Does size matter in predicting SMES failure?

Gupta, Jairaj; Barzotto, Mariachiara; Khorasgani, Amir

DOI:
10.1002/ijfe.1638

Does Size Matter in Predicting SMEs Failure?

Jairaj Gupta
Department of Finance, University of Birmingham
Birmingham, B15 2TY, UK; email: J.Gupta@outlook.com

Mariachiara Barzotto
Essex Business School, University of Essex
Colchester, CO4 3SQ, UK; email: m.barzotto@essex.ac.uk

Amir Khorasgani
School of Economics, Finance and Accounting, Coventry University
Coventry, CV1 5FB, UK; email: ab8394@coventry.ac.uk

June, 2018

Abstract

This study acknowledges the diversity between micro, small and medium-sized firms while predicting bankruptcy and financial distress of United States small and medium-sized enterprises. Empirical findings suggest that survival (failure) probability increases (decreases) with increasing firm size and firms in different size categories have varying determinants of bankruptcy, while factors affecting their financial distress are mostly invariant. Magnitude of significant covariates changes across the size categories of both bankrupt and financially distressed firms. Further, operating cash flow information does not add any marginal increment in prediction performance of multivariate hazard models above baseline models developed using information from income statements and balance sheets. This result holds for failure likelihood of SMEs as well as their respective size categories.

Keywords: bankruptcy; SMEs; survival analysis; financial distress; operating cash flow

JEL Classification Codes: G32; G33

Acknowledgements: The authors are sincerely grateful to the Editor and anonymous referees for their insightful comments and suggestions that improved this paper significantly. An earlier version of this paper was circulated under the title "*Bankruptcy and Financial Distress in US Small and Medium Sized Enterprises*".

1. Introduction

In developed economies, small and medium-sized enterprises (SMEs) are, relative to turnover, disproportionately linked to employment rates. In the United States (US), for instance, almost half of all employees work for enterprises with fewer than 250 employees. SMEs play a crucial role in the global economy, and are pivotal to the economic growth and development of a country (Bosma and Levie 2010), as well as to poverty reduction (Koshy and Prasad 2007). Therefore, a detailed understanding of the factors affecting SMEs survival is important for policy makers, firms and capital suppliers. The attention devoted to SMEs survival has constantly increased over the years, particularly after the financial crisis in 2008-09. Indeed, the revised Basel capital accords and national governments have placed greater emphasis on understanding the credit risk attributes of SMEs.

Notwithstanding the extensive literature on the performance and financial distress of SMEs, the factors and the extent to which they affect SMEs failure likelihood across size categories are still overlooked. As argued in a recent study by Altman *et al.* (2017), bankrupt and non-bankrupt firms have different boundaries due to their size (small and large), which affect the accuracy of prediction when data from one size category is used for another size category. Building upon the previous evidence showing that size affects access to finance (e.g. Beck and Demirguc-Kunt 2006), we propose a development in modelling financial distress and bankruptcy in the US. More specifically, we address this issue by exploring whether insolvency and financial distress likelihood varies across size categories of the US SMEs, by looking at the factors affecting SMEs failure likelihood in three sub-categories of SMEs (namely, micro, small, and medium[1]). Few studies explore the differences amongst the sub-categories of SMEs. Analysing financial and non-financial factors affecting UK SMEs bankruptcy across company size, Gupta *et al.* (2015) show that the credit risk characteristics of micro firms significantly differ from SMEs as a whole. Accordingly, they suggest that they should be treated separately

[1] This paper classifies SMEs into three size categories (micro, small, and medium firms) as defined by the latest European Union classification. According to this classification, a firm is considered 'micro' if it has less than 10 employees with an annual turnover of less than €2 million (about $2.6 million); 'small' if it has less than 50 employees with an annual turnover of less than €10 million (about $13 million); and 'medium' if it has less than 250 employees with an annual turnover of less than €50 million (about $65 million). Further information can be found at: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition/index_en.htm (accessed on July 28, 2015).

for better pricing of credit risk and devising effective credit policies. In the light of this discussion, we expect the default characteristics to vary across SMEs size categories. We draw upon and advance this study by: i) using firms' annual sales turnover, which is a preferred/more appropriate proxy of firms' size than number of employees; ii) providing distinct evidence for two default definitions: bankruptcy (based on Chapter 7/11 filings), and financial distress (based on a firm's ability to honour its financial commitments, and the value of its net worth); iii) exploring operating cash flow marginal discriminatory power across size categories; and, iv) examining the presence of statistically significant differences in the magnitude of mutually significant covariates in the model for all SMEs, and models for respective size categories, via the statistical (Wald) test of equivalence of coefficients.

Based on our empirical analysis in the context of US SMEs, using annual firm level financial information obtained from the Compustat database (from 1990 to 2014), we conclude that all SMEs are not the same. More specifically, the determinants of bankruptcy vary across different size categories of SMEs. Earnings are only found to be important for the largest size category, as is also the case for the ratio of assets to liabilities. Financial expenses are almost always found to be significant, but the size of its effect varies, especially in reference to micro firms. We also present compelling evidence that estimated coefficients differ between models for financial distress estimated across SMEs as a whole, and the varying size categories. Forecasters would therefore be advised that distinct models for bankruptcy or financial distress should be specified not in reference to SMEs as a whole, but rather in consideration of the different size classifications. In contrast to the work by Gentry *et al.* (1987) and Gilbert *et al.* (1990), we do not find that cash-flow contributes to an understanding of bankruptcy. However, the results do compliment the findings of Charitou *et al.* (2004) for the UK in explaining financial distress using cash flow from operations (CFO).

There are a number of differences in the estimated determinants of financial distress as opposed to bankruptcy. Firms with greater holdings of cash and short-term investments are less likely to face financial distress. Taxes are consistently found to have an effect on financial distress, but this is not the case for bankruptcy, where only the model across SMEs as a whole provides evidence of a significant effect. There is also evidence that the effect of the ratio of current assets to current liabilities is different across different classes of firm when predicting financial distress. The value of trade debt predicts financial distress, concordant with the findings of Hudson (1986) and Beck *et al.* (2006). It is possible that the value of trade credit is reduced as a firm appears more likely to file for bankruptcy, explaining the different result. Bankruptcy and financial distress are distinct events and separate modeling of them shall lead to improved risk pricing. A similar conclusion might be reached in reference to the consideration of different size categories of firms.

The remainder of the paper is structured as follows: Section 2 presents a literature review on bankruptcy prediction and survival analysis, which is the foundation for SMEs failure prediction analysis, and includes the potential effect of firm size and operating cash flow on SMEs likelihood of entering financial distress and bankruptcy. Section 3 outlines the empirical methods, including an explanation of the dataset and covariates. Results and discussion are reported in Section 4. Finally, Section 5 draws conclusions and policy implications.

2. Literature Review and Hypotheses Development

This section reviews the past studies on bankruptcy prediction and survival analysis and is the foundation for our SMEs failure prediction analysis. The discussion also includes the potential effect of firm size and operating cash flow on SMEs likelihood of entering financial distress and bankruptcy.

2.1 Approaches to SMEs Failure Prediction

The principal source of external funding for SMEs is debt and, more specifically, bank lending. However, lenders face problems in forecasting loan performance. This issue has been exacerbated over the years due to the presence of less favourable economic environments, particularly after the financial crisis in 2008-09. Such conditions also lead to restricted and over-priced credit. Credit risk incorrectly or inadequately measured can generate detrimental effects for SMEs, banks and the wider economy. Notwithstanding the importance of understanding and forecasting insolvency for SMEs, until the last decade, research in this area has been scant compared to the study on larger firms. This debate, to which the present paper aims to contribute, has mainly focused on improving banks’ estimation and treatment of credit risk for SMEs.

There is an extensive literature, spanning more than three decades, on business failure prediction (Balcaen and Ooghe 2006). This literature includes various credit risk models, which are mainly derived from two approaches: the Altman (1968) model, which uses accounting-based indicators, and the Merton (1974) model, based on market information. Although the Merton (1974) model has significant advantages, the unavailability of market information in the case of unlisted companies deems it inapplicable for the majority of SMEs (e.g. Pompe and Bilderbeek 2005). Prediction of bankruptcy using accounting information began with the seminal work by Beaver (1966), who employed financial (accounting) ratios in an univariate model to predict failure. Shortly thereafter, the seminal multivariate (Z-score) model was developed by Altman (1968). Altman's (1968) study concludes that traditional ratio analysis is not a reliable approach and should be replaced by multivariate discriminant analysis (MDA), as a more sophisticated tool for predicting default events. Following Altman (1968), a vast number of studies has applied the MDA statistical method to predict firms' default. More recently, Altman *et al.* (2017) analysed the performance of the Z-score model in 31 European and 3 non-European countries. The authors argue that the Z-score model performs well in most countries, reaching a prediction accuracy of approximately 90% (when associated with additional country specific variables, or 75% otherwise). However, Ohlson (1980) challenged Altman's (1968) Z-score model and raised some critical issues with the predictive efficiency of the MDA technique. To mitigate/overcome technical issues of previous models, Ohlson (1980) proposed logistic regression technique instead of MDA and thereafter it remains a popular choice (e.g. Altman and Sabato 2007, Gupta *et al.* 2014).

Most bankruptcy prediction models are based on single period classification, with multiple period bankruptcy data. Given the fact that firms change through time, the bankruptcy probabilities produced by MDA or logistic models might be biased and inaccurate. Zavgren (1985) finds that in traditional default prediction models, the coefficients' signs of the bankruptcy indicators may change in the years prior to failure. Luoma and Laitinen (1991) extend this claim by showing that not only the coefficients' signs change before failure, but also the values of the coefficients. Evidence provided in these studies seems to suggest that traditional cross-sectional default prediction models are not valid, as the underlying failure process does not remain stable over time. Conversely, survival analysis models have the ability to address these changes, and hence are more suitable to modelling the dynamic process such as bankruptcy prediction. However, Luoma and Laitinen (1991) conclude that the survival analysis approach slightly underperforms compared with discriminant analysis and logistic analysis in bankruptcy prediction. Laitinen and Kankaanpaa (1999) implemented a comparative study to test the performance of various bankruptcy prediction models. Their analysis indicates that hazard models have better predictive power for two and three-year predictions, while logistic analysis shows superior performance for one-year prediction. However, they conclude that the differences in models' predictive powers are not statistically significant. Nevertheless, more recent studies shed light on the superior performance of the hazard models. Shumway's (2001) study was one of the first to employ a large sample of about 2000 firms, spanning over 31 years. He found that half of the accounting variables used in previous models by Altman (1968) and Zmijewski (1984) are not significant indicators of bankruptcy likelihood. Moreover, the accuracy of the hazard model substantially increased when using both market-based and accounting-based indicators to predict business failures. Laitinen (2005) also found that the classification accuracy of the proportional hazard model in the years prior to the firms' default is superior to other statistical models used by credit institutions. Employing the complete database of UK listed firms between 1979 and 2009, Bauer and Agarwal (2014) tested the performance of two hazard models (Shumway 2001, Campbell *et al.* 2008) against the traditional accounting-based Z-score model (Taffler 1983) and Merton's contingent claims-based model (Bharath and Shumway 2008). They report clear evidence regarding the mis-calibration of the Z-score model and contingent-claim based model, while the average default probabilities of hazard models are closer to observed default rates. They also find that the Z-score model and contingent claim-based approach clearly underperform, while the receiving operator characteristics (ROC) analysis highlights no significant differences between the two hazard models.

The use of qualitative information presented a further development in modelling firms' credit risk (e.g. Lehmann 2003). Analogously, Grunert *et al.* (2005) and Tsai *et al.* (2009) report that non-financial factors present a useful supplement to financial factors in credit rating. In the context of SMEs, Altman *et al.* (2010) report improvement in models' classification performance after accounting for qualitative information of UK SMEs.

Until recently, less academic attention has been devoted to SMEs in the failure prediction literature. This calls for scholarly contributions for a deeper understanding of the factors and their magnitude in affecting SMEs failure likelihood. Previous studies show that the use of annual report variables two years and one year before failure improves the predictive power of financial variables, while annual report variables do not contain incremental information three years before failure (Laitinen 1993). In an analysis of SMEs in the US, Altman and Sabato (2007) use financial measures to develop default prediction models using logistic regression, reporting a significant improvement over standard credit scoring models. They conclude that banks' capital requirements should be (slightly) lower if SMEs are treated as a distinct corporate category. However, the introduction of non-financial information - missing in their study - has been seen by the authors themselves as a necessary future line of investigation. The relevance of the addition of non-financial information in improving model performance has been confirmed by Altman *et al.'s* (2010) approach in an analysis of UK SMEs.

2.2 SMEs and Size Factor

The strength of old-large firms often represents the weakness of new-small firms and vice-versa (Aldrich and Auster, 1986). The liabilities of smallness create various problems, such as: competing for labour, meeting government requirements, innovation performance, and raising capital. These issues lead to their high mortality rate. Indeed, as the number of employees increases, companies adopt different formal human resource practices (e.g. Kotey and Slade 2005), organisational structures, and innovation strategy. In particular, exploring innovation within different SMEs size categories, De Mel *et al.* (2009) find that more than one quarter of micro firms engage in innovation, with marketing innovation being the most common. The authors show that firm size exhibits a stronger positive effect on process and organisational innovations than on product innovations.

Size has a strong direct impact on the financial and economic performance of SMEs. Literature has reported on heterogeneity in the characteristics of firms, their access to finance and, in turn, the company's potential for growth. The smaller the size the more firms' growth is constrained by: i) corruption of bank officials; ii) financial and legal issues (Beck *et al.* 2005); and iii) obstacles to accessing external finance (Beck and Demirguc-Kunt 2006), especially if the firm is young (Beck *et al.* 2006). Leverage decisions (Ramalho and Da Silva 2009) and capital structure choices also vary significantly between micro, small and medium-sized firms. Investigating capital structure choices, Mateev *et al.* (2013) find that medium-sized firms are mainly dependant on long term bank loans as their preferred method of external financing, while short-term loans and trade credits are the main sources of external finance for micro and small businesses. Holmes *et al.* (2010) estimate hazard functions separately for micro-enterprise and SMEs and find that the effect of variables on the survival of these two types of firms is substantially different. The empirical literature also argues that the stability of cash flow and diversity increases with firm size (Gill *et al.* 2009), leading to a negative relationship between firm size and default likelihood (Pettit and Singer 1985). The relationship between SMEs asset size and insolvency risk is not linear, when controlling for company size using total asset value (Altman *et al.* 2010). This is mainly due to creditors being less likely to force companies with low asset value into insolvency, as they do not benefit from the recovery process. After a certain threshold, the insolvency risk ultimately declines with company size, as shown in the study by Gupta *et al.* (2015), in which the authors demonstrate that risk and default characteristics of 'micro' firms differ from those of larger SMEs size classifications. Notwithstanding the effort spent in assessing a firm's financial situation, it would appear that there is room in the literature (e.g. Gupta et al. 2018) for approaches to directly identify factors affecting both financial distress and bankruptcy across different size categories of SMEs. A more comprehensive picture of the factors affecting SMEs failure could help companies, financial institutions, and policy makers to make better informed choices. On this backdrop, we test the following three hypotheses:

H1: Failure rate of SMEs varies across micro, small and medium size categories.

