

# Mind the tail, or risk to fail

## **Jairaj Gupta**

Department of Finance, University of Birmingham  
Birmingham, B15 2TY, United Kingdom  
Email: J.Gupta@outlook.com

## **Sajid Chaudhry**

Department of Finance, Aston University  
Birmingham, B4 7ER, United Kingdom  
Email: s.chaudhry9@aston.ac.uk

January, 2019

### **Abstract**

In this study we hypothesise that more frequent extreme negative daily equity returns result in higher tail risk, and this subsequently increases firms' likelihood of entering financial distress. Specifically, we investigate the role of Value-at-risk and Expected Shortfall in aggravating firms' likelihood of experiencing financial distress. Our results show that longer horizon (three- and five-year) tail risk measures contributes positively toward firms' likelihood of experiencing financial distress. Additionally, considering the declining number of bankruptcy filings, and increasing out-of-court negotiations and debt reorganisations, we argue in favour of penalising firms for becoming sufficiently close to bankruptcy that they have questionable going-concern status. Thus, we propose a definition of financial distress contingent upon firms' earnings, financial expenses, market value and operating cash flow.

*Keywords:* tail risk; value-at-risk; downside risk; expected shortfall; bankruptcy; financial distress

*JEL Classification Codes:* G17, G33, C53

*Acknowledgements:* The authors are sincerely grateful to the Associate Editor Dr. Claire Crutchley Lending, anonymous referees, and seminar participants at the 49<sup>th</sup> Money, Macro and Finance Research Group Annual Conference (London, 2017), British Accounting & Finance Association 2018 Annual Conference (London, 2018), and 8<sup>th</sup> International Conference of the Financial Engineering and Banking Society (Rome, 2018) for their insightful comments and suggestions that improved this paper significantly. We also gratefully acknowledge financial support from the Lloyds Banking Group Centre for Responsible Business within the Birmingham Business School, United Kingdom. An earlier version of this paper was circulated under the title "*Financial Distress and Tail Risk*".

# 1. Introduction

*Financial Distress* and *Tail Risk* are two apparently diverse topics that are gaining increasing attention in the corporate finance literature. The financial crisis of 2007-08 was the alarm bell that augmented global awareness toward tail risk among financial risk managers. Since then, we have witnessed increasing concern among stakeholders toward firms' risk of bankruptcy or financial distress. Although tail risk has been an active area of investigation in the domain concerned with large financial institutions and financial stability, to the best of our knowledge this study is the first academic attempt to address the relationship between firms' extreme negative daily equity returns and their likelihood of experiencing financial distress. We hypothesise that more frequent extreme negative daily equity returns result in higher tail risk, and this subsequently increases firms' likelihood of entering financial distress.

The vast majority of academic literature on bankruptcy prediction gravitates around either the choice of explanatory variables (e.g. Campbell *et al.* 2008, Jones 2017), or modelling methodologies (e.g. Shumway, 2001, Gupta *et al.* 2018) targeted towards optimising models' classification performance. However, a model's performance is significantly dependent on how the distress or bankruptcy event is defined in the first place. Unfortunately, this aspect has not received sufficient attention in bankruptcy literature. Moreover, legal bankruptcy filings in the United States are becoming an increasingly rare phenomenon,<sup>1</sup> and this might provide false signals on firms' bankruptcy likelihood to external stakeholders. Waiting until the bankruptcy event might lead to significant erosion in firms' value (see Glover, 2016), losses to creditors and unfavourable business decisions in the years prior to the bankruptcy filing. Thus, we also argue for the need for an alternative mechanism to identify firms in financial difficulties, and penalise them for getting close to questionable going-concern status. This might be considered

---

<sup>1</sup> Out of about 7,538 firms in our sample, between 1990 and 2016, only 216 exit the Compustat database due to Chapter 7/11 filings.

analogous to identifying the symptoms of a disease (financial distress) and starting the treatment, rather than waiting until the fully-grown disease (bankruptcy). This is expected to improve/stabilise credit pricing and minimise losses arising due to mispricing of credit risk.

As a consequence, we propose a definition of financial distress contingent upon firms' earnings, financial expenses, market value and operating cash flow. Our proposed new definition of financial distress essentially builds on that offered by Pindado *et al.* (2008), which is considered to be more functional than its predecessors. We define a firm as financially distressed if it satisfies three simple conditions: (i) there is negative growth in the average market value (AMV) of the firm in years  $t - 1$  and  $t - 2$ , where AMV is the geometrically declining weighted average of a firm's monthly market values over the past twelve months; (ii) its earnings before interest tax depreciation and amortisation (EBITDA) is less than its financial expenses (FE) in the years  $t - 1$  and  $t - 2$ ; and (iii) its operating cash flow is less than its financial expenses in the years  $t - 1$  and  $t - 2$ .

Subsequently, we explore the information content of tail risk measures, namely Value-at-risk (VaR) and Expected Shortfall (ES), in explaining firms' likelihood of entering financial distress. Downside or tail risk measures focus only on the risk of underperforming a defined benchmark return and ignore the risk of outperforming the same. We consider the relationship between downside risk measures and financial distress for several reasons. First, there exists substantial literature pertaining to safety-first investors, whose objective is to minimise the probability of financial distress/bankruptcy. A safety-first investor's portfolio is intended to maximise expected return subject to downside risk constraint (see, among others, Roy, 1952, Arzac and Bawa, 1977). Second, as part of efficient risk management practices, financial and non-financial firms are increasingly required to quantify the amount of risk their portfolio or assets may incur over specified time horizons. For instance, banks assess their potential losses in order to maintain adequate levels of buffer capital. Similarly, credit rating agencies track the

value of firms' assets in order to provide them with an accurate credit rating. Finally, empirical evidence typically shows a skewed distribution of stock returns, with peaks around the mode and fat tails (see, among others, Jansen and De Vries, 1991, Conrad *et al.* 2013). This implies that occurrence of negative extreme events is more frequent than suggested by the normal distribution. Therefore, traditionally used measures of market risk (i.e. variance or standard deviation) might be insufficient to approximate the likelihood of maximum loss that a firm may witness under highly volatile or normal periods.<sup>2</sup>

There is a growing literature on downside risk and expected stock returns (see, for example, Bali *et al.* 2009, Huang *et al.* 2012). However, we did not come across any suitable literature that explores the relationship between tail risk measures and financial distress. Thus, we explore this relationship and, in addition to proposing a definition of financial distress, we assess the marginal discriminatory power of tail risk measures, namely VaR and ES, in predicting the financial distress of publicly-traded firms in the United States. The relationship we explore is intertemporal as we use three-month, six-month, one-year, three-year and five-year daily, and sixty-month monthly returns to measure the VaR and ES of all firms in our sample. Following Zangari (1996), Jaschke (2002) and Liang and Park (2010), we add the Cornish and Fisher (1938) expansion into these risk measures. Contrary to the normal assumption of these risk measures, the Cornish and Fisher (1938) expansion brings skewness and kurtosis into the equation, which is more appropriate for non-normal financial returns. We estimate the risk measure of each firm at the year-end to predict firms' financial distress in the following year. In terms of tail risk measures, our study closely relates to Bali *et al.* (2009), who study the intertemporal relationship between tail risk and expected return, and Liang and

---

<sup>2</sup> Empirical literature reports that VaR is reasonable in predicting catastrophic financial market events. It also efficiently captures the rate of occurrence of such events, including extreme cases. On the other hand, the traditional measures of market risk, such as standard deviation and conditional variance, provide an inaccurate prediction of extreme events in financial markets (Jorion, 2000, Longin, 2000, Neftci, 2000, Bali, 2003).

Park (2010), who explore the relationship between downside risk measures and hedge funds' failure.

We test our hypothesis using a sample of publicly-traded firms in the United States obtained from the Compustat database, covering an analysis period from 1990 to 2016. To gauge the statistical significance of respective tail risk measures in predicting likelihood of financial distress, we begin the empirical validation by reporting univariate regression estimates for them. Tail risk measures and their respective average marginal effects (AME) are highly significant in predicting financial distress likelihood.<sup>3</sup> In support of our hypothesis, the univariate regression results establish the positive relationship between firms' distress likelihoods and extreme negative returns. We also test the marginal discriminatory power of tail risk measure in the multivariate setup. First, we develop our one-year financial distress prediction model using panel logistic regression technique and standard financial ratios found significant in recent literature (e.g. Campbell *et al.* 2008, Gupta *et al.* 2018). Next, we supplement respective tail risk measures to this baseline distress model and analyse its significance in the presence of other competing covariates.

Empirical results suggest that VaR estimates over longer duration (five years) are significant in the multivariate setup and are superior to standard deviation as a proxy for risk measure (as AME of standard deviation is much lower than AME of significant VaR measures). ES also exhibit significant explanatory behaviour similar to VaR in predicting financial distress. Longer duration (three and five years) estimates are significant and show AMEs higher than standard deviation, but unlike VaR estimates they show lower values of AMEs. Our results are also robust to financial distress definition of varying intensities. Considering the theoretical superiority of ES over VaR (see Righi and Ceretta (2016) for additional discussion) we suggest

---

<sup>3</sup> In unreported results, we find VaR measures estimated at the 5% significance level show highest values of AME for all respective durations.

using longer duration ES measures as a proxy for risk measure. Finally, we find strong support for our hypothesis that firms that show more frequent extreme negative daily equity returns are more likely to experience the risk of financial distress.

The remainder of the paper is organised as follows: Section 2 defines and provides the rationale behind our proposed definition of financial distress; Section 3 provides definitions of downside risk measures employed in this study; and Section 4 discusses data and financial covariates. Section 5 presents our empirical results; Section 6, 7 and 8 present additional results on robustness checks; and Section 9 concludes this study.

## 2. Defining Financial Distress

Traditionally, most academic studies on bankruptcy prediction gravitated around the choice of explanatory variables and modelling methodologies targeted towards improving prediction models' classification performance. However, these performances depend significantly on how the distress or bankruptcy event is defined in the first place. Unfortunately, this aspect has not received sufficient attention in the bankruptcy literature.

The vast majority of studies employ some legal definition of bankruptcy in line with the relevant bankruptcy code (e.g. Chapter 7/11 filings in the United States), or legal bankruptcy filings supplemented with other related events, such as delisting from stock exchanges due to financial reasons (e.g. Shumway, 2001), or receiving below investment grade credit ratings (e.g. Campbell *et al.* 2008). Plausible reasons behind this additional use of proxy bankruptcy events are attributed to the lack of a standard dataset of bankruptcy events, supplemented with the very small frequency of such events. Since these events significantly threaten firms' degree

of solvency, these may be used to model firms' default risk. However, these definitions of default may suffer from several noteworthy issues.<sup>4</sup>

Firstly, by combining all of these different forms of default, we are trying to predict the probability of a heterogeneous outcome variable, which may be inappropriate. Additionally, since events like bankruptcy resolution involve lengthy legal processes, there often exists a significant time gap between real/economic default date and legal default date (see Theodossiou 1993, Balcaen and Ooghe 2006, Tinoco and Wilson 2013).<sup>5</sup> Moreover, these events are significantly affected by differences in bankruptcy codes between different countries, thus cross comparison of default prediction models between different countries might be inappropriate. Even in cases of uniform codes, instances of judicial discretion cannot be completely ignored where bankruptcy outcomes are often biased toward debtors or creditors (see Gennaioli and Rossi, 2010 for additional discussion). Additionally, differences in creditors' rights also play a significant role in bankruptcy or default outcomes. Davydenko and Franks (2008) report large differences in creditors' rights across France, Germany and the United Kingdom, leading banks to adjust their lending and reorganisation practices in an attempt to avoid costly proceedings of bankruptcy filings. This might lead to false impressions about firms' financial health, which might otherwise have questionable going-concern value. A recent study by Franks and Loranth (2014) confirms that the large majority of bankrupt Hungarian firms in their sample were maintained as going-concerns despite these firms having

---

<sup>4</sup> The issues discussed in the subsequent paragraph are drawn from studies that employ sample of listed or unlisted firms of the United States, France, Germany, Hungary, United Kingdom, etc. Since, we employ a sample of the U.S. listed firms in our study, we do not try to imply that all these issues exist or might exist in the U.S. as well. We merely present an exposition in support of financial distress as an appropriate default definition.

<sup>5</sup> To gain protection from the bankruptcy code, in the U.S., firms usually file for Chapter 11 before they actually default on a significant debt obligation. This gives them the opportunity to renegotiate the outstanding debt obligation or related credit terms with its creditors. However, this may cause significant cost to creditors in the form of administration/recovery cost, subsidised loan, partial loan waiver, etc. Thus, if lenders can identify the timing of economic default or financial distress (which usually precedes legal default), they may undertake pre-emptive measures to minimise subsequent costs arising in the event of any bankruptcy litigation.

suffered large operating losses, primarily due to varying controlling rights of creditors as per the Hungarian bankruptcy code.

However, if we define financial default as debtors' inability to honour the terms of debt contract (specifically failure to make payment when it is due), then this usually precedes events such as legal bankruptcy filing, delisting, or rating downgrade. Frequent occurrence of financial default events might lead to a state of financial distress, where a firm's going-concern aspect is seriously threatened. Once a firm is financially distressed, it may opt for formal reorganisation involving the court system, or an informal reorganisation through the market participants (see John *et al.* 2013 for a comparative discussion). Debt restructuring, asset sale and infusion of new capital from external sources are the three most commonly used market-based or 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 creditors, 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 the high risk involved in financing distressed firms and the 'debt overhang' problem as discussed by Myers (1977). As a consequence, infusion of new capital through informal reorganisation involving market participants is rarely observed in the resolution of financial distress.<sup>6</sup> Thus,

---

<sup>6</sup> However, in case of a formal reorganisation involving the court system, new debt financing may be made available through the debtor-in-possession (DIP) provision intended to encourage new lending to firms in bankruptcy reorganisation. Once a firm files for bankruptcy, the DIP provision allows the financially distressed firm to avail new debt financing usually on a seniority basis (violating any absolute priority rule by placing the new financing ahead of the firm's existing debts for payment).

we cannot rule out the possibility that a financially distressed firm may not file for Chapter 7 or Chapter 11 protection and, instead, choose a private workout method for 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, it should be noted that such out-of-court negotiations are nearly impossible if the firm under formal restructuring has any public debt. Thus, under the binary classification based on legal consequences, a financially distressed firm which has not filed for Chapter 7 or Chapter 11 is not considered to have questionable going-concern value. There is, therefore, a clear need for an *ex-ante* mechanism to identify firms vulnerable to bankruptcy risk in the near future. Waiting until the bankruptcy event might lead to significant erosion in firms' value, losses to creditors and unfavourable business decisions in the years prior to the bankruptcy filing year. It is in the interest of most stakeholders to raise the flag when the symptoms of financial distress are visible, in order to avoid or prepare for the forthcoming crisis. In this regard, lending or credit decisions based on firms' *financial distress* likelihood may be more relevant in addressing respective stakeholders' objectives.

