UNIVERSITY OF Hull

Essays on SMEs Insolvency Risk

being a Thesis submitted for the Degree of
Doctor of Philosophy in Finance
in the University of Hull

by

Jairaj Gupta
BBA, MSc.

Hull University Business School
Student No. 201201551
October, 2014

CHAPTERS PUBLISHED AND UNDER REVIEW

The following chapters of my thesis are either published in refereed journal or are under review:

CHAPTER 3

Gupta, J., Wilson, N., Gregoriou, A., & Healy, J. (2014). *The Value of Operating Cash Flow in Modelling Credit Risk for SMEs.* **Applied Financial Economics**, 24(9), 649–660. (DOI: 10.1080/09603107.2014.896979) (ABS 2*)

CHAPTER 4

Gupta, J., Wilson, N., Gregoriou, A., & Healy, J. (2014). *The Effect of Internationalization on Modelling Credit Risk for SMEs: Evidence from UK Market*. **Journal of International Financial Markets, Institutions & Money**, 31, 397–413. (DOI: 10.1016/j.intfin.2014.05.001) (ABS 3*)

CHAPTER 5

Gupta, J., Gregoriou, A., & Healy, J. (2014). *Forecasting Bankruptcy for SMEs Using Hazard Function: To What Extent Does Size Matter?* **Review of Quantitative Finance and Accounting**. (DOI: 10.1007/s11156-014-0458-0) (ABS 3*)

CHAPTER 6

Gupta, J., Gregoriou, A., & Healy, J. (2014). *Using Hazard Models Correctly: A Comparison Employing Different Definitions of SMEs Financial Distress*. Under Review: **European Journal of Operational Research**. (SSRN: ssrn.com/abstract=2457917) (ABS 3*)

ABSTRACT

In light of the new Basel Capital Accord, Small and medium size enterprises (SMEs) play a fundamental role in the economic performance of major economies. Several lending communities proposed to treat SMEs as retail clients to optimize capital requirements and profitability. In this context, it is becoming critically important to have a detailed understanding of its risk behavior for appropriate pricing of credit risk. Thus, this thesis presents *four essays* on SMEs insolvency risk starting from chapter 3 through chapter 6 that investigates different dimensions of their default risk. My *first essay* makes distinction among SMEs that report operating cash flow and those which do not while modeling their default risk. However, I do not report any significant improvement in model's classification performance when operating cash flow information is made available. Similarly, my *second essay* considers domestic and international SMEs separately while modelling their default risk and report almost identical classifications performance of the models' developed for both the groups. The *third essay* compares the default risk attributes of micro, small and medium-sized firms respectively with SMEs. Test results suggest significant difference in the default risk attributes of only micro firms and SMEs. On a different line, my *fourth essay* deals with the methodological issues that have been witnessed recently in the bankruptcy literature that use hazard models for making bankruptcy predictions. This essay highlights the critical issues and provides appropriate guidance for the correct use of hazard models in making bankruptcy predictions. Here, I also propose a default definition for SMEs which considers both legal bankruptcy laws and firms' financial health while defining the default event. Empirical results show that my default definition performs significantly better than its respective counterparts in identifying distressed firms with superior goodness of fit measures across all econometric specifications. Detailed abstract of respective essays are as follows.

Evidence pertaining to SMEs financing strongly motivates me to believe that firms which are unable to generate sufficient operating cash flow (OCF) are more susceptible to bankruptcy. However, the role of OCF in bankruptcy of SMEs lacks empirical validation. Thus, my *first essay* (chapter 3) investigates the role of operating cash flow information as predictors in assessing the creditworthiness of SMEs. One-year distress prediction model developed using significant financial information of United Kingdom SMEs over a period of 2000 to 2009 confirm that the presence of operating cash flow information does not improve the prediction accuracy of the distress prediction model.

My *second essay* (chapter 4) considers domestic and international small and medium-sized enterprises (SMEs) of the United Kingdom separately while modelling their default risk. To establish the empirical validation, separate one-year default prediction models are developed using dynamic logistic regression technique that encapsulates significant financial information over an analysis period of 2000 to 2009. Almost an identical set of explanatory variables affect the default probability of domestic and international SMEs, which contradicts the need for separate default risk models. However, the lower predictive accuracy measures of the model developed for international SMEs motivate me to compare the weights of regression coefficients of the models developed for domestic and international firms. Test results confirm that four out of the nine common predictors display significant statistical differences in their weights. However, these differences do not contribute to the discriminatory performance of the default prediction models, given that I report very little difference in each model's classification performance.

A huge diversity exists within the broad category of Small and medium size enterprises (SMEs). They differ widely in their capital structure, firm size, access to external finance, management style, numbers of employees etc. Thus, my *third essay* (chapter 5) contributes to the literature by acknowledging this diversity while modeling credit risk for them, using a relatively large UK database, covering the analysis period between 2000 and 2009. My analysis partially employs the definition provided by the European Union to distinguish between 'micro', 'small', and 'medium' sized firms. I use both financial and non-financial information to predict firms' failure hazard. I estimate separate hazard models for each sub-category of SMEs, and compare their performance with a SMEs hazard model including all the three sub-categories. I test my hypotheses using discrete-time duration-dependent hazard rate modelling techniques, which controls for both macro-economic conditions and survival time. My test results strongly highlight the differences in the credit risk attributes of 'micro' firms and SMEs, while it does not support the need to consider 'small' and 'medium' firms' category separately while modelling credit risk for them, as almost the same sets of explanatory variables affect the failure hazard of SMEs, 'small' and 'medium' firms.

My *fourth essay* (chapter 6) considers all serious and neglected concerns while developing discrete and continuous time duration dependent hazard models for predicting failure of US SMEs. I compare theoretical and classification performance aspects of three popular hazard models, namely discrete hazard models with logit and clog-log links and the extended Cox model. I report that discrete hazard models are superior to extended Cox models in making default predictions. I also propose a default definition for SMEs which considers both legal bankruptcy laws and firms' financial health while defining the default event. My empirical results show that my default definition performs significantly better than the default definitions which are only based on legal consequence or firms' financial health in identifying distressed firms. In addition, my default definition also shows superior goodness of fit measures across all econometric specifications.

ACKNOWLEDGEMENTS

I will always feel privileged and grateful to Hull University Business School for providing me their much valued PhD Studentship to undertake PhD in Finance study in their Accounting and Finance department. Since the beginning of this precious and valuable journey, Professor Andros Gregoriou has been my supervisor, mentor, strategic advisor and friend. I have been extremely fortunate to work and develop as a researcher under his valuable, expert and professional guidance. I am extremely grateful to him for continuously motivating me, and pushing me to attain higher benchmarks. He took me through the entire process of research and publication which has been a valuable learning experience. I would also like to give my sincere gratitude to my second supervisor Dr. Jerome Healy. He has always been extremely supportive toward all our decision and all these years we enjoyed working together as a team.

I am also very thankful to the Graduate Research Director of Hull University Business School, Dr. Mark Rhodes for his continuous cooperation and motivation throughout my PhD programme. All the staff members of the Graduate Research Office and Finance Office also deserve my gratitude for their valuable and efficient administrative support.

Finally, I would love to render my gratitude to my friends, fellow colleagues and family members for their unconditional support and motivation throughout this journey.

TABLE OF CONTENTS

FIGURES AND TABLES

1. Future Champions: Small and Medium-Sized Enterprises

Small and medium-sized enterprises (SMEs) constitute the most vibrant sector of corporations in both the United Kingdom (UK) and the United States (US). They possess the potential to play the role of a powerful economic engine of growth and development as the UK and the US forge their paths toward economic recovery. However, until recently their potential has been neglected and overlooked by policy makers, the banking community and other stakeholders. SMEs represent a highly diversified group of companies that shows low asset correlation unlike large firms (Dietsch and Petey 2004) and hence does not contribute significantly to the systematic risk. If the UK and the US are successful in creating an environment and conditions favorable for the growth and development of their SMEs, domestically and in international markets, their economic landscape would be transformed by the resulting contribution to sustainable growth and employment creation. But to bring about such conditions and environments requires a deeper understanding of the challenges to growth and complex business dynamics of SMEs. Empirical literature highlight the diversity that exists within the broad SMEs category (micro, small and medium firms) in respect of access to finance (Beck *et al.* 2006), number of employees, management characteristics (Wager 1998) etc. This diversity has finally come to the notice of the banking community which has started developing customized solution to tap the growth potential of this profitable business segment (IFC 2010). Next, we highlight some of the dimensions pertaining to importance, performance and functioning of small business units.

1.1 SMEs Definition

The aftermath of the recent global financial crisis has generated renewed interest in the importance of SMEs to the growth and stability of the global economy. Consequently, agencies responsible for countries economic prosperity and stability are endeavoring to create the best combinations of policy and support to realize their growth potential (see for details Doing Business 2013). However, the definition of SMEs is critical to the overall process of policy formulation, and the creation of a conducive business environment. Currently there is a consensus that the SMEs market is significantly important in size and performance. However, there is considerable variation in their definition across the globe. The definitions of SMEs vary across countries based upon differing qualitative and quantitative parameters such as turnover, number of employees, industrial sector, capital investment, independence, legal status, and asset size. A common definition of SMEs, however, is a registered business with less than two hundred and fifty employees, and this incorporates the vast majority of business units within the SMEs segment. This category is sometimes, further narrowed down by distinguishing firms within the SMEs category as micro, small and medium. Micro enterprises are usually defined as having less than 5 or 10 employees; small firms usually as having less than 50 employees and medium as having less than 250 employees; however there is variation on the precise dividing line. It is worth noting the widely acceptable definition of SMEs offered by the European Union to the extent we arrive at a mutually agreed consensus. It defines a firm as ‘micro’ if it has less than 10 employees with an annual turnover of less than € 2 million; ‘small’ if it has less than 50 employees with an annual turnover of less than € 10 million and ‘medium’ if it has less than 250 employees with an annual turnover of less than € 50 million.

Under the UK regulatory requirements, companies are required to file annual summary financial accounts (income statement, balance sheet, and cash-flow statement) with the

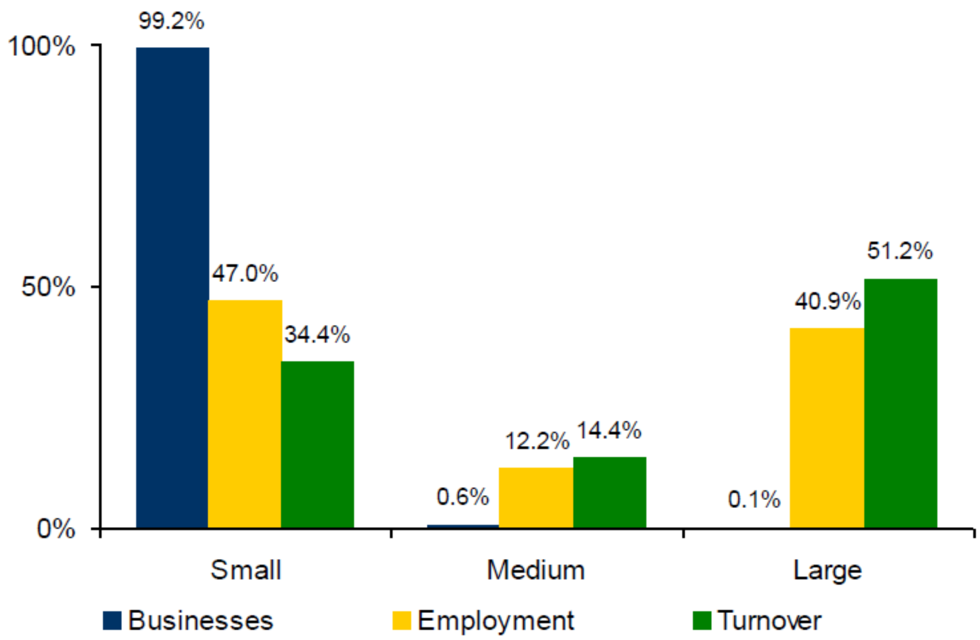
'Companies House' (www.companieshouse.gov.uk). The UK registrar of companies defines a small company as one for which at least two of the following conditions are satisfied: (i) Annual turnover is £6.5 million or less; (ii) the balance sheet total is £3.26 million or less; (iii) the average number of employees is 50 or fewer. It defines medium-sized company as one for which at least two of the following conditions are satisfied: (i) annual turnover must be no more than £25.9 million; (ii) the balance sheet total must be no more than £12.9 million; (iii) the average number of employees must be no more than 250.

The majority of banks serving the SMEs sector use rigid quantitative definitions of the types described above. However, this type of classification is inefficient for several reasons. For example, in developing or developed high-income countries the majority of the SMEs segment comprises of firms having less than 5 employees, and thus as per several definitions they are classified as micro firms. On the other hand, an SME classed as medium size in a high-income country may be the same size as one defined as a large business corporation in a low-income nation. Further complications arise because most SMEs function within the informal sector and hence are often not counted within the SMEs sector, but they nevertheless represent a potential profitable market for banks serving the SMEs community. Conceptually the SMEs banking sector represent the missing position between the large formal business corporations and highly informal micro enterprises. The development of commercial banking model primarily took into account these large corporate enterprises while developing their business models and historically they have been found managing very high-value transactions for a small segment of low-risk corporate clients. Recently we see the micro finance institutions (MFI) emerging as a significant player to tap this market segment which has been neglected by the traditional banking community. Traditionally the banking community have viewed this segment as too small to serve coupled with high risk and business complexities, as their operation methodology is significantly different than the large enterprises with less financial sophistication, lack of business planning and cash flow management skills. However, they have realized the importance of SMEs portfolio and are developing customized banking solutions to serve this sector effectively (IFC 2010).

1.2 Overview of the SME sector in the UK

The United Kingdom faced adverse economic conditions in 2011/12, with the possibility of entering a double-dip recession, and facing low/negative growth rates, as well as sovereign debt downgrades. However, SMEs have emerged as resilient performers throughout this period. There is an estimated 4.8 million firms employing about 23.9 million individuals operating in the UK (BIS 2012). Small business units (having less than 50 employees) comprise 99.2 percent of this total. While, there are 30,000 medium-sized firms (having number of employees between 50 and 249) and 6,000 large firms (having 250 or more employees). Together, small and medium firms contribute to about 49% of the economy's turnover (see Figure 1.1), signifying their importance in the growth and development of the UK economy.

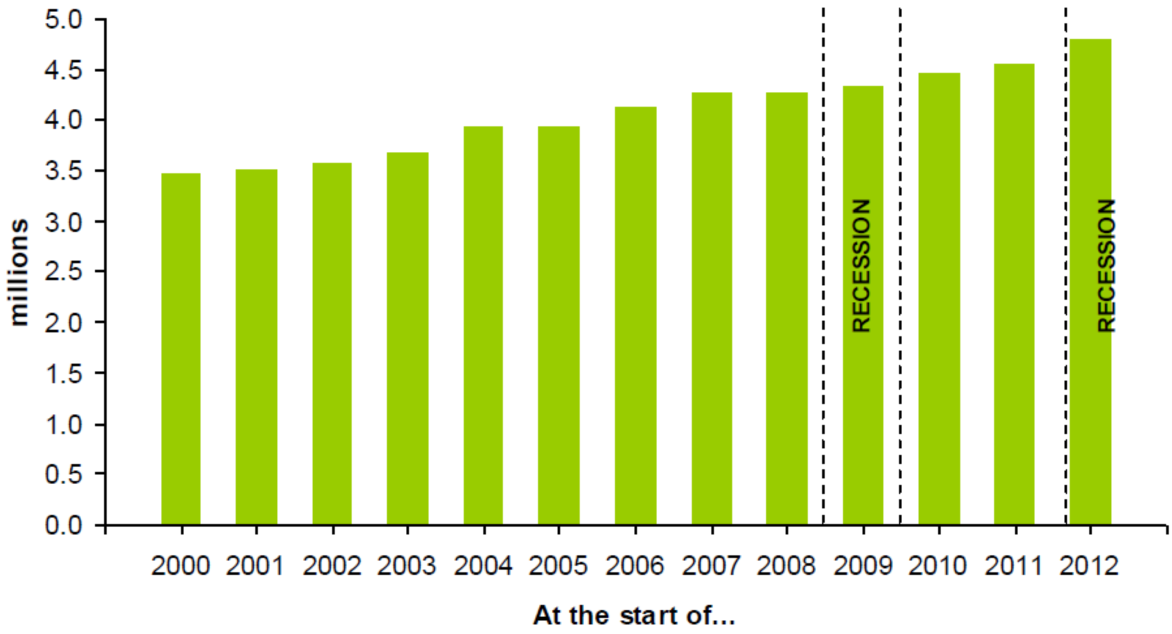
Figure 1.1: Share of businesses in the UK private sector and their associated employment and turnover, by the size of businesses, start of 2012.



Source: BIS (2012)

Changes witnessed in UK business population: There exist about 4.8 million registered private sector business units in the UK (as of the beginning of 2012), the highest number since the year 2000. As highlighted in Figure 1.2, despite unstable economic conditions, private business units have witnessed growth in all years from 2000 to 2012. In the recession period (2008-09) about 80,000 new firms were registered accounting for 1.9 percent growth (BIS 2012). By the beginning of 2012 there were 39 percent more firms in operation compared to the start of year 2000. This increase is primarily driven by increasing SME numbers. The majority of this increase consists of private firms having no employees (see Figure 1.3).

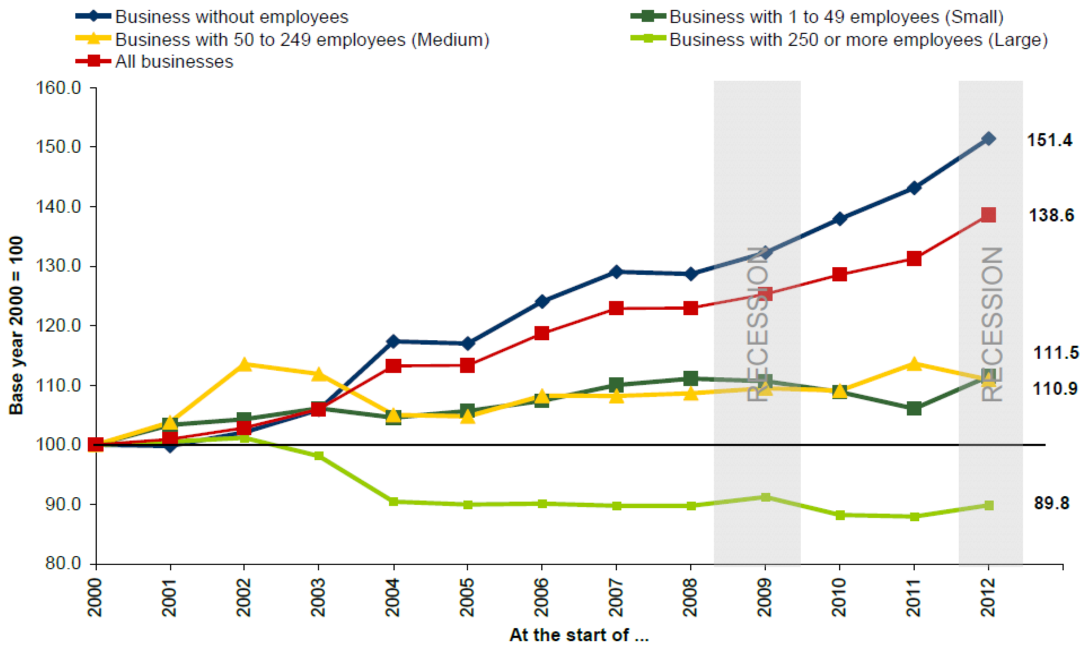
Figure 1.2: Estimated number of businesses in the UK private sector, start of 2000 - start of 2012



Source: BIS (2012

Compared to the beginning of year 2000, there has been a growth of about 51 per cent (about 1.2 million) in the number of firms with no employees by the start of the year 2012. Over this twelve year period, growth is positive, and reached its highest level in the year 2012. Simultaneously, there was a decline in firms employing more than 250 employees, with a net decrease of 10.2 per cent compared to the base year 2000. The medium and small firms also witnessed a positive growth of about 11 per cent, but clearly the pace of growth was dominated by the firms having no employees.

Figure 1.3: Change in the number of UK private sector businesses by size band, 2000-2012 (indexed)

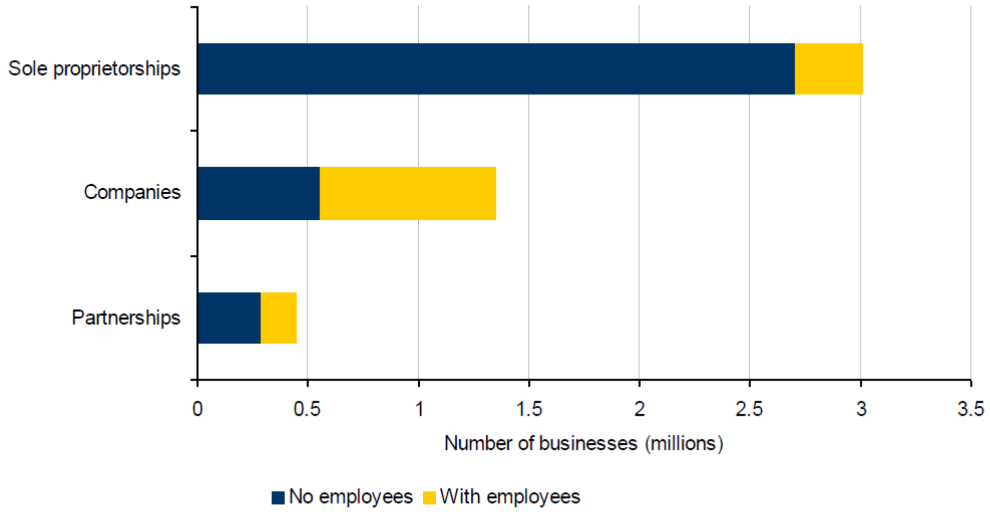


Source: BIS (2012

This diversity across the size category, provides a strong motivation to believe that firms characteristic might significantly vary among micro firms (having less than 10 employees) and the remainder of the SMEs cohort (small and medium firms).

Legal status of business units: Under UK law, firms can have one of three different legal identities: (i) sole proprietorship – business units run by single self-employed individual; (ii) partnership – business units being run by two or more self-employed individuals and (iii) limited liability companies (including both nationalized entities and public corporations) – this type of business units are owned by its shareholders and run by directors who are employees of the company.

Figure 1.4: Number of businesses in the UK private sector with and without employees, by legal status, start of 2012

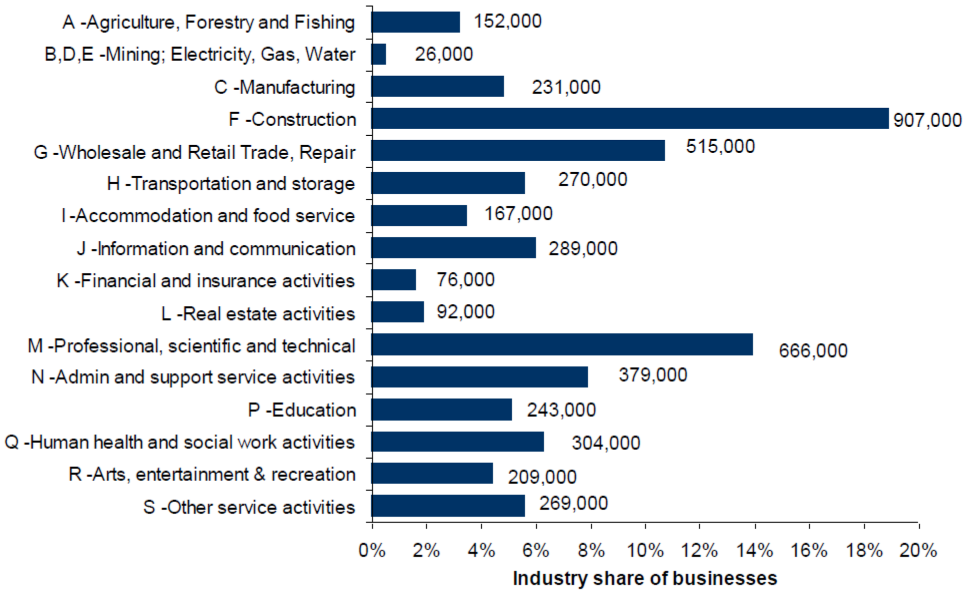


Source: BIS (2012)

At the beginning of year 2012 about 63% of private sector firms were sole proprietorship, whereas companies and partnership represented only 28% and 9% respectively. Out of the 3 million sole proprietorship firms operating at the beginning of 2012, approximately 10% had employees, company and partnership firms by contrast, significantly contribute to the generation of employment (see Figure 1.4).

Business units by industrial sector: As highlighted in Figure 1.5, the largest number of business units in the UK operates in the construction sector, followed by professional, scientific and technical services, and then wholesale and retail trade. These three sectors constitute about 44% of private firms and hence we can understand the significance of their contribution to the overall performance of UK private firms as a whole.

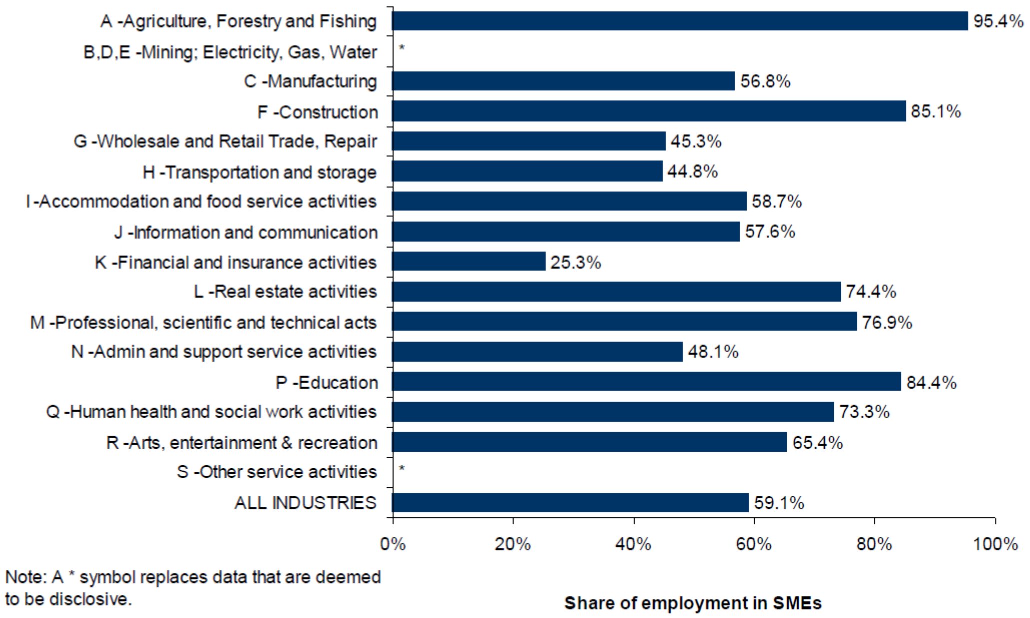
Figure 1.5: Share of businesses in the UK private sector (and numbers) by industry, start of 2012



Source: BIS (2012)

At the beginning of 2012 about 60% of private sector employment was generated by the SMEs sector (firms having less than 250 employees), with the proportion varying by industrial sector. In agriculture, forestry and fishing sector for example, SMEs generate 95% of employment. Figure 1.6 reveals that, about 50% or more of employment is generated by SMEs in various industrial sectors, with a minimum in financial and insurance activities.

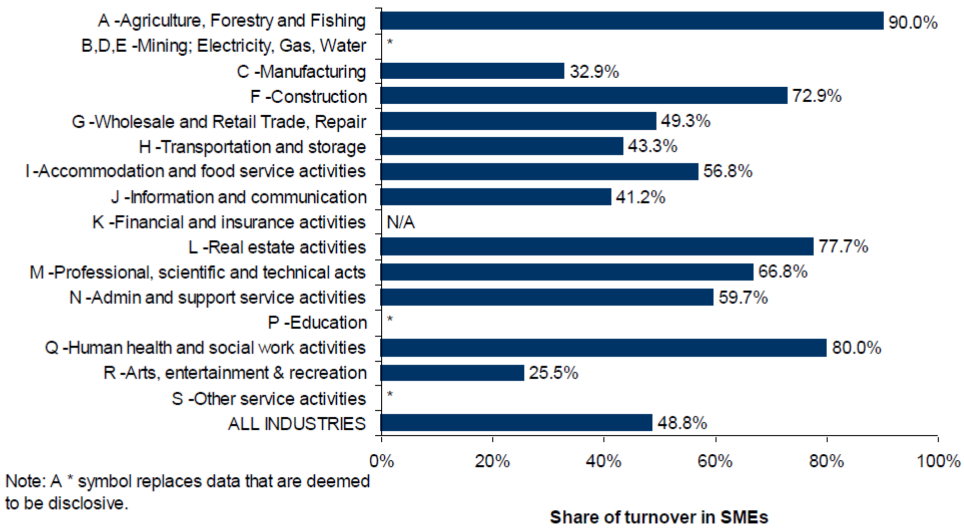
Figure 1.6: SME share of employment in the UK private sector, by industry, start of 2012



Source: BIS (2012)

Again, as highlighted in Figure 1.7 about 49% of total private sector turnover was generated by SMEs as of the beginning of 2012, varying by industrial sector.

Figure 1.7: SME share of turnover in the UK private sector, by industry, start of 2012



Source: BIS (2012)

Considering the relative importance of SMEs to the growth and development of the UK economy; policy makers, regulators and financial institutions are endeavoring to devise strategies to harness the tremendous potential of this sector, and are developing customized solutions to the problems of growth, finance, development and innovation, which they face.

1.3 Challenges to SMEs Performance

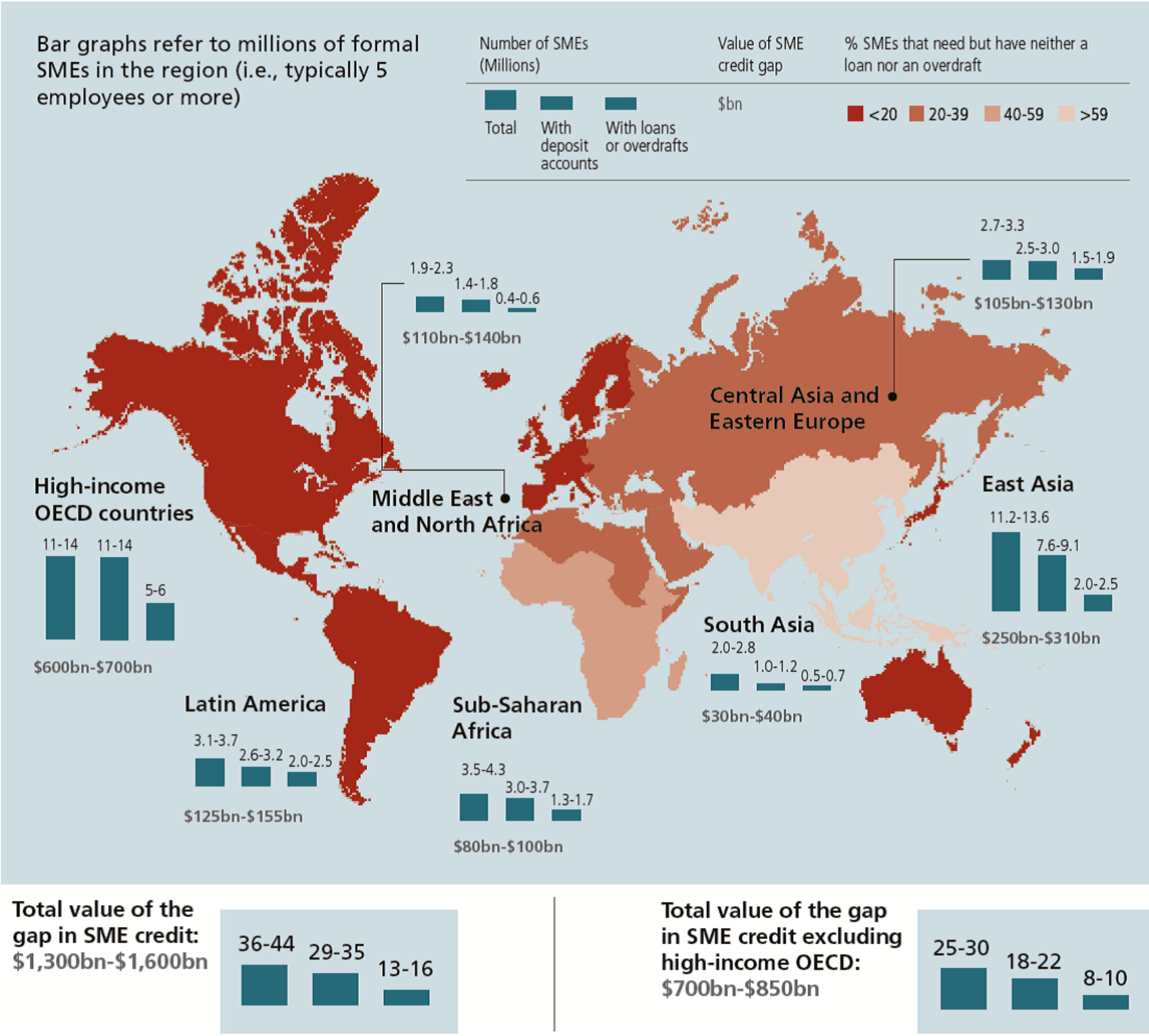
The positive impact of growing small business units on the economy's performance in terms of generating employment, wealth creation and innovation is a widely acknowledged fact. Traditionally, studies investigating drivers of firm's growth, have identified barriers to their performance, and have primarily focused on understanding the characteristics of growing firms (e.g. Orser et al. 2000) to predict their likelihood of growth (e.g. Carter et al. 2006). In a broad sense, barriers may be defined as those conditions or factors (internal or external) that constrain aspiring firms' growth potential (Storey 1994). Empirical literature primarily reports two approaches related to studying barriers related to SMEs growth and development. One approach considers comparison of macroeconomic development indicators as highlighted in Astrakhan and Chepurenko (2003). They investigate the prospect of small businesses growth and development in Russia by comparing its official statistics with Europe and report that Russia lags Europe on number of key indicators of SMEs development (e.g. sector's contribution to economy's employment, number of SMEs per 1000 population etc.). Such studies conclude that pace of development is slower in the presence of higher macro-level barriers. The other approach advocates that key stakeholders (owners, high level managers etc.) concerned with the growth and development of SMEs, directly communicate their views regarding potential threats that their business unit is experiencing as it develops. The usual technique is to provide the participants with a list of potential barriers (e.g. access to finance, government policies, unskilled workforce etc.) and to have their view on how much each of the barriers affects their business performance. From the early to mid-1990 period, wide-scale surveys conducted by World Bank and OECD identified various financial, fiscal, institutional and regulatory factors, negatively affecting SMEs development (Hashi 2001). Since then researchers have continued attempting to identify potential factors which are harmful to firms' growth and development (e.g. Beck et al. 2005; Beck et al. 2006; Hutchinson and Xavier 2006). Zehir et al. (2006) report that SMEs are adversely affected by their weak marketing, management, and information technology skills, which acts as barriers to their development.

Finally, from empirical findings obtained from different parts of the world we summarize that lack of access to finance, poor management skills, unfavorable regulatory environment, market challenges, corruption, lack of training and poor infrastructure are the most prominent hindrances to SMEs growth and development (see among others Zehir et al. 2006; Okpara and Wynn 2007; Robson and Obeng 2008; Hassanien and Adly 2008; Alam 2011).

1.4 Banking the SMEs Market

In both developed and emerging economies, fostering a vibrant SMEs sector is now being seen as a priority as countries seek to promote economic development. However, SMEs experience numerous barriers (as highlighted in section 1.3) to development. Arguably, lack of access to finance is the most serious dampener to the development of SMEs. This have traditionally been underserved in terms of basic financial requirements by banks, which perceive them as costly and risky to serve (IFC 2010). Despite the recognized importance of the SMEs sector, the supply of financial services, which is critical to their growth, is not optimal (see Figure 1.8 for a brief understanding). SMEs are in genuine need of banking services as they lack cash-flows required to make large investments. Unlike large firms, the capital market is inaccessible to them, and often they lack qualified talent to manage complex financial functions (Beck *et al.* 2005). They can use the long-term debt provided by banks to make capital investment without loss of ownership, while short-term financial lending may help them to attain higher operating efficiency and gradual incremental growth. At present, the SMEs banking sector is in a transitional phase. What was once viewed as complex and difficult to serve has now become a strategic business opportunity for the global banking industry. The gap that exists between large corporations and small business units in terms of availability of financial services is shrinking at a rapid pace, as banks develop customized products to overcome the challenges that they traditionally faced in serving the SMEs sector. By using credit analytics tools such as credit scoring, banks are effectively predicting credit risk associated with SMEs clients, without requiring a complete set of reliable financial information. They are also adapting information technology (IT) and management information systems (MIS), and thus building capacity to manage and process SMEs information, and thus better understanding the risk-return tradeoff. Although there is no unique formula for successfully banking the SMEs sector but as per IFC (2010), banks and financial institutions willing to exploit the SMEs potential may needs lessons in the following five key operating dimensions: (i) developing strategy to harness SMEs potential; (2) use of market segmentation for effective development of product and services; (3) reviving sales culture and use of appropriate delivery channels; (4) credit risk management and (5) adapting to IT and MIS.

Figure 1.8: Overview of formal SMEs' access to finance by regions



Source: IFC and McKinsey & Company (2010)

1.5 SMEs Failure

SME bankruptcies have always been difficult to track and measure, as failed businesses are often difficult to locate and if located it's again difficult to determine the reason for their failure. Despite this, recent literature (see among others Headd 2003; Carter and Auken 2006) has focused on understanding the rate and cause of such failures. Carter and Auken (2006) report that the principal reasons for firm failure can be categorized into, lack of knowledge, constraints to debt financing, and the economic climate. Besides the direct costs, the bankruptcy of small firms also causes indirect cost such as loss of personal collateral, self-esteem, self-employment etc. to the owners. A growing body of empirical literature suggests that financial constraints are the strongest reason for small business failures (see Hutchinson and Xavier 2006). Some recent studies also highlight poor management skills as a potential factor for small firm's failure (Peacock 2000).

Knott and Posen (2005) argue that, though the failure of new firms is considered to be wasteful, it enhances social welfare and reduce industry cost. However, all business failures are not purely due to financial difficulties. Empirical studies (see among others J. Watson and Everett 1996; Headd 2003; Bates 2005) suggest many "business failures" involve planned exit strategies, with the business actually being healthy enough to continue operation. The decision to discontinue business operations may be due to several reasons such as, change of ownership, opportunity cost, limiting losses, non-economic considerations, switching cost, etc. Sometimes the decision is made to close a successful business, thus a careful distinction needs to be drawn between failures due to purely financial difficulties, and firm's closure due to some strategic gain. To improve the quality of our analysis we take into account only those small firms where business failure is purely due to financial difficulties[1], and we exclude other form of business closure.

Considering the discussion presented above in this chapter, I understand that SMEs play a critical role in the growth, development and stability of an economy. Hence, I decided to undertake series of empirical studies (presented in this thesis) that might help us in better understanding of SMEs credit risk behaviour. This might eventually lead to better access to finance for SMEs.

[1] Once a firm has become insolvent, the UK Act provides to choose one from the five courses of action: administration, company voluntary arrangement (CVA), receivership, liquidation and dissolution. In this study to represent the failed sample group we take under consideration only those SMEs whose failure followed any of the three common routes, i.e. administration, receivership or liquidation.

