Managerial Fi

On the Determinants and Prediction of Corporate Financial Distress in India

Journal:	*Managerial Finance*
Manuscript ID	MF-06-2020-0332.R3
Manuscript Type:	Original Article
Keywords:	Financial distress prediction, Logit Model, Support Vector Machine, Artificial Neural Networks

On the Determinants and Prediction of Corporate Financial Distress in India

Abstract

Purpose – The main aim of the study is to identify some critical microeconomic determinants of financial distress and to design a parsimonious distress prediction model for an emerging economy like India. In doing so, we also attempt to compare the forecasting accuracy of alternative distress prediction techniques.

Design/methodology/approach – In this study, we use two alternatives accounting information-based definitions of financial distress to construct a measure of financial distress. We then use the binomial logit model and two other popular machine learning based models, namely Artificial Neural Network and Support Vector Machine, to compare the distress prediction accuracy rate of these alternative techniques for the Indian corporate sector.

Findings – Our empirical results suggest that five financial ratios, namely return on capital employed, cash flows to total liability, asset turnover ratio, fixed assets to total assets, debt to equity ratio and a measure of firm size (log total assets) play highly significant role in distress prediction. Our findings suggest that machine learning based models namely SVM and ANN are superior in terms of their prediction accuracy compared to the simple binomial logit model. Results also suggest that one year ahead forecasts are relatively better than the two year ahead forecasts.

Originality/value – This study is one of the first comprehensive attempts to investigate and design a parsimonious distress prediction model for the emerging Indian economy which is currently facing high levels of corporate financial distress. Unlike the previous studies, we use two different accounting information-based measures of financial distress in order to identify an effective way of measuring financial distress. Some of the determinants of financial distress identified in this study are different from the popular distress prediction models used in the literature. Our distress prediction model can be useful for the other emerging markets for distress prediction.

Keywords: Financial distress prediction, Logit Model, Support Vector Machine, Artificial Neural Networks, Corporate Profitability

JEL Classification: G32, G33, C45

1. Introduction

The main aim of the study is to identify some critical microeconomic determinants of financial distress and to design a parsimonious distress prediction model for an emerging economy like India. The issue of financial distress and the case of Indian corporate sector are important and interesting for the fact that Indian economy is facing a somewhat perplexing situation best described as the coexistence of a relatively high growth rate and considerably high financial distress in the corporate sector. The present study is an attempt to provide some new evidence on financial distress and contribute to the existing literature by using a comprehensive dataset of publicly listed non-financial companies in India.

Corporate financial distress has now become a worrying economic reality for policy makers both in advanced as well as emerging economies. The level of corporate distress in the post global financial crisis period has increased to the extent that sometimes investors and lenders look suspicious about the old adage “too big to fail” (Altman and Hotchkiss, 2006). This is also appearing to be true for India in the sense that many big business firms have expressed their inability to repay their debts in the recent past. In short, financial distress refers to a situation in which a firm's cash flows are not sufficient enough to meet contractually required payment obligations. There are large direct and indirect costs of financial distress and relatively high levels of distress can destabilize the overall financial system by gradually impairing the balance sheet of lending institutions (Economic Survey, 2017-2018).

In normal times, the birth and death of firms are in fact not a completely unnatural phenomenon which demands immediate policy intervention. It is rather considered to be a part of the overall economic process governing the production, distribution and consumption activities during different phases of the business cycles. During the normal progression of economic activities, firms take birth, gradually grow and become unicorns and many a times they exit the production process without significantly affecting the overall production. But sometimes the unusual exit of firms due to financial distress imposes huge direct and indirect cost on the economy in terms of output, employment, demand and revenue (see Altman and Hotchkiss, 2006, pp. 93). Further, it also leads to under investment and misallocation of resources as distressed firms have the tendency to under-invest by only focusing on some investment project that will only help them in avoiding bankruptcy (López-Gutiérrez et al. (2015). Hence, it is of great policy importance to investigate and understand the dynamics of financial distress by focusing on an emerging economy like India.