Moreover, from the literature pertaining to bankruptcy costs, we understand that indirect costs of bankruptcy (which are generally in the form of opportunity costs) usually precede direct costs (including legal and liquidation expenses). Thus, leading to declining degrees of solvency much before the actual bankruptcy filing event date. This is in line with the theoretical arguments of Purnanandam (2008), who suggests that financial distress is a state in which firms are unable to meet their debt obligations, ultimately leading them to insolvency as the debt matures. Prominent signs of indirect costs include loss of valuable employees, restrictive terms of trade credit and decline in consumer confidence and investment growth, which, over time, may lead to significant loss in market value even if the firm never files for legal bankruptcy (Chen and Merville, 1999). Thus, a financially distressed firm is left with the challenge of

persuading employees, customers, suppliers and trade creditors to engage in business with it. The closer to bankruptcy a firm is, the stronger this challenge becomes. Stakeholders begin abandoning distressed firms, leading to even faster depletion of operating profits and shareholder wealth. Thus, rather than considering a firm bankrupt when it files for legal bankruptcy, it is appropriate to penalise it for being sufficiently close to bankruptcy/default for their going-concern status to become questionable.

One might argue the economic rationale for imposing this additional cost on firms, which apparently seems to be counterproductive. However, we know that a penalty on rash driving is expected to make drivers more careful, which, in turn, is expected to reduce the number of accidents and, thus, reduce the overall cost associated with such accidents. Similarly, if a firm is penalised for approaching financial distress in the form of higher cost of credit, reduced access to credit or receiving simply a warning bell, the firm may be more mindful towards its deteriorating financial conditions and take proactive measures to avoid potential costly bankruptcy.<sup>7</sup> Penalising firms for approaching financial distress will add additional marginal cost to these firms, but it is expected to reduce cost to other stakeholders (lenders, judiciary, trade creditors, etc.).

Through this, to an extent, we try to imply that the threat of a penalty would decrease the likelihood of firms entering financial distress and eventually bankruptcy. However, we cannot rule out the possibility of firms filing for bankruptcy to gain strategic advantage. Gupta and Barzotto (2018) report evidence of strategic bankruptcy resolution among large corporations in the United States and argue that firms evaluate their likelihood of emerging from Chapter 11 bankruptcy filings before make filing decisions.

---

<sup>7</sup> Through this statement we do not imply that lenders are currently unable to assess credit risk and aren't imposing higher costs of credit to riskier borrowers. We rather suggest that our proposed definition of financial distress will aid them in better assessment and pricing of credit risk.

Considering the discussion above, we believe that financial distress is practically more relevant than legal bankruptcy events and, in the following discussion, we propose a new definition of financial distress that builds on suggestions and findings from earlier related studies. In this study, we define a firm as financially distressed in the year  $t$  if it satisfies the following three conditions in years  $t - 1$  and  $t - 2$ :

**Condition 1:** There is negative growth in the average market value (AMV) of the firm in years  $t - 1$  and  $t - 2$ . Where, AMV is the geometrically declining weighted average of a firm's monthly market values (MV) over the past twelve months computed as follows:

$$AMV_{t-1,t-12} = \frac{1 - \varphi}{1 - \varphi^{12}} (MV_{t-1} + \varphi MV_{t-2} + \dots \dots \dots \varphi^{11} MV_{t-12}) \quad (1)$$

Here,  $\varphi = 2^{-\frac{1}{3}}$ , implying that the weight is halved each month. To avoid losing observations, missing values are replaced with AMV computed using the available number of observations at time  $t$ .

Although decline in firms' market value in two consecutive time periods has been suggested in past literature (e.g. Pindado *et al.* 2008), we suggest considering the average market value rather than the end of the analysis year market value, as the former represents a static snapshot of a firm's market capitalisation at the end of a given trading day, which may be significantly influenced by negative or positive developments around that trading day. Thus, instead of market value on a given day, we use AMV, as it is a better representation of the average state of a firm. Further, by imposing geometrically declining weighted average in Equation 1, we impose higher weight to more recent observations similar to Campbell *et al.* (2008) so that our average is not flawed/biased by rapid growths or declines in firms' monthly market values.

**Condition 2:** the firm's earnings before interest tax depreciation and amortisation (EBITDA) is less than financial expenses (FE) in the years  $t - 1$  and  $t - 2$ .

Traditionally, excessive leverage is believed to be the primary cause of bankruptcy or financial difficulties (e.g. John *et al.* 2013), however significant volumes of literature argue that firms with a higher cost of default choose a lower level of leverage to mitigate potential default likelihood (e.g. George and Hwang, 2010, Glover, 2016), and, thus, we cannot rule out the possibility of financial distress triggered by reasons other than leverage. By including a condition on firms' ability to meet their financial expenses, we are essentially trying to capture those firms that enter into financial difficulties due to their inability to meet financial commitments to the providers of debt capital. Pindado *et al.* (2008) also suggest that financially distressed firms should report financial expenses higher than their earnings in two consecutive years. This effectively implies that the ratio between earnings to financial expenses should be less than one in these two time periods.

**Condition 3:** the firm's operating cash flow (OCF) is less than its financial expenses in the years  $t - 1$  and  $t - 2$ .

If we define financial default as debtors' inability to make payment when it is due, then cash flow is of prime importance. Since the adoption of an accrual accounting system for recording and reporting business transactions, balance sheets and income statements have emerged as dominant sources of information for business decision-making. However, accruals are the non-cash component of earnings and represent adjustments made to cash flows to generate a profit measure largely unaffected by the timing of receipts and payments of cash (Ball *et al.* 2016). Thus, just having earnings above financial expenses ( $EBITDA > IE$  or  $EBITDA/IE > 1$ ) does not guarantee a firm's ability to repay its debt on time. Inefficient management of cash may also result in premature closure of even profitable businesses, as

loans, most operating expenses, taxes etc. are required to be paid in cash. However, this important cash flow information is often overlooked in the academic literature. For instance, Pindado *et al.* (2008) and Keasey *et al.* (2015) did not consider the role of cash flow information while defining financial distress.

Ball *et al.* (2016) argue that cash-based measures of operating profitability (a measure that excludes accruals) outperform measures of profitability that include accruals. Additionally, Grullon *et al.* (2017) report that, for large firms, cash flow remains the primary driver of investment spending and the marginal effect of cash flow on aggregate investment has actually increased over the past 30 years. Recent literature also highlights the importance of financing constraints and free-cash-flow problems in making sound investment decisions (Lewellen and Lewellen, 2016). The relative importance of cash flow in times of financial distress is also realised in recent literature. Lee *et al.* (2016) argue that, in times of financial distress, investors place significantly higher weight on OCF information than on earnings information, as they find that a firm's stock returns show stronger association with its OCF than its earnings when the firm is in financial distress. Considering the discussion above, and to be consistent with the two previous conditions, we include *Condition 3* in our proposed definition of financial distress, with an aim to overcome some of the limitations of firms' failure definitions discussed earlier.

### 3. Estimating Downside Risk Measures

We use daily three-month, six-month, one-year, three-year, five-year and monthly sixty-month returns to measure VaR and ES of all firms in our sample. We estimate the risk measure of each firm as of the month-end (of the latest available daily/monthly return data) in a given year to predict firms' financial distress in the next year. For instance, if the date on which financial statements were filed is June 2015, we calculate its risk measures as of May 2015 to predict a firm's financial distress likelihood in the next one year.

**Value-at-risk:** Estimation of VaR requires the definition of two basic parameters, the confidence level ( $1 - \alpha$ ) and the time horizon ( $\tau$ ), along with the estimation model. These are then used to estimate the worst loss that may occur over the specified time horizon and confidence level.

Traditionally, the risk horizon is measured in the number of trading days,<sup>8</sup> rather than the number of calendar days for measuring VaR, and it is this period over which we measure the expected loss. In the absence of any defined rules or guidelines, time horizon is usually chosen by considering the liquidity of the risk and the duration of exposure to that risk (Alexander, 2008). The higher the assets' liquidity, the lower the time required to close or fully hedge the exposure. Thus, the time horizon required is shorter for the VaR model. A longer time horizon is suggested when measuring VaR in stressful and volatile market conditions, as markets tend to lose liquidity during such economic conditions. Credit rating agencies assign top ratings to those companies that exhibit extremely low default probability (such as 0.03%) over the next one-year horizon. Hence, companies looking for AA or above credit rating should apply a confidence level of 99.97% to measure their enterprise-wide VaR over the next one-year horizon. Considering the arguments presented above, and the nature of our application, we estimate downside risk measures using five different time horizons to detect the presence of any intertemporal differences. We estimate the risk measure of each firm as of the year-end using the past three-month, six-month, twelve-month, three-years and five-years daily returns, and past five-year monthly returns to predict firms' financial distress the following year.

The choice of confidence level is also highly subjective; it primarily depends on the area of application and users' attitude towards risk. For instance, the Basel II Accord requires banks using internal VaR models to assess their capital requirements to measure VaR at a 99%

---

<sup>8</sup> In this study, our risk estimates are computed using 252 trading days in a year.

confidence level, while a credit rating agency may maintain a more stringent confidence level of about 99.97% to assess the probability of firms' insolvency hazard (Jorion, 2000). On the other hand, a risk manager may allow a lower confidence level (95%) and shorter risk horizon (1 day) while setting VaR based trading limits. Thus, we follow the most popular choice recommended by the Basel Accord, and consider probability levels of 99% ( $\alpha = 0.01$ ) to estimate the downside risk measures. However, we also perform robustness check using 95% ( $\alpha = 0.05$ ) confidence level estimates and report the results in Section 8.

Let  $\tau$  denote the time horizon,  $R_{t+\tau}$  denote a firm's return between the time period  $t$  and  $t + \tau$ , and let  $F_{R,t}$  represent the cumulative distribution function (CDF) of  $R_{t+\tau}$  conditional upon the set of information available at time  $t$ . Then,  $F_{R,t}^{-1}$  represents the inverse function of  $F_{R,t}$ . Given this, the VaR of a firm's return as of time  $t$  with a time horizon  $\tau$  and at  $(1 - \alpha)$  confidence level can be estimated as follows:

$$VaR_t(\alpha, \tau) = -F_{R,t}^{-1}(\alpha) \quad (2)$$

Unlike normal VaR, the semi-parametric Cornish-Fisher VaR ( $VaR_{CF}$ ) considers higher moments in the return distribution, thus relaxing the normality assumption. The Cornish and Fisher (1938) approximation incorporates higher moments (skewness and kurtosis) into the estimation process and, thus, accounts for non-normality in the return distribution. Zangari (1996) is the first to introduce this in his study that estimates the VaR of option portfolios. The justification behind the use of this formula is that any distribution can be approximated with these known moments, in terms of any other given distribution (Johnson *et al.* 1994, Jaschke, 2002). Thus, we use  $VaR_{CF}$ <sup>9</sup> to accommodate non-symmetrical and fat-tailed returns

---

<sup>9</sup> We use VaR and  $VaR_{CF}$  interchangeably in this manuscript.

distribution. The fourth order Cornish and Fisher (1938) expansion for  $\alpha$  percentile of  $(R - \mu)/\sigma$  is shown in equation (2), while equation (3) defines the Cornish-Fisher VaR ( $VaR_{CF}$ ).

$$\Omega(\alpha) = Z(\alpha) + \frac{1}{6}(Z(\alpha)^2 - 1)S + \frac{1}{24}(Z(\alpha)^3 - 3Z(\alpha))K - \frac{1}{36}(2Z(\alpha)^3 - 5Z(\alpha))S^2 \quad (3)$$

$$VaR_{CF} = -(\mu + \Omega(\alpha) \times \sigma) \quad (4)$$

where  $\sigma$  is the standard deviation,  $\mu$  is the average return,  $S$  is the measure of skewness,  $K$  is the excess kurtosis of past  $n$ -month daily returns,  $(1 - \alpha)$  is the confidence level, and  $Z(\alpha)$  is the critical value obtained from the standardised normal distribution.<sup>10</sup>

**Expected Shortfall:** Although VaR is the maximum loss that can take place over a defined time horizon and confidence level, there still exists a very small probability that a loss greater than VaR may be experienced. Thus, VaR does not reveal any information about the size of the loss in the event when that level is breached. Expected Shortfall (ES) provides this information. It is the conditional expected loss that is greater than or equal to the VaR. ES expressed in terms of return rather than dollar amount is formulated as follows:

$$\begin{aligned} ES_t(\alpha, \tau) &= -E_t[R_{t+\tau} | R_{t+\tau} \leq -VaR_t(\alpha, \tau)] \\ &= -\frac{\int_{v=-\infty}^{-VaR_t(\alpha, t)} v f_{R,t}(v) dv}{F_{R,t}[-VaR_t(\alpha, \tau)]} \\ &= -\frac{\int_{v=-\infty}^{-VaR_t(\alpha, t)} v f_{R,t}(v) dv}{\alpha} \end{aligned} \quad (5)$$

Here,  $R_{t+\tau}$  represents a firm's return in the time period  $t$ , and  $t + \tau$  and  $f_{R,t}$  denotes the conditional probability density function (PDF) of  $R_{t+\tau}$ .  $F_{R,t}$  is the conditional CDF of  $R_{t+\tau}$

---

<sup>10</sup>Note that standard deviation is always positive, while the original VaR and ES are usually negative. To avoid confusion, the original VaR and ES numbers are multiplied by  $-1$  in equations (3) and (4). Therefore, the VaR and ES numbers presented in this paper are usually positive.

conditional upon the information set available at time  $t$ ,  $F_{R,t}^{-1}$  represents the inverse function of  $F_{R,t}$ , and  $(1 - \alpha)$  is the confidence level. Artzner *et al.* (1999) provide the theoretical argument that VaR is inferior to ES as a risk measure. They argue that, unlike VaR, ES possesses mathematical properties like continuity and subadditivity which are desirable to a coherent measure of risk. We use Cornish-Fisher VaR ( $VaR_{CF}$ ) calculated using equations (3) and (4) to estimate Expected Shortfall, denoted as  $ES_{CF}$ .<sup>11</sup>

## 4. Data and Covariates

This section presents discussion on our sources of data and choice of covariates used in this study to perform required empirical analysis.