2. Enterprise Risk Assessment

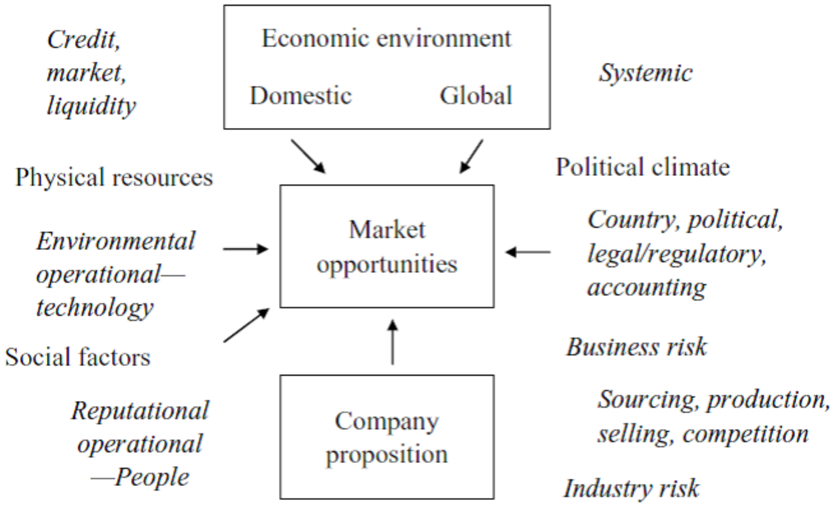
In common parlance, uncertainty related to any future outcome is called risk. Traditionally, for about 400 years insurance companies have been the only companies which were actively involved into the business of risk management, mostly in form of assets and life insurance. But over the past century, this horizon has broadened to the assessment and management of market and credit risk at both individual and enterprise levels. Any business venture is accompanied by risk, which may originate from different sources that essentially requires different kinds of data and modeling techniques to assess and manage it. It also demands various types of tools to control it and even generate profit out of it by applying suitable strategy. The following discussion is primarily focused on credit risk, which is one of the many potential risks that a business enterprise may encounter. However, I do provide a broad overview of the enterprise risk framework to enable a comprehensive understanding of the subject matter.

2.1 The Theory of Risk

Olsson (2002) describes the various types of risk that a business enterprise faces. Figure 2.1 highlights the various kinds of risk that a firm faces from different sources. However, operational, business, market and credit are the four types of risks that are most commonly encountered while running an enterprise. *Operational risk* may arise due to any event that has adverse material impact upon the operating activity of the firm. In the worst case, such risks may lead to the failure of internal processes of the firm. These risks may originate from inefficiencies related to staff, infrastructure, internal policies, technology, fraud, security etc.

The failure of Barings Bank and Orange County due to insufficient oversight of staff members are appropriate examples of consequences that may arise due to failure of operational risk management.

Figure 2.1: Source of Enterprise Risk



Source: Anderson (2007)

Business risk may arise due to inappropriate strategies, misunderstanding of overall economic climate, unfavourable government regulations etc., which may lead to lower than expected profit margins. The factors which are beyond the control of a firm like unanticipated volatility in market prices, foreign exchange rate, commodity prices, interest rate, real estate prices, terror attacks etc. may lead to potential losses to investors and are widely acknowledged as factors that contribute to *market risk*. Finally, *credit risk* arises whenever any uncertainty is involved pertaining to the fulfilment of financial commitments (in part or fully) by counterparties within an agreed time period. Beside this, it also covers changes in risk grades that may adversely affect the market value of the debt instrument and may lead to additional debt recovery cost.

In addition to the four major types of risks discussed above, there also exist other kinds of risks that may impact a firm's performance in a significant way. Being part of a particular industrial group a firm may face *industry risk* due to factors that affect all the firms within that industrial group. The risk may arise due to market concentration, trade barrier, technological changes, product cycle, economic cycle etc. and affect all related stakeholders, including investors, government, employees, creditors etc. Non-compliance with government regulations leads to *legal/regulatory risk*; while any adverse impact on the natural environment that is due to a firm's economic activity culminates to *environmental risk*. In normal business dealings, *counterparty risk* is very common that arises due to the uncertainty related to the fulfilment of credit transaction by the respective counterparties involved. As a consequence, persistent and high level of counterparty risk may lead to *liquidity risk*, in which it becomes difficult for a firm to honour its short term financial obligations. As multinational firms are exposed to multiple businesses, geographical and political environments, they face some additional types of risks like *exchange rate risk, political risk, country risk, transfer risk* etc. Further, at process level we can also differentiate between the risk that may arise due to factors within the system called *endogenous risk* and factors outside the system known as *exogenous risk*.

2.2 Credit Risk

In simple words we may view credit risk as the chance that our debtors may not honour their financial commitments in part or full. It is considered to be the one of the oldest form of risk seen in financial and business markets with growing complexity and importance due to the rapid development and integration of financial markets. Broadly speaking, any business unit or individual which offers goods or services on credit or is involved into the business of lending money to other business entities or individuals is exposed to credit risk, as there always exist uncertainty related to the repayment of debt within the mutually agreed time period.

Virtually every type of credit transactions are exposed to credit risk but the understanding and management of credit risk is critically important in the context of functioning and survival of financial institutions. Traditionally, banks have been pooling savings of the economy to provide it to the business houses which were considered to be the growth engines of the economy and hence they have been the ones being primarily exposed to the threat called credit risk. However, with the development of financial markets we see the intervention from new providers of funds, who grew gradually and emerged to be significant players of this modern financial market. Because of the associated economic benefits, modern financial instruments like *bonds*, *commercial papers*, *asset based lending* etc. became preferred financing options. Although banks are witnessing decreasing market share in terms of capital lending to large enterprises but they are still dominant players in certain segments like mortgage market, small business lending, micro finance, consumer finance etc. Thus, modern banks are significantly exposed to challenging business environment with greater challenges of understanding and managing credit risk of varied and complex financial instruments. Considering their diverse lending activity with added business complexities, it won't be an exaggeration to say that they are the one who are most prone to default due to inefficient credit risk management.

In developed financial markets like United States and United Kingdom, institutional lenders can diversify their risk by selling appropriate amount of credit assets out of their portfolio to interested counterparties. However, the market for such assets is niche, illiquid and at an evolving stage. Thus development of better credit information system seems to be an attractive alternative. In recent times we have seen lending institutions investing heavily in processes related to collection of credit information and analysis of the same across various lending sectors. Changing regulatory environment primarily imposed by Basel I, Basel II and now Basel III have made them develop innovative techniques of managing risk. Major financial turmoil like real estate crash of 1989, Asian crisis of 1997, the financial slowdown of 2007 and more recently the Euro-zone crisis have challenged the understanding of regulators and lending institutions. They have been forced to revisit their credit risk analysis techniques and re-work toward the goal of developing sound risk management tools and policies. In response to that, several innovative financial products and policy guidelines have been developed to manage such crisis efficiently, but at this stage it would be inappropriate to comment on their true contribution. Significant innovation have been witnessed in the area of structured finance transaction like asset-backed securities and collateralized mortgage obligations in which the risk originator is authorized to transfer full or part of its risk by pooling the credit assets and selling them to interested counterparties. Clearing houses and exchanges play a significant role in restoring strong confidence within the market participants with particular emphasis on counterparty credit risk. Recent time have also seen the development of a relatively young and complex mechanism called *credit derivatives* which is now being actively used to manage credit risk by the market participants. Using this kind of tools, they create some kind of insurance mechanism by entering into appropriate derivative contract, thus the lender can mitigate its credit risk exposure without selling its credit assets.

2.3 Classic Credit Analysis

Among other functions of bank, taking and management of credit risk is considered to be its fundamental function and banks were the primary provider of funds to support the industrial engines of growth and development. However, in this modern economic setting the roles of banks have undergone changes with diminishing primacy in certain key areas. In particular they have lost significant market share in lending business to capital market and other non-banking financial institutions. Despite these challenges, banks remain dominant providers of liquidity to large corporate house and other form of financial institutions. They are also dominant participants in areas where access to capital market is difficult like small business financing, project financing etc. Also banks play a significant role in resolution of critical credit problems when firms file for bankruptcy proceedings or when firms are in financial difficulties. These are the areas where banks have developed their competitive edge over many years and have organized their long rich experience into a formal lending process known as *classic credit analysis*. Next we discuss some of the key aspects of classic credit analysis which if used properly may reveal detailed insights, but such analysis is not cost effective and may be myopic.

2.3.1 Analysis of Credit: An Expert System

Classic credit analysis can be viewed as an expert system that relies heavily on the subjective judgement of trained credit lending professionals. The credit officer or the relationship managers are generally empowered to the credit decisions. Individuals eventually turn out to be an expert over his career span; as a consequence they keep on gaining additional authority as they acquire experience and demonstrate established skills. Lending decisions are based on the common sense and experienced judgements of the lending officers, they put weights on some key factors which they believe are significant in making credit decisions. Technically there may be infinite number of such potential factors but the following “5 Cs” of credit (Anderson 2007) gives fair understanding:

i. *Character* – it accounts for the credit history and reputation of a firm or an individual, it's been empirically verified that older firms tends to have better repayment reputation than younger firms.
ii. *Capital* – this information helps the credit officer to assess the optimum leverage limit of the borrower to minimize the default probability. The higher is the proportion of debt, the higher is the probability that the borrower would default on its financial obligations.
iii. *Capacity* – it accounts for volatility of earnings and cash-flow, higher the volatility the higher is the default propensity.
iv. *Collateral* – it provides a kind of insurance mechanism to the lender and is a significant factor which determines the debt size. The higher the value of collateral the higher is the debt amount that a borrower is eligible to have but in event of default the collateral may be sold to recover the debt.
v. *Cycle Conditions* – this factor becomes critically important if the borrowers' cash-flow and earnings are sensitive to economic cycles.

The senior lending officers are in charge of industry analysis, customer and product portfolio analysis, screening of the underwriting process etc. They also play a critical role in determination of loan size, pricing of the debt and determination of credit terms and conditions.

2.3.2 Classic Credit Analysis and Modern Banking Practices

The founding principles of classic credit analysis approach still remains as part of the standard banking practices, but it is more actively employed to small business lending. For large corporate lending, things have undergone wide irrevocable structural changes. These forces of change have been primarily driven by *effectiveness* and *economic challenges*

(Caouette *et al.* 2008). Banks have been severely affected by the disintermediation due to the emergence of other form of raising funds in large corporate lending sector. Non-banking lending institutions with their innovative analytical techniques, flexible and tailor-made terms and conditions, effective sales strategy and commoditized lending instruments have been successful in providing debt capital to modern corporate houses at much lower competitive prices than the banks. Challenged by price reductions and diminishing margins, banks started downsizing their underwriting activities followed by cost reduction to match up with the reducing revenue. As a result of this process, the quality of bankers has deteriorated along with the decline in the number of trainings and trainees. The second issue with the classic approach is subjectivity in its effectiveness. Irrespective of the expertise level and expertise range of lending officers, some misjudgements will be there. In addition to the misjudgements, the problem gets multiplied due to cooked out-dated financial statements and poor customer quality, specifically witnessed in small business lending. Thus banks are left with higher uncertainty in terms of credit risk and often concentration risk.

In response to this, several banking institutions have come up with innovative business models to participate as a facilitator in the modern business setting. Apart from the traditional sources of generating revenues, now they are actively looking at other non-traditional activities within the banking domain which will help them to integrate better with the modern dynamic capital market and generate additional revenues. Now, they are actively involved in creation, packaging and trading of credit products, thus deciding effectively which risk to retain and which risk to hedge. In formal terms, this phenomenon of securitizations and structured finance got significance primarily due to the evolution of assets based lending which modern banks find is an effective way to compete in the modern capital market (Caouette *et al.* 2008). However, credit risk management once believed to be an art, is now being felt like science with the development of newer techniques of measuring and managing credit risk. But many academic scholars and industry practitioners prefers to call it a form of engineering to prevent financial distress or else to provide protection from such potential event.

2.4 Credit Risk Analysis in Modern Setting

In this modern economic setting it is quite difficult to draw a clear line demarcating the classical and modern approaches of credit analysis, as the modern approach highly draws from most of the better ideas of the classic credit analysis. However, it’s worth mentioning the following approaches of credit analysis:

2.4.1 Artificial Neural Networks

In this approach the system simulates the learning process of human beings through learning the relationship between inputs and outputs. The input/output information sets are repeatedly sampled to have a close approximation of the real phenomena. This method is primarily very useful when the input information is noisy or incomplete; it makes an “educated guess” similar to as a human expert system does. However, its major disadvantage is lack of transparency due to the hidden structures of the internal networks, which cannot be mimicked even if we use the same dataset.

2.4.2 Rating Systems

Credit Rating or Rating Systems involves estimation of creditworthiness of the creditor for the purpose of advancing credit. The creditor may be an individual, a private company, a public company, government bodies or even country. The rating system may be further classified into two broad categories namely *external rating systems* and *internal rating systems*. External ratings are provided by organizations which are external to the organization for which the rating is being provided. These are specialized agencies which comment on the creditworthiness of economic entities (generally companies and countries) through rigorous analysis of quantitative and qualitative factors. To name a few, S&P, Moodys, CRISIL, Fitch etc. are the agencies actively involved in the business of providing external credit rating to the potential users. On the other hand, internal ratings are done by the lending institutions (e.g. banks) themselves using methods quite similar employed by external rating agencies. Since this method is highly skill based, expensive and time-consuming, the lending institutions generally employ this for large corporate lending. However, the final lending decision is based upon their own internal assessment as well as the independent external credit ratings.

2.4.3 Credit Scoring Systems

This is the most widely used method of credit analysis virtually applied in all different types of credit analysis like consumer lending, mortgage, small business lending etc. In this method, the key factors which affect the default propensity are identified to build a quantitative model, which is then used to generate a quantitative score reflecting the creditworthiness of the borrower. The score may be viewed analogous to the default probability or as in most cases; this score is compared with a standard cut-off score to differentiate between the good and bad borrowers to make the lending decision. Multivariate discriminant analysis and logistic regression technique are the traditionally preferred methods of estimating the credit score. This approach of credit analysis is generally applied to high volume and low value lending products like car loan, housing loan, education loan, consumer loan etc.

2.5 Data Source for Credit Analysis

The starting point of credit analysis is the data source; the following exhibits comprehensive sources of data that is being used for enterprise risk assessment (Anderson 2007):

Human inputs: the potential borrowers are asked to provide subjective opinion about some question, which the lender thinks are critical in assessing their credit risk. However, most of the lending institutions try to make their assessment as objective as possible but in high values lending, subjective information about the workforce plan, expansion plan, business plan are often helpful in precise estimation of credit risk.

Market value of traded securities: this information is extensively used for assessing credit risk of publicly traded corporate houses. The daily closing stock price, daily price volatility, trading frequency, buy/sell spread are forward-looking information and reflects the view of market participants on firms' future prospects and creditworthiness. Similarly, information about debt instruments traded in the market are also used to assess credit risk of the issuer like government bodies, corporate enterprise, municipals, public utilities etc.

Financial Statements: a detailed review of borrowers past financial statements is very useful in understanding any trend in terms of income, assets value, cash flow etc. Income statement, balance sheet and cash flow statements are analysed to understand any consistent improvement or deterioration in the financial position of firms and thus to assess the creditworthiness of the firm.

Payment history – past payment history serves as a very useful proxy about the future *character* of the borrower in terms of repayment intentions. This information is particularly useful is assessing borrowers who apply for low value and high volume debt products. However, for large companies their payment behaviour toward their creditors, working capital loans, honouring of debt instruments at maturity provide valuable insights about the *character* of firms.

Environment assessment: identification and assessment of external factors like political environment, business cycles, labour market conditions, government policies, regional and industrial factors etc., which are external to the firm but may bear significant impact on the firm's performance is very helpful in identifying the risk that may arise due to such unfavourable changing circumstances. Proper assessment of these risk factors may lead to better credit pricing and avoidance of unwanted risk from the credit portfolio.

The various type of data mentioned above does not represent an exhaustive list of information that are used for credit risk analysis, in particular the data being used for analysis also depends upon the size of borrowing amount and the borrower. However, all information provided corresponds to one of the 5 Cs of credit (see Table 2.1). In case of small size borrowers and small borrowing amount, less time and effort is spent on credit analysis and vice versa. Indeed, it is a well-established fact that SMEs are the key drivers of the economic growth engines especially in light of the recent financial crisis and the kind of information that is used to assess the creditworthiness varies across the firm size. Table 2.2 provides a comprehensive overview of the various types of information that are used to assess the credit risk of firms having various assets sizes.

Table 2.1: Data versus the five Cs

Data Source	Capacity	Capital	Conditions	Character	Collateral
Human Inputs	✔	✔	✔	✔	✔
Traded Securities Prices	✔	✔	✔	✔	✔
Financial Statements	✔	✔			
Environment Assessments			✔		
Payment History				✔	

Source: Anderson (2007)

Table 2.2: Company size versus data

Company Size	Market Prices	Judgemental Assessment	Financial Statements	Payment History	Personal Assessment
Very Large	✔	✔	✔		
Large		✔	✔		
Middle		✔	✔	✔	
Small			✔	✔	✔
Very Small				✔	✔

Source: Anderson (2007)

2.6 Credit Analysis Tools

We have witnessed significant development in the financial market over the past two decades, which lead to the emergence of new types of risks, that have never been witnessed before. As a result several innovative techniques are developed to understand and manage these new risks. The development has been innovative and remarkable in the credit risk management domain. This development has primarily been led by forces of deregulation, securitization, evolution of new borrowing sectors, enhanced emphasis on cash flow based lending, development of OTC derivative market, introduction of inclusive regulatory reforms like Solvency II, Basel II and more recently Basel III.

Tools and techniques from statistics and operation research have tremendously advanced our knowledge and understanding pertaining to development and analysis of credit risk management in this modern financial setting. Tools like mathematical programming, survival analysis, deterministic and probabilistic simulation, neural networks, game theory and stochastic calculus have been remarkable. Recent advances in the understanding of financial markets like capital asset pricing model (CAPM) and its extensions, option pricing model, arbitrage pricing theory (APT), prospect theory etc. have also contributed significantly toward the development and understanding of advanced risk pricing tools. However, Falkenstein et al. (2000) highlights the following tool for the analysis of credit risk of business enterprises:

Rating agency grades: certified credit rating agencies like S&P, Moodys etc., undertake credit risk assessment of very large enterprises for a fee[2] using qualitative and quantitative information and provides letter grades to them. Different letter grades connote different level of risk, which are obtained by using appropriate statistical and qualitative analysis tools. E.g. the rating grades may be AAA, AA and BBB, which are stated in order of the creditworthiness. Here, AAA signifies highest level of creditworthiness and BBB the lowest.

Public-firm models: this kind of assessment is based on Merton (1974)'s concept of *economic default*, which states that the firm will default on its financial obligation when the market value of the firm is less than its current outstanding liabilities. Assuming that the markets are efficient and the current stock price reflects the true market value of the firm, the stock price measures and its return volatility measures could be combined with the outstanding liabilities to estimate the default probability of respective firms, which can be further used to estimate credit risk grades.

Private-firm models: these tools are primarily applied to small companies; their default probabilities are estimated based upon some qualitative and quantitative information and which are effectively converted into some sort of credit score which reflects the credit risk of firms.

Hazard Models: this model is applied to large firms that are rated by credit rating agencies and have liquid traded debt in the financial market. This method of credit assessment is very similar to Merton's model, other than bond spreads are analysed to estimate the credit risk instead of default rates.

[2] Sometimes they undertake voluntary assessment without any fees.

Portfolio Models: these models are used to model debt advanced as a group, using exposure and default estimates for individual asset class. This kind of models relies upon the correlation among various asset classes and estimate worst possible outcome at a given confidence interval.

Exposure Models: these models assume that the account has already defaulted and rather than the default propensity, magnitude of loss is estimated to assess the worst case scenario. The two widely used such measures are exposure-at-default (EAD) and loss-given-default (LGD).

Business Report Scores: these are provided by credit referencing or financial agencies like Dun & Bradstreet (D&B), Experian etc. They account for various qualitative and quantitative factors like court actions, county court judgements, age, size, financial status etc. to assess the trade credit.

Table 2.3: Models versus data

Model Type	Traded Securities	Financial Statements	Environment Assessments	Payment History	Judgemental Assessment
Rating grades	✔	✔	✔		✔
Public Firms	✔				
Private Firms		✔	✔		
Hazard	✔				
Exposure		✔		✔	
Portfolio	✔			✔	
Credit Bureau				✔	

Source: Anderson (2007)

Table 2.3 summarizes an imperfect relationship between the various credit analysis models that are being used and the data sources. Credit scores are not mentioned here but it can be viewed analogous to private firms and credit bureau.

2.7 SME LENDING

The lending community globally, shares the consensus that the SME sector is a profitable market segment, and are developing ways to unlock its potential with specific focus on the problem of high credit risk and service cost (IFC 2010). SMEs are often referred as job creation engines; a 2009 Economist Intelligence Unit (EIU) study reveals that they continued to generate job opportunities throughout the economic slowdown, and along with entrepreneurs they are viewed as key drivers of economic development (Bosma and Levie 2010). Despite of the economic significance of SMEs across different countries, the manner in which the banks maintain relationship with SMEs does vary across different countries.

Allen et al. (2004) argue that, over the past several years modern banks are moving away from traditional *relationship based lending* (old-fashioned method of credit analysis based customers relationship history with the bank and few other key information) towards *transactional lending* (it is relatively advanced method and may be viewed analogous to credit scoring method of credit analysis) in small business lending space. In effect, the method of credit analysis have moved to the other end of the spectrum which has accelerated the growth of banks, as credit analysis cost and time has come down significantly. However, there still exists niche market in developed economies that finds relationship based lending more effective and large market for the same is also favored in transition or developing economies.

2.7.1 RELATIONSHIP LENDING

This is an old-fashioned credit assessment method primarily based on the 5Cs of credit. The lending officer or bank manager is generally in charge of risk assessment of the borrower. The assessment is based on assessor’s personal knowledge about the applicant, his/her past reputation, networks, position in the social community, trade history with other organizations and so on. Berger and Udell (2002) report that, due to the opaque structure of small companies, they attract fewer financing opportunities than their larger counterparts (both in areas of trade credit and institutional credit). The borrowers having long and strong relationship with banks benefits from easy access to credit, lower interest rates, lower requirement of collaterals, flexible terms etc. However, relationship lending might suffer from the following disadvantages: i) sub-optimal assessment of credit risk, ii) discrimination among large and small borrowers and iii) cross-subsidization[3] among borrowers. The borrowers having favorable length and strength of relationship with the banks are rewarded with lower credit pricing, particularly in competitive markets. But, the amount of time and effort required to develop such relationship is not cost effective for banks other than niche banks with limited growth potential. In contrast, most large lending institutions eyes for larger market share with lower debt servicing cost. Thus in long run, relationship based lending may not be cost efficient in managing growing loan books and optimum utilization of capital.

2.7.2 Transactional Lending

The most important difference between relationships based lending and transactional lending lies in the kind of data that is being used for credit analysis. Transactional lending is primarily based on 'hard' (quantitative) information, while relationship lending primarily employs 'soft' (qualitative and subjective) information (Berger and Udell 2002). Rather than relying on the judgment of lending officer, transactional lending utilizes other technologies, specially the method of credit scoring and credit rating. However, in certain cases both the methods may complement each other, particularly for large size lending.

The problem related to SMEs lending is primarily two-fold. First, there exist significant correlation between firm size and transparency. Smaller firms tend to be more opaque than

[3] In this practice, higher price is charged to one group of borrowers to subsidize the price for the other group.

their larger counterparts. The quality and authenticity of the financial statements are often challenged, and the collateral provided as part of security for loan is often illiquid with no market value. Hence, banks primarily rely on trustworthy qualitative information with particular focus on unsecured lending. Second, banks often complain of higher service cost for the SMEs segment, which eventually makes lending cost inefficient and expensive. In this situation, transactional lending seems to be some sort of respite for large lenders who use credit scoring technique to make their lending decisions. It seems to be an appropriate strategy for large lenders as they primarily focus on diversification obtained by a large portfolio of small size loans, rather than paying attention on individual loan (Allen *et al.* 2004).

The impact of credit scoring techniques on the SMEs lending segment has been significant. The ready availability of "hard" information has made the credit assessment process faster with reduced cost and increased volume. This has resulted in greater economies of scale with enhanced geographic diversification but has also lead to increased competition. In this setting the borrowers having better credit scores are rewarded with lower borrowing cost. Risk-based pricing is also a growing phenomenon in this segment; borrowers that are declined on one credit product are often accepted on another credit product at higher price and stringent terms and conditions.

Although, credit scoring is an efficient method of credit assessment for institutional lenders, but there is still some reluctance to move from relationship based lending to transactional lending. The reluctance is more dominant in developing countries like India, China, Brazil etc., where the quality of "hard" information often face some serious challenges.

2.8 Credit Rating Agencies

Credit rating agencies are business organizations which specialize in assessing the creditworthiness of publicly traded debt securities for investors. The securities are primarily issued by corporates, municipalities, government, financial institutions, structured finance etc. Loosely speaking, rating agencies provide an estimate about the likelihood that an investor would receive the interest and principal repayment within the agreed time period for a given debt security. In order to avoid any bias that may arise in the assessment process, the rating agencies ensure that the individuals in charge of carrying on the assessment do not have any stake or contact in the economic entity to be assessed, but instead they collect information from other external sources.

In developed markets like United States and United Kingdom the capital market has emerged as a primary source of debt capital for corporate houses, thus replacing the tradition banking channels and promoting the importance of rating agencies in credit risk management. The three most influential rating agencies in the world are Fitch, Moody’s and S&P. However, the rating agencies do not provide any recommendation related to investment decisions, they express only informed opinion about the creditworthiness of a given debt security. Yet, it has gained widespread acceptance in the modern capital market and investors shows confidence in the convenient, cost effective and accurate information provided by them.

2.8.1 Rating Grades

As mentioned earlier, credit rating agencies provide their assessment of the borrower’s credit risk in form of letter grades which reflect the level of risk associated with the borrower. Each agency uses a method of alphanumeric letter grades that locate an issue or issuer on a scale of creditworthiness from the highest (AAA/Aaa implying extremely low chances of default on its financial obligations) to lowest (C/D implying the borrower has default on its financial obligations). See Table 2.4 for a brief understanding of how the creditworthiness varies across the letter grades. As highlighted in the table below, each letter grade has three notches, the lower the grade the higher is the default risk. The table further differentiates between investment grade, i.e. BBB/Baa or above it and non-investment grade, i.e. below BBB/Baa. Since the risk of default is higher for non-investment grade securities, the expected returns associated with such investments are also higher than the investment grade securities. The rating agencies use different methods of grading for short-term debt securities but the same analysis principles apply for them too.

Table 2.4: Long-Term Senior Debt Rating Symbols

Investment Grade Ratings	
Rating	**Interpretation**
AAA/Aaa	Highest quality; extremely strong, highly unlikely to be affected by foreseeable events.
AA/Aa	Very high quality; capacity for repayment is not significantly vulnerable to forseeable events.
A/A	Strong payment capacity; more likely to be affected by changes in economic circumstances.
BBB/Baa	Adequate payment capacity; a negative change in environment may affect capacity for repayment.
Below Investment Grade Ratings	
Rating	**Interpretation**
BB/Ba	Considered speculative with possibility of developing credit risks.
B/B	Considered very speculative with significant credit risk.
CCC/Caa	Considered highly speculative with substantial credit risk.
CC/Ca	Maybe in default or wildly speculative.
C/C/D	In bankruptcy or default.

Source: Caouette et al. (2008)

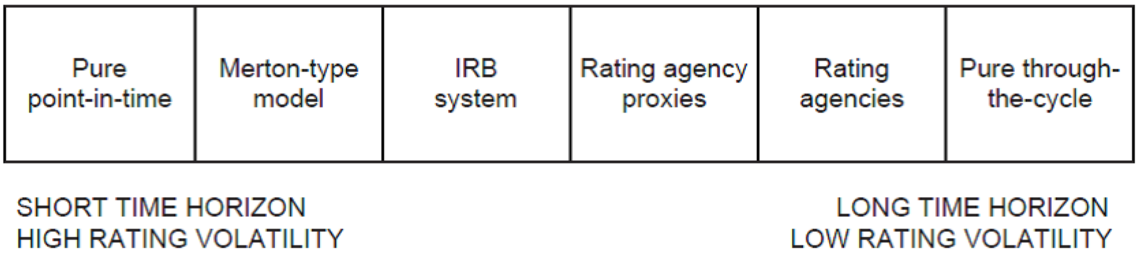
Since the credit quality of the borrower may change with the changing business dynamics over time, the ratings provided by the credit rating agencies are accompanied by *credit outlooks*, which are subjected to revision on a continuous basis to foster transparency. If the credit outlook is *positive*, it implies the credit rating may be raised, *negative* implies rating may be lowered, *developing/evolving* implies rating correction may take place in either direction while a *stable* outlook implies no changes in near future. To revise the ratings on a continuous basis, the rating agencies maintain contact with the issuer of the debt securities and continuously monitor their financial information or any other internal or external economic development that may have an impact on the creditworthiness of the firms.

Despite of their effort to provide transparent and unbiased information to the financial market, switching in payment terms from users or subscribers to debt issuers have raised concerns about the independent functioning of the agencies. However, most of the market participants strongly believe that rating agencies functions more like an academic research unit than corporate business houses and to certify their transparency, market regulators in various countries have introduced functional guidelines for them. For example, *International Organization of Securities Commissions* (IOSCO) introduced a *Code of Conduct Fundamentals for Rating Agencies* in December 2005 to undertake sound and transparent practices. Agencies also undertake *unsolicited* (these are agency-initiated and they don't receive any payment for it) ratings primarily of high profile corporate houses, sovereign debts instruments etc.

2.8.2 The Rating Process and Performance

The rating agencies use many of the tools used by equity analyst to assess debt issue or the issuer itself. But unlike equity analyst, they focus on long-term determinants depending upon the maturity of the debt instrument and the nature of the issuer. As illustrated in Figure 2.2, there are various approaches to credit rating, ranging from very sensitive short-term indicator to long-term agency ratings. We do not see uniform transparency in their detailed rating processes across various agencies; however the processes are broadly similar for all rating agencies. For e.g., to rate an industrial bond the agencies may look at business risk, competitive positioning, industry characteristics, management, financial characteristics, financial risk, financial policies, financial flexibility, profitability, cash flow protection, capitalization etc. (Caouette *et al.* 2008).

Figure 2.2: Point-in-time versus through-the-cycle credit assessments

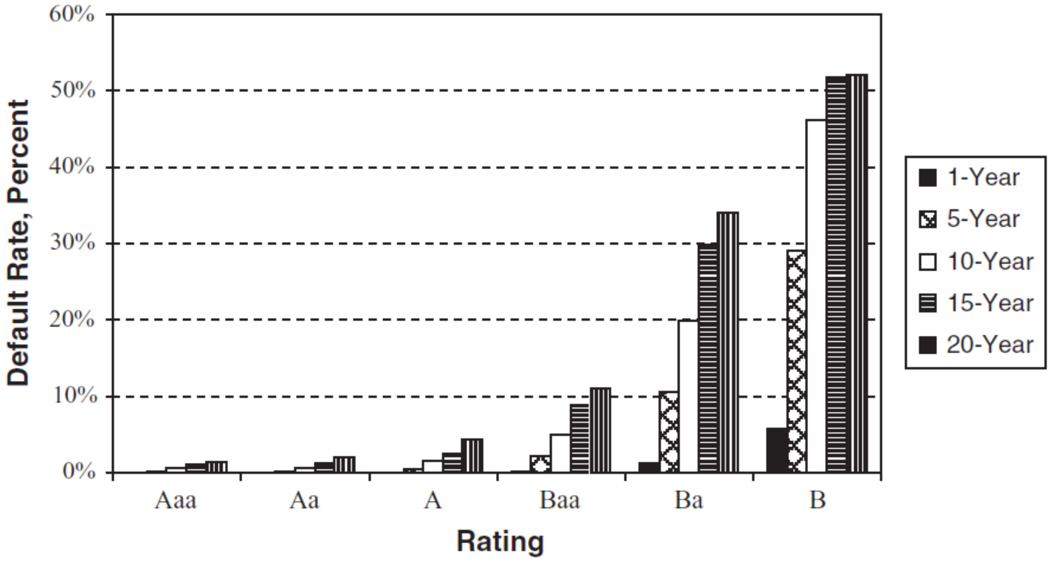


Source: Gonzalez et al. (2004)

All of these characteristics do not receive equal weight in the rating decision. As per S&P – they assign highest weight to industry factors within which the firm operates; in particular they regard strength and stability to the industry as key determinants. To analyze the financial strength, risk, profitability a number of financial ratios are computed covering dimension like leverage, profitability, assets utilization, coverage etc. They also focus on qualitative information like management style, growth plans, industry outlook, regulatory environment etc. during the rating process.

Credit ratings have been effectively used to communicate opinion related to the creditworthiness of debt issues or issuers, which is a condensed version of many small pieces of information, including information about default propensity and severity. As highlighted in Figure 2.3, the rating agencies seem to have pretty good track record. Clear inverse relationship exists between the quality of rating and the default intensity over both short and long time horizon.

Figure 2.3: Corporate Default Rates Based on 1970–2005 Experience



Source: Caouette et al. (2008)

2.8.3 Application Issues

The credit ratings provided by the rating agencies cannot be taken for granted, it does suffer from several imperfections. Several problems have been cited by Schönbucher (2003) and others, especially in light of using rating grades while pricing of securities. Some of the inefficiencies documented by Anderson (2007) are as follows:

i. Small numbers – due to the high cost associated with obtaining such specialized service, only small number of obligors go for it. This in turn makes comprehensive analysis using tools like survival models and transition matrices inefficient and biased (see Schuermann and Jafry (2003) for further details).

ii. Ratings delay and momentum – companies like Enron and WorldCom went bankrupt without any early warnings; such incidence makes it clear that the current ratings did not reflect the latest developments affecting the firms' creditworthiness. This is referred to as rating delay; as a result changes in the market prices may lag by several months (see Schönbucher (2003) for an empirical discussion). It is often seen that the current direction of rating changes follows the immediate previous direction; this phenomenon is called ratings momentum.

iii. Population drift – it refers to the changing nature of the underlying population which is being assessed (see Schuermann and Jafry (2003) for an empirical discussion). It's an established fact the number of firms being assessed has increased dramatically over time and major chunk of the growth is due to the small size firms, which are risky and often complex to understand.

iv. Downward ratings drift – it is observed that the likelihood of downward movement of risk grades is higher than the likelihood of upward movement. In addition, the averaging rating quality of issues is deteriorating over time (Carty and Fons 1994).

v. Business cycle sensitive – the assessors assume that the ratings transitions are not affected by the business cycles, but empirical literature does highlight their varying nature over business cycles (Schuermann and Jafry 2003).

vi. Risk heterogeneity – it is assumed that the credit rating and the credit spread would be constant within a particular risk grade. However, in reality the profiles may be quite different (see Schönbucher (2003) for details).

Although we provide an overview of the philosophy and methods involved in enterprise risk assessment, but practitioners face may obstacles in generation of credit assessments that are close to the reality, primarily because defaults are extremely low frequency events and getting data covering all the dimensions of risks is quite difficult, at least for the academic researchers. However, modest sized bankers manage to get large pool of data from their internal resources to generate more statistically robust measure of enterprise credit risk. Cases become more challenging for practitioners especially in emerging markets, where the dynamic nature of the economy, high level of cooked accounting and business information, coupled with poor quality of data makes the risk assessments biased and unstable.

3. THE VALUE OF OPERATING CASH FLOW IN MODELLING CREDIT RISK FOR SMES

3.1 INTRODUCTION

Traditionally, the banking system has perceived SME (small and medium enterprise) market segment as risky and costly in comparison to large enterprises. This may be primarily due to its inherent complexities at the socio-economic front, but rather than trying to understand the complexities of this market they preferred to overlook it and chose large enterprises as the primary profit serving sector. However the recent financial crisis has proved that no matter how large a firm is, efficient risk management is one of the key components of its long run survival. Hence, a thorough understanding of the market complexities and application of efficient risk management tools and techniques can make SMEs a significant profit serving sector for banks. Mounting evidence (see *The SME Banking Knowledge Guide*, 2010 for further details) suggest that by using tools like market assessment and operational diagnostics, banks are finding solutions to the problems of determining credit risk backed by lower operational cost, and have become more competent in serving the SME sector profitably. From the credit risk perspective, SMEs differ significantly from the large enterprises. Although considered to be riskier, but they exhibit lower asset correlation unlike large firms (Dietsch and Petey, 2004).

Conventionally, the literature of credit risk management gravitates toward large firms and use of financial ratios in predicting their bankruptcy. Seminal contribution came from Beaver (1966) and Altman (1968) who use univariate and multivariate models respectively, to predict business failures using a set of financial ratios. Since Altman's study, and with the advancement in technology and methodology, the number and complexity of bankruptcy prediction models has increased significantly (see among others Anderson, 2007; Servigny and Renault, 2004 for an overview). However, studies addressed to understanding credit risk behavior of small companies are relatively scarce. This may be primarily due to non-availability of required information, as they enjoy concession pertaining to the amount of financial data they need to report to the regulatory authorities. I believe Edmister (1972) to be the first to explore the credit risk behaviour of SMEs using a sample over the period of 1954 to 1969. Using multivariate discriminant analysis technique, he analyzed 19 financial ratios and developed a model to predict default of small business units. Recently, using logit regression technique, Altman and Sabato (2007) developed a distress prediction model for US SMEs by employing a panel of over 2000 firms over a period of 1994 – 2002. However, they acknowledge that their model's performance can be improved by inclusion of qualitative variables. Recent literature also highlights the importance of qualitative variables, like firms' age, type of business, industrial sector, location etc. (see Lehmann, 2003; Grunert et al., 2005) in the study of financial distress. Altman et al. (2010) take cognizance of this issue and employ non-financial information pertaining to firms' characteristic and operational risk as additional regressors in their distress prediction model. As a result, they report about 13% improvement in their model's prediction performance in comparison to Altman and Sabato (2007). Besides this, empirical evidence pertaining to capital structure and trade credit of small firms strongly motivates me to believe that firms which are unable to generate sufficient operating cash flow (OCF) are more prone to financial distress, but lack supporting empirical evidence.

Casey and Bartczak (1985) are the earliest to study the incremental information content of operating cash flow (OCF) in predicting bankruptcy; however they conclude that OCF fails to provide any incremental default information. In contrast, Gentry et al. (1987) propose that classification of failed and non-failed firm improves significantly by studying cash based fund flow components and financial ratios. Additionally, Gilbert et al. (1990) emphasize that the explanatory power of their insolvency prediction model improves significantly by adding cash flow variables. Some studies also argue that, the utility of cash flow information in predicting financial distress is industry specific and does a good job for mining, oil and gas industries (e.g. Ward, 1994). Bernard and Stober (1989) argue that accruals can be manipulated by the managers to represent favorable situations; hence it provides only indirect links to expected cash flows. Whereas, cash flow from operation provides direct links to the liquidity position of the firm and gives a clear understanding about the firm's ability to meet its debt and interest obligation. Recent studies exploring the role of OCF information also prove mixed suggestions. Charitou et al. (2004) conclude that OCF information exhibit significant discriminatory power in predicting financial distress of UK industrial firms. While, Mazouz et al. (2012) study the information content of cash flow information using neural networks and report that cash flow information does not improve the performance of business default prediction models. Although we see some mixed empirical evidence exploring the role of operating cash flow information in predicting business failure, but all these studies were conducted on a sample of large firms. Hence, it would be interesting to examine the role of OCF in predicting financial distress of SMEs, specially in light of the following discussion.

It's been consistently reported in empirical studies that small firms in the US tend to have lower leverage ratios than large firms. Modigliani and Miller (1963) argue that firms value increases with increasing leverage, however this generalized conclusion is not convincing for small firms. Obert and Olawale (2010), Sogorb-Mira (2005), Michaelas et al. (1999), Pettit and Singer (1985) report that the use of debt has negative impact on the profitability of small firms and they reap less benefit from tax shield unlike large firms. Örtqvist et al. (2006)

emphasize that, the expected cost of bankruptcy is higher in small firms, which may outweigh any potential benefit from tax shield. With respect to the pecking order theory, López-Gracia and Sogorb-Mira (2008) report that internal resources represent the primary source of financing for SMEs; even Watson and Wilson (2003) report the existence of pecking order preference in UK SMEs. Thus, SMEs may have the tendency to use capital which minimizes intrusion into their business and as a consequence, the use of retained profits is expected to increase with firm size (Mac An Bhaird and Lucey, 2010). Considering the empirical evidence pertaining to the financing preferences of SMEs, I get strong motivation to believe that SMEs are primarily dependent upon their internal resource to meet their financing requirements.

