but it regains significance in 2010. Therefore, the changing economy has an impact on the effect of the industry performance and industry classification becomes insignificant during the ‘credit crunch’. Hence whilst generally *SIC* has an effect on performance it disappears during the height of the crisis.

*Oldest Age of Current Directors/Proprietors supplied (Years)* has the highest chi-square statistic during the ‘credit crunch’, however, it loses significance in 2010. In summary, although the directors knowledge could help the ‘start-ups’ to survive during the ‘credit crunch’, yet, their experience loses its advantage beyond the peak of the ‘credit crunch’.

*Total Assets* is not significant in 2007 but becomes significant from 2008. It means for ‘start-ups’ their assets size is not a significant explanatory variable of default in the normal economic period. However, when the ‘credit crunch’ hits the ‘start-ups’ *Total*

---

*Asset* becomes much more influential and plays a significant role in helping SME's survive during the changing economy.

*Proportion of Current Directors to Previous Directors in the Last Year* has the highest chi-square value in 2007 and 2009 and lower value in 2008 and 2010. This variable demonstrates the change of board and is most significant when the economy situation is relatively stable, regardless the direction, such as 2007 and 2009. However, when the economy changes, the change causes more noise and makes the board change less important.

The following table provides the results for 'non-start-ups':

non-start-ups				
	2007	2008	2009	2010
Intercept	3.13***	1.90***	0.53***	0.28**
	29675.37	7255.66	129.71	10.03
Legal Form	0.51***	0.59***	0.47***	0.15*
	22.57	40.90	49.08	2.63
1992 SIC Code	0.54***	0.50***	0.47***	0.52***
	109.10	177.66	315.61	152.05
Region	0.05***	0.28***	0.21***	0.13***
	0.20	8.96	18.47	1.20
No. Of 'Current' Directors	0.40***	0.30***	0.25***	0.27***
	83.96	81.69	84.51	45.53
Proportion Of Current Directors To Previous Directors In The Last Year	0.40***	0.59***	0.28***	0.59***
	16.07	45.06	8.62	29.57
PP Worst (Company DBT - Industry DBT) In The Last 12 Months	0.42***	0.31***	0.28***	0.47***
	65.26	45.48	62.22	120.10
Total Value Of Judgements In The Last 12 Months	0.41***	0.37***	0.44***	0.36***
	149.23	89.22	97.30	69.70
Number Of Previous Searches (last 12m)	1.29***	0.73***	0.41***	1.00***
	70.14	79.20	60.49	115.19
Time since last derogatory data item (months)	0.35***	0.62***	0.65***	0.62***
	160.90	1751.42	3426.08	2047.53
Lateness Of Accounts	0.65***	0.60***	0.47***	0.43***
	1408.03	1642.83	1222.93	541.03
Time Since Last Annual Return	0.65***	0.61***	0.50***	0.54***
	898.80	1262.37	1196.36	666.75
Total Fixed Assets As A Percentage Of Total Assets	0.74***	0.64***	0.48***	0.48***
	143.72	340.42	383.34	110.63
Debt Gearing (%)	0.40***	0.34***	0.40***	0.39***
	10.38	10.80	18.46	11.04
Percentage Change In Shareholders Funds	0.46***	0.30***	0.33***	0.21***
	124.32	74.47	212.53	26.36
Percentage Change In Total Assets	0.64***	0.60***	0.49***	0.55***
	306.55	455.86	422.45	263.12

Table 5.2 logistic regression results for non-start-ups.

Different significant level is represented by stars followed: \*90%, \*\*95%, \*\*\*99.9%

Collinear variables have generally been removed, and no negative correlation exists for 'non-start-up' SMEs. The only noticeable change is that *legal form* loses its significance at level 95% in 2010. Hence there is more stability amongst the variables

---

for ‘non-start-up’ segments and ‘non-start-ups’ performance is more predictable during the ‘credit crunch’.

In addition, SME’s coefficients change during the ‘credit crunch’. Therefore, a single logistic model would not fit the SMEs performance through time. The limitations of logistic model means that one could not explore further the influence of changing economy and details of variables influence. Logistic regression model is also used as a benchmark, therefore its separation measures are presented in the following chapter to demonstrate other models separation ability. Separation measures are more meaningful to give comparison of different models.

The transformation by WoE means it is difficult to interpret the influence of independent variables directly indicated by the corresponding coefficients. Further discussion of independent variable’s trend is even less clear due to the same reason. The next section will explain how to use panel data model to build one model through the crisis and explore the MVs’ influence. The third section will present the use of GAM which provides more insight into the detailed trend of each variable.

## **5.2 Panel data**

### **5.2.1 Introduction:**

The previous section presents logistic model for ‘start-ups’ and ‘non-start-ups’. Although the models can fit the performance well, there remains several issues to resolve. First of all, the single time period models do not allow time series effect,

---

therefore such influences would be lost and forecasting accuracy may be reduced. Secondly, in such a dramatic economic switching period, SME's face a hard time in such a tough environment and may have greater difficulty surviving. However, using the firm specific variables only, one is actually assuming that the SME's performance is independent of the business cycles. This assumption is challenged by the 'credit crunch'.

To solve those disadvantages encountered in logistic regression, this research develops SME's credit risk models to multi-time period models. As discussed in the Methodology Chapter, panel data is used due to the short and discrete time line of SME's data available during the 'credit crunch'. Previously, a thorough discussion of specification choice between fixed effect and random effect has been given. Initially, panel data models only use firm specific variables as a logistic model. Time and obligor are identified by year and customer reference number respectively. Coefficients for 'start-ups' and 'non-start-ups' are presented in the first column of Table 5.5 and Table 5.6 correspondingly, while separation measures of those models for training samples and holdout samples are given in Table 5.7 and Table 5.8.

As discussed in the Methodology Chapter, temporary correlation could cause misleading results in panel data analysis. It is especially important to consider it in this research since the observed period contains such a significant macroeconomic switch. Additionally, 'bad' rates as well as statistics of explanatory variables shift sharply from year to year. As a result, several coefficients of 'non-start-ups' show



---

unexpected coefficients. The annual difference during this period will be dealt with by the use of dummy variables or MVs, as discussed in following sections.

### **5.2.2 Adding year dummies**

Results are presented in second column of Table 5.5 and Table 5.6 for ‘start-ups’ and ‘non-start-ups’ correspondingly. Not only are the time dummies significant throughout the ‘credit crunch’ for both segments, but also the separation accuracy has been improved. This means if one built a panel data model with firm specific variables only, their results could be misleading given the interaction of temporary dependence. Surprisingly, the influence of annual difference is especially important for ‘non-start-up’ segments, instead of ‘start-ups’, as several variables switch sign after controlling the annual change and the separation accuracy has been improved much more than for ‘start-ups’ segment.

However, time dummies have a clear disadvantage: “we cannot predict the value of next period” (Beck, 1998). Therefore, the following section discusses the use of macroeconomic variables as an alternative way of controlling temporary differences.

### **5.2.3 Adding Macroeconomic Variables**

Another way to control annual shift is to introduce macroeconomic variables (MVs). MVs can replicate market movements and give a clear explanation on how SMEs performance are influenced by the economic dynamics during the past crisis. As mentioned in the previous section, GDP growth rate, CPI, unemployment rate, FTSE

---

all share index and FTSE 100 index are employed to represent three major aspects of economics: direction of economy, general economic condition and the financial market movements.

Although the most related MVs are carefully chosen, MVs covariance cannot be totally eliminated. Following the framework of Friglewski *et al*, this research inputs MVs into the model one at time to test their influence on SME's performance. This procedure provides clear evidence on how each MVs influence SMEs' performance without uncertainty caused by interaction among MVs. Using AIC, MVs with the best performance are selected to present different aspects of economies (Friglewski et al's 2012). Then, selected MVs are added to the model to replicate the economic condition and improve the estimation accuracy.

#### **i)Lags of MVs**

An essential issue in adding control variables is how to choose the lag of MVs. As previously discussed, time averaged MVs are used to give the best control of economic influence. However, to provide more empirical evidence about MVs influence, different lags of MVs are used in modelling to show how MVs' influence may vary as different lags are used. In addition, the different effects received from different lags also support the use of time averaged MVs as it contains all past information.

For example, during the financial crisis annual shift is particularly sizeable, i.e. 2009' 'bad' rate reaches 14.68% yet 2010's 'bad' rate drop sharply to 8.5%. Therefore, this

synchronous movement makes non-lagged MVs attractive as they can control the sizable annual economic changes. Also, some other economic variables such as unemployment rate can influence SME's with a lag as unemployed labour could in turn establish SME's. This research tested the influence of non-lagged MVs, one year's lagged MVs and MVs cumulative effect. Results reported that not only can it explain MVs influence on SMEs during the past 'credit crunch', but also provide evidence on how to choose MVs lag for future studies.

**ii) The magnitude of MVs' effect**

This section discusses the sign of MVs when adding one MV at a time. All of the listed MVs are significant in this procedure. Therefore, MVs control the annual difference of SME's performance caused by a macroeconomic switch. Results are listed as following:

Start-up SMEs			
	no lag	one year's lag	weighted average of lagged MVs
GDP growth rate	+	+	+
unemployment	-	+	-
CPI	-	-	-
FAI	+	+	+
F1I	+	+	+
Non-start-up SMEs			
	no lag	one year's lag	weighted average of lagged MVs
GDP growth rate	+	+	+
unemployment	-	-	-
CPI	-	-	-
FAI	-	+	+
F1I	-	+	+

*Table 5.3 MVs sign for both segments when adding one MV to the model*

- 
- Economic direction: GDP growth rate

For both segments, GDP growth rate constantly shows a positive correlation with the SMEs performance, no matter which lag length is used. As GDP growth rate reflects the economic direction, the results mean that the SMEs performance is improved if the economy is strong, while the downward economic condition pulls down those enterprises' performance. This research provides clear evidence that SME's performance is positively correlated with the direction of macroeconomics.

- Economic conditions: unemployment rate and CPI

CPI growth rate and unemployment rates represent general economic conditions. Generally speaking, they are negatively related with the 'good' rates of SME's.

An interesting result is that one year lagged unemployment rates has a positive effect on 'start-ups'. One possible explanation is the movement from unemployed labour to entrepreneurs. As the high unemployment rate keeps rising during business downside, more experienced and professional people lose their jobs and turn to starting their own business. These people's skills and experience helps their 'start-ups' to survive. This phenomena may result in a boom on SME's performance.

However, as the cumulative effect of unemployment rates presents a negative coefficient, it implies that the influence of the lagged effect is not as strong as its on-time influence. This finding also criticizes previous research, such as Bellotti and Crook (2007) which only considers MVs effect without lag. Malik and Thomas

---

follow the Figlewski's framework considering the weighted average of MVs. However, their focus is on consumer loans. This research provides results for SME's credit risk and presents how different lag could result in contradictory influences.

Theoretically, CPI could affect enterprises performance in either direction, as inflation can boost the worthiness of their business, increase their cost and decrease customer's buying power. This research shows that CPI has influenced SME's in a negative manner during the past financial crisis regardless which lag is used. Therefore, SME's suffer more from their increasing costs and loss of customers from the 'credit crunch'.

- Financial markets

Similar to general economic conditions, financial market's influence is hard to predict by theory. On one hand, if financial market returns increase, it can benefit listed SMEs. However, high financial market returns will cause investment flows to financial markets and reduce the investment in loan issuing. Additionally, negative financial market returns will lead to more prudent lending strategies by the supervisors which also decreases financial accessibility for SMEs.

For the sake of financial market variables, this research tested the FTSE 100 index (F1I) and FTSE All-Share Index (FAI). Financial market performance is always positively correlated with start-up SME's performance. However, for non-start-up SME's, the influence is not consistent with different time lags. When no time lag is considered, financial markets present a negative coefficient with 'non-start-ups'

performance. This result indicates that during a bear market ‘non-start-ups’ are benefited by the flow of investments and receive better financial support which improves their performance. Its magnitude changes when one year’s lag is taken. It means the investment flow only influences ‘non-start-ups’ in a short time frame, for a longer time period, the falling financial markets has a negative effect on ‘non-start-ups’. When adding those effects on a weighted average form, the benefit of incoming cash flow is covered by the negative effects of falling financial markets.

### iii) Variable selection for best performance model

Akaike information criteria (AIC) is used to choose MV to represent each economic aspect. AIC shows the model fit by taking into account the additional parameters. The best model is the one with lowest AIC value. Models’ AIC values are listed below:

	Start-up SMEs	Non-start-up SMEs
GDP growth rate	152299.23	130047.09
unemployment	153025.14	132925.46
CPI	151206.19	132576.57
FAI	149914.67	134387.58
FII	149649.96	133325.95

*Table 5.4 AIC value when adding one averaged MV into panel model*

Due to the large sample size, AIC presents a high value for all of those models. In general, non-start-up SME’s models have a lower AIC value than ‘start-ups’. The following part will discuss MVs influence by category and select one MV for each category.

The influence of different time lagged MVs are discussed in the previous section showing how MVs influence can change overtime. However, as this section aims to

---

find the best fitted models, MVs with cumulative time effects are used since this form contains more information to improve the model fitting.

GDP growth rate is chosen to indicate the direction of economics. GDP growth rate is the most influential variable to improve AIC value for 'non-start-ups' segments, but not for 'start-ups'. Therefore, the economic direction has a stronger influence on mature companies. The economic condition MVs, such as unemployment rate and CPI, have similar AIC for 'non-start-ups', however, the unemployment rate has the weakest ability to improve the fit for 'start-ups'.

As mentioned previously, the unemployment rate could influence the SME's performance in contradictory ways: on one hand, high unemployment rate indicates worse economic conditions; on the other hand, more professionals could leave their jobs and start their own business. The mixture of both influences may reduce the separation ability of unemployment rate for 'start-ups'. Therefore, CPI is chosen since it always provides a better performance.

The third category is that the financial market variables which group has the lowest AIC value for 'start-ups'. Hence, 'newly established firms are heavily affected by the financial markets' movements. FTSE 100 is the most widely used variable to represent financial market movements which contain large global companies, whilst FTSE all share index contain all firms listed, including some SMEs'. However, the results show that both indexes always influence SME's with the same magnitude and FTSE 100 index always has a better model performance with a lower AIC value.

---

Therefore, GDP growth rate, CPI and FTSE 100 share index are chosen to replicate the macroeconomic situation during the ‘credit crunch’ and to improve the panel data models.

Section 5.2.3 provides a detailed discussion on how MVs influence SME’s performance during the ‘credit crunch’. MVs influence may change as different time lags is considered and their cumulative effects provide the most information to improve credit risk models. For ‘start-up’ SME’s, financial market variables influence their performance most significantly, while the least influential variable is unemployment rates. Meanwhile, it is the GDP growth rate that gives the most information to improve models for ‘non-start-ups’, yet financial market MVs seem to be less appealing for this segment.

#### **5.2.4 Results for ‘start-up’ SMEs**

This section and the following section will discuss fitted coefficients and panel data models’ separation measures for ‘start-ups’ and ‘non-start-ups’. Coefficients of different panel models are listed for start-up SME’s in Table 5.5, while their separation measures are listed in Table 5.7 and Table 5.8 for training sample and holdout sample respectively. Several measures can be used to demonstrate the credit scoring models separation ability, such as Kolmogorov–Smirnov statistic, Gini coefficient, area under the receiver operating characteristic curve (AUROC) and the



---

newly established Hand's measure. This research mainly focusses on the most widely accepted measure AUROC.

For 'start-ups', even when using firm specific variables only, there is no explanatory variable negatively correlated with the dependent variable. It suggests the temporary dependence caused by the 'credit crunch' is less significant for 'start-ups'. In addition, different panel data models provide very similar results. In addition, although MVs are tested significantly in 'start-ups' PD model, they do not have a strong impact on separation measures. Logistic regression performs better than panel models. Hence, the current industry standard model is robust even when the economy is in deep recession. As logistic models only consider the cross-sectional difference, it indicates that time series effect does not have a significant influence on improving 'start-ups' credit scoring.

When comparing different panel models, little improvement is found when dummies or MVs are added to control the influence of economy. It further supports that 'start-ups' performance during the 'credit crunch' can be well explained by their firm specific variables without considering the economic conditions.

### **5.2.5 Results for 'non-start-up' SMEs**

For 'non-start-up' SME's, their coefficients of different panel models are listed in Table 5.6 and their separation measures are listed in Table 5.7 and Table 5.8 for training samples and holdout samples respectively. In this segment, five out of fifteen

---

independent variables exhibit negative coefficient when using firm specific variables only. Hence, temporary dependence is more significant for ‘non-start-ups’, which means ‘non-start-ups’ are significantly influenced by economic switches for firms which have been established for a longer time.

Compared to the panel model with firm specific variables, panel data models with dummies or MVs can significantly improve model fitting for ‘non-start-up’ credit scoring. In addition, logistic regression underperforms panel models with dummies or MVs for the holdout sample. Hence, unlike ‘start-ups’, non-start-up’ performance could not be fully explained by firm specific variables, and temporary dependence could be explained by MVs during the past ‘credit crunch’. To conclude, MVs directly influence the ‘non-start-ups’ performance and cannot be eliminated in PD models.

In section 5.2, panel data results are presented to help demonstrate how MVs influence the ‘start-ups’ and ‘non-start-ups’ performance during the ‘credit crunch’.

The results are summarized as follows:

1. Panel data model should not use firm specific variables only, especially for ‘non-start-ups’ as firm specific variables could not control temporary dependence. Panel data model is more suitable for ‘non-start-ups’. Considering the particular time period this research focuses on, ‘start-ups’ performance can be well explained by cross-sectional analysis without considering time-series effects. Temporary dependence caused by the ‘credit crunch’ could be controlled either by

---

time dummies or sets of MVs. In addition, MVs enjoy the advantage of providing more of an explanation and forecasting future performance.

2. All MVs are significant when added to panel models for both segments. A selected set of MVs in a panel model can replicate influences from the crisis. It is the GDP growth rate which has the most significant influence for ‘non-start-ups’, however, it is financial market variables for ‘start-ups’.

For reference, panel data results are presented in the following tables.

Start-up SMEs			
	Panel model with firm specific variables only	Panel model adding year dummies	Panel model adding set of MVs
Intercept	1.65*** 239.71	2.56*** 84.81	-12.18*** -23.46
Legal Form	1.96*** 57.41	2.24*** 51.54	2.24*** 51.54
1992 SIC Code	0.29*** 17.07	0.28*** 14.77	0.28*** 14.77
Region	0.66*** 25.42	0.70*** 23.54	0.70*** 23.54
Proportion Of Current Directors To Previous Directors In The Last Year	0.66*** 18.39	0.68*** 15.83	0.68*** 15.83
Oldest Age Of Current Directors/Proprietors supplied	0.49*** 32.92	0.58*** 30.76	0.58*** 30.76
Number Of Directors Holding Shares	0.14*** 8.62	0.20*** 10.65	0.20*** 10.65
Total Value Of Judgements In The Last 12 Months	0.63*** 18.92	0.72*** 18.34	0.72*** 18.34
Number Of Previous Searches (last 12m)	0.78*** 33.46	0.81*** 30.97	0.81*** 30.97
Time since last derogatory data item (months)	0.63*** 93.74	0.76*** 70.19	0.76*** 70.19
Lateness Of Accounts	0.64*** 70.32	0.71*** 55.52	0.71*** 55.52
Time Since Last Annual Return	0.61*** 72.49	0.69*** 60.66	0.69*** 60.66
Total Assets	0.15*** 12.68	0.19*** 14.2	0.19*** 14.2
y2		-1.14*** -43.5	
y3		-1.51*** -41.8	
y4		-0.82*** -21.13	
GDP growth rate _ weighted average			1.02*** 28.74
unemployed rate _ weighted average			2.19*** 26.27
FTSE all-share index _ weighted average			0.02*** 14.24

Table 5.5 start-up SMEs random effect panel data models parameter estimation

Note: 1. \*\*\*refers to significant at level 1%; 2. estimated coefficient is listed first and the z-statistic is listed below.

