USING HAZARD MODELS CORRECTLY: A COMPARISON EMPLOYING DIFFERENT DEFINITIONS OF SMES FINANCIAL DISTRESS

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

Andros Gregoriou
Department of Finance and Economics, University of Brighton,
Brighton, BN2 4AT, UK; email: A.Gregoriou@brighton.ac.uk

Tahera Ebrahimi
Department of Finance, University of Hull,
Hull, HU6 7RX, UK; email: T.Ebrahimi@2011.hull.ac.uk

June, 2015

Abstract

We consider neglected issues while developing discrete and continuous-time duration-dependent hazard models for predicting default of US SMEs. Comparing theoretical and performance aspects of discrete hazard models with logit and clog-log links, and continuous Cox model, we report superior performance of discrete-time models. However, involving into complications of hazard models is unrewarding, as discrete hazard model with logit link is panel logistic regression that controls firms' age. Additionally, our proposed default definition based on bankruptcy laws and firms' financial health performs superior than its alternative counterparts in identifying distressed firms, with superior goodness of fit measures across all econometric specifications.

Keywords: Bankruptcy; SMEs; Discrete Hazard Models; Cox Proportional Hazard; Financial Distress

JEL Classification Codes: G33; C25; C41; C53

1. INTRODUCTION

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

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

If there are no differences, then the Cox model shall be a reasonable and convenient choice, as it does not require any *baseline hazard* specification unlike discrete-time models (see Rabe-Hesketh and Skrondal 2012). Baseline hazard is defined as the hazard rate when the value of all covariates is set to zero. The baseline hazard is estimated using time dummies (Beck *et al.* 1998) or some other functional form of time (Jenkins 2005). However, recent studies seems to have distorted this idea of baseline hazard and have established their own version of baseline hazard that includes macroeconomic variables (Nam *et al.* 2008), insolvency risk (Gupta *et al.* 2014) etc., in the baseline hazard function, while many prefer not to include any baseline hazard function in their model (see among others Campbell *et al.* 2008, Bauer and Agarwal 2014). In light of the basic theory of survival analysis, this is inappropriate. Thus, we address this misleading concern in this study and show the steps that need to be followed in specifying the baseline hazard function while developing a discrete hazard model. On the other hand, studies which employ continuous-time Cox models are silent on the critical test of proportional hazards (PH) assumptions for time-independent covariates (e.g. Liang and Park 2010). The PH assumption implies that the hazard rate of any particular subject is a constant proportion of the hazard rate of any other subject across time (Mills 2011). The violation of this assumption might lead to overestimation (the covariate violates this assumption and exhibit an increasing hazard ratio over time) or underestimation (the covariate violates this assumption and exhibit a decreasing hazard ratio over time) of hazard risk (Mills 2011). It also results in incorrect standard errors and decrease in the power of significance tests (Box-Steffensmeier and Zorn 2002). The violation of PH assumption is a frequent phenomenon and thus, it should always be checked and reported in studies. Having said that, Allison (2010) warns that, it is not enough to worry only about the violation of the PH assumption but also about other basic requirements, such as incorporation of relevant explanatory variables. Although all the covariates that we employ in this study are time-dependent, if one also employs time-independent covariates, then one should take cognizance of this serious and neglected concern and use appropriate methods to test, report and rectify any violation of the proportional hazards assumptions[1].

Another highly neglected area of concern is *frailty* and *recurrent* events. Correlation of event time occurs when firms experiencing default event belong to a particular cluster or groups like industry, geographic location etc. or in case of recurrent events, where a firm experiences a default event more than once in its lifetime. In the United States (US), the Bankruptcy Reform Act of 1978 (Code) governs the legal processes involved in dealing with corporate financial distress. It allows firms facing financial distress for a liquidation process (Chapter 7) or a reorganization process (Chapter 11) [2]. Chapter 7 leads to permanent shut down of a financially distressed firm, while Chapter 11 aims at rehabilitation of financially distressed but economically viable firms. Hotchkiss (1995) examines 197 publicly traded firms that filed for Chapter 11 protection during 1979 to 1988 and later recovered from Chapter 11 as publicly traded firms. He reports that 40% of the firms continue to experience operating losses and 32% either restructure their debt or re-enter bankruptcy in the three years

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

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

following the acceptance of reorganization plans. Thus a firm may witness multiple distress events in its lifetime. Given that these issues of clustering and recurrent events are an integral part of the real-life environment, they should be made an essential and standard part of contemporary event history analysis (see Box-Steffensmeier and De Boef (2006) and Mills (2011) for advanced discussion). The solution is to introduce a *frailty* term in the hazard models. *Frailty* is an unobserved random proportionality factor that modifies the hazard function to account for random effects, association and unobserved heterogeneity into hazard models (Mills 2011). Not including a frailty term implicitly assumes that all firms are homogeneous, which implies that all the firms are prone to experience default in the same way, with the duration of defaults considered as independent from one another. However, in real-life some firms are more 'frail' and thus provide a higher likelihood to experience default. Therefore, our empirical analysis also accounts for this neglected concern while developing the hazard models.

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

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

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

Our test results obtained by employing firm-year observations of the US SMEs provide convincing evidence. To establish the empirical validation, we calculate a wide range of financial ratios that gauges a firm's performance from liquidity, solvency, profitability, leverage, activity, growth and financing dimensions. Then following the suggestion of Hosmer Jr *et al.* (2013) we use appropriate strategy to narrow down our list of covariates, and develop multivariate models. First, in line with the theoretical arguments discrete-time duration-dependent hazard models that we develop with logit and complementary log-log (clog-log) links provides superior model fit than Extended Cox models, as they have much lower AIC values than Cox models across all default definitions. However, all three econometric specifications lead to almost identical within sample and hold-out sample classification performance which contradicts the findings of Shumway (2001). Thus one might be tempted to be indifferent in her choice of hazard specification. However, the coefficients of most of the covariates in all multivariate models show the pattern $|logit| > |cloglog| > |cox|$. This states that the absolute value of a covariate's coefficient is highest when estimated using discrete-hazard logit specification and lowest for continuous cox specification. Thus, for a unit change in the value of a covariate, logit estimates leads to highest change in the outcome probability than its alternative counterparts. Hence, we suggest the use of discrete-time hazard model with logit link to model interval censored data. But, if the event of interest is not duration dependent (i.e. some functional form of time or time dummies are not significant in the multivariate model) with the hazard rates being invariant or varies mildly across different time periods, then getting involved into the complications of hazard models is not rewarding considering the marginal gain one would obtain using such models. While developing our multivariate models we find that, in presence of other financial covariates about 90% of time dummies that we use as baseline hazard specification are insignificant with very high values of standard errors. Thus we follow Shumway (2001) and use natural logarithm of firms annual age (AGE) as baseline hazard specification. It is significant in most of our multivariate hazard models, but such objective can easily be achieved by developing regression models using logistic regression techniques that use some functional form of time to capture any duration dependency. Although Shumway (2001) argue that hazard models are superior to competing static models but ln(age) variable in his multivariate models are insignificant, then how can it be used to reliably predict duration specific hazard rate, which is why hazard models are primarily used. Unlike areas like medicine or health economics, duration specific prediction of hazard rates is not common in bankruptcy/financial distress prediction, thus we do not see any real need of hazard models if similar objective can be achieved using much simpler logistic regression that controls for any duration dependencies.

Second, the default definition that we propose (Event 3) performs best in classifying defaulted firms. Thus a default definition based on firms' financial health is superior to default definition based on legal consequences, while a default definition that considers both legal consequence and firms' financial health is best. These differences in classification performance emphasises the fact that all firms that file for legal bankruptcy are not based purely due to financial difficulties. A significant number of firms do consider this as a planned exit strategy (Bates 2005). Additionally, we also test the efficiency and stability of covariates suggested by a celebrated study on US SMEs bankruptcy prediction by Altman and Sabato (2007). Based on our test results, we conclude that the covariates suggested by them fail to exhibit satisfactory discriminatory across all default definitions and lagged time periods (T-1, T-2 and T-3), and there are several other financial ratios which are better performers. Their suggestion might be biased due to 'cherry picking' being involved in their sample selection process, while our study employs near population data of US SMEs.

The rest of this paper is organized as follows: section 2 discusses various default definitions that we consider in our study; section 3 provides discussion related to our dataset, choice of covariates and methodology; in section 4 we report and discuss our empirical findings and finally, section 5 concludes our findings.

2. Different Default Definitions for SMEs

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

Further, a financially distressed firm may go for a formal reorganization involving the court system or an informal reorganization through the market participants. *Debt restructuring*, *asset sale* and *infusion of new capital from external sources* are the three commonly used market-based/private methods of resolving financial distress (Senbet and Wang 2010). Debt restructuring allows a financially distressed firm to renegotiate the outstanding debt obligation or related credit terms with its creditor/s but is critically subject to whether the debt obligation is due to private or public entity. As an alternative to this, a distressed firm may sell-off some of its existing assets to reduce its outstanding liability or may undertake new profitable investment opportunities, which may eventually help it to overcome its misery. Despite having profitable investment opportunities, a financially distressed firm might not be able to generate additional funding due to high risk involved in financing distress firms and the "debt overhang" problem as discussed in Myers (1977). As a consequence, infusion of new capital from external sources is rarely observed in the resolution of financial distress. Thus, we cannot rule out the possibility that a financially distressed firm may not file for Chapter 7 or Chapter 11 protection and choose a private workout method of resolving financial distress. Gilson et al. (1990) and Gilson (1997) report that firms avoid legal bankruptcy processes by out of court negotiation with creditors. However, under the binary classification based on legal consequences, a finically distressed firm which has not filed for Chapter 7 or Chapter 11 is not considered as a financially distressed firm. Thus, there is a clear need of a mechanism to identify finically distressed firms beyond the legal definitions. In this context, we find the argument of Pindado et al. (2008) highly relevant and thus we explore the following definitions of SMEs' default events:

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

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

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

3. EMPIRICAL METHODS

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

3.1 DATASET

To predict default events over the next one year horizon, our empirical analysis employs annual firm-level accounting data from the Compustat database. We consider a relatively long analysis period that includes all SMEs that entered the Compustat database after January 1950 but before April 2015. We define SMEs as firms having less than 250 employees. In Compustat, a company has "TL" footnote on status alert (data item STALT) indicating that the company is in bankruptcy or liquidation (i.e. Chapter 7/11). Generally, a company will have a "TL" footnote on status alert - quarterly (and annual) for the first and following quarters (and years) the company is in Chapter 11. An "AG" footnote will appear on Total Assets (AT_FN) – quarterly, on the quarter the company emerges from Chapter 11. Thus, within its lifetime a firm may go for multiple bankruptcy filings in form of Chapter 11 and may remain in the bankruptcy state until it emerges. Consequently, taking the bankruptcy filing date as the bankruptcy indicator ignores the possible subsequent bankruptcy states. Thus, our first definition (Event 1) consider a firms under *bankruptcy* when its status alert is “TL” and healthy otherwise. This, classification is consistent with the basic notion of survival analysis in which a subject may remain in a given risky state for more than one time period and thus experience an event of interest for more than one time period.

Table 1 reports the age-wise distribution of censored and distressed firms under respective default events (see section 2 for definitions of various default events). We proxy a firm’s age as the earliest year for which, financial information for that firm is available in the Compustat database. In Compustat, 1950 is the earliest point in time for which financial information is available. Thus, in order to get the complete financial history of a firm, we selected only those firms which entered the Compustat database after 1950. Further, firms belonging to Transportation, Communications & Public Utilities; Finance; Insurance & Real Estate; and Public Administration industrial sectors have been excluded from our empirical analysis (see Table 2 for details). It should be noted that same firms might have multiple entry and exits in our database. For instance, when an existing SME reports number of employees over 250, it exits our sample and returns only when its number of employees falls below 250. We also exclude subsidiary firms (if ‘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).

[Insert Table 1 Here]

[Insert Table 2 Here]

3.2 Selection of Variables

Dependent Variable: Considering the discussion presented in section 2, in this study we consider *Event 1*, *Event 2* and *Event 3* as dependent variables for respective hazard models.

Independent Variables: To develop multivariate hazard models we employ wide range of financial ratios that have established reputation in predicting firms’ default risk. Our choice of covariates gauges firms’ performance from leverage, liquidity, solvency, activity, profitability and interest coverage dimensions. Specifically, we incorporate covariates from popular studies on SMEs bankruptcy like Altman and Sabato (2007), Lin *et al.* (2012), Gupta *et al.*(2014) and a recent book on credit risk management by Joseph (2013). To get detailed understanding pertaining to our choice of covariates and their relationship with the default probability, we strongly recommend one to go through these references. Table 3 lists all the covariates along with their respective definition. To eliminate the influence of extreme outliers on our statistical estimates, we restrict the range of all our financial ratios between 5th and 95th percentiles.

Control Variables: Considering the suggestion of Gupta, Gregoriou, *et al.* (2014) we control for the diversity between micro, small and medium firms by employing dummy variables for micro (firms with less than 10 employees) and small (firms having greater than 9 but less than 50 employees) firms into our multivariate models. To control the volatile macroeconomic environment affecting specific industrial sector, we calculate an additional measure of industry risk (*RISK*) as the failure rate (number of firms experiencing the event of interest in respective industrial sector in a given year/total number of firms in that industrial sector in that year) in each of the seven industrial sectors in a given year. Higher values indicates higher risk and vice versa.

[Insert Table 3 Here]

3.3 HAZARD MODEL

3.3.1 BASIC HAZARD MODEL

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

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

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

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

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

3.3.2 EXTENDED COX MODEL

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

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

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

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

Where $x(t)$ is the covariate vector at time t. The rate of change of time-varying covariates is different for different subjects and hence the estimated hazard ratio does not remain constant over time. However, the inclusion of time-varying covariates is not problematic for the partial likelihood estimation (Allison 2010) and hence CPH model can be easily improved to allow for non-proportional hazard risks. It implies that a general hazard model which does not have the restrictive assumption of constant proportional hazard ratio can be generalized by

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

inclusion of both duration-dependent and duration-independent covariates in the same model. However, a CPH model with time-varying covariates is no longer a proportional hazards model and a CPH model with time-varying covariates is appropriately called *Extended Cox* model (see Kleinbaum and Klein 2012). Additionally the time-varying covariates do not need to satisfy the proportional hazards assumption. However, if the model also includes time-independent covariates, then appropriate test of proportionality is suggested (see Kleinbaum and Klein 2012). One major advantage of Cox method is that it easily addresses the problem of right censoring but it suffers from a major disadvantage of proportional hazards assumption if time-independent covariates are also included in the model. One may consider to test[4] this restrictive proportional hazard assumption, that is being neglected in most empirical studies by using the *scaled Schoenfeld residual* (Grambsch and Therneau 1994) rather than the *Schoenfeld residual* (Schoenfeld 1982). While estimating our empirical model we also control for *unobserved heterogeneity* and *recurrent events* by including a *shared frailty* term into our model via a multiplicative scaling factor α_i (Cleves *et al.* 2010). These signifies group-level frailty and are unobservable positive values assumed to follow the *Gamma distribution* with mean 1 and variance θ to be estimated using the development sample (Jenkins 2005). Also, the time at which the distress event occurs is not really relevant for hazard risk analysis using Cox method, but the ordering of the distress event is critically important. In situations where multiple firms experience the event of interest at the same time, exact ordering of distress event is difficult. Thus we use *Efron*[5] (1977)'s method to handle cases of tied failure times.