From trade credit perspective, significant evidence motivates me to believe that operating cash flow is an important source of financing for SMEs. They face more difficulty in getting funding through formal credit institutions, and thus unlike large firms they are primarily dependent on internal financing (Hutchinson and Xavier, 2006). They also face very limited access to capital market (Petersen and Rajan, 1997); as a consequence they use less financing through financial institutions (Beck et al., 2008) and prefer to rely more on short term debt financing (Peel et al., 2000). Supporting this argument, García-Teruel and Martínez-Solano (2010) report that large European SMEs with more growth opportunities receive more trade credit from their supplier, but they use less trade credit when their capacity to generate internal resources increases, or when they have an opportunity to obtain external financing at lower cost. Thus, trade creditors form the major chunk of the liability of small firms and their bankruptcy is primarily influenced by the trade creditors rather than the institutional lenders (Hudson, 1986). The above arguments clearly demonstrate the significance of trade credit and internal financing in the long term survival of SMEs. In this context operating cash flow seems to be a prime factor in determining the long-term survival of small firms.

In this study, I am the first to examine the usefulness of operating cash flow information in explaining financial distress of UK SMEs[4]. I do this by including significant operating cash flow ratios as an enhancement to other accrual ratios which are obtained from balance sheet and income statement. I apply logistic regression technique to develop one year default prediction models (SME1 and SME2; SME1 model include significant accrual ratios obtained from balance sheet and income statement, while SME2 model employs significant operating cash flow ratios as an enhancement to SME1 model) using a relatively large sample of UK SMEs that survived between the analysis period of 2000 to 2009, and 2,666 firms that failed in the same time period. I also examine the incremental information content of OCF ratios in predicting financial distress of UK SMEs over the accrual ratios obtained from income statement and balance sheet. Finally to test the validity of the model developed I use receiver operating characteristics (ROC) curve and related classification accuracy measures, keeping in mind the concerns of Basel Committee (2000) on model validation.

My contribution to the existing literature is that I explore the information value of operating cash flow in modeling credit risk for SMEs. My empirical findings show that all the OCF ratios exhibits significant discriminatory power in univariate analysis but only one operating cash flow ratio (Cash flow from operation/current liabilities; CFOCL) exhibit significant discriminatory power in identifying failed and non-failed firms in the multivariate setup. The prediction power of the distress prediction model does not improve when operating cash flow information is included as an additional regressor. This is because both the models exhibit almost identical classification accuracy measures, and there is no statistically significant

[4] UK companies are required to file accounts at 'Companies House' (www.companieshouse.gov.uk) which defines a small company as one for which at least two of the following conditions are met: (i) Annual turnover is £6.5 million or less; (ii) the balance sheet total is £3.26 million or less; (iii) the average number of employees is 50 or fewer. It defines medium company as one for which at least two of the following conditions are met: (i) Annual turnover must be no more than £25.9 million; (ii) the balance sheet total must be no more than £12.9 million; (iii) the average number of employees must be no more than 250.

difference between the predicted probabilities of failure estimated using SME1 and SME2 models.

The remainder of the paper is organized in the following way. The methodology used to conduct the empirical analysis is shown in Section 3.2. The empirical findings are reported in Section 3.3. Finally Section 3.4 provides a conclusion.

3.2 EMPIRICAL METHODS

Studies related to SMEs financial health gained momentum during the past decade and researchers became particularly interested after the introduction of the Basel Capital Accord and recent financial crisis. As a consequence of Basel II, a firm is considered to be in financial distress if it has 90 days overdue on credit agreement payments. As far as the definition of SMEs is concerned, there is no common definition and it varies across different countries based upon some quantitative[5] and qualitative[6] aspects. In this study I accept the definition of SMEs as provided by the European Union, which classifies a firm as SME if it has less than € 50 million in annual sales revenue and less than 250 employees.

Since closing while successful can be a possible outcome (see among others, Bates, 2005; Headd, 2003; J. Watson and Everett, 1996), careful demarcation needs to be made among the firms which fail because of financial difficulties and firms which close down due to some strategic benefits. Hence, to improve the quality of my data, I take into account only those

[5] Total assets, annual turnover, number of employees are commonly used quantitative differentiators.
[6] Industry type, independent legal existence etc. are examples of qualitative definitions.

firms that have entered into liquidation due to pure business failure[7] (failure due to financial difficulties) and avoided any other form of business closure.

Further, in this section I discuss the following: (a) the dataset; (b) selection of predictor variables and (c) statistical methodology employed.

3.2.1 Sample Description

I perform the statistical analysis on a unique heterogeneous panel-data available to me from the Credit Management Research Center of the University of Leeds. The sample (with non-missing data) consists of 116,212 UK SMEs that survived during the analysis period of 2000 to 2009 and 2,666 firms that failed in the same time period. The data of analysis year 2008 and 2009 have been retained as a test sample (hold-out sample). Additionally, following standard academic practice I exclude utility and finance firms from my sample (see among others Altman et al., 2010). Table 3.1 shows detail of the analysed dataset.

Table 3.1: Dataset for the UK SMEs

Year	Failed	Non-Failed	Total	Failed/Total %
2000	251	6508	6759	3.71
2001	335	8084	8419	3.98
2002	305	8297	8602	3.55
2003	292	10832	11124	2.62
2004	282	14348	14630	1.93
2005	267	13663	13930	1.92
2006	212	13126	13338	1.59
2007	178	14055	14233	1.25
2008	294	13943	14237	2.07
2009	250	13356	13606	1.84
Total	2666	116212	118878	2.24

This tables shows the composition of the development and test sample used in my study that produce complete set of financial statements, i.e. balance sheet, income statement and cash flow statement. The first column shows the year when the financial statements were submitted. The second and third column shows the total number of failed and non-failed firms in respective financial years. The fourth column shows the total number SMEs analyzed in a particular financial year and last column shows the respective annual bad rate.

[7] Once a firm has become insolvent, the UK Act provides to choose one from the five courses of action: administration, company voluntary arrangement (CVA), receivership, liquidation and dissolution. In this study to represent the failed sample group we take under consideration only those SMEs whose failure followed any of the three common routes, i.e. administration, receivership or liquidation.

3.2.2 Selection of Predictor Variables

In this study the dependent variable has a binary outcome, i.e. failed or non-failed. To find which set of independent variables have more significant explanatory power, I develop two different default prediction models named SME1 and SME2 respectively.

3.2.2.1 Covariates Selection for SME1 Model

In this model the independent variables are the significant financial ratios obtained from income statement and balance sheet. The ratios capture firms' performance in terms of profitability, debt servicing capacity, liquidity, leverage and asset utilization. I include most of the financial ratios found successful in prior bankruptcy prediction studies (see among others Altman and Sabato, 2007; Altman et al., 2010). Specifically, I follow Altman et al. (2010), as their selection of explanatory variables is non-overlapping with strong theoretical underpinning. Considering the findings of Jones (2011), I also investigate the effect of intangible assets on firms' default propensity. He suggests that, higher proportions of intangible assets in a firm's capital structure signals higher default probability, as he finds firms approaching bankruptcy accumulate intangible assets more aggressively than non-failed ones. The details of the financial ratios along with their respective definitions are given in Table 3.2.

Table 3.2: Table of Independent Variables

Category	Variable Name	Variable Definition
Non-Cash Flow Ratios	EBITDATA	Earnings Before Interest Taxes Depreciation and Amortization / Total Assets
	STDEBV	Short Term Debt / Equity Book Value
	RETA	Retained Earnings / Total Assets
	CTA	Cash / Total Assets
	EBITDAIE	Earnings Before Interest Taxes Depreciation and Amortization / Interest Expense
	CETL	Capital Employed / Total Liabilities
	TCTL	Trade Creditors / Total Liabilities
	TDTA	Trade Debtors / Total Assets
	TTA	Taxes / Total Assets
	TCTA	Trade Creditors / Total Assets
	IATA	Intangible Assets / Total Assets
Cash Flow Ratios	CFOTA	Cash Flow From Operation / Total Assets
	CFOTL	Cash Flow From Operation / Total Liabilities
	CFOCL	Cash Flow From Operation / Current Liabilities
	CFOIE	Cash Flow From Operation / Interest Expense
	CFOAP	Cash Flow From Operation / Accounts Payable
	CFOAR	Cash Flow From Operation / Accounts Receivable

This table represent list of financial ratios tested in this study. (For each ratio the variable name along with the respective definition are presented; moreover, the ratios are broken down into two major categories).

In line with the previous studies I expect EBITDATA, RETA, CTA and EBITDAIE to have negative relationship with firms' default probability. Higher EBITDATA signifies more earnings per unit of asset employed, hence low probability of default. Similarly higher values of CTA, RETA and EBITDAIE are considered to be characteristics of healthy firm and signals low probability of default. On the other hand, a high value of STDEBV signifies more debt per unit of equity employed and signals financial distress. Hence, I expect a positive relationship between STDEBV and probability of default. Firms in financial distress are expected to have higher value of liabilities, and hence the leverage ratio CETL is expected to exhibit negative relationship with its default probability. Healthier firm are expected to have better liquidity position than distressed firms, and hence lower value of TCTL, TDTA and TCTA. I expect TTA to have negative relationship with default probability, as healthy firms with good liquidity position generated more revenue and hence pay more tax that the distressed firms. To capture the influence of intangible assets on firm's default probability, I calculate intangible assets/total assets (IATA) and expect it to have a negative relationship with default probability. As firms approaching distress capitalize intangibles more aggressively than their healthy counterparts (Jones 2011).

3.2.2.2 Covariates Selection for SME2 Model

SME2 model employs significant operating cash flow ratios as an enhancement to the list of covariates employed for developing SME1 model. Initially I select the operating cash flow ratios found significant in prior default prediction studies, that capture the efficiency of a firm in asset utilization, leverage, liquidity and debt repayment capacity (see among others Charitou et al., 2004; Mazouz et al., 2012). Table 3.2 contains the details of selected operating cash flow ratios along with their respective definitions. A high value of CFOTL and CFOCL signifies high cash flow from operation generated per unit of liability, hence lower would be the chance of default. Similarly, a high value of CFOTA signifies better asset utilization which leads to lower default risk. High value of CFOIE signifies high debt servicing capacity and hence, lower default probability. Additionally, higher value of CFOAP and CFOAR signifies better management of working capital and hence lower default probability. Thus, I expect all the operating cash flow ratios to have negative relation with probability of default, as higher value of these ratios signifies better financial health and hence low probability of default.

Further, to identify the significant cash flow ratios having prediction power I carry out univariate logistic analysis of each ratio in turn. I also perform two groups mean comparison (TGMC) test of each ratio to find if there is any significant difference in the mean of failed and non-failed groups of SMEs.

3.2.3 Logistic Regression

Traditionally multiple discriminant analysis technique (MDA) and logistic regression techniques are the most widely used statistical techniques for modeling credit risk. Altman (1968) is amongst the first researchers to apply MDA technique to predict the default probabilities of firms by calculating his well-known Z-Score[8]. There after MDA remained the popular statistical technique used by researchers for developing default prediction studies, until Ohlson (1980) challenged the restrictive assumptions of MDA[9]. Considering the problems of MDA, Ohlson (1980) for the first time applied conditional logit techniques for default prediction studies. Unlike MDA, logit techniques do not require the restrictive assumptions of MDA and could be applied to disproportional samples. Some other methodologies like expert system, neural networks, smoothing non-parametric methods etc. have been developed beside these basic models and are now being widely used for modelling and understanding credit risk (see Caouette et al., 2008 for further details).

Logistic regression seems to be the appropriate choice for default prediction studies, since the dependent variable is binary (default/non-default) and the groups being non-overlapping, discrete and identifiable. The logit model provides a score between zero and one which can be conveniently interpreted as the probability of default. Since the work of Ohlson (1980), substantial volume of the academic literature (see among others Keasey and R. Watson, 1987; Altman and Sabato, 2007; Altman et al., 2010) have used logit regression technique for default prediction studies.

[8] It is a multivariate model developed using five financial ratios: working capital/total assets, retained earnings/total assets, earnings before interest and tax/total assets, market value of equity/book value of total debt and sales/total assets.

[9] The two restrictive assumptions of MDA analysis are: i) the independent variables included in the model are multivariate normally distributed; ii) the group dispersion matrices (or variance-covariance matrices) are equal across the failing and the non-failing group. See Barnes (1982) and Karels & Prakash (1987) for further discussions about this topic.

Considering the nature and objective of my study, I use logistic regression technique as an appropriate statistical technique for my study. This technique gives the score for each company to be classified either as healthy or failed. My firm-level observations are pooled over time and the covariates are time-varying for individual firms until its year of failure. I assume that, the marginal probability of a firm's default over the next time period follows a logistic distribution represented as:

$$P(Y_{it} = 1) = \frac{e^{\,\beta X_{i,t-1}}}{1 + e^{\,\beta X_{i,t-1}}} \qquad (1)$$

Where $P(Y_{it} = 1)$ is the probability of default of firm *i* at time *t*, $X_{i,t-1}$ is the vector of time-varying covariates made available at the end of the previous time period and β is the vector of coefficients.

To evaluate the performance of the models developed, I report the receiver operating characteristics (ROC) curves. The ROC curve is obtained by plotting the true positive[10] against the false positive[11] rate as the threshold to discriminate between non-failed and failed firm's changes, while the area under ROC curves (AUROC) is a measure of prediction accuracy of the model, with AUROC equal to 1 representing a perfect model (see Anderson, 2007). The Gini coefficient and Kolmogorov–Smirnov (K-S) statistic are often used to evaluate the performance of a scoring model can be easily calculated from AUROC. The Gini coefficient calculated using the relation G = 2(AUROC – 0.5) is used to assess the consistency in the prediction of the model developed, while the K-S statistics measures the distance between the failed and non-filed distributions at the optimal cut-off point and is about 0.8 × Gini coefficient. A model having K-S statistics value below 20 should be questioned, whereas a model having value above 70 is regarded as too good to be true (see

[10] A firm actually defaults and the model has classified it as expected default.
[11] A firm does not default but the model has classified it as expected default.

Anderson, 2007). Further, to assess the validity of the models developed I report the classification accuracy measures of the test sample.

3.3 Empirical Results and Discussion

In this section I develop one year default prediction models named SME1 and SME2 respectively using logistic regression technique. SME1 model is developed using significant financial ratios obtained from income statement and balance sheet, while SME2 model employs significant operating cash flow ratios as an enhancement to SME1 model. Here I illustrate the steps involved in my analysis along with comparison and validation of the results obtained using appropriate statistical techniques. I correct the financial ratios for extreme values by restricting their ranges between 1st and 99th percentiles. Considering the extreme variability of STDEBV I restrict its range between 3rd and 97th percentiles.

3.3.1 Analysis of Descriptive Statistics

Table 3.3 contains descriptive statistics of the financial ratios employed in this study. An initial analysis of descriptive statistics is useful in understanding the variability of variables and the potential biasness that may arise due to extreme and unexpected variability. The mean and standard deviation of all the variables are fairly as per my expectation without any extreme variability, as the studied variables have already been winsorized. The variables which bear positive relationship with the default probability, I expect them to have higher mean value for failed group of firms than their non-failed counterparts and vice versa. The measures of all the variables are as per my expectation with expected sign of mean and standard deviation except RETA. The mean of RETA for the failed group is negative and has a minimum value of negative 0.57, which is quite surprising; given that I expected it to have a positive mean and minimum value. RETA can be negative if the value of total assets is negative, which is very unusual and I don't have any firm with negative total assets in my database. The other reason may be negative retained earnings. A company records negative retained earnings when the amount of loss exceeds the amount of retained profit in the previous accounts, which could be the potential reason behind the negative mean of failed firms. Approximately 30% of the firms in my database have reported negative retained earnings and out of which about 50% are failed firms.

Table 3.3: Key Descriptive Statistics

Variable		Mean	S.D.	Minimum	Maximum	Range
EBITDATA	Failed	.080318	.165835	-.3506738	.7700214	1.120695
	Non-failed	.1333057	.1490204	-.3506738	.7700214	1.120695
STDEBV3	Failed	2.675195	3.792773	-1.907838	15.01064	16.91848
	Non-failed	1.946625	3.180322	-1.907838	15.01064	16.91848
RETA	Failed	-.0240983	.1578841	-.5709065	.4395202	1.010427
	Non-failed	.0284329	.1186871	-.5709065	.4395202	1.010427
CTA	Failed	.0914446	.1650581	0	.997489	.997489
	Non-failed	.1388078	.1849743	0	1	1
EBITDAIE	Failed	45253.08	268756.9	-114000	3354000	3468000
	Non-failed	128536.6	452602.2	-114000	3354000	3468000
CETL	Failed	.7428339	1.550536	-.340678	22.21459	22.55527
	Non-failed	1.522591	2.871834	-.340678	22.21459	22.55527
TCTL	Failed	.315103	.2207553	0	1	1
	Non-failed	.2807083	.2319892	0	1	1
TDTA	Failed	.2928032	.2136701	0	.9954661	.9954661
	Non-failed	.2531184	.2125032	0	1	1
TTA	Failed	.0110534	.0290851	-.0406614	.171612	.2122735
	Non-failed	.0216559	.0332586	-.0406614	.171612	.2122735
TCTA	Failed	.251035	.1909932	0	.774744	.774744
	Non-failed	.1815932	.1763054	0	.774744	.774744
IATA	Failed	.0442048	.1211628	-.0022307	.636151	.6383818
	Non-failed	.0331015	.1065245	-.0022307	.636151	.6383818
CFOTA	Failed	.1555275	.1681985	.002933	.9534515	.9505185
	Non-failed	.1669361	.1642146	.002933	.9534515	.9505185
CFOTL	Failed	.2418883	.3803825	.0044222	3.06383	3.059408
	Non-failed	.3390422	.4452106	.0044222	3.06383	3.059408
CFOCL	Failed	.3242861	.5206171	.0057143	4.18721	4.181496
	Non-failed	.4641382	.6020674	.0057143	4.18721	4.181496
CFOIE	Failed	66395.17	331791.7	.2	3707000	3707000
	Non-failed	146781.4	498717.8	.2	3707000	3707000
CFOAP	Failed	10878.42	112340.9	.0165433	1699000	1699000
	Non-failed	31882.88	196154.8	.0165433	1699000	1699000
CFOAR	Failed	25380.18	183961.5	.0123967	2052000	2052000
	Non-failed	42850.14	239029.7	.0123967	2052000	2052000

First column lists the covariates studied followed by the failed and non-failed groups in the second column. Third, fifth, seventh and ninth columns report the mean and fourth, sixth, eighth and tenth columns report the standard deviation of micro firms, small firms, medium firms and SMEs respectively.

3.3.2 Univariate Analysis

Table 3.4 shows the summary of the results obtained in univariate logistic analysis and two groups mean comparison test. As expected, all the variables are highly significant in discriminating between the failed and non-failed firms in the univariate logistic analysis and two groups mean comparison test with expected sign of the respective coefficients. I use the predictor variables that are significant in the univariate analysis to develop the multivariate model. However, it should be noted that variables which appear to be significant in the univariate analysis may not be significant in the multivariate model due to multicollinearity between the explanatory variables.

Table 3.4: Univariate Logistic Analysis and Two-Group Mean Comparison Test

Variable Name	Sign	Coefficient	Sig.
EBITDATA	-	-2.656694***	0.0000
STDEBV	+	.0561219***	0.0000
RETA	-	-2.730698***	0.0000
CTA	-	-1.853709***	0.0000
EBITDAIE	-	-8.96e-07***	0.0000
CETL	-	-.3749753***	0.0000
TCTL	+	.6097756***	0.0000
TDTA	+	.8295972***	0.0000
TTA	-	-14.16301***	0.0000
TCTA	+	1.838183***	0.0000
IATA	+	.8204672***	0.0000
CFOTA	-	-.4581193***	0.0016
CFOTL	-	-.8197578***	0.0000
CFOCL	-	-.6710535***	0.0000
CFOIE	-	-5.96e-07***	0.0000
CFOAP	-	-1.02e-06***	0.0000
CFOAR	-	-4.22e-07***	0.0008

*** (**) [*] significant at the 1% (5%) [10%] level (two-sided test). The first column list the variables studied. The second column report the expected sign of the coefficients in logistic analysis and third column report the coefficient obtained in univariate logistic analysis. The fourth columns report the significance level (p-value) of two group mean comparison test (TGMCT).

3.3.3 Developing Bankruptcy Prediction Models

The two bankruptcy prediction models are estimated using a logistic regression technique. The dependent variable has binary outcome (failed and non-failed) and independent variables are the set of financial ratios analysed in univariate analysis section. The relationship between firm size and insolvency risk follows a non-linear relationship; insolvency risk being an increasing and decreasing function of firm size (see Altman et al., 2010). Following Altman et al. (2010), I used a quadratic term in total assets value as a control variable [ln (1+Total Assets) and (ln (1+Total Assets))^2][12] to control this size effect. To control for the macro-economic conditions facing the firms I construct an industry "weight of evidence" variable , that expresses the previous year sector failure rate as a log odds of failure in each of 51 industrial sectors (INDWOE) (see Altman et al., 2010). The population data of each sector is being used to calculate it as number of insolvencies in relation to number of active firms in each industrial sector. This serves as a useful proxy for controlling the volatile macro-economic conditions during the sampling period. The details of the two models developed are as follows.

3.3.3.1 SME1 Model

In this model the independent variables are the financial ratios obtained from income statement and balance sheet which are found significant in the univariate analysis. In order to select the best set of explanatory variables for developing the multivariate model, I employ stepwise logistic estimation with failed=1 and non-failed=0 under the 5% significance level. This technique selects variables based on the likelihood ratio test, taking into account multicollinearity problems (Dielman, 2000) and is particularly useful when dealing with large numbers of potentially correlated explanatory variables. However, I did not include the variables EBITDATA and TCTL in the stepwise estimation, as EBITDATA exhibits strong positive correlation of about 0.7 with TTA and RETA, while TCTL exhibits very strong positive correlation of about 0.85 with TCTA (see Table 3.5). This strong relation between EBITDATA and RETA is consistent with the empirical findings which argues that access to

[12] Size (log) and Size squared (log)

external financing is the major hurdle for the growth of SMEs and hence they are primarily dependent on their internal resources for financing needs (Beck and Demirguc-Kunt, 2006). However, Altman and Sabato (2007) and Altman et al. (2010) use the ratio EBITDATA to develop their multivariate model along with RETA but based upon my analysis I have clear motivation to exclude EBITDATA from my multivariate models. The estimated SME1 model is reported in Table 3.6.

Table 3.5: Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
EBITDATA	1	1.00																
STDEBV	2	-0.02	1.00															
RETA	3	0.66	-0.03	1.00														
CTA	4	0.18	0.04	0.15	1.00													
EBITDAIE	5	0.26	-0.02	0.20	0.21	1.00												
CETL	6	0.01	-0.14	0.11	0.09	0.23	1.00											
TCTL	7	-0.04	0.09	0.04	0.01	-0.02	-0.09	1.00										
TDTA	8	0.04	0.17	0.01	-0.13	-0.08	-0.30	0.40	1.00									
TTA	9	0.70	-0.01	0.32	0.23	0.21	-0.06	0.01	0.11	1.00								
TCTA	10	-0.11	0.20	-0.08	-0.06	-0.11	-0.31	0.85	0.48	-0.04	1.00							
IATA	11	-0.09	-0.04	-0.07	-0.10	-0.06	-0.05	-0.10	-0.07	-0.03	-0.06	1.00						
CFOTA	12	0.45	0.08	0.10	0.33	0.13	-0.09	-0.07	-0.01	0.41	-0.03	-0.03	1.00					
CFOTL	13	0.40	-0.06	0.22	0.33	0.30	0.51	-0.07	-0.17	0.31	-0.22	-0.07	0.64	1.00				
CFOCL	14	0.32	-0.10	0.17	0.22	0.24	0.46	-0.20	-0.27	0.20	-0.30	-0.03	0.54	0.85	1.00			
CFOIE	15	0.14	0.04	0.10	0.28	0.80	0.16	-0.03	-0.09	0.14	-0.08	-0.06	0.25	0.34	0.29	1.00		
CFOAP	16	0.04	0.00	0.03	0.05	0.12	0.14	-0.20	-0.16	0.00	-0.17	-0.04	0.10	0.17	0.24	0.16	1.00	
CFOAR	17	0.05	0.01	0.03	0.08	0.13	0.13	-0.14	-0.22	0.01	-0.13	-0.03	0.17	0.22	0.27	0.19	0.42	1.00

Table 3.6: Multivariate Logistic Models

Variable	SME1 Model	SME2 Model	Sig. of LR Test
EBITDATA	—	—	—
	—	—	—
STDEBV3	0.021774***	0.0220501***	—
	(0.0061713)	(0.0061742)	—
RETA	-1.862509***	-1.836027***	—
	(0.1688758)	(0.1683339)	—
CTA	-0.8686636***	-0.9087984***	—
	(0.1634587)	(0.1645556)	—
EBITDAIE	—	—	—
	—	—	—
CETL	-0.1417444***	-0.1553732***	—
	(0.0236297)	(0.0247869)	—
TCTL	—	—	—
	—	—	—
TDTA	—	—	—
	—	—	—
TTA	-9.845205***	-10.18302***	—
	(1.015491)	(1.021576)	—
TCTA	0.8466765***	0.8827736***	—
	(0.1264114)	(0.1275499)	—
IATA	1.107457***	1.106152***	—
	(0.1839902)	(0.1838781)	—
CFOTA	—	—	0.2351
	—	—	—
CFOTL	—	—	—
	—	—	—
CFOCL	—	-0.1214243**	0.0352**
	—	(0.0558004)	—
CFOIE	—	—	0.3974
	—	—	—
CFOAP	—	—	0.6658
	—	—	—
CFOAR	—	—	0.3791
	—	—	—
INDWOE	-0.4845057***	-0.4885412***	—
	(0.0459722)	(0.0460071)	—
Insolvency Rate	2.63876***	2.638863***	—
	(0.1971856)	(0.1972021)	—
Size (log)	1.1003***	1.162657***	—
	(0.3190475)	(0.3198593)	—
Size Squared (log)	-0.0413904***	-0.0434191***	—
	(0.0106627)	(0.0106865)	—
Constant	-13.51393***	-14.02185***	—
	(2.390808)	(2.399131)	—

*** (**) [*] significant at the 1% (5%) [10%] level. First column lists the covariates studied followed by the coefficient of covariates employed in SME1 (second column) and SME2 (third column) model respectively. The values within the parenthesis are the standard errors of respective covariates. The fourth column reports the p-values obtained from likelihood ratio test. '—' represents omission of the respective covariate from the statistical analysis due to reasons discussed in section 3.3.

The stepwise estimation suggests seven covariates for building the multivariate model. The variable EBITDAIE reported significant in prior empirical studies (e.g. Altman and Sabato, 2007) fail to be significant in my multivariate model, which is consistent with the empirical motivations. As access to finance through formal lending institutions is difficult for small firms, hence they primarily rely on trade creditors for their external financial needs (Hudson, 1986). Considering this argument, interest expenses are not expected to have significant influence on their default probability, whereas short-term liquidity measures are expected to play an important role. The variable TDTA also failed to exhibit significant discriminatory power in the stepwise estimation. The covariates which capture the liquidity position of a firm (STDEBV, CTA and TCTA) are highly significant in my multivariate model, highlighting the importance of sound liquidity position in the survival of small firms. The variables RETA, CETL, TTA and IATA are also highly significant with the expected sign of respective coefficients. The significance of RETA highlights the importance of retained earnings as an important source of finance and long-term survival. Consistent with the findings of Jones (2011), the proportion of intangible assets (IATA) is a highly significant predictor of small business failure but its impact may vary with the assets size, as micro firms (firms having less than 10 employees) hardly report any intangible assets. The covariates introduced to control the size effect and macroeconomic conditions are also highly significant with expected sign of respective coefficients.

The overall classification accuracy is calculated to evaluate the percentage of failed and non-failed firms correctly classified by the bankruptcy prediction model developed using the sample bad-rate[13] as the cut-off rate (Anderson, 2007). The within sample overall classification accuracy of SME1 model developed is about 66%, which is not too impressive and may substantially improve by addition of qualitative information (Altman et al., 2010). The model shows area under the ROC curve (AUROC) of about 0.72, which indicates that the scoring model performs fairly well (see Table 3.7).

3.3.3.2 SME2 Model

The next step is to test whether the financial information obtained from cash flow statement adds to the discriminatory power of the accrual based prediction model developed (SME1

[13] Percentage of failed firms.

model). Hence, I test whether the operating cash flow ratios found significance in my univariate analysis adds any additional explanatory power to the SME1 model. This is done by performing the likelihood ratio test (see Pampel, 2000). The likelihood ratio (LR) test is useful in comparing two logistic models. It is done by comparing the log likelihood between full model and restricted model. I perform LR test for each operating cash flow ratios in turn which are found significant in the univariate analysis. The full model employs all the variables of the SME1 model and one operating cash flow ratio in turn, while the restricted model employs only the variables of SME1 model. Although all the CFO ratios are significant in the univatiate analysis but as reported in Table 3.6, only the variable cash flow from operation/current liabilities (CFOCL) is significant in the LR test. Thus I get strong motivation to believe that operating cash flow information when made available, may not significantly improve the prediction performance of default prediction model.

The SME2 model is developed by employing all the financial ratios of SME1 model along with the significant operating cash flow variables obtained in the LR test, i.e. CFOCL (see Table 3.6). The model is estimated using a logistic regression technique with failed = 1 and non-failed = 0. CFOTL have negative coefficients which signals lower probability of failure for higher value of operating cash flow per unit of current liability, which is fairly as per my expectation. The SME2 model has an area under ROC curve of about 0.72 (within sample) and overall classification accuracy of 66.14% (within sample) (see Table 3.7).

Both the models developed, SME1 and SME2 exhibit identical classification performance which confirm that, operating cash flow information does not add to the discriminatory power of the bankruptcy prediction models developed above ratios obtained from income statement and balance sheet. Hence, by analysing cash flow statement lending institutions and policymakers are not expected gain any significant insight about the credit risk behaviour of

SMEs; as I do not see any statistically significant difference between the predicted probabilities of failure, estimated using SME1 and SME2 models.

3.3.4 Results Validation

Validation tests are used to examine the ability of the classification models developed to predict failure among a new set of companies. I use out-of-sample period as a validation method which is widely used in bankruptcy studies (see among others Altman et al., 2010; Charitou et al., 2004). A hold-out sample of 544 failed and 27,299 non-failed SMEs for the two analysis periods, 2008 and 2009 have been used to test the validation performance of the models developed. Table 3.7 summarizes the accuracy measures obtained by using the hold-out sample.

Table 3.7: Misclassification Rates and Accuracy Performance of Models Developed

		Overall Classification Accuracy	Area Under ROC Curve
SME1 Model	Within	66.09%	0.7210
	Hold-out	66.06%	0.6884
SME2 Model	Within	66.14%	0.7208
	Hold-out	66.12%	0.6882

This table shows the classification accuracy of the models developed using sample bad-rate as cut-off rate for development and hold-out samples. The third column shows the area under the ROC curve (AUROC) which is equal to the probability that the rating for a true positive (a firm actually defaults and the model has classified it as expected default) will be less than that for a true negative (a firm does not default and the model has classified it as expected non-default) plus 50 per cent of the probability that the two ratings will be equal.

I report the third and fourth columns in Table 3.7 to compare my results. The third column reports the overall classification accuracy of the models in correctly identifying failed and non-failed firms, being measured as complement of the weighted average of Type I and Type II error rates of the forecasting models. The fourth column show the area under the ROC curve (AUROC) which is equal to the probability that the rating for a true positive (a firm actually defaults and the model has classified it as expected default) will be less than that for a true negative (a firm does not default and the model has classified it as expected non-default) plus 50 per cent of the probability that the two ratings will be equal. AUROC of 50% implies a random (uninformed) model, whereas AUROC of 100% implies a perfect model but a credit scoring model would give something in between the perfect and the uninformative model. Hence, the higher the area under the ROC curve, the better is the model's performance assuming that the information may be misleading if the class distribution is skewed.

The overall classification accuracy of my models (for both SME1 and SME2) is about 66% and has AUROC of about 0.69 for the test-sample. I see that, both the models performs equally well in out of the sample forecasting. Thus, financial information obtained from cash flow statement does not improve our understanding of credit risk behaviour of SMEs over the information available from income statement and balance sheet. Further, comparison of the predicted probabilities estimated using SME1 and SME2 models reveal no significant difference which is consistent with my findings pertaining to their classification measures.

3.4 CONCLUSION

The primary objective of this study is to examine the incremental information content of operating cash flow information in predicting bankruptcy of UK SMEs. To examine this, I develop one year failure prediction model using the significant financial ratios obtained from income statement and balance sheet, along with significant operating cash flow ratios obtained from cash flow statement. Empirical evidence pertaining to trade credit and capital structure of SMEs motivate me to believe that, operating cash flow information could add significant discriminatory power to the models developed using accrual ratios obtained from income statement and balance sheet.

One year default prediction models (SME1 and SME2) were developed using a sample (with non-missing data) of 116,212 UK SMEs that survived in the period of 2000 to 2009 and 2,666 firms that failed in the same time period. The data of analysis year 2008 and 2009 have been retained as a test sample (hold-out sample). SME1 model corresponds to the model developed using significant financial ratios obtained from income statement and balance sheet, while SME2 model employs significant operating cash flow ratios as an enhancement to SME1 model.

Although, all the operating cash flow ratios exhibit significant discriminatory power in the univariate analysis, but test result shows that only one of my operating cash flow ratios (cash flow from operation/current liabilities; CFOCL) exhibit statistically significant discriminatory power in identifying failed and non-failed firms in the multivariate setup. However, classification accuracy measures obtained for SME1 and SME2 models are identical for within sample and hold-out sample, which motivate me to believe that the policymakers and lending institutions may not gain significant benefit in understanding the credit risk behaviour of SMEs by analysing an additional set of financial statement (i.e. cash flow statement).

My findings clearly confirm that operating cash flow information does not improve the prediction performance of the default models, as both SME1 and SME2 models exhibit identical classification performance measures. Gaining access to operating cash flow information for SMEs is a real challenge as firms are not obliged by law to submit cash flow statement. Hence considering my finding I do not see any marginal gain in understanding the credit risk behaviour of SMEs by analysing information obtained from cash flow statement.

4. The Effect of Internationalisation on Modelling Credit Risk for SMEs: Evidence from UK Market

4.1 Introduction

SMEs are the predominant type of business units in all OECD economies and account for about two-third of the total employment. Over the past decade, we have witnessed momentum in the study of their financial health, particularly after the introduction of Basel Capital Accord. Recent studies show that, SMEs demonstrate capacity to drive economic development at domestic and international levels. The International Trade Association (ITA) reports that 286,661 SMEs exported from the United States (US) in 2010, constituting about 98 percent of the total number of US exporting firms. This was approximately 34 percent of all US export revenue for that year (ITA, 2010). US SMEs also accounted for about 98 percent of total number of importers in the year 2010. Hence, understanding the effect of internalisation is of considerable relevance for SMEs. The OECD-APEC study (Secretariat 2009) aimed at identifying the major barriers to internalisation facing SMEs provides renewed impetus on the importance of SMEs in the global economic platform (Lloyd-Reason et al., 2009). Lloyd-Reason et al. (2009) report that growth and knowledge-related motives are most influential factors in driving SMEs towards internationalisation. Also, Acs et al. (2001) and Gjellerup (2000) report that explosive growth of low-cost technology, better information processing and communication technology, and reducing trade barriers, along with financial deregulation, are the key forces driving internationalisation of SMEs.

Previous literature broadly agrees that internationalisation has a positive influence on firms' performance (see among others Mallick and Yang 2013). Caves (2007) and Rugman (2009) argue that unlike their domestic counterparts, international firms enjoy less volatility in their revenue earnings due to diversified revenue streams and face lower business risk due to integrated international markets. Hout et al. (1982) report that they enjoy greater cost efficiency as they gain ability to exploit benefits from economies of scale due to higher volume of business. Benefits may also arise from differential input prices across different locations (Ghoshal, 1987; D. E. Thomas and Eden, 2004), tax saving from appropriate transfer pricing to subsidiary entitities, and arbitrage (Kogut, 1993; Allen and Pantzalis, 1996). International firms also enjoy valuable learning experience while serving diverse customer needs and competing in the international markets (Kostova and Roth, 2002; Zahra et al., 2000). On the darker side, foreign exchange risk (D. E. Thomas and Eden, 2004) and increased coordination and transaction costs have adverse impacts on international firms' performance. However, the majority of empirical studies report that the benefits of internationalisation outweigh the associated costs (see among others Ghoshal, 1987).

Similarly, SMEs that export may gain from economies of scale, enhanced labour productivity and management efficiency (Kogut, 1993; Grant et al., 1988), which potentially leads to cost savings and enhanced profitability. Burgman (1996) on the other hand argues that, through diversification of operations, international firms do not benefit from reduced earnings variability, but are exposed to higher level of risk (Michael et al., 2009) arising from exposure to multiple political environments, variability of exchange rates etc. This may ultimately result in a higher credit risk. Although international SMEs face higher credit risk, they are financially more transparent to lenders and suppliers than their domestic counterparts (Beck and Demirguc-Kunt, 2006). Thus, they may have better access to finance than their domestic counterparts, and fewer problems of financial distress. Lee and Kwok (1988) report that US based multinational and domestic corporations exhibit different capital structures, and different factors lead to their default risk (Doukas and Pantzalis, 2003). This context motivates my examination of the effect of internationalisation on the default risk of SMEs. This may help lending institutions and trade creditors in better understanding and pricing of credit risk. Considering the mixed empirical arguments discussed above, at this stage it is difficult to assess the impact of internationalisation on the default risk of SMEs.

There is extensive empirical literature on modelling default risk for large firms; primarily Altman (1968)'s Z-Score model which predicts firms' default risk using historic accounting information, and Merton (1974)'s subsequent approach, which employs security market information are the predominant ones. Thereafter, we witness a substantial increase in the number and complexity of default prediction studies due to the rapid advancement in technology and methodology. Recent empirical literature also shows momentum in understanding the credit risk behaviour of small firms. Using multivariate discriminant analysis, Edmister (1972) is the first to develop a distress prediction model for small businesses by analysing 19 financial ratios over the period of 1954 to 1969. Recently, Altman and Sabato (2007) study a panel of over 2000 US SMEs from 1994 to 2002 and develop a distress prediction model using logistic regression technique. Their prediction model employs a set of significant accrual ratios; however they acknowledge the need to also employ qualitative information to improve the predictive performance of their model. Empirical literature also highlight the significance of qualitative information such as business type, industrial sector, location, age, etc. (see among others Lehmann, 2003; Grunert et al., 2005) in understanding of firms' credit risk behaviour. Altman et al. (2010) take account of this issue. They study about 5.8 million UK SMEs and report that the prediction performance of Altman's (2007) model improve by about 13% when qualitative information pertaining to firms' non-financial characteristics and compliance information are made available. The only empirical study I am aware of which explores the credit risk behaviour of domestic and international SMEs separately is that of Arslan and Karan (2009). They employ a sample of 1,166 Turkish SMEs, 1,097 of which are domestic, with 87 defaults, and 192 of which are international, with 10 defaults. They report differences in the credit risk attributes of international and domestic SMEs using logistic regression techniques on a set of financial ratios. They suggest the two groups should be considered separately while estimating their default probabilities. However, their findings may be biased due to changing economic conditions of emerging economies over their sampling period and the extremely small number of defaulted international SMEs in their sample.