Non-start-up SMEs			
	Panel model with firm specific variables only	Panel model with year dummies	Panel model with set of MVs
Intercept	2.51*** 238.14	3.53*** 92.04	-11.87*** -24.26
Legal Form	-0.22*** -8.16	0.47*** 9.94	0.47*** 9.94
1992 SIC Code	0.41*** 23.29	0.55*** 26.21	0.55*** 26.21
Region	-0.81*** -24.04	0.20*** 4.53	0.20*** 4.53
No. Of 'Current' Directors	0.47*** 28.32	0.35*** 17.48	0.35*** 17.48
Proportion Of Current Directors To Previous Directors In The Last Year	-0.69*** -21.37	0.49*** 9.5	0.49*** 9.5
PP Worst (Company DBT - Industry DBT) In The Last 12 Months	0.25*** 11.9	0.41*** 16.82	0.41*** 16.82
Total Value Of Judgements In The Last 12 Months	0.08*** 3.96	0.47*** 20.95	0.47*** 20.95
Number Of Previous Searches (last 12m)	-0.35*** -9.87	0.65*** 15.17	0.65*** 15.17
Time since last derogatory data item (months)	0.62*** 88.9	0.72*** 63.48	0.72*** 63.48
Lateness Of Accounts	0.54*** 71.55	0.64*** 56.38	0.64*** 56.38
Time Since Last Annual Return	0.55*** 62.85	0.63*** 56.55	0.63*** 56.55
Total Fixed Assets As A Percentage Of Total Assets	0.50*** 26.89	0.62*** 29.27	0.62*** 29.27
Debt Gearing (%)	-1.04*** -29.1	0.44*** 7.4	0.44*** 7.4
Percentage Change In Shareholders Funds	0.31*** 20.33	0.33*** 18.46	0.33*** 18.46
Percentage Change In Total Assets	0.60*** 40.69	0.63*** 37.14	0.63*** 37.14
y2		-1.36*** -45.97	
y3		-3.18*** -60.07	
y4		-3.34*** -52.42	
GDP growth rate_weighted average			1.62*** 43.58
unemployed rate_ weighted average			2.03*** 25.91
FTSE all-share index_weighed average			0.1*** -48

Table 5.6 non-start-up SMEs random effect panel data model parameter estimation

Note: 1. \*\*\*refers to significant at level 1%; 2. estimated coefficient is listed first and the z-statistic is listed below.

3For each variable estimated coefficient is listed first and the z-statistic is listed below.

**Separation measures**

year	Model	start-ups				non-start-ups			
		H	Gini	AUC	KS	H	Gini	AUC	KS
2007	logistic regression	0.33	0.654	0.827	0.523	0.164	0.639	0.819	0.485
	panel data	0.312	0.642	0.821	0.519	0.154	0.603	0.802	0.458
	panel+dummies	0.313	0.644	0.822	0.526	0.162	0.639	0.819	0.483
	panel +set of averaged MVs	0.313	0.644	0.822	0.526	0.162	0.639	0.819	0.483
2008	logistic regression	0.431	0.736	0.868	0.602	0.311	0.709	0.855	0.553
	panel data	0.425	0.731	0.866	0.599	0.297	0.686	0.843	0.539
	panel+dummies	0.425	0.732	0.866	0.599	0.311	0.71	0.855	0.554
	panel +set of averaged MVs	0.425	0.732	0.866	0.599	0.311	0.71	0.855	0.554
2009	logistic regression	0.554	0.805	0.902	0.697	0.491	0.788	0.894	0.641
	panel data	0.548	0.802	0.901	0.695	0.48	0.775	0.888	0.633
	panel+dummies	0.55	0.803	0.901	0.696	0.49	0.786	0.893	0.641
	panel +set of averaged MVs	0.55	0.803	0.901	0.696	0.49	0.786	0.893	0.641
2010	logistic regression	0.412	0.705	0.852	0.573	0.327	0.718	0.859	0.558
	panel data	0.398	0.699	0.85	0.568	0.316	0.7	0.85	0.544
	panel+dummies	0.412	0.705	0.852	0.573	0.326	0.72	0.86	0.562
	panel +set of averaged MVs	0.399	0.7	0.85	0.568	0.326	0.72	0.86	0.562

*Table 5.7 model fittings of training sample*

**Separation measures**

year	Model	start-ups				non-start-ups			
		H	Gini	AUC	KS	H	Gini	AUC	KS
2010	logistic regression	0.412	0.705	0.852	0.573	0.323	0.674	0.837	0.524
	panel data	0.398	0.699	0.85	0.568	0.314	0.657	0.828	0.51
	panel+dummies	0.399	0.7	0.85	0.568	0.326	0.685	0.843	0.529
	panel +set of averaged MVs	0.399	0.7	0.85	0.568	0.326	0.685	0.843	0.529

*Table 5.8 model fittings of holdout sample*

### 5.3 Residual check

As discussed in the Methodology Chapter, if panel data model could fully capture SMEs credit risk, the residual would follow a normal distribution. In this section, histograms and QQ plot are presented for the estimated residuals contrast against a normal distribution. As four year residuals have similar patterns, only 2007's figure is demonstrated in this data feature.

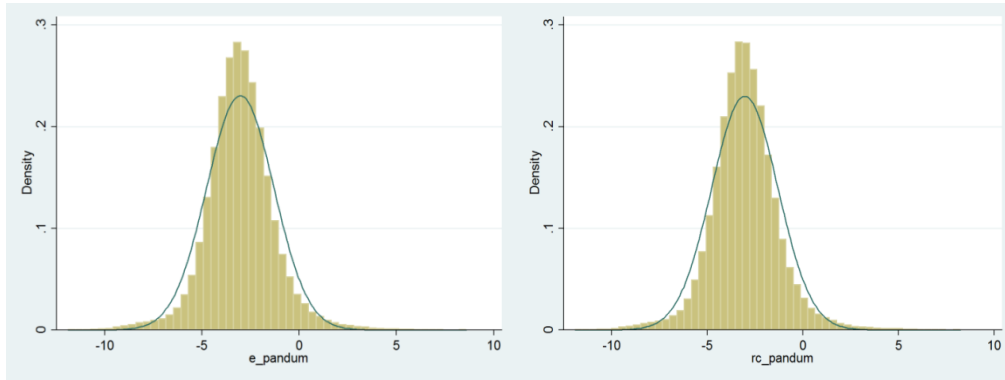


Figure 5.1 histogram plotting for estimated residual V.S. normal distribution (2007) (left)

Figure 5.2 histogram plotting for estimated random effects V.S normal distribution. (2007) (right)

For a normal distribution, histograms should be symmetric with kurtosis value 3 and skewness value zero. Clearly, residuals' real distribution could not support the normal assumption. The residuals present a slightly asymmetric pattern with fatter tails and are much more peaked center.

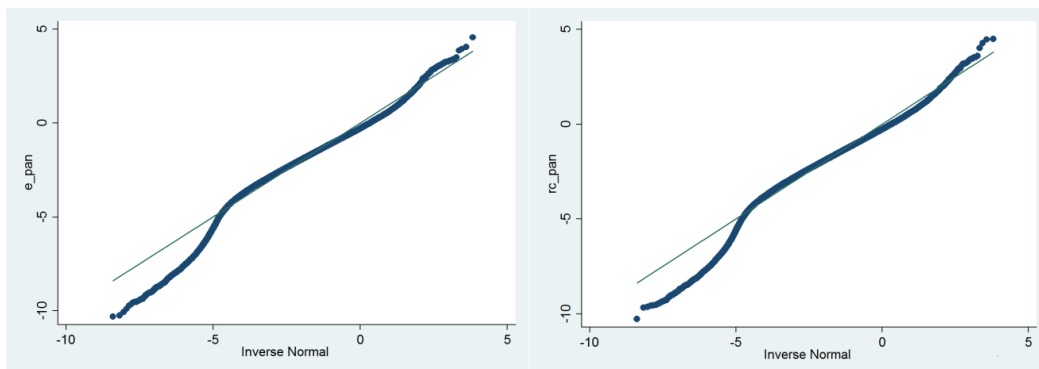


Figure 5.3 QQ plot for estimated residual V.S. normal distribution (2007) (left)

Figure 5.4 QQ plot for estimated random effects V.S normal distribution. (2007) (right)

The horizontal axis stands for quantiles of normal distribution while vertical axis is that of empirical distribution. If empirical distribution is normal, line  $y = x$  will be received. The residuals' QQ plot also rejects the normal assumption as the empirical residual stays away from the assumed normal distribution. QQ plot shows clearly that

---

the residuals are non-linear. Hence, the ‘black swan’ event is experienced, such as what happened in ‘credit crunch’, this research employs the semi-parametric model which allow non-parametric estimation.

## **5.4 GAM results**

As shown in the last section, the derived residuals suggest that parametric models used previously could not fully capture SMEs’ performance during the ‘credit crunch’. Considering the ‘credit crunch’ as a ‘black swan event’, knowledge of such cases is limited. It may be that a non-parametric method would be more capable of capturing the behaviour by empirical means. The main disadvantage of a pure non-parametric model is the estimation efficiency. In the research multiple variables are employed to analyse the data, which has heavy tails with outliers. Given both the size of the sample and the data’s structure, the efficiency of estimation could be reduced, especially due to a slow-down in the convergence rate. This problem is called the *curse of dimensionality*.

Hence, this research uses General Additive Models (GAM), combining non-parametric effects with parametric regressions, using a logistic link function. This section firstly presents results of GAM using WoE data to investigate whether the parametric models employed could capture the non-linear features of SMEs’ behaviour. Then a GAM using the original format data is presented to allow examination of each variable’s trend during the ‘credit crunch’.



### 5.4.1 GAM with WoE Data

In this section, GAM results are fitted by WoE format data sets. Results show not only which independent variables have significant non-parametric effects, but it could also support model fitting could be improved by adding those non-parametric effects.

Significance of non-parametric effects

This section tests for the existence of non-parametric effects in the panel model. As the exact data has been used to build parametric models, such as logistic regression and logit panel data model, a significant smoother means previous models previous model could be improved by incorporate those non-parametric effects. In GAM results, most independent variable show significant non-parametric effects. Table 5.9 summarises a number of insignificant smoothers for both segments:

segment	total no. of variables	significant level	no. of insignificant smoothers			
			2007	2008	2009	2010
start-ups	13	5%	3	3	2	3
non-start-ups	16	5%	5	3	3	3

Table 5.9 Summary of insignificant smoothers for both segments at level 5%

After performing GAM on WoE data, the majority of variables present significant non-parametric effects. Generally speaking, the number of variables with insignificant non-parametric effects follows a relatively consistent pattern through the ‘credit crunch’. Although SMEs exhibit high default rates during a financial crisis, the number of variables showing non-parametric effects is not larger than in normal

---

economic conditions. On the contrary, due to the crisis SMEs' performance actually becomes more predictable by linear models with parametric assumptions. Another interesting finding is that there are not more non-parametric effects found for start-up segments compared to non-start-ups. For example, 10 out of 13 variables exhibit significant non-parametric effects for start-ups in 2010; in the same year, 13 out of 16 variables have significant non-parametric effects for non-start-ups. Therefore, the following conclusions about the variables with non-parametric effects can be made:

1. The majority of variables exhibit non-parametric effects;
2. The 'credit crunch' does not exhibit significantly more or fewer non-parametric effects;
3. Start-ups do not exhibit more non-parametric effects than non-start-ups.

**i) Separation measures**

Separation measures for GAM are presented in **Error! Reference source not found.** and **Error! Reference source not found.** for the training sample and holdout sample respectively. For the holdout sample, the estimated coefficients and smoothers are matched to corresponding variable. GAM always results in a better separation measure than parametric models. It means independent variables even with the use of WoE are not linearly correlated with dependent variables, and ignoring the non-linear effect reduces prediction accuracy. The improvement could also be due to GAM's flexibility, as it allows the use of link functions.

---

As mentioned previously, the data employed contains three separate types of economic conditions: 2007 is usually regarded as normal economic conditions; the following two years are ‘credit crunch’ periods, with SMEs performing worst in 2009; significant recovery in the UKs economy appears in 2010, which is also reflected by SMEs’ performance. The influence of non-parametric effects varies in different economic conditions. For start-up SMEs, adding non-parametric effects achieves the most significant improvement in 2007 and 2010, which are before and after the financial crisis. The improvement made by adding the non-parametric effect is weakest in 2009. These findings align with previous conclusion that there are fewer variables with significant non-parametric smoothers in this year. Therefore, non-parametric effects are less significant for start-ups during a financial crisis. For the ‘non-start-ups’ segment, GAM improvement is more consistent through the crisis. It is in 2010 that GAM has greatest separation power. It means that it is after the ‘credit crunch’ that ‘non-start-up’ SMEs exhibit more non-linear performance.

In conclusion, the ‘credit crunch’ reduces the non-parametric effect for start-up SMEs while non-parametric effects are most noticeable after the ‘credit crunch’ for ‘non-start-ups’. There is no significant increase of non-parametric effects during ‘credit crunch’ for all UK SMEs

## **ii) Implementations**

SMEs perform non-parametrically and GAM can improve model fitting by capturing those non-parametric effects. Significant non-parametric effects may indicate that

---

SMEs' credit risk can diverge from linear predicted models. The successful survivor firms can gain unexpected results which may be captured by non-parametric effects. The 'credit crunch' hits 'start-ups' by making them lose their 'swimming' ability, while for 'non-start-ups' GAM's effects are more significant when economic conditions return to normal.

Results for GAM using data in WoE format are presented below. As explained in the previous chapter, the smoothing component is estimated by kernel smoother. The roughness of the smoother is described by the so-called smoothing parameter. The smoothing parameter is not a coefficient but an indicator of how smoothy the kernel curves are. The non-parametric effect is more smooth if the smoothing parameter becomes close to one. Adding the linear influence to variables' non-parametric effects, the additive trends are the marginal influence of the corresponding independent variables.

The following table summarize estimated parameters for GAM model. It includes the estimated smoothing parameter for variables' non-parametric effects. And the coefficients for the linear component. For simplicity, variables' significance is also marked in the same table by asterisk. The significance of the smoothing component is marked following the smoothing parameter while the significance of linear part is marked following the linear coefficients.

firm character	smoothing component				regression models			
	2007	2008	2009	2010	2007	2008	2009	2010
Parent Company – Derog Details					0.52***	0.49***	0.30*	-0.05
Debt Gearing (%)					0.44***	0.36***	0.35***	0.28**
Legal Form	0.01***	0.01***	0.01***	0.01**	0.51***	0.48***	0.46***	0.29***
1992 SIC Code	0.80***	0.77***	0.75***	0.63***	0.51***	0.55***	0.47***	0.63***
Region	0.04	0.19***	0.09***	0.05***	0.08	0.31***	0.25***	0.15
No. of 'Current' Directors	0.16***	0.39***	0.38***	0.32***	0.59***	0.45***	0.34***	0.47***
Proportion of Current Directors to Previous Directors in the Last Year	0.01**	0.01***	0.01***	0.01***	0.34***	0.56***	0.24***	0.51***
PP Worst (Company DBT - Industry DBT) in the Last 12 Months	0.16	0.13	0.06***	0.11***	0.41***	0.34***	0.26***	0.43***
Total Value Of Judgements in the Last 12 Months	0.01***	0.01	0.01***	0.01***	0.41***	0.34***	0.43***	0.40***
Number of Previous Searches (last 12m)	0.03	0.12***	0.04***	0.10***	1.26***	0.64***	0.42***	0.84***
Time since Last Derogatory Data Item (months)	0.06***	0.01***	0.02***	0.05***	0.52***	0.49***	0.30***	- 0.05***
Lateness of Accounts	0.04***	0.06***	0.04***	0.03***	0.44***	0.36***	0.35***	0.28***
Time Since Last Annual Return	0.01***	0.01***	0.01***	0.01***	0.51***	0.48***	0.46***	0.29***
Total Fixed Assets as a Percentage of Total Assets	0.80***	0.77***	0.75***	0.63***	0.51***	0.55***	0.47***	0.63***
Percentage Change in Shareholders' Funds	0.04***	0.19***	0.09	0.05	0.08***	0.31***	0.25***	0.15***
Percentage Change in Total Assets	0.16***	0.39***	0.38***	0.32***	0.59***	0.45***	0.34***	0.47***
Intercept					0.34***	0.56***	0.24***	0.51**

Table 5.10 GAM via WoE parameter estimation for non-start-up SMEs Note: 1. \*\*\*refers to significant at level 1%; 2. \*\*refers to significant at level 5%; 3. \*refers to significant at level 10%

year	model	start-ups				non-start-ups			
		H	Gini	AUROC	KS	H	Gini	AUROC	KS
2007	logistic R	0.330	0.654	0.827	0.523	0.164	0.639	0.819	0.485
	panel data	0.312	0.642	0.821	0.519	0.154	0.603	0.802	0.458
	panel+dummies	0.313	0.644	0.822	0.526	0.162	0.639	0.819	0.483
	panel +set of averaged MVs	0.313	0.644	0.822	0.526	0.162	0.639	0.819	0.483
	panel+averaged GDP growth rate	0.313	0.643	0.821	0.519	0.162	0.639	0.819	0.484
	GAM with WoE data	<b>0.343</b>	<b>0.666</b>	<b>0.837</b>	<b>0.552</b>	<b>0.174</b>	<b>0.649</b>	<b>0.826</b>	<b>0.491</b>
2008	logistic R	0.431	0.736	0.868	0.602	0.311	0.709	0.855	0.553
	panel data	0.425	0.731	0.866	0.599	0.297	0.686	0.843	0.539
	panel+dummies	0.425	0.732	0.866	0.599	0.311	0.710	0.855	0.554
	panel +set of averaged MVs	0.425	0.732	0.866	0.599	0.311	0.710	0.855	0.554
	panel+averaged GDP growth rate	0.425	0.731	0.865	0.598	0.311	0.710	0.855	0.553
	GAM with WoE data	<b>0.441</b>	<b>0.745</b>	<b>0.873</b>	<b>0.610</b>	<b>0.325</b>	<b>0.727</b>	<b>0.864</b>	<b>0.566</b>
2009	logistic R	0.554	0.805	0.902	0.697	0.491	0.788	0.894	0.641
	panel data	0.548	0.802	0.901	0.695	0.480	0.775	0.888	0.633
	panel+dummies	0.550	0.803	0.901	0.696	0.490	0.786	0.893	0.641
	panel +set of averaged MVs	0.550	0.803	0.901	0.696	0.490	0.786	0.893	0.641
	panel+averaged GDP growth rate	0.549	0.802	0.901	0.696	0.490	0.786	0.893	0.640
	GAM with WoE data	<b>0.560</b>	<b>0.814</b>	<b>0.908</b>	<b>0.696</b>	<b>0.500</b>	<b>0.801</b>	<b>0.901</b>	<b>0.649</b>
2010	logistic R	0.412	0.705	0.852	0.573	0.327	0.718	0.859	0.558
	panel data	0.398	0.699	0.850	0.568	0.316	0.700	0.850	0.544
	panel+dummies	0.412	0.705	0.852	0.573	0.326	0.720	0.860	0.562
	panel +set of averaged MVs	0.399	0.700	0.850	0.568	0.326	0.720	0.860	0.562
	panel+averaged GDP growth rate	0.398	0.700	0.850	0.569	0.326	0.720	0.860	0.562
	GAM with WoE data	<b>0.427</b>	<b>0.723</b>	<b>0.863</b>	<b>0.585</b>	<b>0.343</b>	<b>0.741</b>	<b>0.872</b>	<b>0.574</b>

Table 5.11 GAM model separation measures for training sample

year	Model	start-ups				non-start-ups			
		H	Gini	AUROC	KS	H	Gini	AUROC	KS
2010	logistic R	0.412	0.705	0.852	0.573	0.323	0.674	0.837	0.524
	panel data	0.398	0.699	0.850	0.568	0.314	0.657	0.828	0.510
	panel+dummies	0.399	0.700	0.850	0.568	0.326	0.685	0.843	0.529
	panel +set of averaged MVs	0.399	0.700	0.850	0.568	0.326	0.685	0.843	0.529
	panel+averaged GDP growth rate	0.398	0.700	0.850	0.569	0.326	0.683	0.841	0.528
	GAM with WoE data	<b>0.281</b>	<b>0.557</b>	<b>0.778</b>	<b>0.423</b>	<b>0.358</b>	<b>0.726</b>	<b>0.863</b>	<b>0.562</b>

Table 5.12 GAM model separation measures for holdout sample

## 5.4.2 GAM using original format of continuous variables

The previous section confirms not only that the majority of variables have non-parametric effects, but also that ignoring these effects would reduce models' fit.