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

3.3.3 Discrete Hazard Model

When an event may be experienced at any instant in continuous-time (exact censoring and survival times are recorded in relatively fine time scales such as seconds, hours or days) and

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

there are no *tied* survival time periods, then continuous-time survival model is an appropriate choice (Rabe-Hesketh and Skrondal 2012). However, if the data has relatively few censoring or survival times with *tied* survival time periods, then discrete-time survival model is more appropriate where coarse times-scales are generally used, for instance, expressing time to event in weeks, months or years (Rabe-Hesketh and Skrondal 2012). Interval-censoring[6] leads to discrete-time data, which is the case with our database. Here, the beginning and end of each time interval is same for all the SMEs in analysis time, as the information is recorded on annual basis. Thus, the event of interest may take place at any time within the year but it cannot be known until the information is provided at the end of the year. Hence, considering the discussion above we also estimate our hazard models in discrete-time framework with *random effects* (α_i), thus controlling for *unobserved heterogeneity* or *shared frailty*.

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

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

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

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

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

Where *α(t)* captures baseline hazard rate and $P_{i,t}$ is the probability of experiencing the event by subject *i* at time *t*. This will produce very similar results as long as the time intervals are small (Rabe-Hesketh and Skrondal 2012) and sample bad rate (% of failed to non-failed) is very low (Jenkins 2005). One may also choose probit link function, if one strongly believes that the underlying distribution of the process being modelled is normal, or if the event under study is not a binary outcome but a proportion (e.g. proportion of population at different income levels). While these specifications will generally yield results that are quite similar

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

but there are significant differences in terms of non-proportionality (see Sueyoshi (1995) for detiled discussion). Thus considering this discussion, we estimate our discrete hazard models with clog-log and logit links and analyse any differences in the classification performance of the models developed.

3.3.4 Specification of Baseline Hazard Rate

The final step before estimation of discrete-time hazard model is the specification of baseline hazard function, i.e. the hazard rate when all the covariates are set to zero. This can be done by defining time-varying covariates that bears functional relationship with survival times. Popular specifications are log(survival time), polynomial in survival time, fully non-parametric and piece-wise constant (Jenkins 2005). Duration-interval-specific dummy variables need to be created for specifying a fully non-parametric baseline hazard. The number of dummy variables needs to be equal to the maximum survival time in the dataset. For instance, if the maximum survival time is fifty years, then fifty dummy variables are required for model estimation[7] (e.g. 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 reasonably convenient alternative way of specifying the baseline hazard function is to use piece-wise constant method. In this, the survival times are split into different time intervals that are assumed to exhibit constant hazard rate (Jenkins 2005). However, one must note that if there exist time intervals or time dummies with no events then one must drop the relevant firm-time observation with no event from the estimation or else duration specific hazard rate cannot be estimated for these time intervals/dummies (Jenkins 2005; Rabe-Hesketh and Skrondal 2012). Considering the estimation convenience one might be tempted to use the piece-wise constant specification of baseline hazard rate. However, if the hazard curve shows frequent and continuous steep rises and falls, then fully non-parametric baseline hazard is an appropriate choice.

3.4 Performance Evaluation

To gauge the classification performance of the models developed in identifying distressed firms, we estimate area under the Receiver Operating Characteristic (ROC) curve (AUROC). This curve originates from the signal detection theory, which shows how the receiver detects existence of signal in presence of noise. It is obtained by plotting the probability of detecting true-positive (sensitivity) (a firm actually defaults and the model classifies it as *expected*

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

default) and false-negative (1–specificity) (a firm actually defaults but the model classifies it as *expected non-default*) for an entire range of possible *cutpoints* (these are probability values). *Cutpoints*, *c*, is defined to obtain derived binary variable by comparing each estimated probability with *c*. If the estimated probability is grater than *c* then we let the value of the derived binary variable be equal to 1 or 0 otherwise. AUROC is now considered to be the standard non-parametric method for evaluating a fitted prediction model's ability to assign, in general, higher probabilities of the *event of interest* to the subgroup which develops the event of interest (dependent variable = 1) than it does to the subgroup which do not develop the event of interest (dependent variable = 0). The AUROC provides a measure of the prediction model's ability to discriminate between those firms who experience the event of interest versus those who do not. 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

We begin this section with analysis of key measures of descriptive statistics of our covariates along with relevant discussion pertaining to correlation among them. We perform univariate regression analysis of each covariate in turn using respective event definition, and respective econometric specification, to understanding any unexpected behaviour in their discriminatory performance. Then we discuss development and performance evaluation of multivariate discrete-time hazard models developed using logit and clog-log links. Finally, we develop multivariate extended Cox models and provide a comparative discussion on the performance of multivariate models developed using different default definitions. Additionally, to gauge any temporal variation in the explanatory power all our covariates, we perform our regression analysis using covariate that are lagged by *T–1*, *T–2* and *T–3* time periods.

4.1 DESCRIPTIVE STATISTICS AND CORRELATION

Inspection of descriptive statistics gives us an initial understanding about the variability of covariates and the potential biasness that may arise in the multivariate setup due to any unexpected extreme variability. We expect the mean of covariates that exhibit positive relationship with the insolvency/distress hazard to be higher for distressed group (status indicator = 1) than their healthy or censored counterparts (e.g. STDEBV in Table 4). On the contrary, the mean of covariates that shows negative relationship with the insolvency hazard is expected to be lower for distressed group than their healthy counterparts (e.g. CTA in Table 4). A closer look at Table 4 reveals that the mean, median and standard deviation of most of the covariates under respective event definitions are as per our expectation without any extreme variability. However, EBITDAIE and STDEBV raise some serious concerns. The mean of EBITDAIE is very high, as most of the firms in our sample do not incur or acquire very little interest expenses. This leads to very high difference between its mean and median values, resulting in a highly skewed distribution and very high value of standard deviation. We expect this covariate to be highly problematic in the multivariate setup. Although STDEB and TLNW are positively related to firms' default probability but their mean of the default group is significantly lower than their censored group under *Event 3*, which is quite surprising. This may lead to opposite sign of their coefficient in regression analysis. Additionally we observe that, the mean of respective covariates across different default definitions in Table 4 reveal very little variation in their values. This might signal little variation in the classification performance of multivariate models developed.

[Insert Table 4 Here]

The correlation matrix in Table 5 shows that some of the covariates exhibit moderate to strong correlation with other covariates. RETA shows strong positive correlation of approximately 0.65 with EBITDATA, supporting the belief that SMEs primarily rely on internal sources for their funding requirements, thus end up retaining significant portion of their income. This correlation among the covariates might cause some unexpected variation during multivariate modelling.

[Insert Table 5 Here]

4.2 UNIVARIATE REGRESSION AND AVERAGE MARGINAL EFFECTS

It is always advisable to do some univariate analysis before proceeding to estimation of multivariate models. In survival analysis the standard approach is to initially look at Kaplan-Meier survival curves of all categorical covariates to get an insight about their shape of survival functions and proportionality of each group[8]. Popular non-parametric tests of equality of survival functions like log-rank test and the Wilcoxon–Breslow–Gehan test (see Cleves *et al.* 2010) are also widely reported. However, it is not feasible to calculate Kaplan-

[8] See Cleves *et al.* (2010) for a detailed description of Kaplan-Meier curves.

Meier curves or conduct these non-parametric tests for continuous predictors as continuous predictors have too many different levels[9]. However, Nam *et al.* (2008) report log-rank test and the Wilcoxon–Breslow–Gehan test for their continuous predictor, which, to the best of our knowledge is inappropriate and misleading. Considering this constraint, we perform univariate regression of each covariate in turn to have an initial insight about their effects on respective default events.

However, in order to narrow down our list of covariates at first we obtain univariate regression estimates using *Event 2* as dependent variable and Equation (6) as regression methodology (discrete-time hazard model with logit link). Here we use financial distress based definition rather than legal bankruptcy with the presumption that it is the primary reason behind bankruptcy and always precedes the bankruptcy filing event. Further, filing for legal bankruptcy is the least efficient exit strategy for SMEs (Balcaen *et al.* 2012). Additionally, to gauge the temporal variation in the explanatory power of covariates we obtain regression estimates for $T – 1$, $T – 2$ and $T – 3$ lagged time periods (see Table 6). At this stage we exclude covariates from further empirical analysis that (i) are not significant in all three time periods (this ensures that the selected covariates are consistent predictors of firms' financial health over a sufficiently long time interval to allow for developing a reasonable *early warning system*), or, (ii) are significant but exhibit *Average Marginal Effects*[10] (AME) of less than 5% in all three time periods. The rational being, a unit change in the value of significant covariates must induce sufficient change in the magnitude of the outcome probability to clearly distinguish between distressed and healthy firms. An interesting thing to observe in Table 6 is the AME of Altman and Sabato's (2007) covariates that are widely employed in modelling default risk of SMEs. Out of the five covariates that they suggest, three of them (STDEBV, EBITDAIE and RETA) exhibit AME of less than 5% with AME of EBITDAIE being almost zero. The other two covariates CTA and EBITDATA have AME values less than 17%. This suggests that, although these covariates are significant predictors but a unit change in their value does not transmit significant change in the

---

[9] See for example http://www.ats.ucla.edu/STAT/stata/seminars/stata_survival/default.htm. Also see Cleves *et al.* (2010) for a more thorough understanding.

[10] In non-linear regression analysis, Marginal Effects is an useful way to examine the effect of changes in a given covariate on changes in the outcome variable, holding other covariates constant. These can be computed as marginal change (it is the partial derivative for continuous predictors) when a covariate changes by an infinitely small quantity and discrete change (for factor variables) when a covariate changes by a fixed quantity. Whereas, Average Marginal Effects (AME) of a given covariate is the average of its marginal effects computed for each observation at its observed values. Alternatively, AME can be interpreted as the change in the outcome (financial distress = 1; in our case) probabilities due to unit change in the given covariate, provided other covariates are held constant. See Long and Freese (2014) for detailed discussion on this topic.

probability of outcome variable. Although three of Altman and Sabato's (2007) covariates has AME less than 5%, but we include them for further empirical analysis to have further understanding about their explanatory power in the multivariate setup. From 27 variables, this helps us to narrow down to 16 variables that we use for further empirical analysis. Table 7 reports the final list of covariates that we use for further univariate and multivariate regression analysis.

[Insert Table 6 Here]

[Insert Table 7 Here]

Univariate Regression of Event 1: Section A of Table 8 reports the univariate regression estimates for Event 1 using discrete and continuous-time hazard models. Magnitude of coefficients of respective covariates (β in Table 8) obtained using discrete-time hazard specification with logit and cloglog links, and extended cox model are close to each other with some variation for covariates TLTA, FETA, CAG, SAG, TTA and RETA. For most of the covariates for respective time lags, absolute value of the magnitude of their respective coefficients is highest for logit estimates followed by cloglog estimates and least for cox estimates ($|logit| > |cloglog| > |cox|$). However, logit and cloglog estimates exhibit almost identical model fit as their AIC values are almost identical, but it's about three times higher for Cox estimates. This suggests that discrete-time model with logit/cloglog links offer better model fit than extended Cox model. We also see that CAG and SAG are significant in all time periods when estimated using discrete hazard model, but becomes insignificant (in T-2 and T-3) when estimated using Cox model. Further, the statistical significance of TLTA, NIS and RETA also varies with their econometric specification. Altman and Sabato's (2007) covariate EBITDATA loses its statistical significance beyond T-1; STDEBV shows unstable explanatory power (sign is opposite to expectation for T-1 logit and cloglog estimates), its insignificant in T-1 but significant in T-2 and T-3; EBITDAIE is significant but its coefficients are almost 0; RETA is significant in T-1 thought but is insignificant in T-2 and T-3 when estimated using discrete-hazard specification. Only CTA shows consistent and reliable explanatory power in all three time periods across all econometric specifications. *Event 1* is the same event definition that they use in their default prediction study; however our results do not approve the covariates suggested by them. Their suggestion might be biased due to 'cherry picking' involved in their sample selection process, while we use near-population data to establish our empirical validation.

Univariate Regression of Event 2: Section B of Table 8 reports univariate regression estimates obtained using Event 2 as dependent variable. All covariates are significant across all econometric specifications for all lagged time periods. We also see that coefficients of covariates obtained using logit, cloglog and cox hazard specification reasonably vary from each other, however most covariates follows the pattern |*logit*| > |*cloglog*| > |*cox*| for respective lagged time periods. However, the AIC values of logit and cloglog estimates are about three to six times lower than values obtained using cox specification. This asserts that discrete hazard models offers better model fit than its continuous counterpart. Additionally, in T-3 FETA and WCTA fail to remain significant when estimated using cox specification. In this case, all of Altman and Sabato's (2007) covariates are significant with expected sign across all econometric specification for respective lagged time periods except RETA, for cox estimate at T-3. All of their covariates also has reasonable magnitude of respective coefficients except EBITDAIE, which is again almost 0. However, the real litmus test of their covariates will be performed in the multivariate section.

Univariate Regression of Event 3: Section C of Table 8 reports univariate regression estimates obtained using *Event 3* as dependent variable. Unlike *Event 2* estimates, many of the covariates (OPCE, NIS, CAG, STDEBV, EBITDAIE and RETA) show varying (insignificant) explanatory power across different time periods. However, here also we see the AIC values of discrete-time estimates are much lower than cox estimates and the pattern |*logit*| > |*cloglog*| > |*cox*| also holds good for most of the covariates. Two of Altman and Sabato's (2007) covariates (STDEBV and EBITDAIE) also fail miserably in discriminating distressed and censored firms across T-1 and T-2 lagged time periods.

[Insert Table 8 Here]

4.3 Developing Multivariate Hazard Models

In this section, we develop and discuss multivariate hazard models for respective default definitions discussed in section 2. We begin with our choice for specification of the baseline hazard rate, which is required for developing discrete-time duration-dependent hazard models, followed by development and discussion of multivariate discrete-time and continuous-time hazard models.

4.3.1 DETECTION OF BASELINE HAZARD RATE

Before developing multivariate discrete-time hazard models it is important to choose a baseline specification for the hazard rate. Figure 1 shows the table of hazard curves[11] estimated using Kaplan-Meier[12] estimator for different distress events. The hazard curves of all three events exhibit fairly different functional relationship with firms' age. The hazard curves of all three events show increasing and decreasing relationship with firms' age, and the shape of hazard curves of *Event 1* and *Event 3* are quite similar. From the surface it might look that the distress events are highly duration-dependent. However, one might turn sceptical after having a looks at the magnitude of hazard rates on the y-axis. For *Event 1* it ranges approximately between 0.006 and 0.013; *Event 2* between 0.05 and 0.13; and, *Event 3* between 0.00175 and 0.00325. Considering these tight intervals of hazard rates, piece-wise specification of baseline hazard might fail to reflect the differences in the hazard rates between respective age groups. Additionally, all three hazard curves show steep rises and falls with some flatness in couple of time intervals, thus it's inappropriate to assume the hazard rates to be constant for any defined age group. In this situation it may be appropriate to go for fully non-parametric baseline hazard specification and use age specific dummy variables to specify the baseline hazard rate. To statistically test our intuition, we estimated multivariate discrete hazard models (with logit link) with *Event 1*, *Event 2* and *Event 3* respectively as dependent variable and only age dummies as independent variables. Regression results[13] confirm that about 90% of age dummies are significant (p-value < 0.05) in explaining respective outcome of interest. However, when supplemented with financial covariates, only about 10% of the age dummies remain statistically significant with large value of standard errors of their coefficients. This suggests that in presence of financial covariates, hazard rates do not show duration dependence. Additionally, one also needs to consider that too many variables may make the multivariate model numerically unstable. Thus, following Shumway (2001) we re-estimate these models using ln(firms' annual age)

---

[11] Table 1 show that the earliest age that a firm experiences distress event under all three default definitions is one years. However, the hazard curves start from somewhere around five years. This difference is due to the fact that "sts graph" command in Stata performs an adjustment of the smoothed hazard near the boundaries. In case of the default kernel function of -sts graph- (Epanechnikov kernel), the plotting range of the smoothed hazard function is restricted to be within one bandwidth of each endpoint. The same is true for other kernels, except the epan2, biweight, and rectangular kernels, in which case the adjustment is performed using boundary kernels. If we wish to plot an estimate of the hazard for the entire range, we could use a kernel without a boundary correction. Alternative, we can use then -noboundary- option, but this will produce an estimate that is biased near the edges. See "help sts graph" in Stata and Silverman (1986) for further details. This will not affect the empirical analysis if one uses fully non-parametric method of baseline hazard specification. However, one needs to be little careful while using piecewise-constant specification.