I contribute to the growing literature on SMEs by analysing the impact of internationalisation on the default risk of SMEs in the mature UK market. My empirical analysis employs a large dataset, made available to me from the Credit Management Research Centre of the University of Leeds. I control for macroeconomic conditions using measures similar to Altman et al. (2010). Although internationalisation may be achieved through multiple avenues, the principal avenue is exporting (Sullivan 1994, Ramaswamy *et al.* 1996). Thus, following the existing literature (Fatemi, 1988; Arslan and Karan, 2009), I classify a firm as international if it makes sales abroad, and domestic if it makes sales only in the domestic market. I apply dynamic logistic regression to develop separate default prediction models for domestic and international SMEs by employing a set of financial ratios, and compare the attributes that lead to a firm's failure for the respective groups. Finally, to test the validity of the models developed, I report receiver operating characteristics (ROC) curves and related summary statistics, keeping in mind the concerns of the Basel Committee (2000) (see Sahajwala and Van den Bergh, 2000) on model validation.

In the real world, building credit risk model for SMEs is highly limited by data availability. The comprehensive database made available to me from the Credit Management Research Centre (CMRC) of the University of Leeds contains financial information of 342,711 domestic SMEs (with 8,525 defaulted and 334,186 non-defaulted firms) and 344,205 international SMEs (with 9,114 defaulted and 335,091 non-defaulted firms) ranging over an analysis period of 2000 to 2009. All the firms in my sample have filed at least two sets of financial statements, i.e. balance sheet and income statement. Further I retain data of analysis years 2008 and 2009 as a test-sample to validate the predictive performance of the default prediction models developed.

My empirical findings show that all the attributes which affect the default probability of international SMEs are highly significant in also explaining the default probability of domestic SMEs, except the short-term debt/equity book value (STDEBV), which contradicts the suggestion of Arslan and Karan (2009). It should be noted that the accuracy measures I obtained by employing the same set of covariates, are slightly lower for international SMEs than their domestic counterparts. This motivates me to compare the weights of the regression coefficients of the default prediction models I developed. My test results confirm that the coefficients of four out of the nine common predictors (cash/total assets, capital employed/total liabilities, tax/total assets and trade creditors/total liabilities) exhibit significant statistical difference in their weights. Hence, I conclude that although the same set of financial ratios are significant in predicting the financial distress of domestic and international SMEs, they perform better for domestic SMEs than for their international counterparts. I also investigate the role of intangible assets as predictor in assessing the creditworthiness of SMEs. My test results confirm a significant positive relation between the proportion of intangible assets and firms’ default probability.

The remainder of the paper is organised as follows. Section 4.2 discusses the empirical methods that I employ in my study. In Section 4.3, I report my empirical findings and Section 4.4 presents my conclusion.

4.2 Empirical Methods

Here I describe the following; (a) the dataset, (b) the selection of predictor variables, (c) my choice of multivariate techniques and (d) performance evaluation of the multivariate models developed.

4.2.1 Dataset

I perform the statistical analysis on a sample (with non-missing data) of 686,916 UK SMEs (having annual turnover of less than £45 million) that survived over the period of 2000 and 2009, and 17,639 firms that failed in the same time period. Out of the total surviving firms, 334,186 are domestic and 335,091 are international with 8,525 and 9,114 defaults respectively (see Table 4.1 for more details). I retain the data of analysis years 2008 and 2009 as a hold-out sample to validate the predictive performance of the models developed. I lag my firm-year observations by one period to perform my empirical analysis. Finally, I use a set of available accounting information to estimate the probability of firms' default over the next time period.

Table 4.1: Dataset of UK SMEs

Year	Domestic SMEs				International SMEs			
	Failed	Non-failed	Total	% Failed	Failed	Non-failed	Total	% Failed
2000	652	14683	15335	4.25	776	18020	18796	4.13
2001	784	19197	19981	3.92	1052	23156	24208	4.35
2002	788	20063	20851	3.78	971	23458	24429	3.97
2003	764	27215	27979	2.73	893	29408	30301	2.95
2004	790	36964	37754	2.09	794	37945	38739	2.05
2005	763	39044	39807	1.92	791	38409	39200	2.02
2006	800	43964	44764	1.79	727	37205	37932	1.92
2007	818	45882	46700	1.75	717	40466	41183	1.74
2008	1132	44923	46055	2.46	1105	42919	44024	2.51
2009	1234	42251	43485	2.84	1288	44105	45393	2.84
Total	8525	334186	342711	2.49	9114	335091	344205	2.65

This table shows the composition of the development and test sample used in my study; all firms produce at least two sets of financial statements, i.e. balance sheet and income statement. The first column shows the analysis year. The next four columns list the details of domestic SMEs sample and the last four columns list the details of my international SMEs sample.

Table 4.1 reveals that the sample bad rate (percentage of defaulted firms) for both domestic and international SMEs moves in tandem throughout the sampling period. Since the bankruptcy rates are similar for both the groups, this initially suggests that similar factors might affect the insolvency hazard of international and domestic SMEs.

For this study I adopt the definition of SMEs provided by the European Union, i.e. less than € 50 million in annual sales revenue, with less than 250 employees. The UK Insolvency Act 1986 states, ‘a company is said to be insolvent if it either does not have enough assets to cover its debts (i.e. the value of assets is less than the amount of its liabilities), or it is unable to pay its debts as they fall due’. Once a firm has become insolvent, the Act requires it to choose one from the five courses of action available: administration, company voluntary arrangement (CVA), receivership, liquidation and dissolution. In my paper I define SMEs as defaulted, where failure follows any of the three common routes, i.e. administration, receivership or liquidation. I exclude utility, insurance and finance firms from my sample as they have different asset-liability structure.

4.2.2 SELECTION OF COVARIATES

I consider only accounting information that can be obtained from the income statement and balance sheet. I do not consider cash flow information, as recent empirical findings suggest that cash flow information does not add significant discriminatory power to the distress prediction models developed for UK SMEs (Gupta, Wilson, *et al.* 2014a). Moreover, the majority of SMEs do not file cash flow statement as part of their financial reporting due to the regulatory concession that they receive. I select the financial ratios found successful in prior default prediction studies. These essentially reflect a firm's profitability, leverage, liquidity and solvency conditions. In particular I employ most of the covariates found significant in the Altman et al. (2010) study, which is based on a sample of UK firms, and has a well justified and non-overlapping selection of explanatory variables. Table 3.2 lists my final selection of covariates along with their respective definition.

Table 4.2: Table of Explanatory Variables

Variable Name	Domestic SMEs	International SMEs	Variable Definition
EBITDATA	No	No	Earnings Before Interest Taxes Depreciation and Amortization / Total Assets
STDEBV	No	Yes	Short Term Debt / Equity Book Value
RETA	Yes	Yes	Retained Earnings / Total Assets
CTA	Yes	Yes	Cash / Total Assets
EBITDAIE	Yes	Yes	Earnings Before Interest Taxes Depreciation and Amortization / Interest Expense
CETL	Yes	Yes	Capital Employed / Total Liabilities
QACA	No	No	Quick Assets / Current Assets
lnCR	Yes	Yes	ln(Current Assets / Current Liabilities)
TCTL	Yes	Yes	Trade Creditors / Total Liabilities
TDTA	Yes	Yes	Trade Debtors / Total Assets
SWC	No	No	Stock / Working Capital
TTA	Yes	Yes	Taxes / Total Assets
TCTA	No	No	Trade Creditors / Total Assets
STA	No	No	Stock / Total Assets
IATA	Yes	Yes	Intangible Assets / Total Assets
EI	No	No	Export / Sales
EIL	No	No	EI < 0.2
EIM	No	No	0.2 ≤ IE < 0.5
EIH	No	No	0.5 ≤ IE ≤ 1

This table lists the predictor variables studied (for each predictor the variable name along with the respective definition is presented). The second and third columns list the variables used to develop credit risk model for domestic SMEs (No = variable not included in the model; Yes = variable included in the model) and international SMEs (No = variable not included in the model; Yes = variable included in the model) respectively.

Considering the fact that internationalisation through export remains the dominant internationalisation strategy for SMEs, I define a firm as international if it reports export revenue, and domestic otherwise. To capture the impact of export intensity on firms' default risk, I calculate the export intensity (EI) as export to sales ratio which is the most commonly used method of measuring export intensity (see Katsikeas et al., 2000). I also employ three dummy variables (export intensity low (EIL); export intensity medium (EIM) and export intensity high (EIH)) (see Table 4.2) to capture any dependency of an international firm on its export earnings.

A higher value of the accounting ratio short-term debt/equity book value (STDEBV) reflects higher debt per unit of equity employed, and hence signals higher default probability. High value of capital employed/total liabilities (CETL) on the other hand, reflects a low value of total liabilities, and therefore signals lower default probability. The profitability ratio, retained earnings/total assets (RETA), measures the cumulative profitability of the firm and its capacity to accumulate profit from sales. A financially distressed firm is expected to have declining retained earnings, and thus RETA is expected to have a negative relationship with default probability. Higher value of earnings before interest tax depreciation and amortization/total assets (EBITDATA), cash/total assets (CTA), and earnings before interest tax depreciation and amortization/interest expense (EBITDAIE), are considered to be characteristics of a healthy going concern, thus I expect them to show negative relationship with firms' default risk.

Empirical literature pertaining to trade credit of small firms reports that, firms' facing financial difficulties demand extended credit from their suppliers and they provide extended credit to their customers. Hudson (1986), argues that trade creditors form a significant portion of a firm's liabilities, and bankruptcy is primarily led by trade creditors, rather than institutional lenders. My variable selection also reflects the importance of short term leverage on firms' default risk. The accounting ratio quick assets/current assets (QACA), reflects the proportion of liquid assets with respect to current assets. A healthier firm is expected to have a better liquidity position and hence higher QACA ratio than a financially distressed firm. Similarly, I expect a negative relationship between log of current ratio (lnCR), and default probability. I expect trade creditors/total liabilities (TCTL), trade debtors/total assets (TDTA), stock/working capital (SWC), trade creditors/total assets (TCTA) and stock/total assets (STA), to have a positive relationship with a firm's default probability, as higher values of these ratios signals financial distress. A firm having good liquidity position is not expected to default on its tax obligations, and the more profit it reports the higher the amount of tax it pays. Thus, tax/total asset (TTA) is expected to be negatively related to a firm's default probability. Recent empirical evidence, shows that firms approaching failure capitalize intangible assets more aggressively than their non-failed counterparts (Jones 2011). Hence, higher proportions of intangible assets signal a higher default probability. In order to capture the impact of intangibles on firms' financial health I calculate intangible assets/total assets (IATA) and expect it to have negative relationship with the probability of default.

4.2.3 Statistical Model Applied

Multiple discriminant analysis (MDA) and logistic regression are the traditionally preferred statistical techniques for modelling firms' default risk. Altman (1968) is the earliest to apply MDA technique to predict firms' default risk by calculating his celebrated Z- Score[14]. Thereafter MDA remained a widely preferred statistical methodology for default prediction studies, until Ohlson (1980) challenged its restrictive assumptions[15]. The MDA technique does not allow us to determine the relative importance of covariates, as the standardized coefficients are not interpreted as the slope coefficients of a regression equation. In view of the restrictive assumptions of MDA, Ohlson (1980) employed a conditional logit technique in a default prediction study for the first time. The Logit technique does not require the restrictive assumptions of MDA, and works fairly well with disproportional samples. Since the work of Ohlson (1980), a substantial proportion of the academic literature (see among others Gentry et al., 1987; H. D. Platt and M. B. Platt, 1991; Becchetti and Sierra, 2003) has used logit regression technique in default prediction studies. Shumway (2001) recently proposed a dynamic approach to measuring default probability. He termed the single-period classification models employed by Altman (1968) and others, the "static approach of

[14] It is a multivariate model developed using five financial ratios: working capital/total assets, retained earnings/total assets, earnings before interest and tax/total assets, market value of equity/book value of total debt and sales/total assets.

[15] The two restrictive assumptions of MDA analysis are: i) the independent variables included in the model are multivariate normally distributed; ii) the group dispersion matrices (or variance-covariance matrices) are equal. across the failing and the non-failing group. See Barnes (1982) and Karels and Prakash (1987) for further discussions about this topic.

estimating default rates". He argues that static models ignore the fact that firm characteristics changes over time, hence the default probabilities estimated are biased and show poor out-of-sample performance. Shumway's approach by contrast, uses multi-period default data. He introduced time-varying covariates, and argues the superiority of this approach in modelling default rates over static models. Methodologies such as neural networks, smoothing non-parametric technique, expert system etc. have also been widely applied for measuring and understanding credit risk (see Caouette et al., 2008 for further details).

Given the nature and objective of my study, I use logistic regression technique as an appropriate statistical technique. It is an appropriate choice where the dependent variable is binary, as with default/non-default. This technique allows the score (probability) for each company to be classified either as default or non-default. It uses maximum likelihood estimation (MLE), which; (i) transforms the outcome or dependent variable into a log function; (ii) estimates the quantitative value of the coefficients; (iii) determines changes to the coefficient, to maximize the log likelihood function. My firm-level observations are pooled over time, and the covariates are time-varying for each individual firm until its year of failure. The marginal probability of a firm’s default over the next time period is assumed to follow a logistic distribution represented as:

$$P(Y_{it} = 1) = \frac{e^{\beta X_{i,t-1}}}{1 + e^{\beta X_{i,t-1}}} \qquad (1)$$

Where $P(Y_{it} = 1)$is the probability of default of firm *i* at time *t*, $X_{i,t-1}$ is the vector of time-varying covariates, made available at the end of the previous time period, and β is the vector of coefficients.

4.2.4 Performance Evaluation

Numerous tools are available to evaluate the predictive performance of a scoring model. However, in line with the previous empirical literature (see among others Altman et al., 2010), I focus on the *misclassification matrix* and *receiver operating characteristic (ROC) curve*, thereby addressing the concerns of the Basel Committee (2000) on model validation.

A very simple and intuitive way of evaluating the predictive performance of a model with binary outcomes is to calculate the percentage of outcomes that the predictive model has correctly classified. The percentage of outcomes correctly classified is obtained from a *misclassification matrix* created by: a) choosing a cut-off score, which is generally the score corresponding to the sample bad rate; b) marking outcomes below the cut-off scores as expected default and above the cut-off score expected non-default; c) cross-tabulating the expected failure and non-failure against the actual outcomes; d) calculating the percentage of failure and non-failure correctly identified by the predictive model, and finally the overall classification accuracy is measured as a complement of the weighted average of Type I and Type II error rates of the scoring models (Anderson 2007). The correctly classified outcomes are called *true positive* (the firm has defaulted and the model has classified it as expected default) and *true negative* (the firm has not defaulted and the model has classified it as expected non-default) respectively. On the other hand, wrongly classified outcomes are labelled as *false positive* (false alarm; Type I error; the firm has actually defaulted and the model has classified it as expected non-default) and *false negative* (Type II error; the firm has not defaulted and the model has classified it as expected default) respectively. However, it is desirable to also account for the various misclassification costs, before setting the cut-off score, as in my case the cost associated with Type I error is much higher than for Type II error. Hence, maintaining lower Type I error over Type II error is a natural choice.

Another commonly used tool to measure the predictive performance of a scoring model is the ROC curve. The ROC curve is a plot of *sensitivity* (model's ability to identify true positives) against *specificity* (model's ability to identify true negatives) (Anderson 2007). The area under ROC curves (AUROC) is a measure of prediction accuracy of a default prediction model. AUROC equal to 0.5 represents an uninformative model and AUROC equal to 1 represent a perfect model. Thus an informative prediction model should have AUROC between 0.5 and 1. Gini coefficient, and Kolmogorov–Smirnov (K-S) statistic are often used to evaluate the performance of a scoring model, and these can be easily estimated from AUROC. The Gini coefficient estimated using the relation G = 2(AUROC – 0.5), is used to assess the prediction consistency of the model developed. The K-S statistics measures the distance between the failed and non-failed distributions at the optimal cut-off point, and is about 0.8 times the Gini coefficient. A model having K-S statistic values below 20 is questionable, whereas a value above 70 is regarded as too good to be true (Anderson 2007).

4.3 Results and Discussion

I start this section with the analysis of descriptive statistics of selected covariates, to understand any unexpected variability or potential bias that may arise due to extreme variability. Next, I conduct univariate analysis of each covariate in turn, before employing them in the multivariate framework. Finally, I use dynamic logistic regression to develop separate default prediction models for domestic and international SMEs. I also discuss the steps involved in building the models, and the comparison and validation of the obtained results. To avoid the influence of extreme outliers, I restrict the range of the selected financial ratio between the 1st and 99th percentiles. An exception is STDEBV, where I restrict its range between the 3rd and 97th percentiles because of its extreme variability.

4.3.1 Analysis of Descriptive Statistics and Correlation

The initial analysis of descriptive statistics is useful in understanding the variability of the covariates employed in the study, and any potential bias that may arise in the multivariate estimation due to their unexpected and extreme variability. Table 4.3 reports the key descriptive statistics of the variables employed in my study. The mean and standard deviation of all the variables are as expected, since the required covariates have already been winsorized. Where variables are positively related to the default probability, I expect the mean of the failed group to be higher than that of the non-failed group; both for domestic and international SMEs (e.g. see the variable STDEBV in Table 4.3). For variables which exhibit negative relationship with the probability of default, I expect the mean of the failed group to be lower than that of their non-failed counterparts (e.g. see the variable CTA in Table 4.3). I see that the mean and standard deviation of variables which capture the impact of export (EI, EIL, EIM and EIH) on the probability of default are very close for both failed and non-failed groups. This suggests that these measures may be insignificant in discriminating between failed and non-failed firms in the multivariate framework. Finally, the mean and standard deviation of EBITDAIE is very high, as a significant number of firms in my database incur no interest expense. Thus, all the earnings are available[16] to meet such a financial obligation, leading to very high value of earnings to interest ratio. Further, inspection of the correlations among the covariates reveals strong positive correlation of about 0.75 between EBITDATA and RETA (see Table 4.4). This is consistent with the view that SMEs find difficulty in accessing external finance, and are primarily dependent on internal sources for their financing needs. I also observe a strong positive correlation of about 0.60 between EBITDATA and TTA and approximately 0.8 between TCTA and TCTL (see

[16] If a firm reports EBITDA as 25,000 GBP and no interest expense, then the ratio EBITDAIE is 25000.

Table 4.4), suggesting that these covariates may be problematic in the multivariate framework.

Table 4.3: Key Descriptive Statistics

Variables		Domestic SMEs		International SMEs	
		Mean	Standard Deviation	Mean	Standard Deviation
EBITDATA	Failed	.0573551	.3774481	.0314151	.353931
	Non-failed	.1411848	.3321505	.1069662	.3098826
STDEBV	Failed	4.941976	13.53655	4.481691	12.7993
	Non-failed	4.286182	12.73009	3.9089	11.94138
RETA	Failed	-.0717598	.3493299	-.0885644	.3388518
	Non-failed	.0060786	.2825843	-.0173345	.2811975
CTA	Failed	.1027573	.1857104	.0910337	.1663
	Non-failed	.1802752	.236127	.1334606	.1962424
EBITDAIE	Failed	30641.67	249347.8	35700.05	288332.1
	Non-failed	98292.06	399519.8	111026.3	451876.6
CETL	Failed	.9366238	4.423782	.8776951	4.197403
	Non-failed	2.507379	7.950033	1.636794	6.043295
QACA	Failed	.783762	.2615726	.7755536	.2432354
	Non-failed	.8224892	.2599354	.7967032	.2603316
lnCR	Failed	-.0288736	.8172542	-.0481977	.7884853
	Non-failed	.2406938	1.032938	.1326042	.9996527
TCTL	Failed	.2975213	.2443079	.2865416	.2244876
	Non-failed	.2390211	.2495808	.2309505	.2334513
TDTA	Failed	.3136536	.2622952	.3033715	.2398819
	Non-failed	.2356297	.2492184	.2509065	.2373963
SWC	Failed	.4804538	5.768245	.629056	5.881221
	Non-failed	.5452661	4.728147	.5709797	4.968579
TTA	Failed	.0156676	.0441626	.0104995	.0390723
	Non-failed	.0231501	.0470204	.0203259	.0447914
TCTA	Failed	.264391	.2356712	.2506477	.2127614
	Non-failed	.166308	.2021198	.1648345	.1875209
STA	Failed	.1589481	.2145833	.1626417	.1987546
	Non-failed	.1255211	.2102469	.1430974	.2102131
IATA	Failed	.0270881	.1015657	.0265687	.1025117
	Non-failed	.0240295	.10073	.0246749	.10454
EI	Failed	----	----	.1044578	.2331564
	Non-failed	----	----	.1060829	.2398633
EIL	Failed	----	----	.8383386	.3681674
	Non-failed	----	----	.8394132	.3671502
EIM	Failed	----	----	.0760496	.2650973
	Non-failed	----	----	.0701161	.255343
EIH	Failed	----	----	.0856118	.2798109
	Non-failed	----	----	.0904708	.2868556

First column lists the covariates, followed by failed and non-failed groups in the second column. The third and the fourth columns report the mean and standard deviation of domestic SMEs, while fifth and sixth columns report the mean and standard deviation of international SMEs.

Table 4.4: Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
EBITDATA	1	1														
STDEBV	2	0.0650	1													
RETA	3	0.7340	0.0141	1												
CTA	4	0.1363	0.0188	0.0849	1											
EBITDAIE	5	0.2308	-0.0075	0.1754	0.0589	1										
CETL	6	0.0440	-0.0892	0.0720	0.0662	0.1321	1									
QACA	7	0.0565	0.0044	0.0250	0.2858	0.0832	0.0963	1								
lnCR	8	0.1689	-0.0899	0.2373	0.2963	0.1174	0.4586	0.0935	1							
TCTL	9	0.0177	0.0319	0.0593	0.0326	-0.0523	-0.0773	-0.0492	0.0917	1						
TDTA	10	0.0170	0.0840	0.0095	-0.1448	-0.0812	-0.1890	0.1929	0.0114	0.3584	1					
SWC	11	0.0123	0.1764	0.0166	-0.0307	-0.0094	-0.0232	-0.1432	0.0177	0.0262	-0.0049	1				
TTA	12	0.6042	0.0218	0.2504	0.1791	0.1800	-0.0427	0.0727	0.1008	0.0117	0.0701	-0.0002	1			
TCTA	13	-0.0978	0.1068	-0.0964	-0.0535	-0.1094	-0.1908	-0.0610	-0.1301	0.8261	0.4156	0.0243	-0.0381	1		
STA	14	-0.0448	0.0338	0.0059	-0.2171	-0.0798	-0.1075	-0.8696	0.0804	0.0988	-0.1091	0.1805	-0.0457	0.1127	1	
IATA	15	-0.0472	-0.0260	-0.0342	-0.0821	-0.0338	-0.0233	0.0056	-0.1024	-0.0522	-0.0671	-0.0187	-0.0115	-0.0426	-0.0642	1

4.3.2 UNIVARIATE ANALYSIS

Table 3.4 reports the estimates obtained from univariate logistic analysis, and two-group (failed and non-failed) mean comparison tests. To perform univariate logistic analysis, I select each covariate in turn as independent variable and run the logistic regression to determine the direction and significance of relationship with my binary (failed = 1 and non-failed = 0) dependent variable. I run separate sets of estimates for domestic and international SMEs, to understand any differences due to SMEs' exposure to international markets. I expect the coefficient of stock/working capital (SWC) to be positive, but domestic SMEs exhibits negative and insignificant, while international SMEs shows a positive and insignificant relationship with the binary default indicator. The variable IATA shows a positive coefficient for both domestic and international groups as per my expectation, but the univariate tests are significant only for domestic SMEs. All the variables employed to capture the impact of export exposure on firm's financial distress, are highly insignificant in univariate logistic analysis and two-group mean comparison tests, except EIM, suggesting no significant differences in the credit risk attributes of domestic and international SMEs.

Moreover, the same set of covariates, excepting IATA, show highly significant discriminatory power in identifying failed and non-failed firms for both domestic and international SMEs. Thus, my initial findings contradict the suggestion of Arslan and Karan (2009), as I see that a similar set of factors affects the default risk of both the groups. The remaining covariates (EBITDATA, STDEBV, REAT, CTA, EBITDAIE, CETL, QACA, lnCR, TCTL, TDTA, TTA, TCTA and STA) are highly significant in differentiating between failed and non-failed groups of firms (both domestic and international), along with the expected sign of the respective coefficients.

Table 4.5: Univariate Logistic Analysis and Two-Group Mean Comparison Test

Variable Name	Domestic SMEs			International SMEs		
	Sign	p-value	TGMCT	Sign	p-value	TGMCT
EBITDATA	-	0.000	0.000	-	0.000	0.000
STDEBV	+	0.000	0.000	+	0.000	0.000
RETA	-	0.000	0.000	-	0.000	0.000
CTA	-	0.000	0.000	-	0.000	0.000
EBITDAIE	-	0.000	0.000	-	0.000	0.000
CETL	-	0.000	0.000	-	0.000	0.000
QACA	-	0.000	0.003	-	0.000	0.003
lnCR	-	0.000	0.000	-	0.000	0.000
TCTL	+	0.000	0.000	+	0.000	0.000
TDTA	+	0.000	0.000	+	0.000	0.000
SWC	-	0.291	0.291	+	0.347	0.347
TTA	-	0.000	0.000	-	0.000	0.000
TCTA	+	0.000	0.000	+	0.000	0.000
STA	+	0.000	0.000	+	0.000	0.000
IATA	+	0.018	0.018	+	0.141	0.143
EI	---	---	---	-	0.584	0.584
EIL	---	---	---	-	0.813	0.813
EIM	---	---	---	+	0.061	0.061
EIH	---	---	---	-	0.171	0.171

The first column list the variables studied. The second and third columns report the sign of the coefficient and the significance level respectively, obtained from univariate logistic analysis of domestic firms. The fourth column reports the significance level of two group mean comparison test (TGMCT) of domestic firms. The fifth and sixth columns report the sign of the coefficients and the significance level respectively, obtained from univariate logistic analysis of international firms. The seventh column reports the significance level of two group mean comparison test (TGMCT) of international firms.

I use all of the significant explanatory variables from my univariate analysis to develop the multivariate models, except EBITDATA, QACA and TCTA, as they exhibit strong correlation with other covariates. I expect that some of the covariates might lack significant explanatory power in the multivariate framework, due to multicollinearity.

4.3.3 Developing Multivariate Logistic Models

I develop two bankruptcy prediction models, one for my domestic, and one for my international SMEs samples. The dependent variable in both the models has binary outcomes (failed and non-failed), and the explanatory variables are the set of accounting ratios with significant discriminatory power in identifying failed and non-failed firms in the prior univariate analysis. Furthermore, I perform stepwise logistic estimation under the 5% significance level, to identify the best set of covariates to build the multivariate model. Both the forward selection and backward elimination methods of stepwise estimation suggest the same set of covariates. A non-linear relationship between insolvency rate and firm size has been established in previous empirical literature, with insolvency risk being an increasing and decreasing function of firm size (Altman *et al.* 2010). Thus, following Altman et al. (2010), I control the size effect by employing a quadratic term in total assets value [(natural logarithm of (1 + total assets) and (natural logarithm of (1 + total assets))^2]. To control for macro-economic[17] conditions facing the firms I use previous year's sector failure rate (Insolvency Rate) in each of 51 industrial sectors and construct an industry "weight of evidence" variable. This expresses the previous year sector failure rate (Insolvency Rate) as a log odds of failure in each of the 51 industrial sectors (INDWOE) (see Altman et al., 2010). I use the population data of each sector to calculate this variable, as number of insolvencies in relation to number of active firms in each industrial sector. This serves as a useful proxy for

[17] The classification performance of our default prediction models (for both domestic and international SMEs) estimated without macro-economic control variables decrease by about 3%, while rest of the covariates maintain their statistical significance with expected signs. This confirms that the prediction performance of our default prediction models is substantially due to firms' characteristics.

controlling the volatile macro-economic conditions during the sampling period. I expect Insolvency Rate to be positively related to a firm's default risk as higher sector level failure enhances a firm's default likelihood and vice versa. Although INDWOE is the insolvency rate in each industrial sector, it is calculated as an index (log odds of failure in each sector). It has zero as the base rate. Negative values of INDWOE indicate higher insolvency risk, while positive values indicate lower insolvency risk. Thus, I expect INDWOE to exhibit negative relationship with firms' default probability. The details of the models developed are discussed below.

4.3.3.1 Default Prediction Model for Domestic SMEs

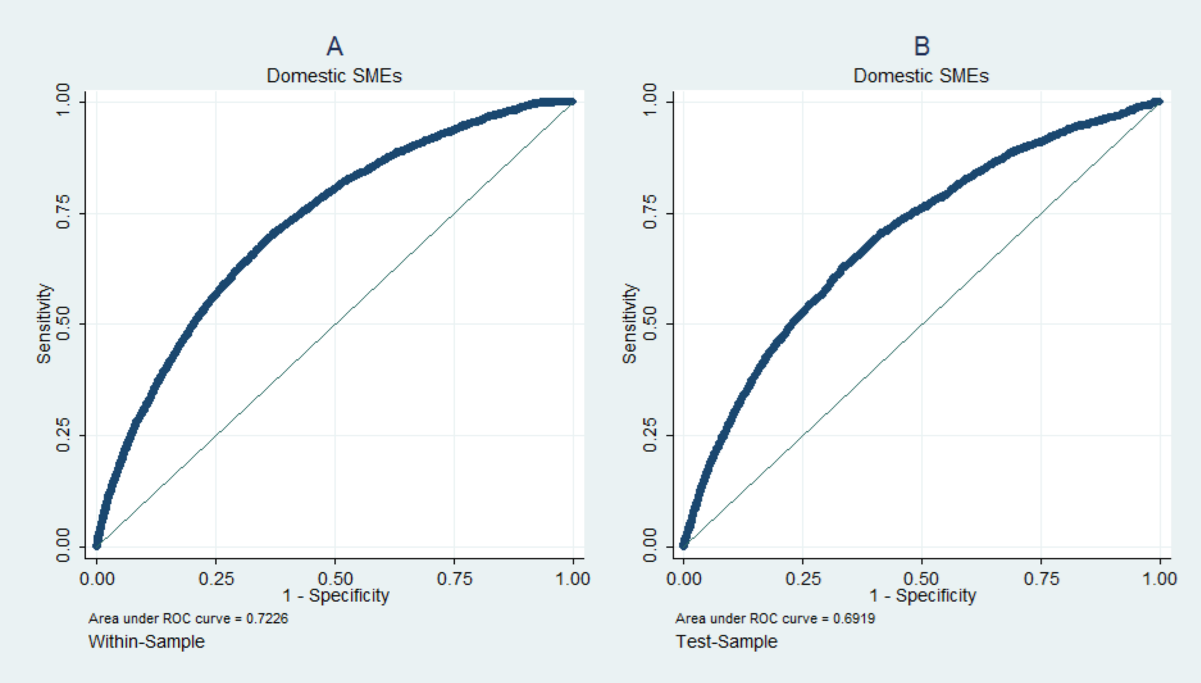
I use my sample of domestic SMEs to develop the default prediction model for domestic UK SMEs by employing logistic regression with failed = 1 and non-failed = 0. I eliminate the covariates EBITDATA, QACA and TCTA from my list of potential explanatory variables because, as discussed above, they are strongly correlated with other covariates. I then perform stepwise estimation to detect the best set of covariates to explain the default propensity of domestic SMEs. The variables short term debt/equity book value (STDEBV), stock/working capital (SWC) and stock/total assets (STA) are eliminated in the stepwise estimation process. Thus, the final model for domestic SMEs is estimated using nine highly significant financial ratios, with expected signs of their respective coefficients (see Table 4.6 for more details). The in-sample overall classification accuracy of the model developed is about 65% (see Table 4.6) and has an area under the ROC curve (AUROC) of about 0.72 (see Figure 4.1).

Table 4.6: Multivariate Logistic Model

Variable	Domestic SMEs			International SMEs			Sig.
	Coefficient	Z	p-value	Coefficient	Z	p-value	
STDEBV	----	----	----	.0034541	3.52	0.000	----
RETA	-.4883401	-11.76	0.000	-.4757847	-11.78	0.000	0.8260
CTA	-1.702634	-18.14	0.000	-1.120779	-12.02	0.000	0.0001
EBITDAIE	-2.76e-07	-4.06	0.000	-3.15e-07	-5.91	0.000	0.6628
CETL	-.0713958	-7.40	0.000	-.0176677	-3.46	0.001	0.0055
lnCR	-.1041718	-5.30	0.000	-.1481872	-8.48	0.000	0.0767
TDTA	.3543306	6.25	0.000	.2826852	4.90	0.000	0.3865
TTA	-3.045346	-8.30	0.000	-4.47203	-11.93	0.000	0.0102
TCTL	.3559783	6.25	0.000	.6236815	10.99	0.000	0.0005
IATA	.3328435	2.61	0.009	.3898429	3.24	0.001	0.7397
INDWOE	-.6180949	-22.45	0.000	-.586922	-22.28	0.000	0.4418
Insolvency Rate	2.708748	22.80	0.000	2.68772	24.20	0.000	0.8982
Size (log)	.5170512	7.37	0.000	.507237	7.02	0.000	----
Size Squared (log)	-.02308	-9.04	0.000	-.0214534	-8.35	0.000	----
Constant	-9.022245	-18.17	0.000	-9.333734	-17.91	0.000	----
Goodness of Fit Tests	Value			Value			
Pseudo R^2	0.0674		----	0.0515		----	
Log Likelihood	-26623.72		0.0000	-29225.091		0.0000	
Number of Observations	253171			254788			

This table shows the multivariate logistic models developed for domestic and international SMEs. The first column lists the variables studied, second and fourth columns report the coefficients, third and sixth columns report the Z statistics, while fourth and seventh columns report the statistical significance of the variables in the respective multivariate models. The last column report the p-value obtained by comparing the regression coefficients of domestic and international credit risk models using "suest" command in STATA 12. However, the last four rows of this table reports goodness of fit measures and number of observations.

Figure 4.1: Receiver Operating Characteristic Curves for Domestic SMEs

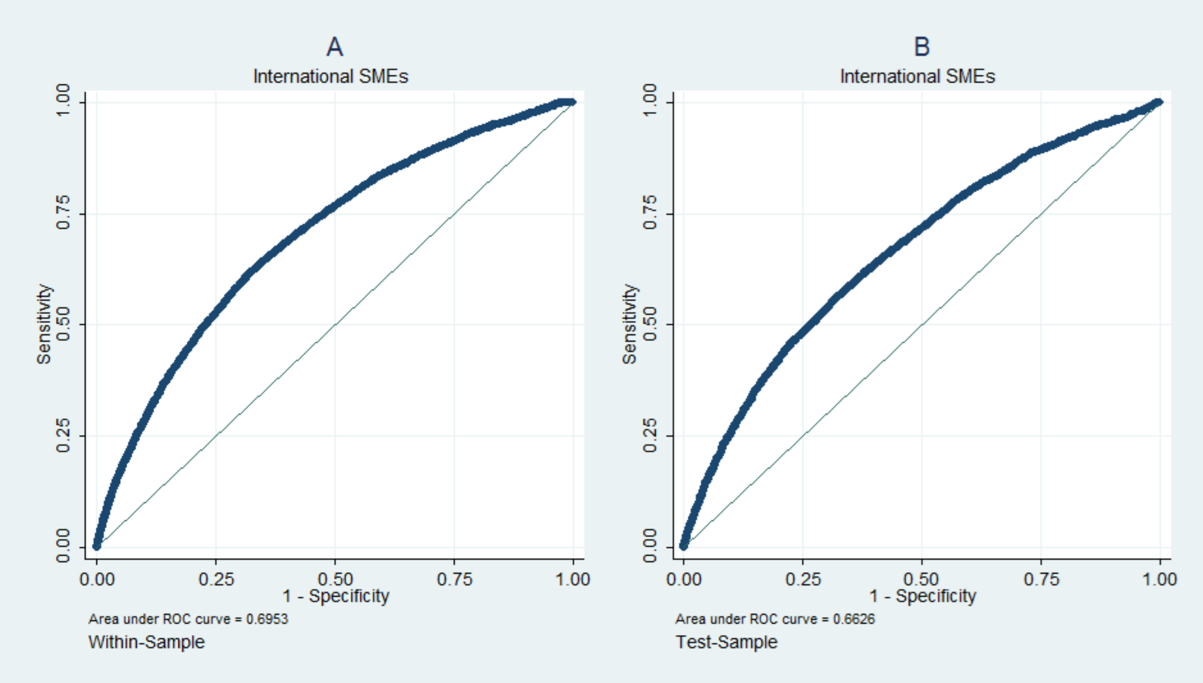


This figure shows Receiver Operating Characteristics Curves for (A) within-sample and (B) test-sample model performance estimated using the sample of domestic UK SMEs. The area under the ROC curve (AUROC) is equal to the probability that the rating for a true positive (a firm actually defaults and the model has classified it as expected default) will be less than that for a true negative (a firm does not default and the model has classified it as expected non-default) plus 50 per cent of the probability that the two ratings will be equal.

4.3.3.2 Default Prediction Model for International SMEs

I estimate the model for international SMEs using logistic regression with failed = 1 and non-failed = 0 as the dependent variable. To estimate this model I employ the same approach as used in building the model for domestic SMEs. Again, I omit the covariates EBITDATA, QACA and TCTA as they are highly correlated with other covariates. I also omit all the export intensity measures except EIM, as they are highly insignificant in the prior univariate analysis. The stepwise estimation also prompt me to eliminate EIM, STA and SWC from the multivariate model. The final model for international SMEs is estimated using ten highly significant financial ratios with the expected sign of their respective coefficients. I report the estimated model in Table 4.6. The in-sample overall classification accuracy of the model is about 61% (see Table 4.7) and has an area under the ROC curve (AUROC) of about 0.7 (see Figure 4.2).

Figure 4.2: Receiver Operating Characteristic Curves for International SMEs



This figure shows Receiver Operating Characteristics Curves for (A) within-sample and (B) test-sample model performance estimated using the sample of international UK SMEs. The area under the ROC curve (AUROC) is equal to the probability that the rating for a true positive (a firm actually defaults and the model has classified it as expected default) will be less than that for a true negative (a firm does not default and the model has classified it as expected non-default) plus 50 per cent of the probability that the two ratings will be equal.

4.3.4 Model Validation

I conduct validation tests to examine the prediction ability of models developed on different set of companies. I use out-of-sample validation, which is the most widely used model validation technique in bankruptcy studies (see among others Altman et al., 2010; Gupta et al., 2014a). For domestic SMEs my hold-out sample has 2,366 failed and 87,174 non-failed SMEs, while for international SMEs there are 2,393 failed and 87,024 non-failed firms for the analysis period 2008 and 2009. The hold-out sample contains a total of 89,540 domestic and 89,417 international firms.

Table 4.7 reports the validation test results on the hold-out sample. The type I and type II error rates shown in the table are calculated using the sample bad-rate as the cut-off rate and the error rates for domestic firms are lower than those for international firms. Hence, I can argue that my data provides a better fit for domestic SMEs than their international counterparts.

Table 4.7: Misclassification Rates and Accuracy Performance of Models Developed

		Type I Error Rate	Type II Error Rate	Overall Classification Accuracy	Area Under ROC Curve
Domestic SMEs	Within	31.71%	35.23%	64.85%	0.7226
	Hold-out	34.83%	36.44%	63.60%	0.6919
International SMEs	Within	32.07%	39.19%	61.00%	0.6953
	Hold-out	37.57%	38.97%	61.06%	0.6626

This table shows the misclassification rates and classification accuracy of the models developed using the sample bad-rate as cut-off rate for development and hold-out samples. The third column report the Type I error rate which measure the percentage of failed firms classified as non-failed. The fourth column reports the Type II error rate which measure the percentage of non-failed firms classified as failed. The average accuracy of the models developed calculated as 1 minus average of the error rates is reported in the fifth column. The sixth column shows the area under the ROC curve (AUROC) which is equal to the probability that the rating for a true positive (a firm actually defaults and the model has classified it as expected default) will be less than that for a true negative (a firm does not default and the model has classified it as expected non-default) plus 50 per cent of the probability that the two ratings will be equal.