H2: Factors affecting SMEs failure likelihood vary across micro, small and medium size categories.

H3: Factors that are mutually significant in predicting failure likelihood of SMEs, and micro, small or medium firms respectively, exhibit significant differences in the magnitude of their coefficient in respective models.

2.3 Cash Flow from Operation and SMEs failure

In this paper, CFO is also introduced as an augmentation to the existing empirical work on the US SMEs. Building upon previous studies, CFO has been acknowledged as a factor in explaining SMEs bankruptcy (Gentry *et al.* 1987, Gilbert *et al.* 1990) as well as financial distress (Charitou *et al.* 2004). Literature on trade credit and capital structure of small firms has shown that firms with insufficient cash flow are more susceptible to financial distress. Casey and Bartczak (1984) find that the accrual-based MDA model has superior predictive power than any single operating cash flow ratio. Traditional cash flow may be a more reliable predictor of failure compared to operating cash flow (Laitinen 1994) but operating cash flow seems to be more sensitive to recession, as it declines in non-failed firms when other firms are approaching the failure date. As CFO relates directly to the liquidity position of the firm, its inclusion provides an indication of a firm's ability to meet its short-term obligations. Accordingly, stakeholders value the interrogation of cash flow variables as they embed information on the potential "early warning" of financial difficulties (Mossman *et al.* 1998).

In studies that do not explicitly model SMEs, there are mixed findings on the impact of CFO on failure. Turetsky and McEwen (2001) model financial distress rather than bankruptcy, making explicit reference to cash-flow. Conversely, Bernard and Stober (1989) do not consider CFO to be useful in forming expectations on future cash-flows, arguing that the figures are too easy for managers to manipulate. This may assist in explaining why Mazouz *et al.* (2012) did not find CFO to be useful in predicting default. Additionally, analysing UK SMEs, Gupta *et al.* (2014) report that operating cash-flow[2] information does not contribute to an improved understanding of SMEs failure likelihood. Preceding studies have examined the effects of CFO in the context of larger firms only or in different institutional-economic environments (Gupta *et al.* 2014). In moving to an examination of smaller firms, we might pose this question again in the US context. SMEs are less able to manipulate accruals; this could determine less measurement error or bias for smaller firms. This might also commend the analysis of SMEs by size class, as micro firms in particular may have less control over the reporting of CFO. In light of this discussion, we test the following hypothesis:

[2] This provides a useful picture of the cash holdings of a firm for use in insolvency studies as it is not conflated with factors such as depreciation.

H4: Marginal information content of CFO information above income statement and balance sheet is significant in predicting failure likelihood of SMEs and their respective size categories.

3. Empirical Methods

This section provides discussion pertaining to the source and use of dataset, selection of explanatory variables as well as statistical model employed in this study.

3.1 Dataset

We employ annual firm-level data from the Compustat database. A relatively long sample period is employed, running from 1990 through 2014 inclusive, although note should be taken that the specified model seeks to predict bankruptcy or financial distress in the following year. SMEs are defined as firms having annual sales turnover less than $65 million (or €50 million), broadly consistent with the definition adopted by the European Union. The United States formally defines an SME as having fewer than 500 employees, although there is a debate as to whether employee numbers are the most appropriate definition. The definition adopted here better facilitates comparison with previous works discussed above. One of the contributions of this work is the investigation of insolvency hazard across size categories of SMEs. This naturally requires a determination of size category. In this, we are guided by the European Union's definition of micro, small and medium firms. Micro firms are defined as those with sales of less than $2.6 million, small firms are those with sales above this but less than $13 million, and medium firms being the remainder, with sales of less than $65 million.

Financial distress and exit are different events as well. As discussed by Keasey *et al.* (2014) in a study of UK firms, exit may occur for reasons other than financial distress, or in anticipation of future distress. However, the greater economic costs and a greater concern for banks is where the restructuring or exit of a firm is less ordered or involuntary. In common with Keasey *et al.* (2014) the focus of part of the analysis presented here is the prediction of whether a firm will be in financial distress, rather than a prediction of exit for all possible reasons. Thus, following Keasey *et al.'s* (2014) definition of financial distress, we consider a firm as financially distressed if: i) its EBITDA (earnings before interest tax depreciation and amortization) is less than its financial expenses for two consecutive years; and ii) the net worth/total debt is less than one and the net worth experienced negative growth between the two periods. Additionally, a firm is also recorded as financially distressed in the year immediately following these distress events.

We further analyse the determinants of exit as a consequence of legal bankruptcy. In the US, a firm may either be liquidated (Chapter 7) or enter bankruptcy proceedings for the purposes of financial restructuring (Chapter 11). In the case of a Chapter 11 filing, the firm may remain in a bankrupt state for a period of time and re-enter, possibly several times, following emergence. Data on bankruptcy filings is available from Compustat. In Compustat, a company has status alert indicator (data item “stalt”) "TL" when it is in bankruptcy. Generally, a company will have a "TL" indicator for the quarter/year in which it files for Chapter 11 or Chapter 7, and it remains "TL" in subsequent the quarter/year until it emerges from Chapter 11 or is liquidated. A further “AG” footnote on total assets appears during the quarter/year the company emerges from Chapter 11. Consequently, taking the bankruptcy filing date as the bankruptcy indicator ignores the possible subsequent bankruptcy states. Thus, our definition considers a firm to be *bankrupt* 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. For ease of exposition (and in want of a better term), the proceeding discussion of explanatory variables collects these two legal default events under the term ‘bankruptcy’, without any distinctions being drawn between Chapter 11 and Chapter 7 filings.

As the age of a firm might be expected to influence the probability of entering bankruptcy or financial distress, we consider the age of firms in our empirical analysis. Age is proxied by the earliest year for which financial information is available in the Compustat database. In Compustat, 1950 is the earliest point in time for which financial information is available. In order to avoid measurement error, we selected only those firms that entered the Compustat database after 1950. Further, financial firms (with Standard Industrial Classification – SIC codes from 6,000 through 6,999) and regulated utilities (codes 4,900 through 4,949) have been excluded from our analysis. This is a common approach in the literature, as financial firms are likely to have very different capital structures and regulated industries may be constrained such that a reliable model of behaviour is more difficult to obtain. Some firms might have multiple entry and exits in our database. For instance, when an existing SME reports sales revenue over $65 million, it exits our sample and returns only when its sales revenue drops below $65 million. The age variable is therefore created before applying other filters. We also excluded subsidiary firms (if the 'stock ownership' code - Compustat data item 'stko') is '1' (subsidiary of a publicly traded company) or '2' (subsidiary of a company that is not publicly traded) in the Compustat database.

3.2 Explanatory Variables

A large number of explanatory variables are discussed in the academic literature on firm default, financial distress and insolvency. We follow this literature in collating variables and conducting a number of first stage tests for their (joint) inclusion in the determination of financial distress or bankruptcy (see for example Altman and Sabato 2007, or Lin *et al.* 2012). This will be discussed further below, following a description of the explanatory variables and the reasons for their consideration in the modelling process. The variables included in the model are listed in Table 1[3].

Liquid assets might be drawn upon by a firm to meet immediate payments. Therefore, we expect that a firm with lower cash and short-term investments relative to total assets (CTA) will have a higher probability of default. A parallel argument might be made in relation to the scale of trade creditors to total assets (TCTA), with a larger value signifying greater immediate short-term claims on the firm and, therefore, a greater probability of default. Hudson (1986) reports that trade creditors are an important component of short term liabilities for many SMEs, and we might therefore expect this variable to have an effect on the probability of default. A further measure of the liquidity position of the firm is the log of current assets to current liabilities (LCR), where a higher value is expected to have a negative effect on the probability of default. A higher income tax paid in relation to total assets (TTA) is expected to have a negative effect on the probability of default, as firms having higher income are expected to pay higher taxes. The level of short term debt relative to equity book value (STDEBV) is also included as a measure proxying the short-term claims on the firm relative to capital employed. Short-term debt must be refinanced or repaid more immediately than is the case for other

[3] While calculating respective financial ratios, we replace denominators having value 0 with small positive values of $1 to ensure that ratios of respective firm-year observations does not produce any missing value.

sources of financing, and may be critical to the (involuntary) default of the firm if it is large. We take the capital employed divided by total liabilities (CETL) as an (inverse) measure of leverage. The larger the capital of the firm relative to the liabilities, the lower the expected default probability. Capital growth (CAG) is included as a higher rate of growth might indicate a growing capacity to meet financial obligations or to finance futures operations. This would suggest a lower probability of default. Total liabilities to total assets (TLTA) is also included as an explanatory variable, again as a measure of the financial fragility of the firm. Tangible assets are often found to be important in the determination of both capital structure and default. *Ceteris paribus* a company with a higher level of tangible assets is more likely to obtain external funding and to be offered re-financing in the event of difficulties. Further, Jones (2011) argues that firms in financial distress are more likely to capitalise intangible assets. We therefore include the ratio of intangible assets to total assets (IATA) as a predictor of default.

We employ a number of measures linked to earnings. The first of these is retained earnings relative to total assets (RETA). If a firm is nearing default, then we expect that this will be reflected in lower retained earnings. Conversely, even where the liquidity of the firm is under pressure, we would expect that a firm with relatively high retained earnings would find (re)financing easier and hence the probability of default lower. As we might expect earnings to be an important predictor of default, this is further explored through the inclusion of earnings before interest, tax, depreciation and amortization, again relative to total assets (EBITDATA). There is relatively little choice on the part of the firm in the determination of the value of this variable, the inclusion of which may therefore compliment the inclusion of RETA, where greater choice on the part of the firm's managers may be evident. We also include financial ratios related to financial expense. Financial expenses relative to total assets (FETA) is included as a means to proxy the financial claims on the company, with a larger size of this variable expected to increase the probability of default. A similar argument holds for the financial expenses relative to cash flow from operations (FECFO).

An innovation in the analysis of default in US SMEs is the inclusion of variables related to cash flow from operations (henceforth CFO). We include a measure for cash-flow divided by total assets (CFOTA), with a high value suggesting a better asset utilization and hence lower default risk. Where current liabilities are used as the denominator, the resultant variable

CFOCL is presented as a proxy for the ability of the firm to meet its obligations from CFO. With sales as the denominator we have the variable CFOS, a proxy for how effectively the firm manages its receipts/payments. We further include the growth in CFO (CFOG) as an indicator that the firm's short term financial position and, hence, ability to meet its obligations will be improving with higher values of CFOG. The lower the rate at which earnings are translated to cash-flows, the worse, relatively speaking, the liquidity position of the firm. As we expect liquidity (rather than earnings) to influence default, we therefore anticipate that higher values of ECFO (earnings divided by operating cash-flows) lower the probability of default.

[Insert Table 1 Here]

3.3 Hazard Model

3.3.1 Basic Hazard Model

The analysis of default is derived from a survival function. Survival analysis addresses the time to the occurrence of an event, which in this study is the time until a default, being either a bankruptcy or financial distress. Suppose T is a non-negative random variable that denotes the time of a default event and t represents time itself. T then has a probability density function $f(t)$ or a cumulative distribution function (CDF) such that $F(t) = \Pr(T \le t)$. A survival function, $S(t)$ is then given by the probability that $T > t$, the reverse CDF of T.

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

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

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

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

3.3.2 Discrete-Time Hazard Model

Some events may be experienced at any instant in continuous-time. This results in exact censoring and survival times, which will be recorded in relatively fine time scales such as seconds, hours or days. If there are no *tied* survival time periods, then under such circumstances a continuous-time survival model is an appropriate choice (Rabe-Hesketh and Skrondal 2012). However, in many cases the events are discrete or recorded at discrete intervals; for instance, expressing time to event in weeks, months or years. Where there are relatively few censoring or survival times with *tied* survival time periods, then a discrete-time survival model is more appropriate (again see Rabe-Hesketh and Skrondal 2012). Interval-censoring[4] leads to discrete-time data, which is the case with our database. Here, the beginning and end of each time interval is the same for all of the SMEs, as the information is recorded on an 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, we estimate our hazard models in a discrete-time framework with *random effects* (α_i), 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* link (Grilli 2005, Jenkins 2005, Rabe-Hesketh and Skrondal 2012), specified as follows:

$$cloglog\big(h_i(t)\big) \equiv ln\{-\,ln(1 - h_i(t))\} = \beta x(t)_i' + \lambda_t \tag{3}$$

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 4.

$$P_{i,t} = \frac{e^{\alpha(t) + \, x(t)_i' \beta}}{1 + e^{\alpha(t) + \, x(t)_i' \beta}} \tag{4}$$

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

Where $P_{i,t}$ is the probability of experiencing the event by subject *i* at time *t* and *α(t)* captures the baseline hazard rate. This will produce very similar results as long as the time intervals are small (Rabe-Hesketh and Skrondal 2012) and the sample bad rate (% of default to non-default) is very low (Jenkins 2005). Thus, considering this discussion, we estimate our discrete hazard models using a logit link estimator.

3.3.3 Specifying Baseline Hazard rate

The baseline hazard rate is a necessary stage of analysis for the discrete hazard model. The specification of the baseline hazard function defines the probability of default given baseline values for the explanatory variables. Here we set all baseline values equal to zero. Time-varying variables are then identified that bear a functional relationship with survival times.

There are a number of alternate specifications for the baseline hazard function including log(survival time), polynomial in survival time, fully non-parametric, and piece-wise constant (see Jenkins 2005). For a fully non-parametric baseline hazard function, duration-interval-specific dummy variables need to be created (see Beck *et al.* 1998). However, this method becomes cumbersome if the maximum survival time in the dataset is very high, as in case of bankruptcy databases. A more parsimonious way of specifying the baseline hazard function is to use a piece-wise constant method. In this, the survival times are split into different time intervals that are assumed to exhibit a constant hazard rate (see Jenkins 2005). If there are time intervals (dummies) with no events, then the relevant observations from the estimation should be dropped as duration specific hazard rates cannot be estimated for them (Jenkins 2005, Rabe-Hesketh and Skrondal 2012). Considering the estimation convenience, the piece-wise constant specification of baseline hazard rate could be applied. However, if the hazard curve shows frequent and continuous steep rises and falls, then the piecewise approach may not be used and the fully non-parametric baseline hazard might be an appropriate choice.

3.4 Model Validation using ROC Curves

To evaluate the classification performance of the default prediction models developed, we present measures of their predictive accuracy. Out-of-sample validation regression models are first estimated up to 2010, with predictions made for bankruptcy and financial distress in

2011. The estimation period is then updated for a further year, ending in 2011, again with predictions made for the following year, with the updating process continuing until the final prediction year of 2014.

Model predictive performance is reported using Receiving Operator Characteristics (ROC) curves. These plot the true positives, where the model predicted a default which actually occurred, also known as sensitivity, on the ordinate axis. The false positives are plotted on the abscissa axis, where the firm defaults but the model failed to predict this, also known as fall out or (1-specificity). These are plotted because the discrimination threshold between defaulted and non-defaulted firms is varied. A 45^o line would indicate no discriminatory power in the model; accordingly, deviations (above this) might be taken as an indication of predictive/discriminatory success. Therefore, the area under the ROC (AUROC) might be measured to provide a numerical value of model performance. Its value ranges from 0.5 to 1.0, which encapsulates the classification performance of the model developed. AUROC of 1 denotes a model with perfect prediction accuracy and 0.5 suggest no discrimination ability. In general there is no 'golden rule' regarding the value of AUROC, however anything between 0.7 and 0.8 is acceptable, while above 0.8 is considered to be excellent (Hosmer Jr *et al.* 2013).