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

[13] These results are not reports in this paper; however, it may be made available from the authors.

(variable AGE in Table 9) as baseline hazard specification. In contrast to Shumway's (2001) results, AGE is significant in most of our multivariate hazard models. In light of this discussion we use natural logarithm of firms' age (AGE) to proxy the baseline hazard rate for all our multivariate models developed.

[Insert Figure 1 Here]

4.3.2 Model-Building Strategy

The criteria for including a covariate in the multivariate model often vary across scientific disciplines, but they all strive to develop the 'best' model that is numerically stable and can be easily adapted for real life applications. The standard error of a model increases with the increase in the number of covariates, and this also makes the model more dependent on the observed data. Thus the objective should be to employ a minimum number of covariates for a desired accuracy level. A good start is to perform univariate regressions of each covariate in turn and consider the covariates with p-values of less than 0.25 for developing multivariate models (see chapter 4 of Hosmer Jr *et al.* 2013). Another school of thought suggests inclusion of all theoretically motivated covariates in the multivariate model irrespective of their significance level in the univariate analysis. Some studies exclude insignificant predictors (p-value > 0.05) from their multivariate models, yet insignificant predictors may explain some of the variation of the dependent variable. Multicollinearity can be a serious issue that may make the model unstable if not addressed effectively. Thus, at first we rank (covariate having highest value of |AME| gets rank 1 and so on) the covariates in Table 7 based on the magnitude of their AME and then introduce each covariate in turn into the multivariate setup starting with the covariate having the highest rank (rank = 1 in Table 7). The rationale being, the higher the value of AME, the higher the change in the predicted probability due to unit changes in the covariate's value. Thus covariates with higher values of AME (e.g. FETA in Table 7) are more efficient in discriminating between distressed and censored firms than covariates with lower values of AME (e.g. TLTA in Table 7). Further, we exclude a covariate from the multivariate model if, when introduced, (i) it affects the sign[14] of any previously added covariate, (ii) it bears the opposite sign than expected, (iii) it bears the expected sign but has a p-value greater than 0.25 and (iv) it makes a previously added covariate insignificant with a p-value greater than 0.25. These primarily arise due to multicollinearity among covariates, thus our screening mechanism seems to be a reasonable

[14] Coefficients with negative sign become positive and vice versa.

choice. Moreover, we believe that this method of covariate introduction while developing multivariate models reasonably addresses the multicollineariy problem and leaves us with a 'best' set of covariates that explain the variance of the dependent variable. Using discrete hazard model with logit link, this process is applied to *Event 1*, *Event 2* and *Event 3* respectively for all three (T-1, T-2 and T-3) respective lagged time periods. Then, multivariate hazard models with cloglog link and extended Cox are estimated using the same set of covariates selected using logit link to see any differences that may arise due to different estimation methods.

Final set of multivariate hazard models reported in Table 9 are estimated using all observation available to us covering the entire sampling period, thus we do not have separate test and hold-out samples. In order to assess within-sample classification performance of the models developed we estimate area under ROC curve for respective models using the full estimation sample. For out-of-sample validation, we first estimate multivariate hazard model using observation till the year 2011 and using these estimates we predict the default probabilities for the year 2012; then we include 2012 in the estimation sample and predict default probabilities for 2013 and so on, till the year 2015. Then we use these predicted probabilities from the year 2012 through 2015 to estimate out-of-sample AUROC for respective multivariate hazard models.

[Insert Table 9 Here]

4.3.3 Hazard Models for Event 1

The binary dependent variable used is *Event 1*, i.e. firms that filed for legal bankruptcy proceedings are considered to have experienced the default event and censored otherwise (please see section 2 for detailed discussion). Section A of Table 9 reports multivariate hazard models estimated for T-1, T-2 and T-3 lagged time periods developed with respective econometric specification. As we see, the logit estimates of factors affecting the outcome probability of *Event 1* vary considerably across time periods, except FETA. However, the control variables Micro, Small, RISK1 and AGE are strongly significant across all time periods. Among five Altman and Sabato's (2007) covariates, EBITDATA and RETA fail to find a place in our multivariate models for all three time periods. Additionally, STDEBV, CTA and EBITDAIE do not show consistency in their explanatory power. As seen in univariate regression, here as well coefficients of EBITDAIE are almost 0. This clearly shows the inefficiency of Altman and Sabato's (2007) covariate in predicting corporate bankruptcies for SMEs. The statistical significance of most of the covariates do not vary considerably when estimated using cloglog and cox specification except STDEBV, CAG and SAG, however most of the covariates follows the pattern |*logit*| > |*cloglog*| > |*cox*|. This suggests cloglog and cox estimates undermines the effect of covariates on the outcome probability. However, AIC values of logit and cloglog estimate are almost identical and are about half of cox estimates. This clearly suggests that discrete-time hazard models offer much superior model fit than continuous extended cox model. However, the within-sample AUROC for all econometric specifications are almost identical with slight variation among estimates of hold-out sample (see Figure 2). This suggests no significant loss in the classification performance if one uses cox specification over discrete-time. Additionally, the AUROC of all our multivariate models developed are around 0.8 or higher, which is considered to be excellent. But, the shape of ROC curves of hold-out sample estimates are steps rather than concave due to very low number of events in out-of-sample estimations.

[Insert Figure 2 Here]

4.3.4 Hazard Models for Event 2

Unlike *Event 1*, multivariate models developed for *Event 2* using logit specification show consistent explanatory power of most covariates over all three lagged time periods (see Section B of Table 9). Here as well the pattern |*logit*| > |*cloglog*| > |*cox*| holds good for most of the explanatory variables. However, the statistical significance of SAG (T-1) and AGE (T-1 and T-3) varies with the estimation technique. All control variables (Small, Medium and RISK2) are also highly significant across all lagged time periods. Among Altman and Sabato's (2007) covariates, STDEBV, EBITDATA and EBITDAIE exhibit significant explanatory power across all lagged time periods and econometric specifications. However, the coefficient of EBITDAIE is almost 0 here as well. The variable CTA finds place only in the models developed for T-2 time periods, while RETA fails to meet our screening criteria for inclusion into the multivariate model. Thus, Altman and Sabato's (2007) covariates are not efficient predictors of financial distress unlike some of the other financial ratios reported in Section B of Table 9. Here as well AIC values of discrete hazard models are about three to four times lower than cox models, thus discrete-time hazard models offer superior model fit than their continuous counterpart. The within sample and hold-out sample AUROC estimated for different multivariate models are about or higher than 0.80 suggesting excellent classification performance of our multivariate models across all time periods and econometric specifications.

4.3.5 Hazard Models for Event 3

The set of hazard models that we estimate is based on the default definition (*Event 3*) that we propose in this study, which considers both legal bankruptcy filing and firms' financial health while classifying SMEs as default (please see section 2 for details). Section C of Table 9 reports multivariate regression estimates for *Event 3* across all three lagged time periods and econometric specification. A casual look the results reveal that the factors affecting the outcome probability varies reasonably across the time periods. Even the statistical significance of six covariates (STDEBV, OPCE, RETA, CAG, SAG and TTA) is sensitive to estimation technique. Among Altman and Sabato's (2007) covariates STDEBV finds palce in T-2 and T-3, while RETA finds place in T-1 only. EBITDATA, CTA and EBITDAIE fail meet our inclusion criteria into the multivariate setup. This reinforces the inefficiency of covariates suggested by Altman and Sabato's (2007) in predicting SMEs financial distress. Here as well the AIC values are in favour of discrete-time models, which are about 0.8 times lower than continuous cox estimates. Both within sample and hold-out sample classification of all the multivariate models across all time periods and econometric specifications are close to or above 0.9, which is superior to Event 1 and Event 2 models' classification performance (see Figure 2).

4.3.6 Comparative Performance of Hazard Models

As reported in Table 9, the extended Cox model performs almost identical to discrete-time models with logit and clog-log links as its shows almost identical classification performance across all default definitions. Thus one might be indifferent in her choice of hazard specification. However, the coefficients of most covariates in all multivariate models show the pattern $|logit| > |cloglog| > |cox|$. This means that the absolute value of a covariate's coefficient is highest when estimated using discrete-hazard logit specification and lowest when estimated continuous cox specification. Thus, for a unit change in the value of a covariate, logit estimates leads to highest change in the outcome probability than its alternative counterparts. Hence, we suggest the use of discrete-time hazard model with logit link to model interval censored data. But, if the event of interest is not duration dependent (i.e. some functional form of time or time dummies are not significant in the multivariate model) with the hazard rates being invariant or varies mildly across different time periods, then getting involved into the complications of hazard models is not rewarding considering the marginal gain one would obtain using such models. As reported earlier, in presence of other financial covariates about 90% of time dummies that we use as baseline hazard specification are insignificant with very high values of standard errors. Thus we follow Shumway (2001) and use natural logarithm of firms annual age (AGE) as baseline hazard specification, but such objective can easily be achieved by developing regression models using logistic regression techniques that use some functional form of time to capture any duration dependency. Although Shumway (2001) argue that hazard models are superior to competing static models but ln(age) variable in his multivariate models are insignificant, then how can it be used to reliably predict duration specific hazard rate, which is why hazard models are primarily used. Unlike other scientific disciplines areas like medicine or health economics, duration specific prediction of hazard rates is not common in bankruptcy/financial distress prediction studies, thus we do not see any real need of hazard models if similar objective can be achieved using much simpler logistic regression that controls for any duration dependencies, as both involve identical statistical estimation methods. However, another interesting observation is the classification performance measures across different default definitions. Based on the AUROC measures, Event 1 is the weakest definition of default while Event 3 is the strongest as it has highest values of AUROC across all time periods. Also, the AIC measure of Event 3 models are the lowest among the three default definitions, which indicates that Event 3 default definition provides a vastly improved fit than the other two default definitions.

5. CONCLUSION

The growing popularity of hazard models in making bankruptcy prediction motivated us to undertake this empirical investigation. Almost every study in the empirical literature suffers from at least one of the following issues: (i) reason behind their choice between *discrete-time* or *continuous-time* hazard model (ii) inappropriate specification of baseline hazard rate (iii) no test of proportional hazards assumption when using *Extended Cox* model with time-independent covariates (iv) ignore *frailty* and *recurrent* events (v) explanation on how they dealt with the issues of *delayed entry* (vi) explanation on treatment of time periods/intervals having no events. Therefore, we contribute to the literature by acknowledging all these serious and neglected concerns in our study and intend to be the first academic paper to report performance comparison of popular hazard models (discrete hazard models with logit and clog-log links and extended Cox model) used in the recent literature (e.g. Campbell *et al.*

2008, Chen and Hill 2013). We also contribute to the literature by undertaking an empirical investigation which compares various default definitions of US SMEs. Three default definitions that we compare are based on legal bankruptcy laws (Event 1), firms' financial health (Event 2) and the third definition (Event 3) that we propose in this study considers both legal bankruptcy and firms' financial health. Finally, we also contribute by testing the efficiency of covariates suggested by Altman and Sabato's (2007) in predicting corporate default across varying default definitions and lagged time periods.

Our empirical results highlight almost identical classification performance of both discrete hazard models and continuous cox model across all three default definitions, suggesting invariance of classification performance to econometric specification. Based on comparison of AIC measures, discrete hazard models provide considerably superior fit than Cox model. However, the AIC measures for both discrete-time hazard models (logit and clog-log links) are almost identical; hence the choice between them is left on the personal preference of the users. But if one considers magnitude of coefficient of covariates, event outcome probabilities are more sensitive to logit estimates than cloglog and cox estimates. Also, Altman and Sabato's (2007) covariates are unstable and inefficient in predicting event outcome across different default events and lagged time periods in comparison to other competing financial ratios. Further, comparison of default definitions leads us to a striking conclusion. Based on the classification performance and AIC values of the models developed using different default definitions, we understand that the default definition that we propose performs best in identifying distressed firms. This emphasises the fact that, significant number of firms choose legal bankruptcy routes as part of their planned exit strategy.

Given the importance of hazard models in predicting bankruptcy and the robustness of our results in dealing with neglected econometric issues in all previous empirical research in survival analysis, this paper makes a significant contribution to SMEs and corporate failure literature.

References

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

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

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

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

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

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

Balcaen, S., Manigart, S., Buyze, J., and Ooghe, H., 2012. Firm Exit After Distress: Differentiating Between Bankruptcy, Voluntary Liquidation and M&A. *Small Business Economics*, 39 (4), 949–975.

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

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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

Joseph, C., 2013. *Advanced Credit Risk Analysis and Management*. 1st ed. West Sussex: John Wiley & Sons.

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

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

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

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

Long, J.S. and Freese, J., 2014. *Regression models for categorical dependent variables using Stata*. 3rd ed. Texas: Stata press.

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

Table and Figures

Table 1: Survival Table

Age	Event 1			Event 2			Event 3		
	1	0	% 1	1	0	% 1	1	0	% 1
1	8	3,931	0.20	316	3,623	8.02	2	3,937	0.05
2	13	6,373	0.20	0	6,386	0.00	1	6,385	0.02
3	28	6,749	0.41	607	6,170	8.96	1	6,776	0.01
4	41	6,412	0.64	894	5,559	13.85	6	6,447	0.09
5	52	6,242	0.83	831	5,463	13.20	12	6,282	0.19
6	47	5,791	0.81	910	4,928	15.59	12	5,826	0.21
7	56	4,941	1.12	812	4,185	16.25	16	4,981	0.32
8	48	4,277	1.11	716	3,609	16.55	20	4,305	0.46
9	54	3,776	1.41	697	3,133	18.20	17	3,813	0.45
10	50	3,283	1.50	618	2,715	18.54	14	3,319	0.42
11	35	2,564	1.35	476	2,123	18.31	11	2,588	0.43
12	25	2,233	1.11	371	1,887	16.43	11	2,247	0.49
13	24	2,039	1.16	358	1,705	17.35	6	2,057	0.29
14	19	1,817	1.03	323	1,513	17.59	5	1,831	0.27
15	15	1,625	0.91	273	1,367	16.65	6	1,634	0.37
16	10	1,460	0.68	248	1,222	16.87	3	1,467	0.20
17	11	1,285	0.85	224	1,072	17.28	2	1,294	0.15
18	7	1,128	0.62	195	940	17.18	3	1,132	0.27
19	8	1,017	0.78	199	826	19.41	3	1,022	0.29
20	12	900	1.32	157	755	17.21	4	908	0.44
21	10	786	1.26	132	664	16.58	1	795	0.13
22	6	715	0.83	123	598	17.06	0	721	0.00
23	6	642	0.93	111	537	17.13	1	647	0.15
24	6	573	1.04	107	472	18.48	0	579	0.00
25	9	483	1.83	95	397	19.31	3	489	0.61
26	10	445	2.20	93	362	20.44	3	452	0.66
27	6	411	1.44	74	343	17.75	5	412	1.21
28	4	379	1.04	62	321	16.19	0	383	0.00
29	4	329	1.20	62	271	18.62	1	332	0.30
30	5	271	1.81	50	226	18.12	2	274	0.73
31	5	235	2.08	41	199	17.08	1	239	0.42
32	5	201	2.43	38	168	18.45	1	205	0.49
33	4	178	2.20	23	159	12.64	1	181	0.55
34	4	163	2.40	20	147	11.98	1	166	0.60
35	4	145	2.68	15	134	10.07	0	149	0.00
36	3	127	2.31	16	114	12.31	0	130	0.00
37	2	115	1.71	16	101	13.68	1	116	0.86
38	1	111	0.89	11	101	9.82	0	112	0.00
39	0	102	0.00	13	89	12.75	0	102	0.00
40	0	91	0.00	9	82	9.89	0	91	0.00
41	1	69	1.43	6	64	8.57	0	70	0.00
42	0	45	0.00	0	45	0.00	0	45	0.00
43	0	46	0.00	3	43	6.52	0	46	0.00
44	0	41	0.00	3	38	7.32	0	41	0.00
45	0	36	0.00	2	34	5.56	0	36	0.00
46	0	30	0.00	3	27	10.00	0	30	0.00
47	0	27	0.00	2	25	7.41	0	27	0.00
48	0	23	0.00	0	23	0.00	0	23	0.00
49	0	23	0.00	1	22	4.35	0	23	0.00
50	0	20	0.00	1	19	5.00	0	20	0.00
51	0	19	0.00	1	18	5.26	0	19	0.00
52	0	14	0.00	2	12	14.29	0	14	0.00
53	0	11	0.00	0	11	0.00	0	11	0.00
54	0	8	0.00	0	8	0.00	0	8	0.00
55	0	6	0.00	1	5	16.67	0	6	0.00
56	0	0	0.00	0	0	0.00	0	0	0.00
57	0	0	0.00	0	0	0.00	0	0	0.00
58	0	1	0.00	0	1	0.00	0	1	0.00
Total	658	74,764	0.87	10,361	65,061	13.74	176	75,246	0.23

Notes: This table shows the age wise distribution of firm-year observations for respective default events discussed in section 2. Numeric '0' signifies censorship and '1' signifies that a firm has experienced the respective default event.