The fifth and sixth columns in Table 4.7 enable me to compare my results. The fifth column reports the overall classification accuracy of the respective models in correctly identifying failed and non-failed firms, which I measure as the complement of the weighted average of Type I and Type II error rates. The sixth column shows the area under the ROC curve (AUROC) which is equal to the probability that the rating for a true positive will be less than that for a true negative plus 50 per cent of the probability that the two ratings will be equal (Anderson 2007). AUROC of 50% implies a random (uninformed) model, whereas AUROC of 100% implies a perfect model. But, a credit scoring model would give something in between the perfect and the uninformative model. Hence, the higher the area under the ROC curve, the better the model's classification performance, assuming that the information may be misleading if the class distributions are skewed.

The overall classification accuracy of the model developed for domestic SMEs is about 64%, while that for international SMEs is about 61%. I also see that the AUROC for domestic SMEs is higher than for international SMEs by about 3 points (see Figure 4.1 and Figure 4.2).

Hence I can say that, although the same set of explanatory variables except STDEBV are significant in explaining the credit risk behaviour for both the groups, the financial ratios perform as better predictors for domestic SMEs than their international counterparts. This supports Arslan and Karan (2009)'s suggestion of considering domestic and international SMEs separately while modelling their credit risk behaviour. However, given the classification performance of the models developed, one may not gain significantly by treating domestic and international SMEs separately while modelling their default risk.

4.3.5 Comparison of Regression Coefficients

I find that an identical set of explanatory variables are significant in explaining the credit risk behaviour of both domestic and international SMEs. However, they lead to slightly different classification accuracy measures for each group. It is possible the impact of covariates may vary between the groups. To determine if this is the case, I performed a chi-square test of each financial ratio in turn to compare[18] the regression coefficients of the models developed for domestic and international SMEs. As reported in the last column of Table 4.6, my test results show that out of the nine common predictors four predictors (CTA, CETL, TTA and TCTL) exhibit significant statistical difference in their weights between domestic and international SMEs. Hence I have some motivation to believe that domestic and international SMEs need to be considered separately while developing bankruptcy models for them.

4.4 Conclusion

In this study, I investigate the effect of internationalisation on modelling credit risk for UK SMEs. Following Fatemi (1988) I classify a firm as international if it makes sales abroad and domestic if it makes sales only in the domestic market. The empirical literature on the

[18] Comparison of regression coefficients is done using "suest" command in Stata 12.

performance of international SMEs is somewhat contradictory, which motivate me to undertake this study. Ramaswamy (1992) reports that international SMEs exhibit lower risk due to revenue and cash flow diversification, while Michael et al. (2009) report that international SMEs exhibit higher default probability due to exposure to multiple political and financial environments. To examine the impact of internationalisation on the default propensity of SMEs, I estimate separate default prediction models for domestic and international firms using a set of financial ratio.

I develop one-year distress prediction models using a dynamic logistic regression technique, and implement appropriate measures to control for the effect of macroeconomic conditions. The unique database available to me from the Credit Management Research Centre of the University of Leeds contains financial information of 342,711 domestic SMEs (with 8,525 defaulted and 334,186 non-defaulted firms) and 344,205 international SMEs (with 9,114 defaulted and 335,091 non-defaulted firms) ranging over an analysis period of 2000 to 2009. I retain the data of analysis year 2008 and 2009 as a hold-out sample.

My empirical findings are somewhat mixed. In my multivariate models, all the factors which affect the default probability of international SMEs are also highly significant in explaining the default probability of domestic SMEs, except short-term debt/equity book value. Furthermore, all the variables capturing the impact of exports on default probability of international firms are highly insignificant in the univariate analysis, thus contradicting the suggestion of Arslan and Karan (2009) to consider domestic and international firms separately while modelling their credit risk behaviour. However, the predictive accuracy measures obtained by employing the same set of variables are lower for international SMEs than for their domestic counterparts. Chi-square tests performed to compare the weights of regression coefficients of the models developed, confirm that the coefficients of four out of the nine common predictors (CTA, CETL, TTA and TCTL) exhibit significant statistical difference. I make a further significant contribution by being the only study to measure the impact of intangible assets on the defaults probability of SMEs. My test results confirm that the ratio intangible assets/total assets (IATA) is highly significant in assessing credit risk for both domestic and international SMEs.

My findings clearly show that almost the same set of factors affect the default probability of both the groups, hence there is no potential need to treat domestic and international SMEs separately while modelling credit risk. This indifference may be due to the recent effort undertaken by the policy makers and business community to understand and mitigate the factors adversely affecting the export performance of small firms (Secretariat 2009).

However, in view of the low predictive power of the model developed for international SMEs, I suggest that modelling credit risk for international SMEs would benefit from further work to understand the inherent complexities. Non-financial factors may play an important role in understanding their credit risk behaviour. In particular the effect of changing government policies, firm specific non-financial characteristics, and changing macroeconomic conditions may play an important role in understanding their credit risk behaviour. These are possible avenues for further research in the field of modelling credit risk behaviour of SMEs.

5. Forecasting Bankruptcy for SMEs Using Hazard Function: To What Extent Does Size Matter?

5.1 Introduction

Small and medium-sized enterprises (SMEs) are widely considered to be the backbone of the global economy, and are viewed as an important route to recovery in the aftermath of the global financial crisis of 2008-2009. The contribution of SMEs to economic performance varies substantially across nations: wealthier nations typically have large organized SME sectors, and fewer informal business sectors, than poorer nations. However, the informal sector results in a higher level of economic activity and there is considerable overlap between both of these sectors (Batini et al. 2010). SMEs are often referred to job creation engines; a 2009 Economist Intelligence Unit study reveals that they continued to generate job opportunities throughout the economic slowdown (Economist Intelligence Unit 2009), and along with entrepreneurs they are viewed as key drivers of economic development (Bosma and Levie 2010). Koshy and Prasad (2007) report that, the SME sector globally acts as an important partner in eradication of poverty. In consequence of this realization of their economic importance; there has been a significant growth in output of academic literature pertaining to SMEs in the last decade.

The lending community globally, shares the consensus that the SME sector is a profitable market segment, and are developing ways to unlock its potential with specific focus on the problem of high credit risk and service cost (IFC 2010). The empirical literature on modelling credit risk for large enterprises is extensive and gravitates toward Altman's Z-Score model

(Altman 1968), which use historical accounting information to predict bankruptcy and alternative approaches which use securities market information to predict financial distress (Merton 1974). Seminal contributions came from Beaver (1966), and Altman (1968), who use univariate and multivariate models respectively, to develop distress prediction models using accounting information. The advancement in methodology and technology since their studies has resulted in a substantial increase in the number and complexity of bankruptcy prediction models. However, recent empirical literature has improved our understanding of credit risk behaviour of small firms. Using multivariate discriminant analysis, Edmister (1972) was the first to develop a default prediction model specifically for SMEs. Recently, Altman and Sabato (2007) suggest the requirement for a separate default prediction model for SMEs. They use logistic regression techniques on a sample of US SMEs, and report that their distress prediction model performs better than generic credit scoring models and leads to slightly lower capital requirements for banks. However, they acknowledge that their model's performance could be improved by inclusion of qualitative information. The empirical literature also reports that qualitative information, such as industrial sector, business type, age, location, auditors' opinion etc. are significant in explaining firm's credit risk (see among others Lehmann 2003; Grunert et al. 2005; Tsai et al. 2009). Altman et al. (2010) took account of such non-financial characteristics as well as compliance information while developing distress prediction models from a sample of about 5.8 million UK SMEs. They report about 13% improvement in their model's performance when significant qualitative information is incorporated along with traditional financial ratios. Recent literature also explores the significance of operating cash flow information in predicting financial distress of SMEs. Gupta et al. (2012a) suggest that lending institutions and policymakers may not gain better understanding of the credit risk behaviour of SMEs by exploring cash flow statements in addition to income statements and balance sheets, as they find no improvement in their models performance by inclusion of significant operating cash flow ratios. Gupta et al. (2012b), separately modelled the credit risk behaviour of both domestic and international SMEs. They report that virtually the same set of predictor variables exhibits significant discriminatory power for both domestic and international groups, suggesting no need for separate credit risk models for each group. However, the lower predictive power of the model for international SMEs, led them to compare the regression coefficients of the two models. Test results confirm significant statistical differences in the weights of four out of nine common predictors. Thus, they suggest that considering domestic and international SMEs separately while modelling credit risk for them may lead to better lending decision.

There exists huge diversity within the broad SMEs category (which contains micro, small and medium-sized firms) in terms of access to finance (Beck et al. 2006), management style (Wager 1998), number of employees etc. However, the empirical literature pertaining to understanding of the credit risk behaviour of SMEs does not take into account this diversity. I address this gap in the literature by distinguishing among micro, small, and medium firms while developing my distress prediction model for SMEs. Using a set of financial and non-financial information, I apply a discrete time duration-dependent hazard rate modelling technique to develop separate bankruptcy prediction models for each of the three categories, and compare their performance with a single model encompassing all three categories[19]. Finally, to test the prediction performance of the models developed, I report receiver operating characteristics (ROC) curves, and related summary statistics, bearing in mind the concerns of the Basel Committee (2000) (Sahajwala and Van den Bergh 2000) on model validation.

My statistical analysis is performed using a heterogeneous-panel available to me from the Credit Management Research Center of the University of Leeds. The database (with non-

[19] It includes all micro, small and medium firms.

missing data) contains financial and non-financial information of 8,162 failed and 385,733 non-failed UK SMEs[20], covering recent analysis periods between 2000 and 2009. I use the available financial and non-financial information of firms reported between the analysis years 2000 to 2007 as my development sample and develop separate failure propensity models for SMEs, micro, small, and medium firms. I retain the corresponding data of years 2008 and 2009 as a hold-out sample to validate the out-of-sample prediction performance of the models developed.

My empirical findings show that, all the multivariate models developed exhibit fairly strong prediction performance[21] with area under ROC curve (AUROC) of about 0.74 for micro firms, 0.77 for small firms and 0.76 for medium and SMEs. Further comparison of insolvency hazard models for micro firms and SMEs strongly suggest that these categories should be considered separately for credit risk modeling purposes, as I see wide variation in the factors affecting their default risk. As almost the same sets of explanatory variables affect the default probability of both small firms and SMEs, I do not expect material impact on the decision making process, by treating each group separately. Finally, I compare the models developed for medium firms and SMEs. Once again, almost the same set of explanatory variables is highly significant in explaining the default probability of both models. My test results support the hypothesis that the credit risk characteristics of firms within the broad SMEs category do vary, and hence I suggest micro firms need to be treated separately while modeling credit risk for them.

In the United States micro firms comprise about 78% of the total employer firms (Alsaaty 2013), whilst in the United Kingdom, about 96% of all businesses are micro firms that employ less than 10 employees (Rhodes 2012). Thus, my findings impact the vast majority of

[20] Which we further classify into micro, small and medium firms as listed in Table 5.1.
[21] AUROC calculated using hold-out sample.

the firms within the SMEs segment (i.e. micro firms), even though I do not report differences in credit risk attributes between small sized firms, medium sized firms and SMEs. My results are of particular interest to commercial lending institutions, which primarily use credit scoring based techniques to make their lending decisions. By not treating micro firms separately from the SMEs, they attempt to predict a heterogeneous group of financially distressed firms from the database. This could produce biased estimates of firms' lifetimes and insolvency hazard rates. I believe that my study can lead them to change the way in which they model the data, leading to more robust econometric estimates.

The remainder of the chapter is structured as follows. Section 5.2 presents an overview of the existing related literature and the motivation for the study. Section 5.3 describes the methodology used to conduct my empirical analysis. The empirical findings and supporting discussion are reported in Section 5.4. Finally Section 5.5 presents my conclusions.

5.2 LITERATURE REVIEW AND MOTIVATION FOR THE STUDY

A growing body of empirical literature reports that SMEs play a major role in macroeconomic growth (Ayyagari et al. 2007) and hence small-business lending has received considerable attention from practitioners and researchers in the last decade. To date, there has been no single agreed definition of small and medium-sized enterprises (SMEs)[22]. A number of key variables, for example; independence, legal status, number of employees, industrial sector, employment, capital investment, and asset size are considered in most working definitions. A widely accepted and useful set of definitions are those used by the European

[22] UK companies are required to file accounts at 'Companies House' (www.companieshouse.gov.uk) which defines a small company as one for which at least two of the following conditions are met: (i) Annual turnover is £6.5 million or less; (ii) the balance sheet total is £3.26 million or less; (iii) the average number of employees is 50 or fewer. It defines medium company as one for which at least two of the following conditions are met: (i) Annual turnover must be no more than £25.9 million; (ii) the balance sheet total must be no more than £12.9 million; (iii) the average number of employees must be no more than 250.

Union. They defines a firm as 'micro' if it has less than 10 employees and an annual turnover of under € 2 million; 'small' if it has less than 50 employees with an annual turnover of less than € 10 million, and 'medium' if it has less than 250 employees with an annual turnover of less than € 50 million. I partially adopt these definitions for the purposes of the present work.

5.2.1 Diversity between Micro, Small and Medium Firms

In recent years, much empirical literature pertaining to understanding the determinants of capital structure and debt policy of SMEs has been published (see among others, Sogorb-Mira 2005; Watson and Wilson 2003). Few of the studies distinguish among micro, small and medium firms however. Thus, the wide diversity within the overall SME category is ignored. However, the determinants of capital structure choices may have different effect on different groups. Ramalho and Da Silva (2009) take account of this diversity and report differences in the determinants of financial leverage decisions among micro, small, medium and large firms. Recently, Mateev et al. (2013) report significant differences in the way micro, small and medium-sized firms make their capital structure choices. They state that micro and small firms are primarily dependent on short-term loans and trade credits for their external sources of finance, while long term bank loans are the preferred route of external financing for medium-sized firms. Beck et al. (2006) report that the probability that a firm rates financing as a major obstacle toward its growth is 39% for small, 38% for medium and 29% for large firms respectively, showing that the larger the size of the firm, the less access to finance is seen as a problem. Besides differences in capital structure and financial constraints, Beck et al. (2005) report that the extent to which the corruption of bank officials, and financial and legal issues, constrain a firm's growth is highly dependent on its size. The smallest firms are the most adversely affected by these constraints. Studies also report differences within the SME segment from personnel management and innovation dimensions. Using data of micro, small and medium-sized firms in Australia, Kotey and Slade (2005) report that the rate of adaptation of formal personnel management practices increases with firm size. Their findings suggest a move toward division of labour, increased documentation, hierarchical structures, and increasing administrative processes, with increasing number of employees. They suggest taking account of diversity across various size categories, while providing advice and management training to SMEs. De Mel et al. (2009) find that more than one quarter of micro firms are engaging in innovation, mostly marketing innovation, and firm size has a stronger positive relationship with processes and organizational innovations, than with product innovations.

Although few studies make a distinction between micro, small, and medium firms, however considering the diversity that exist within the broad SMEs segment, I expect some factors leading to the failure of micro, small, and medium firms may differ, or the common factors may have varying degrees of influence, within this category. Empirical literature reports a negative relationship between firm size and default probability (Pettit and Singer 1985), since large firms are more diversified and have more stable cash flow than small firms (Gill *et al.* 2009). Thus, I examine the credit risk behaviour of the three groups separately, and compare their performance with the inclusive SMEs group as a whole.

5.2.2 SME Failure

SME bankruptcies have always been difficult to track and measure, as failed businesses are often difficult to locate and if located it’s again difficult to determine the reason for their failure. However, recent literature (see among others Headd 2003; Carter and Auken 2006) has focused on understanding the rate and cause of such failures. Carter and Auken (2006) report that the principal reasons for firm failure can be categorized into, lack of knowledge, constraints to debt financing, and the economic climate. Besides the direct costs, the bankruptcy of small firms also causes indirect costs such as loss of personal collateral, self-esteem, self-employment etc. to the owners. A growing body of empirical literature suggests that financial constraints are the strongest reason for small business failures (e.g. Hutchinson and Xavier 2006). Some recent studies also highlight poor management skills as a potential factor for small firm's failure (Peacock 2000).

Knott and Posen (2005) argue that, though the failure of new firms is considered to be wasteful, it enhances social welfare and reduce industry cost. Hence all business failures are not purely due to financial difficulties. Empirical studies (see among others J. Watson and Everett 1996; Headd 2003; Bates 2005) suggest many "business failures" involve planned exit strategies, with the business actually being healthy enough to continue operation. The decision to discontinue business operations may be due to several reasons such as, change of ownership, opportunity cost, limiting losses, non-economic considerations, switching cost, etc. Sometimes the decision is made to close a successful business, thus a careful distinction needs to be drawn between failures due to purely financial difficulties, and firm's closure due to some strategic gain. To improve the quality of my analysis I take into account only those small firms where business failure is due purely to financial difficulties[23] and I exclude other form of business closure.

Altman et al. (2010) report that there exists a non-linear relationship between firm size (measured by assets) and insolvency risk, with insolvency risk being an increasing and decreasing function of firm size (see Figure 1 in Altman et al. 2010). They argue that companies with lower asset size are less likely to be chased by creditors for bankruptcy proceedings, as the creditors are unlikely to benefit (as hardly any assets will be left to recover debt). However, as asset size increases, bankruptcy proceedings become more

[23] Once a firm has become insolvent, the UK Act provides to choose one from the five courses of action: administration, company voluntary arrangement (CVA), receivership, liquidation and dissolution. In this study to represent the failed sample group we take under consideration only those SMEs whose failure followed any of the three common routes, i.e. administration, receivership or liquidation.

attractive to creditors. Thus, insolvency risk increases with increasing asset size, but after reaching a threshold level, it starts to decline with further increase in firm's asset size. In consideration of this behaviour of SMEs insolvency risk with different assets size segments, I partitioned my sample into micro, small, and medium firms, while developing my insolvency hazard models. Since the insolvency rate changes with assets size, I expected the factors affecting default probability to vary across these three size segments. I believe myself to be the first to make a distinction between micro, small, and medium firms while modelling credit risk for SMEs. However, a recent study by Holmes et al. (2010) on the survival of SMEs resembles my work in a few respects. They study a sample of 781 manufacturing SMEs in north-east England between 1973 and 2001 using hazard function technique. They also distinguish between 'micro firms' and 'small and medium firms', and report that each is differently affected by macro-economic and firm specific factors. However, the study of Holmes et al. (2010) differs from my work in several important respects; their data covers only 781 firms and they did not use any financial information in their analysis. Their sample is limited to one industrial sector, namely manufacturing, which accounts for only around 12% of firms in the UK. Moreover, their sample is highly concentrated in geographic location. Finally, their sample covers a too wide and back dated sampling period. My study by contrast uses both financial and non-financial data from 393, 865 firms from all industries and regions, for the more recent period 2000 – 2009.

5.3 Empirical Methods

In this section I discuss the following: (a) the dataset, (b) selection of explanatory variables and (c) statistical models applied in my analysis.

5.3.1 DATASET

I perform the statistical analysis on a unique heterogeneous panel-data available to me from the Credit Management Research Center of the University of Leeds. The sample (with non-missing data) contains financial and non-financial information of 8,162 failed and 385,733 non-failed UK firms covering recent analysis periods between 2000 and 2009. Further details pertaining to the sub-samples are reported in Table 5.1. Using a discrete time duration-dependent hazard modeling technique, I develop default prediction models by employing available financial and non-financial information between the analysis years 2000 to 2007. To validate out-of-sample prediction performance of the models developed, I retain the data of analysis year 2008 and 2009 as a hold-out sample. For surviving firms, I use the financial and non-financial information filed in the previous year to conduct my analysis, and for failed firms I use the last set of information reported before failure.

Table 5.1: Dataset of UK SMEs

Firm Category	Failed	Non-failed	Total	Failed/Total %
Micro	1489	83447	84936	1.75
Small	3251	150648	153899	2.11
Medium	3422	151638	155060	2.21
SME	8162	385733	393895	2.10

This table shows the sub-classification of my database among micro, small and medium firms, containing firm level information of UK firms from analysis year 2000 to 2009.

In this study, I partially follow the definition of the European Union while separating my sample into sub-samples of micro, small, and medium firms. Specifically, I define a firm as ‘micro’ if it has less than 10 employees; ‘small’ if it has greater than or equal to 10 but less than 50 employees; ‘medium’ if it has greater than or equal to 50 but less than 250 employees. The overall encompassing definition of SME includes all firms with less than 250 employees. The SMEs model is developed using the full sample of all micro, small, and medium firms. In terms of the UK Insolvency Act 1986, ‘a company is said to be insolvent if it either does not have enough assets to cover its debts (i.e. the value of assets is less than the amount of its liabilities), or it is unable to pay its debts as they fall due'. The Act provides for an insolvent firm to choose between five defined courses of action: administration, company voluntary arrangement (CVA), receivership, liquidation, and dissolution. In line with prior bankruptcy studies (e.g. Altman et al. 2010) I accept corporate failure as a firm entering into liquidation, administration, or receivership, and exclude finance, insurance, and utility firms from my sample.

5.3.2 Selection of Predictor Variables

The database available to me contains a wide range of both financial and non-financial information of UK SMEs. In my study I consider only those firms which report both summary financial statements (balance sheet and income statement), and I analyse only those financial ratios which can be obtained from these two financial statements. The selection of the financial ratios is such that it captures a firm's performance in the dimensions of asset utilization, solvency, liquidity, and debt coverage. I include most of the financial ratios found successful in prior bankruptcy prediction studies. Specifically, I follow Altman et al. (2010), as their selection of explanatory variables is non-overlapping with strong theoretical underpinning. In consideration of the findings of Jones (2011), I also investigate the effect of intangible assets on a firm's default probability. He reports that higher proportions of intangible assets in a firm's capital structure signal a higher default probability, due to firms approaching bankruptcy accumulating intangible assets more aggressively than non-failed ones. I also analyse some of the non-financial data reported in Altman et al. (2010) and test their statistical significance across the broad SMEs category. Finally, to select among the wide range of competing explanatory variables, I calculate their variance inflation factor (VIF), and variables having VIF value of less than 10 have been selected for this study. Table 5.2 presents a list of the variables included in my analysis along with their respective definition.

Table 5.2: Table of Independent Variables

Variable Name	Variable Definition
EBITDATA	Earnings Before Interest Taxes Depreciation and Amortization / Total Assets
STDEBV	Short Term Debt / Equity Book Value
RETA	Retained Earnings / Total Assets
CTA	Cash / Total Assets
EBITDAIE	Earnings Before Interest Taxes Depreciation and Amortization / Interest Expense
CETL	Capital Employed / Total Liabilities
QACA	Quick Assets / Current Assets
lnCR	log (current assets / current liabilities)
TCTL	Trade Creditors / Total Liabilities
TDTA	Trade Debtors / Total Assets
TTA	Taxes / Total Assets
TCTA	Trade Creditors / Total Assets
STA	Stock / Total Assets
IATA	Intangible Assets / Total Assets
AUDIT	Financial Statements Audited (Dummy Variable; Yes = 1 & 0 otherwise)
LFD	Financial statements filed late (Dummy Variable; Yes = 1 & 0 otherwise)
LLF	Log (Number of days late in filing financial reports)
CFS	Cash Flow Statement reported (Dummy Variable; Yes = 1 & 0 otherwise)
CCJ Number	Number of county court judgements (CCJ) pending
CCJ Amount	Outstanding CCJ amount

This table list the predictor variables studied (for each predictor the variable name along with the respective definition is presented).

I expect the leverage ratios STDEBV, to have positive relationship with the default probability as higher value of STDEBV represent higher value of debt per unit of capital employed and signal higher probability of failure. On the contrary, a firm in financial distress is expected to have higher value of liabilities and hence the leverage ratio CETL is expected to exhibit negative relationship with its default probability. The cumulative profitability of a firm and its capacity to retain earnings from its current income is measured by the profitability ratio RETA. A firm approaching financial distress exhibits a declining trend of retained earnings, which ultimately leads to lower value of RETA and higher default probability. Similarly, a healthy firm is expected to have higher value of EBITDATA, CTA and EBITDAIE than a distressed entity and hence lower default probability. Hudson (1986) reports that trade creditors form a significant proportion of small firm’s liabilities and are the primary cause of their bankruptcy. My selection of explanatory variables also takes into account the impact of the short-term liquidity position of firms on their survival. Many firms fail due to lack of sufficient liquid assets and sustainable level of working capital. Healthier firm are expected to have better liquidity position than distressed firms and hence higher value of QACA and lnCR but lower value of TCTL, TDTA, TCTA and STA. I expect TTA to have negative relationship with default probability as healthy firms with good liquidity position are not expected to default on their financial obligations. To capture the influence of intangible assets on firm's default probability, I calculate intangible assets/total assets (IATA) and expect it to have a negative relationship with default probability, as firms approaching distress capitalize intangibles more aggressively than their healthy counterparts (Jones 2011).

The potential power of adding qualitative information in modelling credit risk for SMEs is explicitly highlighted in Altman et al. (2010). Thus I also include qualitative information such as firms' audit information, compliance information, firm's age, accounts filing history, legal action taken by lenders to recover debt, and other firm specific characteristics in my analysis. A creditor may take court action to recover debts by way of county court judgement (CCJ). If a CCJ is issued against a debtor, it means that the court has formally decided that the debtor owes money to the creditor and orders settlement of the debt. Failure to settle debts leads to an accumulation of CCJ's. Records are kept for six years, which makes accessing external credit hard for the defaulters. Hence accumulation of CCJ's signals poor financial health, and is expected to be positively related to default probability. CJJ's as a potential insolvency indicator may not work for companies having significant market power, as they may use their bargaining power to make slow payments despite having sound financial health forcing creditors to make CCJ applications (Altman *et al.* 2010). To exploit its discriminatory power I study the influence of both the number of CJJ's (CJJ Number) and outstanding CCJ amounts (CCJ Amount) on insolvency risk across the broad SME category. The next type of information that I consider relates to timely filings of financial reports to the regulatory authority. A private company needs to file its accounts within ten months of the end of the relevant accounting reference period, or else it incurs penalties for late filing. Late filing days in the UK are calculated as the number of days following the ten month period till the accounts is filed. The late filing of accounts may be due to deliberate attempt by firms' directors to delay disclosure of unfavourable information, or disagreement between auditors and directors regarding the firm's "true" financial position. However, in both these cases late filing of accounts acts as an indicator of financial distress, and hence to capture any inherent discriminatory power I use the log[24] of the number of days by which a firm is late as an explanatory variable. To account for the reliability of financial information made available by the firm, I employ a dummy variable. I include the variable AUDIT which takes the value of 1 if the firm has been audited and 0 otherwise. Small companies with annual turnover of less than £ 6.5 million and assets of less than £ 3.26 million[25] qualify for a total audit exemption. Hence, accounts which are non-audited primarily relate to small companies and are expected to be less reliable than their audited counterparts. Considering this information asymmetry I expect non-audited firms to exhibit higher insolvency risk than audited firms. I expect firms which submit cash flow statements to be more transparent and hence less risky. I exploit this information as a dummy variable (CFS) which takes the value 1 if a cash flow statement is provided and 0 otherwise. I expect the dummy variable CFS to be a significant insolvency indicator of small and medium firms only, as micro companies rarely submit full sets of accounts. Further, following Altman et al. (2010), I control[26] for the size effect by using a quadratic term in total asset values [(natural logarithm of (1 + total assets) and (natural logarithm of (1 + total assets))^2], to allow for their reported non-linear relationship between

[24] We take 'log' to capture any non-linear relationship.
[25] Source: https://www.gov.uk/audit-exemptions-for-private-limited-companies
[26] We apply size control only for SMEs sample.

insolvency rate and firm's asset size (insolvency risk follow an increasing and decreasing function of firm size).

5.3.3 Discrete Hazard Model

The use and estimation of default probability across various lending sectors has gained higher importance since the introduction of the Basel II capital accord, requiring the banks to maintain risk-based capital reserves. In the empirical literature, multiple discriminant analysis (MDA) and logistic regression are the traditionally preferred statistical techniques for estimating default probability. A majority of the distress prediction methodologies that I see in recent empirical studies gravitate towards Altman (1968), Ohlson (1980), Zmijewski (1984) and more recently Shumway (2001). Altman (1968) was the earliest to apply MDA technique to make default predictions of US manufacturing firms using a set of financial ratios by calculating the celebrated Z-Score. Thereafter MDA remain the most widely used statistical technique for bankruptcy prediction studies, until Ohlson (1980) challenged its restrictive assumptions[27] and for the first time employed conditional logit regression techniques in bankruptcy prediction studies. Since this pioneering work of Ohlson (1980), substantial volume of empirical literature using logit regression technique for bankruptcy prediction studies with one firm-year observation for each firm available has appeared (e.g. Altman et al. 2010; Gupta et al. 2012a). In such single-period models, single firm-year observations for each non-failed firm are randomly selected from the available firm-year database. Whereas, the (non-random) firm-year information immediately preceding the bankruptcy filing year is selected (Hillegeist *et al.* 2004) for *failed* firms? The ordinary single-period logistic (logit) regression has the following form; P_i is the probability of default of firm i, α is a constant, X is the vector of covariates and β is the vector of coefficients:

[27] The two restrictive assumptions of MDA analysis are: i) the independent variables included in the model are multivariate normally distributed; ii) the group dispersion matrices (or variance-covariance matrices) are equal across the failing and the non-failing group. See Barnes (1982) and Karels and Prakash (1987) for further discussions about this topic.

$$P_i = \frac{e^{\alpha + X_i\beta}}{1+e^{\alpha + X_i\beta}} \quad (1)$$

However equation (1) will result in understated values of standard errors (Beck et al. 1998), and possible sample selection bias, along with failure to capture time-varying changes (Hillegeist *et al.* 2004). Considering the misspecification between the covariates and the logit of predicted bankruptcy probabilities, Hwang et al. (2007) suggest a robust semi-parametric logit model with smaller hold-out sample error rates. Recently, in a study using cross-sectional sample of German corporate credit defaults, Kukuk and Rönnberg (2013) extended the popular logit model that allows for varying stochastic parameters (mixed logit) and non-linearity of covariates. Alternatively, Shumway (2001) proposed a dynamic approach to measuring default probability, unlike the single-period classification models of Altman (1968), Altman et al. (2010) etc., which Shumway calls *static approaches*. Shumway argues that static models ignore the fact that firm characteristics changes over time, hence the default probabilities estimated are biased and show poor out-of-sample performance. He introduced time-varying covariates, and suggests that default prediction models should be specified as duration dependent models with time-varying covariates. By improving on Shumway's (2001) suggestion, Chava and Jarrow (2004) report superior forecasting performance of Shumway's (2001) model compared to popular static models. More recently, Hwang (2012) use discrete-time duration-dependent hazard rate modelling techniques and report superior performance over the discrete-time hazard model without a time-varying specification.

The conditional probability of discrete time hazard function (λ) for firm *i* to default in the time interval *t*, given it survives up to this time interval, is stated as follows:

$$\lambda_t\left(t \middle| X_{i,t}\right) = \Pr(T = t | T \geq t, X_{i,t}) \quad (2)$$

In equation (2), T is discrete failure time; T=t states failure within the time interval t and $X_{i,t}$ is the value of covariates of firm i up to time interval t. Recently Nam et al. (2008) suggest that the duration dependent hazard model;

$$h(t|x_{i,t}) = h(t|0).\exp\{x'_{i,t}\beta\} \tag{3}$$

where, $h(t|x_{i,t})$ represent the individual hazard rate of firm i at time t and $x_{i,t}$ is the vector of covariates of each company i at time t; can be estimated, and stated in the form of a panel logistic model that incorporate macro-dependent base line hazard. The dependent variable in the hazard rate model is the time that a firm spends within the healthy group (time spent before it's failure) and the moment a firm leaves the healthy group due to reason other than failure (e.g. acquisition or merger) it is considered to be censored (no longer observed). The default risk may be a function of the latest financial information, macroeconomic variables, and firm's age, which change as the values of explanatory variables changes with time.

Following the suggestions of Hwang (2012), Beck et al. (1998), Shumway (2001) and Nam et al. (2008), I use the discrete-time duration-dependent hazard model to overcome the econometric limitations discussed above. The discrete hazard modelling technique is well suited to analyse data that consists of binary dependent variables and exhibit both time-series and cross-sectional characteristics, such as bankruptcy data. It can be estimated analogous to the logit model having the following form, where *α(t)* is the time-varying covariate introduced to captures the baseline hazard rate.

$$P_{i,t} = \frac{e^{\alpha(t)+X_{i,t}\beta}}{1+e^{\alpha(t)+X_{i,t}\beta}} \tag{4}$$

Beside this, other methodologies such as linear programming (Kwak *et al.* 2012), support vector machines (Trustorff *et al.* 2011), neural networks (Wu and Wang 2000, Chen *et al.*

2006), nearest neighbour (Yip 2006), non-parametric smoothing techniques, expert systems etc. have also been developed and are being widely used for measuring and understanding credit risk (see Caouette et al. 2008 for an overview).

5.3.3.1 Specification of Baseline Hazard Rate

I use equation (4) as the final form for this empirical study. The term $\alpha(t)$ captures the baseline hazard rate which can take various forms with varying specification of this baseline hazard rate. For a duration-independent model, the baseline hazard rate is assumed to be constant, as in Shumway (2001), who uses a time-invariant constant term, (natural logarithm of firm's age). For a duration-dependent model it is assumed to be time-varying, as highlighted in Beck et al. (1998), who use time dummies to proxy the baseline hazard rate. Indirect measures (use of time dummies) for estimating the baseline hazard rate may be less efficient in capturing the time-varying macro dependencies. Hence I follow the suggestions of both Shumway (2001) and Nam et al. (2008) while specifying my baseline hazard rate, thus controlling for firm's survival time[28] and a volatile macroeconomic environment. To capture the macroeconomic impact facing the firms I construct an industry "weight of evidence" variable , expressing the previous year's sector failure rate as a log odds of failure in each of 51 industrial sectors (INDWOE) (see Altman et al. 2010). I use the population data of each sector to calculate INDWOE as the number of insolvencies relative to number of active firms, for each of the industrial sectors. This serves as a useful proxy to control for the volatile macroeconomic conditions during the sampling period. Finally, I regress the variable Ln(Age) (natural logarithm of firm's age) with the variable capturing macroeconomic impact, to detect the baseline hazard rate at the individual firm level.

[28] We also use natural logarithm of firm's age [Ln(Age)].

5.3.3.2 MODEL'S PERFORMANCE EVALUATION

To evaluate the prediction performance of the models developed I report the receiver operating characteristics (ROC) curves. The ROC curve is obtained by plotting the true positive[29] against the false positive[30] rate, as the threshold to discriminate between non-failed and failed firm's changes. The area under ROC curves (AUROC) is a measure of prediction accuracy of the model with AUROC equal to 1 representing a perfect model. The Gini coefficient and Kolmogorov–Smirnov (K-S) statistic are often used to evaluate the performance of a scoring model, and can be easily calculated from AUROC. The Gini coefficient calculated using the relation $G = 2(AUROC - 0.5)$, is used to assess the consistency in the predictions of the model developed, while the K-S statistics measures the distance between the failed and non-failed distributions at the optimal cut-off point, and is about 0.8 × Gini coefficient (Anderson 2007). A model having a K-S statistic value below 20 should be questioned, whereas a model having a value above 70 should be treated with caution as possibly too good to be true (Anderson 2007).

5.4 RESULTS AND DISCUSSION

I open this section with the analysis of descriptive statistic of explanatory variables selected for this study to understand any extreme variability and potential bias that may arise in my multivariate models due to such variability. Next I estimate the model to proxy the baseline hazard rate, followed by univariate analysis of selected covariates to obtain an initial understanding of their discriminatory power across the broad SMEs category. Finally, I develop separate multivariate insolvency hazard models for micro, small, and medium firms and compare their performance with the insolvency hazard model developed for SMEs as a

[29] A firm actually defaults and the model has classified it as expected default.
[30] A firm actually defaults and the model has classified it as expected non-default.

whole. I also illustrate the steps involved in developing the multivariate models with relevant analysis related to comparison and validation of the results obtained. To ensure that my statistical estimates are not heavily influenced by extreme outliers, I restrict the required covariates between 1st and 99th percentiles except STDEBV. Considering its extreme variability across the entire broad SMEs category I restricted the later between 3rd and 97th percentiles.

5.4.1 Analysis of Descriptive Statistics

For an initial understanding of the variability of the covariates and the potential bias that may arise due to extreme variability, I report the key descriptive statistics as shown in Table 5.3. The mean and standard deviation of all the variables are as per my expectation with no extreme variability, as the required variables have already been winsorized to control any extreme variability and potential bias that may arise due to such variability. I expect the mean of the covariates which are positively related to default probability, to be higher for the failed group than the non-failed group for all firms across the broad SMEs category (e.g. see the variable STDEBV in Table 5.3). On the contrary, I expect the mean of the covariates, which are negatively related to default probability to be lower for failed groups than their non-failed counterparts (e.g. see the variable RETA in Table 5.3). A casual cross-section comparison of covariates across the SMEs category reveals differences in their mean value across micro, small, and medium firms. Hence there is some initial evidence to support my hypothesis that the factors affecting the default probability of SMEs may vary across its broad category. An initial inspection of correlation among the covariates reveals strong positive correlation among TCTL and TCTA of about 0.8, and very strong positive correlation of about 0.98 between log of late filing days and late filing dummy across all categories. The variables EBITDATA and RETA also exhibit strong positive correlation of about 0.81 which supports the fact that SMEs are primarily dependent on their internal resources for financing. I expect these variables to be problematic in the multivariate framework. Finally, the mean and standard deviation of EBITDAIE is very high across all the categories as a significant number of firms in my database incur no interest expense. Hence all the earnings are available[31] to meet such financial obligation, which ultimately lead to very high value of earnings to interest ratio.

[31] If a firm has reported EBITDA as 35,000 GBP and no interest expense, then the ratio EBITDAIE is 35,000.

Table 5.3: Key Descriptive Statistics

Variable		Micro		Small		Medium		SMEs	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
EBITDATA	Failed	0.028	0.420	-0.041	0.400	0.009	0.286	-0.008	0.361
	Non-failed	0.095	0.364	0.073	0.309	0.088	0.224	0.084	0.293
STDEBV3	Failed	3.938	9.762	3.114	7.953	2.902	6.488	3.167	7.744
	Non-failed	3.546	8.935	2.462	6.549	2.259	5.637	2.611	6.822
RETA	Failed	-0.086	0.408	-0.144	0.409	-0.089	0.299	-0.111	0.367
	Non-failed	-0.019	0.343	-0.029	0.302	-0.013	0.222	-0.021	0.283
CTA	Failed	0.151	0.228	0.103	0.174	0.073	0.141	0.099	0.174
	Non-failed	0.191	0.257	0.168	0.218	0.112	0.164	0.151	0.211
EBITDAIE	Failed	46375	342619	15859	293509	38564	361502	30759	332471
	Non-failed	102465	457868	89425	407980	159463	599567	119614	502217
CETL	Failed	1.474	5.772	0.807	3.069	0.727	2.441	0.889	3.488
	Non-failed	2.759	8.549	2.255	6.549	1.678	4.855	2.135	6.468
QACA	Failed	0.820	0.268	0.803	0.235	0.767	0.219	0.791	0.236
	Non-failed	0.833	0.281	0.834	0.225	0.796	0.226	0.819	0.239
lnCR	Failed	0.121	0.998	-0.030	0.786	-0.047	0.668	-0.011	0.784
	Non-failed	0.254	1.181	0.296	0.922	0.168	0.808	0.237	0.944
TCTL	Failed	0.264	0.284	0.277	0.223	0.285	0.200	0.278	0.226
	Non-failed	0.187	0.256	0.235	0.231	0.244	0.210	0.228	0.230
TDTA	Failed	0.299	0.292	0.301	0.235	0.281	0.203	0.292	0.234
	Non-failed	0.213	0.266	0.251	0.230	0.239	0.203	0.239	0.229
TTA	Failed	0.017	0.048	0.007	0.039	0.005	0.029	0.008	0.037
	Non-failed	0.021	0.049	0.016	0.042	0.015	0.035	0.017	0.041
TCTA	Failed	0.236	0.264	0.239	0.203	0.231	0.174	0.235	0.204
	Non-failed	0.138	0.207	0.157	0.176	0.160	0.157	0.154	0.176
STA	Failed	0.152	0.240	0.145	0.188	0.153	0.164	0.150	0.189
	Non-failed	0.131	0.242	0.121	0.180	0.132	0.167	0.127	0.190
IATA	Failed	0.017	0.081	0.028	0.098	0.041	0.114	0.032	0.103
	Non-failed	0.013	0.071	0.021	0.085	0.034	0.105	0.024	0.091
AUDIT	Failed	0.904	0.295	0.975	0.156	0.997	0.058	0.972	0.165
	Non-failed	0.911	0.284	0.973	0.163	0.997	0.055	0.969	0.172
LLF	Failed	1.206	2.024	1.044	1.929	0.997	1.879	1.052	1.926
	Non-failed	0.438	1.357	0.394	1.294	0.363	1.241	0.391	1.287
LFD	Failed	0.273	0.446	0.235	0.424	0.226	0.418	0.238	0.426
	Non-failed	0.098	0.297	0.088	0.283	0.081	0.273	0.087	0.282
CFS	Failed	0.158	0.365	0.236	0.425	0.427	0.495	0.303	0.460
	Non-failed	0.157	0.364	0.258	0.437	0.442	0.497	0.309	0.462
CJJ Number	Failed	0.197	0.962	0.313	1.695	0.204	1.890	0.247	1.683
	Non-failed	0.013	0.191	0.015	0.218	0.008	0.183	0.012	0.199
CJJ Amount	Failed	1797.795	13720.7	2449.678	16724.410	3261.267	20907.130	2680.016	18184.150
	Non-failed	63.394	1592.31	90.754	2900.674	126.503	3323.166	99.017	2865.305

First column lists the covariates studied followed by the failed and non-failed groups in the second column. Third, fifth, seventh and ninth columns report the mean and fourth, sixth, eighth and tenth columns report the standard deviation (S.D.) of micro firms, small firms, medium firms and SMEs respectively.