At present, the Indian economy, one of the fastest growing large economies in the world, is passing through a tough business environment. Unlike many advanced and emerging countries, the Indian economy has witnessed impressive growth with the gross domestic product (GDP) consistently above 7 percent per annum between 2011 and 2017 with some moderation in 2018–2019. However, notwithstanding with this impressive growth rates, many firms in the corporate sector have expressed their inability to service their debt and revealed severe financial distress in their respective balance sheets. The level of financial distress has increased considerably and now it has starting impacting the balance sheet of lending institutions.[1] In response to the growing financial distress, the government of India implemented the Insolvency and Bankruptcy Code, 2016 (IBC). After its implementation around 14,000 applications had been filed within first 27 months for initiation of Corporate Insolvency Resolution Process (CIRPs) by February 2019 (see Economic Survey, 2017-18). This clearly indicates the seriousness of financial distress in the Indian corporate sector where a large number of firms are

[1] It is noteworthy that by the year 2013 nearly one-third of corporate debt was owed by firms with an interest coverage ratio (ICR) less than 1.

waiting to exit. Therefore, aclear understanding of distress dynamics of the Indian corporate sector and identification of key determinants of financial distress can be useful in developing a sound distress prediction system.

Against this background, the present study attempts to contribute to the existing literature in multiple ways. First, building on the findings of previous studies and moving a step further, we considered two different measures of financial distress to classify firms in distressed and healthy categories. This exercise will help in identifying a more effective way of measuring financial distress for an emerging market like India. Second, while some of the previous studies have focused on establishing a relationship between financial distress and accounting ratios (see for example, Altman 1968, Mselmi, 2017, Charalambakis and Garrett, 2019), some other studies have mainly focused only on the market factors (Merton 1974, Rees 2005). In this study, we attempt to examine the usefulness of a combined model by using the both accounting and market factors to evaluate their usefulness in predicting financial distress (see Campbell et al., 2008 and Tinoco and Wilson, 2013). Three, we attempt to identify some critical determinants of financial distress from a list of 34 initial factors to develop a parsimonious distress prediction model. Four, we attempt to estimate our empirical distress prediction models on two-time horizons, one year ahead as well as two year ahead, to compare the predictive accuracy of models on different time horizons. Five, we attempt to compare the predictive abilities of three forecasting techniques namely, binomial Logit, Artificial Neural Network (ANN) and Support Vector Machine (SVM). Finally, to the best of our knowledge, this is one of the first extensive efforts to develop a parsimonious distress prediction model for publicly listed non-financial companies in India. Hence, we attempt to contribute to the literature by providing some new evidence from an emerging economy.

The remainder of the paper is organized as follows: the next section provides a brief review of literature. Section three provides description of data and database along with the selection and construction procedure of variables used in the empirical analysis. Section four presents a detailed discussion on the empirical methodology adopted in the study. The empirical results are presented in the fifth section. And in the last section we provide summary and conclusions.

2. Review of Literature

Although the issue of financial distress has been studied extensively, empirical studies so far have only focused on the advanced economies and the empirical evidences from emerging economies are very limited in number and scope.[2] In this regard, important early studies on prediction of corporate financial distress include Beaver (1966), Atlman (1968) and Deakin (1972) that focused on the estimation of univariate or multivariate discriminative functions for a sample of distressed and healthy firms. Empirical findings of early studies collectively suggested that accounting ratios of failing firms are significantly different from those of healthy firms and accounting ratios can be useful in investigation and identification of financial distress. The financial ratios of distressed firms were found to be very poor compared to the healthy firms and all were facing unstable financial situations. For example, Beaver (1966) used univariate analysis to analyze the ability of accounting data for distress and bankruptcy prediction. This approach is based on the comparison of a financial ratio of interest with a benchmark ratio to distinguish between a failed and non-failed firm. Altman (1968) used the multiple discriminant analysis to constructed Z-score which is now widely used for predicting financial distress. Dambolena and Khoury (1980) used Multivariate Discriminate Analysis (MDA) to predict bankruptcy with

[2] To conserve space, we only provide a brief discussion and review of literature. See Bhattacharjee and Han (2014), Tinoco and Wilson (2013), Mselmi et al. (2017) and Charalambakis and Garrett (2019) for further discussion.

prediction accuracy of 87%, 85% and 78% from one, three and five years prior to bankruptcy, respectively.

Rees (1995) argued that information regarding market prices might be helpful in prediction of bankruptcy because they reflect ability of firms to generate cash flows. Shumway (2001) showed that accounting ratios employed in earlier work on bankruptcy were statistically insignificant whereas market variables like security returns are found to be highly correlated with bankruptcy. Christidis and Gregory (2010) tested the usefulness of market variables in predicting financial distress for quoted companies of UK and found that inclusion of market variables enhances the predictive ability of their model. Similarly, Chava and Jarrow (2004) showed that the power of predictive model can be enhanced by accounting for industry classification.