Table 2: Sample Industrial Classification

Industry Code	SIC Code	Industry	Included/Excluded
1	< 1000	Agriculture, Forestry, Fishing	Included
2	1000 to < 1500	Mining	Included
3	1500 to < 1800	Construction	Included
4	2000 to < 4000	Manufacturing	Included
5	5000 to < 5200	Wholesale Trade	Included
6	5200 to < 6000	Retail Trade	Included
7	7000 to < 8900	Services	Included
Excluded	4000 to < 5000	Transportation, Communications & Public Utilities	Excluded
Excluded	6000 to < 6800	Finance, Insurance & Real Estate	Excluded
Excluded	9100 to < 10000	Public Administration	Excluded

Notes: This table reports Standard Industrial Classification (SIC) of US firms. SIC Code is a four digit code that represents a given industrial sectors. The last column reports the industrial sectors that we included or excluded form our sample.

Table 3: List of Covariates

Category	Variable	Definition	Compustat Data Item
Leverage	STDEBV	Short term debt/equity book value	DLC/SEQ
	TLTA	Total liabilities/tangible total assets	LT/(AT – INTAN)
	TLNW	Total liabilities/net worth	LT/(AT - LT)
	CETL	Capital employed/total liabilities	(AT – LCT)/LT
Liquidity	CTA	Cash and short-term investments/total assets	CHE/AT
	CR	Current Ratio; current assets/current liabilities	ACT/LCT
	QR	Quick Ratio; (current assets – stocks - prepayments)/current liabilities	(ACT – INVT – XPP)/LCT
	CHR	Cash Ratio; (cash + bank + marketable securities)/current liabilities	CHE/LCT
Financing	FETA	Financial expenses/total assets	XINT/AT
	FES	Financial expenses/sales	XINT/SALE
	RETA	Retained earnings/total assets	RE/AT
	EBITDAIE	Earnings before interest taxes depreciation and amortization/interest expense	EBITDA/XINT
Profitability	EBITDATA	Earnings before interest taxes depreciation and amortization/total assets	EBITDA/AT
	OPCE	Operating profit/capital employed	EBIT/(AT - LCT)
	ROE	Return on equity; Net profit/equity	NI/SEQ
	NIS	Net income/sales	NI/SALE
	OPNI	Operating profit/net income	EBIT/NI
Activity	SHP	Stock holding period; (stock × 365)/sales	(INVT × 365)/SALE
	DCP	Debtor collection period; (trade debtors × 365)/sales	(RECTR × 365)/SALE
	TCP	Trade creditors payment period; (trade creditors × 365)/sales	(AP × 365)/SALE
	WCTA	Working capital/total assets	WCAP/AT
	WCS	Working capital/sales	WCAP/SALE
	STA	Sales/tangible assets	SALE/(AT – INTAN)
Growth	CAG	Capital growth; calculated as ($Capital_t$ / $Capital_{t-1}$) - 1	(AT - LCT)
	SAG	Sales growth; calculated as ($Sale_t$ / $Sale_{t-1}$) - 1	SALE
	ERG	Earnings growth; calculated as ($EBIT_t$ / $EBIT_{t-1}$) - 1	EBIT
Other	TTA	Income taxes/total assets	TXT/AT
Control	Micro	No. of employees < 10	
	Small	10 =< No. of employees < 50	
	RISK	Event rate in a given industrial sector in a given year (calculated separately for different *Event* definition)	

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

Table 4: Descriptive Statistics

Variable	Status Indicator	Event 1			Event 2			Event 3		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
STDEBV	0	0.1889	0.0144	0.4235	0.1843	0.0176	0.3977	0.1893	0.0143	0.4245
	1	0.1967	0.0000	0.5262	0.2163	0.0000	0.5558	0.0442	-0.0571	0.4291
OPNI	0	1.1200	1.0058	1.0524	1.1707	1.0338	1.0748	1.1177	1.0050	1.0533
	1	0.6870	0.6729	1.0861	0.7976	0.8270	0.8516	0.5416	0.4644	0.9635
TLTA	0	0.6390	0.4978	0.5124	0.5481	0.4325	0.4430	0.6404	0.4993	0.5130
	1	0.9685	0.8343	0.5879	1.1870	1.0742	0.5578	1.2802	1.3848	0.5885
TLNW	0	1.0663	0.4885	2.4897	1.0672	0.4964	2.2490	1.0681	0.4890	2.4963
	1	1.0459	0.3762	3.3770	1.0595	0.1858	3.6364	0.2452	-1.1327	3.4305
CETL	0	2.5593	1.5343	2.5867	2.8603	1.8361	2.6255	2.5520	1.5274	2.5852
	1	1.1824	0.5146	1.8534	0.7110	0.3667	1.2149	0.6954	0.1984	1.2239
CTA	0	0.2346	0.1357	0.2351	0.2425	0.1475	0.2362	0.2343	0.1352	0.2352
	1	0.1856	0.0660	0.2335	0.1853	0.0747	0.2230	0.1681	0.0578	0.2272
CR	0	2.8370	1.9516	2.4008	3.0807	2.2228	2.4266	2.8311	1.9443	2.4002
	1	1.6869	0.9452	1.9695	1.3366	0.8283	1.5489	1.3542	0.6025	1.6870
QR	0	2.1085	1.2028	2.1818	2.3227	1.3980	2.2275	2.1027	1.1982	2.1804
	1	1.1411	0.4881	1.6776	0.8220	0.3966	1.2752	0.9004	0.3045	1.4141
CHR	0	1.4273	0.4502	1.9092	1.5658	0.5720	1.9647	1.4240	0.4478	1.9081
	1	0.8188	0.0984	1.5851	0.5769	0.0900	1.2250	0.6384	0.0572	1.4420
FETA	0	0.0294	0.0172	0.0324	0.0244	0.0136	0.0285	0.0295	0.0173	0.0324
	1	0.0507	0.0474	0.0381	0.0583	0.0591	0.0375	0.0602	0.0689	0.0402
FES	0	0.0611	0.0192	0.0953	0.0509	0.0157	0.0854	0.0613	0.0193	0.0954
	1	0.0982	0.0510	0.1087	0.1227	0.0624	0.1237	0.1235	0.0608	0.1222
EBITDAIE	0	99.092	-0.0423	537.480	120.948	1.2368	576.057	98.612	-0.0524	536.500
	1	22.388	-0.7447	333.325	-24.179	-3.723	166.108	20.568	-0.6260	312.090
EBITDATA	0	-0.2248	-0.0043	0.6142	-0.1498	0.0266	0.5294	-0.2250	-0.0045	0.6141
	1	-0.3060	-0.0438	0.6628	-0.6531	-0.2799	0.8446	-0.4394	-0.0765	0.7822
OPCE	0	-0.1001	-0.0052	0.4090	-0.0905	0.0025	0.3881	-0.1003	-0.0057	0.4091
	1	-0.1207	-0.0373	0.4296	-0.1573	-0.1130	0.5120	-0.0597	-0.0003	0.4276
ROE	0	-0.1204	0.0105	0.5801	-0.1293	0.0080	0.5361	-0.1202	0.0104	0.5809
	1	-0.0163	0.0411	0.6791	-0.0624	0.0754	0.7918	0.1536	0.1829	0.6011
NIS	0	-0.4445	-0.0374	0.7868	-0.3881	-0.0069	0.7603	-0.4451	-0.0382	0.7866
	1	-0.5511	-0.2287	0.7520	-0.7995	-0.4063	0.8510	-0.5537	-0.2568	0.7432
RETA	0	-1.5313	-0.4904	2.2310	-1.2142	-0.3045	2.0203	-1.5314	-0.4929	2.2298
	1	-2.0233	-1.0498	2.2862	-3.3941	-3.1440	2.4774	-3.2022	-2.7933	2.4772
SHP	0	50.132	35.532	51.986	49.777	35.764	51.378	50.134	35.536	51.986
	1	49.579	31.523	52.770	52.287	33.687	55.590	46.837	22.446	54.655
DCP	0	63.907	55.077	49.284	64.987	56.112	48.985	63.873	55.017	49.304
	1	59.123	46.274	53.136	57.293	47.770	50.741	60.949	49.949	55.460
TCP	0	168.668	35.971	532.046	146.338	33.444	485.965	168.499	35.982	531.456
	1	187.043	40.940	558.262	306.849	58.592	742.050	308.623	50.766	801.227
WCTA	0	0.2370	0.2774	0.3594	0.2857	0.3245	0.3327	0.2357	0.2760	0.3598
	1	-0.0152	-0.0288	0.3790	-0.0645	-0.0845	0.3707	-0.1111	-0.2199	0.3810
WCS	0	0.4157	0.2371	0.6494	0.4689	0.2735	0.6404	0.4140	0.2357	0.6496
	1	0.0812	-0.0269	0.5831	0.0657	-0.0388	0.5965	-0.0504	-0.1886	0.4929
STA	0	1.0735	0.8915	0.9189	1.0633	0.8915	0.8993	1.0737	0.8912	0.9192
	1	1.1171	0.8864	0.9830	1.1347	0.8914	1.0262	1.1383	0.9604	1.0089
CAG	0	0.1639	0.0292	0.6464	0.1924	0.0469	0.6229	0.1621	0.0283	0.6465
	1	-0.0889	-0.2126	0.6427	0.0053	-0.2042	0.7368	-0.0191	-0.1314	0.7247
SAG	0	0.1788	0.0815	0.4423	0.1906	0.0955	0.4327	0.1776	0.0804	0.4424
	1	-0.0170	-0.1309	0.4161	0.1033	-0.0326	0.4852	-0.0466	-0.1897	0.4138
ERG	0	-0.0394	-0.0681	1.3143	-0.0439	-0.0420	1.3478	-0.0407	-0.0695	1.3150
	1	-0.2671	-0.3645	1.3192	-0.0295	-0.1764	1.1325	-0.3475	-0.4978	1.1091
TTA	0	0.0130	0.0000	0.0282	0.0148	0.0000	0.0296	0.0130	0.0000	0.0282
	1	0.0061	0.0000	0.0222	0.0020	0.0000	0.0140	0.0037	0.0000	0.0167

Notes: This table reports the mean, median and standard deviation for healthy (censored) and unhealthy (firms which experienced default event) groups for respective covariates under different definitions of default events as discussed in section 2.

Table 5: Correlation Matrix

Variable		1	2	3	4	5	6	7	8
STDEBV	1	1							
TLTA	2	0.0780	1						
CETL	3	-0.2986	-0.6772	1					
CTA	4	-0.2670	-0.3276	0.4829	1				
FETA	5	0.1943	0.7093	-0.5344	-0.3343	1			
FES	6	0.0084	0.4339	-0.2462	-0.0563	0.6230	1		
EBITDAIE	7	-0.1004	-0.1983	0.2729	0.1330	-0.2071	-0.1508	1	
EBITDATA	8	0.1278	-0.4223	0.1548	-0.1694	-0.2899	-0.3963	0.2129	1
OPCE	9	-0.1097	0.0783	-0.0390	-0.2278	0.0482	-0.1084	0.2402	0.3127
NIS	10	0.1112	-0.2128	-0.0345	-0.2941	-0.172	-0.5105	0.2517	0.6920
RETA	11	0.1292	-0.5031	0.2299	-0.1590	-0.3409	-0.3353	0.1872	0.6588
WCTA	12	-0.1851	-0.7413	0.5870	0.5451	-0.5782	-0.4074	0.1994	0.3148
WCS	13	-0.2144	-0.5619	0.6392	0.7005	-0.4397	-0.1252	0.0593	0.0360
CAG	14	-0.0992	-0.1505	0.1613	0.1336	-0.1323	-0.0293	0.0540	0.1490
SAG	15	-0.0117	-0.0651	0.0448	0.0833	-0.0868	-0.0430	0.0225	0.0375
TTA	16	-0.0680	-0.1870	0.0894	-0.0085	-0.1813	-0.2153	0.2993	0.3199
		9	10	11	12	13	14	15	16
OPCE	9	1							
NIS	10	0.4705	1						
RETA	11	0.3057	0.5371	1					
WCTA	12	-0.0633	0.1543	0.3499	1				
WCS	13	-0.2574	-0.2318	0.1047	0.7403	1			
CAG	14	0.2161	0.0921	0.1658	0.1969	0.1783	1		
SAG	15	0.0424	0.0061	0.0497	0.0554	0.0676	0.2730	1	
TTA	16	0.4175	0.3232	0.2985	0.2158	-0.0259	0.1345	0.0971	1