4. Results and Discussion

To eliminate the influence of extreme outliers on our statistical estimates, we restrict the range of all our financial ratios between the 5th and 95th percentiles. In addition, we lag all our variables by one-year to ensure that all information employed in prediction is available at the beginning of each year. Note that default is recorded during the year in which it takes place to ensure that fitted values for the dependent are not based on financial information presented in the same year, as some information would post date the default event.

4.1 Failure Rate and Descriptive Statistics

Information on the sample for the dependent variables is presented in Table 2, disaggregated by SMEs size classification and failure type (bankrupt and financially distressed firms).

Notwithstanding the number of SMEs filing for bankruptcy is small in absolute terms, bankruptcy aggregated data delineates an apparent decline in incidence over the time window analysed (especially after around 2001). Data seems to suggest different trends when bankruptcy is disaggregated by SMEs size categories. Indeed, the 'micro firm' set represents the largest single component of the total (with only a few exceptions). In particular, this tendency is confirmed in this century. The same might also be said for firms that are financially distressed. Indeed, micro firms are always the largest group of financially distressed firms and their relative importance is growing over time, especially during this century (in line with the bankruptcy trend). Distress rates for this group have been above 40% for much of the time since the dotcom bubble in 2002. Although cohort sizes may vary, these simple descriptive statistics reinforce the interest in examining the determinants of default events separately. This evidence provides strong support for our hypothesis H1. Default numbers and proportions do not appear to be consistent across the different size classifications of SMEs, whether firms are filing for legal bankruptcy or entering financial distress. Note might also be taken of the very large number of firms that are facing financial distress, and the challenges and importance of predicting distress for external providers of capital. Further details on sample description and descriptive statistics for the explanatory variables can be found in Appendix A[5].

[Insert Table 2 Here]

4.2 Analysis of Survival and Hazard Curves

As discussed earlier, it appears that the likelihood of default changes over time and, consequently, a simple statistical test (such as a *t* test) may be misleading. In the context of survival analysis, many of the alternative measures were specifically developed for continuous-time models and may also suffer from bias when applied in a discrete-time framework. Consequently, we estimate univariate hazard models and report average marginal effects (AME) for each variable to facilitate the specification of the later multivariate model. The usual marginal effects report on the change in the conditional mean default in the baseline model as a consequence of a marginal change in one of the explanatory variables[6]. However, this paper

[5] Table A1 reports descriptive statistics outlining any potential discriminatory power of a set of explanatory variables, according to their type of failure (bankruptcy vs financial distress).
[6] Note that none of the variables in question is a dummy variable.

first estimates a marginal effect for each observation and, then, the average across the marginal effects to obtain AME. The absolute values of the AME are ranked, as a guide to later model specification, and their statistical significance is reported. In case the variables for the SME category are statistically significant, we reported the p-value for a Wald test of the equivalence of the estimated coefficient for each of the size categories. The rationale behind this analysis is two-fold: first, it provides early evidence of whether the determinants of default are similar across different categories of SMEs; and second, it informs the proceeding model specification.

Figure 1 presents survival and hazard curves for our sample of micro, small and medium firms. These show the hazard of bankruptcy and financial distress for the different size classes of SMEs. Reflecting on the discussion of the previous section, the hazard rates for bankruptcy are lower than for financial distress; conversely, the survival rates are higher for bankrupt firms in comparison to firms experiencing financial distress. This reinforces our argument that, at a given age, the likelihood of financial distress is much higher and significantly different from bankruptcy likelihood. The top part of Figure 1 presents survival curves for the bankrupt and distressed groups of firms for different SMEs size categories. From these graphs, it emerges clearly that survival rates with respect to firms' age vary with firms' size. Larger firms tend to have better survival rates than smaller firms. In particular, for the bankrupt group of firms, the survival rate of micro firms is significantly lower (conversely hazard curves are significantly higher) than small and medium-sized firms. Although the survival curve of small firms is lower than that of medium firms, they are very close to each other throughout. This suggests that both small and medium-sized firms exhibit quite similar bankruptcy attributes. However, for financially distressed firms all three size groups exhibit sufficiently different survival rates. Also, as might be expected from the preceding statistics, a higher hazard rate is found for financially distressed firms in Figure 1 than for their bankrupt counterparts, although the patterns differ. In line with the previous results, micro firms are most vulnerable to financial distress, followed by small and medium firms respectively. For micro firms, the hazard rate for financial distress rises but then falls from around 30 years. By contrast, the hazard rate for small firms shows a rise until an age of around 20, then plateaus before rising steeply again. The pattern to the hazard rate for financial distress for medium firms reports general moderate rise. These differences in hazard rates across size categories of bankrupt and financially distressed firms reinforces our hypothesis (H1) that default attributes of SMEs vary with firms' size.

Further, the hazard curves do not follow any consistent parametric shape. Thus, a fully non-parametric baseline hazard specification seems to be an appropriate choice. In light of this discussion, in the following section, we employ firm age specific dummy variables in our multivariate models to proxy the baseline hazard rate.

[Insert Figure 1 Here]

4.3 Univariate Hazard Analysis

The results of univariate modelling are presented in Table 3, with columns reporting on findings for SMEs as a whole and for each SMEs size sub-categories, disaggregated by bankrupt and financially distressed firms respectively. For each variable, the first row shows the coefficient estimates, together with an indication of statistical significance. The second row indicates the standard error. The AME are presented in the third row with their ranking (for the column concerned). Appropriate Wald statistics are then presented in the final row.

4.3.1 Non-Cash Flow Covariates

Section A of Table 3 reports the results for non-cash flow covariates. Earnings (EBITDATA) demonstrate discriminatory power over the baseline model for both bankrupt and financially distressed firms, with the exception of micro and small firms in the bankrupt part. All cases in which this variable presents a statistically significant coefficient, it has also a rank for the AME of 6 or better. In the case of bankrupt firms, the effect appears to be driven by the medium category, and this result informs the further model specification. In the case of short term debt (STDEBV) only the medium category of financially distressed firms indicates discriminatory power over the baseline model, though with the expected positive sign for the estimated coefficient. The estimated coefficients for the variable cash scaled to total assets (CTA) have the expected negative signs and are all statistically significant. Ranks for the AME are high, but multicollinearity issues restrict the use of this variable in the full model when considering bankrupt firms. Retained earnings (RETA) also have the expected negative and statistically significant signs in univariate estimation but have low rank for the AME, with the possible exception of bankrupt micro firms. Similarly, capital employed (CETL) has the expected negative and statistically significant effect across the board. However, the rank of the AME is low for financially distressed firms; this is reflected in the multivariate model specification. Liabilities relative to total assets (TLTA) have a strong effect in a univariate setting but the variable is correlated with other variables in the full model, where its use is somewhat restricted. In considering the results on capital growth (CAG), which is a variable included in most of the multivariate models, the estimated coefficients confirm the expected sign. The rank for the AME, especially in the case of bankrupt firms, is high. The Wald statistics suggest a different response by size classification. In an examination of the effect of taxes (TTA), there are two cases where the estimated coefficients are of an unexpected sign (micro in bankrupt firms) or are statistically insignificant (small in bankrupt firms). However, in the other univariate results for this variable, the expected positive effect also has a highly ranked discriminatory power, commending its inclusion in the multivariate setting. The ratio of current assets to liabilities (LCR) ranks most highly for small bankrupt firms, when examining the AME, with the lowest rank being for medium-sized financially distressed firms. Trade creditors (TCTA) are also found to have significant discriminatory power against the baseline model, and the variable ranks highly on the basis of AME. However, the Wald statistics do not suggest that the effects differ greatly across different categories of bankrupt firms. Greater levels of financial expense (FETA) have the expected positive effect on default probability (relative to the estimated baseline). The variable also ranks highest on the basis of AME for all but one case, and the response appears to be different across size categories. The variable measuring intangible assets (IATA) has no discernible discriminatory power above the baseline model for bankrupt firms and the AME are ranked low where this is the case of financially distressed firms.

4.3.2 Cash flow Covariates

The remaining sub-set of variables reference operating cash-flows and the results for the univariate hazard analyses on these variables may be found in Section B of Table 3. A relatively large number of estimated coefficients presents either an unexpected sign and/or is statistically insignificant in estimation. For bankrupt firms, operational cash-flows divided by assets (CFOTA), liabilities or sales have little discriminatory power relative to the baseline model. However, earnings divided by cash-flow (ECFO) provides some evidence of effect for bankrupt firms. Cash-flow growth and financial expenses divided by cash-flow (CFOG and FECFO) respectively provide little evidence of discriminatory power over the baseline hazard model, however. This result is consistent with the findings of similar studies, such as Mazouz *et al.* (2012) and Gupta *et al.* (2014).

Cash-flow variables demonstrate some further discriminatory power over the baseline model in predicting financial distress. For the variable cash-flow divided by total assets (CFOTA) all coefficients are of the expected sign and statistically significant. The AME are also significant, however the correlation coefficients were rather high (especially with earnings variables). For this reason, this variable was not included in the multivariate models. The level of operating cash-flow divided by current liabilities (CFOCL) is of the expected sign and is statistically significant both overall and for most classes of financially distressed firms. This merits inclusion of the variable in the later analysis, again in conjunction with a consideration of any multicollinearity. Where cash-flow is divided by sales (CFOS) the evidence is mixed with statistical significance of a coefficient only found in the case of medium sized firms and an insignificant Wald statistic. The results are similar where earnings are divided by cash-flow from operations (ECFO). However, cash-flow growth (CFOG) provides a more interesting case. For all cases of financially distressed firms, the estimated coefficients are of the expected sign and are statistically significant. The AME are also significant and there is evidence that the effect is different for different size classes of firm. The results for financial expenses (FECFO) are counter intuitive and, unlike the CFOG variable, are therefore excluded from the multivariate models.

In line with our observation in Figure 1, an identical set of covariates is significant in explaining bankruptcy and financial distress for different SMEs size groups in the majority of the cases. However, overall, the coefficients of covariates for micro firms are significantly different from SMEs, both in the case of the bankrupt group of firms and in those financially distressed. This also holds true for a few covariates within the small and medium size groups (e.g. CTA, TLTA, FETA, LRC). As reported in Table 3, the magnitude of coefficients of a given covariate also varies significantly across micro, small and medium firms for both bankrupt and financially distressed sample groups.

These results provide preliminary support for our hypotheses H2, H3 and H4.

[Insert Table 3 Here]

4.4 Multivariate Discrete Hazard Models

In this section, we estimate multivariate discrete-time duration-dependent hazard models with logit link across size categories (respectively for SMEs, micro, small and medium firms) for bankruptcy and financial distress respectively. The dependent variable has a binary outcome with financially distressed/bankrupt equal to '1' and '0' otherwise, while independent variables are the set of covariates found to be significant in the univariate regression analysis. Considering the multicollinearity among the covariates, we introduce each significant covariate in turn into the multivariate setup based on the magnitude (sign is ignored) of their AME. For this, we first rank[7] all the covariates found to be significant (with the expected sign) in the univariate analysis, based on the absolute value of their AME. Following Gupta *et al.* (2018), we then introduce each covariate in turn into the multivariate model in increasing order of the rank of their AME, the rationale being the higher the value of AME, the higher the change in the predicted probability due to a unit change in the covariate. Thus, a covariate with a higher value of AME (e.g. FETA) is more efficient in discriminating between distressed and censored firms than covariates with lower values of AME (e.g. TLTA). Further, if the introduction of a covariate flips the sign of any previously added covariate, then that covariate is excluded from the multivariate model. This situation can possibly occur due to multicollinearity among covariates; accordingly, excluding these covariates seems to be a reasonable choice. We believe that this method of covariate introduction, while developing the multivariate models, leaves us with the best set of covariates with the expected sign of respective coefficients. Initially the multivariate models are estimated using financial ratios obtained from the income statement and balance sheet only, as, amongst others, we examine the information content of cash-flow information. Subsequently, we estimate an additional set of models supplemented with significant operating cash flow ratios. We also control for a volatile macroeconomic environment affecting specific industrial sectors. For this, we calculate an additional measure of industry risk (*Risk*) separately for SMEs, micro, small, and medium firms as the failure (bankrupt/financial distress) rate (number of firms experiencing the event of interest in respective industrial sectors and respective size categories in a given year/total number of firms in that industrial sector and size category in that year) in each of the seven industrial sectors in

[7] However, cash-flow ratios are not included in the ranking process to assess their incremental information content above information obtained from income statement and balance sheet.

a given year[8]. Higher values indicate a higher failure risk, and vice versa. To assess the classification and validation performance of the models developed, area under ROC curves are also estimated (further details of these analyses can be found in Graph A1 and Graph A2 in the Appendix).

4.4.1 Multivariate Hazard Models for Bankrupt Firms

SMEs: We start with a report on the multivariate hazard model for bankrupt SMEs without cash-flow variables in the first column of Table 4. The last rows of this column report on the goodness of fit measures. The Wald Chi Squared measure indicates that the included variables are jointly statistically significant at a 99% level. Both within sample and out-of-sample AUROC statistics confirm that the model performs well, with both values over 70%.

Capital employed divided by total liabilities (CETL) and capital growth (CAG) show the expected negative sign, which is statistically significant at the 99% level. Their AMEs are also statistically significant. Accordingly, firms are less likely to file for bankruptcy when in the presence of higher levels of capital. Similarly, firms with stronger capital growth are also less likely to file for bankruptcy. Taxes (TTA) also have the expected negative effect on the probability of bankruptcy. Although the estimated coefficient is statistically significant at a 95% level, this cannot be said for the AME. Thus, the conclusion on the effect of taxes is mixed, certainly in terms of economic significance. Trade creditors (TCTA) do not have an effect on the basis of the results presented for SMEs, however financial expenses (FETA) do. Both the estimated coefficient and the AME have a statistically significant effect at a 99% level in their positive effect on the probability of bankruptcy. The accuracy of results as well as the signs and significance of few variables for SMEs are compatible with the results of similar studies such as Altman and Sabato (2007) and Altman *et al.* (2010). However, as shown in the next sections, the impact of the variables and their significance vary when we divide SMEs into respective size categories.

The addition of a cash-flow variable, namely earnings divided by cash-flow (ECFO)[9], does bring further predictive power. This is not to the extent that there are any changes in the

[8] Further information on the industrial classification can be found in the Appendix (Table A2).
[9] The Wald test confirms the joint significance of the variables and the AUROC results are equivalent.

interpretation of the variables from the first model. The estimated coefficient for ECFO is negative as expected and the result statistically significant at a 99% level. This provides further support for concluding that earnings mediated by cash-flow is a significant influence on the probability of bankruptcy, as the AME is also statistically significant (though only at a 90% confidence level).