5.4.2 Detection of Baseline Hazard Rate

To proxy the baseline hazard rate, first I compute the natural logarithm of firm's age[32] and then regress it with my macroeconomic covariates INDWOE and Insolvency Rate. Using the development sample, I estimate the baseline model using random-effects panel data logistic

[32] Age is computed in years.

regression[33] techniques, where the dependent variables has binary outcome (failed = 1 and 0 otherwise). Table 5.4 report the baseline models separately estimated for micro, small, medium and SMEs. As expected all the factors are highly significant in explaining the survival of firms across the broad SMEs category. The significance of macro covariates confirms that the macroeconomic factors affect the survival of firms across the entire SMEs category. The significance of Ln(Age) supports my view that insolvency hazard is duration dependent. Hence I employ all the three covariates to estimate my final baseline hazard rate, thus controlling for both individual firm's survival time and macroeconomic conditions.

Table 5.4: Detection of Baseline Hazard Rate

Covariates	Micro	Small	Medium	SMEs
Ln(Age)	-0.1871***	-0.2668***	-0.2774***	-0.2426***
INDWOE	-0.7075***	-0.6160***	-0.7845***	-0.7227***
Insolvency Rate	2.8695***	2.8473***	2.9230***	2.8770***
Constant	-6.4351***	-6.0213***	-6.1018***	-6.1653***
*** (**) [*] significant at the 1% (5%) [10%] level (two-sided test).				

5.4.3 Initial Discriminatory Analysis

In order to obtain an initial understanding about the discriminatory power of my explanatory variables I perform a simple t-test for differences in the mean of failed and non-failed groups. The first five columns of Table 5.5 report the t-test results for micro, small, medium and SMEs respectively. As I see, most of the variables are highly significant in differentiating among failed and non-failed groups across the broad SMEs segment. However, I do observe some cross-section differences in their discriminatory power. STDEBV is insignificant for micros firms but significant for other categories, which supports the empirical finding that difficulty in access to finance decreases with firm size. The variables AUDIT is insignificant throughout, whereas QACA, CFS and AGE1 show mixed discriminatory power across the broad SMEs category. However, we need to acknowledge that time-varying models have a

[33] We re-organize our database to incorporating the effect of time-varying covariates in our logistic model as stated in equation (4).

different data structure, and a simple t-test may exhibit biased discriminatory power. In multi-period models like the one I use, we need a statistical test that incorporates temporal changes rather than a simple two groups mean comparison test. Nam et al. (2008) use a log-rank test and the Wilcoxon–Breslow–Gehan test, which are widely used methods to test for equality of two hazard functions, but were primary developed for continuous-time models. I believe using tests which are primarily designed for continuous-time models may give biased discriminatory results when applied to discrete-time models. To the best of my knowledge there is no discrete-time counterpart of such a test. Hence I decided to observe the average marginal effect of each covariate in turn on the baseline model. Marginal effect measures the impact on the conditional mean of probability of default due to change in one of the explanatory variables, whereas with average marginal effect, a marginal effect is estimated for each observation, and then all the estimated effects are averaged. The last four columns of Table 5.5 report the average marginal effects for each covariate across the broad SMEs category. Nearly all those covariates which are significant in the simple t-test show a highly significant average marginal effect (AME), confirming that these variables have significant discriminatory power above the baseline model and vice versa. However, I do see some differences between the two methods. The variables QACA and STA are significant in the mean comparison test, but fail to add significant discriminatory power over the baseline model. I get mixed results for AUDIT and CFS when I compare their discriminatory power obtained by employing the two different methods. Although I don't see cross-category differences in the significance level of AME, a close observation reveals that for most of the covariates, the magnitude of AME changes across the broad SMEs category. I expect this behaviour of covariates to be reflected in the multivariate framework. Finally, I expect the variables with significant AME to exhibit significant discriminatory power in the multivariate framework. However problems may arise due to multicollinearity among the explanatory variables and that may render some of the covariates insignificant in the multivariate model.

Table 5.5: Simple t-test and Average Marginal Effect

Variable Name	Two Group Mean Comparison Test				Average Marginal Effect (dy/dx)			
	Micro	Small	Medium	SMEs	Micro	Small	Medium	SMEs
EBITDATA	0.0000	0.0000	0.0000	0.0000	-0.4115***	-0.8754***	-1.1204***	-0.8235***
STDEBV3	0.1317	0.0000	0.0000	0.0000	0.0015	0.0081***	0.0143***	0.0071***
RETA	0.0000	0.0000	0.0000	0.0000	-0.4102***	-0.8085***	-0.9676***	-0.7508***
CTA	0.0000	0.0000	0.0000	0.0000	-0.5649***	-1.4949***	-1.4037***	-1.1930***
EBITDAIE	0.0000	0.0000	0.0000	0.0000	-3.0e-7***	-6.6e-7***	-5.4e-7***	-5.4e-7***
CETL	0.0000	0.0000	0.0000	0.0000	-0.0164***	-0.1167***	-0.2146**	-0.0745***
QACA	0.1010	0.0000	0.0000	0.0000	-0.0062	-0.1602*	-0.0396	-0.0741
lnCR	0.0021	0.0000	0.0000	0.0000	-0.0713***	-0.3369***	-0.3062***	-0.2518***
TCTL	0.0000	0.0000	0.0000	0.0000	0.6799***	0.4663***	0.5983***	0.5576***
TDTA	0.0000	0.0000	0.0000	0.0000	0.6911***	0.5388***	0.6912***	0.6082***
TTA	0.0013	0.0000	0.0000	0.0000	-2.6076***	-6.4276***	-9.1383***	-6.5213***
TCTA	0.0000	0.0000	0.0000	0.0000	1.2889***	1.6158***	1.8472***	1.5558***
STA	0.0027	0.0000	0.0000	0.0000	0.1369	0.1794*	0.1412	0.1218*
IATA	0.0368	0.0001	0.0002	0.0000	0.5714	0.4418**	0.2385	0.4593***
AUDIT	0.3492	0.4448	0.7004	0.2380	-0.1047	0.0723	-.3968	0.0561
LFD	0.0000	0.0000	0.0000	0.0000	1.1993***	1.1178***	1.1783***	1.1494***
LLF	0.0000	0.0000	0.0000	0.0000	0.2478***	0.2327***	0.2483***	0.2399***
CFS	0.9093	0.0106	0.0934	0.3261	0.0115	-0.1530***	-0.0948***	-0.0680***
CJJ Number	0.0000	0.0000	0.0000	0.0000	0.6309***	0.7198***	0.6772***	0.6957***
CJJ Amount	0.0000	0.0000	0.0000	0.0000	0.00006***	0.00003***	0.00003***	0.00003***

*** (**) [*] significant at the 1% (5%) [10%] level (two-sided test). The first column list the covariates studied followed by p-values of two group mean comparison test of micro, small, medium and SMEs in the next four columns respectively. The last four columns report the average marginal effect of each covariate above the baseline model for the broad SME category along with their respective significance levels.

5.4.4 Developing Discrete-time Duration-dependent Hazard Model

I separately estimated four bankruptcy models[34] for SMEs, micro, small and medium firms' respectively using equation (4). The response variable in all four models have binary outcome (failed and non-failed) and the explanatory variables are the set of covariates analysed in section 5.4.3 along with the variables specified in the baseline model. In order to select the best set of explanatory variables I initially select the variables having significant AME after considering correlation among the covariates (I exclude EBITDATA, LLF and TCTL, as they

[34] We re-organize our database to incorporating the effect of time-varying covariates in our logistic model as stated in equation 4.

exhibit very strong correlation with other variables). To estimate final models I use only those covariates which are significant in the multivariate setup. I perform this selection procedure on each sample separately and as a result I obtained different sets of explanatory variables which best explained the outcome variable across the broad SMEs category. I control for the size effect but only for the SMEs model. As for my other models, their definitions already take into account differences that may arise due to their varying size category. I employ a quadratic term in total asset value [(natural logarithm of (1 + total assets) and (natural logarithm of (1 + total assets))^2] in line with Altman et al. (2010) to control the size effect, as there exists a non-linear relationship between insolvency rate and firm size (insolvency risk follow an increasing and decreasing function of firm size). Table 5.6 report the final estimated models for SMEs, micro, small and medium firms respectively. I provide further discussion on the individual models in the following sections.

Table 5.6: Discrete-Time Duration-Dependent Hazard Models

Variable Name	Expected Sign	Coefficients			
		Micro (1)	Small (2)	Medium (3)	SMEs (4)
STDEBV	+	—	0.0050**	0.0072***	0.0043***
RETA	-	-0.2240***	-0.4256***	-0.4301***	-0.3450***
CTA	-	-0.3628***	-0.9048***	-0.7181***	-0.8085***
EBITDAIE	-	—	-1.47e-07**	-2.34e-07***	-1.59e-07***
CETL	-	—	-0.0255***	-0.0348***	-0.0159***
lnCR	-	—	-0.0583**	-0.0861***	-0.0462***
TDTA	+	0.3127***	—	—	—
TTA	-		-3.2342***	-5.6485***	-3.672***
TCTA	+	0.9851***	1.1904***	1.5929***	1.1755***
STA	+	—	—	—	—
IATA	+	—	0.4300**	—	0.6400***
LFD	+	1.0853***	0.9486***	1.0705***	1.0191***
CJJ Number	+	0.3362***	0.4923***	0.3898***	0.4103***
CJJ Amount	+	0.00003***	0.00001***	0.00002***	0.00001***
Ln(Age)	-	-0.1603***	-0.1820***	-0.2239***	-0.1640***
Size log	+	—	—	—	0.3559***
Size Squared log	-	—	—	—	-0.0144***
INDWOE	-	-0.4749***	-0.4176***	-0.6467***	-0.5552***
Insolvency Rate	+	2.5247***	2.5993***	2.7887***	2.5965***
Constant		--6.5562***	-6.2152***	-6.4077***	-8.4065***

*** (**) [*] significant at the 1% (5%) [10%] level (two-sided test). The first column lists the set of covariates significant in the multivariate models followed by second column listing the expected sign of the covariates. The estimated coefficients and their significance for SMEs, micro, small and medium firms are reported in the next four columns respectively.

5.4.4.1 Bankruptcy Prediction for SMEs

I started by estimating a bankruptcy prediction model for my SMEs sample, i.e. all the firms having less than 250 employees. Table 5.6 reports the final distess prediction model estimated for SMEs. All seventeen covariates in the final model are highly significant with their expected sign. Unlike previous findings (Altman and Sabato 2007; Altman et al. 2010), the variable EBITDATA is excluded from the multivariate framework, as it shows strong positive correlation with RETA. This high correlation also supports the view that SMEs face difficulty in accessing external finance and are primarily dependent on internal sources of finance like retained earnings. The variables AUDIT and CFS fail to add any significant discriminatory power in the multivariate setup, which contadicts the findings of Altman et al.

(2010). Finally all the variables in the baseline hazard function appear highly significant in the multivariate hazard model with expected sign of their respective coefficients.

5.4.4.2 Bankruptcy Prediction for Micro Firms

I estimate the bankruptcy prediction model for micro firms with my sample of firms having less than 10 employees. Table 5.6 report the final model developed using seven covariates along with the baseline model. All the variables of the baseline hazard function are highly significant with expected sign of the respective coefficients. As I see, the three key covariates EBITDATA, STDEBV and EBITDAIE which are reported significant in prior SMEs bankruptcy studies, (see among others Altman and Sabato 2007; Altman et al. 2010), are insignificant in explaining the financial distress of micro firms, which clearly outline their differences and suggest financial reports do not provide material information about their likelihood of failure. Variables capturing the financial requirements of micro firms (RETA and CTA) are highly significant, highlighting the importance of internal sources of finance and liquidity for their survival. Also, consistent with the suggestion of Hudson (1986) I find short term leverage variables TCTA significant, which support the fact that small firm bankruptcy is primarily influenced by trade creditors. Similarly, their survival is also dependent upon how efficiently they manage their debtors as I see TDTA bear significant positive relationship with default probability. Jointly the variables TCTA and TDTA highlight the importance of working capital management on the survival of micro firms which is in line with prior empirical literature. The variable TTA shows significant AME but fails to be significant in the multivariate setup, which supports the fact that most of the micro firms enjoy tax concession and hence TTA does not exhibit significant explanatory power for this category. LFD, CJJ number and CJJ amount are highly significant as per my expectation. Table 5.6 clearly highlight that the factors affecting the hazard risk of micro firms and SMEs are significantly different. The variables STDEBV, EBITDAIE, CETL, lnCR and IATA are significant only for the SMEs group. Finally, in view of my empirical findings, I have strong motivation to believe that the credit risk characteristics of micro firms and SMEs do vary substantially and micro firms need to be considered separately when modeling credit risk for them.

5.4.4.3 Bankruptcy Prediction for Small Firms

I estimate this model with my small firms sample, i.e. firms having 10 or more but less than 50 employees. Table 5.6 reports the final distress prediction model for small firms using twelve highly significant covariates, along with the baseline hazard function. Unlike micro firms most of the financial ratios (RETA, CTA, EBITDAIE, CETL etc.) are significant in explaining the failure propensity of small firms, suggesting that the explanatory power of financial reports increases with the size of the firm. The variable STDEBV is significant, which emphasize that access to external finance is easier with increasing firm size. All twelve significant covariates are significant in explaining the financial distress of my SMEs model as well. Considering the test results, there may not be material impact on the decision making process, by treating both small firms and SME's overall separately.

5.4.4.4 Bankruptcy Prediction for Medium Firms

I estimated my final model for the sample of medium firms (firms having 50 or more but less than 250 employees), using similar method as applied for my other models. I estimate the final model using twelve highly significant covariates (see Table 5.6), along with the baseline model. The models developed for small and medium firms share the same set of significant explanatory variable. The variable STDEBV is insignificant in models developed for micro firms but is highly significant in models developed for small and medium firms and SMEs, which again supports the view that difficulty in access to external finance decreases with increase in firm size. All the twelve predictors that my SMEs model employs are significant predictors in my medium firm's model. Hence, based upon my finding I suggest that medium firms need not be considered separately when modeling credit risk for them.

5.4.5 MODEL VALIDATION

In order to evaluate the prediction performance of the hazard models developed I estimated the receiver operating characteristics (ROC) curves. The hold-out sample estimates of area under ROC curves (AUROC) are reported in Table 5.7. The models developed for small firms exhibit highest AUROC of about 0.77, SMES and medium firms, achieve fairly strong classification performance with AUROC of about 0.76. While the model developed for micro firms reveals slightly poor classification performance with AUROC of about 0.74. All the models have AUROC close to 0.75 which suggest that all the models developed exhibit fairly robust prediction performance.

Table 5.7: Area under ROC Curves

Hazard Model	Area Under ROC Curve
Micro	0.7414
Small	0.7758
Medium	0.7631
SME	0.7610

This table reports the area under the ROC curve (AUROC) of hold-out samples. The AUROC is equal to the probability that the rating for a true positive (a firm actually defaults and the model has classified it as expected default) will be less than that for a true negative (a firm does not default and the model has classified it as expected non-default) plus 50 per cent of the probability that the two ratings will be equal.

5.5 CONCLUSION

There is huge diversity that exists within the broad SMEs category (micro, small, and medium firms) in the form of capital structure (Ramalho and Da Silva 2009; Mateev et al. 2013), access to external finance (Beck *et al.* 2006), management style (Wager 1998), default probability (Pettit and Singer 1985), firm size, number of employees etc. In this paper I investigate the financial and non-financial factors influencing SMEs failure propensity, in order to identify any differences that may exist within this broad SMEs category. I apply discrete-time duration-dependent hazard rate modeling techniques to develop separate bankruptcy prediction models for micro, small, and medium firms respectively, using a relatively large database of UK firms. I compare their performance with the model developed for SMEs, as a whole, which includes micro, small, and medium firms. I estimate the insolvency hazard models after taking account of correlation among the covariates. Finally, I compare the estimated models (for micro, small, and medium firms) with my SMEs model in turn, to identify the common default attributes.

To undertake the statistical estimations, I use a heterogeneous-panel that contains financial and non-financial information of 8,162 failed and 385,733 non-failed UK SMEs covering the analysis period between 2000 and 2009. The data between analysis year 2000 to 2007 is used as my development sample, while I retain the data of analysis year 2008 and 2009 as a hold-out sample to validate out-of-sample prediction performance of the models developed. To validate the robustness of the models developed I estimate the area under the receiver operating characteristics curves (AUROC) for hold-out samples.

All the multivariate models demonstrate fairly strong classification performance with AUROC of about 0.77 for holdout samples except the model developed for micro firms; which shows slightly lower AUROC of approximately 0.74. Further comparison of default prediction models for micro firms and SMEs strongly suggest that they need to be considered separately while modeling credit risk for them. Three of the financial ratios reported significant in Altman and Sabato (2007) are insignificant in my micro model but significant in my SMEs model. The variables capturing working capital, and short term leverage position (trade creditors/total assets and trade debtors/total assets), are highly significant in line with prior empirical findings (Hudson 1986, Beck *et al.* 2006). Given that a near identical set of explanatory variables affect the default probability of small firms and SMEs, I do not expect a material impact on the decision making process by treating each of these groups separately. Finally, I make comparisons between hazard models developed for medium firms and SMEs. Once again almost the same set of explanatory variables are highly significant in both the models, hence I suggest that considering both the groups separately may not lead to superior risk pricing.

Based upon my findings, I see that financial reports do not provide sufficient information about the likelihood of default of micro firms and hence there is a clear need of separate treatment, as globally the vast majority of business units are very small (Di Giovanni et al. 2011) with median business units having no employees.

6. THE CORRECT USE OF HAZARD MODELS: A COMPARISON EMPLOYING DIFFERENT DEFINITIONS OF SMEs FINANCIAL DISTRESS

6.1 INTRODUCTION

Survival or *event history analysis* is the umbrella term for the set of statistical tools that are used to answer questions related to timing and the occurrence of an event. The statistical models examine the hazard rate, which is defined as the conditional probability that an event of interest occurs within a particular time interval (t). The growing popularity of *hazard* models in predicting corporate failure has motivated me to undertake this empirical study. Since the seminal work of Shumway (2001), the use of the hazard rate modelling technique (also called *survival analysis*) has become a standard methodology in firms' default prediction studies (see among others Chava and Jarrow 2004; Campbell *et al.* 2008; Gupta *et al.* 2014). However, this growing popularity of hazard models in bankruptcy prediction seems to be trend or momentum driven rather than strong theoretical understanding. Although the superiority of hazard models in predicting binary outcomes is well documented in the literature (see among others Beck *et al.* 1998; Shumway 2001; Allison 2014), but its recent use in predicting corporate failure seems to dilute the primary notion behind the use of survival models. Most of the existing studies suffer from at least one of the following issues: (i) reasons behind their choice between *discrete-time* or *continuous-time* hazard model (ii) inappropriate specification of baseline hazard rate (iii) no test of proportional hazards assumption when using *Extended Cox* models with time-independent covariates (iv) ignore *frailty* and *recurrent* events (v) explanation on how they dealt with the issues of *delayed entry* (vi) explanation on treatment of time periods/intervals having no events.

The variable of primary interest in survival analysis is the time to some event, which in my case is the incorporation of a firm to bankruptcy filing. A firm is said to be at risk of the event of interest (bankruptcy) after the initial event (incorporation) has taken place. Alternatively, the response variable can be viewed as the time duration that a firm spent in healthy state until transition to bankruptcy state takes place. Survival analysis demands special methods primarily due to *right-censoring*, where the time to the occurrence of an event is unknown for some subjects because the event of interest has not taken place by the end of the sampling or observation period. A remarkable feature of hazard models is that *time-varying* covariates can be included. The *survival time*, which is the duration or time to event is generally measured in quarterly or annual units in bankruptcy studies. Furthermore, the time scale used may be discrete or continuous. If the time of occurrence of an event is precisely known, *continuous-time* hazard models are employed, otherwise *discrete-time* hazard model is an appropriate choice when the event takes place within a given time interval and the precise time is unknown (Rabe-Hesketh and Skrondal 2012). Thus, from a theoretical point of view discrete-time hazard models are an appropriate choice as a firm may file for bankruptcy anytime within a quarter or year. However, in both models the probability of occurrence of an event at time *t* is being modelled. The dependent variable in a continuous-time model is the *hazard rate* but in a discrete-time model it is the *odds ratio* (if modelling is done using standard logit/probit models). However, in recent studies the choice between discrete-time (eg. Campbell *et al.* 2008, Gupta *et al.* 2014) and continuous-time model (eg. Bharath and Shumway 2008, Chen and Hill 2013) seems to be random without any satisfactory explanation behind their choice. Furthermore, the required precision of the timing to an event is significantly dependent on the research question and data restrictions. Studies also suggest that results obtained from continuous-time and discrete-time methods are virtually identical in most models (Yamaguchi 1991; Allison 2014). However, the performance of a bankruptcy prediction model is evaluated based on some non-parametric classification measures like *misclassification matrix*, area under *receiver operating characteristic (ROC) curve* etc. (see Anderson (2007) for further details). Thus, despite the theoretical differences between continuous-time and discrete-time models, if they lead to identical classification performance, then this theoretical difference is of no practical relevance. Thus, I compare the classification performance of most widely used discrete-time duration-dependent hazard models (see among others Shumway 2001, Nam *et al.* 2008) with the most popular continuous-time duration-dependent *Cox* model (see among others Bharath and Shumway 2008, Chen and Hill 2013) to find any differences in their classification performance.

If there are no differences, then the Cox model shall be a reasonable and convenient choice, as it does not require any *baseline hazard* specification unlike discrete-time models (see Rabe-Hesketh and Skrondal 2012). Baseline hazard is defined as the hazard rate when the value of all the covariates is zero. The baseline hazard is estimated using time dummies (Beck *et al.* 1998) or some other functional form of time (Jenkins 2005). However, recent studies seems to have distorted this idea of baseline hazard and have established their own version of baseline hazard that includes macroeconomic variables (Nam *et al.* 2008), insolvency risk (Gupta *et al.* 2014) etc., in the baseline hazard function, while many prefer not to include any baseline hazard function in their model (see among others Campbell *et al.* 2008, Bauer and Agarwal 2014). In light of the basic theory of survival analysis, this is inappropriate. Thus, I address this misleading concern in this study and show the steps that need to be followed in specifying the baseline hazard function while developing a discrete hazard model. On the other hand, studies which employ continuous-time Cox models are silent on the critical test of proportional hazards (PH) assumptions for time-independent covariates (e.g. Liang and Park 2010). The PH assumption implies that the hazard rate of any particular subject is a constant proportion of the hazard rate of any other subject across time (Mills 2011). The violation of this assumption might lead to overestimation (the covariate violates this assumption and exhibit an increasing hazard ratio over time) or underestimation (the covariate violates this assumption and exhibit a decreasing hazard ratio over time) of hazard risk (Mills 2011). It also results in incorrect standard errors and decrease in the power of significance tests (Box-Steffensmeier and Zorn 2002). The violation of PH assumption is a frequent phenomenon and thus, it should always be checked and reported in studies. Having said that, Allison (2010) warns that, it is not enough to worry only about the violation of the PH assumption but also about other basic requirements, such as incorporation of relevant explanatory variables. Although all the covariates that I employ in this study are time-dependent, if one also employs time-independent covariates, then one should take recognition of this serious and neglected concern and use appropriate methods to test, report and rectify any violation of the proportional hazards assumptions[35].

Another highly neglected area of concern is *frailty* and *recurrent* events. Correlation of event time occurs when firms experiencing default event belong to a particular cluster or groups like industry, geographic location etc. or in case of recurrent events, where a firm experiences a default event more than once in its lifetime. In the United States (US), the Bankruptcy Reform Act of 1978 (Code) governs the legal processes involved in dealing with corporate financial distress. It allows firms facing financial distress for a liquidation process (Chapter 7) or a reorganization process (Chapter 11) [36]. Chapter 7 leads to permanent shut down of a financially distressed firm, while Chapter 11 aims at rehabilitation of financially distressed

[35] See Kleinbaum and (Klein 2012) for detailed understanding about various tests of proportional hazards assumption for time-independent covariates. A Cox model with time-dependent covariates does not need to satisfy the proportional hazards assumption and is called an *Extended Cox* model. However, if the model employs both time-dependent and time-independent covariates, then PH assumption for time-independent covariates must be satisfied (Kleinbaum and Klein 2012).

[36] Although the law provide other provisions but we consider only Chapter 11 and Chapter 7, as vast majority of the financially distressed firms file for either of these two.

but economically viable firms. Hotchkiss (1995) examines 197 publicly traded firms that filed for Chapter 11 protection during 1979 to 1988 and later recovered from Chapter 11 as publicly traded firms. He reports that 40% of the firms continue to experience operating losses and 32% either restructure their debt or re-enter bankruptcy in the three years following the acceptance of reorganization plans. Thus a firm may witness multiple distress events in its lifetime. Given that these issues of clustering and recurrent events are an integral part of the real-life environment, they should be made an essential and standard part of contemporary event history analysis (see Box-Steffensmeier and De Boef (2006) and Mills (2011) for advanced discussion). The solution is to introduce a *frailty* term in the hazard models. *Frailty* is an unobserved random proportionality factor that modifies the hazard function to account for random effects, association and unobserved heterogeneity into hazard models (Mills 2011). Not including a frailty term implicitly assumes that all firms are homogeneous, which implies that all the firms are prone to experience default in the same way, with the duration of defaults considered as independent from one another. However, in real-life some firms are more 'frail' and thus provide a higher likelihood to experience default. Therefore, my empirical analysis also accounts for this neglected concern while developing the hazard models.

Furthermore, in time to event studies the origin of time scale is an important consideration, as at this point in time a firm starts being at risk of experiencing the financial distress event. This needs to be firms' incorporation date in bankruptcy studies. However in cases where incorporation dates are unknown, firms' age or the earliest available date of information in the databases serves as useful proxy. A firm's incorporation date may differ from the start date of sampling period; as a result the time firms become at risk do not coincide with the start of the sampling period. This leads to *delayed entry*, which means that a firm become at risk before entering the study. Thus the appropriate likelihood contribution under delayed entry is obtained by allowing the firm to start contributing observations from time period $t_k + 1$ and discarding the prior time periods (see section 14.2.6 of Rabe-Hesketh and Skrondal 2012). Where t_k is the time period for which a firm has already been at risk when it enters the research study.

In light of the discussion presented above, I contribute to the literature by presenting a comprehensive analysis of the use of hazard models in predicting corporate failure, which takes into account all the serious and neglected concerns discussed above. I expect this study to be an essential guide to bankruptcy and social science researchers interested in using hazard models for making binary predictions. I also intend to be the first paper to provide a comparison between the prediction performance of discrete-time and continuous-time hazard models in the context of SMEs insolvency hazard prediction.

In addition, I also contribute to the fast growing literature on SMEs bankruptcy, by providing a comprehensive comparison of SMEs failure prediction models developed using different definitions of default events. In particular, my comparison involves default definitions based on: (i) legal consequences (Chapter 7/11 bankruptcy filings), (ii) financial health, as discussed in Pindado *et al.* (2008) and Keasey *et al.* (2014) and (iii) both legal and financial health of an SME, which I propose in this study. My legal definition classifies a firm a default when it files for bankruptcy under the bankruptcy law (*Event 1*), which is usually Chapter 7/11 in the US. My second definition partially follows the distress definition provided by Keasey *et al.* (2014) and classifies a firm as financially distressed if it reports earnings less than its financial expenses for two consecutive years, has net worth/total debt less than one and experiences negative growth in net worth for the same two consecutive periods (*Event 2*). The definition of SMEs default that I propose combines *Event 1* and *Event 2*, and classifies a firm as default when it files for legal bankruptcy besides being financially distressed (*Event 3*). The detailed analogy behind this default definition is discussed in the following section. However, a recent study by Lin *et al.* (2012) on SMEs default prediction follows a similar line and predicts SMEs default using different definitions of financial distress, but my study differs from them in several respects. First, I present my analysis based on sample of US SMEs, whereas their study employs sample of UK SMEs. They use static binary logistic regression to establish their empirical validations, while I use much superior dynamic hazard models. Finally, they use a flow-based (earnings/interest payable) and stock-based (1 – total liabilities/total assets) insolvency indicators to group the firms in their sample into four groups of financial health (which corresponds to their four different definitions of financial distress), however my distress definitions are more realistic and arguably superior (see Tinoco and Wilson (2013) and Keasey *et al.* (2014) for relevant discussion).

My test results obtained by employing 3,951 firm-year observations of the US SMEs provide convincing evidence. First, in line with the theoretical arguments discrete-time duration-dependent hazard models that I develop with logit and complementary log-log (clog-log) links perform marginally better than Extended Cox models in identifying defaulted firms across all default definitions. For respective default definitions, almost an identical set of covariates explain the financial distress of the US SMEs when estimation is done using discrete-hazard model with logit and clog-log links. Both these econometric specifications also lead to almost identical classification performance. Thus, one is left to their personal taste when choosing among these two discrete-time specifications. However, I report some variation in the significance of covariates across different default definitions when estimation is done using extended Cox model. Based on the Akaike information criterion (AIC) values I understand that discrete-hazard models provide much superior fit than continuous Cox model across all default definitions. Second, the default definition that I propose (Event 3) performs best in classifying defaulted firms. Event 1 classifies about 52% of defaulted firms in the top three deciles, while for Event 2 and Event 3 this value increases by about 33% and 44%

respectively. Thus a default definition based on firms' financial health is superior to default definition based on legal consequences, while a default definition that considers both legal consequence and firms' financial health is best. These differences in classification performance emphasises the fact that all firms that file for legal bankruptcy are not based purely due to financial difficulties. A significant number of firms do consider this as a planned exit strategy (Bates 2005).

The rest of this paper is organized as follows: section 6.2 discusses various default definitions that I consider in my study; section 6.3 provides discussion related to my dataset, choice of covariates and methodology; in section 6.4 I report and discuss my empirical findings and finally, section 6.5 concludes my findings.

6.2 DIFFERENT DEFAULT DEFINITIONS FOR SMES

Traditionally, the debate about financial distress has been rooted in the literature pertaining to firms' capital structure with particular relevance to the cost of financial distress (see Altman and Hotchkiss (2006) for an overview). However, current studies also highlight its growing importance in the context of modelling firms' insolvency hazard (e.g. Keasey *et al.* 2014). Recent literature pertaining to firms' default prediction argue that a 'financial distress' based definition of default contingent upon a firm's earnings and market value is more appropriate than the definition based on legal consequence (Pindado et al. 2008; Tinoco and Wilson 2013). We see a range of definitions in the empirical literature that have been successfully used to define/proxy firms' default/distress risk. Most of the empirical models employ a definition of default that is in line with some legal consequence (e.g. Chapter 11/7 Bankruptcy Code in United States; United Kingdom Insolvency Act), which lead to a well-defined and clearly separated population of bankrupt versus non-bankrupt firms. This remains the most widely used method of classifying financially distressed firms in the empirical literature, that employ binary choice statistical models to predict firms' financial distress (see among others Altman 1968, Ohlson 1980, Hillegeist *et al.* 2004, Gupta *et al.* 2014a). However, legal definition of default may suffer from noteworthy issues. Since insolvency involves lengthy legal processes, often there exists a significant time gap between real/economic default date and legal default date. UK companies exhibit a significant time gap of up to 3 years (average of about 1.17 year) between the time they enter into the state of financial distress and legal default dates (Tinoco and Wilson 2013), while companies in US stop reporting their financial statements about two years before filing for bankruptcy (Theodossiou 1993). Recent changes in insolvency legislation (for instance, the Enterprise Act 2004 in the UK or Chapter 11 in the US) do consider this problem and suggested several stages of financial distress based upon the severances of financial distress.

Further, a financially distressed firm may go for a formal reorganization involving the court system or an informal reorganization through the market participants. *Debt restructuring*, *asset sale* and *infusion of new capital from external sources* are the three commonly used market-based/private methods of resolving financial distress (Senbet and Wang 2010). Debt restructuring allows a financially distressed firm to renegotiate the outstanding debt obligation or related credit terms with its creditor/s but is critically subject to whether the debt obligation is due to private or public entity. As an alternative to this, a distressed firm may sell-off some of its existing assets to reduce its outstanding liability or may undertake new profitable investment opportunities, which may eventually help it to overcome its misery. Despit having profitable investment opportunities, a financially distressed firm might not be able to generate additional funding due to high risk involved in financing distress firms and the "debt overhang" problem as discussed in Myers (1977). As a consequence, infusion of new capital from external sources is rarely observed in the resolution of financial distress.

Thus, we cannot rule out the possibility that a financially distressed firm may not file for Chapter 7 or Chapter 11 protection and choose a private workout method of resolving financial distress. Gilson et al. (1990) and Gilson (1997) report that firms avoid legal bankruptcy processes by out of court negotiation with creditors. However, under the binary classification based on legal consequences, a finically distressed firm which has not filed for Chapter 7 or Chapter 11 is not considered as a financially distressed firm. Thus, there is a clear need of a mechanism to identify finically distressed firms beyond the legal definitions. In this context, I find the argument of Pindado et al. (2008) highly relevant and thus I explore the following definitions of SMEs' default events:

Event 1 - Any firm which files for bankruptcy under Chapter 7/11 is considered default and is said to have experienced Event 1. Vast majority of the empirical literature on SMEs default prediction employ this kind of binary classification based on some legal consequences to classify a firm a health or bankrupt (see among others Altman and Sabato 2007, Gupta, *et al.* 2014b).

Event 2 – I partially follow the financial distress definition provided by Keasey *et al.* (2014) while classifying a SME as default under Event 2. In particular, I consider a firm as financially distressed if it's EBITDA (earnings before interest tax depreciation and amortization) is less than its financial expenses for two consecutive years; the net worth/total debt is less than one and the net worth experienced negative growth between the two periods.

Event 3 – The third default definition that I propose considers both legal and finance-based definition of distress while classify a firm as default. A firm is classified as default under Event 3 if it satisfies the conditions of Event 1 and Event 2 simultaneously. That is, besides being financially distressed it should also file for bankruptcy under Chapter 7/11. The rationale being, all business closures are not due to financial difficulties, many file for legal bankruptcies as part of their planned exit strategies (see among others Bates 2005). Thus, this definition identifies firms which follow legal exit routes due to pure financial difficulties.

6.3 Empirical Methods

This section provides discussion related to the source and use of dataset, selection of explanatory variables and statistical models that I use in my research.

6.3.1 Dataset

To predict the financial distress over the next one year horizon, my empirical analysis employs annual firm-level accounting data from the Compustat database. I consider a relatively long analysis period that runs from the year 1952 through 2013. Small and medium-sized enterprises having less than 250 employees and sales turnover less than $ 65 million, that filed for legal bankruptcy proceeding between January 1952 to December 2013 have been chosen for this study. Thus, the bankruptcy data that I use includes only those firms which filed for bankruptcy (Chapter 11/7) within this time period. In Compustat, a company has "TL" footnote on status alert (data item STALT) indicating that the company is in bankruptcy or liquidation (i.e. Chapter 7/11). Generally, a company will have a "TL" footnote on status alert - quarterly (and annual) for the first and following quarters (and years) the company is in Chapter 11. An "AG" footnote will appear on Total Assets (AT_FN) – quarterly, on the quarter the company emerges from Chapter 11. Thus, within its lifetime a firm may go for multiple bankruptcy filings in form of Chapter 11 and may remain in the bankruptcy state until it emerges. Consequently, taking the bankruptcy filing date as the bankruptcy indicator ignores the possible subsequent bankruptcy states. Thus, my first definition (Event 1) consider a firms under *bankruptcy* when its status alert is "TL" and healthy otherwise. This, classification is consistent with the basic notion of survival analysis in which a subject may remain in a given risky state for more than one time period and thus experience an event of interest for more than one time period. However, it is not a matter of concern as I am modelling the transition hazard.

Table 6.1 reports the age-wise distribution of censored and distressed firms under respective default events (see section 6.2 for definitions of various default events). I proxy a firm's age as the earliest year for which, financial information for that firm is available in the Compustat database. In Compustat, 1950 is the earliest point in time for which financial information is available. Thus, in order to get the complete financial history of a firm, I selected only those firms which entered the Compustat database after 1950. Further, firms with Standard Industrial Classification (SIC) codes from 6,000 through 6,999 (financial firms) and 4900 through 4949 (regulated utilities) have been excluded from my analysis. This leads me to a total of 3,951 firm-year observations for 398 US SMEs. It should be noted that same firms might have multiple entry and exits in my database. For instance, when an existing SME reports sales revenue over $ 65 million it exits my sample and returns only when its sales revenue drops below $ 65 million. Thus, the age variable needs to be created before applying all filters.