Table 6: Event 2 Univariate Regression

Variable	*Sign*	*T – 1*			*T – 2*			*T – 3*		
		β	*SE*	*dy/dx* %	*β*	*SE*	*dy/dx* %	*β*	*SE*	*dy/dx* %
STDEBV	+	0.1700[a]	0.0298	1.05[a]	0.3928[a]	0.0312	2.84[a]	0.3225[a]	0.0329	2.72[a]
OPNI	-	-0.3627[a]	0.0147	-2.32[a]	-0.3323[a]	0.0150	-2.48[a]	-0.1608[a]	0.0146	-1.36[a]
TLTA	+	2.4762[a] [a]	0.0332	16.59 [a]	2.2925[a]	0.0340	17.77[a]	0.6454[a]	0.0317	5.78[a]
TLNW	+	0.0096[b]	0.0048	0.05[b]	0.0349[a]	0.0050	0.25[a]	0.0294[a]	0.0053	0.24[a]
CETL	-	-0.9832[a]	0.0175	-7.60[a]	-0.8332[a]	0.0155	-7.20[a]	-0.2073[a]	0.0076	-1.87[a]
CTA	-	-1.9079[a]	0.0753	-11.67[a]	-2.2780[a]	0.0882	-16.20[a]	-0.5436[a]	0.0770	-4.55[a]
CR	-	-0.5862[a]	0.0113	-4.09[a]	-0.5718[a]	0.0113	-4.60[a]	-0.1559[a]	0.0077	-1.35[a]
QR	-	-0.6774[a]	0.0179	-4.69[a]	-0.6328[a]	0.0171	-4.86[a]	-0.1790[a]	0.0113	-1.51[a]
CHR	-	-0.5532[a]	0.0132	-3.55[a]	-0.5760[a]	0.0138	-4.30[a]	-0.1436[a]	0.0098	-1.22[a]
FETA	+	32.1570[a]	0.4631	248.01[a]	24.0993[a]	0.4610	219.22[a]	8.8175[a]	0.4786	84.55[a]
FES	+	7.2175[a]	0.1558	50.52[a]	6.3882[a]	0.1665	51.69[a]	3.7784[a]	0.1733	33.55[a]
EBITDAIE	-	-0.0007[a]	0.0000	-0.00[a]	-0.0020[a]	0.0001	-0.02[a]	-0.0014[a]	0.0000	-0.01[a]
EBITDATA	-	-1.0990[a]	0.0233	-7.98[a]	-1.5416[a]	0.0289	-11.39[a]	-0.8150[a]	0.0276	-7.22[a]
OPCE	-	-0.1528[a]	0.0320	-0.94[a]	-1.0082[a]	0.0351	-7.50[a]	-1.1616[a]	0.0373	-9.96[a]
ROE	-	0.2635[a]	0.0210	1.64[a]	-0.2591[a]	0.0219	-1.90[a]	-0.5588[a]	0.0238	-4.75[a]
NIS	-	-0.5768[a]	0.0200	-3.66[a]	-1.0970[a]	0.0237	-7.89[a]	-0.7226[a]	0.0229	-5.97[a]
RETA	-	-0.4931[a]	0.0078	-3.68[a]	-0.4340[a]	0.0079	-3.76[a]	-0.1911[a]	0.0077	-1.81[a]
SHP	+	-0.0002	0.0000	-0.00	0.0034[a]	0.0003	0.02[a]	0.0048[a]	0.0003	0.04[a]
DCP	+	-0.0039[a]	0.0003	-0.02[a]	-0.0023[a]	0.0003	-0.02[a]	0.0005	0.0003	0.00
TCP	+	0.0003[a]	0.0000	0.00[a]	0.0003[a]	0.0000	0.00[a]	0.001[a]	0.0000	0.00[a]
WCTA	-	-3.2430[a]	0.0500	-22.62[a]	-3.1762[a]	0.0517	-25.47[a]	-0.8462[a]	0.0462	-7.44[a]
WCS	-	-1.5370[a]	0.0335	-9.60[a]	-1.4041[a]	0.0336	-10.04[a]	-0.1892[a]	0.0276	-1.50[a]
STA	+	0.2387[a]	0.0184	1.44[a]	0.0406[b]	0.0199	0.28[b]	-0.2551[a]	0.0215	-2.10[a]
CAG	-	-0.4236[a]	0.0210	-3.12[a]	-1.1145[a]	0.0271	-9.68[a]	-0.4648[a]	0.0240	-3.78[a]
SAG	-	-0.4975[a]	0.0321	-3.42[a]	-0.8480	0.0351	-6.76[a]	-0.3526[a]	0.0356	-2.70[a]
ERG	-	-0.0145	0.0104	-0.10	-0.0009	0.0108	-0.00	-0.0331[a]	0.0116	-0.26[a]
TTA	-	-24.5294[a]	0.8286	-166.40[a]	-46.0535[a]	1.2150	-370.58[a]	-28.3887[a]	0.9206	-255.68[a]

Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table reports univariate regression estimates of respective covariates at respective lagged time periods, estimated using discrete-time hazard model with logit link and Event 2 = 1 as outcome event. 'Sign' represents expected sign of regression coefficients, *β* is the regression coefficient, SE is standard error and *dy/dx* is Average Marginal Effects in percentage.

Table 7: Final Set of Covariates

Variable	*Sign*	*T – 1*			*T – 2*			*T – 3*		
		β	*Rank*	*dy/dx* %	*β*	*Rank*	*dy/dx* %	*β*	*Rank*	*dy/dx* %
STDEBV	+	0.1700[a]	14	1.05[a]	0.3928[a]	15	2.84[a]	0.3225[a]	11	2.72[a]
TLTA	+	2.4762[a]	5	16.59 [a]	2.2925[a]	5	17.77[a]	0.6454[a]	8	5.78[a]
CETL	-	-0.9832[a]	9	-7.60[a]	-0.8332[a]	12	-7.20[a]	-0.2073[a]	13	-1.87[a]
CTA	-	-1.9079[a]	6	-11.67[a]	-2.2780[a]	6	-16.20[a]	-0.5436[a]	9	-4.55[a]
FETA	+	32.1570[a]	1	248.01[a]	24.0993[a]	2	219.22[a]	8.8175[a]	2	84.55[a]
FES	+	7.2175[a]	3	50.52[a]	6.3882[a]	3	51.69[a]	3.7784[a]	3	33.55[a]
EBITDAIE	-	-0.0007[a]	16	-0.00[a]	-0.0020[a]	16	-0.02[a]	-0.0014[a]	16	-0.01[a]
EBITDATA	-	-1.0990[a]	8	-7.98[a]	-1.5416[a]	7	-11.39[a]	-0.8150[a]	6	-7.22[a]
OPCE	-	-0.1528[a]	15	-0.94[a]	-1.0082[a]	11	-7.50[a]	-1.1616[a]	4	-9.96[a]
NIS	-	-0.5768[a]	11	-3.66[a]	-1.0970[a]	10	-7.89[a]	-0.7226[a]	7	-5.97[a]
RETA	-	-0.4931[a]	10	-3.68[a]	-0.4340[a]	14	-3.76[a]	-0.1911[a]	14	-1.81[a]
WCTA	-	-3.2430[a]	4	-22.62[a]	-3.1762[a]	4	-25.47[a]	-0.8462[a]	5	-7.44[a]
WCS	-	-1.5370[a]	7	-9.60[a]	-1.4041[a]	8	-10.04[a]	-0.1892[a]	15	-1.50[a]
CAG	-	-0.4236[a]	13	-3.12[a]	-1.1145[a]	9	-9.68[a]	-0.4648[a]	10	-3.78[a]
SAG	-	-0.4975[a]	12	-3.42[a]	-0.8480	13	-6.76[a]	-0.3526[a]	12	-2.70[a]
TTA	-	-24.5294[a]	2	-166.40[a]	-46.0535[a]	1	-370.58[a]	-28.3887[a]	1	-255.68[a]

Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table reports the final set of covariates that we use for multivariate hazard analysis. This excludes covariates reported in Table 6 that are not significant in all three time periods or are significant but exhibit Average Marginal Effects (AME) of less than 5% in all three time periods. It also includes all covariates of Altman and Sabato's (2007) study irrespective of their significance or AME values. 'Sign' represents expected sign of regression coefficients, β is the regression coefficient, SE is standard error and *dy/dx* is Average Marginal Effects in percentage. *Rank* is based on the absolute values of AME, where highest value gets 1, second highest get 2 and so on.

Table 8: Univariate Regression

Section A: Event 1									
Variable	logit			clog-log			Cox		
	T – 1	*T – 2*	*T – 3*	*T – 1*	*T – 2*	*T – 3*	*T – 1*	*T – 2*	*T – 3*
TLTA									
β	1.6496[a]	0.9402[a]	0.5690[a]	1.4897[a]	0.8714[a]	0.5411[a]	1.2130[a]	0.5792[a]	0.2296[c]
SE	0.1164	0.1183	0.1230	0.1031	0.1060	0.1168	0.1111	0.1153	0.1312
AIC	4866.94	4711.79	4394.56	4865.63	4713.43	4398.85	16902.86	16511.26	15864.87
CETL									
β	-0.4373[a]	-0.3022[a]	-0.2104[a]	-0.4068[a]	-0.2801[a]	-0.1931[a]	-0.3974[a]	-0.2629[a]	-0.1569[a]
SE	0.0424	0.0371	0.0346	0.0393	0.0340	0.0316	0.0396	0.0324	0.0337
AIC	5125.48	4966.65	4650.35	5125.60	4971.34	4663.57	17225.83	17114.84	16301.24
CTA									
β	-1.5618[a]	-2.4716[a]	-2.0885[a]	-1.3537[a]	-2.0737[a]	-1.7839[a]	-1.6333[a]	-2.3980[a]	-1.9371[a]
SE	0.2933	0.3325	0.3420	0.2617	0.2903	0.3031	0.2818	0.3004	0.3151
AIC	5546.80	5240.32	4859.45	5552.50	5253.27	4870.86	18900.08	18084.53	16885.00
FETA									
β	21.9980[a]	16.2405[a]	12.2258[a]	19.737[a]	14.708[a]	11.454[a]	17.330[a]	11.251[a]	7.526[a]
SE	1.6891	1.7580	1.9064	1.5101	1.5475	1.6837	1.572	1.598	1.753
AIC	5028.53	4975.84	4634.43	5033.99	4976.34	4636.87	16649.44	16779.55	16034.27
FES									
β	4.6120[a]	4.2092[a]	3.4310[a]	4.2783[a]	3.9209[a]	3.3885[a]	4.0081[a]	3.4730[a]	2.6573[a]
SE	0.5815	0.6150	0.6842	0.5280	0.5521	0.6161	0.601	0.6159	0.6763
AIC	4907.00	4804.39	4458.37	4913.86	4812.60	4466.04	16158.28	16259.79	15558.52
EBITDATA									
β	-0.4419[a]	-0.1463	-0.0290	-0.4109[a]	-0.1540	-0.0349	-0.3719[a]	-0.0523	0.0821
SE	0.0952	0.1106	0.1293	0.0878	0.1021	0.1184	0.0987	0.1137	0.1354
AIC	4943.13	4525.44	3924.66	4944.80	4531.46	3928.27	16900.65	15599.45	13849.46
OPCE									
β	-0.3015[a]	-0.5390[a]	-0.0830	-0.2527[b]	-0.4796[a]	-0.1038	-0.2982[b]	-0.4521[a]	-0.0438
SE	0.1351	0.1428	0.1560	0.1236	0.1302	0.1389	0.1304	0.1338	0.1444
AIC	5180.19	5010.50	4688.66	5185.94	5020.43	4698.28	17676.46	17420.71	16541.07
NIS									
β	-0.4056[a]	-0.3281[a]	-0.1338	-0.3923[a]	-0.3229[a]	-0.1613[b]	-0.4466[a]	-0.3397[a]	-0.1647[b]
SE	0.0759	0.0809	0.0887	0.0696	0.0732	0.0795	0.0752	0.0779	0.0841
AIC	5130.54	5021.24	4681.91	5138.88	5032.81	4696.54	17108.88	17131.33	16317.38
WCTA									
β	-2.3753[a]	-1.5757[a]	-1.0011[a]	-2.1715[a]	-1.4630[a]	-0.9621[a]	-2.0940[a]	-1.2861[a]	-0.6302[a]

	SE	0.1790	0.1790	0.1188	0.1616	0.1615	0.1717	0.1738	0.1723	0.1805
	AIC	5004.96	4924.59	4641.70	5012.94	4929.73	4649.67	16388.07	16688.53	16071.58
WCS										
	β	-1.1173[a]	-0.7756[a]	-0.4955[a]	-1.0632[a]	-0.7286[a]	-0.4557[a]	-1.1180[a]	-0.7139[a]	-0.3625[a]
	SE	0.1230	0.1185	0.1190	0.1155	0.1083	0.1089	0.1262	0.1190	0.1210
	AIC	4827.48	4722.42	4436.58	4834.08	4730.81	4450.76	15642.71	15998.45	15491.13
CAG										
	β	-0.7359[a]	-0.5442[a]	-0.2251[b]	-0.6653[a]	-0.5101[a]	-0.2557[a]	-0.4378[a]	-0.2648[a]	-0.0032
	SE	0.0991	0.1015	0.0986	0.0910	0.0914	0.0883	0.0863	0.0840	0.0826
	AIC	4939.88	4600.72	4203.41	4947.41	4606.36	4208.70	17167.37	16333.60	15133.20
SAG										
	β	-1.2002[a]	-0.7332[a]	-0.4815[a]	-1.1010[a]	-0.6863[a]	-0.4584[a]	-0.5733[a]	-0.1759	0.1238
	SE	0.1472	0.1444	0.1465	0.1348	0.1289	0.1290	0.1312	0.1249	0.1245
	AIC	4850.35	4613.32	4226.02	4857.08	4624.41	4237.87	16259.93	15966.82	15100.79
TTA										
	β	-15.279[a]	-19.145[a]	-14.104[a]	-12.914[a]	-16.382[a]	-12.404[a]	-9.151[a]	-10.854[a]	-5.7572[a]
	SE	2.6195	2.8173	2.7590	2.3253	2.5141	2.455	2.396	2.433	2.3351
	AIC	5364.37	5199.73	4846.44	5373.75	5214.21	4854.49	18084.87	17845.73	16942.59
STDEBV										
	β	-0.1305	0.3249[a]	0.2026[c]	-0.1293	0.2497[b]	0.1838[c]	0.0581	0.3164[a]	0.1792[c]
	SE	0.1211	0.1127	0.1226	0.1075	0.0984	0.1070	0.1023	0.0949	0.0972
	AIC	5536.37	5259.03	4875.63	5540.88	5267.90	4883.93	18894.16	18123.03	16986.19
EBITDAIE										
	β	-0.0005[a]	-0.0007[a]	-0.0007[a]	-0.0005[a]	-0.0006[a]	-0.0007[a]	-0.0005[a]	-0.0007[a]	-0.0007[a]
	SE	0.0001	0.0002	0.0002	0.0001	0.0001	0.0002	0.0001	0.0001	0.0002
	AIC	5116.50	4967.77	4579.99	5119.08	4972.20	4587.21	17237.08	17021.55	15996.35
RETA										
	β	-0.2179[a]	-0.0403	0.0579	-0.2123[a]	-0.0545[c]	0.0342	-0.0963[a]	0.0828[b]	0.2082[a]
	SE	0.0285	0.0313	0.0363	0.0259	0.0285	0.0326	0.0292	0.0328	0.0377
	AIC	5276.77	5164.54	4789.04	5277.01	5173.25	4802.39	17632.39	17494.17	16569.14
Section B: Event 2										
TLTA										
	β	2.4763[a]	2.2925[a]	0.6454[a]	1.9131[a]	1.7385[a]	0.4722[a]	1.5491[a]	1.2181[a]	-0.0936[a]
	SE	0.0332	0.0340	0.0317	0.02433	0.0249	0.0254	0.0269	0.0263	0.0268
	AIC	39819.01	38592.44	39138.22	40351.88	39033.42	39367.36	169228.88	169676.96	151787.27
CETL										
	β	-0.9832[a]	-0.8333[a]	-0.2073[a]	-0.8996[a]	-0.7392[a]	-0.1744[a]	-0.7573[a]	-0.5932[a]	-0.0674[a]
	SE	0.0175	0.0155	0.0076	0.0154	0.0135	0.0067	0.0153	0.0128	0.0071
	AIC	42936.67	40933.98	41487.03	42838.21	40938.75	41690.80	178868.12	179742.98	163152.60
CTA										
	β	-1.9080[a]	-2.2780[a]	-0.5437[a]	-1.5666[a]	-1.8406[a]	-0.4203[a]	-1.3751[a]	-1.5422[a]	-0.1930[a]
	SE	0.0753	0.0821	0.0770	0.0626	0.0667	0.0632	0.0692	0.0689	0.0691
	AIC	50466.9	47135.76	43106.02	50713.44	47333.51	43285.52	187157.31	186678.96	166773.82
FETA										
	β	32.157[a]	24.0993[a]	8.8175[a]	25.545[a]	18.576[a]	6.5496[a]	19.732[a]	11.720[a]	0.3837
	SE	0.4632	0.4610	0.4786	0.3534	0.3537	0.3826	0.3847	0.3743	0.3990
	AIC	44019.66	43860.67	41314.35	44370.10	44184.51	41531.49	181845.91	185102.80	163025.31
FES										
	β	7.2176[a]	6.3882[a]	3.7784[a]	5.5920[a]	4.7768[a]	2.8066[a]	4.8881[a]	3.7060[a]	1.8270[a]