Tinoco and Wilson (2013) used ex-ante models for distress identification and prediction one and two years prior to the distress event for United Kingdomand highlighted the importance of a combined model in distress prediction. Most recently, Mslemi et al. (2017) examined the issue for a sample of French firms using recently developed machine learning based techniques. The results of the study indicated that for one-year prior to financial distress, SVM is the best classifier with an overall accuracy of 88.57%. Charalambakis et al. (2019) investigated the determinants of corporate financial distress by using a multi-period logit model and concluded that profitability, leverage, size and output growth rate have significant prediction power of financial distress for Greek firms.

In short, previous empirical studies have identified several determinants of financial distress with significant explanatory power and a large part of the available empirical evidences are for advanced countries. To the best of our knowledge, empirical studies for the Indian corporate sector are negligible. Therefore, in this study we attempt to develop an efficient distress prediction model for the Indian corporate sector by using most recent information on publicly listed Indian firms. Lastly, most of the previous studies have focused on the industry specific approach. But in the present study, we aim to perform an extensive empirical analysis of financial distress by covering companies across many industries covered in cooperate sector.

3. Data and Measurement of Variables

The information on the sample firms used in the study is taken from Prowess Database.[3] The study is based on the annual financial data of 1,957 companies listed on the National Stock Exchange (NSE). The primary reason behind considering companies listed on the NSE is that it accounts for the largest part of trading activity in India as compared to the Bombay Stock Exchange (BSE). Following the standard practice in the related literature, we excluded companies related to financial services from the sample because of their unique operating, financial and risk characteristics (see Bhattacharjee and Han, 2014; Mslemi et al., 2017). Hence, we initially included observations for a total of 1,742 non-financial publicly listed companies. The sample period of the study spans from 2010 to 2016. We select 2010 as the starting point of study in order to eliminate the impact of global financial crisis on our results (see Ahmed et al., 2013 for further discussion). Further, we restrict the sample period by the end of 2016 to avoid impact of the implementation of the Insolvency and Bankruptcy code (IBC) by the Government of Indian on May 28, 2016.The aftermath period witnessed sudden increase in number of firms that applied for the corporate insolvency resolution process (see Economic Survey, 2018 – 2019).[4] In order to enhance the reliability of the data, in each year, we excluded those firms from

[3]Prowess database is provided by the Centre for Monitoring Indian Economy (CMIE).
[4] The first case under the IBC was admitted by the National Company Law Appellate Tribunal (NCLT) on January 17, 2017 and the first insolvency resolution plan was approved on August 2, 2017. It is noteworthy that by February

the sample that were suspended by the stock exchange from trading. Thus, the actual sample size in each year is 1,742 less the number of companies suspended by the stock exchange in that particular year. This means that the number of companies varies (between 1681 to 1705) on year to year basis in our sample. Table 1 presents information on total number of companies in each year after eliminating financial firms and suspended companies.

[Insert Table 1 about here]

The selection of accounting ratios and market variables considered for empirical analysis is based on prior literature on financial distress (see Min and Lee, 2005; Chen, 2011; Mslemi et al., 2017). Initially we collected data on 34 variables (i.e. 29 financial ratios and 5 market variables) for empirical analysis. Further, because of inconsistencies or missing values in the data, or unavailability of data points for some of the years or high correlation between variable, a multi-stage refinement and elimination process is adopted to identify most critical determinants of financial distress. A list of all financial ratios and market variables initially selected for analysis is presented in table no. 2. We classify companies into different industries on the basis of Global Industry Classification Standard (GICS), Thomson Reuters, to test if classification of firms on the basis of industry can enhance the predictive power of the model. Further, we also include an industry dummy in the empirical analysis.

2019, that is, within 27 months of the implementation of the IBC, as many as 14,000 applications had been filed for initiation of corporate insolvency resolution process (CIRPs) (See Economic Survey, 2018-19).

[Insert Table 2 about here]

4. Empirical Methodology

4.1 Financial Distress Measures

The first issue in the empirical analysis of financial distress is the definition and measurement of financial distress. We need to clearly define financial distress using some easily observable financial measures or indicators that can be used for the classification of firms in financially distressed or healthy category. Previous studies have used different measures of financial distress for classification and distress prediction (see Altman, 1993; Allen and Saunders, 2003 for extensive reviews).