Table 6.1: Survival Table

Age	Event 1			Event 2			Event 3		
	0	1	% 1	0	1	% 1	0	1	% 1
3	161	23	12.50	156	28	15.22	177	7	03.80
4	210	46	17.97	209	47	18.36	236	20	07.81
5	221	47	17.54	218	50	18.66	252	16	05.97
6	231	46	16.61	233	44	15.88	264	13	04.69
7	226	48	17.52	233	51	18.61	252	22	008.03
8	223	39	14.89	219	43	16.41	249	13	04.96
9	191	60	23.90	204	47	18.73	229	22	08.76
10	169	48	22.12	181	36	16.59	207	10	04.61
11	166	37	18.23	159	44	21.67	189	14	06.90
12	145	29	16.67	140	34	19.54	162	12	06.90
13	140	33	19.08	146	27	15.61	159	14	08.09
14	135	25	15.63	136	24	15.00	158	8	05.00
15	122	17	12.23	115	24	17.27	136	3	02.16
16	115	13	10.16	105	23	17.97	123	5	03.91
17	97	15	13.39	86	26	23.21	103	6	05.36
18	79	10	11.24	69	20	22.47	86	3	03.37
19	73	11	13.10	70	14	16.67	80	4	04.76
20	64	10	13.51	66	8	10.81	71	3	04.05
21	59	13	18.06	62	10	13.89	71	1	01.39
22	56	9	13.85	52	13	20.00	63	2	03.08
23	53	6	10.17	50	9	15.25	57	2	03.39
24	53	4	7.02	49	8	14.04	57	0	00.00
25	44	8	15.38	41	11	21.15	48	4	07.69
26	35	10	22.22	33	12	26.67	41	4	08.89
27	36	8	18.18	35	9	20.45	41	3	06.82
28	30	6	16.67	34	2	5.56	35	1	02.78
29	24	5	17.24	25	4	13.79	28	1	03.45
30	20	3	13.04	18	5	21.74	20	3	13.04
31	20	2	9.09	19	3	13.64	22	0	00.00
32	16	3	15.97	18	1	5.26	18	1	05.26
33	14	2	12.50	14	2	12.50	15	1	06.25
34	12	1	7.69	13	0	00.00	13	0	00.00
35	10	2	16.67	12	0	00.00	12	0	00.00
36	8	2	20.00	10	0	00.00	10	0	00.00
37	9	1	10.00	8	2	20.00	10	0	00.00
38	7	1	12.50	5	3	37.50	10	0	00.00
39	5	1	16.67	4	2	33.33	6	0	00.00
40	5	0	00.00	3	2	40.00	5	0	00.00
41	4	1	20.00	3	2	40.00	4	1	20.00
42	3	0	00.00	3	0	00.00	3	0	00.00
43	2	0	00.00	1	1	50.00	2	0	00.00
44	3	0	00.00	2	1	33.33	3	0	00.00
45	2	0	00.00	2	0	00.00	2	0	00.00
46	1	0	00.00	1	0	00.00	1	0	00.00
47	1	0	00.00	1	0	00.00	1	0	00.00
48	1	0	00.00	1	0	00.00	1	0	00.00
49	1	0	00.00	1	0	00.00	1	0	00.00
50	1	0	00.00	1	0	00.00	1	0	00.00
51	1	0	00.00	1	0	00.00	1	0	00.00
52	1	0	00.00	1	0	00.00	1	0	00.00
53	1	0	00.00	1	0	00.00	1	0	00.00
Total	3306	645		3259	692		3732	219	

Notes: This table shows the age wise distribution of firm-year observations for respective default events discussed in section 6.2. Numeric ‘0’ signifies censorship and ‘1’ signifies that a firm has experienced the respective default event.

6.3.2 Selection of Covariates

To develop the hazard models I employ financial ratios that have established reputation in predicting SMEs default risk. My choice of covariates gauges firms' performance from liquidity, solvency, activity, profitability and interest coverage dimensions. Specifically, I incorporate the covariates from popular studies on SMEs bankruptcy like Altman and Sabato (2007), Lin *et al.* (2012), Gupta *et al.*(2014) and similar others. Table 6.2 lists all the covariates along with their respective definition. To get detailed understanding pertaining to my choice of covariates and their relationship with the default probability, I strongly recommend one to see Altman *et al.* (2010) and Gupta *et al.* (2014).

Table 6.2: List of Covariates

Variable	Definition	Compustat Data Item
EBITDATA	Earnings before interest taxes depreciation and amortization/total assets	EBITDA/AT
EBITDAIE	Earnings before interest taxes depreciation and amortization/interest expense	EBITDA/XINT
STDEBV	Short term debt/equity book value	DLC/SEQ
CTA	Cash and short-term investments/total assets	CHE/AT
RETA	Retained earnings/total assets	RE/AT
CETL	Capital employed/total liabilities	(AT – LCT)/LT
TLTA	Total liabilities/total assets	LT/AT
CAG	Capital growth; calculated as ($Capital_t$ / $Capital_{t-1}$) - 1	(AT - LCT)
TTA	Taxes/total assets	TXT/AT
LCR	ln(current assets/current liabilities)	ln(ACT/LCT)
TCTA	Trade creditors/total assets	AP/AT
FETA	Financial Expense/total assets	XINT/AT

Notes: This table lists the set of covariates along with their respective definition that I use for the empirical analysis. The last column lists the specific Compustat data items that I use to calculation the financial covariates.

6.3.3 Hazard Models

6.3.3.1 Basic Hazard Model

Survival analysis deals with the analysis of the time to the occurrence of an event, which in this study is the time until a financial distress event. Suppose T is a non-negative random variable which denotes the time to a distress event and *t* represent survival of a firm beyond time *t*. If instead of referring to *T*s probability density function as $f(t)$ or its cumulative distribution function (CDF) as $F(t) = \Pr(T \leq t)$, we think of *T*s survivor function, $S(t)$ or its hazard function $h(t)$ the understanding of survival analysis becomes much more convenient (Cleves *et al.* 2010). The survivor function estimates the probability of survival beyond the time *t*, which is simply the reverse CDF of *T*, i.e.:

$$S(t) = 1 - F(t) = \Pr(T > t) \tag{1}$$

At $t = 0$ the survivor function is equal to one and moves toward zero as t approaches infinity. The relationship between survivor function and hazard function (also known as the conditional failure rate at the time t) is mathematically defined as follows:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{-d \ln S(t)}{dt}; \tag{2}$$

In simple words, hazard rate is the (limiting) probability that the failure event occurs within a given time interval, given that the subject have survived to the start of that time interval, divided by width of the time interval. The hazard rate varies from zero to infinity and may be increasing, decreasing or constant over time. Hazard rate of zero signifies no risk of failure at that instant, while infinity signifies certainty of failure at that instant.

6.3.3.2 Extended Cox Model

An elegant and computationally feasible way to estimate the hazard function (2) is to use the semi-parametric Cox proportional hazards (CPH) model (Cox 1972, 1975) as shown in equation (3). Here, x_i' is the transpose of covariates vector x_i, β is the vector of regression parameters and $h_0(t)$ is the arbitrary unspecified baseline hazard function (hazard risk that the subject *i* faces in absence of covariates; i.e. x = 0). The regression parameters (βs) are estimated using partial likelihood function which takes into account censored survival times and eliminates the unspecified baseline hazard term $h_0(t)$. CPH model treats time as continuous, and is semi-parametric in the sense that the model does not make any assumption related to the shape[37] of the hazard function over time.

$$h_i(t) = h_0(t). \exp(x_i'\beta) \tag{3}$$

Some of the factors (leverage ratio, profitability ratio, volatility etc.) affecting firms' survival vary with time but the fixed CPH model as highlighted in equation (3) does not allow for time-varying covariates. However, inclusion of time-varying covariates in CPH framework is relatively easy and thus enables us to predict dynamic survival probability over the life of the firm. The CPH model can be generalized to allow for the covariate vector x to be time-varying as follows:

$$h_i(t) = h_0(t). \exp(x(t)_i'\beta) \tag{4}$$

Where $x(t)$ is the covariate vector at time t. The rate of change of time-varying covariates is different for different subjects and hence the estimated hazard ratio does not remain constant over time. However, the inclusion of time-varying covariates is not problematic for the partial likelihood estimation (Allison 2010) and hence CPH model can be easily improved to allow for non-proportional hazard risks. It implies that a general hazard model which does not have the restrictive assumption of constant proportional hazard ratio can be generalized by inclusion of both duration-dependent and duration-independent covariates in the same model. However, a CPH model with time-varying covariates is no longer a proportional hazards model and a CPH model with time-varying covariates is appropriately called *Extended Cox* model (see Kleinbaum and Klein 2012). Additionally the time-varying covariates do not need to satisfy the proportional hazards assumption. However, if the model also includes time-independent covariates, then appropriate test of proportionality is suggested (see Kleinbaum

[37] It could be increasing, decreasing, decreasing and then increasing or any shape we may imagine. But it assumes that whatever is the general shape of the hazard function, it's same for all the subjects.

and Klein 2012). One major advantage of Cox method is that it easily addresses the problem of right censoring but it suffers from a major disadvantage of proportional hazards assumption if time-independent covariates are also included in the model. One may consider to test[38] this restrictive proportional hazard assumption, that is being neglected in most empirical studies by using the *scaled Schoenfeld residual* (Grambsch and Therneau 1994) rather than the *Schoenfeld residual* (Schoenfeld 1982). While estimating my empirical model I also control for *unobserved heterogeneity* and *recurrent events* by including a *shared frailty* term into my model via a multiplicative scaling factor α_i (Cleves *et al.* 2010). These signifies group-level frailty and are unobservable positive values assumed to follow the *Gamma distribution* with mean 1 and variance θ to be estimated using the development sample (Jenkins 2005). Also, the time at which the distress event occurs is not really relevant for hazard risk analysis using Cox method, but the ordering of the distress event is critically important. In situations where multiple firms experience the event of interest at the same time, exact ordering of distress event is difficult. Thus I use *Efron*[39] (1977)'s method to handle cases of tied failure times.

Recent empirical literature highlight the use of CPH in default prediction studies (see among others Bharath and Shumway 2008; Chen and Hill 2013) but it is inappropriate to use CPH model in discrete-time framework for the reasons I discuss shortly. Both, Bharath and Shumway (2008) and Chen and Hill (2013) are silent on issues pertaining to shared frailty and tied failure times, which I consider are important aspects and should be addressed in empirical studies if one choose to use CPH modelling technique.

[38] In particular we use Stata 12, -stphtest- command to perform this test for all the covariates simultaneously (globally) and for each covariates separately.
[39] In our analysis the risk set keeps on decreasing with successive failures, Efron (1977)'s method reduces the weight of contributions to the risk set from the subjects which exhibit tied event times in successive risk sets.

6.3.3.3 DISCRETE HAZARD MODEL

When an event may be experienced at any instant in continuous-time (exact censoring and survival times are recorded in relatively fine time scales such as seconds, hours or days) and there are no *tied* survival time periods, then continuous-time survival model is an appropriate choice (Rabe-Hesketh and Skrondal 2012). However, if the data has relatively few censoring or survival times with *tied* survival time periods, then discrete-time survival model is more appropriate where coarse times-scales are generally used, for instance, expressing time to event in weeks, months or years (Rabe-Hesketh and Skrondal 2012). Interval-censoring[40] leads to discrete-time data, which is the case with my database. Here, the beginning and end of each time interval is same for all the SMEs in analysis time, as the information is recorded on annual basis. Thus, the event of interest may take place at any time within the year but it cannot be known until the information is provided at the end of the year. Hence, considering the discussion above I also estimate my hazard models in discrete-time framework with *random effects* (α_i), thus controlling for *unobserved heterogeneity* or *shared frailty*.

The discrete-time representation of the continuous-time proportional hazard model with time-varying covariates leads to a generalized linear model with *complementary log-log* (Grilli 2005; Jenkins 2005; Rabe-Hesketh and Skrondal 2012) link, specified as follows:

$$cloglog\big(h_i(t)\big) \equiv \ln\{-\ln(1 - h_i(t))\} = \beta x(t)_i' + \lambda_t \qquad (5)$$

Here, λ_t is time-specific constant which is estimated freely for each time period *t*, thus making no assumption about the baseline hazard function within the specified time interval.

However, in most empirical studies logit link is used over complementary log-log (clog-log) link as specified in equation 6.

[40] The event is experienced in continuous-time but we only record the time interval within which the event takes place.

$$P_{i,t} = \frac{e^{\alpha(t)+\, x(t)_i'\beta}}{1 + e^{\alpha(t)+\, x(t)_i'\beta}} \tag{6}$$

Where *α(t)* captures baseline hazard rate and $P_{i,t}$ is the probability of experiencing the event by subject *i* at time *t*. This will produce very similar results as long as the time intervals are small (Rabe-Hesketh and Skrondal 2012) and sample bad rate (% of failed to non-failed) is very low (Jenkins 2005). One may also choose probit link function, if one strongly believes that the underlying distribution of the process being modelled is normal, or if the event under study is not a binary outcome but a proportion (e.g. proportion of population at different income levels). While these specifications will generally yield results that are quite similar but there are significant differences in terms of non-proportionality (see Sueyoshi (1995) for detiled discussion). Thus considering this discussion, I estimate my discrete hazard models with clog-log and logit links and analyse any differences in the classification performance of the models developed.

6.3.3.4 Specification of Baseline Hazard Rate

The final step before estimation of discrete-time hazard model is the specification of baseline hazard function, the hazard rate when all the covariates are set to zero. This can be done by defining time-varying covariates that bears functional relationship with survival times. Popular specifications are log(survival time), polynomial in survival time, fully non-parametric and piece-wise constant (Jenkins 2005). Duration-interval-specific dummy variables need to be created for specifying a fully non-parametric baseline hazard. The number of dummy variables needs to be equal to the maximum survival time in the dataset. For instance, if the maximum survival time is fifty years, then fifty dummy variables are required for model estimation[41] (e.g. Beck *et al.* 1998). However, this method becomes

[41] The model is run using forty nine dummies to avoid perfect multicolinearity arising from the dummy variable trap.

cumbersome if the maximum survival time in the dataset is very high as in case of bankruptcy databases. A reasonably convenient alternative way of specifying the baseline hazard function is to use piece-wise constant method. In this, the survival times are split into different time intervals that are assumed to exhibit constant hazard rate (Jenkins 2005). However, one must note that if there exist time intervals or time dummies with no events then one must drop the relevant firm-time observation with no event from the estimation or else duration specific hazard rate cannot be estimated for these time intervals/dummies (Jenkins 2005; Rabe-Hesketh and Skrondal 2012). Considering the estimation convenience one might be tempted to use the piece-wise constant specification of baseline hazard rate. However, if the hazard curve shows frequent and continuous steep rises and falls, then fully non-parametric baseline hazard is an appropriate choice.

6.3.4 Performance Evaluation

To gauge the classification performance of the models developed in identifying the distress firms I follow an approach nearly similar to Shumway (2001). After estimating the hazard models, I group the firms into deciles based on their computed event probabilities. The firms most likely to experience the distress event in the subsequent year are grouped in the first decile, the next most likely to experience the distress event in the second decile and so forth. Then for each decile I report the percentage of firms that experience the distress event. Higher the percentage of firms that experienced the distress event in the top deciles (usually top three), better is the model's classification performance.

6.4 Results and Discussion

I begin this section with the analysis of key measures of descriptive statistics of my covariates along with relevant discussion pertaining to correlation among the covariates. I perform univariate analysis of each covariate in turn using respective default definitions and econometric specification as discussed earlier to understanding any unexpected behaviour in their discriminatory performance. Then I discuss the development and performance of multivariate discrete-time hazard models using logit and clog-log links along with the baseline hazard specification. Finally, I develop multivariate extended Cox models and provide a comparative discussion on the performance of the multivariate models using different default definitions. I also illustrate the steps involved in developing various multivariate hazard models along with relevant analysis pertaining to classification performance of the models developed. To eliminate the influence of extreme outliers on my statistical estimates, I restrict the range of all my financial ratios between 5th and 95th percentiles. In addition, I lag all my covariates by one-time period so that the information is available in the beginning of the time period.

6.4.1 Descriptive Statistics and Correlation

Inspection of descriptive statistics is useful in giving us an idea about the variability of covariates and the potential biasness that may arise in the multivariate setup due to any unexpected extreme variability. I expect the mean of covariates that exhibit positive relationship with the insolvency hazard to be higher for default group than for healthy or censored group (e.g. see the variable STDEBV in Table 6.3). On the contrary, the mean of covariates that shows negative relationship with the insolvency hazard is expected to be lower for default group than their healthy counterparts (e.g. see variable CTA in Table 6.3). A closer look at Table 6.3 reveals that the mean, median and standard deviation of most of the covariates under all default definitions are as per my expectation without any extreme variability. However, EBITDAIE and STDEBV raise some serious concerns. The mean of EBITDAIE is very high, as most of the firms in my sample do not incur or incur very little interest expenses. This leads to very high differences between its mean and median values, resulting in a highly skewed distribution and very high value of standard deviation. I expect this covariate to be highly problematic in the multivariate setup. Additionally, although STDEBV is positively related to firms' default probability but the mean of the default group is lower than the censored group under all default definitions, which is quite surprising. This may lead to opposite sign of the coefficient in the multivariate setup. The mean of respective covariates across different default definitions in Table 6.3 reveal very little variation in their values. This might signal little variation in the classification performance of the multivariate models developed.

Table 6.3: Descriptive Statistics

Variable	Status Indicator	Event 1			Event 2			Event 3		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
EBITDATA	Healthy	-0.1326	0.0429	0.5328	-0.0855	0.0627	0.4932	-0.1395	0.0391	0.5431
	Unhealthy	-0.2879	-0.0309	0.6390	-0.4993	-0.1998	0.6848	-0.4721	-0.1995	0.6456
EBITDAIE	Healthy	-3761.2	1.2016	16767.4	-4049.2	1.85	17317.2	-4018.0	1.1012	17264
	Unhealthy	-5167.0	-0.5615	19167.8	-3715.1	-3.9131	16572.0	-3526.1	-3.6058	15844
STDEBV	Healthy	0.1863	0.0574	0.5777	0.1804	0.0503	0.5470	0.1789	0.0497	0.5931
	Unhealthy	0.1086	0	0.7682	0.1420	0.0126	0.8595	0.0847	0	0.8883
CTA	Healthy	0.1638	0.0775	0.2021	0.1686	0.0814	0.2079	0.1661	0.0773	0.2073
	Unhealthy	0.1585	0.0584	0.2289	0.1365	0.0494	0.1989	0.1098	0.0359	0.1884
RETA	Healthy	-3.5893	-0.2242	10.438	-3.2536	-0.1832	10.0667	-3.9052	-0.2725	10.9793
	Unhealthy	-6.4210	-0.9326	13.923	-7.8099	-1.3740	14.5994	-6.5463	-1.3060	13.218
CETL	Healthy	1.8893	1.1501	2.3645	2.0229	1.1880	2.4802	1.8488	1.1041	2.4024
	Unhealthy	1.1469	0.5311	2.2762	0.5681	0.4668	1.1176	0.3923	0.3335	0.8027
TLTA	Healthy	0.8831	0.5814	1.1617	0.8361	0.5656	1.0839	0.9268	0.5954	1.2016
	Unhealthy	1.3737	0.8562	1.4685	1.5615	0.8923	1.6366	1.5825	0.9968	1.5196
CAG	Healthy	0.0835	0.0264	0.8181	0.1086	0.0422	0.8234	0.0661	0.0174	0.8343
	Unhealthy	-0.1812	-0.1805	0.9254	-0.2813	-0.2862	0.8550	-.4002	-0.4207	0.8535
TTA	Healthy	0.0131	0	0.0286	0.0150	0.0003	0.0293	0.0128	0	0.0284
	Unhealthy	0.0066	0	0.02424	-0.0019	0	0.0144	-0.0007	0	0.0169
LCR	Healthy	0.4049	0.4855	1.0358	0.4528	0.5302	1.0310	0.3680	0.4625	1.0700
	Unhealthy	-0.0946	0.0068	1.2254	-0.2860	-0.0709	1.1240	-0.4365	-0.2455	1.0468
TCTA	Healthy	0.1501	0.0969	0.1762	0.1377	0.0893	0.1653	0.1517	0.0950	0.1816
	Unhealthy	0.1995	0.1060	0.2361	0.2543	0.1579	0.2496	0.2678	0.1710	0.2537
FETA	Healthy	0.0477	0.0270	0.0684	0.0432	0.0254	0.0613	0.0490	0.0275	0.0695
	Unhealthy	0.0704	0.0452	0.0836	0.0899	0.0549	0.0988	0.0918	0.0638	0.0923

Notes: This table reports the mean, median and standard deviation for healthy (censored) and unhealthy (firms which experienced default event) groups for respective covariates under different definitions of default events as discussed in section 6.2.

The correlation matrix presented in Table 6.4 provides evidence that some of the covariates are strongly correlated with each other. Out of the twelve covariates, FETA exhibits moderate to strong correlation with six other covariates. This is also the case with TCTA and LCR, while RETA shows strong positive correlation of approximately 0.7 with EBITDATA, supporting the fact that SMEs primarily rely on internal sources for their funding requirements. Considering the discussion presented above, I expect some volatility in my econometric estimates of the multivariate models.

Table 6.4: Correlation Matrix

Variable		1	2	3	4	5	6	7	8	9	10	11	12
EBITDATA	1	1.00											
EBITDAIE	2	0.18	1.00										
STDEBV	3	0.22	0.07	1.00									
CTA	4	-0.19	-0.34	-0.15	1.00								
RETA	4	0.66	0.17	0.21	-0.23	1.00							
CETL	6	0.19	-0.29	-0.05	0.40	0.18	1.00						
TLTA	7	-0.57	0.01	-0.26	0.00	-0.70	-0.44	1.00					
CAG	8	0.18	0.03	0.05	0.01	0.11	0.12	-0.10	1.00				
TTA	9	0.27	0.09	-0.01	0.07	0.14	0.15	-0.16	0.14	1.00			
LCR	10	0.39	-0.14	0.07	0.34	0.39	0.61	-0.61	0.17	0.24	1.00		
TCTA	11	-0.55	0.00	-0.18	-0.04	-0.53	-0.40	0.61	-0.11	-0.14	-0.52	1.00	
FETA	12	-0.49	0.16	-0.19	-0.07	-0.50	-0.40	0.72	-0.16	-0.16	-0.53	0.44	1.00

6.4.2 UNIVARIATE ANALYSIS OF COVARIATES

It is always advisable to do some univariate analysis before proceeding to estimation of multivariate models. In survival analysis the standard approach is to initially look at Kaplan-Meier survival curves of all categorical covariates to get an insight about the shape of survival functions and proportionality of each group[42]. Popular non-parametric tests of equality of survival functions like log-rank test and the Wilcoxon–Breslow–Gehan test (see Cleves *et al.* 2010) are also widely reported. However, it is not feasible to calculate Kaplan-Meier curves or conduct these non-parametric tests for continuous predictors as continuous predictors have too many different levels[43]. However, Nam *et al.* (2008) report log-rank test and the Wilcoxon–Breslow–Gehan test for their continuous predictor, which, to the best of my knowledge is inappropriate and misleading. Considering this constraint, I perform univariate regression of each covariate in turn to have an initial insight about their effects on

[42] See Cleves *et al.* (2010) for a detailed description of Kaplan-Meier curves.
[43] See for example http://www.ats.ucla.edu/STAT/stata/seminars/stata_survival/default.htm. Also see Cleves *et al.* (2010) for a more thorough understanding.

respective dependent variables. Table 6.5 reports the results I obtain from the univariate regressions.

Table 6.5: Univariate Test

Variable	Event 1			Event 2			Event 3		
	logit	clog-log	CPH	logit	clog-log	CPH	logit	clog-log	Cox
EBITDATA	-.6117***	-.5246***	-0.4197***	-1.1243***	-.7939***	-0.5357***	-.8688***	-.7900***	-0.9669***
EBITDAIE	-3.26e-06	-3.58e-06	7.2e-06***	3.23e-06	2.94e-06	7.8e-06***	3.36e-06	3.07e-06	0.0000**
STDEBV	-.2711***	-.2607***	-0.0976	-.0448	-.0545	0.1494***	-.2981***	-.3025***	-0.1369
CTA	-.2122	-.1392	-0.8235***	-1.4751***	-1.2455***	-1.436***	-1.8492***	-1.7897***	-2.31***
RETA	-.0247***	-.0218***	0.0003	-.0210***	-.0157***	0.0157***	-.0187***	-.0176***	-0.0019
CETL	-.2367***	-.2287***	-0.2623***	-.7164***	-.6370***	-0.4201***	-1.0675***	-1.0061***	-1.094***
TLTA	.3272***	.2694***	0.1767***	.3054***	.2204***	0.0246	.3026***	.2755***	0.2068***
CAG	-.4417***	-.3995***	-0.2876***	-.6159***	-.5129***	-0.3402***	-.7919***	-.7476***	-0.6725***
TTA	-11.511***	-10.185***	-7.024***	-40.738***	-34.681***	-32.18***	-27.543***	-26.078***	-25.15***
LCR	-.4996***	-.4393***	-0.4281***	-.6469***	-.5171***	-0.3358***	-.7069***	-.6514***	-0.8023***
TCTA	1.6774***	1.4560***	1.444***	2.4727***	1.9016***	0.962***	2.5567***	2.3603***	2.806***
FETA	5.1395***	4.3400***	3.539***	6.6515***	4.9264***	1.857***	6.6758***	6.1383***	6.629***

Notes: *** (**) [*] significant at the 1 % (5 %) [10 %] level (two-sided test). This table reports the coefficients obtained from univariate regression analysis of respective covariates for different default definitions (Event 1, Event 2 and Event 3) as discussed in section 6.2. For each default event this table report coefficients estimated using discrete-time duration-dependent hazard function (with logit and clog-log links respectively) and Extended Cox model.

I observe from Table 6.5 that for all three distress events, identical set of covariates display significant discriminatory power when estimated using discrete-hazards models (logit and clog-log). However, there is variation when estimation is done using extended Cox model. In particular, EBITDAIE (Event 1, 2 and 3), STDEBV (Event 1, 2 and 3), CTA (Event 1), RETA (Event 1 and 3) and TLTA (Event 2) shows contradictory effects on firms' insolvency hazard when estimation is done using discrete-hazard and Cox model respectively. Supporting my discussion on EBITDAIE and STDEBV in section 6.4.1, we see that the discriminatory power of both these variables is sensitive to varying default definitions. EBITDAIE exhibits significant discriminatory power only when estimated using extended Cox model, while STDEBV is significant in most cases but the sign of the coefficients are negative (except for Cox estimation for Event 2) which is expected to be positive. The variable RETA also shows mixed sign of coefficients under different distress definitions. I expect these covariates to have deterrent effects in the multivariate setup. In addition, multicollinearity among the covariates may also influence their significance or sign in the multivariate models.

6.4.3 Developing Multivariate Hazard Models

In this section of the paper, I develop and discuss multivariate hazard models developed for my respective default definitions. I begin with the choice of my specification for the baseline hazard rate, which is required for developing discrete-time duration-dependent hazard models, followed by development and discussion of discrete-time and continuous-time hazard models. I exclude FETA from all my multivariate models as it exhibits moderate to strong correlation with six other covariates (see Table 6.4 for more details) and affects the sign of other covariates when entered into the multivariate setup. Considering the enormously high value of standard deviation and skewness of EBITDAIE, I also exclude it from all multivariate models developed. This is further justified in the univariate analysis given that it is only significant when estimated using extended Cox model and the values the coefficients generated are virtually zero (see Table 6.5 for more details).

6.4.3.1 Detection of Baseline Hazard Rate

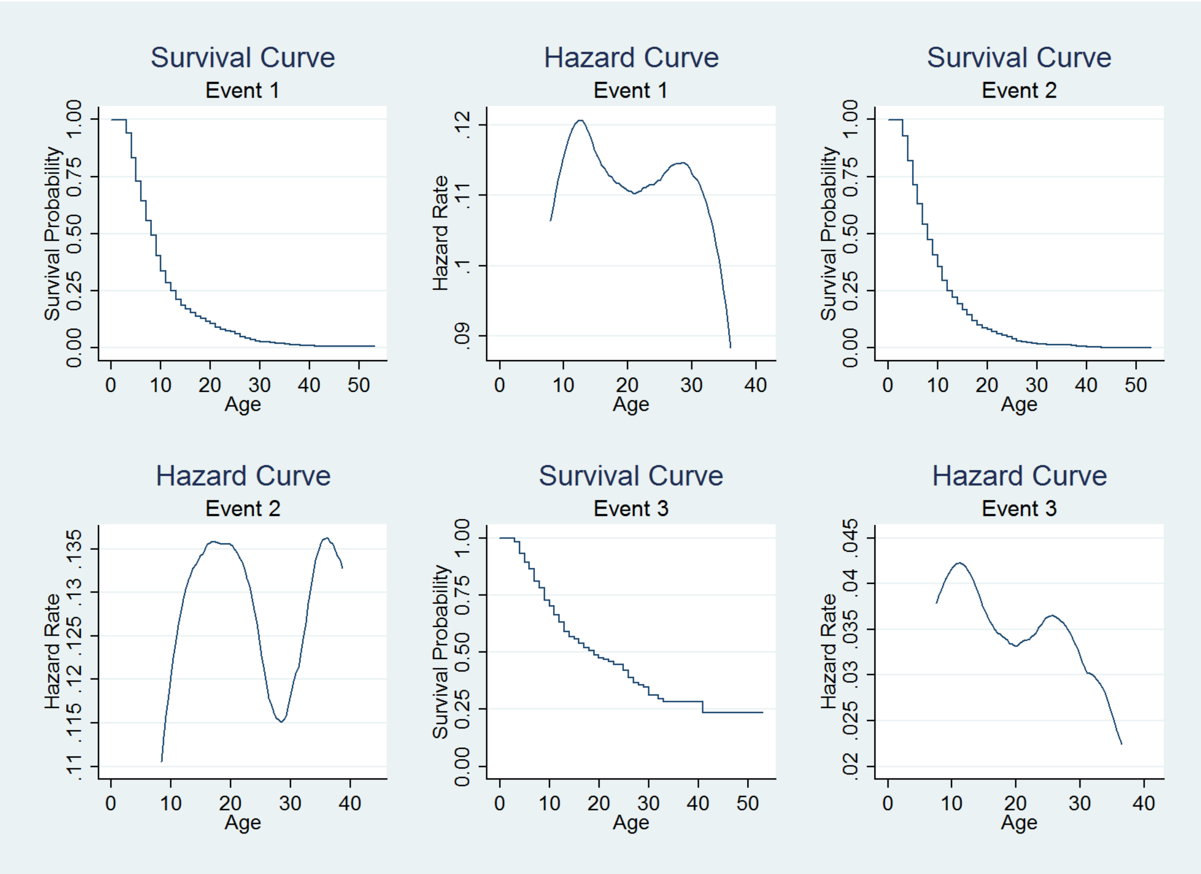
Before developing multivariate discrete-time hazard models it is important to choose a baseline specification for the hazard rate. Table 6.6 shows the table of survival and hazard curves estimated using Kaplan-Meier estimator for different distress events. [44] The survival curves of Event 1 and Event 2 are almost identical, while for Event 3 it's slightly different. The survival probability under Event 1 and 2 grazes toward zero as firms' age approaches forty, while it freezes around 0.25 under Event 3. This gives me an initial motivation to believe that the survival attributes under Event 3 may be different from the other two events.

[44] See among others Cleves *et al.* (2010) and Mills (2011) for details regarding Kaplan-Meier estimator.

However, my prime interests are the hazard curves[45]. Although, we see some similarity in the survival curves but the hazard curves for all three events exhibit different functional relationship with firms' age. Thus, for all three different distress events different baseline hazard rate specifications are required under piecewise-constant method. Additionally, all three hazard curves show steep rises and falls, thus it's inappropriate to assume the hazard rate to be constant for any defined age group. Under this situation I think it's appropriate to go for fully non-parametric baseline hazard specification, and thus I use age specific dummy variables to specify the baseline hazard rate. Table 6.1 lists the age-wise distribution of firm year-observations in my sample. Here, we see that the minimum age of a firm in my sample is 3 years, while the maximum age is 53 years. Thus I create 51 age specific dummies that represent respective age category. Further, we also see in Table 6.1 that there are no events for some age dummies under all three respective events. Thus, considering my discussion in section 6.3.3 I exclude firm-year observation for age dummies having no events under respective default definitions from my multivariate models.

[45] Table 6.1 shows that, the earliest age that a firm experiences distress event under all three default definitions is three years. However, the hazard curves start somewhere around seven years. This difference is due to the fact that "sts graph" command in Stata performs an adjustment of the smoothed hazard near the boundaries. In case of the default kernel function of -sts graph- (Epanechnikov kernel), the plotting range of the smoothed hazard function is restricted to be within one bandwidth of each endpoint. The same is true for other kernels, except the epan2, biweight, and rectangular kernels, in which case the adjustment is performed using boundary kernels. If we wish to plot an estimate of the hazard for the entire range, we could use a kernel without a boundary correction. Alternative, we can use then -noboundary- option, but this will produce an estimate that is biased near the edges. See "help sts graph" in Stata and Silverman (1986) for further details. However, this will not affect our empirical analysis as we use fully non-parametric method of baseline hazard specification. However, one needs to be little careful while using piecewise-constant specification.

Table 6.6: Survival and Hazard Curves



Notes: This table shows survival and hazard curves estimated using the development sample for different definitions of financial distress events as discussed in section 6.2. Here 'Age' represents the age of firms in years.

6.4.3.2 Hazard Models for Event 1

The binary dependent variable used is Event 1, i.e. firms that have filed for legal bankruptcy proceedings are considered to have experienced the default event and censored otherwise (please see section 6.2 for detailed discussion). I estimate the models using 645 defaulted and 3,283 censored firm-year observations, which exclude observation with age dummies having no events (see Table 6.1). The discrete-time duration-dependent hazard models developed using logit and clog-log links and continuous-time duration-dependent Extended Cox models are reported in Part A of Table 6.7. In developing the multivariate models I employ all the covariates found significant in the univariate analysis section other than EBITDAIE and FETA (excluded due to reasons discussed earlier). All the significant covariates under respective econometric specification bear expected sign of the coefficients except TLTA under Cox model, which might be due to multicollinearity among the covariates. A comparison across the hazard models reveal that some of the covariates significant in the univariate section (EBITDATA, CTA and TCTA) fail to exhibit significant discriminatory power in the multivariate setup. Further, an identical set of covariates explains the insolvency hazard of Event 1 under all econometric specifications, with little variation in the weight of the coefficients of respective covariates. This leads me to assume that Event 1 is not sensitive to my econometric specifications. However, I report very high differences between the values of Akaike information criterion (AIC) of discrete-hazard models and Cox model. [46] All my discrete hazard models show value of AIC around 3290, while Cox model gives a value of 8918.48. Thus considering the AIC values, discrete-time models provide a better fit than Cox model. In terms of classification performance we do not see any difference between the three models as all of them encapsulate about 52 % of distressed firms in the top three deciles (see Part A of Table 6.7).

[46] The AIC is used to test whether we have the appropriate model fit between the competing non-nested statistical models. This simple rule is that, lower the value of AIC better is the model's fit (see Mills (2011) for details).

Table 6.7: Multivariate Hazard Models

Part A		Event 1					
Variable	Expected Sign	logit		clog-log		Extended Cox	
		Coefficient	SE	Coefficient	SE	Coefficient	SE
EBITDATA	-	-.1348	.1244	-.0963	.1055	-.0762	.1159
EBITDAIE	-	---	---	---	---	---	---
STDEBV	+	---	---	---	---	---	---
CTA	-	---	---	---	---	.2834	.2917
RETA	-	.0027	.0070	.0014	.0061	---	---
CETL	-	-.0767**	.0348	-.0761**	.0320	-.1668***	.0377
TLTA	+	.1324**	.0658	.1008*	.0557	-.1010**	.0524
CAG	-	-.3044***	.0606	-.2628***	.0527	-.1628***	.0515
TTA	-	-6.6812***	2.1820	-5.9211***	1.9463	-3.8241**	2.0812
LCR	-	-.2674***	.0708	-.2333***	.0614	-.2634***	.0700
TCTA	+	-.0753	.3423	-.0374	.2877	.3587	.3149
FETA	+	---	---	---	---	---	---
Age Dummies						---	---
Goodness of Fit		Value	p-value	Value	p-value	Value	p-value
Wald chi2		181.59	0.0000	192.48	0.0000	121.31	0.0000
Log likelihood		-1599.6038		-1597.3004		-4451.2402	
AIC		3293.208		3288.601		8918.48	
BIC		3588.174		3583.567		8968.687	
Number of observations		3,928		3,928		3,928	
	Default	645		645		645	
	Censored	3,283		3,283		3,283	
Decile	1	2	3	4	5	6 – 10	Total
logit %	20.47(20.47)	17.83(38.30)	13.33(51.63)	11.63(63.26)	9.15(72.41)	27.59(100)	100
clog-log %	21.09(21.09)	17.05(38.14)	13.64(51.78)	11.47(63.25)	9.61(72.86)	27.13(100)	100
Cox %	20.00(20.00)	18.91(38.91)	13.64(52.55)	9.31(61.86)	8.07(69.93)	30.07(100)	100
Part B		Event 2					
EBITDATA	-	-.8695***	.1216	-.6035***	.0903	-.3897***	.0875
EBITDAIE	-	---	---	---	---	---	---
STDEBV	+	---	---	---	---	.1618***	.0592
CTA	-	-.3279	.3428	-.2788	.2715	-1.0380***	.2734
RETA	-	.0214***	.0074	.0156***	.0057	---	---
CETL	-	-.4678***	.0691	-.4362***	.0617	-.3299***	.0552
TLTA	+	-.1044	.0667	-.1071**	.0517	---	---
CAG	-	-.3259***	.0615	-.2254***	.0493	-.2282***	.0471
TTA	-	-36.282***	3.5275	-31.279***	2.8684	-29.698***	3.2347
LCR	-	-.1494**	.0787	-.1135*	.0643	.0412	.0649
TCTA	+	.2966	.3301	.2281	.2518	-.1983	.2658
FETA	+	---	---	---	---	---	---
Age Dummies						---	---
Goodness of Fit		Value	p-value	Value	p-value	Value	p-value
Wald chi2		375.22	0.0000	415.12	0.0000	236.75	0.0000
Log likelihood		-1421.6452		-1430.3595		-4592.5542	
AIC		2939.29		2956.719		9201.108	
BIC		3240.226		3257.655		9251.264	
Number of observations		3,903		3,903		3,903	
	Default	692		692		692	
	Censored	3,211		3,211		3,211	
Decile	1	2	3	4	5	6 – 10	Total
logit %	27.31(27.31)	21.97(49.28)	19.80(69.08)	12.86(81.94)	7.95(89.89)	10.11(100)	100
clog-log %	27.46(27.46)	22.11(49.57)	18.50(68.07)	13.15(81.22)	7.51(88.73)	11.27(100)	100
Cox %	26.16(26.16)	20.52(46.68)	18.06(64.74)	13.58(78.32)	9.10(87.42)	12.58(100)	100
Part C		Event 3					
EBITDATA	-	-.5937***	.1792	-.5096***	.1589	-.6578***	.1805
EBITDAIE	-	---	---	---	---	---	---
STDEBV	+	---	---	---	---	---	---
CTA	-	-.9052*	.5546	-.8520*	.5063	-1.1617**	.5342
RETA	-	.0263**	.0113	.0225**	.0102	---	---
CETL	-	-.8243***	.1704	-.7918***	.1589	-.9085***	.1690
TLTA	+	-.1067	.1020	-.1057	.0929	-.3603***	.0839

CAG	-	-.3517***	.0977	-.2988***	.0873	-.2810***	.0889
TTA	-	-22.166***	4.9345	-20.470***	4.5646	-21.548***	5.3546
LCR	-	-.1708	.1208	-.1458	.1112	-.2428*	.1284
TCTA	+	.3960	.4702	.3339	.4165	.4991	.4657
FETA	+	---	---	---	---	---	---
Age Dummies						---	---
Goodness of Fit		Value	p-value	Value	p-value	Value	p-value
Wald chi2		153.80	0.0000	171.90	0.0000	153.79	0.0000
Log likelihood		-689.02993		-689.79272		-1427.0373	
AIC		1458.06		1459.585		2870.075	
BIC		1707.665		1709.19		2919.996	
Number of observations		3,790		3,790		3,790	
	Default	219		219		219	
	Censored	3,571		3,571		3,571	
Decile	1	2	3	4	5	6 – 10	Total
logit %	38.36(38.36)	24.20(62.56)	12.78(75.34)	11.42(86.76)	3.65(90.41)	9.59(100)	100
clog-log %	37.44(37.44)	25.57(63.01)	12.33(75.34)	10.50(85.84)	4.57(90.41)	9.59(100)	100
Cox %	30.59(30.59)	25.57(56.16)	16.43(72.59)	10.96(83.55)	5.47(89.02)	10.98(100)	100

Notes: *** (**) [*] significant at the 1 % (5 %) [10 %] level (two-sided test). Part A, Part B and Part C of this table reports estimations corresponding to Event 1, Event 2 and Event 3 respectively. For each Part (A, B and C) this table reports the results obtained from respective multivariate hazard analysis (logit, clog-log and Extended Cox) followed by goodness of fit measures and classification performance measures as discussed in section 6.4.3. Values in the parenthesis are cumulative classification measures over the ten deciles.