	SE	0.1558	0.1665	0.1733	0.1181	0.1252	0.1355	0.1368	0.1371	0.1472
	AIC	42745.25	40838.36	37532.89	43045.40	41128.05	37750.06	161702.51	162744.67	144887.85
EBITDATA										
	β	-1.0991[a]	-1.5416[a]	-0.8150[a]	-0.8326[a]	-1.1046[a]	-0.5830[a]	-0.6991[a]	-0.8538[a]	-0.3375[a]
	SE	0.0233	0.0288	0.0275	0.0173	0.0200	0.0206	0.0195	0.0203	0.0221
	AIC	46126.68	39795.64	36760.93	46499.05	40255.74	37028.58	177955.61	165151.81	143739.48
OPCE										
	β	-0.1529[a]	-1.0081[a]	-1.1616[a]	-0.0848[a]	-0.7633[a]	-0.8937[a]	-0.1837[a]	-0.6900[a]	-0.7763[a]
	SE	0.0320	0.0351	0.0374	0.0268	0.0284	0.0296	0.0269	0.0274	0.0291
	AIC	49958.13	46026.44	41098.81	50193.82	46327.37	41345.69	183139.45	182568.16	161874.04
NIS										
	β	-0.5768[a]	-1.0970[a]	-0.7226[a]	-0.4656[a]	-0.8509[a]	-0.5663[a]	-0.5472[a]	-0.9142[a]	-0.5837[a]
	SE	0.0201	0.0237	0.0229	0.0165	0.0181	0.0181	0.0192	0.0197	0.0203
	AIC	45352.45	40878.51	38124.25	45592.83	41159.19	38350.81	164494.07	162248.50	146980.02
WCTA										
	β	-3.2430[a]	-3.1762[a]	-0.8462[a]	-2.6079[a]	-2.4827[a]	-0.6387[a]	-2.1781[a]	-1.8792[a]	-0.0260
	SE	0.0490	0.0516	0.0461	0.0379	0.0391	0.0376	0.0413	0.0404	0.0403
	AIC	44180.67	41937.58	41577.89	44532.85	42262.79	41793.63	178733.60	179140.35	161417.71
WCS										
	β	-1.5368[a]	-1.4041[a]	-0.1892[a]	-1.3067[a]	-1.1611[a]	-0.1411[a]	-0.9721[a]	-0.7765[a]	0.1827[a]
	SE	0.0336	0.0336	0.0276	0.0282	0.0276	0.0227	0.0286	0.0270	0.0245
	AIC	41981.14	39752.62	37887.54	42170.25	39923.35	38056.91	158354.53	158024.93	142668.54
CAG										
	β	-0.4236[a]	-1.1145[a]	-0.4648[a]	-0.3359[a]	-0.9067[a]	-0.3640[a]	-0.1901[a]	-0.6554[a]	-0.1948[a]
	SE	0.0209	0.0271	0.0240	0.0177	0.0230	0.0199	0.0170	0.0218	0.0194
	AIC	46446.14	40032.58	36510.56	46667.55	40334.54	36707.40	182447.86	161571.4	142418.85
SAG										
	β	-0.4975[a]	-0.8480[a]	-0.3526[a]	-0.4018[a]	-0.6950[a]	-0.2773[a]	-0.0007	-0.2608[a]	0.0784[a]
	SE	0.0321	0.0351	0.0356	0.0266	0.0292	0.0291	0.0263	0.0286	0.0288
	AIC	42940.24	38469.84	34404.26	43119.91	38641.54	34566.08	163908.56	147657.20	130515.75
TTA										
	β	-24.529[a]	-46.053[a]	-28.388[a]	-21.440[a]	-39.567[a]	-25.191[a]	-16.111[a]	-32.700[a]	-17.570[a]
	SE	0.8287	1.2150	0.9206	0.7345	1.0230	0.8199	0.7956	1.0520	0.8362
	AIC	49823.67	45309.29	41671.94	50062.04	45472.27	41817.08	186556.15	185582.80	165637.98
STDEBV										
	β	0.1701[a]	0.3928[a]	0.3226[a]	0.1142[a]	0.2924[a]	0.2537[a]	0.2214[a]	0.3352[a]	0.3159[a]
	SE	0.0299	0.0312	0.0329	0.0250	0.0252	0.0261	0.0248	0.0243	0.0261
	AIC	50832.49	47532.22	42826.97	51068.9	47749.17	43001.94	186510.11	185977.43	165771.45
EBITDAIE										
	β	-0.0007[a]	-0.0020[a]	-0.0014[a]	-0.0006[a]	-0.0016[a]	-0.0013[a]	-0.0008[a]	-0.0021[a]	-0.0014[a]
	SE	0.0001	0.0001	0.0001	0.00004	0.0001	0.0001	0.0001	0.0001	0.0001
	AIC	49315.06	45838.27	41007.13	49550	46036.04	41161.93	184043.20	185078.73	162333.72
RETA										
	β	-0.4932[a]	-0.4340[a]	-0.1911[a]	-0.3929[a]	-0.3403[a]	-0.1501[a]	-0.2244[a]	-0.1383[a]	0.0965[a]
	SE	0.0078	0.0079	0.0077	0.0059	0.0061	0.0062	0.0068	0.0068	0.0072
	AIC	45778.64	44205.43	42160.60	46152.58	44510.55	42364.95	185095.29	185404.33	165212.43

Section C: Event 3	
TLTA	

	β	2.2613^{a}	2.2702^{a}	1.7594^{a}	2.0836^{a}	2.0658^{a}	1.6174^{a}	1.9450^{a}	1.7950^{a}	1.0670^{a}
	SE	0.2104	0.2194	0.2057	0.1874	0.1927	0.1835	0.2118	0.2594	0.2529
	AIC	1672.94	1605.55	1607.53	1678.16	1606.31	1612.21	8793.33	11036.43	11439.28
CETL										
	β	-0.8100^{a}	-1.3677^{a}	-1.0503^{a}	-0.7819^{a}	-1.2917^{a}	-0.9926^{a}	-0.5795^{a}	-1.1940^{a}	-0.8392^{a}
	SE	0.1256	0.1756	0.1501	0.1204	0.1596	0.1352	0.1162	0.1653	0.1334
	AIC	1807.22	1713.77	1705.97	1812.53	1715.99	1709.98	11896.03	8558.01	8911.28
CTA										
	β	-1.2774^{b}	-2.7134^{a}	-3.1083^{a}	-1.0860^{b}	-2.3802^{a}	-2.7352^{a}	-1.3870^{a}	-2.8200^{a}	-2.7380^{a}
	SE	0.4999	0.5819	0.6234	0.4575	0.5207	0.5522	0.5584	0.6180	0.6329
	AIC	1956.25	1879.62	1829.70	1964.16	1889.12	1839.38	13023.02	12597.72	12360.73
FETA										
	β	24.262^{a}	29.334^{a}	28.935^{a}	21.696^{a}	27.067^{a}	26.671^{a}	20.470^{a}	24.470^{a}	21.970^{a}
	SE	2.8578	3.0460	3.2041	2.5710	2.7135	2.8383	3.0910	3.2460	3.3690
	AIC	1782.76	1768.11	1734.50	1792.13	1772.51	1737.39	11905.15	11712.32	11686.16
FES										
	β	5.9683^{a}	5.9914^{a}	6.2774^{a}	5.6128^{a}	5.5103^{a}	5.8787^{a}	5.8460^{a}	5.7780^{a}	5.9680^{a}
	SE	0.9425	0.9578	1.0459	0.8731	0.8526	0.9317	1.2450	1.2390	1.2920
	AIC	1683.05	1720.33	1663.62	1688.37	1726.73	1670.39	11989.54	12111.67	11787.72
EBITDATA										
	β	-0.6521^{a}	-0.7622^{a}	-0.8135^{a}	-0.6212^{a}	-0.6576^{a}	-0.7441^{a}	-0.4909^{a}	-0.4434^{a}	-0.7848^{a}
	SE	0.1534	0.1749	0.1794	0.1428	0.1497	0.1628	0.1799	0.1988	0.2319
	AIC	1719.03	1598.20	1437.74	1724.62	1599.91	1444.79	12248.42	11576.53	10713.39
OPCE										
	β	0.4351^{c}	-0.0031	-0.5827^{b}	0.4065^{c}	-0.0273	-0.5849^{a}	0.2023	-0.0342	-0.6290^{c}
	SE	0.2456	0.2434	0.2499	0.2257	0.2239	0.2246	0.2418	0.2353	0.2484
	AIC	1842.11	1840.29	1798.55	1848.99	1848.20	1807.13	12810.22	12672.58	12406.50
NIS										
	β	-0.1690	-0.5182^{a}	-0.6758^{a}	-0.1570	-0.4651^{a}	-0.6240^{a}	-0.2152	-0.5197^{a}	-0.7127^{a}
	SE	0.1341	0.1321	0.1387	0.1235	0.1172	0.1219	0.1566	0.1499	0.1585
	AIC	1748.85	1759.28	1689.10	1757.05	1766.78	1697.26	12644.75	12542.26	12144.18
WCTA										
	β	-2.9201^{a}	-3.4000^{a}	-2.8521^{a}	-2.7048^{a}	-3.1075^{a}	-2.6474^{a}	-2.1300^{a}	-2.5091^{a}	-1.6840^{a}
	SE	0.3286	0.3533	0.3375	0.2987	0.3107	0.3000	0.3571	0.3645	0.3442
	AIC	1745.31	1701.38	1700.15	1751.42	1706.69	1705.59	12114.67	11805.99	11934.96
WCS										
	β	-1.8870^{a}	-1.8229^{a}	-1.6438^{a}	-1.7958^{a}	-1.7241^{a}	-1.5782^{a}	-1.5530^{a}	-1.3580^{a}	-1.0540^{a}
	SE	0.2770	0.2724	0.2729	0.2609	0.2501	0.2482	0.2890	0.2652	0.2595
	AIC	1638.45	1658.85	1613.76	1643.09	1664.47	1620.71	11665.17	11757.54	11502.82
CAG										
	β	-0.2799^{c}	-0.5790^{a}	-1.2346^{a}	-0.2719^{b}	-0.5426^{a}	-1.1636^{a}	-0.0643	-0.2673^{c}	-0.7301^{a}
	SE	0.1512	0.1731	0.2203	0.1390	0.1586	0.2026	0.1342	0.1440	0.1791
	AIC	1830.94	1775.43	1610.59	1838.31	1783.90	1616.51	12647.23	12316.50	11161.17
SAG										
	β	-1.1656^{a}	-1.6190^{a}	-1.7849^{a}	-1.1151^{a}	-1.4808^{a}	-1.4904^{a}	-0.5766^{b}	-0.9741^{a}	-0.7436^{a}
	SE	0.2665	0.2951	0.3119	0.2478	0.2693	0.2711	0.2661	0.2680	0.2664
	AIC	1719.37	1666.91	1556.88	1725.12	1675.49	1566.20	12345.43	11927.15	11210.35
TTA										
	β	-20.974^{a}	-20.448^{a}	-48.177^{a}	-17.538^{a}	-17.843^{a}	-42.345^{a}	-17.620^{a}	-16.050^{a}	-45.040^{a}

SE	5.5795	5.5359	8.4749	4.8925	4.9486	7.4022	5.6440	5.4830	9.1990
AIC	1890.55	1877.57	1799.09	1898.95	1886.24	1808.33	12943.40	12792.58	12340.86
STDEBV									
β	-1.7364[a]	-0.0330	0.5541[a]	-1.5598[a]	-0.0460	0.5348[a]	-0.8312[a]	0.0242	0.4216[b]
SE	0.3239	0.1968	0.1824	0.3018	0.1769	0.1596	0.2529	0.1814	0.1770
AIC	1901.23	1897.58	1846.94	1908.93	1905.26	1854.31	12948.38	12805.56	12503.92
EBITDAIE									
β	-0.0002	-0.0003	-0.0009[b]	-0.0002	-0.0003	-0.0008[b]	-0.0003	-0.0003	-0.0009[b]
SE	0.0002	0.0003	0.0004	0.0002	0.0003	0.0004	0.0003	0.0003	0.0004
AIC	1827.81	1851.02	1791.41	1833.82	1858.40	1799.65	12565.97	12561.62	12210.71
RETA									
β	-0.4576[a]	-0.3484[a]	-0.2025[a]	-0.4209[a]	-0.3201[a]	-0.1846[a]	-0.2789[a]	-0.1585[a]	0.0303
SE	0.0511	0.0484	0.0493	0.0453	0.0431	0.0441	0.0598	0.0596	0.0627
AIC	1788.77	1813.79	1811.38	1793.14	1819.79	1820.26	12422.61	12528.49	12529.59

Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table reports univariate regression estimates of Event 1, Event 2 and Event 3 using respective hazard models and lagged time periods. Section A reports regression estimates of Event 1, Section B reports Event 2 and Section C reports Event 3.

Table 9: Multivariate Regression

Section A: Event 1

Variable	logit			clog-log			Cox		
	T – 1	*T – 2*	*T – 3*	*T – 1*	*T – 2*	*T – 3*	*T – 1*	*T – 2*	*T – 3*
STDEBV									
β		0.3008[b]			0.2010[c]			0.3709[a]	
SE		0.1295			0.1201			0.0884	
p-value		0.0200			0.0940			0.0000	
TLTA									
β	0.9948[a]			0.8597[a]			0.7369[a]		
SE	0.2226			0.1913			0.1384		
p-value	0.0000			0.0000			0.0000		
CETL									
β	-0.2530[a]			-0.2388[a]			-0.2259[a]		
SE	0.0700			0.0618			0.0448		
p-value	0.0000			0.0000			0.0000		
CTA									
β	-0.3425		-2.2829[a]	-0.2665		-1.9289[a]	-0.2804		-1.7416[a]
SE	0.4572		0.4935	0.3970		0.4305	0.2786		0.2893
p-value	0.2440		0.0000	0.2320		0.0000	0.2100		0.0000
FETA									
β	7.3617[a]	12.369[a]	6.1982[b]	6.0372[a]	11.740[a]	6.4750[a]	3.0878[a]	9.1737[a]	5.5547[a]
SE	2.7061	2.3270	2.5504	2.3321	2.2416	2.2326	1.8124	1.6206	1.5421
p-value	0.0070	0.0000	0.0150	0.0100	0.0000	0.0040	0.0000	0.0000	0.0030
FES									
β									
SE									

p-value									
EBITDAIE									
β		-0.0005[b]	-0.0006[a]		-0.0005[b]	-0.0006[a]		-0.0003[b]	-0.0005[a]
SE		0.0002	0.0002		0.0002	0.0002		0.0001	0.0001
p-value		0.0200	0.0090		0.0120	0.0090		0.0420	0.0019
EBITDATA									
β									
SE									
p-value									
OPCE									
β	-0.5270[a]	-0.2243		-0.4234[a]	-0.2546		-0.2847[b]	-0.1002	
SE	0.1812	0.1809		0.1550	0.1697		0.1185	0.1277	
p-value	0.0040	0.2150		0.0060	0.1340		0.0160	0.2300	
NIS									
β									
SE									
p-value									
RETA									
β									
SE									
p-value									
WCTA		-0.5958[b]			-0.6578[a]			-0.5090[a]	
β		0.2543			0.2462			0.1748	
SE		0.0190			0.0080			0.0036	
p-value									
WCS									
β									
SE									
p-value									
CAG									
β	-0.2328[c]	-0.1335		-0.2025[c]	-0.1082		-0.3474[a]	-0.2175[b]	
SE	0.1206	0.1120		0.1063	0.1039		0.0913	0.0862	
p-value	0.0540	0.2330		0.0570	0.2480		0.0000	0.0120	
SAG									
β	-0.5326[a]	-0.2226	-0.2023	-0.4826[a]	-0.1808	-0.2368[c]	-0.8411[a]	-0.4764[a]	-0.2778[b]
-SE	0.1748	0.1528	0.1643	0.1551	0.1438	0.1415	0.1351	0.1200	0.1155
p-value	0.0020	0.1450	0.2180	0.0020	0.2090	0.0940	0.0000	0.0000	0.0160
TTA									
β	-4.1461	-10.394[a]	-10.428[a]	-3.4981	-9.5294[a]	-8.9473[a]	-1.1812	-7.6240[a]	-7.0481[a]
SE	3.5669	3.6154	3.5802	3.1214	3.3587	3.0952	2.4601	2.5793	2.3600
p-value	0.2450	0.0040	0.0040	0.2420	0.0050	0.0040	0.6300	0.0031	0.0000
Micro									
β	2.0176[a]	1.9982[a]	2.5646[a]	1.7140[a]	2.0089[a]	2.2413[a]	0.7910[a]	1.1579[a]	1.5554[a]
SE	0.2419	0.2119	0.2412	0.2078	0.2038	0.2116	0.1295	0.1254	0.1257
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Small									
β	0.7552[a]	0.8696[a]	1.2075[a]	0.6548[a]	0.8792[a]	1.0438[a]	0.1142	0.3240[a]	0.5422[a]
SE	0.2064	0.1899	0.2060	0.1829	0.1802	0.1861	0.1266	0.1225	0.1275