Given the fact that the government of India has recently implemented the IBC and many distressed firms are in the early stages of legal proceedings, we attempt to adopt a measure of distress that can be applied regardless of legal consequences of bankruptcy and liquidation process. Specifically, we aim to develop an accounting-based definition of financial distress by mainly focusing on the ability of firms to repay its debt and other financial obligations (see Asquith et al. 1994). Following Pindado et al. (2008), Hernandez and Wilson (2013) and Bhattacharjee and Han (2014), we considered the following two definitions of financial distress to classify firms in distressed and healthy category.

The first definition, hereafter called two factor based definition of financial distress, is based on two factors which focuses on the Interest Coverage Ratio and change in market capitalization.[5] Using this two factor definition a firm is classified as financially distressed: first, when its EBITDA is lower than its interest expenses in the year under consideration (i.e. ICR < 1) and second, when the firm experiences negative growth in market value in the same year. If the first condition is satisfied, it can be inferred that the profits generated by the firm from operations are not sufficient enough to cover interest expense. And if the second condition is satisfied, it implies that equity holders are likely to lose confidence in the firm which could be attributed to poor operational efficiency. Therefore, negative growth in market value can be interpreted as a sign that a firm is perceived negatively by its equity holders, and hence, a decline in the market value reflects that a firm is in the state of financial distress (see Tinoco and Wilson, 2013).

The second definition, hereafter called the three-factor based measure of financial distress, is formulated by using one additional criterion that is, change in total assets of a firm (see Bhattacharjee and Han, 2014). According to this definition, a firm is classified as financially distressed, if it satisfies the above mentioned two conditions as discussed under two-factor definition, and it also suffers from a negative growth in assets. We include this additional criterion with the aim to design a more appropriate classification of firms and for the reason that a low interest coverage ratio may arise when debt is used as a major source of finance. In such a situation, the assets of a firm must increase. And if it does not increase, it can be inferred that the debt borne by a firm is not employed in productive asset building. Any firm that fails to satisfy all the criteria in a given definition is considered to be a part of the middle group that is, neither distressed nor healthy, and hence, it is excluded from the analysis. For simplicity, in the remainder of this paper, the binary dependent variable constructed using both the definitions of

[5]Interest Coverage Ratio (ICR) that is defined as the EBITDA (Earnings before Interest, Tax, Depreciation and Amortization) divided by the Interest Charge. In literature, interest cover is a frequently employed as a measure of financial distress and an important determinant of bankruptcy (seeKam et al., 2008).

financial distress will be referred to as 'financial distress indicator' or dependent variable. Financially distressed firms are assigned value 1 and healthy firms are assigned value 0.The year wise details and process of elimination adopted to select the final sample of distressed and healthy firms classified using both the definitions of distress are presented in Table 3.

[Insert Table no. 3 & 4 about here]

The classification of firms according to both the definitions varies with different time horizons because of the fact that firms with missing data points are eliminated from the sample. Panel A of Table 3 presents the details of firms during the process of elimination at each stage when the dependent variable is defined using the 2-factor criteria. And panel B reports the same for the dependent variable defined using the 3-factor criteria. The actual number of firms classified as financially distressed and healthy (whose status is predicted using 1-year and 2-year lagged values of independent variables) is also reported in Table 3. The summary of firms identified and included in financially distressed and healthy category for one and two years ahead prediction are presented in table 4. It is clearly observable that the percentage of firms in distressed category is highest in the year 2011 using both the definitions of financial distress. For example, by using the 2-factor definition 53% firms are classified as financially distressed in the year 2011 whereas the 39% firms are classified as distress using the 3-factor definition during the same year.