6.4.3.3 Hazard Models for Event 2

The definition of my binary predictor is based on firms' financial health in terms of its earnings and net worth (see section 6.2 for details). To develop the multivariate models I employ all the covariates found significant in explaining Event 2 financial distress in the univariate section for each respective econometric specification. This takes us to Part B of Table 6.7, which reports the respective estimated results, goodness of fit measures and classification performance measures. I estimate the models using 692 defaulted and 3,211 censored firm-year observations that exclude observation having age dummies with no events (see Table 6.1). All significant covariates in respective hazard models bear the expected sign of the covariates except RETA and TLTA under logit and clog-log specifications, which might be due to multicollinearity among the covariates. A closer look at the table also reveals variation in the significance and coefficients' weight of the covariates across all econometric specifications. CTA is significant only under Cox specification, while LCR is significant only under discrete-time specification. The value of AIC is around 2950 for discrete-time models but for Cox model it's around 9200, which emphasise that discrete-time hazard models provide a better fit to my sample. The percentage of distressed firms classified in the top three deciles is around 69% for both discrete-hazard models but for Cox model it's slightly lower at about 65%. Based on my reported results, I understand that both my discrete-time specifications exhibit identical classification performance, while Cox specification performs marginally poorer.

6.4.3.4 HAZARD MODELS FOR EVENT 3

The set of hazard models that I estimate is based on the default definition (Event 3) that I propose in this study, which considers both legal bankruptcy filing and firms' financial health while classifying SMEs as default (please see section 6.2 for details). The econometric models are estimated using 219 defaulted and 3,591 censored firm-year observations that exclude observations with age dummies having no events (see Table 6.1). I employ all covariates found significant in respective univariate hazard analysis to develop the multivariate hazard models. Part C of Table 6.7 reports the estimated results. All significant covariates in the respective hazard models bear the expected sign of the covariates except RETA in discrete-time models and TLTA in Cox model. The weights of the coefficients of respective covariates in respective hazard models show very little variation, which indicates little variation in their respective classification performance measures. The logit and clog-log exhibits identical classification performance of about 75%, while Cox model performs slightly lower at about 73%. The AIC values of both the discrete-time models are identical at around 1460, however Cox model shows almost double value of AIC of around 2870. Thus, here also the discrete-time models provide superior model fit than their continuous counterpart.

6.4.3.5 Distress Definitions and Performance of Hazard Models

As reported in Table 6.7, the extended Cox model performs marginally poorer than discrete-time models with logit and clog-log links as its shows marginally poorer classification performance across all default definitions. However, the interesting thing to note is the classification performance measures across different default definitions. Based on the classification of defaulted firms in the top three deciles, Event 1 is the weakest definition of default while Event 3 is the strongest. The default definition based on legal bankruptcy filing classifies about 52% of defaulted firms in the top three deciles, while the distress definition based on firms' financial heath (Event 2) encapsulates about 69% in the top three deciles. However, the default definition that I propose which is based on both legal bankruptcy filing and SMEs' financial health is superior. It identifies about 75% of the firms in the top three deciles, which is 44% higher than Event 1 and about 9% higher than Event 2. The AIC measure of Event 3 models are the lowest among the three default definitions, which indicates that Event 3 default definition provides a vastly improved fit than the other two default definitions.

6.5 Conclusion

The growing popularity of hazard models in making bankruptcy prediction motivated me to undertake this empirical investigation. Almost every study in the empirical literature suffers from at least one of the following issues: (i) reason behind their choice between *discrete-time* or *continuous-time* hazard model (ii) inappropriate specification of baseline hazard rate (iii) no test of proportional hazards assumption when using *Extended Cox* model with time-independent covariates (iv) ignore *frailty* and *recurrent* events (v) explanation on how they dealt with the issues of *delayed entry* (vi) explanation on treatment of time periods/intervals having no events. Therefore, I contribute to the literature by acknowledging all these serious and neglected concerns in my study and intend to be the first academic paper to report performance comparison of popular hazard models (discrete hazard models with logit and clog-log links and extended Cox model) used in the recent literature (e.g. Campbell *et al.* 2008, Chen and Hill 2013). I also contribute to the literature by undertaking an empirical investigation which compares various default definitions of US SMEs. Three default definitions that I compare are based on legal bankruptcy laws (Event 1), firms' financial health (Event 2) and the third definition (Event 3) that I propose in this study considers both legal bankruptcy and firms' financial health.

My empirical results highlight almost identical classification performance of both discrete hazard models across all three default definitions, while Cox model performs marginally poorer than their discrete counterparts across all default categories. Based on my comparison of AIC measures, discrete hazard models provide considerably superior fit than Cox model. However, the AIC measures for both discrete-time hazard models (logit and clog-log links) are almost identical; hence the choice between them is left on the personal preference of the users.

Comparison of default definitions leads me to a striking conclusion. Based on the classification performance of the models developed using different default definitions, I understand that the default definition that I propose performs best in identifying distressed firms. This emphasises the fact that, significant number of firms choose legal bankruptcy routes as part of their planned exit strategy.

Given the importance of hazard function models in predicting bankruptcy in light of the financial crises, and the robustness of my results in dealing with neglected econometric issues in all previous empirical research in survival analysis, this study cannot be ignored.

7. CONCLUSION

This thesis presents four essays on SMEs insolvency risk. The primary objective of my *first essay* is to examine the incremental information content of operating cash flow information in predicting bankruptcy of UK SMEs. To examine this, I develop one year failure prediction model using the significant financial ratios obtained from income statement and balance sheet, along with significant operating cash flow ratios obtained from cash flow statement. Empirical evidence pertaining to trade credit and capital structure of SMEs motivated me to believe that, operating cash flow information could add significant discriminatory power to the models developed using accrual ratios obtained from income statement and balance sheet.

One year default prediction models (SME1 and SME2) were developed using a sample (with non-missing data) of 116,212 UK SMEs that survived in the period of 2000 to 2009 and 2,666 firms that failed in the same time period. The data of analysis year 2008 and 2009 have been retained as a test sample (hold-out sample). SME1 model corresponds to the model developed using significant financial ratios obtained from income statement and balance sheet, while SME2 model employs significant operating cash flow ratios as an enhancement to SME1 model.

Although, all the operating cash flow ratios exhibit significant discriminatory power in the univariate analysis, but test result shows that only one of my operating cash flow ratios (cash flow from operation/current liabilities; CFOCL) exhibit statistically significant discriminatory power in identifying failed and non-failed firms in the multivariate setup. However, classification accuracy measures obtained for SME1 and SME2 models are identical for within sample and hold-out sample, which motivate me to believe that the policymakers and lending institutions may not gain significant benefit in understanding the credit risk behaviour of SMEs by analysing an additional set of financial statement (i.e. cash flow statement).

My findings clearly confirm that operating cash flow information does not improve the prediction performance of the default models, as both SME1 and SME2 models exhibit identical classification performance measures. Gaining access to operating cash flow information for SMEs is a real challenge as firms are not obliged by law to submit cash flow statement. Hence considering my finding I do not see any marginal gain in understanding the credit risk behaviour of SMEs by analysing information obtained from cash flow statement.

My *second essay* investigates the effect of internationalisation on modelling credit risk for UK SMEs. Following Fatemi (1988), I classify a firm as international if it makes sales abroad and domestic if it makes sales only in the domestic market. The empirical literature on the performance of international SMEs is somewhat contradictory, which motivate me to undertake this study. Ramaswamy (1992) reports that international SMEs exhibit lower risk due to revenue and cash flow diversification, while Michael et al. (2009) report that international SMEs exhibit higher default probability due to exposure to multiple political and financial environments. To examine the impact of internationalisation on the default propensity of SMEs, I estimate separate default prediction models for domestic and international firms using a set of financial ratios.

I develop one-year distress prediction models using a dynamic logistic regression technique, and implement appropriate measures to control for the effect of macroeconomic conditions. The unique database available to me from the Credit Management Research Centre of the University of Leeds contains financial information of 342,711 domestic SMEs (with 8,525 defaulted and 334,186 non-defaulted firms) and 344,205 international SMEs (with 9,114 defaulted and 335,091 non-defaulted firms) ranging over an analysis period of 2000 to 2009. I retain the data of analysis year 2008 and 2009 as a hold-out sample.

My empirical findings are somewhat mixed. In my multivariate models, all the factors which affect the default probability of international SMEs are also highly significant in explaining the default probability of domestic SMEs, except short-term debt/equity book value. Furthermore, all the variables capturing the impact of exports on default probability of international firms are highly insignificant in the univariate analysis, thus contradicting the suggestion of Arslan and Karan (2009) to consider domestic and international firms separately while modelling their credit risk behaviour. However, the predictive accuracy measures obtained by employing the same set of variables are lower for international SMEs than for their domestic counterparts. Chi-square tests performed to compare the weights of regression coefficients of the models developed, confirm that the coefficients of four out of the nine common predictors (CTA, CETL, TTA and TCTL) exhibit significant statistical difference. I make a further significant contribution by being the only study to measure the impact of intangible assets on the defaults probability of SMEs. My test results confirm that the ratio intangible assets/total assets (IATA) is highly significant in assessing credit risk for both domestic and international SMEs.

My findings clearly show that almost the same set of factors affect the default probability of both the groups, hence there is no potential need to treat domestic and international SMEs separately while modelling credit risk. This indifference may be due to the recent effort undertaken by the policy makers and business community to understand and mitigate the factors adversely affecting the export performance of small firms (Secretariat 2009).

However, in view of the low predictive power of the model developed for international SMEs, I suggest that modelling credit risk for international SMEs would benefit from further work to understand the inherent complexities. Non-financial factors may play an important role in understanding their credit risk behaviour. In particular the effect of changing government policies, firm specific non-financial characteristics, and changing macroeconomic conditions may play an important role in understanding their credit risk behaviour. These are possible avenues for further research in the field of modelling credit risk behaviour of SMEs.

My *third essay* considers the huge diversity that exists within the broad SMEs category (micro, small, and medium firms) in the form of capital structure (Ramalho and Da Silva 2009; Mateev et al. 2013), access to external finance (Beck *et al.* 2006), management style (Wager 1998), default probability (Pettit and Singer 1985), firm size, number of employees etc. In this essay, I investigate the financial and non-financial factors influencing SMEs failure propensity, in order to identify any differences that may exist within this broad SMEs category. I apply discrete-time duration-dependent hazard rate modeling techniques to develop separate bankruptcy prediction models for micro, small, and medium firms respectively, using a relatively large database of UK firms. I compare their performance with the model developed for SMEs as a whole, which includes micro, small, and medium firms. I estimate the insolvency hazard models after taking account of correlation among the covariates. Finally, I compare the estimated models (for micro, small, and medium firms) with my SMEs model in turn, to identify the common default attributes.

To undertake the statistical estimations, I use a heterogeneous-panel that contains financial and non-financial information of 8,162 failed and 385,733 non-failed UK SMEs covering the analysis period between 2000 and 2009. The data between analysis year 2000 to 2007 is used as my development sample, while I retain the data of analysis year 2008 and 2009 as a hold-out sample to validate out-of-sample prediction performance of the models developed. To validate the robustness of the models developed I estimate the area under the receiver operating characteristics curves (AUROC) for hold-out samples.

All the multivariate models demonstrate fairly strong classification performance with AUROC of about 0.77 for holdout samples except the model developed for micro firms; which shows slightly lower AUROC of approximately 0.74. Further comparison of default prediction models for micro firms and SMEs strongly suggest that they need to be considered separately while modeling credit risk for them. Three of the financial ratios reported significant in Altman and Sabato (2007) are insignificant in my micro model but significant in my SMEs model. The variables capturing working capital, and short term leverage position (trade creditors/total assets and trade debtors/total assets), are highly significant in line with prior empirical findings (Hudson 1986, Beck *et al.* 2006). Given that a near identical set of explanatory variables affect the default probability of small firms and SMEs, I do not expect a material impact on the decision making process by treating each of these groups separately. Finally, I make comparisons between hazard models developed for medium firms and SMEs. Once again almost the same set of explanatory variables are highly significant in both the models, hence I suggest that considering both the groups separately may not lead to superior risk pricing.

Based upon my findings, I see that financial reports do not provide sufficient information about the likelihood of default of micro firms and hence there is a clear need of separate treatment, as globally the vast majority of business units are very small (Di Giovanni et al. 2011) with median business units having no employees.

My *fourth essay* is motivated from the growing popularity of hazard models in making bankruptcy prediction. Almost every study in the empirical literature suffers from at least one of the following issues: (i) reason behind their choice between *discrete-time* or *continuous-time* hazard model (ii) inappropriate specification of baseline hazard rate (iii) no test of proportional hazards assumption when using *Extended Cox* model with time-independent covariates (iv) ignore *frailty* and *recurrent* events (v) explanation on how they dealt with the issues of *delayed entry* (vi) explanation on treatment of time periods/intervals having no events. Therefore, I contribute to the literature by acknowledging all these serious and neglected concerns in my study and intend to be the first academic study to report performance comparison of popular hazard models (discrete hazard models with logit and clog-log links and extended Cox model) used in the recent literature (e.g. Campbell *et al.* 2008, Chen and Hill 2013). I also contribute to the literature by undertaking an empirical investigation which compares various default definitions of US SMEs. Three default definitions that I compare are based on legal bankruptcy laws (Event 1), firms' financial health (Event 2) and the third definition (Event 3) that I propose in this study considers both legal bankruptcy and firms' financial health.

My empirical results highlight almost identical classification performance of both discrete hazard models across all three default definitions, while Cox model performs marginally poorer than their discrete counterparts across all default categories. Based on my comparison of AIC measures, discrete hazard models provide considerably superior fit than Cox model. However, the AIC measures for both discrete-time hazard models (logit and clog-log links) are almost identical; hence the choice between them is left on the personal preference of the users.

Comparison of default definitions leads me to a striking conclusion. Based on the classification performance of the models developed using different default definitions, I understand that the default definition that I propose performs best in identifying distressed firms. This emphasises the fact that, significant number of firms choose legal bankruptcy routes as part of their planned exit strategy.

Given the importance of hazard function models in predicting bankruptcy in light of the financial crises, and the robustness of my results in dealing with neglected econometric issues in all previous empirical research in survival analysis, this study cannot be ignored.

8. Suggestions for Future Research

My *first essay* asserts that the presence of operating cash flow information does not lead to superior classification performance measures of UK SMEs. However, considering the huge diversity that exists within the broad SMEs category (micro, small, and medium firms) in the form of capital structure (Ramalho and Da Silva 2009; Mateev et al. 2013), access to external finance (Beck *et al.* 2006), management style (Wager 1998), default probability (Pettit and Singer 1985), firm size, number of employees etc., it would be an worth exercise to revisit this study considering micro, small and medium-sized firms separately while modelling their default probability. As, larger the firm size, the less access to external finance is seen as a problem (Beck *et al.* 2006). Hence, the discriminatory power of operating cash flow information might also vary within the SMEs category. Besides, it will also be useful to conduct this study on different economies that might also draw comparison between developed and developing economies.

Considering the low predictive power of the model developed for international SMEs in my *second essay*, I suggest that the model for international SMEs need further inspection to understand the inherent complexities. Non-financial factors may play an important role in understanding their credit risk behavior. In particular the effect of changing government policies, firm specific non-financial characteristics and changing macroeconomic dimensions may play an important role in understanding their credit risk behavior. These are very interesting avenues for further research in the field of modeling credit risk behavior of SMEs.

My *third essay* asserts that financial reports do not provide sufficient information about the likelihood of default of micro firms and hence there is a clear need of separate treatment, as globally the vast majority of business units are very small (Di Giovanni et al. 2011) with median business units having no employees. I believe further firm level qualitative additions, such as owner characteristics, business type, business location, bank relationship history, etc. may work as a better qualifier for micro firms. Although the same set of predictors do affect the default probability of both small firms, medium firms and SMEs, but their impacts may vary. Thus it would be interesting to compare the factors affecting the failure propensity of medium firms and large firms using alternative methodology as highlighted in Merton (1974), since many of the medium firms are traded in the financial markets. My results thus support the hypothesis that the credit risk characteristics of firms within the broad SMEs category do vary, and hence I suggest treating them separately for better pricing of credit risk and estimation of default probability. However my hypothesis may get stronger support if tested in developing economies like India, China, Brazil etc., as there exist much greater structural differences and diversity within the broad SMEs category unlike developed economies such as the USA and UK.

My *fourth essay* considers all serious and neglected concerns while developing discrete and continuous time duration dependent hazard models for predicting default of US SMEs. This study might further be extended in context of large firms that are traded in financial markets, as vast majority of the academic literature use hazard models to predict insolvency hazard of large firms. Additionally, my proposed default definition for SMEs based on legal bankruptcy laws and firms' financial health performs significantly better than their alternative counterparts in identifying distressed firms, with superior goodness of fit measures across all econometric specifications. It would also be useful to explore the stability of my proposed default definition across other developed and emerging economies.

BIBLIOGRAPHY

Acs, Z.J., Morck, R.K., and Yeung, B., 2001. Entrepreneurship, globalization, and public policy. *Journal of International Management*, 7 (3), 235–251.

Alam, S.S., 2011. Assessing barriers of growth of food processing SMIs in Malaysia: a factor analysis. *International Business Research*, 4 (1), 252.

Allen, L., DeLong, G., and Saunders, A., 2004. Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance*, 28 (4), 727–752.

Allen, L. and Pantzalis, C., 1996. Valuation of the operating flexibility of multinational corporations. *Journal of International Business Studies*, 27 (4), 633–653.

Allison, P.D., 2010. *Survival analysis using SAS: A practical guide*. 2nd ed. SAS Institute.

Allison, P.D., 2014. *Event History and Survival Analysis: Regression for Longitudinal Event Data*. 2nd ed. USA: Sage.

Alsaaty, F.M., 2013. The Business Density of Employer Micro Firms in the United States. *Research in Business & Economics Journal*, 7, 157.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23 (4), 589–609.

Altman, E.I. and Hotchkiss, E., 2006. *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt*. 3rd ed. USA: Wiley. com.

Altman, E.I. and Sabato, G., 2007. Modelling credit risk for SMEs: Evidence from the US market. *Abacus*, 43 (3), 332–357.

Altman, E.I., Sabato, G., and Wilson, N., 2010. The value of non-financial information in small and medium-sized enterprise risk management. *Journal of Credit Risk*, 2 (6), 95–127.

Mac An Bhaird, C. and Lucey, B., 2010. Determinants of capital structure in Irish SMEs. *Small Business Economics*, 35 (3), 357–375.

Anderson, R., 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. 1st ed. OUP Oxford.

Arslan, Ö. and Karan, M.B., 2009. Credit risks and internationalization of SMEs. *Journal of Business Economics and Management*, 10 (4), 361–368.

Astrakhan, I. and Chepurenko, A., 2003. Small business in Russia: any prospects after a decade? *Futures*, 35 (4), 341–359.

Ayyagari, M., Beck, T., and Demirguc-Kunt, A., 2007. Small and medium enterprises across the globe. *Small Business Economics*, 29 (4), 415–434.

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

Bates, T., 2005. Analysis of young, small firms that have closed: delineating successful from unsuccessful closures. *Journal of Business Venturing*, 20 (3), 343–358.

Batini, N., Levine, P., Kim, Y.B., and Lotti, E., 2010. *Informal labour and credit markets: a survey*. IMF Working Papers.

Bauer, J. and Agarwal, V., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432–442.

Beaver, W.H., 1966. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111.

Becchetti, L. and Sierra, J., 2003. Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking & Finance*, 27 (11), 2099–2120.

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), 1260–1288.

Beck, T. and Demirguc-Kunt, A., 2006. Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, 30 (11), 2931–2943.

Beck, T., Demirgüç-Kunt, A., Laeven, L., and Maksimovic, V., 2006. The determinants of financing obstacles. *Journal of International Money and Finance*, 25 (6), 932–952.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V., 2008. Financing patterns around the world: Are small firms different? *Journal of Financial Economics*, 89 (3), 467–487.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V., 2005. Financial and legal constraints to growth: does firm size matter? *The Journal of Finance*, 60 (1), 137–177.

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), F32–F53.

Bernard, V.L. and Stober, T.L., 1989. The nature and amount of information in cash flows and accruals. *Accounting Review*, 64 (4), 624–652.

Bharath, S.T. and Shumway, T., 2008. Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21 (3), 1339–1369.

BIS, 2012. *Business Population Estimates for the UK and Regions 2012*.

Bosma, N.S. and Levie, J., 2010. *Global Entrepreneurship Monitor 2009 Executive Report.* Global Entrepreneurship Research Association.

Box-Steffensmeier, J.M. and De Boef, S., 2006. Repeated events survival models: The conditional frailty model. *Statistics in Medicine*, 25 (20), 3518–3533.

Box-Steffensmeier, J.M. and Zorn, C., 2002. Duration models for repeated events. *The Journal of Politics*, 64 (04), 1069–1094.

Burgman, T.A., 1996. An empirical examination of multinational corporate capital structure. *Journal of International Business Studies*, 27 (3), 553–570.

Campbell, J.Y., Hilscher, J., and Szilagyi, J., 2008. In search of distress risk. *The Journal of Finance*, 63 (6), 2899–2939.

Caouette, J.B., Altman, E.I., Narayanan, P., and Nimmo, R., 2008. *Managing credit risk: the great challenge for global financial markets*. Wiley.

Carter, R. and Auken, H. V, 2006. Small firm bankruptcy. *Journal of Small Business Management*, 44 (4), 493–512.

Carter, S., Mason, C., and Tagg, S., 2006. *Lifting the barriers to growth in UK small businesses*. Policy Document.

Carty, L. V and Fons, J.S., 1994. Measuring changes in corporate credit quality. *The Journal of Fixed Income*, 4 (1), 27–41.

Casey, C. and Bartczak, N., 1985. Using operating cash flow data to predict financial distress: Some extensions. *Journal of Accounting Research*, 23 (1), 384–401.

Caves, R.E., 2007. *Multinational enterprise and economic analysis*. 3rd ed. Cambridge University Press.

Charitou, A., Neophytou, E., and Charalambous, C., 2004. Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13 (3), 465–497.

Chava, S. and Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. *Review of Finance*, 8 (4), 537–569.

Chen, J. and Hill, P., 2013. The Impact of Diverse Measures of Default Risk on UK Stock Returns. *Journal of Banking & Finance*, 37, 5118–5131.

Chen, J., Marshall, B.R., Zhang, J., and Ganesh, S., 2006. Financial distress prediction in China. *Review of Pacific Basin Financial Markets and Policies*, 9 (02), 317–336.

Cleves, M.A., Gould, W.W., and Gutierrez, R.G., 2010. *An introduction to survival analysis using Stata*. 3rd ed. Stata Corp.

Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34 (2), 187–220.

Cox, D.R., 1975. Partial likelihood. *Biometrika*, 62 (2), 269–276.

Dielman, T.E., 2000. *Applied regression analysis for business and economics*. 3rd ed. Duxbury.

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. *Journal of Banking & Finance*, 28 (4), 773–788.

Doing Business, 2013. *Doing Business 2013 Smarter Regulations for Small and Medium-Size Enterprises*. Washington, DC.

Doukas, J.A. and Pantzalis, C., 2003. Geographic diversification and agency costs of debt of multinational firms. *Journal of Corporate Finance*, 9 (1), 59–92.

Economist Intelligence Unit, 2009. *Surviving the Drought*.

Edmister, R.O., 1972. An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, 7 (2), 1477–1493.

Efron, B., 1977. The efficiency of Cox's likelihood function for censored data. *Journal of the American statistical Association*, 72 (359), 557–565.

Falkenstein, E., Boral, A., and Carty, L., 2000. *RiskCalc for private companies: Moody's default model*. As published in Global Credit Research, May. New York.

Fatemi, A.M., 1988. The effect of international diversification on corporate financing policy. *Journal of Business Research*, 16 (1), 17–30.

García-Teruel, P.J. and Martínez-Solano, P., 2010. Determinants of trade credit: A comparative study of European SMEs. *International Small Business Journal*, 28 (3), 215–233.

Gentry, J.A., Newbold, P., and Whitford, D.T., 1987. Funds flow components, financial ratios, and bankruptcy. *Journal of Business Finance & Accounting*, 14 (4), 595–606.

Ghoshal, S., 1987. Global strategy: An organizing framework. *Strategic Management Journal*, 8 (5), 425–440.

Gilbert, L.R., Menon, K., and Schwartz, K.B., 1990. Predicting Bankruptcy for Firms in Financial Distress. *Journal of Business Finance & Accounting*, 17 (1), 161–171.

Gill, A., Biger, N., Pai, C., and Bhutani, S., 2009. The determinants of capital structure in the service industry: evidence from United States. *The Open Business Journal*, 2, 48–53.

Gilson, S.C., 1997. Transactions costs and capital structure choice: Evidence from financially distressed firms. *The Journal of Finance*, 52 (1), 161–196.

Gilson, S.C., John, K., and Lang, L.H.P., 1990. Troubled debt restructurings: An empirical study of private reorganization of firms in default. *Journal of Financial Economics*, 27 (2), 315–353.

Di Giovanni, J., Levchenko, A.A., and Ranciere, R., 2011. Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics*, 85 (1), 42–52.

Gjellerup, P., 2000. SME support services in the face of globalisation, concerted action seminar. *In*: *Opening address, Conference Proceedings, Danish Agency for Trade and Industry, Copenhagen*. 16–28.

Gonzalez, F., Haas, F., Johannes, R., Persson, M., Toledo, L., Violi, R., Zins, C., and Wieland, M., 2004. Market dynamics associated with credit ratings: a literature review. *Financial Stability Review*, (4), 53–76.

Grambsch, P.M. and Therneau, T.M., 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81 (3), 515–526.

Grant, R.M., Jammine, A.P., and Thomas, H., 1988. Diversity, diversification, and profitability among British manufacturing companies, 1972-84. *Academy of Management Journal*, 31 (4), 771–801.

Grilli, L., 2005. The random-effects proportional hazards model with grouped survival data: a comparison between the grouped continuous and continuation ratio versions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168 (1), 83–94.

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

Gupta, J., Gregoriou, A., and Healy, J., 2014. Forecasting Bankruptcy for SMEs Using Hazard Function: To What Extent Does Size Matter? *Review of Quantitative Finance and Accounting*, Forthcoming.

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

Gupta, J., Wilson, N., Gregoriou, A., and Healy, J., 2014b. The Effect of Internationalization on Modeling Credit Risk for SMEs: Evidence from UK Market. *Journal of International Financial Institutions, Markets & Money*, 31, 397–413.

Hashi, I., 2001. Financial and institutional barriers to SME growth in Albania: results of an enterprise survey. *MOCT-MOST: Economic Policy in Transitional Economies*, 11 (3), 221–238.

Hassanien, A.A.G. and Adly, S.W., 2008. Issues facing small Egyptian construction firms: the financing barrier.

Headd, B., 2003. Redefining business success: Distinguishing between closure and failure. *Small Business Economics*, 21 (1), 51–61.

Hillegeist, S.A., Keating, E.K., Cram, D.P., and Lundstedt, K.G., 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9 (1), 5–34.

Holmes, P., Hunt, A., and Stone, I., 2010. An analysis of new firm survival using a hazard function. *Applied Economics*, 42 (2), 185–195.

Hotchkiss, E.S., 1995. Postbankruptcy performance and management turnover. *The Journal of Finance*, 50 (1), 3–21.

Hout, T.M., Porter, M.E., and Rudden, E., 1982. *How global companies win out*. Graduate School of Business Administration, Harvard University.

Hudson, J., 1986. An analysis of company liquidations. *Applied Economics*, 18 (2), 219–235.

Hutchinson, J. and Xavier, A., 2006. Comparing the impact of credit constraints on the growth of SMEs in a transition country with an established market economy. *Small Business Economics*, 27 (2), 169–179.

Hwang, R., Cheng, K.F., and Lee, J.C., 2007. A semiparametric method for predicting bankruptcy. *Journal of Forecasting*, 26 (5), 317–342.

Hwang, R.-C., 2012. A varying-coefficient default model. *International Journal of Forecasting*, 28 (3), 675–688.

IFC, 2010. *The SME Banking Knowledge Guide*. USA.

IFC and McKinsey & Company, 2010. *Assessing and Mapping the Gap in Micro, Small and Medium Enterprise (MSME) Finance.*

ITA, 2010. Small & Medium-Sized Exporting Companies: Statistical Overview, 2010 [online]. *International Trade Association*. Available from: http://www.trade.gov/mas/ian/smeoutlook/tg_ian_001925.asp#P0_0 [Accessed 20 Nov 2012].

Jenkins, S.P., 2005. Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*.

Jones, S., 2011. Does the Capitalization of Intangible Assets Increase the Predictability of Corporate Failure? *Accounting Horizons*, 25 (1), 41–70.

Karels, G. V and Prakash, A.J., 1987. Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting*, 14 (4), 573–593.

Katsikeas, C.S., Leonidou, L.C., and Morgan, N.A., 2000. Firm-level export performance assessment: review, evaluation, and development. *Journal of the Academy of Marketing Science*, 28 (4), 493–511.

Keasey, K., Pindado, J., and Rodrigues, L., 2014. The determinants of the costs of financial distress in SMEs. *International Small Business Journal*, 0266242614529317.

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), 335–354.

Kleinbaum, D.G. and Klein, M., 2012. *Survival analysis*. 3rd ed. London: Springer.

Knott, A.M. and Posen, H.E., 2005. Is failure good? *Strategic Management Journal*, 26 (7), 617–641.

Kogut, B., 1993. ``Designing global strategies: profiting from operational flexibility''. *Readings in International Business, The MIT Press, Cambridge, MA*, 195–213.

Koshy, P. and Prasad, V.N., 2007. *Small and Micro Enterprises: A tool in the fight against poverty*.

Kostova, T. and Roth, K., 2002. Adoption of an Organizational Practice by Subsidiaries of Multinational Corporations: Institutional and Relational Effects. *Academy of Management Journal*, 45 (1), 215–233.

Kotey, B. and Slade, P., 2005. Formal human resource management practices in small growing firms*. *Journal of small business management*, 43 (1), 16–40.

Kukuk, M. and Rönnberg, M., 2013. Corporate credit default models: a mixed logit approach. *Review of Quantitative Finance and Accounting*, 40 (3), 1–17.

Kwak, W., Shi, Y., and Kou, G., 2012. Bankruptcy prediction for Korean firms after the 1997 financial crisis: using a multiple criteria linear programming data mining approach. *Review of Quantitative Finance and Accounting*, 38 (4), 441–453.

Lee, C.K. and Kwok, C.C.Y., 1988. Multinational corporations vs. domestic corporations: International environmental factors and determinants of capital structure. *Journal of International Business Studies*, 19 (2), 195–217.

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.

Liang, B. and Park, H., 2010. Predicting hedge fund failure: A comparison of risk measures. *Journal of Financial and Quantitative Analysis*, 45 (1), 199–222.

Lin, S.M., Ansell, J., and Andreeva, G., 2012. Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research Society*, 63 (4), 539–548.

Lloyd-Reason, L., Ibeh, K., and Deprey, B., 2009. *Top barriers and drivers to SME internationalisation*. Unknown Publisher.

López-Gracia, J. and Sogorb-Mira, F., 2008. Testing trade-off and pecking order theories financing SMEs. *Small Business Economics*, 31 (2), 117–136.

Mallick, S. and Yang, Y., 2013. Productivity Performance of Export Market Entry and Exit: Evidence from Indian Firms. *Review of International Economics*, 21 (4), 809–824.

Mateev, M., Poutziouris, P., and Ivanov, K., 2013. On the Determinants of SME Capital Structure in Central and Eastern Europe: A Dynamic Panel Analysis. *Research in International Business and Finance*, 27 (1), 28–51.

Mazouz, A., Crane, K., Mohamed eUniversity, H. Bin, and Gambrel, P.A., 2012. The impact of cash flow on business failure analysis and prediction. *International Journal of Business, Accounting and Finance*, 6 (2), 68.

De Mel, S., McKenzie, D., and Woodruff, C., 2009. Innovative firms or innovative owners? Determinants of innovation in micro, small, and medium enterprises.

Merton, R.C., 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29 (2), 449–470.

Michael, H.M., Arthur, I.S., and David, K.E., 2009. *Fundamentals of Multinational Finance*. 3rd ed. New York: Pearson International.

Michaelas, N., Chittenden, F., and Poutziouris, P., 1999. Financial policy and capital structure choice in UK SMEs: Empirical evidence from company panel data. *Small Business Economics*, 12 (2), 113–130.

Mills, M., 2011. *Introducing survival and event history analysis*. London: SAGE Publications.

Modigliani, F. and Miller, M.H., 1963. Corporate income taxes and the cost of capital: a correction. *The American Economic Review*, 53 (3), 433–443.

Myers, S.C., 1977. Determinants of corporate borrowing. *Journal of Financial Economics*, 5 (2), 147–175.

Nam, C.W., Kim, T.S., Park, N.J., and Lee, H.K., 2008. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, 27 (6), 493–506.

Obert, M. and Olawale, F., 2010. Does debt really matter on the profitability of small firms? A perspective on small manufacturing firms in Bulawayo, Zimbabwe. *African Journal of Business Management*, 4 (9), 1709–1716.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 18 (1), 109–131.

Okpara, J.O. and Wynn, P., 2007. Determinants of small business growth constraints in a sub-Saharan African economy. *SAM Advanced Management Journal*, 72 (2), 24.

Olsson, C., 2002. *Risk management in emerging markets*. Financial Times and Prentice Hall. London: Financial Times.

Orser, B.J., Hogarth-Scott, S., and Riding, A.L., 2000. Performance, firm size, and management problem-solving. *Journal of Small Business Management*, 38 (4), 42–58.

Örtqvist, D., ERYADI, K.M., SHEIKH, F.R., and Selvarajah, C., 2006. Determinants of capital structure in new ventures: Evidence from Swedish longitudinal data. *Journal of Developmental Entrepreneurship*, 11 (4), 277–296.

Pampel, F.C., 2000. *Logistic regression: A primer*. 1st ed. Sage Publications.

Peacock, R., 2000. *Failure and assistance of small firms*. Understanding Small Business.

Peel, M.J., Wilson, N., and Howorth, C., 2000. Late payment and credit management in the small firm sector: some empirical evidence. *International Small Business Journal*, 18 (2), 17–37.

Petersen, M.A. and Rajan, R.G., 1997. Trade credit: theories and evidence. *Review of Financial Studies*, 10 (3), 661–691.

Pettit, R.R. and Singer, R.F., 1985. Small business finance: a research agenda. *Financial management*, 14 (3), 47–60.

Pindado, J., Rodrigues, L., and de la Torre, C., 2008. Estimating financial distress likelihood. *Journal of Business Research*, 61 (9), 995–1003.

Platt, H.D. and Platt, M.B., 1991. A note on the use of industry-relative ratios in bankruptcy prediction. *Journal of Banking & Finance*, 15 (6), 1183–1194.

Rabe-Hesketh, S. and Skrondal, A., 2012. *Multilevel and longitudinal modeling using Stata,Volume II: Categorical Responses, Counts, and Survival*. 3rd ed. Texas: Stata Corp.

Ramalho, J.J.S. and Da Silva, J. V, 2009. A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. *Quantitative Finance*, 9 (5), 621–636.

Ramaswamy, K., 1992. Multinationality and performance: a synthesis and redirection. *Advances in International Comparative Management*, 7, 241–267.

Ramaswamy, K., Kroeck, K.G., and Renforth, W., 1996. Measuring the degree of internationalization of a firm: A comment. *Journal of International Business Studies*, 27 (1), 167–177.

Rhodes, C., 2012. *Small businesses and the UK economy*.

Robson, P.J.A. and Obeng, B.A., 2008. The barriers to growth in Ghana. *Small Business Economics*, 30 (4), 385–403.

Rugman, A.M., 2009. *The Oxford handbook of international business*. 2nd ed. Oxford University Press,USA.

Sahajwala, R. and Van den Bergh, P., 2000. *Supervisory risk assessment and early warning systems*. Basle Committee on Banking Supervision.

Schoenfeld, D., 1982. Partial residuals for the proportional hazards regression model. *Biometrika*, 69 (1), 239–241.

Schönbucher, P.J., 2003. *Credit derivatives pricing models: models, pricing and implementation*. 1st ed. West Sussex: John Wiley & Sons.

Schuermann, T. and Jafry, Y., 2003. *Measurement and estimation of credit migration matrices*. Center for Financial Institutions Working Papers 03.

Secretariat, A.P.E.C., 2009. *Removing Barriers to SME Access to International Markets*. APEC Secretariat.

Senbet, L. and Wang, T., 2010. Corporate Financial Distress and Bankruptcy: A Survey. *Foundations and Trends in Finance*, 5 (4), 243–335.

Servigny, A. De and Renault, O., 2004. *Measuring And Managing Credit Risk*. 1st ed. New: McGraw-Hill Companies.

Shumway, T., 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model*. *The Journal of Business*, 74 (1), 101–124.

Silverman, B.W., 1986. *Density estimation for statistics and data analysis*. London: CRC press.

Sogorb-Mira, F., 2005. How SME uniqueness affects capital structure: Evidence from a 1994–1998 Spanish data panel. *Small Business Economics*, 25 (5), 447–457.

Storey, D.J., 1994. *Understanding the small business sector*. Thomson Learning Emea.

Sueyoshi, G.T., 1995. A class of binary response models for grouped duration data. *Journal of Applied Econometrics*, 10 (4), 411–431.

Sullivan, D., 1994. Measuring the degree of internationalization of a firm. *Journal of International Business Studies*, 25 (2), 325–342.

Theodossiou, P.T., 1993. Predicting shifts in the mean of a multivariate time series process: an application in predicting business failures. *Journal of the American Statistical Association*, 88 (422), 441–449.

Thomas, D.E. and Eden, L., 2004. What is the shape of the multinationality-performance relationship? *Multinational Business Review*, 12 (1), 89–110.

Tinoco, M.H. and Wilson, N., 2013. Financial Distress and Bankruptcy Prediction amongst Listed Companies using Accounting, Market and Macroeconomic Variables. *International Review of Financial Analysis*.

Trustorff, J.-H., Konrad, P.M., and Leker, J., 2011. Credit risk prediction using support vector machines. *Review of Quantitative Finance and Accounting*, 36 (4), 565–581.

Tsai, B.-H., Lee, C.-F., and Sun, L., 2009. The impact of auditors' opinions, macroeconomic and industry factors on financial distress prediction: An empirical investigation. *Review of Pacific Basin Financial Markets and Policies*, 12 (03), 417–454.

Wager, T.H., 1998. Determinants of human resource management practices in small firms: some evidence from Atlantic Canada. *Journal of Small Business Management*, 36 (2), 13–23.

Ward, T.J., 1994. Cash flow information and the prediction of financially distressed mining, oil and gas firms: a comparative study. *Journal of Applied Business Research (JABR)*, 10 (3), 78–86.

Watson, J. and Everett, J.E., 1996. Do Small Businesses Have High Failure Rates?: Evidence from Australian Retailers. *Journal of Small Business Management*, 34 (4), 45–62.

Watson, R. and Wilson, N., 2003. Small and medium size enterprise financing: A note on some of the empirical implications of a pecking order. *Journal of Business Finance & Accounting*, 29 (3/4), 557–578.

Wu, C. and Wang, X.-M., 2000. A neural network approach for analyzing small business lending decisions. *Review of Quantitative Finance and Accounting*, 15 (3), 259–276.

Yamaguchi, K., 1991. *Event history analysis*. USA: Sage.

Yip, A.Y.N., 2006. Business failure prediction: a case-based reasoning approach. *Review of Pacific Basin Financial Markets and Policies*, 9 (03), 491–508.

Zahra, S.A., Ireland, R.D., and Hitt, M.A., 2000. International expansion by new venture firms: International diversity, mode of market entry, technological learning, and performance. *Academy of Management journal*, 43 (5), 925–950.

Zehir, C., Acar, A.Z., and Tanriverdi, H., 2006. Identifying organizational capabilities as predictors of growth and business performance. *The Business Review*, 5 (2), 109–116.

Zmijewski, M.E., 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59–82.