### 4.1 Data

Unlike several others (e.g. Chava and Jarrow, 2004), we do not try to predict the financial distress event over the next month or quarter. We believe that an efficient distress prediction model should work as an Early Warning System (Edison, 2003) and should signal any adverse distress event well in advance. This may provide sufficient time to avert or manage the forthcoming crisis. Thus, to predict the financial distress over the next one-year horizon, our empirical analysis employs annual firm-level accounting and market data of the United States firms from the Compustat database. We consider a relatively long analysis period from 1990 to 2016. We also limit our sample to firms which are publicly-traded either on AMEX, NASDAQ or NYSE stock exchanges only. Transportation, Communications & Public Utilities, Finance, Insurance & Real Estate, and Public Administration firms have been excluded from our analysis. The seven industrial sectors included in this study are: 1. Agriculture, Forestry, Fishing (SIC Code < 1000); 2. Mining (SIC Code 1000 to < 1500); 3. Construction (SIC Code

---

<sup>11</sup> We use ES and  $ES_{CF}$  interchangeably in this manuscript.

1500 to < 1800); 4. Manufacturing (SIC Code 2000 to < 4000); 5. Wholesale Trade (SIC Code 5000 to < 5200); 6. Retail Trade (SIC Code 5200 to < 6000); and 7. Services (SIC Code 7000 to < 8900). Where SIC (Standard Industrial Classification) Code is a four-digit code that represents a given industrial sector of the United States firms.

## 4.2 Covariates

**Dependent Variable:** We use the definition of financial distress proposed in Section 2 as the binary dependent variable for this study.

**Independent Variables:** A large volume of empirical literature has evolved over time which efficiently addresses the issues pertaining to prediction of corporate bankruptcy. However, the default prediction models gravitate toward the use of accounting variables (e.g. Altman, 1968, Ohlson, 1980) or market information to predict firms' insolvency hazard (e.g. Bharath and Shumway, 2008). The market-based approach of modelling default risk has become popular since their commercial introduction by KMV, CreditMetrics and others. Hillegeist *et al.* (2004) report that market-based structural models perform better than popular accounting-based models, in particular the Altman Z-score (Altman, 1968) and Ohlson O-score (Ohlson, 1980) models.

In a similar study, Das *et al.* (2009) report that their distress prediction models, built using accounting metrics and market metrics, respectively, exhibit very similar classification performances. Models which employ both accounting and market information substantially outperform both accounting and market-based models, suggesting that accounting and market information are complementary to each other. Findings of other related studies (see, among others, Agarwal and Taffler, 2008; Trujillo-Ponce *et al.* 2014) also support the findings of Das *et al.* (2009). Considering the above arguments, we use accounting and market information to develop our distress prediction models.

In particular, we employ all eight (three accounting-based and five market-based) measures of bankruptcy employed by Campbell *et al.* (2008) to predict the bankruptcy of US firms, as they report that this set of covariates outperforms other standard sets of covariates employed in prior studies. We supplement this with two additional covariates, financial expenses to sales (FES) and tax-to-market valued total assets (TMTA), as they exhibit significant explanatory power and very high average marginal effects in several recent studies on SMEs financial distress (e.g. Gupta *et al.* 2018). Furthermore, we also test the marginal predictive ability of tail risk measures discussed in Section 3, i.e. VaR and ES. Detailed definitions of financial covariates are as follows (see Appendix A1 for complete list of covariates along with their respective Compustat data items):

- I. Total liabilities to Market-valued Total Assets (TLMTA), calculated as:

$$TLMTA_{i,t} = \frac{Total\ Liabilities_{i,t}}{(Market\ Value\ of\ Equity_{i,t} + Total\ Liabilities_{i,t})}$$

Here a Market-valued Total Asset (MTA) is the sum of the market value of a firm's equity and its total liabilities. In case of missing values of Total Liabilities, MTA is equal to the firm's equity, and vice-versa.

- II. Cash and short-term assets to Market-valued Total Assets (CMTA), calculated as:

$$CMTA_{i,t} = \frac{Cash\ and\ Short - Term\ Investments_{i,t}}{(Market\ Value\ of\ Equity_{i,t} + Total\ Liabilities_{i,t})}$$

- III. Weighted average of Net Income to Market-valued Total Assets (*NIMTAAVG*) over the previous three-year period (*NIMTAAVG*), calculated as:

$$NIMTAAVG_{t-1,t-3} = \frac{1}{1.75} NIMTA_{t-1} + \frac{0.5}{1.75} NIMTA_{t-2} + \frac{0.25}{1.75} NIMTA_{t-3}$$

where

$$NIMTA_{i,t} = \frac{Net\ Income_{i,t}}{(Market\ Value\ of\ Equity_{i,t} + Total\ Liabilities_{i,t})}$$

Following Chen and Hill (2013), we modify Campbell *et al.*'s (2008) measure of average *NIMTA* to allow for the fact that we employ annual, rather than quarterly, accounting measures. Missing values of *ANIMTA* are replaced with the cross-sectional mean of *NIMTA*.

- IV. Weighted average of monthly log excess return relative to value-weighted S&P 500 return over the previous 12-month period (*EXRETAVG*), calculated as:

$$EXRETAVG_{t-1,t-12} = \frac{1 - \phi}{1 - \phi^{12}} (MEXRET_{t-1} + \dots + \phi^{11} MEXRET_{t-12})$$

Where;

$$MEXRET_{i,t} = \log(1 + Return_{i,t}) - \log(1 + Return_{S\&P\ 500,t})$$

Missing values of *AMEXRET* are replaced with the cross-sectional mean of *MEXRET*.

- V. Annualised standard deviation of a firm's daily log returns over the previous three months (*SIGMA*), estimated as:

$$SIGMA_{i,m-1,m-3} = \left( 252 \times \frac{1}{N-1} \sum_{k \in \{m-1, m-2, m-3\}} r_{i,k}^2 \right)^{\frac{1}{2}},$$

where  $r_{i,k}^2$  represents the log returns of firm *i* on the day *k*, and *N* represents the number of days in the three-month estimation window. Here, following Campbell *et al.* (2008), we use a proxy centred around zero rather than a three months' rolling mean.

- VI. Firm's Market-to-book ratio (*MB*)
- VII. Log of price per share winsorized above \$15 (*PRICE*)
- VIII. Logarithm of each firm's size relative to S&P 500 market capitalisation (*RSIZE*), calculated as:

$$RSIZE_{i,t} = \log \left( \frac{Firm\ Market\ Equity_{i,t}}{Total\ S\&P\ Market\ Value_t} \right)$$

- IX. Financial Expense to Sale (*FES*), calculated as:

$$FES_{i,t} = \frac{\text{Financial Expense}_{i,t}}{(\text{Sale}_{i,t})}$$

X. Tax to Market-Valued Total Assets (TMTA), calculated as:

$$TMTA_{i,t} = \frac{\text{Total Income Tax}_{i,t}}{(\text{Market Value of Equity}_{i,t} + \text{Total Liabilities}_{i,t})}$$

Healthy firms are expected to pay higher taxes due to higher business volume in comparison to distressed firms and, hence, a negative relationship is expected between TMTA and financial distress likelihood. Further, following the suggestion of Cohen *et al.* (2003), we adjust the book value of equity to avoid the influence of outliers. Specifically, we add 10% of the difference between the market value and the book value of equity to the book value of equity,<sup>12</sup> thus increasing book values that are relatively very small and probably wrongly measured. This ultimately prevents outliers when book equity is used in the denominator when calculating financial ratios. For observations which still exhibit negative book value of equity after this adjustment, we replace negative book values with positive \$1 to ensure that market-to-book value ratios for all observations are positive. Further, we restrict the range of all the covariates within 5<sup>th</sup> and 95<sup>th</sup> percentiles to limit the influence of any outliers.

**Control Variables:** To control the volatility in the macroeconomic environment affecting specific industrial sectors, we calculate a measure of industry risk (*INDRISK*) as the financial distress rate (number of firms experiencing the event of interest in the 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 indicate a higher risk of financial distress, and vice versa.

---

<sup>12</sup>  $\text{Total Assets}(\text{adjusted})_{i,t} = \text{Total Assets}_{i,t} + 0.1(\text{Market Equity}_{i,t} - \text{Book Equity}_{i,t})$

## 5. Empirical Methods

### 5.1 Panel Logistic Regression

Although we see a significant rise in the popularity of hazard models in modelling bankruptcy or financial distress events, in line with the findings of Gupta *et al.* (2018) we use panel logistic regression with random effects to establish our empirical validation. Gupta *et al.* (2018) argue that the discrete-time hazard model with logit link is essentially a panel logistic model that controls for firms' age. Thus, we assume that marginal probability of firms' financial distress over the next year follows a logistic distribution that is estimated as follows:

$$P(Y_{it} = 1) = \frac{1}{1 + \exp(-\alpha - \beta X_{i,t-1})} \quad (7)$$

where  $Y_{it}$  is an indicator variable that equals one if the firm is in financial distress in year  $t$ , and  $X_{i,t-1}$  is a vector of explanatory variables known at the end of the previous year. To capture any duration dependency, we use the natural logarithm of firms' annual age<sup>13</sup> (AGE) as a control variable in our multivariate models. Additionally, since our panel regression analyses use lagged covariates ( $t - 1$ ), any potential endogeneity problem is expected to be mitigated, as lagged explanatory variables and dependent variables have a low likelihood of being jointly determined (Berger and Bouwman, 2013).

However, Reed (2015) raise concern on this widespread practice of dealing with potential simultaneity, and argue that associated estimates may still remain inconsistent and hypothesis testing may be invalid. Rather they suggest an alternative, is to use lagged values of the endogenous variable in instrumental variable estimation. However, to the best of our knowledge, programming codes required to perform such statistical estimations are currently available only for linear regression specifications. Since we use non-linear panel logistic

---

<sup>13</sup> We proxy a firm's age as the earliest year for which financial information for that firm is available in the Compustat database.

specification, we are at a disadvantage to act upon this valuable suggestion. But, to be mindful, we also report regression estimates of our main model using  $t - 2$  and  $t - 3$  lagged covariates. It is unlikely that the dependent and independent variables are simultaneously determined two or three years ahead in advance. Thus, significance of  $t - 2$  and  $t - 3$  lagged covariates shall imply absence of any severe endogeneity bias in our regression estimates arising due to reverse causality.

## 5.2 Summary Statistics and Correlations

To get an initial understanding about the variability of our covariates, we report descriptive statistics for financially distressed and non-distressed/healthy groups of firms. Table 1 provides an overview of mean, median, standard deviation, minimum and maximum of all covariates used in our study. Column 1 presents the list of covariates, while column 2 lists names of descriptive measures. Descriptive measures of the vast majority of covariates are as per our expectation and without any extreme variability.

The mean/median of covariates we report are very similar to those reported by Campbell *et al.* (2008) in their study, with some differences due to the data period and different definition of the default event. We expect the mean/median of the covariates that are positively related to the insolvency/distress risk to be higher for the distressed group than for the group without distress (see, for example, variable FES in Table 1). Contrarily, we expect the mean/median of the covariates that are negatively related to the insolvency/distress risk to be lower for the distressed groups than for the group without distress (see, for example, variable NIMTAAVG in Table 1). However, we also see some exceptions. The mean value of TLMTA for the distressed group is lower than for the non-distressed group of firms, implying that financially distressed firms have lower total liabilities. This is contrary to our traditional understanding, where we expect firms in financial distress to have higher liabilities compared to their total

assets. Similarly, the mean value of CMTA for the distressed group is higher than for the non-distressed group, implying that financially distressed firms have higher cash reserves. This is contrary to intuition, where we expect financially distressed firms to have liquidity problems and, hence, lower reserves of cash and short-term investments. Also, the mean market-to-book (MB) ratio, which is expected to be lower for the distressed group of firms than its non-distressed counterpart, is actually higher. This may suggest active demand of distressed stocks, which consequently pushes their stock price higher. These contrary revelations might affect the sign of respective regression coefficients.

**[Insert Table 1 Here]**

A comparison of the distressed and non-distressed groups also reveals expected differences in Table 1. For example, distressed firms make mean loss of about 18%, compared with non-distressed firms that make mean loss of only about 1.9%, as lower mean of net income relative to market value of total assets shown by NIMTAAVG. The mean volatility of distressed firms is quite high at 80%, relative to only 55% of non-distressed firms, as shown by SIGMA. Distressed firms are relatively small (RSIZE) and have a much lower mean log price per share (PRICE), and lower returns (EXRETAVG) compared to their non-distressed counterpart. Similar to Campbell *et al.* (2008), the mean of market-to-book (MB) ratio of financially distressed firms is slightly higher, because the book value of their equity erodes due to frequent losses. This does not seem to be reflected in the market value of the firms because of the proactive nature of stock markets.

A closer look at extreme measures generated using three-month, six-month, one-year, three-year and five-year daily returns, and sixty-month monthly returns (i.e. VaR and ES) reveal that distressed firms have higher values compared to non-distressed firms. This is to be expected, as distressed firms tend to exhibit extreme movements. VaR is the loss that will not

exceed from that level with a certain confidence interval during a certain period of time. The higher magnitude reflects the higher probability of loss. For example, in Table 1 the mean VAR3M1 of financially distressed firms is 13%. It means that we are 99% confident that the mean loss from financial distressed firms will not exceed 13% during a period of 3 months.

On the other hand, expected shortfall is the expected loss conditional upon that the loss is greater than a certain threshold. For example, in Table 1 the mean ES3M1 of financially distressed firms is 21.03%. It means that the mean loss of financially distressed firms will be 21.03% during a period of 3 months with the assumption that the loss is greater than 99% of the loss distribution. Furthermore, we also find intertemporal differences, especially at the 99% confidence interval. We see the mean/median of extreme measures at shorter rolling windows are less than those at longer rolling windows.