However, the variables' trends cannot be directly estimated since the variables are

---

transformed by WoE. This section provides GAM's results by using the original data format. This may demonstrate better how each variable influences the SME's performance. In the Data Description chapter, missing values are imputed with matching observed values for all selected continuous variables. Hence, the different variables' influence can be explored. Segments are kept as previously.

Categorical variables are usually not numerical variables initially. Take *region* as an example, which is described by the first two letters of the postcodes. To avoid using character format data, *region* is transformed to numerical form. However, this would not reflect firms' geographical locations. Therefore, its explored trend could not demonstrate geographical influence. Considering that the aim of using original format is to explore the independent variables' trend, WoE is used for categorical variables and linear function is assumed for those variables in WoE format. A similar assumption is made for variables that do not have significant non-parametric effects.

#### **i) Non-start-up SMEs**

##### ***Introduction***

Comparing to previous models, several variables lose significance at significance level 5%. Two of them are categorical variables, *legal form* and *region*, which are not significant in 2007 and 2010. Omitting their smoothing components and using WoE format are both reasons for the loss of significance. All continuous variables in their original form have at least one significant component. The following will discuss continuous variables' coefficients, significance, smoothing parameters and variable

---

trends in detail. Variables' coefficients and smoothing parameters are presented in Table 5.13 along with their significance. Each variable's trend is discussed separately.

The smoothing parameter is not a coefficient, yet it represents the smoothness of smoothing components. As the smoothing parameter approximates to one, the curve becomes smoother. In the results, most of those variables can be well estimated by a smooth curve with a larger smoothing parameter. The roughest curve is the smoothing component of *Number of 'Current' Directors*, which is clearly composed of a series of broken lines while the smoothing components of other variables forms smoothing curves.

#### ***Variables with or without a smoothing component***

There are three categorical variables in the 'non-start-ups' model which are: *Legal Form*, *1992 SIC Code* and *Region*. *Legal Form* and *Region* are initially in character format and are then transferred into numerical variables. Although *1992 SIC Code* is a numerical variable originally, its values are codes for categories, and refer to a specific industry. As discussed previously, their numerical values are as not meaningful as those of continuous variables, and smoothing trends on those variables cannot give any further insights into behaviour. In addition, there are two continuous variables that do not have significant smoothing components, which are *Parent Company's Derogatory Details* and *Debt Gearing (%)*. Therefore those five variables are used in WoE form. A problematic variable is *Percentage Change in Shareholders' Funds* as this variable GAM does not converge in 2009. The failure is due to the



---

estimation of its missing category. Even the best estimation of its missing category, as shown in the data chapter, could not represent its performance well enough. The estimation becomes too erratic and so GAM cannot converge. Therefore, its WoE format is included. The rest of variables are estimated by two parts: a linear part added to a smoothing component. Similar as in the previous chapter, the smoothing parameter shows how smoothy the kernel curves are. The following table summarizes estimated parameters using the original data. It includes the estimated smoothing parameter for variables' non-parametric effects. And the coefficients for the linear component. For simplicity, variables' significance is also marked in the same table by asterisk. The significance of the smoothing component is marked following the smoothing parameter while the significance of linear part is marked following the linear coefficients.

Firm Character	Linear Coefficient				Smoothing Parameter			
	2007	2008	2009	2010	2007	2008	2009	2010
Intercept	1.450***	0.358**	-0.413***	-0.415**				
Legal Form	0.199*	0.407***	0.380***	0.094				
Parent Company – Derog Details	0.650***	0.461***	0.329***	0.269***				
1992 SIC Code	0.705***	0.601***	0.543***	0.538***				
Region	0.095	0.306***	0.332***	0.171				
Debt Gearing (%)	0.482***	0.496***	0.472***	0.481***				
Percentage Change in Shareholders' Funds	0.635***	0.560***	0.537***	0.707***				
No. of 'Current' Directors	0.198***	0.256***	0.357***	0.392***	0.668	0.787***	0.846***	0.871*
Proportion of Current Directors to Previous Directors in The Last Year	-0.190*	-0.150	0.534**	0.518***	1.000***	0.999***	1.000***	1.000
PP Worst (Company DBT - Industry DBT) in the Last 12 Months	-0.107***	-0.099***	-0.096***	-0.226***	1.000***	1.000***	1.000***	1.000**
Total Value of Judgements in the Last 12 Months	-0.436***	-0.234***	-0.390***	-0.794***	1.000***	1.000***	1.000***	1.000***
Number of Previous Searches (last 12m)	0.030	0.003	0.073***	-0.055	0.941***	0.953***	0.964***	0.955***
Time Since Last Derogatory Data Item (months)	0.097***	0.689***	0.802***	0.526***	1.000***	1.000***	1.000***	1.000***
Lateness Of Accounts	-0.466***	-3.103***	-2.056***	-3.626***	1.000***	1.000***	1.000***	1.000***
Time Since Last Annual Return	-3.947***	-1.856***	-2.212***	-2.585***	1.000***	1.000***	1.000***	1.000***
Total Fixed Assets as a Percentage of Total Assets	0.131***	0.214***	0.213***	0.163***	1.000***	1.000***	1.000***	1.000***
Percentage Change in Total Assets	0.185*	0.701***	0.868***	0.753	1.000***	1.000***	1.000***	1.000***

Table 5.13 Parameter estimation: GAM with original values for non-start-up SMES

Note: 1. \*\*\*refers to significant at level 1%; 2. estimated coefficient is listed first and the z-statistic is listed below.

explored additive trend. This section exhibits the additive trend of continuous variables which have significant non-parametric effects. Those variables are summarized in the previous sections and the significance of components are presented in the above table. Variables' trends are plotted through four years and shown in the following order: 2007 in the upper left corner; 2008 in the upper right corner; 2009 in the lower left corner; 2010 in the lower right corner. For each year's plotting, the x-axis represents the value of independent variable, while the y-axis is the value of 'bad' rate. The shadow area shows the 95% of confidence band of additive trend. Each variable's trend is explained in detail in the following section.

- *Number of 'Current' Directors*

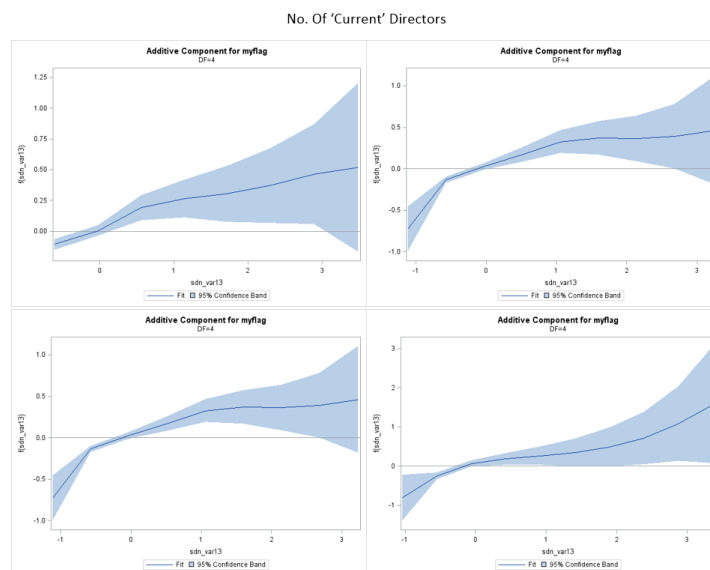


Figure 5.5 additive effects for Number of 'Current' Directors:

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

---

The linear effect of *Number of 'Current' Directors* has a positive parameter which is significant over four years. However, its smoothing component is not significant in 2007 and 2010.

When adding those two parts together, *Number of 'Current' Directors* shows a positive correlation in general. The confidence band is especially narrow from negative half standard deviation (SD) to the sample mean. 'Non-start-up' SMEs falling into this interval are influenced most by increasing board size. In general, as *Number of 'Current' Directors* increases, SMEs have a higher probability of being 'good'. This is a very strong implication suggesting that the larger the director group, the less its probability of being 'bad'. This finding seems contrary to finance research in general. Larger boards raise problems of agency which can significantly reduce a firm's performance (Jensen, 1993 and Lipton *et. al.*, 1992). However, there are two major differences between this research and others with different results:

1. This research only focuses on SMEs while others are usually talking about large corporations. In addition, it is common for SMEs to have very small number of directors on the board, as low as one director. A larger board of directors for SMEs is still a small size compared to large firms, therefore the agency problem is still minor.
2. The dependent variable in this research is SMEs' 'bad' rate. However, finance research is usually interested in a firm's profitability. Although a smaller

board size is found to be more efficient, riskier projects or default decisions might be more easily approved.

Hence, the conclusion is that SMEs' performance can be improved if they enlarge their board size. The possible explanations are: first, a larger board will bring more knowledge to help the firm to survive; second, single-person or small boards can lead to default decisions more easily.

- *Proportion of Current Directors to Previous Directors in the Last Year*

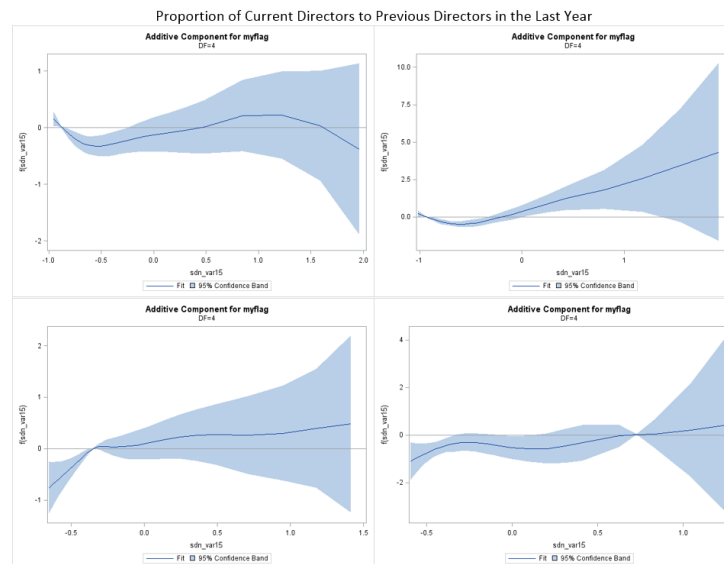


Figure 5.6 additive effects for Proportion of Current Directors to Previous Directors in the Last Year

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

This variable's linear component is not significant at 95% level in 2007 and 2008, while its smoothing component is not significant in 2010. It implies the 'credit crunch' changes this variable's non-linear influence into a linear-like one. Its additive effect has wide confidence limits above SD, especially in 2007. In general, firms with current directors at a proportion equal to SD correspond to those whose number of

---

directors is six times larger than the previous number. As the majority of data falls below SD, it is mainly outliers that are above it. Its additive effect can be divided into three parts for discussion:

1. The first part is below the negative half SD. This part corresponds to SMEs which have a smaller number of directors than the previous year. For those companies, the additive effect shows a negative trend in 2007 and 2008 but an inverse effect in 2009 and 2010. Hence, if non-start-up SMEs reduce their board size faster than the sample mean, this variable impacts the firm's performance in an inverse way before and after the 'credit crunch'.
2. The second part is from negative half SD to SD, which is an interval around the sample mean. This part contains the majority of 'non-start-ups' and this trend is almost consistent through time. The only exception is 2010, in which year this variable's effect becomes more flat. Hence, around the sample mean, if a 'non-start-up' increases its board, the knowledge brought in by the new directors helps the firm to survive. However, after the 'credit crunch' recruiting more directors would not be able to help the firm to take on new challenges. This finding is consistent with *No of 'Current' Directors*.
3. The last part is the part above SD, which mainly contains outliers. For firms that fall into this interval, this variable has a wide confidence band and therefore there is a great deal of uncertainty.

In conclusion, enlarging board size decreases the firm’s probability of facing financial constraints if the increasing ratio falls around the sample mean, except in 2010. It means the increasing board size brings new knowledge to firms and significantly helps them to achieve survival. However, the ‘credit crunch’ has changed the business environment so much that, after it, the knowledge directors gained from the past becomes less precious for helping the firm. This finding coordinates with that of *No of ‘Current’ Directors*. For a firm’s fall away from the sample mean, the variable’s influence is inconsistent.

- *PP Worst (Company DBT - Industry DBT) in the Last 12 Months*

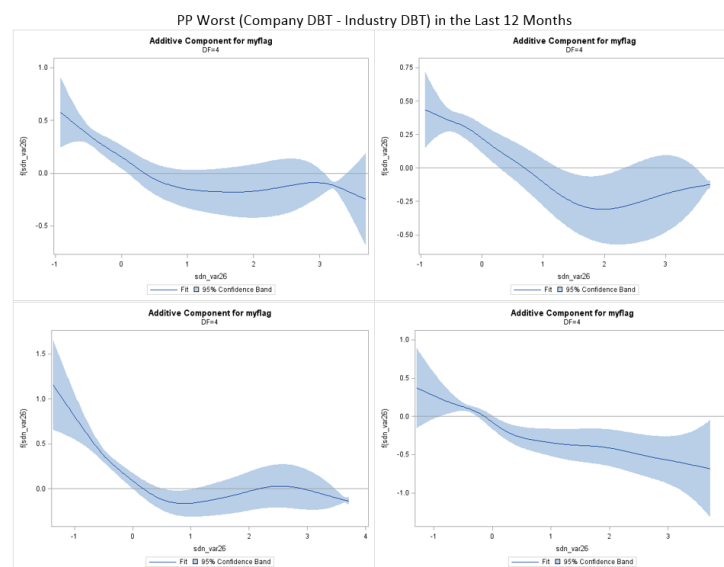


Figure 5.7 additive effects for *PP Worst (Company DBT - Industry DBT) in the Last 12 Months*

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

Both this variable’s linear component and smoothing component are significant over four years. Its additive effect shows a negative influence in general. *DBT* refers to ‘Days Beyond Terms’, which shows how rapidly firms transact their liabilities. *PP*

---

*Worst (Company DBT - Industry DBT) in the Last 12 Months* compares the company's performance with that of the corresponding industry. Its trend can be divided into two parts and the cutting point is one SD, which corresponds to approximately two months beyond term:

1. Below one SD: There is a negative trend consistently over the four years. As the majority of non-start-ups can be classified in this category, one can conclude that the longer one non-start-up takes to pay its invoices back, the higher their credit risk. The fitted line has the highest absolute tangent value in 2009, which means the longer it takes a non-start-up to pay its invoices, the faster its accelerate rate of 'bad' rate is.
2. Above one SD: this variable switches trends over time. DBT can present contradictory influences for SMEs in this category in those years, making those firms' performance more difficult to predict. Larger DBT makes 'non-start-ups' much more sensitive during business cycles, especially when the economy is changing its direction such as in 2008 and 2010.

Around the sample mean, for this variable a higher DBT, compared to its industry average, leads to a higher PD. However, there are still some changes to this variable's trend throughout the 'credit crunch'.



In conclusion, DBT has negative effects for ‘non-start-up’ SMEs falling around the sample mean, regardless of the economic conditions. When economic conditions are changing, extraordinarily long DBT makes firms’ performance less predictable.

- *Total Value of Judgements in the Last 12 Months*

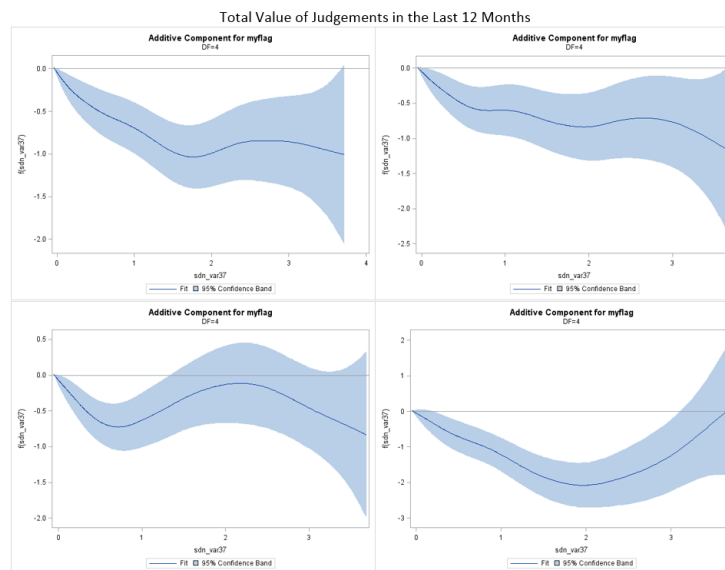


Figure 5.8 Total Value of Judgements in the Last 12 Months

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

If the obligor is not paying loans back, a judgement would be made with regard to the unsettled loan. The judgement record would persist even if payment was made after the judgement. This variable’s linear component and smoothing component are both significant over four years. In addition, most ‘non-start-ups’ have zero value of judgements; for example 98.12% ‘non-start-ups’ are in this category in 2007. Its additive effect is also zero at this point. Except for the zero judgments, the influence of a large value of judgments is sensitive towards business cycles and could be divided into three parts:

1. From the sample mean to 0.5 SD: a consistent negative effect regardless of the business cycle, with narrower confidence interval;
2. From 0.5 SD to 2 SD: negative additive effects, except 2009 in which an inverse effect occurs;
3. Above 3 SD: a larger value of judgements is negatively related to ‘non-start-ups’ performance, except 2010 in which this variable presents a positive effect.