Micro Firms: In the case of the micro category of firms, the previous univariate analysis and the review of the correlation matrix recommended i) the exclusion of taxes (TTA) and trade creditors over total assets (TCTA); and ii) the inclusion of two further variables (namely, short term cash and investments (CTA) and retained earnings (RTA). Both CTA and RETA are not statistically significant for estimation in the multivariate model without the cash-flow variables. Compared to the SMEs category, capital employed (CETL) retains the expected sign and significance for both the coefficient and AME. The result of a Wald test of the equivalence of the coefficients from the micro firms shows that there is no significant difference between the two coefficients. The coefficient on capital growth (CAG) is negative and statistically significant, as in the SME case. Conversely, the result for AME - even with the same sign - loses its statistical significance. The Wald test finds no difference between the value for this coefficient and the one for the SMEs as a whole. The result related to financial expenses (FETA) seems to uncover a different role played by this covariate in predicting bankruptcy. As in the SME case, FETA is found to be statistically significant both in estimation and in the average marginal effect. However, this ratio of financial expenses to cash-flow has an estimated coefficient that is half the size of that for SMEs as a whole, and the AME is much lower for micro firms. This is confirmed by the Wald test for the equality of coefficients, leading to the conclusion that the FETA is significant in predicting bankruptcy but the size of effect is lower.

The results described in this section are not entirely consistent with the relevant studies, such as Gupta *et al.* (2015). This study finds CTA and RETA, two significant indicators of micro firms' financial failure. However, they suggest that micro firms should be considered separately in the process of failure prediction, which is in line with our findings. The inclusion of the previous cash-flow variable (ECFO) does not add to the explanatory power of the model. This result seems to be in contrast to the model for SMEs as a whole.

Small Firms: The third results column for Table 4 references the estimates predicting bankruptcy for small firms, in a model excluding the cash-flow variables. Goodness of fit measures indicate that the included variables are jointly significant and there is a reasonably high predictive power both within and outside sample. Three variables were included based on the univariate analysis: i) capital growth (CAG) and financial expenses (FETA) as in the SMEs model; and ii) ratio of current assets to liabilities (LCR), which has not been considered before. CAG has the expected negative sign for the coefficient and is statistically significant at a 99% level. In scale, the estimated coefficient is larger in absolute terms than that for SMEs as a whole; however, the Wald test does not indicate that there is a significant difference. LCR has the expected negative sign, is statistically significant at a 99% level, and has a significant average marginal effect. FETA appears consistently through our modelling of bankruptcy prediction and has the expected positive sign here with a statistically significant estimated coefficient and AME. However, in contrast to the finding for micro firms, the Wald test for the equality of coefficients does not suggest that there is a difference between small firms and SMEs as a whole.

The addition of the cash-flow variable earnings divided by cash-flow (ECFO) does not add to the explanatory/predictive power of the model in this case, leaving the findings for the first model, immediately above, to stand.

Medium Firms: The model specification[10] for medium firms differs slightly from the univariate analysis. An examination of the correlation coefficients for this sample i) admitted earnings (EBITDATA), liabilities divided by assets (TLTA) and the (log of) current assets divided by current liabilities (LCR) to the model; and ii) excluded the trade creditors over total assets (TCTA). The coefficient on EBITDATA carries the expected negative sign and is statistically significant at a 95% level, although the AME is not statistically significant. Capital employed divided by total liabilities (CETL) is statistically insignificant in estimation. This result seems to be inconsistent with the results of Gupta *et al.* (2015). However, total liabilities divided by total assets (TLTA) has the expected positive coefficient in estimation and is statistically significant at a 95% level. The AME is not statistically significant, suggesting that

[10] The included variables are jointly significant at a 95% level and the model has the highest in sample AUROC value for all of the bankruptcy models.

the finding is not robust across all observations. Capital growth (CAG) was found to have a significant effect on the probability of default in all previous models; conversely, it is not the case for the medium-sized firm category. Taxes (TTA) do not have an effect, although the coefficient is of the expected sign. The (log of) current assets divided by current liabilities (LCR) does have the expected sign for the estimated coefficient, but the AME is not significant. As noted earlier, financial expenses relative to assets (FETA) is of the expected positive sign and also statistically significant. As the coefficient estimate is similar in size to that for SMEs as a whole it is perhaps unsurprising that the Wald test fails to reject the null of equivalence (although of course note should be taken that the test is of a different nature).

No cash-flow variables were suggested as an addition to the above and so no further model results are presented for the medium firms' category.

In summary, from this analysis, we can conclude that factors affecting the bankruptcy of SMEs do vary across size categories. Amongst others, we highlighted the different role play by FETA in predicting bankruptcy of micro-sized firms compared to the SMEs category. Furthermore, in contrast with Gentry *et al.* (1987) and Gilbert *et al.* (1990), CFO information does not provide any marginal improvement in models' classification performance above information obtained from income statements and balance sheets. These results for micro, small and medium SMEs strongly support our two hypotheses H2 and H3. At the same time, they are not entirely consistent with Gupta *et al.* (2015), who conclude that there is no need to consider small and medium categorised firms separately in modelling their credit risk. Finally, the analysis leads us to accept H4 only in the univariate dimension, as operating cash flow information does not add any marginal increment in prediction performance of hazard models above income statement and balance sheet in the multivariate framework. Our results corroborate the empirical evidence provided by Gupta *et al.* (2015) for our sample of UK SMEs.

[Insert Table 4 Here]

4.4.2 Multivariate Hazard Model for Financially Distressed Firms

Firms that have declared bankruptcy have, arguably, experienced greater problems than those that are in some form of financial distress. However, as discussed earlier and following the argument of Keasey *et al.* (2014), financial distress may also be the source of losses to firms and external providers of capital. As noted in Section 3, the proportion of firms of all types that are financially distressed in any year is large relative to the proportion that declare bankruptcy and, perhaps because of the greater number of cases, a larger number of variables enter as significant in the models. Table 5 illustrates multivariate hazard models for financially distressed firms. Goodness of fit measures for all of the estimated models indicate that the included variables are jointly significant at a 99% level. All of the AUROC statistics, both within and out of sample, are over 0.8. These values signal that the estimated models fit well and exhibit a very good performance in predicting financial distress, whether within or out of sample.

SMEs: In line with the analysis performed in Section 4.4.1, we estimate a model in which cash-flow variables have been excluded. Earnings relative to assets (EBITDATA) carry the expected negative coefficient and the variable is statistically significant at a 99% level. This is also the case for the average marginal effect. This conforms to our expectation that higher earnings will forestall financial distress in the following year. The same results apply to i) cash and short-term investments (CTA); ii) capital employed relative to total liabilities (CETL); iii) asset liability ratio (LCR), although the size of effect is relatively small; and iv) capital growth (CAG). The result for CAG carries from the model on bankruptcy. Taxes (TTA) follows the same pattern of the previous variable, showing a particularly robust finding across both financial distress and bankruptcy. This evidence shows that higher taxes reduce the probability of distress with a statistically significant coefficient and AME. Consistent with the preceding set of results, higher financial expenses (FETA) predict an increase in the probability of financial distress. Firms that are not performing well may face greater financial costs, either from restructuring or re-financing, or perhaps as the cost of funds is raised to reflect a perception of increased risk.

The model for financial distress including cash-flow variables adds two further covariates: cash-flow divided by liabilities (CFOCL) and cash-flow growth (CFOG). The effect on the overall fit of the model is minor, with the Wald test and AUROC statistics suggesting that the overall level of explanation is high. The two-included cash-flow variables are statistically significant in estimation and in the AME, both with the expected negative sign.

Amongst all the variables discussed in the previous model, only the log ratio of current assets to liabilities (LCR) has a significant change in the value of the estimated coefficient and AME. These retain the expected negative sign but they more than double in absolute size. Omitting the cash-flow variables would, therefore, bias the finding for LCR.

Micro Firms: The specification of the model for financial distress in micro firms is similar to that for SMEs as a whole. Trade creditors (TCTA) enter the model for micro firms (whereas this was not the case for SMEs), whilst capital employed relative to total liabilities (CETL) was not included this time. Both earnings (EBITDATA) and short-term cash and investments (CTA) have the expected negative sign and are statistically significant in estimation, referencing both the estimated coefficients and the AME. The Wald test on the equality of coefficients rejects the null in both cases, with the conclusion that for both variables the effects are smaller in the case of micro firms. The same conclusions may be drawn in reference to capital growth (CAG) and taxes (TTA). The effects are statistically significant and of the expected sign, but smaller in scale compared to SMEs as a whole. However, when we consider the (log) ratio of current assets to current liabilities (LCR), the finding is different. Although this variable is still statistically significant in estimation, the effect is significantly larger for micro firms (as confirmed by the Wald statistic). Conforming to expectations, the sign on the variable trade creditors (TCTA) is positive and statistically significant in reference to the estimated coefficient and the AME, although only at a 90% confidence level in both cases. It is worthwhile noting that the coefficient is smaller than those for the larger sized classifications. Common across all results, higher financial expenses (FETA) give a greater probability of financial distress. The effect is also larger than the estimated coefficient for the SME model.

The growth in operating cash flow (CFOG) is added to create the further model of financial distress in micro firms. CFOG has the expected negative effect on the probability of distress and is statistically significant in estimation, including AME. There is no qualitative difference between the other results for this model and the previous one.

Small Firms: The model specifications are the same for small and micro firms. Findings on the sign and statistical significance of estimated coefficients and AME are the same, although some differences emerge in the size of effects to which we now turn for the model without the cash-flow variable. Earnings relative to assets (EBITDATA), cash and short-term investments (CTA) and capital growth (CAG) have a larger negative effect on the probability of distress for small firms and the estimated coefficient is statistically different from that of SMEs as a whole. Taxes (TTA) also appear to have a greater negative effect on distress probability than was the case for micro firms, although the Wald test indicates that the effect cannot be distinguished from the effect for SMEs as a whole. The log ratio of current assets to current liabilities (LRC) has very similar effects to those found for micro firms, diminishing the probability of distress as the value rises. Trade creditors (TCTA) have a larger (positive) estimated coefficient but the AME is similar, although we are more certain of this finding in reference to small firms. As in all other cases, financial expenses (FETA) have a significant positive effect on the probability of financial distress.

The above results change slightly with the addition of a cash flow variable (namely, CFOG). CFOG is statistically significant in estimation, as is AME. Therefore, we might conclude that the addition of this variable does add to the predictive power of the model.

Medium Firms: Earnings relative to assets (EBITDATA) remain statistically significant in estimation, with a relatively large coefficient in absolute terms. This suggests the presence of a greater reduction in the likelihood of financial distress in correspondence of increasing earnings. However, this finding carries a greater standard deviation. The low values of AME indicate that the effect is relatively small across all observations. Larger levels of short term debt (STEBV) result in an increase in the probability of financial distress. Both the estimated coefficient and AME are found to be statistically significant. The effect of cash and short-term investments (CTA) is of the expected negative sign, although again with the findings that the relatively large absolute value of the estimated coefficient is reversed in considering the AME. Capital growth (CAG) reduces the chance of financial distress in medium sized firms, as does an incidence of higher tax (TTA). More trade debt results in an increased probability of financial distress, as might be noted from the positive and statistically significant estimated coefficient and AME on trade creditors (TCTA). The consistent finding on financial expenses (FETA) remains. It is worth noticing that the estimated coefficient is statistically different from the result pertaining from SMEs as a whole.

The inclusion of cash-flow variables improves the fit of the model as measured by the log likelihood, and both within and out-of-sample AUROC figures. The lower the conversion rate from earnings into cash-flow (ECFO), the lower the probability of financial distress. In line with previous models, cash-flow growth (CFOG) is found to reduce the distress probability. The scale of effect is similar to that for other size classes.

Although there is not much variation in the factors affecting financial distress of SMEs and its size categories, the impact of those factors varies significantly across size categories. This can be observed from the differences in the magnitude of coefficients and additionally from the significance of Wald statistics. CFO does not seem to play a convincing role in predicting either financial distress or, as described in Section 4.4.1, bankruptcy of the US SMEs. These results strongly support H3, but do not entirely support H2 and H4.

[Insert Table 5 Here]

5. Conclusion

Both bankruptcy and financial distress can impose considerable costs on firms and their suppliers of capital. Consequently, understanding such events is important for firms, financial institutions, shareholders and government. Considerable advances have been made in identifying the determinants as well as the prediction of bankruptcy and financial distress. This paper contributes to this debate by providing important further insights on the determinants and extent to which they affect both bankruptcy and financial distress in different SMEs size categories. Based on our findings, we can conclude that not all SMEs are the same and the determinants of bankruptcy are different across different size categories of SMEs. More specifically, evidence presented in this study offers strong support to the fact that the failure rate of SMEs varies across micro, small and medium size categories (H1). Moreover, factors that are mutually significant in predicting failure likelihood of SMEs and micro, small, or medium firms respectively, exhibit significant differences in the magnitude of their coefficient in respective models (H3). However, our results are not conclusive when we consider the factors affecting SMEs failure (bankruptcy vs financial distress) likelihood across size categories (H2). In particular, earnings and ratio of assets to liabilities are only found to be important for the largest size category. Financial expenses are almost always considered important to predict failure, but the size of the effect varies across different SMEs categories (especially for micro firms). We also present compelling evidence that estimated coefficients are different between models for financial distress estimates concerning SMEs as a whole, as well as across different SMEs size categories. These results seem to suggest that forecasters would therefore be advised that distinct models for bankruptcy or financial distress should be specified not in reference to SMEs as a whole but rather in consideration of the different size classifications.

The presence of cash-flows in explaining SMEs failure leads to mixed results. In contrast to the work of Gentry *et al.* (1987) and Gilbert *et al.* (1990) on SMEs (as a whole), this study does not find that cash-flow contributes to an understanding of bankruptcy (H4). It is, however, in line with the findings of Gupta *et al.* (2014) who also report the inability of CFO information to predict the bankruptcy likelihood of UK SMEs. Our results complement the findings of Charitou *et al.* (2004) for the UK in explaining financial distress using cash-flows in different size categories. Building on Bernard and Stober's (1989) argument, we believe that explanation of these results can be found in the fact that cash-flow figures are easy for managers to manipulate but the incentive to do so may be stronger where the firm is facing bankruptcy, conflating the estimated effect.

There are a number of differences in the estimated determinants of financial distress as opposed to bankruptcy. Firms with greater holdings of cash and short-term investments are less likely to face financial distress. Taxes are consistently found to have an effect on financial distress, but this is not the case for bankruptcy, where only the model across SMEs as a whole provides evidence of a significant effect. There is also evidence that the effect of the ratio of current assets to liabilities is different across different classes of firm in predicting financial distress. The value of trade debt predicts financial distress, concordant with the findings of Hudson (1986) and Beck *et al.* (2006). Possibly the value of trade credit is reduced as a firm appears more likely to file for bankruptcy, explaining the different result.

To conclude, bankruptcy and financial distress are distinct events and separate modeling of them will lead to improved risk pricing. Similarly, our results on US SMEs indicate that separate modeling is required for each SMEs size category. Consideration of both the failure type and the SMEs size category has an impact on the implementation of firm policy on financial structure and cash-flow management, on bank lending policy and government policy relating to support for SMEs, and on legal arrangements surrounding financial restructuring and liquidation.

BIBLIOGRAPHY

Aldrich, E. H. and Auster, R. E., 1986. Even dwarfs started small: Liabilities of age and size and their strategic implications. *Research in Organisational Behaviour*.