p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3700	0.0084	0.0000
AGE									
	0.4964^{a}	0.5408^{a}	0.7246^{a}	0.3949^{a}	0.5447^{a}	0.6130^{a}	-33.253^{a}	-36.575^{a}	-43.231^{a}
β	0.1452	0.1451	0.1835	0.1266	0.1435	0.1631	1.0312	1.2182	1.2320
SE	0.0010	0.0000	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
p-value									
RISK1	90.495^{a}	77.879^{a}	91.854^{a}	77.607^{a}	78.677^{a}	78.7463^{a}	46.871^{a}	44.689^{a}	43.998^{a}
β	7.3345	6.1372	7.3431	6.2999	6.1021	6.4905	3.0780	2.9346	3.3759
SE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p-value									
Model's Goodness of Fit and Performance Measure									
Chi2	403.06^{a}	321.8^{a}	311.8^{a}	411.2^{a}	377.8^{a}	311.9^{a}	2785^{a}	2874^{a}	3222^{a}
likelihood	-1746.9	-1837.2	-1791.8	-1748.6	-1838.5	-1790.4	-3051.2	-3291.5	-3283.1
AIC	3521.9	3702.3	3605.6	3525.3	3705.1	3602.9	6294.5	6753.4	7237.08
N	46927	44400	40882	46927	44400	40882	46927	44400	40882
Event	433	464	469	433	464	469	433	464	469
AUROC-W	0.8209	0.7936	0.7827	0.8212	0.7929	0.7831	0.8192	0.7912	0.7797
AUROC-H	0.7943	0.8339	0.9242	0.7890	0.8937	0.9233	0.7572	0.8784	0.9112
Section B: Event 2									
STDEBV									
β	0.1743^{a}	0.2511^{a}	0.2533^{a}	0.1544^{a}	0.2151^{a}	0.1998^{a}	0.2003^{a}	0.2046^{a}	0.1974^{a}
SE	0.0388	0.0453	0.0456	0.0290	0.0328	0.0348	0.0233	0.0249	0.0290
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TLTA									
β	1.9886^{a}	2.2478^{a}	0.3296^{a}	1.4622^{a}	1.5668^{a}	0.2143^{a}	1.1641^{a}	1.1266^{a}	0.4912^{a}
SE	0.0654	0.0777	0.0723	0.0481	0.0548	0.0549	0.0347	0.0370	0.0418
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CETL									
β									
SE									
p-value									
CTA									
β		-1.3447^{a}			-0.9909^{a}			-0.5477^{a}	
SE		0.1462			0.1082			0.0769	
p-value		0.0000			0.0000			0.0000	
FETA									
β	15.786^{a}	2.6571^{a}	3.3366^{a}	11.798^{a}	2.1859^{a}	2.6998^{a}	8.6179^{a}	3.6960^{a}	4.5587^{a}
SE	0.7083	0.9955	0.9549	0.5354	0.7093	0.7338	0.4185	0.5306	0.6053
p-value	0.0000	0.0080	0.0000	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000
FES									
β		0.3141	0.8577^{a}		0.0920	0.5806^{a}		-0.3210^{b}	0.1233
SE		0.2911	0.2675		0.2110	0.2061		0.1461	0.1607
p-value		0.2410	0.0010		0.6630	0.0050		0.0280	0.4400
EBITDAIE									
β	-0.0003^{a}	-0.0004^{a}	-0.0006^{a}	-0.0004^{a}	-0.0004^{a}	-0.0006^{a}	-0.0005^{a}	-0.0004^{a}	-0.0006^{a}
SE	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
p-value	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EBITDATA									

β	-0.1624[a]	-0.7619[a]	-0.3114[a]	-0.0850[a]	-0.5110[a]	-0.2281[a]	-0.0860[a]	-0.3205[a]	-0.1783[a]
SE	0.0380	0.0467	0.0422	0.0264	0.0312	0.0312	0.0177	0.0205	0.0236
p-value	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
OPCE									
β	-0.9026[a]	-1.4723[a]	-0.8999[a]	-0.6398[a]	-1.0036[a]	-0.6650[a]	-0.5802[a]	-0.7744[a]	-0.6558[a]
SE	0.0471	0.0580	0.0538	0.0336	0.0395	0.0409	0.0260	0.0289	0.0341
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
NIS									
β									
SE									
p-value									
RETA									
β									
SE									
p-value									
WCTA									
β	-0.5875[a]	-0.2946[a]	-0.1058	-0.5189[a]	-0.3310[a]	-0.1064	-0.3858[a]	-0.2745[a]	-0.0722
SE	0.0881	0.1130	0.0948	0.0668	0.0832	0.0733	0.0501	0.0595	0.0566
p-value	0.0000	0.0090	0.2440	0.0000	0.0000	0.1470	0.0000	0.0000	0.2000
WCS									
β									
SE									
p-value									
CAG									
β		-0.7094[a]	-0.1534[a]		-0.4575[a]	-0.1116[a]		-0.4140[a]	-0.1096[a]
SE		0.0373	0.0300		0.0278	0.0240		0.0239	0.0221
p-value		0.0000	0.0000		0.0000	0.0000		0.0000	0.0000
SAG									
β	-0.0569	-0.3881[a]		-0.0652[b]	-0.2993[a]		-0.1152[a]	-0.2601[a]	
SE	0.0395	0.0461		0.0300	0.0342		0.0252	0.0284	
p-value	0.1500	0.0000		0.0300	0.0000		0.0000	0.0000	
TTA									
β	-11.374[a]	-28.651[a]	-14.413[a]	-10.596[a]	-25.191[a]	-13.489[a]	-11.846[a]	-24.508[a]	-14.586[a]
SE	1.0783	1.5922	1.1580	0.9030	1.3307	1.0279	0.7909	1.0302	0.9033
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Micro									
β	0.4154[a]	0.5147[a]	1.0577[a]	0.2928[a]	0.3762[a]	0.8451[a]	0.1853[a]	0.2582[a]	0.6144[a]
SE	0.0662	0.0742	0.0683	0.0504	0.0556	0.0536	0.0342	0.0366	0.0377
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Small									
β	0.2786[a]	0.3265[a]	0.6749[a]	0.2094[a]	0.2675[a]	0.5466[a]	0.1471[a]	0.1845[a]	0.4122[a]
SE	0.0488	0.0551	0.0521	0.0381	0.0423	0.0421	0.0279	0.0301	0.0316
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AGE									
β	-0.0285	-0.1323[a]	-0.0029	-0.0732[a]	-0.1535[a]	-0.0137	-24.865[a]	-27.661[a]	-33.798[a]
SE	0.0353	0.0445	0.0453	0.0277	0.0343	0.0366	0.2734	0.3437	0.4217
p-value	0.4190	0.0030	0.9490	0.0080	0.0000	0.7090	0.0000	0.0000	0.0000
RISK2									
β	5.4673[a]	4.0201[a]	6.5065[a]	4.1230[a]	3.2239[a]	5.2226[a]	2.4035[a]	1.7983[a]	3.3036[a]

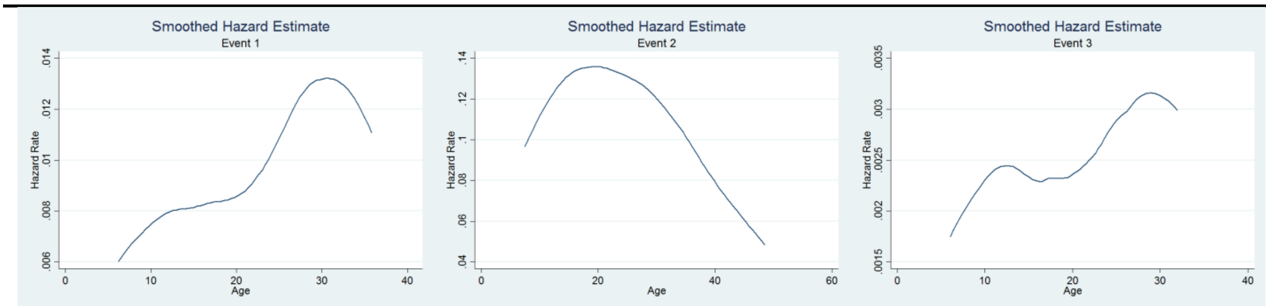
SE	0.4117	0.4662	0.4423	0.3152	0.3513	0.3475	0.2266	0.2531	0.2688
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Model's Goodness of Fit and Performance Measure									
Chi2	5135.9[a]	4381.9[a]	2228.2[a]	5818.2[a]	5039.8[a]	2307.3[a]	42073[a]	37046[a]	29048[a]
likelihood	-13550.1	-10534.2	-12115.9	-13807.9	-10862.8	-12225.9	-52703.6	-42792.4	-39919.4
AIC	27130.1	21104.3	24263.6	27645.9	21761.6	24483.8	105433.1	85616.7	79866.8
N	44740	36907	33396	44740	36907	33396	44740	36907	33396
Event	7553	6390	5721	7553	6390	5721	7553	6390	5721
AUROC-W	0.8739	0.9015	0.7794	0.8721	0.8991	0.7767	0.8699	0.8969	0.7865
AUROC-H	0.8436	0.8714	0.7783	0.8425	0.8686	0.7745	0.8414	0.8705	0.7803
Section C: Event 3									
STDEBV									
β		0.4912[c]	0.6417[b]		0.4266[c]	0.4680[c]		0.6404[a]	0.6278[a]
SE		0.2715	0.2861		0.2250	0.2401		0.1721	0.1593
p-value		0.0700	0.0250		0.0580	0.0510		0.0000	0.0000
TLTA									
β		1.7594[a]	1.5434[a]		1.6420[a]	1.3451[a]		1.4437[a]	1.0166[a]
SE		0.4712	0.4767		0.3943	0.4140		0.2787	0.2551
p-value		0.0000	0.0010		0.0000	0.0010		0.0000	0.0000
CETL									
β	-0.4543[a]			-0.4430[a]			-1.0176[a]		
SE	0.1747			0.1682			0.1237		
p-value	0.0090			0.0080			0.0000		
CTA									
β			-1.3311			-1.0743			-0.9048
SE			1.1703			1.0090			0.6390
p-value			0.2450			0.2870			0.1600
FETA									
β	12.022[a]	16.656[a]	15.631[a]	9.5936[a]	13.686[a]	14.313[a]	5.7686[b]	10.272[a]	7.3060[b]
SE	4.1089	4.9550	5.7427	3.5011	4.1907	4.9884	2.8568	3.3580	3.5285
p-value	0.0030	0.0010	0.0060	0.0060	0.0010	0.0040	0.0440	0.0000	0.0380
FES									
β									
SE									
p-value									
EBITDAIE									
β									
SE									
p-value									
EBITDATA									
β									
SE									
p-value									
OPCE									
β		-0.4816	-0.3651		-0.4501	-0.2209		-0.5053[b]	-0.3928[c]
SE		0.3382	0.3815		0.2829	0.3328		0.2123	0.2202
p-value		0.1540	0.2390		0.1120	0.5070		0.0170	0.0750
NIS									
β									

SE									
p-value									
RETA									
β	-0.2000[a]			-0.1957[a]			0.0104		
SE	0.0751			0.0659			0.0451		
p-value	0.0080			0.0030			0.8199		
WCTA									
β	-0.6195	-0.8440		-0.5472	-0.5571		0.0368	-0.4660	
SE	0.5165	0.6515		0.4475	0.5384		0.3657	0.4109	
p-value	0.2300	0.1950		0.2210	0.3010		0.9200	0.2600	
WCS									
β			-0.4132			-0.3643			-0.2302
SE			0.4486			0.3830			0.2504
p-value			0.2470			0.3410			0.3600
CAG									
β			-0.5432[b]			-0.4291[c]			-0.4133[b]
SE			0.2587			0.2286			0.1692
p-value			0.0360			0.0610			0.0150
SAG									
β	-0.2695	-0.7311[b]	-0.8411[b]	-0.1909	-0.6557[b]	-0.7206[b]	-0.7072[a]	-1.0009[a]	-0.6980[a]
SE	0.2873	0.3311	0.3575	0.2571	0.2863	0.3126	0.2307	0.2441	0.2369
p-value	0.2480	0.0270	0.0190	0.4580	0.0220	0.0210	0.0022	0.0004	0.0032
TTA									
β	-15.152[c]	-3.948	-36.756[a]	-14.043[c]	-3.288	-30.773[a]	-12.173[b]	-3.463	-21.901[a]
SE	8.458	8.164	11.596	7.407	6.970	10.121	5.458	5.151	7.195
p-value	0.0730	0.6290	0.0020	0.0580	0.6370	0.0020	0.0260	0.5000	0.0023
Micro									
β	1.9486[a]	2.4476[a]	2.7400[a]	1.6619[a]	1.9905[a]	2.2511[a]	1.3615[a]	1.3692[a]	1.5166[a]
SE	0.4272	0.4541	0.4979	0.3718	0.3700	0.4144	0.2614	0.2485	0.2334
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Small									
β	0.4747	0.4216	0.6411	0.3976	0.3934	0.5285	0.4086[c]	0.2318	0.2639
SE	0.3809	0.4366	0.4742	0.3376	0.3681	0.4114	0.2418	0.2565	0.2572
p-value	0.2130	0.2340	0.1760	0.2390	0.2850	0.1990	0.0910	0.3700	0.3000
AGE									
β	0.9837[a]	0.7653[b]	0.4242	0.8869[a]	0.6001[b]	0.3010	-42.888[a]	-47.477[a]	-35.227[a]
SE	0.2714	0.3363	0.3910	0.2386	0.2799	0.3439	2.4175	2.0754	3.2681
p-value	0.0000	0.0230	0.2480	0.0000	0.0320	0.3810	0.0000	0.0000	0.0000
RISK3									
β	219.806[a]	226.564[a]	239.428[a]	204.698[a]	189.601[a]	211.302[a]	94.989[a]	87.859[a]	79.818[a]
SE	22.477	24.405	29.131	16.901	17.913	21.444	11.357	11.824	10.728
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Model's Goodness of Fit and Performance Measure									
Chi2	170.1[a]	155.2[a]	132.6[a]	236.5[a]	202.6[a]	167.3[a]	1149[a]	1112[a]	1084[a]
likelihood	-657.02	-590.5	-529.3	-660.6	-592.1	-530.7	-938.6	-881.6	-808.5
AIC	1338.1	1207.1	1088.6	1345.2	1210.3	1091.4	2117.9	2048.1	1808.8
Censored	50126	40639	35327	50126	40639	35327	50126	40639	35327
Event	143	136	131	143	136	131	143	136	131
AUROC-W	0.8840	0.9019	0.9020	0.8840	0.9031	0.9015	0.8783	0.8955	0.8964