This value shows a constant pattern below 0.5 SD and become more volatile above it. By using the original format of data, categorical annual variations are better captured to incorporate the influence of business cycles and improve model performance.

- *Number of Previous Searches (last 12 months)*

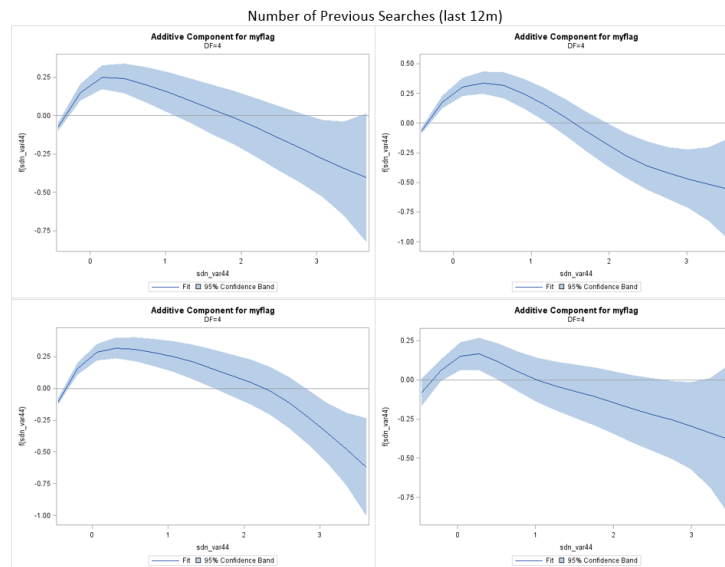


Figure 5.9 additive effect for Number of Previous Searches (last 12 months)

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

---

The number of searches shows how many times the ‘non-start-ups’ sought financial support in the last 12 months. Its linear component is only significant in 2009, yet its smoothing component is significant over four years. When adding the two parts together, a very clear quadratic pattern is produced. The sample mean of this variable increases through time: 1.45 in 2007, 1.56 in 2008, 1.77 in 2009 and 1.88 in 2010. As the mean increases, it indicates the ‘non-start-ups’ searching for funding more often. Dividing the performance into two parts, the detailed discussions are as follows:

1. Below the sample mean: a positive effect is observed, with a narrow confidence limit band. Therefore, a larger number of searches correlates to a better performance of the firm. One explanation is that the more often ‘non-start-ups’ ask for financial support, the more active their business.
2. Above the sample mean: there is a negative effect with a much wider confidence limit band. The higher search numbers are a clear sign of financial constraints. The difficulty of getting financial support can significantly decrease firms’ performance.

In summary, regardless of the business cycle, *Number of Previous Searches* shows a quadratic form with turning point around the sample mean.

- *Time since last derogatory data item (months)*

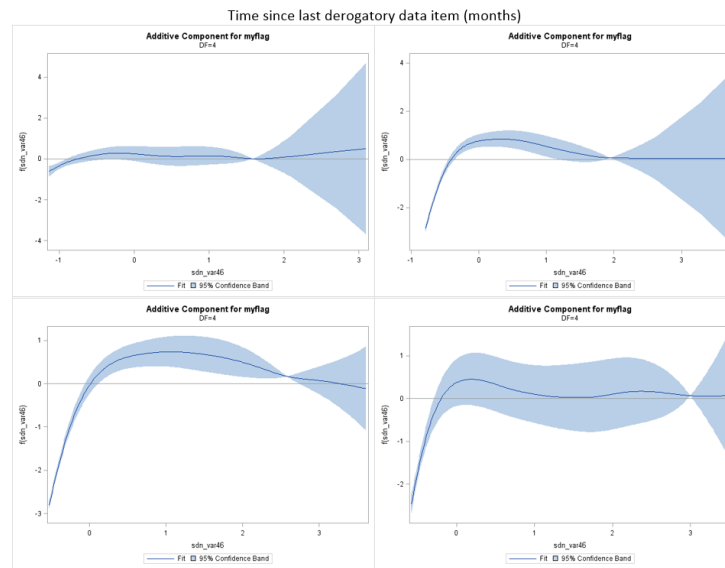


Figure 5.10 additive effects for *Time since last derogatory data item (months)*

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

The firm’s derogatory data is collected from various public sources for a more complete record of the firm’s previous history. This variable’s both linear component and smoothing component are significant over four years. When adding the two parts together, an exponential-like form becomes apparent. The additive component’s scale is smallest in 2007 but much wider in the remaining three years. This suggests the derogatory variable has a higher influence in PD forecasting through the ‘credit crunch’. In detail, the trend of its additive effect could be divided into two parts:

1. Below the sample mean: a positive influence is presented, with a very narrow confidence band. The scale of the negative influence is smallest in 2007 and largest in 2009. Hence, ‘non-start-ups’ performance improves as the last

derogatory data recedes with time. This influence is very significant given the narrow confidence band.

2. Above sample mean to sample mean plus two SD: less significant influence in 2007 and 2010, while there is a negative influence in 2008 and 2009. Possibly those firms had not encountered financial difficulties for such a long time that they lost their ability to handle a crisis. Those firms turned out to have a higher ‘bad’ rate during the ‘credit crunch’.
3. Above two SD: this variable has no significant influence, with very wide confidence limits which may be due to lack of data.

The derogatory data is especially significant if the record is more recent than the average. The effect of a recent derogatory record significant reduces the firm’s performance during the ‘credit crunch’.

- *Lateness of Accounts*

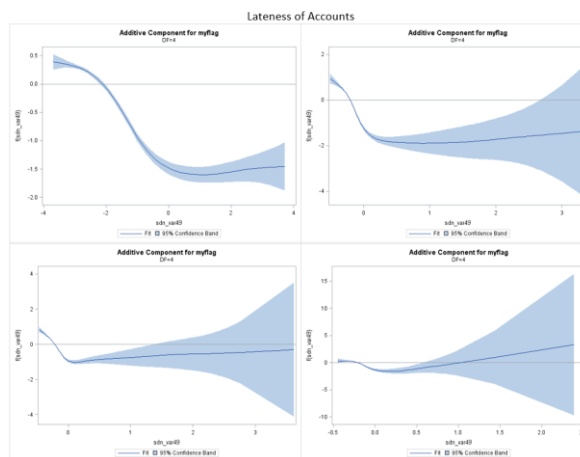


Figure 5.11 additive effects for Lateness of Accounts

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

---

This variable's linear component and smoothing component are significant over all four years. The inverse trend of linear component and smoothing component results in a quadratic form for the additive effect, and the changing point is always around the sample mean. Therefore, the impact of *Lateness of Accounts* can be separated into two parts:

1. Below the sample mean: SMEs' performance decreases as their *Lateness of Accounts* increase. This part has a negative influence with a very narrow confidence band. Hence, as the SMEs' accounts becomes more dated, they tend to exhibit a worse performance. This trend is consistent regardless the business cycle.
2. Above the sample mean, *Lateness of Accounts* has a positive influence with a wider and wider confidence band. This means those 'non-start-ups' becomes less predictable due to the changing economy.

In summary, *Lateness of Accounts* is a strong determinant of separation for 'non-start-ups' below the sample mean. The newer the account information is, the better performance the firm will have. For firms that update their account over a longer period than the sample mean, their performance is influenced by the 'credit crunch' and becomes less predictable.

- *Time since Last Annual Return*

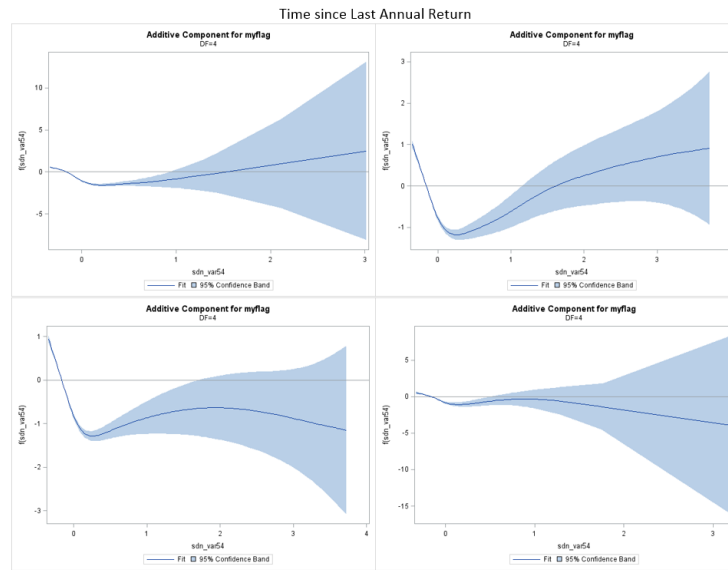


Figure 5.12 additive effects for Time since Last Annual Return

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

According to the UK government, companies are required to send their annual return one year after either *incorporation of the company* or *date you filed your last annual return*. It should be completed up to 28 days after the due date. It mainly contains firms' general information rather than accounting ratios describing the functioning of the firm (GOVUK, 10 Dec., 2014):

- officers' information----firm directors and secretaries general;
- SIC----classification of firm's business type;
- Capital snapshot which is required for firms that have share capital.

*Time since Last Annual Return* tells the duration since a firm last reported its general information. It helps supervisors and banks to gather more information about this firm

---

for the purpose of ‘*communication, influence, training and support, investigation and others*’ (Annual Return 2010, Standards for England). Keasey and Watson (1986) has used a similar variable, lags in reporting to the Companies House, to model UK SMEs’ defaults. It provides a snapshot of the firm and is used to guarantee sufficient information is provided to Companies House. Hence, being able to provide its annual return indicates that the firm is being run under normal circumstances by the known directors with a clearly stated amount of capital.

*Time since Last Annual Return*’s linear component and smoothing component are significant over all four years. Adding them together, a quadratic form is seen in 2007 and 2008, and then a higher order polynomial of degree three in 2009 and 2010. Hence, the additive effect can be divided into three parts:

1. Below the sample mean: a rapid decrease with a narrow confidence band. For firms that fall below the average, the longer duration since their last annual return is correlated with worse performance. The influence of this part stays constant through the financial crisis.
2. From sample mean to sample mean plus one SD: the influence of *Time since Last Annual Return* becomes positive in this part, with wider confidence limits. It indicates that firms falling into this part gain survival ability through time. The longer the duration since their last annual return, the more knowledge they gain to keep their business from financial constraints.



- 
3. Above sample mean plus one SD: the annual return has an almost constant effect, with widest confidence limits especially before and after the financial crisis. The firms have not reported to the Company House for a long time. These firms' information becomes opaque and their performance is therefore difficult to predict given the information.

The scale of additive effects is very large over all four years, with a very narrow confidence band below the sample mean. Although one cannot gather more detailed information about the firms without further investigation, this research shows that *Time since Last Annual Return* is a key variable in judging SMEs' performance. 'Non-start-ups' should regularly release their information to the public. This variable has a similar trend to *Lateness of Accounts*, as both variables describe the frequency with which a company updates its information. The two variables are not highly correlated since they are collected from different sources. *Lateness of Accounts* is usually used by banks or other credit suppliers, and is related to firms' accounting information statutes. Meanwhile it is Companies House that receives firms' annual reports, which contain updates to firms' legal information. Their annual correlation is shown in Table 5.14:

non-start-ups			
2007	2008	2009	2010
0.085	0.317	0.385	0.429

Table 5.14 correlation between lateness of account and annual reports

- *Total Fixed Assets as a Percentage of Total Assets*

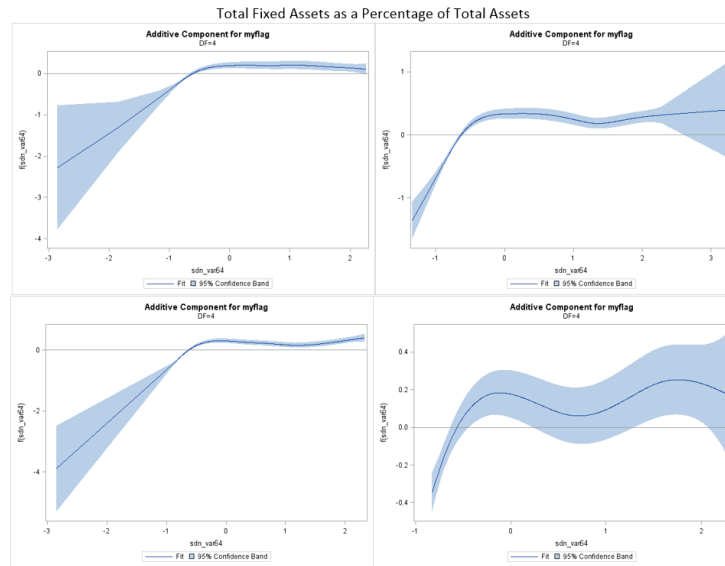


Figure 5.13 additive effects for Total Fixed Assets as a Percentage of Total Assets

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

This variable’s linear component and smoothing component are both significant over all four years. Its additive effect gives an exponential-like pattern in the first three years, yet it has a polynomial pattern in 2010. Sample means usually falls around 26%, which means fixed assets are usually 26% of total assets for ‘non-start-ups’. Its pattern stays rather constant around the sample mean though time, yet it is the tails trends’ change. There are longer tails on the left hand side in 2007 and 2009 but on the right hand side in 2008. The variable’s trend can be divided into two parts:

1. Below the sample mean: positive influence with wide confidence band which is especially wide below the sample mean minus one SD. Hence, for firms in this category, an increase in fixed assets indicates better performance.

However, if ‘non-start-ups’ have a very low percentage of fixed assets, their performance becomes less predictable.

2. Above the sample mean: an almost constant value except 2010. Therefore, if ‘non-start-ups’ have a higher percentage of fixed assets than the average, the firms’ performance would not be improved by having more fixed assets. The performance of those ‘non-start-ups’ becomes less clear after economic shocks, such as what happened in 2010; GAM can help modellers capture the new trend and increase prediction accuracy in those years.

In conclusion, the percentage of fixed assets shows an exponential-like trend with the turning point falling around the sample mean. It is especially informative for ‘non-start-ups’ falling between minus one SD and the sample mean.

- *Percentage Change in Total Assets*

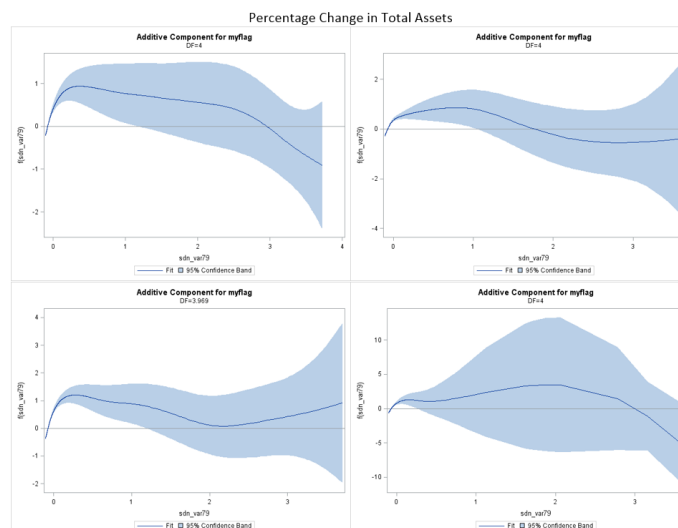


Figure 5.14 additive effects for Percentage Change in Total Assets

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

---

This variable's linear component is not significant, at 95% level in 2007 and 2010. However, its smoothing component is significant over four years. When adding those two parts together, its additive component exhibits wide confidence limits, especially in 2010. GAM has derived a quadratic-like form in 2007 and 2010, while in 2008 and 2009 the additive effect has a polynomial-like form of degree three. Therefore, the influence of this additive effect can be divided into three parts:

1. Below the sample mean: positive influence with narrow confidence band.  
Hence, for 'non-start-ups' with a *Percentage Change in Total Assets* lower than the sample mean, any positive increment of total assets can significantly improve firms' performance in an almost linear fashion.
2. From the sample mean to sample mean plus two SD: there is a negative influence with a wide confidence band in 2007, 2008 and 2009. However, in 2010 this part stays almost constant, with an extremely wide confidence band. Therefore, this variable's influence is less predictable with a bigger bias, especially after the 'credit crunch'.
3. Above sample mean plus two SD: influence is sensitive to the business cycle, with a wide confidence band. Hence, the 'non-start-ups'' performance is less predictable for larger total asset changes.

---

In summary, the change of total assets is more predictable when ‘non-start-ups’ have a smaller *Percentage Change in Total Assets*, giving a positive influence. For a larger change in total assets, SMEs’ performance becomes less predictable.

***Summary and implementations***

In total, there are ten continuous variables analysed in their original format for the ‘non-start-up’ segment. This number is much larger than that of the start-ups since missing categories for ‘non-start-ups’ are easier to replace by observed values. Each variable’s features have been explained in above section and the summary is given below:

<b>intercept</b>	<b>whether kept in categorical format</b>	<b>trend consistency</b>	<b>trend of additive effects</b>	<b>interval with narrow confidence band</b>	<b>sensitive tails interval</b>
Legal Form	categorical variable				
Parent Company – Derog Details	categorical variable				
1992 SIC Code	categorical variable				
Region	categorical variable				
Debt Gearing (%)	too much noise (1)				
Percentage Change in Shareholders Funds	outlier(2)				
No. of 'Current' Directors		constant	positive	(-SD, SD)	
Proportion of Current Directors to Previous Directors in the Last Year		switch sign below -0.5SD	quadratic-like	(-0.5SD, 0.5SD)	above SD
PP Worst (Company DBT - Industry DBT) in the Last 12 Months		switch trend above one SD	quadratic-like	(-SD, SD)	above 3SD
Total Value of Judgements in the Last 12 Months		switch trend above SD	quadratic or polynomial-like	around sample mean	above SD
Number of Previous Searches (last 12m)		constant for four years	quadratic	below sample mean	
Time since Last Derogatory Data Item (months)		almost constant	exponential-like	below sample mean	above 3SD
Lateness of Accounts		almost constant	L' shape-like	below sample mean	above 2SD
Time since Last Annual Return		almost constant	quadratic or polynomial-like	below sample mean	above 2SD
Total Fixed Assets as a Percentage of Total Assets		almost constant	exponential-like	above -SD	varies by year(3)
Percentage Change in Total Assets		almost constant	quadratic or polynomial-like	below sample mean	above sample mean

Notes: 1. too much noise in observed data to find a match for missing category

2. missing category is approximated by an outlier

3. below -SD in 2007 and 2009 but above sample mean in 2008 and 2010

---

*Table 5.15 independent variables' trend summary for non-start-ups*

---

Only two of them have a consistent pattern throughout the ‘credit crunch’: *Number of ‘Current’ Directors* and *Number of Previous Searches in Last 12 Months*. A positive effect is shown by *Number of ‘Current’ Directors*; therefore, non-start-ups with a larger board size will have a lower PD. *Number of Previous Searches in Last 12 Months* has a quadratic form: if the search number is lower than the sample mean, the more often the non-start-up seeks funding, the more active this firm is and it will have a lower PD; above the sample mean, more searching seems to be a warning sign that the non-start-up has difficulties in its financial chain. Although *Lateness of Accounts* and *Time since Last Annual Return* have sensitive tails, those two variables have a large scale of influence with narrow confidence bands below the sample mean. They describe how often firms provide their accounting information and general information to the public respectively. If their updating is faster than the average, a shorter gap they have between the two updates leads to a lower PD in monotonous pattern.