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., Iwanicz-Drozdowska, M., Laitinen, E.K., and Suvas, A., 2017. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting*, 28 (2), 131–171.

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.

Balcaen, S. and Ooghe, H., 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38 (1), 63–93.

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.

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

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.

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

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

Casey, C.J. and Bartczak, N.J., 1984. Cash flow-Its not the bottom line. *Harvard Business Review*, 62 (4), 60–66.

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

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.

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.

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 Ebrahimi, T., 2018. Empirical Comparison of Hazard Models in Predicting SMEs Failure. *Quantitative Finance*, 18 (3), 437–466.

Gupta, J., Gregoriou, A., and Healy, J., 2015. Forecasting Bankruptcy for SMEs Using Hazard Function: To What Extent Does Size Matter? *Review of Quantitative Finance and Accounting*, 45 (4), 845–869.

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

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

Hosmer Jr, D.W., Lemeshow, S., and Sturdivant, R.X., 2013. *Applied Logistic Regression*. 3rd ed. New Jersy: John Wiley & Sons.

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

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.

Keasey, K., Pindado, J., and Rodrigues, L., 2015. The determinants of the costs of financial distress in SMEs. *International Small Business Journal*, 33 (8), 862–881.

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

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

Laitinen, E.K., 1993. The use of information contained in annual reports and prediction of small business failures. *International Review of Financial Analysis*, 2 (3), 155–176.

Laitinen, E.K., 1994. Traditional versus operating cash flow in failure prediction. *Journal of Business Finance & Accounting*, 21 (2), 195–217.

Laitinen, E.K., 2005. Survival analysis and financial distress prediction: Finnish evidence. *Review of Accounting and Finance*, 4 (4), 76–90.

Laitinen, T. and Kankaanpaa, M., 1999. Comparative analysis of failure prediction methods: the Finnish case. *European Accounting Review*, 8 (1), 67–92.

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.

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.

Luoma, M. and Laitinen, E.K., 1991. Survival analysis as a tool for company failure prediction. *Omega*, 19 (6), 673–678.

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., 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.

Mossman, C.E., Bell, G.G., Swartz, L.M., and Turtle, H., 1998. An empirical comparison of bankruptcy models. *Financial Review*, 33 (2), 35–54.

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

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

Pompe, P.P.M. and Bilderbeek, J., 2005. The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business venturing*, 20 (6), 847–868.

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.

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

Taffler, R.J., 1983. The assessment of company solvency and performance using a statistical model. *Accounting and Business Research*, 13 (52), 295–308.

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 (3), 417–454.

Turetsky, H.F. and McEwen, R.A., 2001. An empirical investigation of firm longevity: a model of the ex ante predictors of financial distress. *Review of Quantitative Finance and Accounting*, 16 (4), 323–343.

Zavgren, C. V, 1985. Assessing the vulnerability to failure of American industrial firms: a logistic analysis. *Journal of Business Finance & Accounting*, 12 (1), 19–45.

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

List of Tables and Figures

Table 1: Covariates Analysed

Covariates	Definition	Compustat Data Item
EBITDATA	Earnings before interest taxes depreciation and amortization/total assets	EBITDA/AT
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
IATA	Intangible Assets/total assets	INTAN/AT
CFOTA	Cash flow from operations/total assets	OANCF/AT
CFOCL	Cash flow from operations/current liabilities	OANCF/LCT
CFOS	Cash flow from operations/sales	OANCF/SALE
ECFO	Earnings before interest taxes depreciation and amortization /cash flow from operations	EBITDA/OANCF
CFOG	Operating cash flow growth; calculates as (Cash flow from operations $_t$ / cash flow from operations $_{t-1)}$ – 1	OANCF
FECFO	Financial Expenses/ cash flow from operations	XINT/OANCF

Notes: This table lists the set of covariates, along with their respective definitions, used for our empirical analysis. The last column presents the specific Compustat database items we used to construct the covariates.

Table 2: Sample Description

Year	Bankrupt				Financially Distressed			
	SME	Micro	Small	Medium	SME	Micro	Small	Medium
	D | C | %D	D | C | %D	D | C | %D	D | C | %D	D | C | %D	D | C | %D	D | C | %D	D | C | %D
1990	20|1423|1.39	8|361|2.17	5|452|1.09	7|610|1.13	275|1168|19.06	104|265|28.18	99|358|21.66	72|545|11.67
1991	29|1856|1.54	9|485|1.82	10|582|1.69	10|789|1.25	359|1526|19.05	139|355|28.14	120|472|20.27	100|699|12.52
1992	18|1850|0.96	8|454|1.73	5|596|0.83	5|800|0.62	324|1544|17.34	132|330|28.57	104|497|17.30	88|717|10.93
1993	22|1924|1.13	9|438|2.01	6|622|0.96	7|864|0.80	271|1675|13.93	119|328|26.62	82|546|13.06	70|801|8.04
1994	15|1971|0.76	7|453|1.52	3|499|0.50	5|919|0.54	265|1724|13.34	119|341|25.87	78|524|12.96	68|856|7.36
1995	15|1976|0.75	10|424|2.30	4|586|0.68	1|966|0.10	274|1717|13.76	108|326|24.88	89|501|15.08	77|890|7.96
1996	21|1999|1.04	11|425|2.52	3|568|0.53	7|1006|0.69	301|1719|14.90	120|316|27.52	101|470|17.69	80|933|7.90
1997	22|2215|0.98	11|472|2.28	5|626|0.79	6|1117|0.53	317|1920|14.17	125|358|25.88	101|530|16.01	91|1032|8.10
1998	20|2081|0.95	8|466|1.69	8|574|1.37	4|1041|0.38	369|1732|17.56	129|345|27.22	113|469|19.42	127|918|12.15
1999	18|1916|0.93	7|451|1.53	7|537|1.30	4|933|0.43	413|1521|21.35	152|306|33.19	125|414|23.19	136|801|14.51
2000	14|1996|0.70	9|491|1.80	2|521|0.38	3|984|0.30	393|1617|19.55	151|349|30.20	116|407|22.18	126|861|12.77
2001	18|1962|0.91	13|558|2.28	1|508|0.20	4|896|0.44	431|1549|21.77	119|372|34.85	116|393|22.79	116|784|12.89
2002	12|1978|0.60	8|627|1.26	1|480|0.21	3|871|0.34	588|1402|29.55	299|336|47.09	142|339|29.52	147|727|16.82
2003	12|1854|0.64	8|638|1.24	3|441|0.68	1|775|0.13	514|1325|28.99	282|364|43.65	132|312|29.73	127|649|16.37
2004	11|1802|0.61	8|640|1.23	1|394|0.25	2|768|0.26	475|1338|26.20	261|387|40.28	97|298|24.56	117|653|15.19
2005	13|1693|0.76	12|611|1.93	1|387|0.26	0|695|0.00	384|1322|22.51	215|408|34.51	87|301|22.42	82|613|11.80
2006	10|1545|0.64	6|562|1.06	3|335|0.89	1|648|0.15	389|1166|25.02	220|348|38.73	88|250|26.04	81|568|12.48
2007	5|1443|0.35	4|516|0.77	0|308|0.00	1|619|0.16	360|1088|24.86	194|326|37.21	72|326|23.38	94|526|15.16
2008	5|1235|0.40	3|460|0.65	0|266|0.00	2|509|0.39	330|910|26.61	169|294|36.50	73|193|27.44	88|423|17.22
2009	5|1232|0.40	2|442|0.45	1|258|0.39	2|532|0.37	400|837|32.34	209|235|47.07	82|177|31.66	109|425|20.41
2010	2|1155|0.17	1|426|0.23	1|244|0.41	0|485|0.00	363|794|31.37	196|231|45.90	67|178|27.35	100|385|20.62
2011	1|1055|0.09	0|405|0.00	0|221|0.00	1|429|0.23	279|777|26.42	157|248|38.77	60|161|27.15	62|368|14.42
2012	1|1044|0.10	1|416|0.24	0|218|0.00	0|410|0.00	291|754|27.85	173|244|41.49	57|161|26.15	61|349|14.88
2013	2|1099|0.18	2|494|0.40	0|228|0.00	0|377|0.00	348|753|31.61	173|244|41.49	67|161|29.39	74|303|19.63
2014	0|311|0.00	0|134|0.00	0|76|0.00	0|101|0.00	96|215|30.87	66|68|49.25	15|61|19.74	15|86|14.85

Notes: This table represents the sample of US SMEs used in our analysis. Year-wise description of bankrupt and financially distressed SMEs, micro, small, and medium firms are listed in respective columns. In "D | C | %D", 'D' represents number of events of interest witnessed, 'C' represents number of censored observations and '%D' is calculated as D/(D+C)×100.

Table 3: Univariate Hazard Analysis

		Bankrupt Firms				Financially Distressed Firms			
Section A: Non-Cash Flow Covariates									
Covariates	Sign	SME	Micro	Small	Medium	SME	Micro	Small	Medium
EBITDATA	-								
β		-0.9120[a]	0.0210	-0.4133	-2.5247[a]	-3.0847[a]	-2.3999[a]	-3.2123[a]	-6.0694[a]
SE		0.2014	0.3026	0.4576	0.5616	0.0510	0.0745	0.1133	0.1909
dy/dx [R]		-0.446[a][5]	0.018[.]	-0.169[.]	-0.156[a][5]	-31.18[a][5]	-48.90[a][5]	-30.86[a][6]	-15.73[a][4]
Wald Sig.		----	----	----	0.0275[b]	----	0.0000[a]	0.2959	0.0000[a]
STDEBV	+								
β		-0.4559[b]	-0.2621	0.0262	-.05236	-0.3531[a]	-1.0013[a]	-0.0900	0.6608[a]
SE		0.2146	0.3493	0.4241	0.4338	0.0436	0.0741	0.0821	0.0825
dy/dx [R]		-0.274[b][.]	-0.134[.]	0.017[.]	-0.101[.]	-4.047[a][.]	-21.19[a][.]	-1.08[.]	3.039[a][9]
Wald Sig.		----	----	----	----	----	----	----	----
CTA	-								
β		-1.3072[a]	-0.8276[c]	-4.4336[a]	-3.3977[a]	-2.0163[a]	-2.0799[a]	-2.9400[a]	-2.4626[a]
SE		0.3558	0.4510	1.0984	1.0629	0.0804	0.0998	0.1804	0.2011
dy/dx [R]		-0.84[a][4]	-0.474[c][4]	-0.129[b][5]	-0.109[a][9]	-21.89[a][6]	-43.33[a][6]	-33.16[a][4]	-10.15[a][6]
Wald Sig.		----	0.5034	0.0200[b]	0.0753[c]	----	0.6421	0.0000[a]	0.0189[b]
RETA	-								
β		-0.1386[a]	-0.0774[b]	-0.0210	-0.2568[a]	-0.2969[a]	-0.2662	-0.2753[a]	-0.3923[a]
SE		0.0227	0.0335	0.0559	0.0645	0.0053	0.0079	0.0111	0.0165
dy/dx [R]		-0.07[a][10]	-0.041[b][8]	-0.008[.]	-0.05[b][10]	-3.45[a][10]	-5.636[a][.]	-3.25[a][11]	-1.75[a][11]
Wald Sig.		----	0.1191	----	0.1560	----	----	0.0830[c]	0.0000[a]
CETL	-								
β		-0.4190[a]	-0.2435[a]	-0.8874[a]	-1.5806[a]	-1.0624[a]	-0.7092[a]	-1.0799[a]	-1.8825[a]
SE		0.0528	0.0595	0.1757	0.2474	0.0189	0.0189	0.0387	0.0601
dy/dx [R]		-0.21[a][8]	-0.085[a][6]	-0.127[b][6]	-0.167[4]	-5.512[a][9]	-10.84[a][9]	-5.78[a][10]	-1.38[a][12]
Wald Sig.		----	0.0276[b]	0.0240[b]	0.0000[a]	----	0.0000[a]	0.6846	0.0000[a]
TLTA	+								
β		1.9864[a]	1.2101[a]	1.9772[a]	3.6502[a]	3.0732[a]	2.4592[a]	3.1442[a]	4.7703[a]
SE		0.1672	0.2290	0.3518	0.3859	0.0430	0.0548	0.0949	0.1372
dy/dx [R]		1.18[a][3]	0.596[a][3]	0.116[c][7]	0.123[8]	31.690[a][4]	49.83[a][4]	32.76[a][5]	14.149[a][5]
Wald Sig.		----	0.0137[b]	0.7661	0.0008[a]	----	0.0000[a]	0.4916	0.0000[a]
CAG	-								
β		-0.7420[a]	-0.3330[a]	-1.3020[a]	-1.5398[a]	-0.7450[a]	-0.2824[a]	-1.1131[a]	-1.9619[a]
SE		0.1166	0.1424	0.3114	0.3244	0.0213	0.0263	0.0507	0.0724
dy/dx [R]		-0.426[a][6]	-0.163[b][5]	-0.523[b][3]	-0.332[a][3]	-8.259[a][8]	-5.95[a][10]	-12.11[a][8]	-6.501[a][7]
Wald Sig.		----	0.0272[b]	0.1268	0.0322[b]	----	0.0000[a]	0.0000[a]	0.0000[a]
TTA	-								
β		-11.5546[a]	0.6543	-9.0858	-17.4265[a]	-24.0440[a]	-9.8952[a]	-25.717[a]	-27.6011[a]
SE		4.1909	8.5017	8.4212	6.7336	1.0332	2.0325	2.1411	1.6183
dy/dx [R]		-7.097[a][2]	0.245[.]	-0.244[.]	-0.661[b][2]	-201[a][2]	-208[a][2]	-302[a][1]	-103[a][2]
		----	----	----	0.4435	----	0.0000[a]	0.4766	0.0669[c]
LCR	-								
β		-0.7593[a]	-0.2315[b]	-1.2774[a]	-1.952[a]	-1.3956[a]	-1.1283[a]	-1.4131[a]	-1.8027[a]
SE		0.0900	0.1112	0.2284	0.2602	0.0226	0.0287	0.0484	0.0625
dy/dx [R]		-0.381[a][7]	-0.083[c][7]	-0.288[c][4]	-0.127[c][7]	-12.96[a][7]	-22.04[a][7]	-13.43[a][7]	-4.976[a][8]
Wald Sig.		----	0.0003[a]	0.0892[c]	0.0003[a]	----	0.0000[a]	0.7365	0.0000
TCTA	+								
β		3.9728[a]	2.3930[a]	3.5891[a]	7.4751[a]	7.8283[a]	6.8795[a]	8.5441[a]	7.7549[a]
SE		0.6489	0.8787	1.5202	1.5738	0.1511	0.1926	0.3410	0.3769
dy/dx [R]		0.210[a][9]	0.836[b][2]	0.55[b][2]	0.141[a][6]	87.33[a][3]	144[a][3]	94.87[a][3]	34.68[a][3]
Wald Sig.		----	0.1632	0.5463	0.0769[c]	----	0.0001[a]	0.0530[b]	0.8853
FETA	+								
β		18.7490[a]	9.9275[a]	21.6106[a]	35.7482[a]	21.4210[a]	17.4174[a]	21.328[a]	30.8303[a]
SE		1.8667	2.5761	4.0511	4.1180	0.4052	0.5206	0.8292	1.0895
dy/dx [R]		9.383[a][1]	3.6[a][1]	9.063[c][1]	11.247[c][1]	205[a][1]	370[a][1]	256[a][2]	131[a][1]
Wald Sig.		----	0.0079[a]	0.7503	0.0019[a]	----	0.0000[a]	0.9247	0.0000[a]
IATA	+								
β		-1.1954[c]	-1.4373	0.9257	-1.4606	0.1770	0.5922[a]	0.8760[a]	0.4645[a]
SE		0.7317	1.1445	1.3502	1.4294	0.1305	0.1983	0.2662	0.2589