Panel A of Appendix A2 reports correlation among main (non-tail risk) covariates. As reported, most covariates show low to moderate correlation among each other. However, PRICE shows strong positive correlation with RSIZE (0.7168) and strong negative correlation with SIGMA (-0.6121). Thus, we take into account the issues of multicollinearity while developing our multivariate models excluding tail risk measures. Subsequently, Panel B of Appendix A2 reports correlation of tail risk measures with main variables. As reported, most tail risk measures exhibit moderate to low correlation with main variables, except SIGMA. This is expected, as both of these are competing measures of firms' stock price volatility. However, we address this issue effectively while developing multivariate models (see Section 5.4).

### 5.3 Univariate Regression and Average Marginal Effects

Table 2 reports the univariate regression results of respective covariates. This table reports point estimates ( $\beta$ ), standard errors (SE), z-statistics (Z) and average marginal effects (AME)

in percentages.<sup>14</sup> All financial ratios are highly significant at the 1% level of significance. MB negatively predicts the financial distress of firms, meaning that the stock market is already anticipating the financial distress, driving down the market value of equity and hence the market-to-book ratio. As explained above in the summary statistics, we find another counterintuitive result as CMTA positively predicts the likelihood of financial distress.

Eight out of twelve covariates show AME of less than 5% with AME of TLMTA, MB, RSIZE and AGE even less than 1%. However, the other four covariates reported are highly significant predictors, with more than 5% AME. Two non-Campbell covariates, FES and TMTA, exhibit AME of around 5% and 30%, respectively. Among extreme risk measures, all are highly significant at the 1% level of significance. This implies that all extreme risk measures can predict financial distress, but some of these measures have very low values of AME, which raises scepticism about their predictive power in the multivariate setup. We see intertemporal differences as longer horizon rolling estimates show more statistical and economic significance. For example, sixty-month VaR and sixty-month ES have the highest AME of 1.87% and 1.66%, respectively.

**[Insert Table 2 Here]**

## 5.4 Baseline Multivariate Model

In order to test the marginal discriminatory power of downside risk measures, we first estimate the baseline multivariate model with all financial covariates, as they are all significant in the univariate analysis. However, considering that few cases of moderate to high correlation among

---

<sup>14</sup> In non-linear regression analysis, Marginal Effects are a useful way to examine the effect of changes in a given variable on changes in the outcome variable, holding other variables constant. These can be computed as marginal change (it is the partial derivative for continuous predictors) when a variable change by an infinitely small quantity and discrete change (for factor variables) when a variable change by a fixed quantity. The Average Marginal Effects (AME) of a given variable 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 value of a given variable, provided other covariates are held constant. See Long and Freese (2014) for additional discussion on this topic.

covariates are reported in Table 3, we follow the method suggested by Gupta *et al.* (2018) while developing our baseline multivariate model. Primarily, this requires ranking of the covariates in univariate regression based on the magnitude of their respective AME and, then, introducing each variable at a time in the multivariate setup, in ascending order of their rank. Gupta *et al.* (2018) argue that the higher the value of AME, the higher the change in the predicted probability due to unit changes in the covariate's value. Thus, to develop a parsimonious model, our priority should be to include minimum number of covariates with highest explanatory power/AME. Also, we exclude a covariate from the multivariate model if, when introduced: (i) it reverses the expected sign of any previously added covariate; (ii) it bears the opposite sign to that expected; (iii) it bears the expected sign, but has a p-value greater than 0.10; and (iv) it makes a previously added covariate insignificant with a p-value greater than 0.10.

This allows us to develop a parsimonious baseline multivariate model that excludes PRICE, RSIZE, TLMTA and MB.<sup>15</sup> The remaining covariates exhibit low to moderate correlation with each other with maximum magnitude of around 0.4. Thus, any severe biasness in our regression results due to multicollinearity among covariates is highly unlikely. We also exclude SIGMA from our baseline model, as the purpose of this study is to report the comparative performance of several tail risk measures in comparison to SIGMA, which we do in the subsequent section. Finally, we include INDRISK and natural logarithm of firms' annual age (AGE) as control variables in the model. As reported in column 2 of Table 3, we develop the final parsimonious baseline model with seven covariates, all of which are highly significant and retain their respective signs, as reported in Table 2. We therefore conclude that, all of these covariates are jointly significant in predicting the financial distress of US firms. Also, as reported in Appendix A3, the within-sample and out-of-sample area under ROC curves

---

<sup>15</sup> PRICE makes EXRETAVG insignificant, hence removed; RSIZE makes EXRETAVG insignificant, hence removed; TLMTA makes EXRETAVG insignificant, hence removed; and the coefficient of MB is positive, thus removed.

(AUROC<sup>16</sup>) of this model is about 92%, suggesting excellent performance of our baseline model in classifying financially distressed firms.

Columns 3 and 4 of Table 3 report regression estimates for  $t - 2$  and  $t - 3$  lagged covariates of the main multivariate model reported in column 2 of Table 3. Except for the insignificance of EXRETAVG in  $t - 3$ , all financial covariates are highly significant in explaining a firm's likelihood of experiencing financial distress in time  $t$ . This suggest strong intertemporal explanatory power of our model, with excellent classification performance of around 90% (see Appendix A3). This also implies, absence of any severe endogeneity bias in our regression estimates due to reverse causality, as it is unlikely that the dependent and independent variables are simultaneously determined two or three years ahead in advance.

**[Insert Table 3 Here]**

## 5.5 Multivariate Models with Risk Measures

In order to test the marginal discriminatory power of tail risk measures in comparison to SIGMA, we re-estimate the baseline multivariate model with  $t - 1$  lagged covariates (reported in column 2 of Table 3), supplemented with tail risk measures and report the results in Tables 4 and 5. Rows 3 to 7 in Tables 4 and 5 report regression estimates for respective risk measures as listed in the first row of columns 2 to 8.  $LR\ Chi^2$ , in the seventh row of column (1), is the chi-square for likelihood ratio test that compares the log likelihood of the baseline model supplemented with risk variables with the log likelihood of the baseline model, and tests whether this difference is statistically significant. We estimate separate multivariate models for daily three-month, six-month, one-year, three-year and five-year, and monthly sixty-month tail

---

<sup>16</sup> The higher the AUROC, the better the prediction model. Please see Gupta *et al.* (2018) for additional discussion on this. For out-of-sample validation, we first estimate models using observations until the year 2011, and, using these estimates, we predict financial distress probabilities for the year 2012; we then include 2012 in the estimation sample and predict distress probabilities for 2013 and so on until the year 2016. We then use these predicted probabilities from the year 2012 through 2016 to estimate out-of-sample AUROC.

risk estimates to account for any intertemporal differences that may exist. Following Liang and Park (2010), we compare these models with our baseline multivariate model, supplemented with SIGMA as the risk measure.

We estimate all multivariate models using our proposed definition of financial distress as defined in Section 2. Table 4 reports our multivariate regression results with VaR measures estimated at 1% significance level. The first column gives the names of the covariates. Column (2) reports estimates with SIGMA as the risk measure. Columns (3) to (8) report respective regression estimates employing tail risk measured at the 99% confidence interval.

Table 4 reveals intertemporal differences, as short-duration rolling estimates of VaR measures are insignificant in predicting firms' likelihood of experiencing financial distress. As we measure the extreme downside risk, it might be that short-duration rolling estimates do not have enough observations (Huang *et al.* 2012);<sup>17</sup> therefore, long-duration rolling estimates perform better in predicting the financial distress of firms. The baseline model with SIGMA as a measure of risk is also statistically significant with *LR Chi*<sup>2</sup> of 7.50; however, the AME is very low at 0.45%. Other significant downside risk measures are five-year and sixty-month VaR with higher values of AMEs than SIGMA.<sup>18</sup>

Subsequently, we repeat this multivariate analysis with ES as the risk measure and report the results in Table 5 at 1% significance level. Similar to multivariate regression results with VaR, we find intertemporal differences in that longer-duration rolling estimates of ES perform better in predicting financial distress of firms. However, with ES we find that five-year ES performs better with the *LR Chi*<sup>2</sup> of 23.61. Three-year ES measure also significantly predicts

---

<sup>17</sup> While measuring ES and TR, Bali *et al.* (2009) also use a longer rolling window of 100 days compared to shorter windows for measuring VaR.

<sup>18</sup> Three-month VAR measured at the 99% level of significance are also marginally significance with the reverse sign.

the financial distress likelihood of firms but with slightly lower AME than five-years ES. Our baseline model with SIGMA as a measure of risk also significantly explains the financial distress of firms, but shows lower values of AME and *LR Chi*<sup>2</sup> than ES3Y1, ES5Y1 and ES60M1. Overall, as the superiority is ES over VaR is well-documented in the literature, we conclude that five-year and sixty-month ES measured at the 99% confidence interval (ES5Y1) are superior to SIGMA (higher AME than SIGMA) and other competing tail risk measures.

**[Insert Tables 4 and 5 Here]**

## 6. Multivariate Models with Thresholds

Conditions 1, 2 and 3 of our definition of financial distress presented in Section 2 emphasise only on the decline in respective values in the past two consecutive years. One might argue that the decline should be significant enough to threaten firms' degree of solvency. Thus, imposing a threshold (a decline of X% or higher) might appear reasonable. Addressing this concern relating to model calibration or stress-testing in our existing framework is relatively straightforward. However, imposing such threshold will alter the distress status of a firm and, thus, needs to be based upon the risk appetite of the decision-maker. Imposing 0% threshold will imply most conservative behaviour or lowest risk appetite, while higher values will imply higher risk appetites. For a given threshold, this will also lead to decline in number of distressed firms in the sample, which will keep decreasing with increasing threshold values.

To investigate the effect of imposing such threshold on our distress definition, and its subsequent impact on the performance of the main distress prediction model (presented in column 2 of Table 3), we impose further restrictions on our proposed definition of financial distress and report the results in Table 6. In particular, we report results obtained from arbitrarily imposing 10%, 20% and 30% threshold levels respectively on conditions 1, 2 and 3.

For instance, a 10% threshold implies - *Condition 1*: There is negative growth in the AMV of the firm in years  $t - 1$  and  $t - 2$ , and this decline in AMV is at least 10% or higher in both these respective years; *Condition 2*: the firm's EBITDA is less than financial expenses in the years  $t - 1$  and  $t - 2$ , and this shortfall in earnings is at least 10% or higher in both these respective years; and *Condition 3*: the firm's OCF is less than its financial expenses in the years  $t - 1$  and  $t - 2$ , and this shortfall in OCF is at least 10% or higher in both these respective years. Subsequently we do the same for 20% and 30% threshold levels respectively. Although our choice of threshold levels is arbitrary, one might choose any threshold level subjected to their risk appetite, or any required combination of different thresholds in the same definition of financial distress.

In order to compare the baseline multivariate model with the models with different thresholds, we first report the baseline model of Table 3 in column 2 of Table 6. Columns 3 to 5 of Table 6 report multivariate regression estimates for models with 10%, 20% and 30% threshold levels respectively. As we see in Table 6, all five financial covariates are highly significant and retain their respective expected signs at different threshold levels. This shows consistency of these covariates in identifying different intensities of financial distress. However, the AME of respective covariates is highest for the model without any threshold and declines gradually with increasing threshold levels (see Table 6). This implies that the explanatory power of covariates decreases with the increase in the intensity of financial distress. This appears reasonable as firms experiencing higher intensity of financial distress makes the distress group of firms appear more homogeneous than their healthy counterparts. This is also supported from the fact that the number of financially distressed firms in our sample decreases with increasing threshold levels. From 1,219, it falls to - 696 with the imposition of 10% threshold; 504 with the imposition of 20% threshold; and 350 with the imposition of 30% threshold. However, the  $R^2$  of these models stays around 32%, implying no loss in overall

explanatory power of these models with increasing threshold levels. Additionally, we also see in Table 6 that the log-likelihood values of respective models increase with increasing threshold levels. This implies that the sample quality improves with increasing threshold levels.

Overall, we find our proposed definition of financial distress is robust to varying intensities of distress, with positive association between the sample quality and threshold levels.

**[Insert Table 6 Here]**

## 7. Multivariate Models with Thresholds and Risk Measures

In order to test whether our finding relating to the statistical significance of tail risk measures in explaining financial distress are robust to financial distress definitions of varying intensities, we re-estimate the multivariate models reported in Table 6 supplementing risk measures listed in Appendix A1 (see Tables 7 and 8)<sup>19</sup>.

Similar to what we see in Tables 4, we find that short-duration rolling estimates of VaR measures are insignificant in predicting firms' likelihood of experiencing financial distress, whereas long-duration rolling estimates perform significantly better (see Tables 7 and 8). Similar to our baseline results in Table 4, we find that multivariate models with sixty-month VaR (VAR60M1) is the best at all respective threshold levels of 10%, 20% and 30%, with the highest value of AME, and the highest value of  $LR\ Chi^2$ . Models with SIGMA as a measure of risk are also statistically significant at all threshold levels but with lower AMEs. Our results are also robust for five-year VaR estimates at the 99% confidence level (VAR5Y1) but only at 10% threshold level. Overall, these results at respective threshold levels are broadly identical to what we report in Table 4, with the only difference being insignificance of VAR5Y1 at 20%

---

<sup>19</sup> Please note that, to save space we report only risk variables and models' goodness of fit measures in these table. The statistical significance of unreported covariates remain identical to the one reported in Table 9.

and 30% threshold levels. Additionally, as reported in the earlier section, here also we see decline in the AME of risk measure with increasing threshold levels.

We repeat the above multivariate analysis with ES as the risk measure at three selected threshold levels of 10%, 20% and 30% respectively. We report our results in Table 8. At all threshold levels, our results are robust and are very similar to what we report in Table 5 without any threshold; with three-year, five-year and sixty-month ES estimates being significant predictor of financial distress. These results are also similar to multivariate regression results with VaR, that, long-duration rolling estimates of ES perform better in predicting financial distress than its short-duration counterparts. Our baseline model with SIGMA as a measure of risk also significantly explain financial distress of firms at all threshold levels but show lower AME values. Additionally, as reported in the earlier section and similar to VaR, here also we see decline in the AME of respective ES measures with increasing threshold levels.

Overall, our robustness checks substantiate our earlier conclusion, and show that our proposed definition of financial distress is consistent and robust. Because of the theoretical superiority of ES over VaR, we suggest using five-year or sixty-month ES estimates in predicting financial distress.