Firms’ lack of recent data updates show more uncertainty Therefore, when the economic condition switches, credit suppliers should examine their portfolio and give more attention to those SMEs with less information available.



---

**Results of separation measures**

<b>Year</b>	<b>Model</b>	<b>H</b>	<b>Gini</b>	<b>AUC</b>	<b>KS</b>
<b>2007</b>	<b>GAM</b>	0.153	0.581	0.791	0.443
	<b>Log Reg</b>	0.132	0.553	0.777	0.413
<b>2008</b>	<b>GAM</b>	0.312	0.679	0.840	0.527
	<b>Log Reg</b>	0.212	0.590	0.795	0.434
<b>2009</b>	<b>GAM</b>	0.481	0.769	0.885	0.626
	<b>Log Reg</b>	0.422	0.718	0.859	0.569
<b>2010</b>	<b>GAM</b>	0.238	0.646	0.823	0.490
	<b>Log Reg</b>	0.180	0.594	0.797	0.443
<b>Holdout</b>	<b>GAM</b>	0.358	0.726	0.863	0.864
	<b>Log Reg</b>	0.268	0.621	0.811	0.812

*Table 5.16 non-start-ups separation measure: GAM & logistic regression based on data's original format*

When performing GAM and logistic regression on the same set of data, GAM always provides better separation measures. GAM can better capture ‘non-start-ups’ risk features during the ‘credit crunch’, although one may argue that there are more parameters in GAM. This section provides the effects of continuous variables’ for ‘non-start-ups’ and explains how they influence non-start-ups’ performance through ‘credit crunch’. The variables’ trends can help financial institutions identify their portfolio’s risk in a changing economy. Variables usually have narrow confidence bands with little change of trend around the sample mean through the ‘credit crunch’. The ‘credit crunch’ usually changes tails’ performance by enlarging the confidence band or even changing the direction of influence. For instance, firms which have longer lateness of accounts, or which provided their last annual return some time ago, become opaque with regard to their information from Companies House, and prediction bias is bigger for those firms. ‘Non-start-ups’ with infrequent updates have a wider confidence band and those firms’ performance is less predictive. Considering

---

that all the information used in this research is ‘hard information’, this research suggests that ‘hard information’ is less predictive for firms that update infrequently. In conclusion, GAM can provide better prediction than standard models even when missing categories are replaced with matching observed values.

**ii) Start-up SMEs**

***Variables with or without smoothing components***

Out of 13 variables selected for start-up SMEs, *Legal Form*, *Company is Subsidiary*, *1992 SIC Code* and *Region* are categorical variables. Two continuous variables have insignificant smoothing components: *Number of Directors Holding Shares* and *Total Value of Judgements in the Last 12 Months*. As shown in the previous section, the following variables’ missing categories could not be well replicated: *Time since Last Derogatory Data Item (months)* and *Total Assets*. Hence, only five variables are entered in their original format. Their influence is summarized as follows.

Firm characters	parameters				smoothing parameters			
	2007	2008	2009	2010	2007	2008	2009	2010
interpretation	2.715***	2.486***	1.749***	2.189***				
Legal Form	1.076***	1.114***	1.429***	1.456***				
Company is Subsidiary	2.283***	0.326**	0.080	13.824***				
1992 SIC Code	0.450***	0.216***	0.283***	0.465***				
Region	0.511**	0.304***	0.466***	0.304***				
Number Of Directors Holding Shares	-0.320***	0.104***	0.228***	0.067				
Total Value Of Judgements In The Last 12 Months	0.851***	0.597***	0.733***	0.385***				
Time since last derogatory data item (months)	0.368***	0.642***	0.672***	0.642***				
Total Assets	0.885***	0.303***	0.426***	0.843***				
Proportion Of Current Directors To Previous Directors In The Last Year	0.376	0.168*	0.602***	0.624***	0.999	1.000	1.000***	1.000
Oldest Age Of Current Directors/Proprietors supplied (Years)	0.051	0.143***	0.267***	-0.012	1.000**	1.000***	1.000***	1.000***
Number Of Previous Searches (last 12m)	0.170**	0.196***	0.263***	0.208***	0.570***	0.860***	0.862***	0.911***
Lateness Of Accounts	-0.059	-0.483***	-0.312***	-0.501***	1.000***	1.000***	1.000***	1.000***
Time Since Last Annual Return	-0.586***	-0.615***	-0.568***	-0.452***	0.984***	0.987***	0.987***	0.989***

Table 5.17 Parameter estimation: GAM with original values for start-up SMES

- *Proportion of Current Directors to Previous Directors in the Last Year*

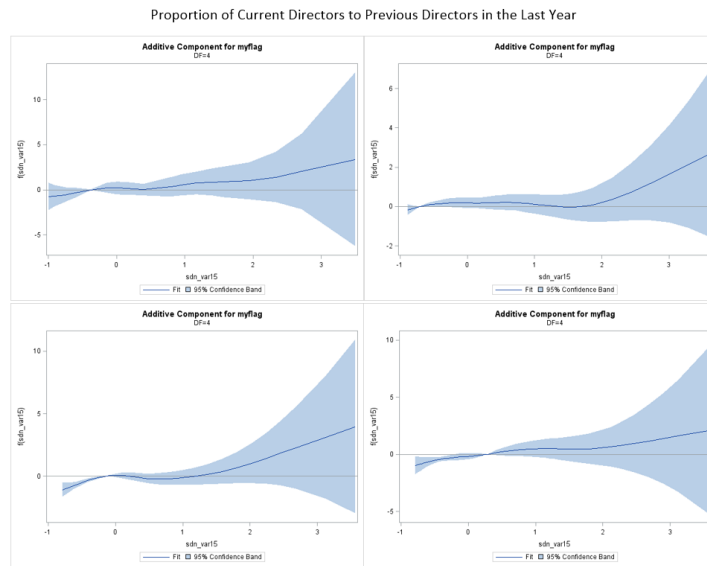


Figure 5.15 additive effects for *Proportion of Current Directors to Previous Directors in the Last Year*

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

The linear component is significant in 2009 and 2010 while the smoothing part is only significant in 2009. Adding them together, the additive effect shows a linear-like trend. As the *Proportion of Current Directors to Previous Directors in the Last Year* increases, the credit risk of ‘start-ups’ decreases and their performance improves. A narrow confidence band exists around the sample mean. Hence, if the board of ‘start-ups’ becomes constrained, the SME is more likely than others to fail. The confidence band becomes much wider subsequently, which means prediction can be highly variable for firms with high liquidity in their board.

- *Oldest Age of Current Directors/Proprietors Supplied (Years)*

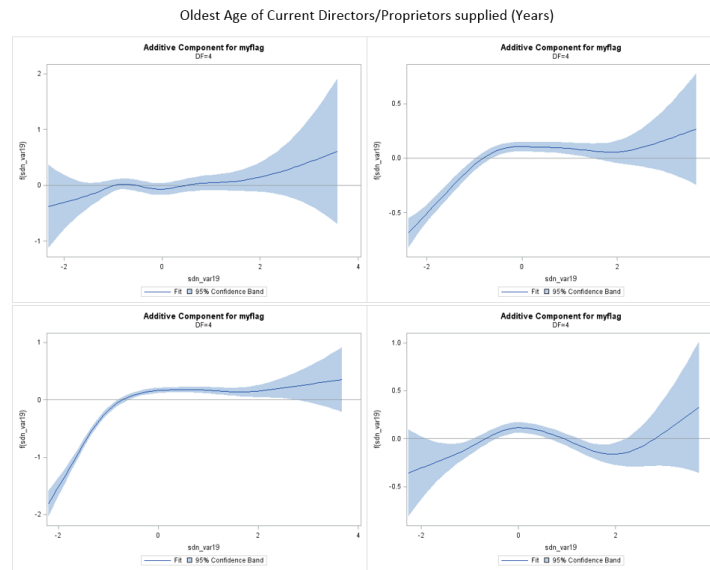


Figure 5.16 additive effects for Oldest Age of Current Directors/Proprietors supplied (Years)

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

Directors' ages describes another aspect of the board of directors. Its linear component is significant in 2008 and 2009 while the smoothing component is significant for all four years. Its additive effect has a linear-like trend in 2007, exponential-like trend in 2008 and 2009 and a polynomial form with a degree of three in 2010. Hence, this variable's influence is divided into three parts:

1. Below sample mean: The sample mean of directors' ages is always around 43. The curves all rise, with the steepest increase in 2009 with a narrow confidence band. It is lower before and after the financial crisis, with a wider confidence band. Therefore, for 'start-ups' that have directors younger than 43, their performance can be improved by recruiting older directors. This

---

improvement is especially significant during the ‘credit crunch’. Above this point, this variable starts to lose its effect.

2. From the sample mean, 43, to mean plus two SD, which is around 66: This part is relatively flat from 2007 to 2009. In 2010, it turns to a clear negative influence. Hence, directors’ experience gained through time becomes less effective after the financial crisis. Older directors, between 43 and 66 years, could not assist the firm to succeed in the post-crisis environment.
3. Above mean plus two SD: positive influence with a wide confidence band. Therefore, ‘start-ups’ with much older directors become less predictable. Considering that directors aged older than 66 are relatively rare, this result could be due to lack of data.

In summary, ‘start-ups’ performance increases as the directors becomes older since they have gained more knowledge to help the firm to survive. However, this impact becomes less clear, or even reverses, when the director’s age becomes considerably greater than the sample mean.

- *Number of Previous Searches (last 12m)*

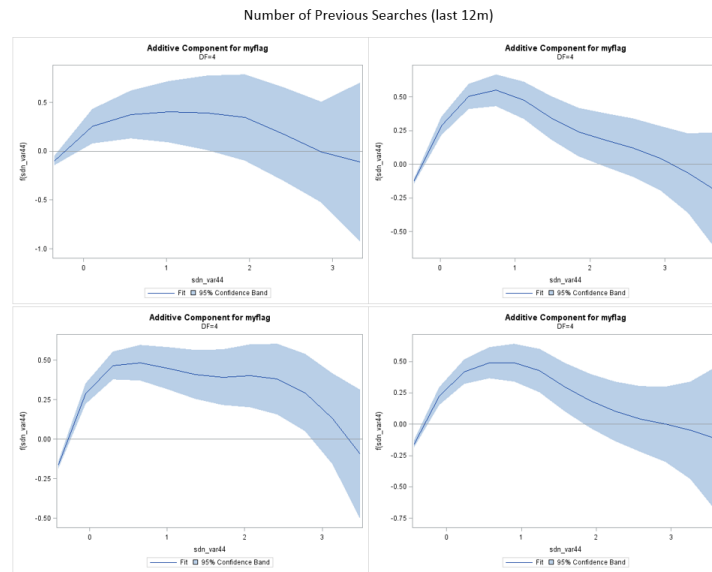


Figure 5.17 additive effects for Number of Previous Searches (last 12m)

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

The linear and smoothing parts are both significant. The summary statistic increases from 2007 to 2010; for example, the sample mean was 0.77 in 2007 and 1.3 in 2010, which means SMEs were seeking for financial support more frequently due to the financial crisis. A quadratic pattern is produced by GAM and the turning point occurs around mean plus SD, which is around four searches within the last 12 months:

1. Below mean minus one SD, a positive influence is observed, with a narrow confidence band. Hence, an increasing number of previous searches results in a better performance. Therefore, if an SME's number of previous searches is less than four, the more searches it had in the last 12 months, the better its performance. Seeking for financial support more than once in the last 12

months improve SMEs' performance most significantly, as the tangent is higher for the part below the sample mean.

2. Above the turning point, the additive effect becomes negative with a wide confidence band. For 'start-ups' which have more than four searches, the more searches results in worse performance. However, there is much more noise in the prediction considering the corresponding wide confidence band.

In summary, the performance of 'start-ups' increases as its number of searches increases, provided it is less than four. Above this point, predictions become less reliable.

- *Lateness of Accounts*

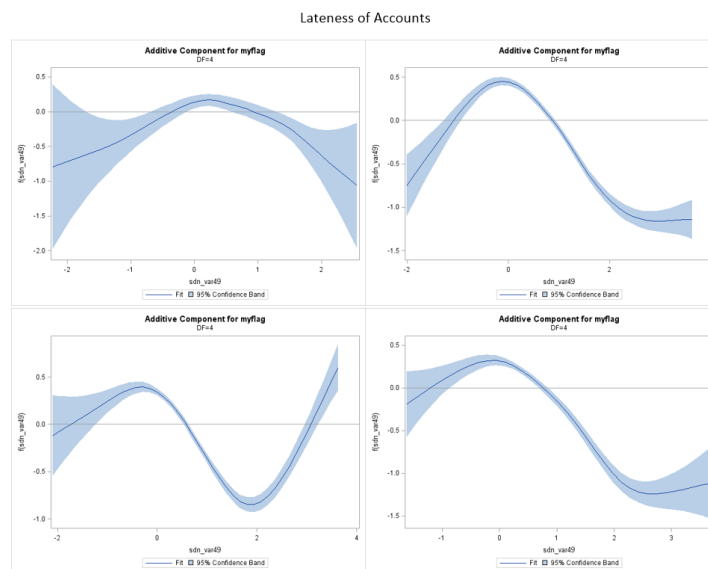


Figure 5.18 additive effects for Lateness of Accounts

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner



---

Its linear part is not significant in 2007 but is significant for the remaining three years, and its smoothing effect is significant over all four years. The additive effect has a clear quadratic form in 2007, a polynomial-like form with a degree of three in 2008 and 2010 and a clear polynomial form with a degree of three in 2009. Hence, the pattern of this variable can be divided into three parts:

1. Below the sample mean, this variable presents a positive influence with a wider confidence band. Hence, if ‘start-ups’ account update duration is shorter than the sample mean, the longer the duration since the last account update, the better performance it will have. It means that if the firm is able to survive with a longer *Lateness of Accounts*, its surviving ability increases as well. However, this prediction comes with higher uncertainty, suggested by the wide confidence band.
2. From sample mean to mean plus two SD, negative influence with narrow confidence band. Hence, *Lateness of Accounts* is most predictive for ‘start-ups’ that fall into this part. ‘Start-ups’ performance decreases as the time since the last accounting update becomes longer.
3. Above mean plus two SD: switched influence through time. This part is shorter, with a negative coefficient and wide confidence band in 2007. It stays flat in 2008 and 2010. However, a clear positive effect is perceived with a narrow confidence band in 2009. It is the most obvious sign of switching for

‘start-up’ SMEs during the financial crisis. ‘Start-ups’ that survived through the ‘credit crunch’ gain ‘swimming’ ability to increase performance.

In summary, *Lateness of Accounts* is most informative for ‘start-ups’ when it falls between sample mean to mean plus two SD, showing a negative influence with narrow confidence band. Below this interval, this variable’s prediction has a higher degree of uncertainty, indicated by the wide confidence band. Above this interval, the ‘start-ups’ performance varies over time.

- *Time since Last Annual Return*

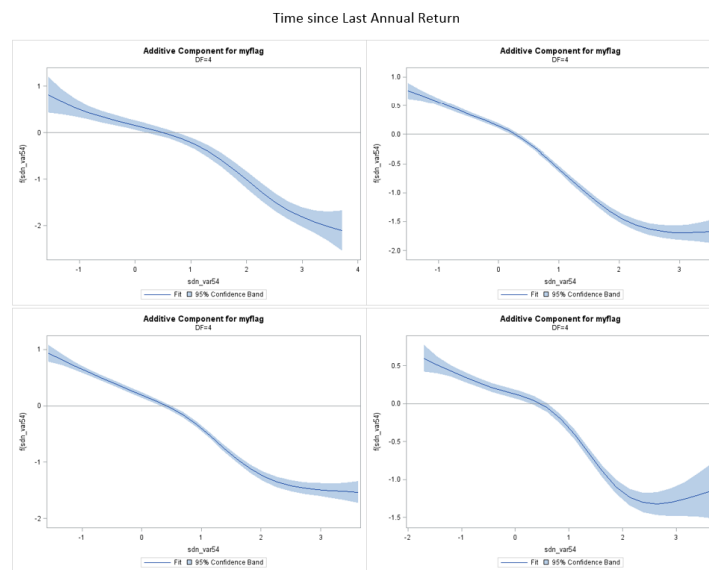


Figure 5.19 additive effects for Time since Last Annual Return

2007 – 2010: the upper left corner; the upper right corner; the lower left corner; the lower right corner

Both its linear component and smoothing component are significant over four years. Its additive effect shows an almost linear pattern. A clear negative coefficient is perceived, with a narrow confidence band. As mentioned previously, *Time since Last Annual Return* marks the duration since the last time the firm reported to Companies

House. The longer the time since last reporting, the more opaque the firm's information. This is a very strong conclusion. It tells banks that even if they cannot collect detailed 'soft' information on SMEs, banks can still separate 'good' SMEs from 'bad' according to the punctuality of their annual returns. This influence is especially strong during the 'credit crunch', which is 2008 and 2009.

In summary, this part presents independent variables trend for 'start-up' SMEs. There are fewer continuous variables analysed in their original format for 'start-ups' due to the distinct performance of missing categories. However, the trend of their additive effects is not more volatile than that of 'non-start-ups'. For example, *Time since Last Annual Return* presents an almost constant decrease pattern. Compared to 'non-start-ups', the tail performance is less volatile since the 'start-ups' records are much more recent.

iii) **Separation measures**

Year	Model	H	Gini	AUC	KS
2007	GAM	0.158	0.520	0.760	0.407
	Log Reg	0.151	0.493	0.747	0.388
2008	GAM	0.316	0.640	0.820	0.503
	Log Reg	0.305	0.618	0.809	0.500
2009	GAM	0.530	0.791	0.896	0.664
	Log Reg	0.517	0.776	0.888	0.652
2010	GAM	0.326	0.654	0.827	0.507
	Log Reg	0.312	0.634	0.817	0.498
Holdout	GAM	0.304	0.615	0.809	0.476
	Log Reg	0.293	0.600	0.802	0.461

Table 5.18 start-ups separation measure: GAM & logistic regression based on data's original format

To test GAM's prediction ability, GAM and logistic regression are applied to the same set of data: eight out of 13 variables are entered in their WoE form, while the

---

other five are used in their original form. Using different separation measures, GAM always provides a better performance than logistic regression. Focusing on AUROC, GAM has the largest improvement in 2007 and its smallest improvement occurs in 2009. Hence, for the ‘start-up’ segment, the non-parametric effects are most powerful in normal time periods, but the financial crisis makes those firms perform more linearly and lose their ability to survive during the financial constraints.

In conclusion, non-parametric effects cause start-up SMEs performance to drift away from the linear predictions before the ‘credit crunch’, while the ‘credit crunch’ makes the non-parametric effects lose their significance. All of the findings coordinate with GAM model using WoE form:

1. non-parametric effects have significant explanatory power in SMEs’ credit scoring;
2. non-parametric effects can improve the prediction of performance;
3. the ‘credit crunch’ makes ‘start-ups’ performance less influenced, not more, by non-parametric effects.

In summary, GAM improves prediction accuracy and demonstrates the marginal improvement in performance of independent variables. The variables’ trends are explained in this section highlighting how SME’s performance varies. For most variables, the additive effect stays constant around the sample mean even through the ‘credit crunch’. Variables’ tail performance is more sensitive to business cycles,

---

especially for 'non-start-ups'. Their performance becomes less stable due to the uncertainty of their supply chains through the 'credit crunch'.