dy/dx [R]		-0.682[.]	-0.76[.]	0.361[.]	-0.172[.]	2.018[.]	12.48^{a}[8]	10.61^{a}[9]	2.151^{c}[10]
Wald Sig.		----	----	----	----	----	----	----	----
Section B: Cash Flow Covariates									
CFOTA	-								
β		-0.4433^{b}	0.1812	0.2735	-1.7382^{a}	-2.3708^{a}	-1.7235^{a}	-2.2602^{a}	-3.9332^{a}
SE		0.2145	0.2944	0.5488	0.6228	0.0482	0.0673	0.1045	0.1498
dy/dx		-0.265^{b}	0.063	0.112	-0.444^{b}	-26.84^{a}	-35.93^{a}	-25.56^{a}	-16.65^{a}
Wald Sig.		----	----	----	0.0682^{c}	----	0.0000^{a}	0.3410	0.0000^{a}
CFOCL	-								
β		0.1224^{c}	0.4140^{a}	0.1858	-0.2102	-.09476^{a}	0.3118^{a}	-0.1124^{a}	-0.6640^{a}
SE		0.0659	0.0931	0.1505	0.1674	0.0144	0.0197	0.0284	0.0365
dy/dx		0.068^{c}	0.142^{a}	0.078	-0.038	-1.087^{a}	6.55^{a}	-1.35^{a}	-3.00^{a}
Wald Sig.		----	----	----	----	----	----	0.5691	0.0000^{a}
CFOS	-								
β		0.0032	0.0117^{a}	0.0013	0.2024	-0.0052^{a}	0.0009	-0.0034	-0.0060^{c}
SE		0.0027	0.0032	0.0146	0.2779	0.0005	0.0006	0.0025	0.0037
dy/dx [R]		0.018	0.004^{a}	0.005	0.037	-0.06^{a}	0.020	-0.041	-0.02^{c}
Wald Sig.		----	----	----	----	----	----	----	0.8754
ECFO	-								
β		-0.1044^{a}	-0.1227^{c}	-0.1849^{b}	-0.1000	0.0321	0.0913^{a}	0.0410^{a}	-0.0690^{a}
SE		0.0404	0.0666	0.0876	0.0766	0.0083	0.0140	0.0168	0.0151
dy/dx		-0.056^{a}	-0.058^{c}	-0.347^{c}	-0.029	0.366^{a}	1.923^{a}	0.400^{b}	-0.320^{a}
Wald Sig.		----	0.7901	0.4273	----	----	----	----	----
CFOG	-								
β		-0.0352	-0.0474	-0.0018	-0.1143	-0.0477^{a}	-0.0878^{a}	-0.0345^{b}	-0.0419^{a}
SE		0.0376	0.0607	0.0764	0.0737	0.0080	0.0136	0.0157	0.0145
dy/dx		-0.021	-0.024	-0.001	-0.023	-0.543^{a}	-1.851^{a}	-0.417^{b}	-0.194^{a}
Wald Sig.		----	----	----	----	----	0.0117^{b}	0.4518	0.7266
FECFO	+								
β		-0.0585	0.1489	0.1800	-0.2574	-0.4518^{a}	-0.4536^{a}	-0.4109^{a}	-0.4253^{a}
SE		0.1239	0.2258	0.2387	0.2210	0.0277	0.0512	0.0515	0.0475
dy/dx		-0.034	0.007	0.112	-0.072	-5.184^{a}	-9.59^{a}	-4.95^{a}	-1.986^{a}
Wald Sig.		----	----	----	----	----	----	----	----

Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table presents the univariates analysis results for bankrupt and financially distressed groups of firms. Further, under each group analysis results are listed for SMEs, micro, small and medium firms respectively. '*β*' is the regression coefficient, '*SE*' is the standard error and '*dy/dx*' is the average marginal effect (AME) obtained from univariate regression estimates of respective covariates for respective groups. *dy/dx* for the bankrupt group of firms are multiplied by 10,000, while for financially distressed firms it is multiplied by 100. '*Wald Sig.*' is the p-value obtained from the Wald test of equality of coefficients of significant covariates of SMEs with other groups (micro, small and medium firms) for respective significant covariates. '[R]' is the rank of covariates obtained by arranging the absolute values of AME of respective covariates in ascending order. The 'Sign' column reports the expected sign of the respective covariates in the regression analysis.

Table 4: Multivariate Hazard Models for Bankrupt Firms

Covariates	Sign	Without CFO Ratios				With CFO Ratio			
		SMEs	Micro	Small	Medium	SMEs	Micro	Small	Medium
EBITDATA	-								
β					-1.9328[b]				
SE					0.9562				
dy/dx					-0.0038				
Wald Sig.									
STDEBV	+								
β									
SE									
dy/dx [R]									
Wald Sig.									
CTA	-								
β			-0.1565				-0.2122		
SE			0.5995				0.6019		
dy/dx			-0.013				-0.0154		
Wald Sig.									
RETA	-								
β			-0.0325				-0.0383		
SE			0.0495				0.0498		
dy/dx			-0.002				-0.0027		
Wald Sig.									
CETL	-								
β		-0.2016[a]	-0.2337[a]		-0.1132	-0.2045[a]	-0.2305[a]		
SE		0.0608	0.0825		0.3242	0.0612	0.0826		
dy/dx		-0.044[a]	-0.0168[b]		-0.0002	-0.040[a]	-0.0167[b]		
Wald Sig.			0.6671		0.8513				
TLTA	+								
β					2.7366[a]				
SE					0.9734				
dy/dx					0.0054				
CAG	-								
β		-0.5451[a]	-0.3118[c]	-1.4726[a]	-0.3651	-0.5441[a]	-0.3111[c]	-1.4657[a]	
SE		0.1209	0.1677	0.4894	0.3489	0.1210	0.1682	0.4899	
dy/dx [R]		-0.121[a]	-0.0225	-1.424[a]	-0.0007	-0.109[a]	-0.0225	-1.108[a]	
Wald Sig.			0.2293	0.1041	0.7390				
TTA	-								
β		-8.7883[b]			-3.5408	-8.2080[c]			
SE		4.7165			9.7533	4.7329			
dy/dx		-1.940			-0.007	-1.643			
Wald Sig.									
LCR	-								
β				-1.6270[a]	-0.9449[b]			-1.6309[a]	
SE				0.4069	0.4472			0.4080	
dy/dx				-1.573[a]	-0.0018			-1.233[a]	
Wald Sig.									
TCTA	+								
β		1.2189				1.3165			
SE		0.8229				0.8274			
dy/dx		0.270				0.263			
Wald Sig.									
FETA	+								
β		15.7840[a]	7.3781[b]	19.999[b]	17.1025[b]	15.8410[a]	7.4598[b]	20.1024[a]	
SE		2.3852	3.5178	7.9746	8.7496	2.3970	3.5242	7.9932	
dy/dx		3.490[a]	0.532[c]	-19.336[b]	-0.038	3.171[a]	0.540[c]	15.202[a]	
Wald Sig.			0.0500[b]	0.6424	0.9453				
IATA	+								
β									

SE								
dy/dx								
Wald Sig.								
CFOTA	-							
β								
SE								
dy/dx								
Wald Sig.								
CFOCL	-							
β								
SE								
dy/dx								
Wald Sig.								
CFOS	-							
β								
SE								
dy/dx								
Wald Sig.								
ECFO	-							
β						-0.1009[a]	-0.0928	-0.1869
SE						0.0429	0.0751	0.1282
dy/dx						-0.020[c]	-0.0067	-0.141
Wald Sig.								
CFOG	-							
β								
SE								
dy/dx								
Wald Sig.								
FECFO	+							
β								
SE								
dy/dx								
Wald Sig.								
Goodness of Fit								
Wald chi2		202.92[a]	73.11[c]	61.65[b]	76.45[b]	206.38[a]	73.91[c]	62.03[b]
Log likelihood		-1339.90	-592.310	-288.406	-340.618	-1337.25	-591.554	-287.360
AUROC								
Within Sample		0.7578	0.7193	0.7943	0.8879	0.7579	0.7226	0.7945
Out-of-sample		0.7065	0.5897	0.8091	0.7275	0.7169	0.5987	0.8224
Number of Observations		40171	11237	8273	14944	40171	11237	8273
Bankrupt		311	165	70	76	311	165	70
Censored		39860	11072	8203	14868	39860	11072	8203

Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table presents multivariate regression results for the bankrupt group of firms for respective size categories with and without operating cash flow information. Here, '*β*' is the regression coefficient, '*SE*' is the standard error and '*dy/dx*' (multiplied by 10,000) is the average marginal effect (AFE). '*Wald Sig.*' is the p-value obtained from the Wald test of equality of coefficients of significant covariates of SMEs with other groups (micro, small and medium firms) for respective significant covariates. The 'Sign' column reports the expected sign of the respective covariates in the regression analysis. The bottom part of this table reports goodness of fit measures and classification performance measures of respective multivariate models developed.

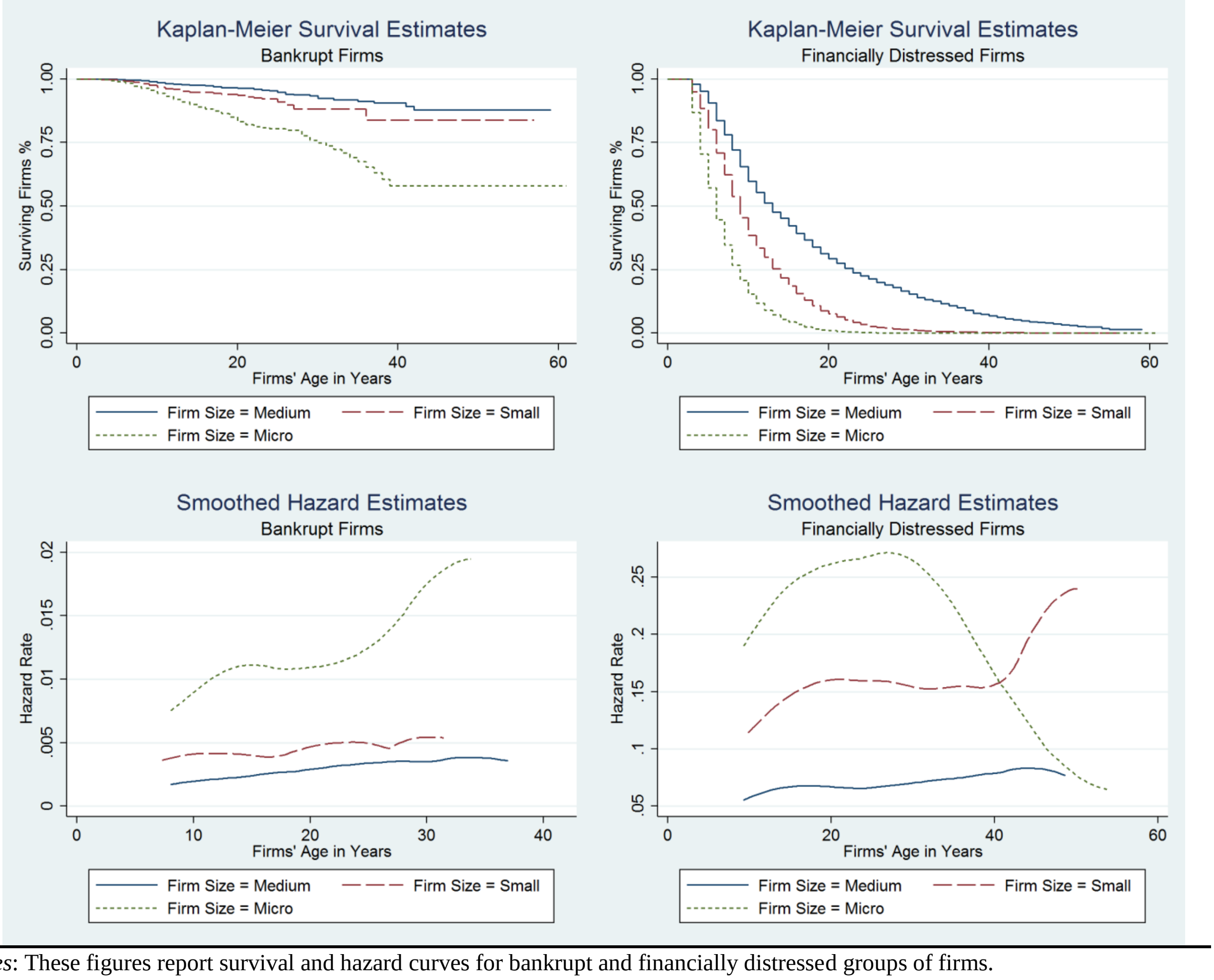
Table 5: Multivariate Hazard Models for Financially Distressed Firms