**[Insert Tables 7 and 8 Here]**

## 8. Multivariate Models with 5% Tail Risk Estimates

As highlighted in Section 3 that the choice of confidence level in tail risk estimations is also highly subjective, we use less stringent confidence level of 5% to estimate VaR and ES, and report the results in Table 9. We only report the results only with longer time horizons as the results for both VaR and ES are significant only at longer time horizons. Columns (3) to (5) reveal that the multivariate models with three-year, five-year and sixty-month VaR estimated

at the 95% confidence interval perform better than that were measured at the 99% confidence interval. Columns (6) to (8) show results with ES, and only sixty-month ES significantly predicts financial distress likelihood of firms. Since ES is the mean of losses greater than VaR, the mean of the losses is lower at the 95% confidence interval because lower values reduce the mean of the losses. The mean is higher at the 99% confidence interval and, therefore, we find stronger results for five-year ES measured at the 99% confidence interval. However, VaR is simply a cut-off point and the cut-off point at the 99% confidence interval covers the cut-off point at the 95% confidence interval. Therefore, we find stronger results for five-year VaR measured at the 95% confidence interval.

Furthermore, we estimate the multivariate regression models at different thresholds with VaR and ES measured at 95% confidence interval and report the results in Table 10. Similar to the results report above, our results are robust with three-year, five-year and sixty-month VaR at all the three threshold levels of 10%, 20% and 30% (see columns (3) to (5)). However, only sixty-month ES significantly predicts the likelihood of financial distress of firms at all the three threshold levels (see column (8)). Five-year ES is significant only at 10% threshold level. Overall, our multivariate results with tail risk measures of VaR and ES estimated at 95% confidence interval without any threshold and with threshold levels of 10%, 20% and 30% are broadly robust.

**[Insert Tables 9 and 10 Here]**

## 9. Conclusion

Appropriate pricing of credit risk is integral to maintaining sustainable and competitive lending practices. This also informs the estimation of capital reserves that lending institutions are required to maintain in proportion to their risk-weighted assets. Overestimation of credit risk will lead to higher capital reserves and, thus, an opportunity loss, whereas underestimation

results in lower capital reserves, but may trigger large shortfalls and significant capital losses (see Demircuc-Kunt *et al.* 2013, for brief discussion). We contribute to this strand of literature by documenting the impact of tail risk measures on firms' likelihood of financial distress.

At first, we argue the need for a mechanism to identify firms in financial difficulties rather than waiting until they file for bankruptcy. Waiting until the bankruptcy filing might lead to significant erosion in firms' value, losses to creditors and unfavourable business decisions. Thus, there is a need for an alternative mechanism to identify firms in financial difficulties and it would be appropriate to penalise them for getting close to questionable going-concern status. As a consequence, we propose a definition of financial distress contingent upon firms' earnings, financial expenses, market value and operating cash flow. Specifically, we define a firm as financially distressed if it satisfies three simple conditions: (i) there is negative growth in the average market value (AMV) of the firm in years  $t - 1$  and  $t - 2$ , where AMV is a geometrically declining weighted average of a firm's monthly market values over the past twelve months; (ii) its earnings before interest tax depreciation and amortisation is less than its financial expenses in the years  $t - 1$  and  $t - 2$ ; and (iii) its operating cash flow is less than its financial expenses in the years  $t - 1$  and  $t - 2$ .

Next, we hypothesise that more frequent extreme negative daily equity returns result in higher tail risk, and this subsequently increases firms' likelihood of entering financial distress. Thus, we assess the marginal discriminatory power of tail risk measures, namely value-at-risk and expected shortfall, in predicting financial distress likelihood of United States publicly-traded firms. Empirical results suggest that longer duration tail risk measures exhibit significant discriminatory powers. Our robustness checks substantiate these findings, and show that our proposed definition of financial distress is consistent and robust. Because of the theoretical

superiority of ES over VaR, we suggest using five-year or sixty-month ES estimates in predicting financial distress.

## Bibliography

- Agarwal, V. and Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32 (8), 1541–1551.
- Alexander, C., 2008. *Market Risk Analysis, Value at Risk Models*. 1st ed. Sussex: John Wiley & Sons.
- Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23 (4), 589–609.
- Artzner, P., Delbaen, F., Eber, J., and Heath, D., 1999. Coherent measures of risk. *Mathematical Finance*, 9 (3), 203–228.
- Arzac, E.R. and Bawa, V.S., 1977. Portfolio choice and equilibrium in capital markets with safety-first investors. *Journal of Financial Economics*, 4 (3), 277–288.
- Balcaen, S. and Ooghe, H., 2006. 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *British Accounting Review*, 38 (1), 63–93.
- Bali, T.G., 2003. An Extreme Value Approach to Estimating Volatility and Value at Risk\*. *The Journal of Business*, 76 (1), 83–108.
- Bali, T.G., Demirtas, K.O., and Levy, H., 2009. Is there an intertemporal relation between downside risk and expected returns? *Journal of Financial and Quantitative Analysis*, 44 (4), 883.
- Ball, R., Gerakos, J., Linnainmaa, J.T., and Nikolaev, V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, 121 (1), 28–45.
- Berger, A.N. and Bouwman, C.H.S., 2013. How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109 (1), 146–176.
- Bharath, S.T. and Shumway, T., 2008. Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21 (3), 1339–1369.
- 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, G.M. and Merville, L.J., 1999. An analysis of the underreported magnitude of the total indirect costs of financial distress. *Review of Quantitative Finance and Accounting*, 13 (3), 277–293.
- 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.
- Cohen, R.B., Polk, C., and Vuolteenaho, T., 2003. The value spread. *The Journal of Finance*, 58 (2), 609–642.
- Conrad, J., Dittmar, R.F., and Ghysels, E., 2013. Ex Ante Skewness and Expected Stock Returns. *The Journal of Finance*, 68 (1), 85–124.
- Cornish, E.A. and Fisher, R.A., 1938. Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique*, 5, 307–320.
- Das, S.R., Hanouna, P., and Sarin, A., 2009. Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking & Finance*, 33 (4), 719–730.
- Davydenko, S.A. and Franks, J.R., 2008. Do bankruptcy codes matter? A study of defaults in France, Germany, and the UK. *The Journal of Finance*, 63 (2), 565–608.
- Demirguc-Kunt, A., Detragiache, E., and Merrouche, O., 2013. Bank Capital: Lessons from the Financial Crisis. *Journal of Money, Credit and Banking*, 45 (6), 1147–1164.
- Edison, H.J., 2003. Do indicators of financial crises work? An evaluation of an early warning system. *International Journal of Finance & Economics*, 8 (1), 11–53.
- Franks, J. and Loranth, G., 2014. A study of bankruptcy costs and the allocation of control. *Review of Finance*, 18 (3), 961–997.
- Gennaioli, N. and Rossi, S., 2010. Judicial discretion in corporate bankruptcy. *Review of Financial Studies*, 23 (11), 4078–4114.
- George, T.J. and Hwang, C.-Y., 2010. A resolution of the distress risk and leverage puzzles in the cross section of stock returns. *Journal of Financial Economics*, 96 (1), 56–79.
- 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.
- Glover, B., 2016. The expected cost of default. *Journal of Financial Economics*, 119 (2), 284–299.
- Grullon, G., Hund, J., and Weston, J.P., 2017. Concentrating on q and cash flow. *Journal of Financial Intermediation*.
- Gupta, J. and Barzotto, M., 2018. *Insights on Bankruptcy Emergence*. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3216433](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3216433)
- Gupta, J., Gregoriou, A., and Ebrahimi, T., 2018. Empirical Comparison of Hazard Models in Predicting SMEs Failure. *Quantitative Finance*, 18 (3), 437–466.

- 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.
- Huang, W., Liu, Q., Rhee, S.G., and Wu, F., 2012. Extreme downside risk and expected stock returns. *Journal of Banking & Finance*, 36 (5), 1492–1502.
- Jansen, D.W. and De Vries, C.G., 1991. On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, 18–24.
- Jaschke, S.R., 2002. The Cornish–Fisher expansion in the context of Delta–Gamma-normal approximations. *Journal of Risk*, 4, 33–52.
- John, K., Mateti, R.S., and Vasudevan, G., 2013. Resolution of financial distress: A theory of the choice between Chapter 11 and workouts. *Journal of Financial Stability*, 9 (2), 196–209.
- Johnson, K., Kotz, S., and Balakrishnan, N., 1994. *Continuous Univariate Distributions*. 2nd ed. New York: John Wiley & Sons.
- Jones, S., 2017. Corporate bankruptcy prediction: a high dimensional analysis. *Review of Accounting Studies*, 22 (3), 1366–1422.
- Jorion, P., 2000. Risk management lessons from long-term capital management. *European Financial Management*, 6 (3), 277–300.
- Keasey, K., Pindado, J., and Rodrigues, L., 2015. The determinants of the costs of financial distress in SMEs. *International Small Business Journal*, 33 (8), 862–881.
- Lee, J.E., Glasscock, R., and Park, M.S., 2016. Does the Ability of Operating Cash Flows to Measure Firm Performance Improve during Periods of Financial Distress? *Accounting Horizons*, 31 (1), 23–35.
- Lewellen, J. and Lewellen, K., 2016. Investment and cash flow: New evidence. *Journal of Financial and Quantitative Analysis*, 51 (4), 1135–1164.
- 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.
- Long, J.S. and Freese, J., 2014. *Regression models for categorical dependent variables using Stata*. 3rd ed. Texas: Stata press.
- Longin, F.M., 2000. From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance*, 24 (7), 1097–1130.
- Myers, S.C., 1977. Determinants of corporate borrowing. *Journal of Financial Economics*, 5 (2), 147–175.
- Neftci, S.N., 2000. Value at risk calculations, extreme events, and tail estimation. *The Journal of Derivatives*, 7 (3), 23–37.
- 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.

- Purnanandam, A., 2008. Financial distress and corporate risk management: Theory and evidence. *Journal of Financial Economics*, 87 (3), 706–739.
- Reed, W.R., 2015. On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics*, 77 (6), 897–905.
- Righi, M.B. and Ceretta, P.S., 2016. Shortfall deviation risk: an alternative for risk measurement. *Journal of Risk*, 19 (2), 81–116.
- Roy, A.D., 1952. Safety first and the holding of assets. *Econometrica: Journal of the Econometric Society*, 20, 431–449.
- 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.
- 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*, 30, 394–419.
- Trujillo-Ponce, A., Samaniego-Medina, R., and Cardone-Riportella, C., 2014. Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management*, 15 (2), 253–276.
- Zangari, P., 1996. A VaR methodology for portfolios that include options. *RiskMetrics Monitor*, 1, 4–12.

## Tables

Table 1: Descriptive Statistics

This table provides descriptive statistics for covariates used in the regressions. To facilitate comparison, descriptive measures are reported separately for financially distressed and healthy groups. Except EXRETAVG, two groups mean comparison test for respective covariates are statistically significant at 1% significance level.

Variable	Status	Mean	Median	Standard Deviation	Minimum	Maximum
(1)	(2)	(3)	(4)	(5)	(6)	(7)
TLMTA	Distressed	0.2816	0.2078	0.2284	0.0352	0.8446
	Healthy	0.3440	0.2970	0.2407	0.0352	0.8446
CMTA	Distressed	0.2025	0.1681	0.1489	0.0028	0.4510
	Healthy	0.1129	0.0648	0.1249	0.0027	0.4510
NIMTAAVG	Distressed	-0.1812	-0.1780	0.1046	-0.3039	0.0750
	Healthy	-0.0192	0.0192	0.0987	-0.3039	0.0750
EXRETAVG	Distressed	-0.0071	-0.0056	0.0572	-0.0984	0.0712
	Healthy	-0.0079	-0.0051	0.0430	-0.0984	0.0712
SIGMA	Distressed	0.8002	0.7433	0.3059	0.1962	1.2851
	Healthy	0.5529	0.4734	0.2993	0.1963	1.2851
AGE	Distressed	2.4457	2.3979	0.5469	1.3863	4.1897
	Healthy	2.3227	2.3979	1.0483	0.0000	4.2047
MB	Distressed	3.1660	2.6513	2.0469	0.6588	6.6199
	Healthy	2.4520	2.0025	1.5767	0.6588	6.6199
PRICE	Distressed	1.0365	0.8755	0.8451	0.1178	2.7081
	Healthy	2.1295	2.6301	0.7971	0.1178	2.7081
RSIZE	Distressed	-11.9120	-12.0701	1.3269	-13.4014	-6.8133
	Healthy	-10.2861	-10.3200	1.8228	-13.4014	-6.8133
FES	Distressed	0.0445	0.0170	0.0516	0.0000	0.1290
	Healthy	0.0242	0.0112	0.0333	0.0000	0.1290
TMTA	Distressed	0.0001	0.0000	0.0079	-0.0174	0.0494
	Healthy	0.0115	0.0090	0.0159	-0.0174	0.0494
RISKFD	Distressed	0.0337	0.0250	0.0227	-0.0009	0.1667
	Healthy	0.0146	0.0094	0.0164	0.0000	0.1667
VAR3M1	Healthy	0.1300	0.0961	0.1106	1.0000	0.0033
	Healthy	0.0880	0.0672	0.0779	1.0000	0.0001
VAR6M1	Distressed	0.1550	0.1111	0.1343	1.0000	0.0004
	Healthy	0.1058	0.0798	0.0942	1.0000	8.2e-05
VAR1Y1	Distressed	0.1867	0.1354	0.1566	1.0000	0.0052
	Healthy	0.1243	0.0922	0.1139	1.0000	4.2e-05
VAR3Y1	Distressed	0.2829	0.2009	0.2240	1.0000	0.0480
	Healthy	0.1689	0.1202	0.1595	1.0000	0.0023
VAR5Y1	Distressed	0.3475	0.2408	0.2605	1.0000	0.0510
	Healthy	0.1945	-0.1366	0.1832	1.0000	0.0011
VAR60M1	Distressed	0.5797	0.5418	0.2339	1.0000	0.0628
	Healthy	0.3758	0.3223	0.2137	1.0000	0.0004
ES3M1	Distressed	0.2103	0.1243	0.2277	1.0000	0.0248
	Healthy	0.1287	0.0841	0.1555	1.0000	0.0007
ES6M1	Distressed	0.2250	0.1439	0.2173	1.0000	0.0009
	Healthy	0.1389	0.0987	0.1405	1.0000	0.0001
ES1Y1	Distressed	0.2717	0.1842	0.2307	1.0000	0.0003
	Healthy	0.1660	0.1191	0.1572	1.0000	4.2E-05
ES3Y1	Distressed	0.3686	0.2828	0.2487	1.0000	0.0034
	Healthy	0.2044	0.1497	0.1778	1.0000	2.9E-05
ES5Y1	Distressed	0.3968	0.3028	0.2605	1.0000	-0.0029
	Healthy	0.2201	0.1603	0.1897	1.0000	2.7E-06
ES60M1	Distressed	0.6828	0.6620	0.2407	1.0000	0.1531
	Healthy	0.4607	0.4004	0.2481	1.0000	0.0363

Table 2: Univariate Regression

This table reports univariate regression estimates of respective covariates using our proposed definition of financial distress as the dependent variable. AME is average marginal effects in percentages. Column (6) reports ranks of main variables based on the magnitude of their average marginal effects (AME).