---

## 6. Conclusion

The previous chapter has provided a detailed discussion of the results of employed models, showing various aspects of SMEs' performance during the 'credit crunch'. As the final chapter of this thesis, this chapter explains how it answers the proposed research questions and summarizes how the employed methodologies which not only could improve SMEs' credit risk modelling during a 'credit crunch' but also explore patterns of association between different predictors and outcomes. Limitations and further research topics are also presented at the end of this chapter.

### 6.1 Research questions answered

This research has answered three research questions proposed in the previous chapter:

1. Could a well-defined logistic model provide accurate prediction of SMEs' performance?

This research has carefully selected significant independent variables to model SMEs' performance. Using those variables, logistic regression can predict SMEs' performance accurately, even during the 'credit crunch'. Hence, the standard PD model would not cause significant prediction bias even during a 'credit crunch'.

2. How should modellers employ multi-stage models to SMEs' performance modelling? Further, how could analysts reflect macroeconomic changes during business cycles?

---

Random effects (RE) panel data models are well-established in this research as efficiently incorporating time series effects for SMEs performance prediction. Using MVs, RE panel model could improve PD prediction during the ‘credit crunch’. GDP growth rate is most influential factor for ‘non-start-up’ SMEs, while for ‘start-ups’, it is financial market movements which are more significant.

3. Could a parametric model capture SMEs’ behaviour well during the ‘credit crunch’? If not, can prediction be improved by involving non-parametric effects? Further, what are the derived marginal trend of independent variables?

Panel models’ residuals indicate that the assumptions made by parametric models do not hold. In addition, when applying GAM, most variables exhibited a significant non-parametric component. Hence, standard linear models are likely to predict SMEs performance with some bias. Adding non-parametric components, GAM can significantly improve model fitting whether the data is transformed by WoE or in its original format. Different aspects of SMEs are discussed in detail, such as board information, DBT information, annual return punctuality, lateness of accounts and others. By analysing variables’ marginal effects on ‘non-start-ups’, there is not too much switch of trend around the sample mean even during the ‘credit crunch’. The ‘credit crunch’ usually has an effect on the tails and makes the prediction for those firms less reliable. For example, the prediction could be significantly biased if a ‘non-start-up’ has a large value for lateness of account, a long time since its last annual return, a very high number of judgements or aged derogatory data. Information

---

opacity could be the main difficulty in analysing those firms' performance. Since those firms update their accounting information less frequently. On the contrary, the variables marginal influence of 'start-ups' have even fewer shifts since their information is updated more recently.

In summary, this research analyses UK SMEs' performance during the 'credit crunch'. Both standard models and innovative models are used to answer questions regarding the different aspects of SMEs' credit risk. The standard model is shown to have sufficient separation ability even during the 'credit crunch'. Hence, although SMEs' 'bad' rate is much higher than under normal conditions, these events could be well predicted even by standard models.

It is believed that SMEs' performance is affected by significant economic shifts. Lack of reflection of business cycles is regarded as a major disadvantage of standard models (EBA, 2015). By employing panel data models, this research added MVs to investigate SMEs' performance prediction. Not only can adding MVs improve prediction, but it can also explore how SMEs are influenced by the economy. This improvement helps financial institutions to build strategy to face different business cycles.

A 'black swan' event such as the 'credit crunch' is a situation in which the standard models with their linear assumptions are challenged by this extreme change. By using smoothing components, the linear assumption is avoided and more flexibility is given to fit the 'black swan' event. Variables' marginal effects derived from GAM conclude



---

that, even during extreme cases such as the ‘credit crunch’, firms’ performance is more predictable if their statistics stay around the sample mean. Firms’ performance becomes less predictable when their statistics fall in the tails. These firms should be given more attention and should be required to present more recent information for analysis, especially during the ‘credit crunch’. ‘Start-ups’ do not encounter more non-parametric effects than ‘non-start-ups’, due in part to their restricted data.

This research contribute to the SMEs credit scoring literature in several ways: it builds through circle credit scoring model and analyses MVs influence for UK SMEs during the ‘credit crunch’; an alternative way of processing missing value is proposed for SMEs credit scoring; it also explores explanatory variables non-monotonous pattern.

## **6.2 Limitations and further research**

Due to data restrictions, this research could not judge whether SMEs’ high ‘bad’ rate is caused by economic shifts or strategy changes by financial institutions. For further research, the researcher would suggest using financial institutions’ data to test the influence of strategy changes. If more data were provided, the researcher would be able to carry out an out-of-time validation test.

Derived from GAM, the performance of firms that fall on tails becomes more volatile and estimation bias becomes larger. Information opacity could be the main difficulty

---

for those firms. For further research, the researcher would be interested in obtaining more information and building separate models to analyse those firms.

## Appendix:

Appendix 1: independent variable reference for non-start-ups

non-start-ups			
variable	Characteristic Description	Format	information type
var1	Legal Form	character	general information
var6	Parent Company – derog details	character	general information
var7	1992 SIC Code	character	general information
var9	Region	character	general information
var13	No. Of 'Current' Directors	numerical	directors information
var15	Proportion Of Current Directors To Previous Directors In The Last Year	numerical	directors information
var26	PP Worst (Company DBT - Industry DBT) In The Last 12 Months	numerical	payment and credit records
var37	Total Value Of Judgements In The Last 12 Months	numerical	payment and credit records
var44	Number Of Previous Searches (last 12m)	numerical	payment and credit records
var46	Time since last derogatory data item (months)	numerical	payment and credit records
var49	Lateness Of Accounts	numerical	financial statement
var54	Time Since Last Annual Return	numerical	financial statement
var64	Total Fixed Assets As A Percentage Of Total Assets	numerical	financial statement
var75	Debt Gearing (%)	numerical	financial statement
var76	Percentage Change In Shareholders Funds	numerical	financial statement
var79	Percentage Change In Total Assets	numerical	financial statement

Appendix 2: independent variable reference for start-ups

start-ups			
variable	Characteristic Description	Format	type
var1	Legal Form	character	general information
var4	Company is Subsidiary	character	general information
var7	1992 SIC Code	character	general information
var9	Region	character	general information
var15	Proportion Of Current Directors To Previous Directors In The Last Year	numerical	directors information
var19	Oldest Age Of Current Directors/Proprietors supplied (Years)	numerical	directors information
var20	Number Of Directors Holding Shares	numerical	directors information
var37	Total Value Of Judgements In The Last 12 Months	numerical	payment and credit records
var44	Number Of Previous Searches (last 12m)	numerical	payment and credit records
var46	Time since last derogatory data item (months)	numerical	payment and credit records
var49	Lateness Of Accounts	numerical	financial statement
var54	Time Since Last Annual Return	numerical	financial statement
var58	Total Assets	numerical	financial statement

Appendix 3: imputation results for missing category

non-start-ups				
firm character	2007	2008	2009	2010
No. Of 'Current' Directors	21	21	21	51
Proportion Of Current Directors To Previous Directors In The Last Year	0.3	0.3	1.4	6.5
Time since last derogatory data item (months)	228	224	222	271
Lateness Of Accounts	261	319	354	287
Time Since Last Annual Return	359	413	438	396
Total Fixed Assets As A Percentage Of Total Assets	18.9	69.2	63.1	97.7
Percentage Change In Total Assets	-0.082	0.070	-0.065	0.015
start-ups				
firm character	2007	2008	2009	2010
Oldest Age Of Current Directors/Proprietors supplied (Years)	21	19	24	34
Time since last derogatory data item (months)	17.8	17.1	2.5	2.7
Lateness Of Accounts	-24	-22	-22.69	-6
Time Since Last Annual Return	19	24.8	6.9	24
Total Assets	-0.368	-0.367	-0.369	-0.370

Appendix 4: GAM via WoE parameter estimation for Start-up SMEs

firm character	smoothing component				regression components			
	2007	2008	2009	2010	2007	2008	2009	2010
Company is Subsidiary					1.41***	0.46***	0.04	12.42***
Number Of Directors Holding Shares					0.15***	0.13***	0.15***	0.13*
Total Value Of Judgements In The Last 12 Months					0.62***	0.62***	0.69***	0.43***
Legal Form	0.11***	0.11***	0.13***	0.12***	1.64***	1.88***	1.64***	1.96***
1992 SIC Code	0.71***	0.28***	0.67***	0.61***	1.40***	0.28***	0.14***	0.47***
Region	0.56***	0.61***	0.37***	0.16***	0.54***	0.59***	0.46***	0.79***
Proportion Of Current Directors To Previous Directors In The Last Year	0.01**	0.01***	0.01***	0.01	0.65***	0.23***	0.40***	0.50***
Oldest Age Of Current Directors/Proprietors supplied (Years)	0.28***	0.06***	0.07***	0.15***	0.77***	0.52***	0.46***	0.09
Number Of Previous Searches (last 12m)	0.07	0.07	0.01***	0.03***	1.15***	0.64***	0.52***	0.69***
Time since last derogatory data item (months)	0.01***	0.01***	0.01***	0.01***	0.64***	0.61***	0.56***	0.62***
Lateness Of Accounts	0.01	0.01***	0.01**	0.01***	1.41***	0.46***	0.04***	12.42***
Time Since Last Annual Return	0.06***	0.06***	0.02***	0.06***	0.15***	0.13***	0.15***	0.13***
Total Assets	0.11***	0.11***	0.13**	0.12***	0.62***	0.62***	0.69***	0.43***
intercept					1.64***	1.88***	1.64***	1.96***

---

## Bibliography

- Acs, Z.J., 1992. Small business economics: A global perspective. *Challenge*, 35(6), pp.38-44.
- Acuna, E. and Rodriguez, C., 2004. The treatment of missing values and its effect on classifier accuracy. In *Classification, clustering, and data mining applications* (pp. 639-647). Springer Berlin Heidelberg.
- Allen, L., DeLong, G. and Saunders, A., 2004. Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance*, 28(4), pp.727-752.
- Anderson, R., 2007. *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*. Oxford University Press.
- Angilella, S. and Mazzù, S., 2015. The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, 244(2), pp.540-554.
- Aggelopoulos, E. and Georgopoulos, A., 2015. The Determinants of Shareholder Value in Retail Banking During Crisis Years: The Case of Greece. *Multinational Finance Journal*, 19(2), pp.109-147.
- Agostino, M., Gagliardi, F. and Trivieri, F., 2011. Bank competition, lending relationships and firm default risk: An investigation of Italian SMEs. *International Small Business Journal*, p.0266242611416681.
- Al-Najjar, B., 2015. The Effect of Governance Mechanisms on Small and Medium-Sized Enterprise Cash Holdings: Evidence from the United Kingdom. *Journal of Small Business Management*, 53(2), pp.303-320.
- Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (pp. 199-213). Springer New York.
- Akhavein, J., Frame, W.S. and White, L.J., 2005. The diffusion of financial innovations: An examination of the adoption of small business credit scoring by large banking organizations. *The Journal of Business*, 78(2), pp.577-596.
- Alan, S. and Loranth, G., 2013. Subprime consumer credit demand: evidence from a lender's pricing experiment. *Review of Financial Studies*, 26(9), pp.2353-2374.
- Ang, J.B. and Madsen, J.B., 2012. Risk capital, private credit, and innovative production. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(4), pp.1608-1639.
- Appleyard, L., 2013. The geographies of access to enterprise finance: the case of the West Midlands, UK. *Regional Studies*, 47(6), pp.868-879.
- Arosa, B., Iturralde, T. and Maseda, A., 2013. The board structure and firm performance in SMEs: Evidence from Spain. *Investigaciones Europeas de Dirección y Economía de la Empresa*, 19(3), pp.127-135.
- Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), pp.589-609.
- Altman, E.I. and Rijken, H.A., 2006. A point-in-time perspective on through-the-cycle ratings. *Financial Analysts Journal*, 62(1), pp.54-70.

- 
- Altman, E.I. and Sabato, G., 2007. Modelling credit risk for SMEs: Evidence from the US market. *Abacus*, 43(3), pp.332-357.
- Altman, E.I. and Sabato, G., 2007. Modelling credit risk for SMEs: Evidence from the US market. *Abacus*, 43(3), pp.332-357.
- Altman, E.I., Sabato, G. and Wilson, N., 2008. The value of qualitative information in SME risk management. *Journal of Financial Services Research*, 40(2), pp.15-55.
- Allen, L., DeLong, G. and Saunders, A., 2004. Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance*, 28(4), pp.727-752.
- Allen, L. and Saunders, A., 2004. Incorporating systemic influences into risk measurements: A survey of the literature. *Journal of Financial Services Research*, 26(2), pp.161-191.
- Anagnostopoulos, C., 2012. Measuring classification performance: the hmeasure package. URL: <http://cran.r-project.org/web/packages/hmeasure/vignettes/hmeasure.pdf>.
- Andreeva, G, Ansell, J & Crook, J 2007, 'Modelling Profitability using Survival Combination Scores' *European Journal of Operational Research*, vol 183, pp. 1537-1549.
- Andreeva, G, Ansell, J & Crook, J 2005, 'Modelling the Purchase Propensity: Analysis of a Revolving Store Card' *Journal of the Operational Research Society*, vol 56, pp. 1041-1050.
- Andreeva, G., Calabrese, R. and Osmetti, S.A., 2016. A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models. *European Journal of Operational Research*, 249(2), pp.506-516.
- Angelini, P., Clerc, L., Cúrdia, V., Gambacorta, L., Gerali, A., Locarno, A., Motto, R., Roeger, W., Van den Heuvel, S. and Vlček, J., 2015. Basel III: Long-term Impact on Economic Performance and Fluctuations. *The Manchester School*, 83(2), pp.217-251.
- Angeriz, A. and Arestis, P., 2007. Monetary policy in the UK. *Cambridge Journal of Economics*, 31(6), pp.863-884.
- Angilella, S. and Mazzù, S., 2015. The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, 244(2), pp.540-554.
- Antão, P. and Lacerda, A., 2011. Capital requirements under the credit risk-based framework. *Journal of banking & finance*, 35(6), pp.1380-1390.
- Arellano, M., 1993. On the testing of correlated effects with panel data. *Journal of Econometrics*, 59(1-2), pp.87-97.
- Arellano, M. and Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, 58(2), pp.277-297.
- Arellano, M. and Carrasco, R., 2003. Binary choice panel data models with predetermined variables. *Journal of Econometrics*, 115(1), pp.125-157.
- Arellano, M. and Honoré, B., 2001. Panel data models: some recent developments. *Handbook of econometrics*, 5, pp.3229-3296.



---

Armstrong, A., Davis, E.P., Liadze, I. and Rienzo, C., 2013. Evaluating Changes in Bank Lending to UK SMEs over 2001-12—Ongoing tight credit?. *National Institute of Economic and Social Research: London*.

Auel, M.C. and de Mendonça, H.F., 2011. Macroeconomic relevance of credit channels: Evidence from an emerging economy under inflation targeting. *Economic Modelling*, 28(3), pp.965-979.

Avery, R.B., Bostic, R.W., Calem, P.S. and Canner, G.B., 1996. Credit risk, credit scoring, and the performance of home mortgages. *Fed. Res. Bull.*, 82, p.621.

Azzalini, A., Bowman, A.W. and Härdle, W., 1989. On the use of nonparametric regression for model checking. *Biometrika*, 76(1), pp.1-11.

Baltagi, B., 2008. *Econometric analysis of panel data*. John Wiley & Sons.

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.

Bangia, A., Diebold, F.X., Kronimus, A., Schagen, C. and Schuermann, T., 2002. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2), pp.445-474.

Bank of England, 2010. Credit risks to the UK banking systems, *Financial Stability Report*

Bartoli, F., Ferri, G., Murro, P. and Rotondi, Z., 2014. Bank support and export: evidence from small Italian firms. *Small Business Economics*, 42(2), pp.245-264.

Beaver, W.H., 1966. Financial ratios as predictors of failure. *Journal of accounting research*, pp.71-111.

Bellotti, T. and Crook, J., 2009. Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12), pp.1699-1707.

Baltagi, B.H. and Song, S.H., 2006. Unbalanced panel data: a survey. *Statistical Papers*, 47(4), pp.493-523.

Banerjee, S., 2012. Basel I and Basel II compliance issues for banks in India. *Macroeconomics and Finance in Emerging Market Economies*, 5(2), pp.228-245.

Baraldi, A.N. and Enders, C.K., 2010. An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), pp.5-37.

Barnes, P., 1982. Methodological implications of non-normally distributed financial ratios. *Journal of Business Finance & Accounting*, 9(1), pp.51-62.

Basel Committee, 1999. Principles for the management of credit risk, *Basel Committee on Banking Supervision, Basel*.

Basel Committee, 2006. Basel II: Revised international capital framework, *Basel Committee on Banking Supervision, Basel*.

Basel Committee, 2010. Basel III: A global regulatory framework for more resilient banks and banking systems. *Basel Committee on Banking Supervision, Basel*.

Bastos, R. and Pindado, J., 2013. Trade credit during a financial crisis: A panel data analysis. *Journal of Business Research*, 66(5), pp.614-620.

BCBS, F., 2010. Assessing the macroeconomic impact of the transition to stronger capital and liquidity requirements. *Interim Report, August. Bank for International Settlements, Basel*.

---

Beaver, W. H. (1966): Financial ratios as predictors of failures. *Empirical Research in Accounting*, Supplement to *Journal of Accounting Research*: 71-111.

Beck, N., Katz, J.N. and Tucker, R., 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science*, 42(4), pp.1260-1288.

Behr, P. and Güttler, A., 2007. Credit risk assessment and relationship lending: An empirical analysis of German small and medium-sized enterprises. *Journal of Small Business Management*, 45(2), pp.194-213.

Bennedsen, M., Kongsted, H.C. and Nielsen, K.M., 2008. The causal effect of board size in the performance of small and medium-sized firms. *Journal of Banking & Finance*, 32(6), pp.1098-1109.

Benzoni, L., Collin-Dufresne, P., Goldstein, R.S. and Helwege, J., 2015. Modeling credit contagion via the updating of fragile beliefs. *Review of Financial Studies*, p.hhv018.

Berg, D., 2007. Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2), pp.129-143.

Berger, A.N. and Frame, W.S., 2007. Small business credit scoring and credit availability. *Journal of small business management*, 45(1), pp.5-22.

Berger, A.N., Frame, W.S. and Miller, N.H., 2002. Credit scoring and the availability, price, and risk of small business credit.

Berger, A.N. and Udell, G.F., 1995. Relationship lending and lines of credit in small firm finance. *Journal of business*, pp.351-381.

Berger, A.N. and Udell, G.F., 2002. Small business credit availability and relationship lending: The importance of bank organisational structure. *The economic journal*, 112(477), pp.F32-F53.

Berger, A.N., Hasan, I. and Zhou, M., 2009. Bank ownership and efficiency in China: What will happen in the world's largest nation?. *Journal of Banking & Finance*, 33(1), pp.113-130.

Berger, A.N. and Udell, G.F., 1995. Universal banking and the future of small business lending.

Berlin, M. and Mester, L.J., 2004. Retail credit risk management and measurement: An introduction to the special issue. *Journal of Banking & Finance*, 28(4), pp.721-725.