Covariates	Sign	Without CFO Ratios				With CFO Ratio			
		SMEs	Micro	Small	Medium	SMEs	Micro	Small	Medium
EBITDATA	-								
β		-2.2485^{a}	-1.587^{a}	-2.5870^{a}	-5.6399^{a}	-1.9957^{a}	-1.6409^{a}	-2.6292^{a}	-5.7394^{a}
SE		0.0588	0.0870	0.1294	0.2186	0.0697	0.0876	0.1305	0.2207
dy/dx		-13.323^{a}	-30.619^{a}	-20.226^{a}	-9.323^{a}	-11.334^{a}	-31.636^{a}	-19.521^{a}	-9.328^{a}
Wald Sig.			0.0000^{a}	0.0171^{b}	0.0000^{b}				
STDEBV	+								
β					0.6538^{a}				0.6531^{a}
SE					0.0945				0.0946
dy/dx [R]					1.080^{a}				1.061^{a}
Wald Sig.									
CTA	-								
β		-0.7884^{a}	-0.3738^{a}	-1.1278^{a}	-2.4368^{a}	-0.7966^{a}	-0.3787^{a}	-1.1218^{a}	-2.4184^{a}
SE		0.1034	0.1297	0.2367	0.2638	0.1031	0.1297	0.2370	0.2635
dy/dx		-4.639^{a}	-7.212^{a}	-8.381^{a}	-4.028^{a}	-4.524^{a}	-7.301^{a}	-8.329^{a}	-3.930^{a}
Wald Sig.			0.0124^{b}	0.1879	0.0000^{a}				
RETA	-								
β									
SE									
dy/dx									
Wald Sig.									
CETL	-								
β		-0.6689^{a}				-0.6860^{a}			
SE		0.0213				0.0216			
dy/dx		-3.936^{a}				-3.896^{a}			
Wald Sig.									
TLTA	+								
β									
SE									
dy/dx									
Wald Sig.									
CAG	-								
β		-0.3922^{a}	-0.1685^{a}	-0.6643^{a}	-1.0562^{a}	-0.3717^{a}	-0.1636^{a}	-0.6596^{a}	-1.048^{a}
SE		0.0231	0.0300	0.0523	0.0688	0.0230	0.0301	0.0523	0.0688
dy/dx [R]		-2.308^{a}	-3.251^{a}	-4.936^{a}	-1.746	-2.111^{a}	-3.154^{a}	-4.897^{a}	-1.704^{a}
Wald Sig.			0.0000^{a}	0.0000^{a}	0.0000^{a}				
TTA	-								
β		-16.012^{a}	-6.7385^{a}	-19.659^{a}	-14.727^{a}	-15.490^{a}	-6.7609^{a}	-19.518^{a}	-14.169^{a}
SE		1.1665	2.3129	2.6044	1.9925	1.1717	2.3107	2.600	1.9950
dy/dx		-94.227^{a}	-130.01^{a}	146.105^{a}	-24.346^{a}	-87.975^{a}	-130.03^{a}	-144.92^{a}	-23.028^{a}
Wald Sig.			0.0003^{a}	0.2012					
LCR	-								
β		-0.0662^{b}	-0.7920^{a}	-0.7630^{a}		-0.1415^{a}	-0.7868^{a}	-0.7613^{a}	
SE		0.0333	0.0395	0.0637		0.0347	0.0395	0.0637	
dy/dx		-0.390^{b}	-15.281^{a}	-5.670^{a}		-0.803^{a}	-15.170^{a}	-5.652^{a}	
Wald Sig.			0.0000^{a}	0.0000^{a}					
TCTA	+								
β			0.4286^{c}	1.2110^{a}	2.0290^{a}		0.3378^{c}	1.1844^{a}	1.9727^{a}
SE			0.2549	0.3989	0.4708		0.2554	0.3990	0.4714
dy/dx			8.270^{c}	9.000^{a}	3.354^{a}		6.513^{c}	8.794^{a}	3.206^{a}
Wald Sig.									
FETA	+								
β		3.4678^{a}	5.4674^{a}	7.0595^{a}	21.8413^{a}	3.4341^{a}	5.4749^{a}	7.0160^{a}	21.7840^{a}
SE		0.5001	0.6383	1.0545	1.2862	0.4981	0.6386	1.0547	1.2868
dy/dx		20.407^{a}	105.48^{a}	52.465^{a}	36.106^{a}	19.503^{a}	105.55^{a}	52.093^{a}	35.404^{a}
Wald Sig.			0.0137^{b}	0.0021^{a}	0.0000^{a}				
IATA	+								

β									
SE									
dy/dx									
Wald Sig.									
CFOTA	-								
β									
SE									
dy/dx									
Wald Sig.									
CFOCL	-								
β						-0.1850[a]			
SE						0.0247			
dy/dx						-1.050[a]			
Wald Sig.									
CFOS	-								
β									
SE									
dy/dx									
Wald Sig.									
ECFO	-								
β									-0.0969[a]
SE									0.0189
dy/dx									-0.157[a]
Wald Sig.									
CFOG	-								
β						-0.0608[a]	-0.0946[a]	-0.0583[a]	-0.0442[a]
SE						0.0094	0.0155	0.0188	0.0183
dy/dx						-0.345[a]	-1.825[a]	-4.328[a]	-0.071[a]
Wald Sig.							0.0313[b]	0.3710	0.1527
FECFO	+								
β									
SE									
dy/dx									
Wald Sig.									
Goodness of Fit									
Wald chi2		5152.95[a]	1811.69[a]	1222.76[a]	1408.23[a]	5134.59[a]	1832.63[a]	1226.34[a]	1411.44[a]
Log likelihood		-13638.3	-5671.45	-3825.34	-4462.15	-13594.7	-5652.85	-3820.51	-4445.62
AUROC									
Within Sample		0.8873	0.8224	0.8613	0.8918	0.8873	0.8243	0.8619	0.8930
Out-of-sample		0.8680	0.8403	0.8425	0.8276	0.8674	0.8416	0.8432	0.8278
Number of Observations		40922	11995	10682	18207	40922	11995	10682	18207
Financially Distressed		8836	4245	2283	2308	8836	4245	2283	2308
Censored		32086	7750	8399	15899	32086	7750	8399	15899

Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table presents multivariate regression results for the financially distressed group of firms for respective size categories with and without operating cash flow information. Here, '*β*' is the regression coefficient, '*SE*' is the standard error and '*dy/dx*' (multiplied by 100) is the average marginal effect (AFE). '*Wald Sig.*' is the p-value obtained from the Wald test of equality of coefficients of significant covariates of SMEs with other groups (micro, small, and medium firms) for respective significant covariates. The 'Sign' column reports the expected sign of the respective covariates in the regression analysis. The bottom part of this table reports goodness of fit measures and classification performance measures of respective multivariate models developed.

Figure 1: Table of Survival and Hazard Curves



Notes: These figures report survival and hazard curves for bankrupt and financially distressed groups of firms.

Appendix

Descriptive Statistics on Failure Rate Across SMEs Size Categories

Information on the sample for the dependent variables, disaggregated by SMEs size classification and failure (bankrupt and financially distressed firms), is presented in Table A1. For each default classification the values for all SMEs are presented, followed by values for micro, small, and medium firms. Each column contains three figures. The first, "D", is the number of firms experiencing a default event in the year/category concerned. The second, "C", is the number of censored (relevant non-default) observations and %D is calculated as D/(D+C) ×100.

Turning to the columns headed "SME", note that the incidence of financial distress is much greater than bankruptcy. There is an apparent decline in incidence from the start of the sample period, especially from around 2001. This is also the case in terms of the percentage of firms that are filing for bankruptcy, falling from 1.39% in 1990 to less than 0.5% in 2007. In contrast, occurrences of financial distress have been rather more stable or increased over time. The largest number of distress events was in the years 2002, 2003 and 2004, although a relatively large increase is also observed in 2009. The former may coincide with the end of the dotcom boom and the latter, with the more recent financial crisis, 2009 being a year that the US experienced negative GDP growth. The percentage of SMEs under financial distress has also generally increased over the sample period, from 19% in 1990 to over 30% in 2014. This percentage has not fallen significantly from the peak in 2009 as the replacement rate has clearly not been sufficient to counter the large proportion of filings that are observed.

Descriptive statistics for the explanatory variables are presented in Table A1, disaggregated by firms filing for bankruptcy and in financial distress. Section A of Table A1 reports on firms filing for bankruptcy, with columns "C" representing the censored group of firms and "D" representing the group that experienced the default event (bankruptcy or financial distress). With the exception of Micro firms, the earnings (EBITDATA) of SMEs that filed for bankruptcy are worse than for non-filing firms, as might be expected. Cash and short-term investments (CTA) are lower for all classes of filing firms. The same is also true for capital growth (CAG) which is negative for all classes of filing firms, in contrast the non-filing. Bankruptcy filing firms also have lower levels of intangible assets relative to total assets (IATA). Overall, cash flow relative to assets (COFTA) is worse for bankrupt firms but this is not a consistent finding across size classes. A similar finding is observed when the denominator is sale (CFOS). Earnings relative to cash-flow (ECFO) is worse for all classes of filing firms. Cash-flow growth (CFOG) is also worse overall for the bankrupt firms, although perhaps surprisingly this is not consistent across groups. Interpretation of the financial expenses variable is more difficult given that a number of cash-flows are negative.

Section B of Table A1 presents basic statistics on variables for financially distressed firms, where the larger sample size facilitates interpretation. Earnings (EBITDATA) are worse for all classes of financially distressed firms. On average short-term debt relative to equity book value (STDEBV) is also lower (although the finding is not consistent). Cash holdings are consistently lower for distressed firms, as are retained earnings, presumably reflecting the greater difficulty that such firms are facing in servicing their financial commitments. Capital employed relative to total liabilities (CETL) is also lower for financially distressed firms. The finding for capital growth is the same as for filing firms, negative for all classes of financially distressed firms. For all but medium sized firms, current assets are lower than current liabilities, with a negative log of the ratio in contrast to firms that are not distressed. This might be expected, but the contrast suggests that this variable might be particularly important in predicting default. Trade credit relative to total assets is also larger for all classes of distressed firm. Financial expenses (FETA) are more than twice as large for distressed firms, although the standard deviation is understandably quite large for values that cut across many different years and types of firm. A finding more consistent with the literature than was the case for bankruptcy is that intangible assets (IATA) are greater for the distressed as opposed to non-distressed firms. Descriptive statistics on cash-flow measures are generally consistent. With only two exceptions, cash-flow relative to assets, liabilities and sales (CFOTA, CFOCL and CFOS) are worse for distressed firms. The exceptions are micro firms where cash-flows are measured relative to liabilities and sales. This therefore suggests the value in considering these ratios in the prediction of default. Further evidence is presented on this matter when considering earnings relative to cash-flow (ECFO), which are higher for distressed firms, except the medium size category. This suggests that distressed firms are less able to convert earnings to cash-flow. Cash-flow growth (CFOG) is also lower for financially distressed firms, as are financial expenses relative to cash-flow (FECFO).

Table A1: Descriptive Statistics

Covariates	Section A: Bankrupt Firms							
	SME		Micro		Small		Medium	
	C	D	C	D	C	D	C	D
EBITDATA								
Mean	-0.2049	-0.3084	-0.5560	-0.4539	-0.1706	-0.1899	0.0042	-0.1017
Median	-0.0510	-0.1327	-0.5176	-0.3435	-0.0641	-0.0806	0.0603	0.0028
SD	0.4130	0.4334	0.4456	0.4625	0.3577	0.3757	0.2288	0.2754
STDEBV								
Mean	0.1034	0.0437	0.0058	-0.0267	0.1231	0.1114	0.1557	0.1345
Median	0.0087	0	0	0	0.0147	0.0033	0.0307	0
SD	0.3441	0.4341	0.3372	0.3313	0.3586	0.4878	0.3261	0.5442
CTA								
Mean	0.2655	0.2198	0.3418	0.2976	0.2585	0.1469	0.2199	0.1183
Median	0.1613	0.0709	0.2535	0.1557	0.1612	0.0446	0.1320	0.0470
SD	0.2599	0.2669	0.2990	0.2987	0.2493	0.2094	0.2247	0.1740
RETA								
Mean	-2.5666	-4.2548	-5.3681	-6.1627	-2.3369	2.4005	-.8716	-1.8208
Median	-0.8798	-2.164	-4.6119	-7.3364	-1.0274	-1.2817	-0.1625	-0.8284
SD	3.5758	4.1384	4.0742	4.1326	3.1649	2.8172	1.9722	2.9872
CETL								
Mean	2.6031	1.3570	2.7868	1.8086	2.5931	1.0003	2.4890	0.7053
Median	1.5743	.4508	1.2492	0.3450	1.5665	0.5401	1.6827	0.4856
SD	2.6965	2.4149	3.2746	3.0640	2.6559	1.3180	2.2605	0.9892
TLTA								
Mean	0.5931	1.0194	0.7652	1.0524	0.5719	0.9310	0.4932	1.0293
Median	0.4495	0.9699	0.5188	1.2092	0.4497	0.8277	0.4308	0.9131
SD	0.4824	0.5710	0.6381	0.6355	0.4496	0.5167	0.3274	0.4572
CAG								
Mean	0.1857	-0.1508	0.1199	-0.1213	0.1728	-0.2106	0.2363	-0.1599
Median	0.0169	-0.2920	-0.0944	-0.3687	0.0044	-0.2432	0.0555	-0.2486
SD	0.7936	0.7379	0.9111	0.8449	0.7888	0.5433	0.7058	0.6424
TTA								
Mean	0.0077	0.0035	0.0012	0.0028	0.0062	0.0059	0.0128	0.0031
Median	0	0	0	0	0	0	0.0012	0
SD	0.0231	0.0199	0.0129	0.0173	0.0218	0.0239	0.0274	0.0211
LCR								
Mean	0.5916	0.0219	0.2854	0.0759	0.6510	-0.0362	0.7568	-0.0416
Median	0.6488	-0.1028	0.1666	-0.2232	0.6644	0.0249	0.7585	-0.0693
SD	1.0465	1.1119	1.3243	1.2963	0.9890	0.9503	0.8020	0.7687
TCTA								
Mean	0.1211	0.1648	0.1573	0.1830	0.1162	0.1360	0.1003	0.1517
Median	0.0766	0.1044	0.0879	0.1124	0.0775	0.0747	0.0730	0.1320
SD	0.1177	0.1481	0.1514	0.1651	0.1090	0.1273	0.0882	0.1204
FETA								
Mean	0.0312	0.0586	0.0433	0.0558	0.0319	0.0611	0.0230	0.0625
Median	0.0139	0.0468	0.0156	0.0327	0.0163	0.0532	0.0123	0.0507
SD	0.0398	0.0513	0.0508	0.0555	0.0385	0.0471	0.0289	0.0453
IATA								
Mean	0.0749	0.0468	0.0618	0.0294	0.0667	0.065	0.0882	0.0669
Median	0	0	0	0	0	0	0.0133	0.0033
SD	0.1350	0.1133	0.1362	0.0981	0.1274	0.1339	0.1373	0.1189
CFOTA								
Mean	-0.1777	-0.2412	-0.4535	-0.3654	-0.1466	-0.1251	-0.0158	-0.0786
Median	-0.0449	-0.0723	-0.3703	-0.2186	-0.0544	-0.0296	0.0264	-0.0237
SD	0.3633	0.4089	0.4271	0.4591	0.3170	0.3319	0.2028	0.2359
CFOCL								
Mean	-0.5031	-0.4221	-1.3040	-0.6389	-0.4990	-0.2382	0.0174	-0.1208
Median	-0.1203	-0.1076	-0.8409	-0.1991	-0.1540	-0.0371	0.0910	-0.0519
SD	1.3030	1.1656	1.4574	1.4102	1.2603	0.9052	0.8909	0.5345
CFOS								