Variables	Coefficient	Standard Error	z-statistics	AME %	Rank
(1)	(2)	(3)	(4)	(5)	(6)
TLMTA	1.2407***	0.1443	8.60	0.41***	8
CMTA	5.0517***	0.2107	23.98	2.78***	5
NIMTAAVG	-9.9555***	0.2553	-38.99	-7.20***	2
EXRETAVG	-11.7030***	0.6643	-17.62	-6.71***	3
SIGMA	1.7099***	0.0952	17.97	1.05***	----
MB	-0.0822***	0.0208	-3.95	-0.03***	9
PRICE	-1.8438***	0.0843	-38.21	-1.47***	6
RSIZE	-1.0774***	0.0376	-28.69	-0.62***	7
FES	10.6394***	0.7990	13.32	4.80***	4
TMTA	-53.7919***	2.7701	-19.42	-30.07***	1
AGE	0.5636***	0.0465	12.12	0.14***	----
INDRISK	22.6630***	1.3730	16.51	8.72***	----
VAR3M1	2.1574***	0.2666	8.09	1.05***	----
VAR6M1	1.7075***	0.2256	7.57	0.84***	----
VAR1Y1	1.4859***	0.1900	7.82	0.72***	----
VAR3Y1	1.7141***	0.1404	12.21	0.91***	----
VAR5Y1	2.1107***	0.1274	16.57	1.20***	----
VAR60M1	2.8442***	0.1303	21.83	1.87***	----
ES3M1	1.2239***	0.1327	9.22	0.59***	----
ES6M1	1.5158***	0.1447	10.48	0.76***	----
ES1Y1	1.5925***	0.1355	11.74	0.81***	----
ES3Y1	2.2144***	0.1239	17.87	1.13***	----
ES5Y1	2.3195***	0.1236	18.77	1.39***	----
ES60M1	2.5213***	0.1231	20.47	1.66***	----

\* p < 0.10.

\*\* p < 0.05.

\*\*\* p < 0.01.

Table 3: Baseline Multivariate Regression Models

This table reports multivariate regression estimates that employ only main variable discussed in Section 4.2 for T-1 (column 2), T-2 (column 3) and T-3 (column 4) lagged covariates. The baseline multivariate model in column 2 is developed following the method suggested by Gupta *et al.* (2018) (see section 5). TLMTA, MB, RSIZE, PRICE and SIGMA are excluded from the model for the reasons mentioned in section 5 (related to the expected coefficient sign and the p-value, or strong correlation with other covariates). N = 0 + 1 represent the total number of firm-year observations for the two groups of firms between 1990 and 2016. N = 1 is the number of observations for the financially distressed firms. This table also reports the chi-square, the likelihood ratio and the coefficient of determination (McKelvey & Zavoina's R<sup>2</sup>) to measure the model's goodness of fit. AWROC-W is within-sample area under ROC curve and AWROC-H is for the hold-out sample.

Variables		(T - 1)	(T - 2)	(T - 3)
(1)		(2)	(3)	(4)
CMTA	$\beta$	3.9430***	3.3276***	1.8415***
	SE	0.2480	0.2598	0.2618
	AME	5.1450***	4.4193***	2.7795***
NIMTAAVG	$\beta$	-9.6073***	-11.7358***	-9.3950***
	SE	0.3362	0.3711***	0.3333
	AME	-12.5361***	-15.5853	-14.1800***
EXRETAVG	$\beta$	-3.6941***	-12.6168***	-0.5815
	SE	0.7194	0.7968	-0.8200
	AME	-4.8202***	-16.7559***	-0.8776
FES	$\beta$	7.0887***	3.8007***	3.2327***
	SE	0.7688	0.8400	0.8155
	AME	9.2497***	5.0475***	4.8792***
TMTA	$\beta$	-26.3068***	-22.0542***	-23.7398***
	SE	2.9939	2.9722	2.9429
	AME	-34.3264***	-29.2894***	-35.8306***
AGE	$\beta$	0.3017***	0.1799***	-0.2586***
	SE	0.0558	0.0522	0.0459
	AME	0.3937***	0.2389***	-0.3903***
INDRISK	$\beta$	18.7770***	-2.5321***	-5.6689***
	SE	1.6898	2.2349	2.0998
	AME	24.5011***	-3.3628***	-8.5562***
Model's goodness of fit and prediction performance measures				
Chi2		2236.11***	2242.69***	1817.70***
Log likelihood		-4150.7302	-3741.3455	-4182.0543
R <sup>2</sup>		0.3917	0.3536	0.2508
AUROC-W		0.9240	0.9431	0.9026
AUROC-H		0.9193	0.9428	0.8995
N = 0+1		59,602	53,003	48,010
N = 1		1,160	1,225	1,168

\* p < 0.10.  
 \*\* p < 0.05.  
 \*\*\* p < 0.01.

Table 4: Multivariate Regression Models with VaR estimated at 1% Significance Level

This table reports multivariate regression estimates obtained by supplementing the baseline model reported in Table 3 with SIGMA and VaR measures. Estimates of the risk measures as listed in columns (2) to (8) are reported in rows (4) to (7). Column (2) reports regression estimates with SIGMA being the risk measure. Columns (3) to (8) report respective multivariate regression estimates supplementing VaR as the risk measure (three-month, six-month, one-year, three-year and five-year daily returns and sixty-month monthly return) at the 99% confidence interval. The chi-square for likelihood ratio test (LR  $Chi^2$ ) reported in row (7) of column (1) tests the existence of any significant difference between the log likelihood of the baseline model supplemented with risk variable, and log likelihood of baseline model. The last five rows of this table provide goodness of fit measures. Chi-square (Chi2), likelihood ratio, and the coefficient of determination (McKelvey & Zavoina's  $R^2$ ) of respective multivariate models with respective risk variables are reported in the first three of the last five rows.  $N = 0 + 1$  represents the total number of firm-year observations for the two groups of firms between 1990 and 2016.  $N = 1$  is the number of firm-year observations for financially distressed group of firms.

Variables	SIGMA	VAR3M1	VAR6M1	VAR1Y1	VAR3Y1	VAR5Y1	VAR60M1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Measure							
$\beta$	0.3446***	-0.6634*	-0.4875	-0.2922	0.2044	0.5732***	0.6907***
SE	0.1251	0.3682	0.3045	0.2512	0.1726	0.1503	0.1667
AME	0.4495***	-0.8697*	-0.2649	-0.3773	0.2662	0.7437***	0.8826***
LR $Chi^2$	7.5000***	3.4400*	2.6500	1.3800	1.3800	14.100***	16.9500***
CMTA							
$\beta$	4.0596***	3.8820***	3.9273***	3.9347***	3.9766***	3.9924***	3.9923***
SE	0.2504	0.2522	0.2511	0.2512	0.2492	0.2511	0.2511
NIMTAAVG							
$\beta$	-9.1909***	-9.8214***	-9.8028***	-9.7502***	-9.5584***	-9.1629***	-9.1629***
SE	0.3721	0.3551	0.3542	0.3536	0.3463	0.3609	0.3608
EXRETAVG							
$\beta$	-3.4113***	-3.8365***	-3.8214***	-3.7963***	-3.5371***	-3.4006***	-3.4005***
SE	0.7217	0.7318	0.7280	0.7283	0.7233	0.7234	0.7234
FES							
$\beta$	7.1077***	7.0961***	7.1864***	7.0929***	7.0809***	6.9784***	6.9783***
SE	0.7692	0.7760	0.7733	0.7769	0.7735	0.7786	0.7786
TMTA							
$\beta$	-25.7929***	-26.0233***	-25.9140***	-25.9666***	-26.1026***	-25.5957***	-25.5957***
SE	3.0013	3.0130	3.0053	3.0144	3.0070	3.0193	3.0192
AGE							
$\beta$	0.3235***	0.2862***	0.2985***	0.2920***	0.3038***	0.3375***	0.3375***
SE	0.0568	0.0565	0.0563	0.0565	0.0563	0.0576	0.0576
RISK							
$\beta$	19.5882***	18.8046***	18.8477***	18.8847***	18.7999***	19.4803***	19.4802***
SE	1.7130	1.7131	1.7045	1.7078	1.6952	1.7022	1.7022
Model's goodness of fit and prediction performance measures							
chi2	2232.06***	2209.66***	2234.30***	2208.21***	2226.58***	2184.84***	2184.84***
Log likelihood	-4128.5475	-4032.4688	-4063.9586	-4066.6788	-4120.0273	-4084.6001	-4084.6001
R <sup>2</sup>	0.3299	0.3267	0.3280	0.3257	0.3270	0.3305	0.3305
N = 0+1	59,601	58,659	57,698	57,869	58,175	58,185	58,185
N = 1	1,215	1,183	1,196	1,194	1,213	1,205	1,205

\* p < 0.10.

\*\* p < 0.05.

\*\*\* p < 0.01.

Table 5: Multivariate Regression Models with ES estimated at 1% Significance Level

This table reports multivariate regression estimates obtained by supplementing the baseline model reported in Table 3 with SIGMA and ES measures. Estimates of the risk measures as listed in columns (2) to (8) are reported in rows (4) to (7). Column (2) reports regression estimates with SIGMA being the risk measure. Columns (3) to (8) report respective multivariate regression estimates supplementing ES as the risk measure at the 99% confidence interval. The chi-square for likelihood ratio test (LR  $Chi^2$ ) reported in row (7) of column (1) tests the existence of any significant difference between the log likelihood of the baseline model supplemented with risk variable, and log likelihood of baseline model. The last five rows of this table provide goodness of fit measures. Chi-square (Chi2), likelihood ratio and the coefficient of determination (McKelvey & Zavoina's  $R^2$ ) of respective multivariate models with respective risk variables are reported in the first three of the last five rows.  $N = 0 + 1$  represents the total number of firm-year observations for the two groups of firms between 1990 and 2016.  $N = 1$  is the number of firm-year observations for financially distressed group of firms.

Variables	SIGMA	ES3M1	ES6M1	ES1Y1	ES3Y1	ES5Y1	ES60M1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Measure							
$\beta$	0.3446***	0.1632	0.0947	0.2682	0.6783***	0.7360***	0.6769***
SE	0.1251	0.1636	0.1840	0.1667	0.1536	0.1489	0.1558
AME	0.4495***	0.2124	0.1238	0.3472	0.8813***	0.9523***	0.8715***
LR $Chi^2$	7.500***	0.9800	0.2600	2.5400	18.7800***	23.6100***	18.7700***
CMTA							
$\beta$	4.0596***	3.9649***	3.9807***	3.9924***	3.9973***	3.9547***	3.9882***
SE	0.2504	0.2492	0.2493	0.2503	0.2491	0.2496	0.2498
NIMTAAVG							
$\beta$	-9.1909***	-9.5841***	-9.5781***	-9.4613***	-9.2437***	-9.2368***	-9.0949***
SE	0.3721	0.3433	0.3486	0.3499	0.3490	0.3476	0.3592
EXRETAVG							
$\beta$	-3.4113***	-3.5959***	-3.7082***	-3.095***	-3.3643***	-3.4783***	-3.4067***
SE	0.7217	0.7226	0.7230	0.7255	0.7199	0.7185	0.7192
FES							
$\beta$	7.1077***	7.1047***	7.1388***	7.0795***	6.8732***	6.7409***	6.9813***
SE	0.7692	0.7708	0.7701	0.7742	0.7734	0.7768	0.7737
TMTA							
$\beta$	-25.7929***	-26.0431***	-26.1313***	-26.1694***	-25.9099***	-25.6348***	-25.3468***
SE	3.0013	3.0031	3.0021	3.0124	3.0168	3.0206	3.0065
AGE							
$\beta$	0.3235***	0.3026***	0.3023***	0.3027***	0.3176***	0.3196***	0.3427***
SE	0.0568	0.0562	0.0563	0.05630	0.0565	0.0566	0.0574
RISK							
$\beta$	19.5882***	18.9429***	19.1011***	19.1011***	18.9332***	18.9319***	19.3062***
SE	1.7130	1.6969	1.6992	1.6992	1.6965	1.6971	1.7005
Model's goodness of fit and prediction performance measures							
chi2	2232.06***	2229.36***	2238.87***	2211.94***	2223.00***	2211.40***	2196.33***
Log likelihood	-4128.5475	-34122.0715	-4110.1135	-4096.2486	-4114.9963	-4115.2167	-4110.5026
$R^2$	0.3299	0.3267	0.3282	0.3269	0.3290	0.3287	0.3313
$N = 0+1$	58,386	58,236	58,131	58,138	58,311	58,331	58,343
$N = 1$	1,215	1,213	1,210	1,202	1,213	1,214	1,212

\*  $p < 0.10$ .  
 \*\*  $p < 0.05$ .  
 \*\*\*  $p < 0.01$ .

Table 6: Baseline Multivariate Model at different threshold levels

This table reports the baseline multivariate regression estimates for  $t - 1$  lagged covariates, similar to the model reported in column 2 of Table 3. Columns 3, 4 and 5 report results obtained from arbitrarily imposing 10%, 20% and 30% threshold levels respectively on conditions 1, 2 and 3 of our proposed financial distress definition (see section 6 for details). All regression estimates are reported for  $t - 1$  lagged covariates.  $N = 0 + 1$  represent the total number of firm-year observations for the two groups of firms between 1990 and 2016.  $N = 1$  is the number of firm-year observations for financially distressed group of firms. This table also reports the chi-square, the likelihood ratio and the coefficient of determination (McKelvey & Zavoina's  $R^2$ ) to measure the model's goodness of fit.