Bernanke, B.S., 1983. Non-monetary effects of the financial crisis in the propagation of the Great Depression.

Berry, A., Grant, P. and Jarvis, R., 2004. European Bank Lending to the UK SME Sector An Investigation of Approaches Adopted. *International Small Business Journal*, 22(2), pp.115-130.

Björklund, A., 1985. Unemployment and mental health: some evidence from panel data. *Journal of Human Resources*, pp.469-483.

Blazy, R. and Weill, L., 2013. Why do banks ask for collateral in SME lending?. *Applied Financial Economics*, 23(13), pp.1109-1122.

Blume, M.E., Keim, D.B. and Patel, S.A., 1991. Returns and Volatility of Low-Grade Bonds 1977–1989. *The Journal of Finance*, 46(1), pp.49-74.

---

Blundell-Wignall, A., Atkinson, P. and Roulet, C., 2014. Bank business models and the Basel system. *OECD Journal: Financial Market Trends*, 2013(2), pp.43-68.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), pp.307-327.

Bond, E., Tybout, J.R. and Utar, H., 2008. *Credit rationing, risk aversion and industrial evolution in developing countries* (No. w14116). National Bureau of Economic Research.

Brahmbhatt, M. and Da Silva, L.P., 2009. The global financial crisis: comparisons with the Great Depression and scenarios for recovery (No. 11110). The World Bank.

Brewer, E., 2007. On lending to small firms. *Journal of Small Business Management*, 45(1), pp.42-46.

Breeden, J.L. and Thomas, L.C., 2008. The relationship between default and economic cycles for retail portfolios across countries. *Journal of Risk Model Validation*, 2(3), pp.11-47.

Bruns, V. and Fletcher, M., 2008. Banks' risk assessment of Swedish SMEs. *Venture Capital*, 10(2), pp.171-194.

Buddelmeyer, H., Jensen, P.H., Oguzoglu, U. and Webster, E., 2008. Fixed effects bias in panel data estimators. *Available at SSRN 1136288*.

Burkhard, J. and De Giorgi, E.G., 2004. An intensity based non-parametric default model for residential mortgage portfolios. *ETH Zurich, RiskLab Working Paper*.

Camino, D. and Cardone, C., 1999. The valuation and cost of credit insurance schemes for SMEs: the role of the loan guarantee associations. *International Small Business Journal*, 17(4), pp.13-31.

Cameron, A.C. and Trivedi, P.K., 2009. *Microeconometrics using stata* (Vol. 5). College Station, TX: Stata press.

Carbó-Valverde, S., Rodríguez-Fernández, F. and Udell, G.F., 2012, March. Trade credit, the financial crisis, and firm access to finance. In *Central Bank of Ireland Conference on the SME Lending Market* (Vol. 2).

Cardone-Riportella, C., Trujillo-Ponce, A. and Briozzo, A., 2013. Analyzing the role of mutual guarantee societies on bank capital requirements for small and medium-sized enterprises. *Journal of Economic Policy Reform*, 16(2), pp.142-159.

Castro, V., 2013. Macroeconomic determinants of the credit risk in the banking system: The case of the GIPSI. *Economic Modelling*, 31, pp.672-683.

Clark, T.S. and Linzer, D.A., 2015. Should I use fixed or random effects?. *Political Science Research and Methods*, 3(02), pp.399-408.

Comitato di Basilea per la vigilanza bancaria, 2004. *International convergence of capital measurement and capital standards: a revised framework*. Bank for International Settlements.

Conaway, M.R., 1990. A random effects model for binary data. *Biometrics*, pp.317-328.

- 
- Corry, D., Valero, A. and Van Reenen, J., 2011. UK economic performance since 1997: growth, productivity and jobs.
- Cowling, M., Liu, W., Ledger, A. and Zhang, N., 2015. What really happens to small and medium-sized enterprises in a global economic recession? UK evidence on sales and job dynamics. *International Small Business Journal*, 33(5), pp.488-513.
- Chamberlain, G., 1982. Multivariate regression models for panel data. *Journal of econometrics*, 18(1), pp.5-46.
- Chalupka, R. and Kopeckni, J., 2009. Modeling bank loan LGD of corporate and SME segments: a case study. *Czech Journal of Economics and Finance*, 59(4), pp.360-382.
- Chavez-Demoulin, V. and Embrechts, P., 2004. Smooth extremal models in finance and insurance. *Journal of Risk and Insurance*, 71(2), pp.183-199.
- Chelley-Steeley, P.L. and Pentecost, E.J., 1994. Stock market efficiency, the small firm effect and cointegration. *Applied Financial Economics*, 4(6), pp.405-411.
- Chelley-Steeley, P.L. and Steeley, J.M., 1995. Conditional volatility and firm size: an empirical analysis of UK equity portfolios. *Applied Financial Economics*, 5(6), pp.433-440.
- Chen, X., Wang, X. and Wu, D.D., 2010. Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China. *Decision Support Systems*, 49(3), pp.301-310.
- Comment, S., 2008. Corporate Default and Recovery Rates, 1920-2007. *MoodyVs Global Corporate Finance*, February.
- Crook, J. and Bellotti, T., 2010. Time varying and dynamic models for default risk in consumer loans. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2), pp.283-305.
- Crook, J.N., Edelman, D.B. and Thomas, L.C., 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), pp.1447-1465.
- Cuylen, A., Kosch, L. and Breitner, M.H., 2016. Development of a maturity model for electronic invoice processes. *Electronic Markets*, 26(2), pp.115-127.
- Daniel, B. And Nicolae, P., 2011. COLLATERAL'S IMPORTANCE IN SMES FINANCING: WHAT IS THE BANKS' RESPONSE? SOME EVIDENCE FOR ROMANIA. *Annals Of The University Of Oradea, Economic Science Series*, 20(1).
- Das, A. and Ghosh, S., 2007. Determinants of credit risk in Indian state-owned banks: An empirical investigation.
- Das, S.R., Duffie, D., Kapadia, N. and Saita, L., 2007. Common failings: How corporate defaults are correlated. *The Journal of Finance*, 62(1), pp.93-117.
- De la Torre, A., Pería, M.S.M. and Schmukler, S.L., 2010. Bank involvement with SMEs: Beyond relationship lending. *Journal of Banking & Finance*, 34(9), pp.2280-2293.
- Demir, E., 2014. A Decision Support Tool for Predicting Patients at Risk of Readmission: A Comparison of Classification Trees, Logistic Regression, Generalized Additive Models, and Multivariate Adaptive Regression Splines. *Decision Sciences*, 45(5), pp.849-880.

- 
- De Leeuw, E.D., 2001. Reducing missing data in surveys: An overview of methods. *Quality and Quantity*, 35(2), pp.147-160.
- Degryse, H., de Goeij, P. and Kappert, P., 2012. The impact of firm and industry characteristics on small firms' capital structure. *Small Business Economics*, 38(4), pp.431-447.
- Devroye, L. and Györfi, L., 1985. *Nonparametric density estimation: the L1 view* (Vol. 119). John Wiley & Sons Incorporated.
- Dichev, I.D., 1998. Is the risk of bankruptcy a systematic risk?. *the Journal of Finance*, 53(3), pp.1131-1147.
- Dietsch, M. and Petey, J., 2002. The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements. *Journal of Banking & Finance*, 26(2), pp.303-322.
- Dietsch, M. and Petey, J., 2004. Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs. Dietsch, M. and Petey, J., 2010. Sector Concentration Risk in SME Credit Portfolios: a Multifactor Approach. *université de Strasbourg, LARGE, Working Paper*.
- Dietsch, M. and Petey, J., 2010. Sector Concentration Risk in SME Credit Portfolios: a Multifactor Approach. *Journal of Banking & Finance*, 28(4), pp.773-788.
- Dick, J.M., Hussinger, K., Blumberg, B. and Hagedoorn, J., 2013. Is success hereditary? Evidence on the performance of spawned ventures. *Small Business Economics*, 40(4), pp.911-931.
- Dionne, G., Laajimi, S., Mejri, S. and Petrescu, M., 2008. Estimation of the default risk of publicly traded companies: Evidence from Canadian data. *Canadian Journal of Administrative Sciences*, 25(2), p.134.
- Dwyer, D., Kocagil, A. and Stein, R., 2004. The Moody's KMV RiskCalc v3. 1 Model: Next-generation technology for predicting private firm credit risk. *Moody's KMV*.
- Duffie, D. and Lando, D., 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69(3), pp.633-664.
- Duffie, D., Saita, L. and Wang, K., 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), pp.635-665.
- Edmister, R.O., 1972. An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative analysis*, 7(02), pp.1477-1493.
- Elsas, R. and Florysiak, D., 2011. Heterogeneity in the speed of adjustment toward target leverage. *International Review of Finance*, 11(2), pp.181-211.
- Emery, K., Ou, S., Tennant, J., Matos, A. and Cantor, R., 2009. Corporate default and recovery rates, 1920–2008. *Moody's Investors Service, Special Comment, February*.
- Enders, C.K., 2010. *Applied missing data analysis*. Guilford Press.
- Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pp.987-1007.
- Europäische Kommission, 2005. *The new SME definition: User guide and model declaration*. European Comm., Publication Office.

---

European Central Bank, 2012. The development of prices and costs during the 2008-09 recession, *European Central Bank Monthly Bulletin*, pp 71-85,

Ewing, B.T., 2001. Cross-effects of fundamental state variables. *Journal of Macroeconomics*, 23(4), pp.633-645.

Falkner, E.M. and Hiebl, M.R., 2015. Risk management in SMEs: a systematic review of available evidence. *The Journal of Risk Finance*, 16(2), pp.122-144.

Fernandes, G.B. and Artes, R., 2016. Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*, 249(2), pp.517-524.

Fidrmuc, J. and Hainz, C., 2010. Default rates in the loan market for SMEs: Evidence from Slovakia. *Economic Systems*, 34(2), pp.133-147.

Figlewski, S., Frydman, H. and Liang, W., 2012. Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1), pp.87-105.

Fink, J., Grullon, G., Fink, K. and Weston, J., 2004. Firm age and fluctuations in idiosyncratic risk. *Available at SSRN 891173*.

Fong, W.M., 1992. The size effect: a multiperiod analysis. *Applied Financial Economics*, 2(2), pp.87-92.

Frame, W.S., Srinivasan, A. and Woosley, L., 2001. The effect of credit scoring on small-business lending. *Journal of Money, Credit and Banking*, pp.813-825.

Fraser, S., 2009. Small firms in the credit crisis: Evidence from the UK survey of SME finances. *Warwick Business School (ESRC grant no. RES-177-25-0007)*.

Fraser, S., 2012. The impact of the financial crisis on bank lending to SMEs. *Department for Business Innovation & Skills, Report, (12/949)*.

Friedman, J.H. and Stuetzle, W., 1981. Projection pursuit regression. *Journal of the American statistical Association*, 76(376), pp.817-823.

Galton, F., 1886. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, pp.246-263.

Gärtner, M., Griesbach, B. and Jung, F., 2011. PIGS or lambs? The European sovereign debt crisis and the role of rating agencies. *International advances in economic research*, 17(3), pp.288-299.

Gonçalves, V.S., Martins, F.V. and Brandão, E., 2014. *The Determinants of Credit Default on Start-Up Firms. Econometric Modelling using Financial Capital, Human Capital and Industry Dynamics Variables* (No. 534). FEP Working Papers.

Gonçalves, T.S., Ferreira, F.A., Jalali, M.S. and Meidutė-Kavaliauskienė, I., 2015. An idiosyncratic decision support system for credit risk analysis of small and medium-sized enterprises. *Technological and Economic Development of Economy*, pp.1-19.

Gozgor, G., 2014. Determinants of domestic credit levels in emerging markets: the role of external factors. *Emerging Markets Review*, 18, pp.1-18.

Gray, C. and Mabey, C., 2005. Management development key differences between small and large businesses in Europe. *International Small Business Journal*, 23(5), pp.467-485.

---

Greene, W., 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, 7(1), pp.98-119.

Grunert, J. and Norden, L., 2012. Bargaining power and information in SME lending. *Small Business Economics*, 39(2), pp.401-417.

Guest, P.M., 2009. The impact of board size on firm performance: evidence from the UK. *The European Journal of Finance*, 15(4), pp.385-404.

Grunert, J., Norden, L. and Weber, M., 2005. The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, 29(2), pp.509-531.

Gupta, J., Gregoriou, A. and Healy, J., 2015. Forecasting bankruptcy for SMEs using hazard function: To what extent does size matter?. *Review of Quantitative Finance and Accounting*, 45(4), pp.845-869.

Gupta, J., Wilson, N., Gregoriou, A. and Healy, J., 2014. The effect of internationalisation on modelling credit risk for SMEs: Evidence from UK market. *Journal of International Financial Markets, Institutions and Money*, 31, pp.397-413.

Gupta, J., Wilson, N., Gregoriou, A. and Healy, J., 2014. The value of operating cash flow in modelling credit risk for SMEs. *Applied Financial Economics*, 24(9), pp.649-660.

Haaland, K.S. and Sworder, D.D., 1987. *Operator Multiple-Tasking Study for Remotely Operated Platforms*. California Univ San Diego La Jolla Dept Of Applied Mechanics And Engineering Sciences.

Han, L., Benson, A., Chen, J.J. and Zhang, S., 2012. The use and impacts of bank support on UK small and medium-sized enterprises. *International Small Business Journal*, p.0266242612455008.

Han, S.R., Guikema, S.D. and Quiring, S.M., 2009. Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. *Risk analysis*, 29(10), pp.1443-1453.

Härle, P., Lüders, E., Pepanides, T., Pfetsch, S., Poppensieker, T. and Stegemann, U., 2010. Basel III and European banking: Its impact, how banks might respond, and the challenges of implementation. *EMEA Banking*, pp.16-17.

Hassan, G.M. and Wu, E., 2015. Sovereign credit ratings, growth volatility and the global financial crisis. *Applied Economics*, 47(54), pp.5825-5840.

Hastie, T.J. and Tibshirani, R.J., 1990. *Generalized additive models* (Vol. 43). CRC Press.

Hausman, J.A., Hall, B.H. and Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship.

Hermalin, B.E. and Weisbach, M.S., 2001. *Boards of directors as an endogenously determined institution: A survey of the economic literature* (No. w8161). National Bureau of Economic Research.

Hernández-Cánovas, G. and Koëter-Kant, J., 2010. The institutional environment and the number of bank relationships: an empirical analysis of European SMEs. *Small Business Economics*, 34(4), pp.375-390.

- 
- Hernandez-Canovas, G. and Martinez-Solano, P., 2007. Effect of the number of banking relationships on credit availability: Evidence from panel data of Spanish small firms. *Small Business Economics*, 28(1), pp.37-53.
- Hernández-Cánovas, G. and Koëter-Kant, J., 2008. Debt maturity and relationship lending an analysis of European SMEs. *International Small Business Journal*, 26(5), pp.595-617.
- Higón, D.A. and Driffield, N., 2011. Exporting and innovation performance: analysis of the annual small business survey in the UK. *International Small Business Journal*, 29(1), pp.4-24.
- Glasserman, P., 2004. Tail approximations for portfolio credit risk. *The Journal of Derivatives*, 12(2), pp.24-42.
- Hoch, I., 1962. Estimation of production function parameters combining time-series and cross-section data. *Econometrica: journal of the Econometric Society*, pp.34-53.
- Holton, S., Lawless, M. and McCann, F., 2013. SME financing conditions in Europe: credit crunch or fundamentals?. *National Institute Economic Review*, 225(1), pp.R52-R67.
- Honoré, B.E. and Lewbel, A., 2002. Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica*, 70(5), pp.2053-2063.
- Honoré, B.E. and Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4), pp.839-874.
- Hou, A.J., 2013. Asymmetry effects of shocks in Chinese stock markets volatility: A generalized additive nonparametric approach. *Journal of International Financial Markets, Institutions and Money*, 23, pp.12-32.
- Iturralde, T., Maseda, A. and San-Jose, L., 2010. Empirical evidence of banking relationships for Spanish SMEs. *International Small Business Journal*, 28(3), pp.274-295.
- Ivashina, V. and Scharfstein, D., 2010. Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, 97(3), pp.319-338.
- Jacobson, T., Lindé, J. and Roszbach, K., 2005. Credit risk versus capital requirements under Basel II: are SME loans and retail credit really different?. *Journal of Financial Services Research*, 28(1-3), pp.43-75.
- Jacobs Jr, M. and Kiefer, N.M., 2010. The Bayesian approach to default risk: A guide. *Rethinking risk measurement and reporting*, 2, pp.319-334.
- Jacobson, T., Lindé, J. and Roszbach, K., 2006. Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies. *Journal of Banking & Finance*, 30(7), pp.1899-1926.
- Jeffrey, M.W., 2006. Introductory Econometrics: A modern approach. *Canada: South-Western Cengage Learning*.
- Jensen, P., Rosholm, M. and Verner, M., 2001. A comparison of different estimators for panel data sample selection models. *U of Aarhus, Economics Working Paper*, (2002-1).
- Jia, Y.G., 2013. Small business loan guarantees as insurance against aggregate risks. *The BE Journal of Macroeconomics*, 13(1), pp.455-479.



- 
- Ju, Y.H. and Sohn, S.Y., 2014. Updating a credit-scoring model based on new attributes without realization of actual data. *European Journal of Operational Research*, 234(1), pp.119-126.
- Ju, Y., Jeon, S.Y. and Sohn, S.Y., 2015. Behavioral technology credit scoring model with time-dependent covariates for stress test. *European Journal of Operational Research*, 242(3), pp.910-919.
- Ju, Y. and Sohn, S.Y., 2015. Stress test for a technology credit guarantee fund based on survival analysis. *Journal of the Operational Research Society*, 66(3), pp.463-475.
- Kagraoka, Y., 2010. A time-varying common risk factor affecting corporate yield spreads. *The European Journal of Finance*, 16(6), pp.527-539.
- Kavvathas, D., 2001. Estimating credit rating transition probabilities for corporate bonds. Job market paper.
- Keane, M.P. and Runkle, D.E., 1992. On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *Journal of Business & Economic Statistics*, 10(1), pp.1-9.
- Keasey, K. and Watson, R., 1987. Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti's Hypotheses. *Journal of Business Finance & Accounting*, 14(3), pp.335-354.
- Keenan, S.C., Sobehart, J.R. and Hamilton, D.T., 1999. Predicting default rates: a forecasting model for Moody's issuer-based default rates. Available at SSRN 1020303.
- Khandker, S.R. and Koolwal, G.B., 2015. How has microcredit supported agriculture? Evidence using panel data from Bangladesh. *Agricultural Economics*.
- Kim, H.S. and Sohn, S.Y., 2010. Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*, 201(3), pp.838-846.
- Kiff, J., Kisser, M. and Schumacher, M.L., 2013. *Rating through-the-cycle: What does the concept imply for rating stability and accuracy?* (No. 13-64). International Monetary Fund.
- Kim, J. and Scott, C.D., 2012. Robust kernel density estimation. *Journal of Machine Learning Research*, 13(Sep), pp.2529-2565.
- Kitazawa, Y., 2001. Recent Development in Panel Data Econometrics. *Ekonomikusu-the bulletin of the Economics Association of the Kyushu Sangyo University*, 6(1).
- Koopman, S.J. and Lucas, A., 2008. A non-Gaussian panel time series model for estimating and decomposing default risk. *Journal of Business & Economic Statistics*, 26(4), pp.510-525.
- Kozarevic, E., Jukan, M.K. and Softic, A., 2015. An Overview of Small and Medium-Sized Banking Development in Bosnia and Herzegovina. *Journal of Economic and Social Studies*, 5(1), p.107.
- Kremp, E. and Sevestre, P., 2013. Did the crisis induce credit rationing for French SMEs?. *Journal of Banking & Finance*, 37(10), pp.3757-3772.