Mean	-10.6050	-15.1467	-33.3286	-27.6337	-2.2182	-2.0802	-0.6751	-0.0718
Median	-0.0492	-0.0894	-3.6520	-0.6408	-0.0555	-0.0205	0.0221	-0.0130
SD	28.749	34.6362	43.6190	43.2090	11.7680	12.4812	7.0432	0.3036
ECFO								
Mean	1.1013	0.8596	1.3699	1.0054	1.0099	0.6738	0.9794	0.7143
Median	1.0732	0.7621	1.1708	0.8612	1.0145	0.7684	1.0105	0.5383
SD	1.8143	1.9500	1.7200	1.8848	1.7601	1.8834	1.8859	2.1412
CFOG								
Mean	-0.1176	-0.2539	0.0956	-0.1598	-0.1837	-0.13561	-0.2182	-0.5673
Median	-0.1806	-0.5192	-0.0590	-0.475	-0.2261	-0.4918	-0.2567	-0.7161
SD	1.9075	2.0441	1.7678	2.0072	1.9329	2.2120	1.9684	1.9561
FECFO								
Mean	-0.0017	-0.0180	-0.11580	-0.09690	0.0044	0.1188	0.0691	0.0270
Median	-0.00002	-0.0077	-0.0158	-0.0035	-0.0021	-0.0005	0.0012	-0.0616
SD	0.5514	0.74830	0.4763	0.5982	0.5695	0.8082	0.5736	0.9475
Section B: Financially Distressed Firms								
EBITDATA								
Mean	-0.1139	-0.5392	-0.4370	-0.7699	-0.1025	-0.4221	0.0378	-0.2308
Median	0.0128	-0.4500	-0.3327	-1.0766	-0.0069	-0.3085	0.0799	-0.1394
SD	0.3568	0.4323	0.4287	0.3936	0.3185	0.3818	0.1984	0.2827
STDEBV								
Mean	0.1175	0.0502	0.0453	-0.0676	0.1294	0.0995	0.1465	0.2184
Median	0.0132	0	0	-0.0509	0.0172	0	0.0286	0.0693
SD	0.3003	0.4690	0.2784	0.4142	0.3160	0.4867	0.2963	0.4857
CTA								
Mean	0.2789	0.2153	0.3902	0.2515	0.2785	0.1815	0.2248	0.1821
Median	0.1816	0.1005	0.3867	0.1224	0.1907	0.0877	0.1393	0.0856
SD	0.2620	0.2462	0.3019	0.2719	0.2539	0.2145	0.2256	0.2145
RETA								
Mean	-1.8372	-5.2749	-4.2463	-7.4520	-1.8051	-4.2977	-0.6780	-2.2373
Median	-0.5192	-4.4052	-2.6929	-10.1678	-0.7138	-3.0994	-0.0869	-1.2023
SD	3.0510	4.0500	3.8988	3.5433	2.7827	3.6660	1.7555	2.7439
CETL								
Mean	3.1292	0.6487	3.9980	0.5322	3.0841	0.7359	2.7289	0.7768
Median	2.1091	0.3408	3.1116	-0.1069	2.0553	0.5055	1.9267	0.7099
SD	2.7523	1.1598	3.3638	1.3621	2.7292	1.0859	2.2995	0.7233
TLTA								
Mean	0.4626	1.0819	0.5035	1.2553	0.4629	0.9843	0.4425	0.8598
Median	0.3587	0.9835	0.2601	1.6921	0.3651	0.8567	0.3878	0.7610
SD	0.3802	0.5117	0.5287	0.5277	0.3710	0.4815	0.2849	0.3837
CAG								
Mean	1.0819	-0.1025	0.1867	-0.0115	0.2660	-0.1822	0.2964	-0.1911
Median	0.9835	-0.2842	-0.0502	-0.2899	0.0537	-0.3204	0.0830	-0.2521
SD	0.5117	0.7667	0.9144	0.8897	0.7887	0.6775	0.7039	0.5578
TTA								
Mean	0.0097	0.0003	0.0018	0.0003	0.0079	1.95e-06	0.0146	0.0006
Median	0	0	0	0	0	0	0.0027	0
SD	0.0247	0.0133	0.0141	0.0105	0.0235	0.0119	0.0281	0.0182
LCR								
Mean	0.8150	-0.2397	0.7414	-0.5573	0.8405	-0.0682	0.8374	0.1747
Median	0.8665	-0.2988	1.008	-1.1981	0.8665	-0.0601	0.8438	0.1779
SD	0.9610	0.9289	1.2780	0.9393	0.9364	0.8448	0.7735	0.7671
TCTA								
Mean	0.0991	0.2026	0.1113	0.2424	0.0983	0.1828	0.0935	0.1488
Median	0.0649	0.1663	0.0512	0.2657	0.0659	0.1483	0.0691	0.1180
SD	0.0985	0.1444	0.1284	0.1542	0.0951	0.1300	0.0815	0.1147
FETA								
Mean	0.0234	0.0607	0.0287	0.0704	0.0253	0.0570	0.0197	0.0465
Median	0.0098	0.0482	0.0055	0.0720	0.0120	0.0447	0.0105	0.0356
SD	0.0328	0.0491	0.0432	0.0530	0.0333	0.0459	0.0254	0.0401
IATA								
Mean	0.0731	0.0803	0.0563	0.0707	0.0626	0.0819	0.0869	0.0964

Median	0	0	0	0	0	0	0.0129	0.0166
SD	0.1322	0.1441	0.1303	0.1449	0.1231	0.1413	0.1362	0.1441
CFOTA								
Mean	-0.1118	-0.4190	-0.3639	-0.6140	-0.1017	-0.3114	0.0058	-0.1669
Median	-0.0096	-0.2868	-0.2605	-0.7710	-0.0232	-0.1899	0.0391	-0.0898
SD	0.3132	0.4274	0.3988	0.4315	0.2843	0.3724	0.1805	0.2735
CFOCL								
Mean	-0.4337	-0.7523	-1.4773	-0.9611	-0.4480	-0.6792	0.0832	-0.4407
Median	-0.0343	-0.3290	-1.3033	0.4573	-0.0816	-0.3025	0.1498	-0.1885
SD	1.3534	1.0583	1.5681	1.1622	1.3168	0.9941	0.8822	0.8026
CFOS								
Mean	-9.0921	-16.2593	-34.0816	-31.7291	-2.0840	-2.7081	-0.5946	-1.2109
Median	-0.0105	-0.40788	-3.9621	-3.0557	-0.0253	-0.1756	0.0331	-0.0760
SD	26.9017	34.2552	43.8976	43.0612	11.4163	12.9902	6.8406	8.1899
ECFO								
Mean	1.0666	1.2190	1.2568	1.5628	0.9922	1.0647	1.0130	0.7393
Median	1.0617	1.1113	1.1259	1.2824	1.0112	1.0189	1.0364	0.8051
SD	1.7882	1.9066	1.6158	1.8873	1.7362	1.8487	1.8863	1.8755
CFOG								
Mean	-0.10333	-0.1746	0.1606	-0.0333	-0.1628	-0.2591	-0.2007	-0.3506
Median	-0.1530	-0.2733	0.0133	-0.1940	-0.1930	-0.3447	-0.2373	-0.4238
SD	1.9529	1.737	1.8537	1.6027	1.9674	1.8075	1.9807	1.8768
FECFO								
Mean	0.0368	-0.1426	-0.0708	-0.1972	0.0354	-0.1061	0.0902	-0.0782
Median	-1.35e-07	-0.0893	-0.0030	-0.0892	-5.94e-07	-0.1000	0.0072	-0.0828
SD	0.5141	0.6573	0.4086	0.5756	0.5273	0.6997	0.5441	0.7417

Notes: This table reports mean, median and standard deviation (SD) of respective covariates for SMEs, micro, small and medium firms. Here, ‘D’ represents firms which have experienced the event of interest and ‘C’ represents the censored group of firms. Section A reports the details for our sample of bankrupt firms, while section B reports details of the financially distressed group of firms.

Sample Industrial Classification

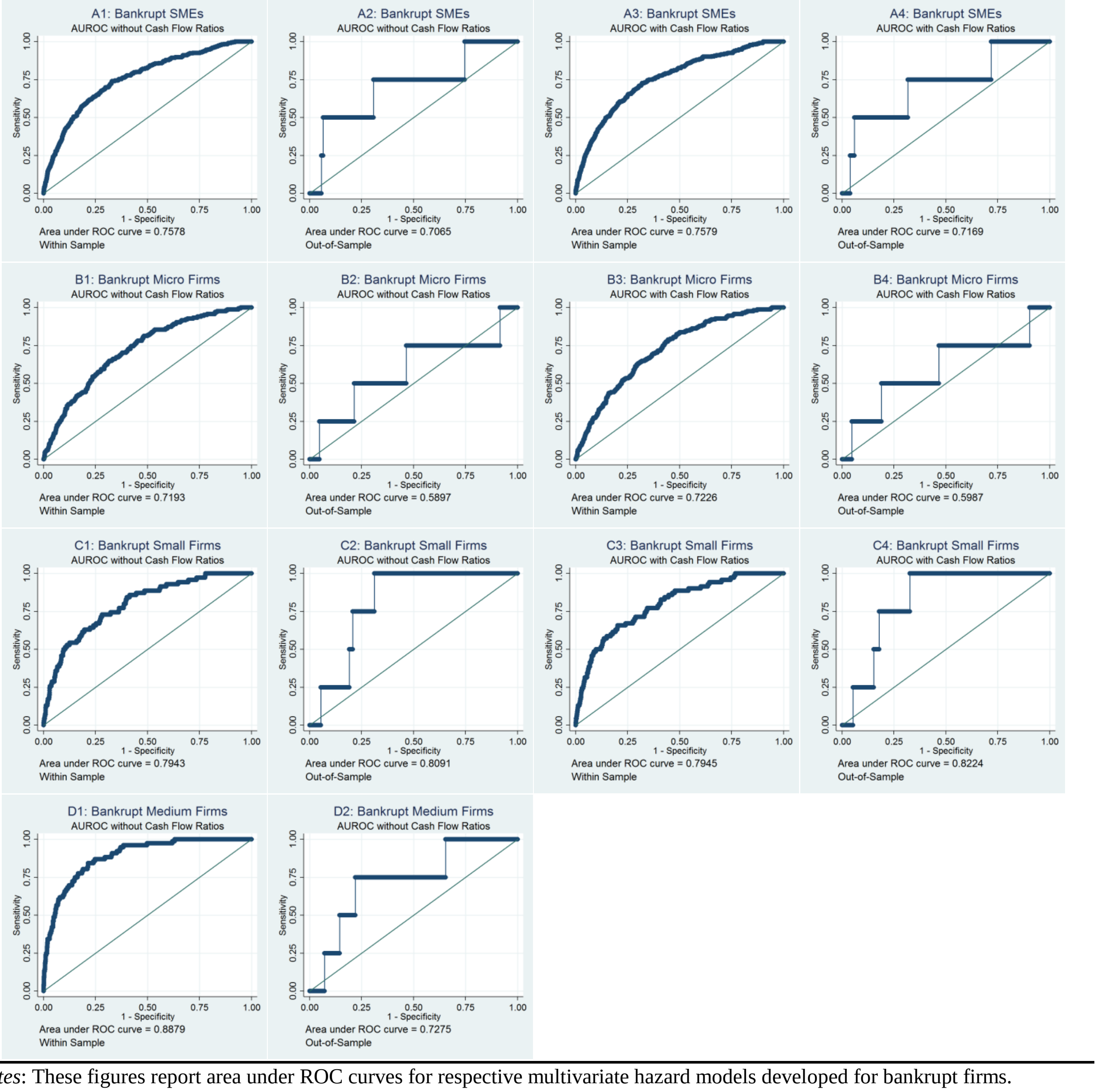
Table A2: Sample Industrial Classification

Industry Code	SIC Code	Industry
1	< 1000	Agriculture, Forestry, Fishing
2	1000 to < 1500	Mining
3	1500 to < 1800	Construction
4	2000 to < 4000	Manufacturing
5	5000 to < 5200	Wholesale Trade
6	5200 to < 6000	Retail Trade
7	7000 to < 8900	Services

Notes: This table reports Standard Industrial Classification (SIC) of US firms. SIC Code is a four-digit code that represents a given industrial sectors.

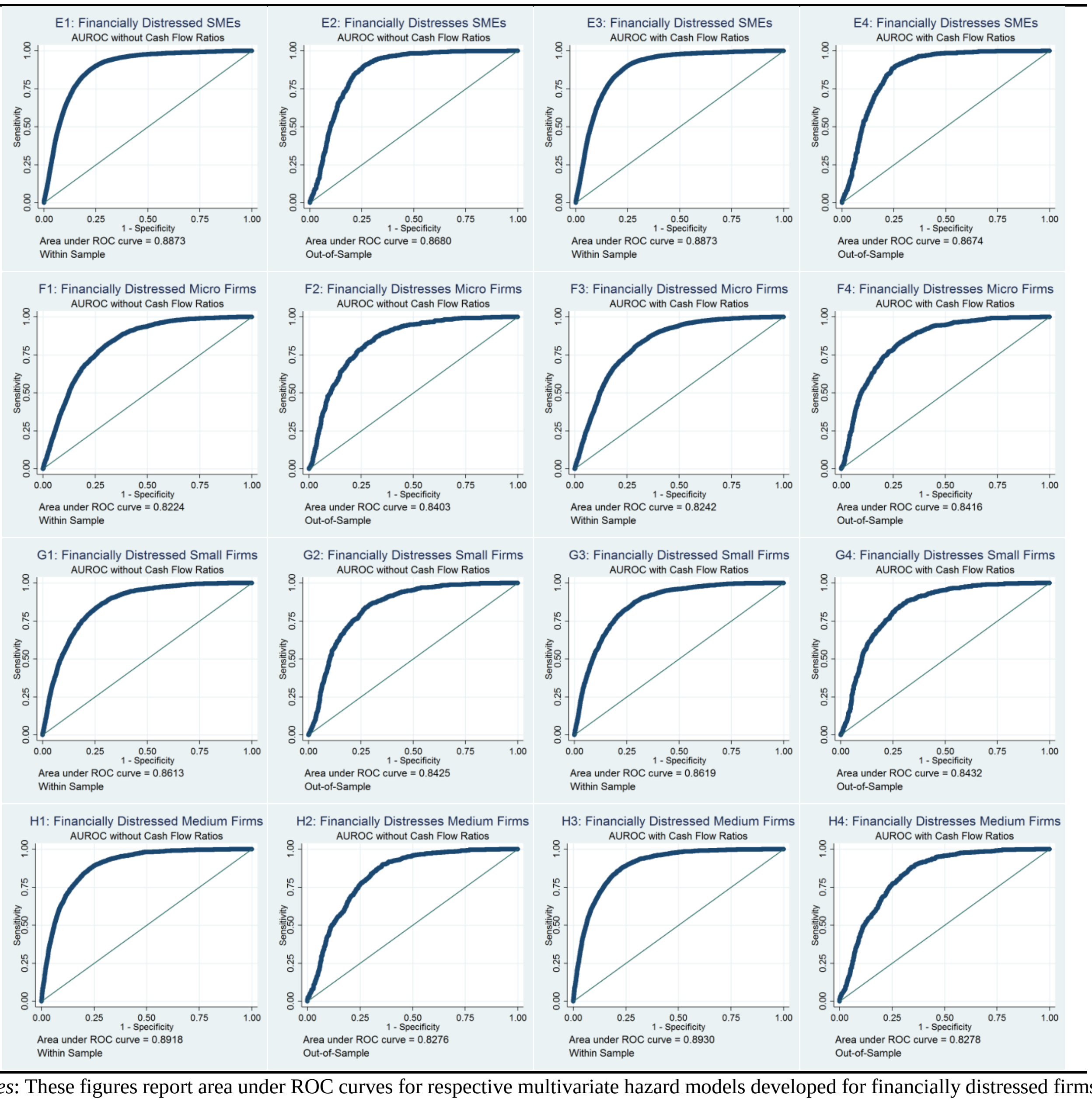
Area Under ROC Curves

Figure A1: Table of Area under ROC curves for Bankrupt Firms



Notes: These figures report area under ROC curves for respective multivariate hazard models developed for bankrupt firms.

Figure A2: Table of Area under ROC curves for Financially Distressed Firms



Notes: These figures report area under ROC curves for respective multivariate hazard models developed for financially distressed firms.