Variables	Main (0% Threshold)	10% Threshold	20% Threshold	30% Threshold
(1)	(2)	(3)	(4)	(5)
CMTA				
$\beta$	3.9430***	2.908***	2.9629***	3.1287***
SE	0.2480	0.3063	0.3432	0.4006
AME	5.1450***	2.4574***	1.9692***	1.4829***
NIMTAAVG				
$\beta$	-9.6073***	-9.1548***	-9.5856***	-10.3909***
SE	0.3362	0.4180	0.4787	0.5766
AME	-12.5361***	-7.7360***	-6.3709***	-4.9247***
EXRETAVG				
$\beta$	-3.6941***	-4.5370***	-4.3899***	-4.1111***
SE	0.7194	0.9034	1.0367	1.2236
AME	-4.8202***	-3.8334***	-2.9177***	-1.9484***
FES				
$\beta$	7.0887***	11.0839***	9.5835***	8.3311***
SE	0.7688	0.8878	0.9944	1.1687
AME	9.2497***	9.3660***	6.3695***	3.9485***
TMTA				
$\beta$	-26.3068***	-22.8561***	-19.1492***	-13.6018***
SE	2.9939	3.8382	4.4013	5.1473
AME	-34.3264***	-19.3139***	-12.7271***	-6.4465***
AGE				
$\beta$	0.3017***	0.1957***	0.1071	0.0378
SE	0.0558	0.0682	0.0779	0.0934
AME	0.3937***	0.1653***	0.0712	0.0179
INDRISK				
$\beta$	18.7770***	192664***	19.4829***	20.9762***
SE	1.6898	2.0835	2.3834	2.7946
AME	24.5011***	16.2805***	12.9489***	9.9415***
Model's goodness of fit and prediction performance measures				
Chi2	2236.11***	1506.61***	1196.95***	903.11***
Log likelihood	-4150.7302	-2730.9877	-2101.0535	-1532.2126
R <sup>2</sup>	0.3263	0.3211	0.3262	0.3340
N = 0+1	59,691	58,995	59,187	59,341
N = 1	1,219	696	504	350

\*  $p < 0.10$ .  
 \*\*  $p < 0.05$ .  
 \*\*\*  $p < 0.01$ .

Table 7: Multivariate Regression Models at different thresholds with VaR estimated at 1% Significance Level

This table reports multivariate regression estimates obtained by supplementing models estimated at different threshold levels of conditions 1, 2 and 3 of our proposed financial distress definition (see section 6 for details) with SIGMA and VaR measures. Estimates of the risk measures as listed in columns (2) to (8) are provided in rows (4) to (7). Column (2) reports regression estimates with SIGMA being the risk measure. Columns (3) to (8) report respective multivariate regression estimates supplementing VaR as the risk measure (three-month, six-month, one-year, three-year and five-year daily returns and sixty-month monthly returns) at the 99% confidence interval. The chi-square for likelihood ratio test ( $LR\ Chi^2$ ) reported in row (7) of column (1) tests the existence of any significant difference between the log likelihood of the baseline model supplemented with risk variable, and log likelihood of baseline model. The last 5 rows of respective panels provide goodness of fit measures. Chi-square ( $Chi^2$ ), log likelihood and the coefficient of determination (McKelvey & Zavoina's  $R^2$ ) of respective multivariate model with respective risk variables are reported in first three of last five rows. All regression estimates are reported for  $t - 1$  lagged covariates. To save space, we do not report the baseline covariates of Table 6 in this table.  $N = 0 + 1$  represent the total number of firm-year observations for the two groups of firms between 1990 and 2016.  $N = 1$  is the number of firm-year observations for financially distressed group of firms.

Panel A: 10 % Threshold Estimates							
Risk Measures	SIGMA	VAR3M1	VAR6M1	VAR1Y1	VAR3Y1	VAR5Y1	VAR60M1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\beta$	0.3724***	-0.4628	-0.3947	-0.3527	0.2804	0.6081***	1.1308***
SE	0.1550	0.4366	0.3639	0.3054	0.2045	0.1774	0.2010
AME	0.3158***	-0.3932	-0.3330	-0.2940	0.2363	0.5113***	0.9312***
LR $Chi^2$	5.7000**	1.1600	1.2100	1.3800	1.8400	11.3200***	31.2100***
Model's goodness of fit and prediction performance measures							
Chi2	1508.75***	1493.39***	1491.64***	1479.09***	1500.05***	1491.80***	1464.23***
Log likelihood	-2716.0946	-2669.8853	-2677.1974	-2671.3670	-2710.5895	-2709.0528	-2680.1741
$R^2$	0.3249	0.3208	0.3215	0.3189	0.3212	0.3224	0.3266
$N = 0+1$	58,907	57,979	58,211	58,384	58,695	58,724	58,702
$N = 1$	694	680	683	679	693	694	688
Panel B: 20 % Threshold Estimates							
$\beta$	0.4900***	-0.4411	-0.6696	-0.5887*	-0.0557	0.2963	1.0120***
SE	0.1761	0.4899	0.4258	0.3555	0.2378	0.2031	0.2280
AME	0.3263	-0.2924	-0.4399	-0.3806*	-0.0369	0.1965	0.6509***
LR $Chi^2$	7.6600***	0.8400	2.6300	2.9100*	0.0600	2.0800	19.4700***
Model's goodness of fit and prediction performance measures							
Chi2	1196.56***	1177.35***	1177.57***	1166.62***	1192.22***	1189.53***	1153.79***
Log likelihood	-2084.8657	-2045.4045	-2045.3282	-2053.0042	-2086.1928	-2084.6411	-2057.9970
$R^2$	0.3322	0.3254	0.3261	0.3241	0.3266	0.3270	0.3312
$N = 0+1$	59,099	58,169	58,403	58,571	58,886	58,916	58,893
$N = 1$	502	490	491	492	502	502	497
Panel C: 30 % Threshold Estimates							
$\beta$	0.3990*	-0.4631	-0.9538*	-0.5987	-0.1221	0.3454	1.2543***
SE	0.2089	0.5694	0.5175	0.4129	0.2782	0.2345	0.2680
AME	0.1891*	-0.2167	-0.4466*	0.2788	-0.0577	0.1628	0.5748***
LR $Chi^2$	3.6200*	0.6900	3.7300*	2.2400	0.2000	2.1100	21.7700***
Model's goodness of fit and prediction performance measures							
Chi2	901.62***	884.89***	888.47***	889.10***	899.74***	895.38***	877.62***
Log likelihood	-1517.1064	-1493.5122	-1492.1717	-1499.9776	-1517.2735	-1515.9201	-1500.6078
$R^2$	0.3398	0.3327	0.3331	0.3334	0.3349	0.3353	0.3416
$N = 0+1$	59,253	58,317	58,552	58,719	59,040	59,070	59,043
$N = 1$	348	342	342	344	348	348	347

\*  $p < 0.10$ .

\*\*  $p < 0.05$ .

\*\*\*  $p < 0.01$ .

Table 8: Multivariate Regression Models at different thresholds with ES estimated at 1% Significance Level

This table reports multivariate regression estimates obtained by supplementing models estimated at different threshold levels of conditions 1, 2 and 3 of our proposed financial distress definition (see section 6 for details) with SIGMA and ES measures. Estimates of the risk measures as listed in columns (2) to (8) are provided in rows (4) to (7). Column (2) reports regression estimates with SIGMA being the risk measure. Columns (3) to (8) report respective multivariate regression estimates supplementing ES as the risk measure (three-month, six-month, one-year, three-year and five-year daily returns and sixty-month monthly returns) at the 99% confidence interval. The chi-square for likelihood ratio test (*LR Chi<sup>2</sup>*) reported in row (7) of column (1) tests the existence of any significant difference between the log likelihood of the baseline model supplemented with risk variable, and log likelihood of baseline model. The last 5 rows of respective panels provide goodness of fit measures. Chi-square (*Chi<sup>2</sup>*), log likelihood and the coefficient of determination (McKelvey & Zavoina's *R<sup>2</sup>*) of respective multivariate model with respective risk variables are reported in first three of last five rows. All regression estimates are reported for *t* – 1 lagged covariates. To save space, we do not report the baseline covariates of Table 6 in this table. *N* = 0 + 1 represent the total number of firm-year observations for the two groups of firms between 1990 and 2016. *N* = 1 is the number of firm-year observations for financially distressed group of firms.

Panel A: 10 % Threshold Estimates							
Risk Measures	SIGMA	ES3M1	ES6M1	ES1Y1	ES3Y1	ES5Y1	ES60M1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>β</i>	0.3724***	-0.0509	-0.1867	0.05307	0.7653***	0.81118***	1.0569***
<i>SE</i>	0.1550	0.2069	0.2311	0.2600	0.1829	0.1759	0.1938
<i>AME</i>	0.3158***	-0.04230	-0.1576	0.0443	0.7653***	0.6807***	0.8826***
<i>LR Chi<sup>2</sup></i>	5.7000**	1.0600	0.6600	0.0700	16.6900***	20.3900***	29.6200***
<b>Model's goodness of fit and prediction performance measures</b>							
<i>Chi<sup>2</sup></i>	1508.75***	1503.03***	1501.64***	1483.50***	1497.98***	1488.17***	1478.02***
Log likelihood	-2716.0946	-2717.0580	-2712.8445	-2697.4317	-2705.2087	-2707.2534	-2694.5402
<i>R<sup>2</sup></i>	0.3249	0.3208	0.3213	0.3198	0.3228	0.3232	0.3311
<i>N</i> = 0+1	58,907	58,755	58,648	58,654	58,831	58,851	58,863
<i>N</i> = 1	694	694	693	686	693	694	692
Panel B: 20 % Threshold Estimates							
<i>β</i>	0.4900***	-0.0444	-0.1753	0.2059	0.6863***	0.6492***	0.9652***
<i>SE</i>	0.1761	0.2322	0.2583	0.2285	0.2052	0.1981	0.2233
<i>AME</i>	0.3263***	-0.0294	-0.1157	0.1347	0.4544***	0.4291***	0.6295***
<i>LR Chi<sup>2</sup></i>	7.6600***	0.0400	0.4700	0.8000	10.6700***	10.3300***	18.7000***
<b>Model's goodness of fit and prediction performance measures</b>							
<i>Chi<sup>2</sup></i>	1196.56***	1193.69***	1189.37***	1179.15***	1189.81***	1183.88***	1166.31***
Log likelihood	-2084.8657	-2087.3259	-2083.2399	-2072.9820	-2082.3212	-2082.4929	-2069.5539
<i>R<sup>2</sup></i>	0.3322	0.3260	0.3260	0.3253	0.3280	0.3277	0.3353
<i>N</i> = 0+1	59,099	58,947	58,840	58,843	59,022	59,043	59,055
<i>N</i> = 1	502	502	501	497	502	502	500
Panel C: 30 % Threshold Estimates							
<i>β</i>	0.3990*	-0.3343	-0.3030	0.1054	0.6213***	0.7005***	1.0475***
<i>SE</i>	0.2089	0.2865	0.3051	0.2681	0.2397	0.2298	0.2677
<i>AME</i>	0.189*	-0.1570	0.1428	0.0497	0.2930***	0.3286***	0.4855***
<i>LR Chi<sup>2</sup></i>	3.6200*	1.4300	1.0200	0.1500	6.4300**	8.9300***	15.4300***
<b>Model's goodness of fit and prediction performance measures</b>							
<i>Chi<sup>2</sup></i>	901.62***	897.95***	898.80***	896.52***	899.25***	891.50***	878.39***
Log likelihood	-1517.1064	-1517.2659	-1516.1788	-1509.3053	-1515.0614	-1513.8117	-1506.6564
<i>R<sup>2</sup></i>	0.3398	0.3340	0.3347	0.3354	0.3360	0.3357	0.3450
<i>N</i> = 0+1	59,253	59,101	58,993	58,994	59,176	59,197	59,208
<i>N</i> = 1	348	348	348	346	348	348	347

\* *p* < 0.10.

\*\* *p* < 0.05.

\*\*\* *p* < 0.01.

Table 9: Multivariate Regression Models with VaR and ES estimated at 5% Significance Level

This table reports multivariate regression estimates obtained by supplementing the baseline model reported in Table 3 with SIGMA, VaR and ES measures estimated at 5% significance level. Estimates of the risk measures as listed in columns (2) to (8) are reported in rows (4) to (7). Column (2) reports regression estimates with SIGMA being the risk measure. Columns (3) to (5) report respective multivariate regression estimates supplementing VaR as the risk measure and columns (6) to (8) report respective multivariate regression estimates supplementing ES as the risk measure (one-year, three-year and five-year daily returns and sixty-month monthly return). The chi-square for likelihood ratio test (LR  $Chi^2$ ) reported in row (7) of column (1) tests the existence of any significant difference between the log likelihood of the baseline model supplemented with risk variable, and log likelihood of baseline model. The last five rows of this table provide goodness of fit measures. Chi-square (Chi2), likelihood ratio, and the coefficient of determination (McKelvey & Zavoina's  $R^2$ ) of respective multivariate models with respective risk variables are reported in the first three of the last five rows.  $N = 0 + 1$  represents the total number of firm-year observations for the two groups of firms between 1990 and 2016.  $N = 1$  is the number of firm-year observations for financially distressed group of firms.

Variables	SIGMA	VAR3Y5	VAR5Y5	VAR60M5	ES3Y5	ES5Y5	ES60M5
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Measure							
$\beta$	0.3446***	2.8789***	6.5853***	1.3093***	-0.4769	0.2433	0.7264***
SE	0.1251	1.0149	1.1449	0.2819	0.3512	0.2876	0.1724
AME	0.4495***	3.6191***	8.1862***	1.6723***	-0.6016	0.3077	0.9334***
LR $Chi^2$	7.5000***	7.7200***	31.7700***	21.2300***	1.9400	0.7000	17.4300***
Model's goodness of fit and prediction performance measures							
chi2	2232.06***	2126.69***	2083.52***	2185.96***	1938.18***	2142.31***	2198.83***
Log likelihood	-4128.5475	-3878.275	-3821.1612	-4089.0856	-3784.0133	-3955.3607	-4111.3689
$R^2$	0.3299	0.3264	0.3301	0.3311	0.3259	0.3218	0.3297
$N = 0+1$	59,601	56,332	55,968	58,258	58,201	56,891	58,348
$N = 1$	1,215	1,140	1,124	1,206	1,108	1,156	1,212

\*  $p < 0.10$ .

\*\*  $p < 0.05$ .

\*\*\*  $p < 0.01$ .