- 
- Lawless, M. and McCann, F., 2013. Determinants of default: Evidence from a sector-level panel of Irish SME loans. *The Economic and Social Review*, 44(4, Winter), pp.473-488.
- Laurent, M. and Mervi, T., 2015. Risk, capital buffers and bank lending: The adjustment of euro area banks. *Journal of Banking and Financial Economics*, (1 (3)), pp.113-129.
- Lee, C.C., Chiu, Y.B. and Chang, C.H., 2013. Insurance demand and country risks: A nonlinear panel data analysis. *Journal of International Money and Finance*, 36, pp.68-85.
- Lee, S.C., Lin, C.T. and Yang, C.K., 2011. The asymmetric behavior and procyclical impact of asset correlations. *Journal of Banking & Finance*, 35(10), pp.2559-2568.
- Lehmann, B., 2003. Is it worth the while? The relevance of qualitative information in credit rating. *The Relevance of Qualitative Information in Credit Rating (April 17, 2003)*. EFMA.
- Lemmen, J.J. and Goodhart, C.A., 1999. Credit risks and European government bond markets: a panel data econometric analysis. *Eastern Economic Journal*, 25(1), pp.77-107.
- Lehmann, B., 2003. Is it worth the while? The relevance of qualitative information in credit rating. *The Relevance of Qualitative Information in Credit Rating (April 17, 2003)*. EFMA.
- Lensink, R., Van Steen, P. and Sterken, E., 2005. Uncertainty and Growth of the Firm. *Small Business Economics*, 24(4), pp.381-391.
- Levratto, N., 1996. Small firms finance in France. *Small Business Economics*, 8(4), pp.279-295.
- Li, Y., Du, M., Armstrong, A. and Clarke, A., 2012. Board size and performance of small firms: a meta-analysis.
- Li, Z., Crook, J. and Andreeva, G., 2014. Chinese companies distress prediction: an application of data envelopment analysis. *Journal of the Operational Research Society*, 65(3), pp.466-479.
- Lin, S.M., 2007. SMEs Credit Risk Modelling for Internal Rating Based Approach in Banking Implementation of Basel II Requirement.
- Littell, R.C., 1996. SAS. John Wiley & Sons, Ltd.
- Lo, A.W., 1986. Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of econometrics*, 31(2), pp.151-178.
- Lugovskaya, L., 2010. Predicting default of Russian SMEs on the basis of financial and non-financial variables. *Journal of Financial Services Marketing*, 14(4), pp.301-313.
- Lussier, R.N., 1995. A nonfinancial business success versus failure prediction mo. *Journal of Small Business Management*, 33(1), p.8.
- Ma, Y. and Lin, S., 2010. 'Credit crunch' and Small-and Medium-sized Enterprises: Aspects affecting survival. *Journal of Financial Services Marketing*, 14(4), pp.290-300.
- Ma, Y., 2012. Modelling credit risk of small and medium sized enterprises using transactional, accounting and market variables.

- 
- Maddala, G.S., 1987. Recent developments in the econometrics of panel data analysis. *Transportation Research Part A: General*, 21(4), pp.303-326.
- Maddala, G.S., 1971. The use of variance components models in pooling cross section and time series data. *Econometrica: Journal of the Econometric Society*, pp.341-358.
- Maes, J., Sels, L. and Roodhooft, F., 2005. Modelling the link between management practices and financial performance. Evidence from small construction companies. *Small Business Economics*, 25(1), pp.17-34.
- Mallin, C. and Ow-Yong, K., 1998. Corporate governance in small companies—the alternative investment market. *Corporate Governance: An International Review*, 6(4), pp.224-232.
- 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.
- Manski, C.F., 1987. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society*, pp.357-362.
- Mari, C. and Renò, R., 2005. Credit risk analysis of mortgage loans: an application to the Italian market. *European Journal of Operational Research*, 163(1), pp.83-93.
- Matias Gama, A.P. and Susana Amaral Geraldes, H., 2012. Credit risk assessment and the impact of the New Basel Capital Accord on small and medium-sized enterprises: An empirical analysis. *Management Research Review*, 35(8), pp.727-749.
- McCann, F. and McIndoe-Calder, T., 2015. Firm size, credit scoring accuracy and banks' production of soft information. *Applied Economics*, 47(33), pp.3594-3611.
- McGuinness, G. and Hogan, T., 2016. Bank credit and trade credit: Evidence from SMEs over the financial crisis. *International Small Business Journal*, 34(4), pp.412-445.
- Mendoza, E.G., Quadrini, V. and Rios-Rull, J.V., 2007. *Financial integration, financial deepness and global imbalances* (No. w12909). National Bureau of Economic Research.
- Merton, R.C., 1973. Theory of rational option pricing. *The Bell Journal of economics and management science*, pp.141-183.
- Miller, M. and Rojas, D., 2004. Improving access to credit for smes: an empirical analysis of the viability of pooled data sme credit scoring models in Brazil, Colombia & Mexico. *World Bank working paper*.
- Mills, T.C., 2009. *Palgrave handbook of econometrics* (Vol. 2). Springer.
- Milošević, N., Rakočević, S.B. and Milanović, N., 2014. One Approach to Risk Management in SMEs Banking. *Management (1820-0222)*, (73).
- Mintrom, M., 1997. Policy entrepreneurs and the diffusion of innovation. *American journal of political science*, pp.738-770.
- Montanari, A. and Viroli, C., 2010. Heteroscedastic factor mixture analysis. *Statistical Modelling*, 10(4), pp.441-460.
- Moon, T.H., Kim, Y. and Sohn, S.Y., 2011. Technology credit rating system for funding SMEs. *Journal of the Operational Research Society*, 62(4), pp.608-615.

---

Moscone, F., Tosetti, E. and Canepa, A., 2014. Real estate market and financial stability in US metropolitan areas: A dynamic model with spatial effects. *Regional Science and Urban Economics*, 49, pp.129-146.

Mityš, L. and Sevestre, P. eds., 2008. *The econometrics of panel data: Fundamentals and Recent Developments in Theory and Practice*. Springer.

Naifar, N., 2011. What explains default risk premium during the financial crisis? Evidence from Japan. *Journal of Economics and Business*, 63(5), pp.412-430.

Nassr, I.K. and Wehinger, G., 2015. Unlocking SME finance through market-based debt. *OECD Journal: Financial Market Trends*, 2014(2), pp.89-190.

Neguriță, O. and Trandafir, A., 2014. The impact of the fourth capital requirements directive on the credit institutions in Romania. *Economics, Management, and Financial Markets*, (4), pp.193-200.

Nickell, P., Perraudin, W. and Varotto, S., 2000. Stability of rating transitions. *Journal of Banking & Finance*, 24(1), pp.203-227.

Nickell, P., Perraudin, W. and Varotto, S., 2001. *Ratings versus equity-based credit risk modelling: an empirical analysis*. Bank of England.

Nickell, P., Perraudin, W. and Varotto, S., 2007. Ratings-based credit risk modelling: An empirical analysis. *International Review of Financial Analysis*, 16(5), pp.434-451.

Nitani, M. and Riding, A., 2014. Risk factors and the Canada Small Business Financing Program. *Journal of Small Business & Entrepreneurship*, 27(3), pp.251-274.

North, D., Baldock, R. and Ekanem, I., 2010. Is there a debt finance gap relating to Scottish SMEs? A demand-side perspective. *Venture Capital*, 12(3), pp.173-192.

Nunes, P.M., Serrasqueiro, Z.S. and Leitao, J., 2010. Are there nonlinear relationships between the profitability of Portuguese service SME and its specific determinants?. *The Service Industries Journal*, 30(8), pp.1313-1341.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pp.109-131.

Ono, A., Hasumi, R. and Hirata, H., 2014. Differentiated use of small business credit scoring by relationship lenders and transactional lenders: Evidence from firm–bank matched data in Japan. *Journal of Banking & Finance*, 42, pp.371-380.

Ou, S., Chlu, D. and Metz, A., 2011. Corporate default and recovery rates, 1920–2010. *Special Comment, Moody's Investors Service*.

Owualah, S.I., 1999. Banking crisis, reforms, and the availability of credit to Japanese small and medium enterprises: A research note. *Asian Survey*, 39(4), pp.656-667.

Paulet, E., Parnaudeau, M. and Abdessemed, T., 2014. The SME struggle for financing: a clampdown in European banks post-crisis. *Journal of Business Strategy*, 35(2), pp.36-45.

Paull, A.E., 1978. A generalized compound Poisson model for consumer purchase panel data analysis. *Journal of the American Statistical Association*, 73(364), pp.706-713.

Pederzoli, C., Thoma, G. and Torricelli, C., 2013. Modelling credit risk for innovative SMEs: The role of innovation measures. *Journal of Financial Services Research*, 44(1), pp.111-129.

- 
- Pesaran, M.H., Schuermann, T., Treutler, B.J. and Weiner, S.M., 2006. Macroeconomic dynamics and credit risk: a global perspective. *Journal of Money, Credit, and Banking*, 38(5), pp.1211-1261.
- Crosbie, P., 2003. Modelling Default Risk Modeling methodology, *Moody's KMV Company*
- Portman-Smith, C. and Harwood, I.A., 2015. 'Only as good as your last gig?': an exploratory case study of reputational risk management amongst self-employed musicians. *Journal of Risk Research*, 18(4), pp.483-504.
- Prati, A., Schindler, M. and Valenzuela, P., 2012. Who benefits from capital account liberalization? Evidence from firm-level credit ratings data. *Journal of International Money and Finance*, 31(6), pp.1649-1673
- Presbitero, A.F., Udell, G.F. and Zazzaro, A., 2014. The home bias and the credit crunch: A regional perspective. *Journal of Money, Credit and Banking*, 46(s1), pp.53-85.
- Reinhart, C.M. and Rogoff, K.S., 2008. *Is the 2007 US sub-prime financial crisis so different? An international historical comparison* (No. w13761). National Bureau of Economic Research.
- Reveley, J., Down, S. and Taylor, S., 2004. Beyond the Boundaries An Ethnographic Analysis of Spatially Diffuse Control in a Small Firm. *International small business journal*, 22(4), pp.349-367.
- Rice, T. and Strahan, P.E., 2010. Does Credit Competition Affect Small-Firm Finance?. *The Journal of Finance*, 65(3), pp.861-889.
- Roach, K., 2008. Measuring The Effects Of Concentration And Risk On Bank Returns: Evidence From A Panel Of Individual Loan Portfolios In Jamaica R. Brian Langrin And.
- Rigby, R.A. and Stasinopoulos, D.M., 2012. A flexible regression approach using GAMLSS in R, 2010.
- Rottmann, H. and Wollmershäuser, T., 2013. A micro data approach to the identification of credit
- Rodríguez-Rodríguez, O.M., 2006. Trade credit in small and medium size firms: an application of the system estimator with panel data. *Small Business Economics*, 27(2-3), pp.103-126.
- crunches. *Applied Economics*, 45(17), pp.2423-2441.
- Ruud, P.A., 1986. Consistent estimation of limited dependent variable models despite misspecification of distribution. *Journal of Econometrics*, 32(1), pp.157-187.
- Sánchez-Vidal, J. and Martín-Ugedo, J.F., 2012. Are the implications of the financial growth cycle confirmed for Spanish SMEs?. *Journal of Business Economics and Management*, 13(4), pp.637-665.
- Sangalli, I., 2013. Inventory investment and financial constraints in the Italian manufacturing industry: A panel data GMM approach. *Research in Economics*, 67(2), pp.157-178.

- 
- Santomero, A.M. and Vinso, J.D., 1977. Estimating the probability of failure for commercial banks and the banking system. *Journal of Banking & Finance*, 1(2), pp.185-205.
- Saridakis, G., Mole, K. and Hay, G., 2012. Liquidity constraints in the first year of trading and firm performance. *International Small Business Journal*, p.0266242611410091.
- Savvopoulos, A., 2010. Consumer Credit Models: Pricing, Profit and Portfolios. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2), pp.468-468.
- Setodji, C.M., Scheuner, M., Pankow, J.S., Blumenthal, R.S., Chen, H. and Keeler, E., 2012. A graphical method for assessing risk factor threshold values using the generalized additive model: the multi-ethnic study of atherosclerosis. *Health Services and Outcomes Research Methodology*, 12(1), pp.62-79.
- Scellato, G. and Ughetto, E., 2010. The Basel II reform and the provision of finance for R & D activities in SMEs: An analysis of a sample of Italian companies. *International Small Business Journal*, 28(1), pp.65-89.
- Schizas, E., 2011. Framing the debate: Basel III and SMEs. *Association of Chartered Certified Accountants (ACCA) publication*.
- Schularick, M. and Taylor, A.M., 2012. Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *The American Economic Review*, 102(2), pp.1029-1061.
- Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model\*. *The Journal of Business*, 74(1), pp.101-124.
- Sims, C.A., 1972. Money, income, and causality. *The American economic review*, 62(4), pp.540-552.
- Slovik, P. and Cournède, B., 2011. Macroeconomic impact of Basel III.
- Slovin, M.B., Johnson, S.A. and Glascock, J.L., 1992. Firm size and the information content of bank loan announcements. *Journal of Banking & Finance*, 16(6), pp.1057-1071.
- Smallbone, D., Deakins, D., Battisti, M. and Kitching, J., 2012. Small business responses to a major economic downturn: Empirical perspectives from New Zealand and the United Kingdom. *International Small Business Journal*, p.0266242612448077.
- So, M.M. and Thomas, L.C., 2010. Modeling and model validation of the impact of the economy on the credit risk of credit card portfolios. *The Journal of Risk Model Validation*, 4(4), p.93.
- Sohn, S.Y., Doo, M.K. and Ju, Y.H., 2012. Pattern recognition for evaluator errors in a credit scoring model for technology-based SMEs. *Journal of the Operational Research Society*, 63(8), pp.1051-1064.
- Sohn, S.Y. and Kim, Y.S., 2013. Behavioral credit scoring model for technology-based firms that considers uncertain financial ratios obtained from relationship banking. *Small Business Economics*, 41(4), pp.931-943.

- 
- Staikouras, S.K., 2005. Multinational banks, credit risk, and financial crises: A qualitative response analysis. *Emerging Markets Finance and Trade*, 41(2), pp.82-106.
- Stepanova, M. and Thomas, L., 2002. Survival analysis methods for personal loan data. *Operations Research*, 50(2), pp.277-289.
- Supervision, B.C.B. and A brief history of the Basel Committee, 2003. International convergence of capital measurement and capital standards—A revised framework—Comprehensive version.
- Sutradhar, B.C., Bari, W. and Das, K., 2010. On probit versus logit dynamic mixed models for binary panel data. *Journal of Statistical Computation and Simulation*, 80(4), pp.421-441.
- Syed, M., 2008. Forecasting financial default of Small Medium Enterprises. *Thesis (5912008)*, university of Amsterdam.
- Tajik, M., Aliakbari, S., Ghaliya, T. and Kaffash, S., 2015. House prices and credit risk: Evidence from the United States. *Economic Modelling*, 51, pp.123-135.
- Thomas, A., 2006. Consistent estimation of binary-choice panel data models with heterogeneous linear trends. *The Econometrics Journal*, 9(2), pp.177-195.
- Thomas, L.C., Edelman, D.B. and Crook, J.N., 2004. *Readings in credit scoring: foundations, developments, and aims*. Oxford University Press on Demand.
- Thomas, R., Hills, S. and Dimsdale, N., 2010. The UK recession in context—what do three centuries of data tell us?. *Bank of England Quarterly Bulletin*, p.Q4.
- Tong, E.N., Mues, C., Brown, I. and Thomas, L.C., 2016. Exposure at default models with and without the credit conversion factor. *European Journal of Operational Research*, 252(3), pp.910-920.
- Tong, E.N., Mues, C. and Thomas, L.C., 2012. Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, 218(1), pp.132-139.
- Vallés, V., 2006. *Stability of a "through-the-cycle" Rating System During a Financial Crisis: FSI Award 2006 Winning Paper*. Financial Stability Inst., Bank for Internat. Settlements.
- Van Gestel, T. and Baesens, B., 2009. *Credit Risk Management: Basic concepts: Financial risk components, Rating analysis, models, economic and regulatory capital*. Oxford University Press.
- Verbano, C. and Venturini, K., 2013. Managing risks in SMEs: A literature review and research agenda. *Journal of technology management & innovation*, 8(3), pp.186-197.
- Vos, E., Yeh, A.J.Y., Carter, S. and Tagg, S., 2007. The happy story of small business financing. *Journal of Banking & Finance*, 31(9), pp.2648-2672.
- Wang, M.S., 2013. Idiosyncratic risk and expected returns: a panel data model with random effects. *Applied Financial Economics*, 23(10), pp.869-880.
- Wehinger, G., 2012. Bank deleveraging, the move from bank to market-based financing, and SME financing. *OECD Journal: Financial Market Trends*, 2012(1), pp.65-79.

- 
- Wehinger, G., 2014. SMEs and the credit crunch. *OECD Journal: Financial Market Trends*, 2013(2), pp.115-148.
- Wetmore, J.L., 2004. Panel Data, liquidity risk, and increasing Loans-to-Core Deposits ratio of large commercial bank holding companies [1992-2000]. *American Business Review*, 22(2), p.99.
- Williams, R., 2015. Panel data 4: Fixed effects vs random effects models. *University of Notre Dame*.
- Wilson, T.C., 1998. Portfolio credit risk. *Economic Policy Review*, 4(3).
- Wilson, N., Wright, M. and Altanlar, A., 2013. The survival of newly-incorporated companies and founding director characteristics. *International Small Business Journal*, p.0266242613476317.
- Winkelmann, L. and Winkelmann, R., 1998. Why are the unemployed so unhappy? Evidence from panel data. *Economica*, 65(257), pp.1-15.
- Wooldridge, J.M., 2002. *Econometric analysis of cross section and panel data*. MIT press.
- Wooldridge, J.M., 2013. *Introductory econometrics: A modern approach*. Nelson Education.
- Wu, J., Luca, A.C. and Jeon, B.N., 2011. Foreign bank penetration and the lending channel in emerging economies: Evidence from bank-level panel data. *Journal of International Money and Finance*, 30(6), pp.1128-1156.
- Xiang, D., 2001. Fitting generalized additive models with the GAM procedure. *SAS Institute, Paper P256-26*.
- Yang, C.H. and Chen, K.H., 2009. Are small firms less efficient?. *Small Business Economics*, 32(4), pp.375-395.
- Yule, G.U., 1897. On the theory of correlation. *Journal of the Royal Statistical Society*, 60(4), pp.812-854.
- Zeng, D. and Cai, J., 2010. Additive transformation models for clustered failure time data. *Lifetime data analysis*, 16(3), pp.333-352.
- Zhang, L., Hu, H. and Zhang, D., 2015. A credit risk assessment model based on SVM for small and medium enterprises in supply chain finance. *Financial Innovation*, 1(1), p.1.