AUROC-H	0.9249	0.8924	0.9668	0.9317	0.9019	0.9214	0.9447	0.8653	0.9556

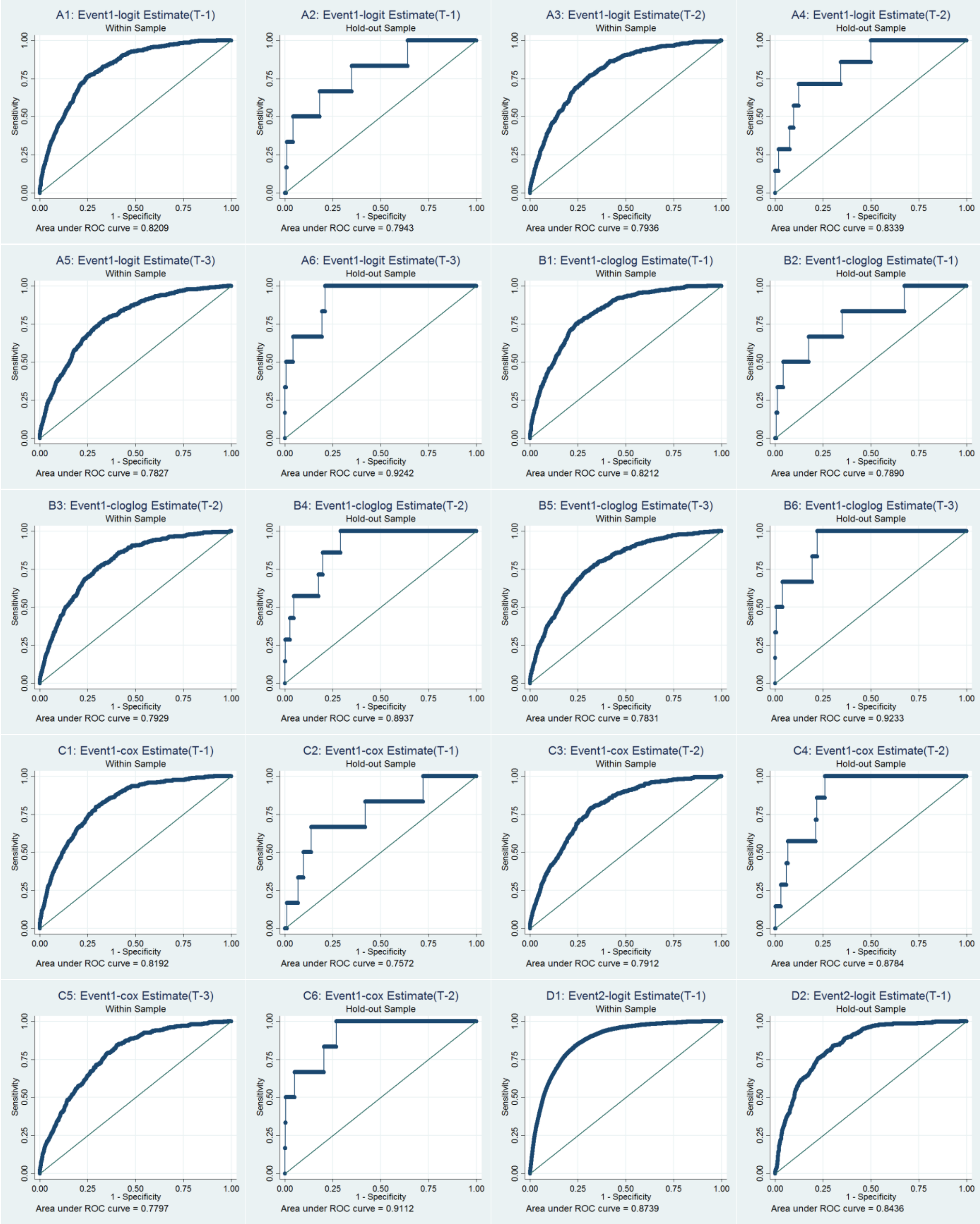
Notes: a (b) [c] significant at the 1 % (5 %) [10 %] level (two-sided test). This table reports multivariate regression estimates of Event 1, Event 2 and Event 3 using respective hazard models and lagged time periods. Section A reports regression estimates of Event 1, Section B reports Event 2 and Section C reports Event 3. The Chi2 values reported for logit and cloglog estimates is obtained using Wald test, while for Cox regression it is obtained using likelihood ratio test. AUROC-W represents within sample and AUROC-H represents hold-out sample area under ROC curves. 'Event' reports total number of observations with dependent variable = 1 and 'censored' reports total number of observations with dependent variable = 0.

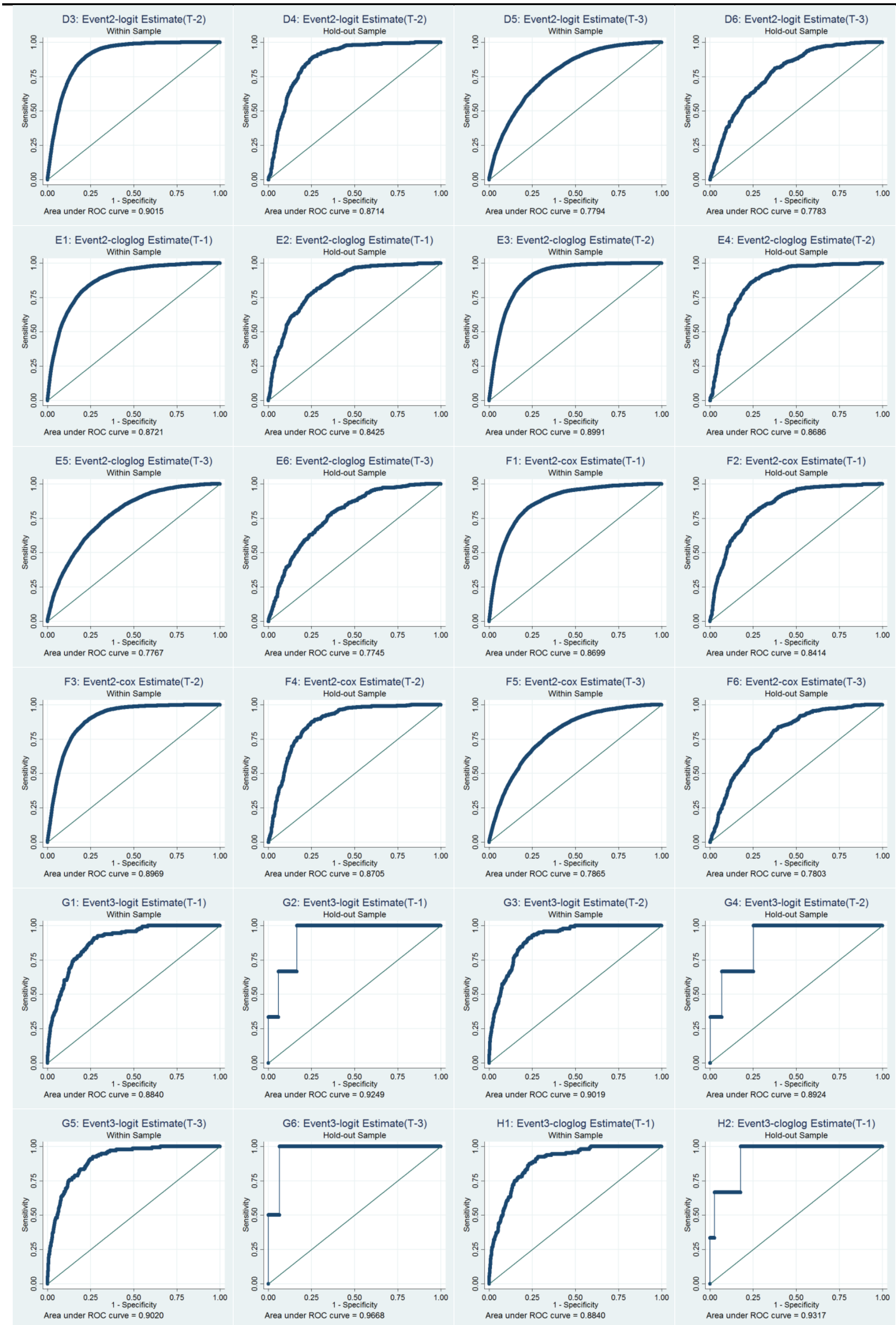
Figure 1: Table of Hazard Curves

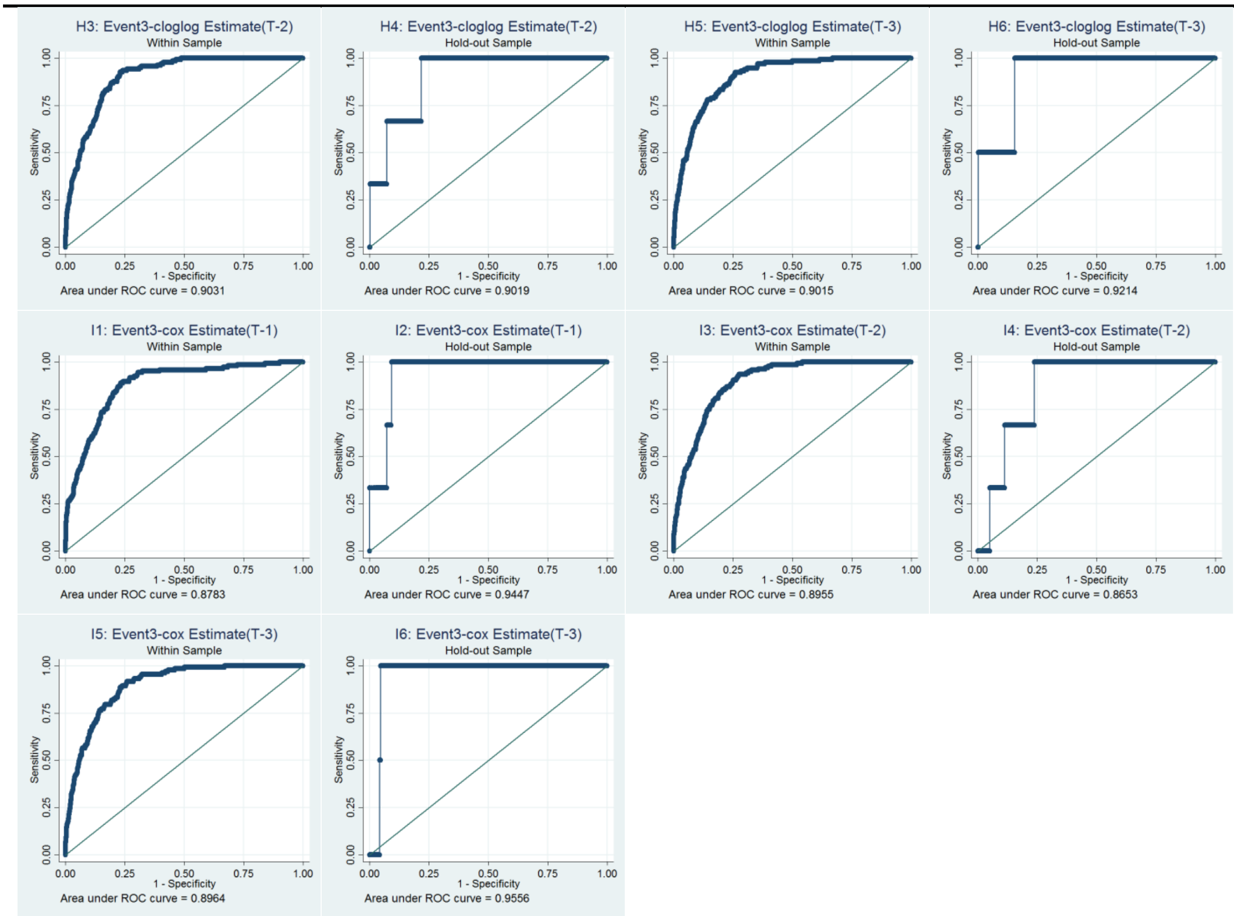


Notes: This table reports smoothed hazard curves estimated using the development sample for different definitions of financial distress events as discussed in section 2. Here 'Age' represents the age of firms in years.

Figure 2: Table of Area under ROC curves




D3: Event2-logit Estimate(T-2)
Within Sample
Area under ROC curve = 0.9015
D4: Event2-logit Estimate(T-2)
Hold-out Sample
Area under ROC curve = 0.8714
D5: Event2-logit Estimate(T-3)
Within Sample
Area under ROC curve = 0.7794
D6: Event2-logit Estimate(T-3)
Hold-out Sample
Area under ROC curve = 0.7783
E1: Event2-cloglog Estimate(T-1)
Within Sample
Area under ROC curve = 0.8721
E2: Event2-cloglog Estimate(T-1)
Hold-out Sample
Area under ROC curve = 0.8425
E3: Event2-cloglog Estimate(T-2)
Within Sample
Area under ROC curve = 0.8991
E4: Event2-cloglog Estimate(T-2)
Hold-out Sample
Area under ROC curve = 0.8686
E5: Event2-cloglog Estimate(T-3)
Within Sample
Area under ROC curve = 0.7767
E6: Event2-cloglog Estimate(T-3)
Hold-out Sample
Area under ROC curve = 0.7745
F1: Event2-cox Estimate(T-1)
Within Sample
Area under ROC curve = 0.8699
F2: Event2-cox Estimate(T-1)
Hold-out Sample
Area under ROC curve = 0.8414
F3: Event2-cox Estimate(T-2)
Within Sample
Area under ROC curve = 0.8969
F4: Event2-cox Estimate(T-2)
Hold-out Sample
Area under ROC curve = 0.8705
F5: Event2-cox Estimate(T-3)
Within Sample
Area under ROC curve = 0.7865
F6: Event2-cox Estimate(T-3)
Hold-out Sample
Area under ROC curve = 0.7803
G1: Event3-logit Estimate(T-1)
Within Sample
Area under ROC curve = 0.8840
G2: Event3-logit Estimate(T-1)
Hold-out Sample
Area under ROC curve = 0.9249
G3: Event3-logit Estimate(T-2)
Within Sample
Area under ROC curve = 0.9019
G4: Event3-logit Estimate(T-2)
Hold-out Sample
Area under ROC curve = 0.8924
G5: Event3-logit Estimate(T-3)
Within Sample
Area under ROC curve = 0.9020
G6: Event3-logit Estimate(T-3)
Hold-out Sample
Area under ROC curve = 0.9668
H1: Event3-cloglog Estimate(T-1)
Within Sample
Area under ROC curve = 0.8840
H2: Event3-cloglog Estimate(T-1)
Hold-out Sample
Area under ROC curve = 0.9317
Sensitivity
1 - Specificity
0.00
0.25
0.50
0.75
1.00


H3: Event3-cloglog Estimate(T-2)
Within Sample
Area under ROC curve = 0.9031
H4: Event3-cloglog Estimate(T-2)
Hold-out Sample
Area under ROC curve = 0.9019
H5: Event3-cloglog Estimate(T-3)
Within Sample
Area under ROC curve = 0.9015
H6: Event3-cloglog Estimate(T-3)
Hold-out Sample
Area under ROC curve = 0.9214
I1: Event3-cox Estimate(T-1)
Within Sample
Area under ROC curve = 0.8783
I2: Event3-cox Estimate(T-1)
Hold-out Sample
Area under ROC curve = 0.9447
I3: Event3-cox Estimate(T-2)
Within Sample
Area under ROC curve = 0.8955
I4: Event3-cox Estimate(T-2)
Hold-out Sample
Area under ROC curve = 0.8653
I5: Event3-cox Estimate(T-3)
Within Sample
Area under ROC curve = 0.8964
I6: Event3-cox Estimate(T-3)
Hold-out Sample
Area under ROC curve = 0.9556
Sensitivity
1 - Specificity