Table 10: Multivariate Regression Models at different thresholds with VaR and ES estimated at 5% Significance Level

This table reports multivariate regression estimates obtained by supplementing models estimated at different threshold levels of conditions 1, 2 and 3 of our proposed financial distress definition (see section 6 for details) with SIGMA, VaR and ES measures estimated at 5% significance level. Estimates of the risk measures as listed in columns (2) to (8) are provided in rows (4) to (7). Column (2) reports regression estimates with SIGMA being the risk measure. Columns (3) to (5) report respective multivariate regression estimates supplementing VaR as the risk measure and columns (6) to (8) report respective multivariate regression estimates supplementing ES as the risk measure (three-month, six-month, one-year, three-year and five-year daily returns and sixty-month monthly returns). The chi-square for likelihood ratio test (LR  $Chi^2$ ) reported in row (7) of column (1) tests the existence of any significant difference between the log likelihood of the baseline model supplemented with risk variable, and log likelihood of baseline model. The last 5 rows of respective panels provide goodness of fit measures. Chi-square (Chi2), log likelihood and the coefficient of determination (McKelvey & Zavoina's  $R^2$ ) of respective multivariate model with respective risk variables are reported in first three of last five rows. All regression estimates are reported for  $t - 1$  lagged covariates. To save space, we do not report the baseline covariates of Table 6 in this table.  $N = 0 + 1$  represent the total number of firm-year observations for the two groups of firms between 1990 and 2016.  $N = 1$  is the number of firm-year observations for financially distressed group of firms.

Panel A: 10 % Threshold Estimates							
Risk Measures	SIGMA	VAR3Y5	VAR5Y5	VAR60M5	ES3Y5	ES5Y5	ES60M5
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\beta$	0.3724***	5.01948***	8.9486***	2.0060***	-0.0045	0.4964	1.1631***
SE	0.1550	1.1076	1.2545	0.3270	0.3872	0.3170	0.2038
AME	0.3158***	4.0686***	7.3382***	1.6607***	-0.0036	0.4137	0.9674***
LR $Chi^2$	5.7000**	18.7400***	47.2600	36.4100***	0.0000	2.3100	31.5600***
Model's goodness of fit and prediction performance measures							
Chi2	1508.75***	1438.60***	1432.81***	1475.48***	1461.82***	1462.74***	1481.95***
Log likelihood	-2716.0946	-2527.0812	-2490.9328	-2684.0485	-2591.0134	-2595.1436	-2693.6768
$R^2$	0.3249	0.3241	0.3327	0.3268	0.3193	0.3204	0.3275
$N = 0+1$	58,907	56,826	56,452	58,775	57,627	57,388	58,868
$N = 1$	694	646	640	689	660	659	692
Panel B: 20 % Threshold Estimates							
$\beta$	0.4900***	4.3923***	8.7093***	1.8207***	-0.5234	0.2915	1.0660***
SE	0.1761	1.2448	1.4112	0.3652	0.4795	0.3682	0.2298
AME	0.3263	2.7035***	5.3361***	1.1833***	-0.3280	0.1840	0.6960***
LR $Chi^2$	7.6600***	11.3000***	35.3800***	24.0000***	1.2800	0.6000	20.8600***
Model's goodness of fit and prediction performance measures							
Chi2	1196.56***	1118.10	1099.30***	1168.44**	1146.37***	1135.97***	1173.36***
Log likelihood	-2084.8657	-1946.1826	-1921.0403	-2062.0946	-1985.8420	-1990.4561	-2068.5422
$R^2$	0.3322	0.3288	0.3346	0.3316	0.3248	0.3231	0.3321
$N = 0+1$	59,099	57,001	56,625	58,966	57,808	57,570	59,060
$N = 1$	502	471	467	498	479	477	500
Panel C: 30 % Threshold Estimates							
$\beta$	0.3990*	4.2145***	8.9213***	2.1655***	-0.4601	-0.1651	1.1884***
SE	0.2089	1.4047	1.5886	0.4192	0.5499	0.4703	0.2689
AME	0.1891*	0.1879***	3.9878***	0.9998***	-0.2053	-0.0754	0.5512***
LR $Chi^2$	3.6200*	8.0800***	29.1900***	25.6100***	0.7500	0.1300	18.9900***
Model's goodness of fit and prediction performance measures							
Chi2	901.62***	857.57***	849.02***	888.23***	866.12***	870.76***	883.40***
Log likelihood	-1517.1064	-1434.2777	-1408.3170	-1499.4390	-1454.9139	-1452.9582	-1504.9120
$R^2$	0.3398	0.3362	0.3450	0.3419	0.3313	0.3313	0.3413
$N = 0+1$	59,253	57,140	56,763	59,117	57,952	57,714	59,213
$N = 1$	348	332	329	347	335	333	347

\*  $p < 0.10$ .

\*\*  $p < 0.05$ .

\*\*\*  $p < 0.01$ .

# Appendix

## Appendix A1: Variable Description

Variable	Definition	Compustat Data Item
TLMTA	Total liabilities to Market-valued Total Assets	LT, PRCC_F, CSHO
CMTA	Cash and short-term assets to Market-valued Total Assets	CHE, PRCC_F, CSHO, LT
NIMTAAVG	Weighted average of Net Income to Market-valued Total Assets over the previous three-year period	NI, PRCC_F, CSHO, LT
EXRETAVG	Weighted average of monthly log excess return relative to value-weighted S&P 500 return over the previous 12-month period	PRCCM, AJEXM, SPTRN
SIGMA	Annualised standard deviation of a firm's daily log returns over the previous three months	PRCCD, AJEXDI
MB	Firm's Market-to-book ratio	PRCC_F, CSHO, SEQ
PRICE	Log of price per share winsorized above \$15	PRCC_F
RSIZE	Logarithm of each firm's size relative to S&P 500 market capitalisation	PRCC_F, CSHO, TOTVAL
FES	Financial Expense to Sale	IE, SALE
TMTA	Tax to Market-valued Total Assets	TXT, PRCC_F, CSHO, LT
AGE	Natural logarithm of firms' annual age	
INDRISK	Number of firms experiencing financial distress in respective industrial sectors in a given year/total number of firms in that industrial sector in that year	
VAR3M1	Value-at-Risk estimated using daily returns over past 3 months at 1% significance level	PRCCD, AJEXDI
VAR6M1	Value-at-Risk estimated using daily returns over past 6 months at 1% significance level	PRCCD, AJEXDI
VARIY1	Value-at-Risk estimated using daily returns over past 1 year at 1% significance level	PRCCD, AJEXDI
VAR3Y1	Value-at-Risk estimated using daily returns over past 3 years at 1% significance level	PRCCD, AJEXDI
VAR5Y1	Value-at-Risk estimated using daily returns over past 5 years at 1% significance level	PRCCD, AJEXDI
VAR60M1	Value-at-Risk estimated using monthly returns over past 5 years at 1% significance level	PRCCM, AJEXDI
VAR3M5	Value-at-Risk estimated using daily returns over past 3 months at 5% significance level	PRCCD, AJEXDI
VAR6M5	Value-at-Risk estimated using daily returns over past 6 months at 5% significance level	PRCCD, AJEXDI
VARIY5	Value-at-Risk estimated using daily returns over past 1 year at 5% significance level	PRCCD, AJEXDI
VAR3Y5	Value-at-Risk estimated using daily returns over past 3 years at 5% significance level	PRCCD, AJEXDI
VAR5Y5	Value-at-Risk estimated using daily returns over past 5 years at 5% significance level	PRCCD, AJEXDI
VAR60M5	Value-at-Risk estimated using monthly returns over past 5 years at 5% significance level	PRCCM, AJEXDI
ES3M1	Expected Shortfall estimated using daily returns over past 3 months at 1% significance level	PRCCD, AJEXDI
ES6M1	Expected Shortfall estimated using daily returns over past 6 months at 1% significance level	PRCCD, AJEXDI
ES1Y1	Expected Shortfall estimated using daily returns over past 1 year at 1% significance level	PRCCD, AJEXDI
ES3Y1	Expected Shortfall estimated using daily returns over past 3 years at 1% significance level	PRCCD, AJEXDI
ES5Y1	Expected Shortfall estimated using daily returns over past 5 years at 1% significance level	PRCCD, AJEXDI
ES60M1	Expected Shortfall estimated using monthly returns over past 5 years at 1% significance level	PRCCM, AJEXDI
ES3M5	Expected Shortfall estimated using daily returns over past 3 months at 5% significance level	PRCCD, AJEXDI
ES6M5	Expected Shortfall estimated using daily returns over past 6 months at 5% significance level	PRCCD, AJEXDI
ES1Y5	Expected Shortfall estimated using daily returns over past 1 year at 5% significance level	PRCCD, AJEXDI
ES3Y5	Expected Shortfall estimated using daily returns over past 3 years at 5% significance level	PRCCD, AJEXDI
ES5Y5	Expected Shortfall estimated using daily returns over past 5 years at 5% significance level	PRCCD, AJEXDI
ES60M5	Expected Shortfall estimated using monthly returns over past 5 years at 5% significance level	PRCCM, AJEXDI

*Notes:* This table reports the set of accounting- and market-based covariates that we use in our empirical analysis. The first column lists names of covariates, while the second column provides their respective definitions. Financial information and stock price information are sourced from the Compustat database, covering an analysis period from 1990 to 2016. The third column states the specific Compustat data items that we use to calculate respective covariates. Value-at-Risk (VAR) and Expected Shortfall (ES) are estimated at 1% and 5% significance levels using three-month, six-month, one-year, three-year and five-year daily returns, and five-year monthly returns.

Appendix A2: Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9	10
Panel A: Correlation among Main Variables										
TLMTA (1)	1.0000									
CMTA (2)	-0.2351	1.0000								
NIMTAAVG (3)	-0.0311	-0.3074	1.0000							
EXRETAVG (4)	-0.2141	-0.1351	0.1993	1.0000						
SIGMA (5)	0.0758	0.1975	-0.4696	-0.2595	1.0000					
MB (6)	-0.3104	-0.1883	-0.1360	0.2007	0.0156	1.0000				
PRICE (7)	-0.2216	-0.2598	0.5217	0.3033	-0.6121	0.1109	1.0000			
RSIZE (8)	-0.1516	-0.2596	0.3490	0.2140	-0.5189	0.2372	0.7168	1.0000		
FES (9)	0.3265	-0.0744	-0.2872	-0.0836	0.1448	0.1631	-0.1706	-0.0569	1.0000	
TMTA (10)	-0.0907	-0.1179	0.4489	0.0757	-0.2525	-0.1162	0.2874	0.1751	-0.2394	1.0000
Panel B: Correlation among Main and Tail Risk Variables										
VAR3M1	-0.0860	-0.1368	0.3489	0.2263	-0.7607	-0.0195	0.4636	0.3784	-0.1142	0.1768
VAR6M1	-0.0743	-0.1441	0.3521	0.2774	-0.6806	-0.0341	0.4320	0.3654	-0.1042	0.1799
VAR1Y1	-0.0490	-0.1538	0.3482	0.2566	-0.5946	-0.0485	0.4035	0.3484	-0.1042	0.1797
VAR3Y1	-0.0166	-0.1604	0.3474	0.1556	-0.4454	-0.0835	0.3493	0.3092	-0.1253	0.2015
VAR5Y1	-0.0096	-0.1620	0.3427	0.1250	-0.3836	-0.0943	0.3341	0.2976	-0.1300	0.1997
VAR60M1	-0.0050	-0.2147	0.4741	0.1698	-0.5438	-0.0934	0.3840	0.3304	-0.1650	0.2747
VAR3M5	-0.1004	-0.1573	0.4021	0.2570	-0.8464	-0.0058	0.5315	0.4274	-0.1309	0.2031
VAR6M5	-0.1012	-0.1661	0.4210	0.3059	-0.8335	-0.0072	0.5475	0.4454	-0.1286	0.2134
VAR1Y5	-0.0858	-0.1746	0.4390	0.2813	-0.8154	-0.0179	0.5688	0.4677	-0.1306	0.2260
VAR3Y5	-0.0419	-0.1740	0.4525	0.1842	-0.7576	-0.0453	0.5664	0.4866	-0.1378	0.2545
VAR5Y5	-0.0049	-0.1803	0.4413	0.1511	-0.7144	-0.0579	0.5598	0.4961	-0.1275	0.2538
VAR60M5	0.0118	-0.2385	0.5485	0.1994	-0.6432	-0.1059	0.4541	0.3896	-0.1764	0.3065
ES3M1	-0.0500	-0.1115	0.2676	0.1478	-0.5639	-0.0311	0.3263	0.2806	-0.0924	0.1423
ES6M1	-0.0667	-0.1336	0.3214	0.2064	-0.5859	-0.0330	0.3920	0.3392	-0.0994	0.1646
ES1Y1	-0.0543	-0.1514	0.3463	0.1939	-0.5485	-0.0371	0.4193	0.3713	-0.1116	0.1820
ES3Y1	-0.0185	-0.1747	0.4060	0.1491	-0.5130	-0.0851	0.4375	0.3889	-0.1460	0.2257
ES5Y1	-0.0042	-0.1760	0.3908	0.1331	-0.4524	-0.0981	0.4085	0.3632	-0.1499	0.2231
ES60M1	-0.0070	-0.2074	0.4590	0.1627	-0.5459	-0.0915	0.4087	0.3651	-0.1633	0.2727
ES3M5	-0.0749	-0.1414	0.3582	0.2088	-0.7499	-0.0251	0.4491	0.3687	-0.1142	0.1777
ES6M5	-0.0718	-0.1337	0.3369	0.2546	-0.6519	-0.0254	0.4096	0.3444	-0.0935	0.1678
ES1Y5	-0.0503	-0.1346	0.3129	0.2390	-0.5811	-0.0280	0.3782	0.3143	-0.0823	0.1580
ES3Y5	-0.0083	-0.1067	0.2509	0.1199	-0.4184	-0.0413	0.2733	0.2345	-0.0713	0.1434
ES5Y5	0.0072	-0.0931	0.2046	0.0757	-0.3091	-0.0432	0.2137	0.1907	-0.0582	0.1177
ES60M5	-0.0009	-0.2105	0.4795	0.1747	-0.5582	-0.0986	0.4074	0.3497	-0.1637	0.2756

Notes: This table reports correlation among the set of covariates. Panel A displays correlations among main variables suggested by Campbell *et al.* (2008) to which we supplemented FES and TMTA. Panel B provides correlations among main variables and the respective tail risk measures (VaR and ES).

Appendix A3: Area under ROC Curves