4.2 Selection of independent variables

Previous studies have considered a range of potential independent variables for the analysis of financial distress. Based on their findings, we initially select 34 accounting ratios as potential explanatory variables to begin the empirical analysis (see Chen, 2011; Sun and Li, 2012; Min and Lee, 2005 and Mselmi et al., 2017). Out of the initial 34variables, there are 29 accounting ratios selected to cover six different aspects (or groups) of firm performance: profitability, liquidity, structure, activity, solvency and size. Each aspect of firm performance includes at least four or more financial ratios that provide some indication of firms' financial position. Previous studies have used different proxies for firm size. In this study, we consider an accounting-based variable for size namely log (total assets) and five market variables (MV 30 – MV 34) as discussed in table 2. The selection of final regressors is based on a stepwise cleaning and testing process. We first eliminate some of the variables mainly due to the non-availability of full data for a sample of firms. Out of 29 accounting ratios and 5 market variables chosen from prior studies (as shown in Table 2), 3 ratios (namely Stock Turnover ratio, No Credit Interval and Return on Net Worth) and 2 market variables (namely Residual Return and PE ratio) were eliminated due to numerous missing values. Second, given the possibility of high correlation among various ratios belonging to one group, we perform the pairwise correlation and the results are presented in Table 5. The accounting ratios that were found to be highly correlated with one or the other ratios in a group were eliminated. Out of the remaining 25 ratios and 4 market variables, we removed 4 ratios (Gross Profit Margin, Quick Ratio, Debt Ratio, and Equity to Total Assets Ratio) and one market variable (market capitalization) to avoid possibility of multicollinearity. Finally, we used the binary dependent variable and the remaining 24 predictor variables and industry dummy for further analysis. The details of all 24 predictor variables identified after the first stage of elimination and selection process are reported as Model – 1 in column 3 of Table 6.

[Insert Table 5 & 6 about here]

4.3 Distress Prediction Models

4.3.1 Binomial Logit Model

The next stage of empirical analysis involves further screening and elimination of variables using the logistic regression in order to identify the most critical determinants of financial distress. The starting point of the empirical analysis is the estimation of Model – 1 that takes into account all 24 variables (21 accounting ratios, 3 market variables and industry dummy) selected on the basis of correlation and missing values in the first stage of screening process (see Table 6, column 2 & 3). We now performed the logistic regression using the dependent binary variable constructed using the two-factor definition of financial distress. The logistic regression was estimated using the values of 1-year lagged regressors and industry dummies. Independent variables statistically insignificant at 20 percent level of significance were eliminated and variables that were significant at 20 percent level, in at least three or more years of the sample period, were considered for further investigation.[6] We call this relatively condensed model as Model – 2 which includes a total of 14 variables (i.e. 12 accounting ratios and 2 market variables) and the details are reported in column 4 & 5 of Table 6. This process helped us in identifying a more summarized model which now includes only the most important regressors from each category and has a relatively high explanatory power.

[6] In a multiple regression model, some of the regressors may be weakly associated with the financial distress and make a small but important contribution in distress prediction. Keeping this in mind, we consider the 20 percent significance level.

We again repeat the same process using the most significant variables within each group, as identified in model–2, and arrived at a more parsimonious final model (termed as Model–3) which includes only six highly significant regressors. It is noteworthy that none of the market variables, included as a proxy for size, is significant and hence we will focus only on log (total assets) as a proxy for size. This final model will be used for further analysis and distress prediction purpose. The estimated results of logistic regression for the final model are reported in Table 7. The signs of the estimated coefficients are consistent with theoretical expectations. Broadly, they suggest that a high ROCE (profitability), CFO/TL (liquidity), ATR (efficiency) and FA/TA (structure) lead to low probability of financial distress. Further, a high DE ratio (solvency) and large total assets (firm size)result in high probability of financial distress in the Indian corporate sector. While the sign of the solvency related coefficient is broadly on expected line indicating a high debt increases financial distress, the impact of firm size needs some discussion. The results suggest that a large firm size is associated with more financial distress. The positive coefficient of size variables is somewhat similar to the findings of Charalambous et al. (2020). This may be because of the fact that efficient management of large firms becomes difficult once the size of firms increases beyond some threshold level. It is also indicative of the fact that very large firms generally exhaust all profitable investment opportunities and new incremental investments can only deliver fewer returns compared to initial investment. Hence, additional capital is less efficiently deployed. The results (positive sign of *size* coefficient) also make sense in the Indian case as recent crisis of NPA is more of a large firm problem. We confirm our argument by comparing the mean size of distressed and healthy firms over the study period. The mean size (log of total assets) of distressed firms (3-factor based definition) is 9.19 which are higher than that of healthy firms which is 8.97. Further, in order to rule out the interaction effect leading to positive size coefficients (see Charalambous et al., 2020), we estimate a univariate model involving size (log of total assets) as only explanatory variable. The estimated results suggest that the coefficient of *size* is significantly positive and hence re-confirmed our findings of multivariate analysis. The results for distressed and healthy firms based on two-factor based definition are broadly similar.

[Insert Table 7 about here]

4.3.2 Other Predictive Models – SVM and ANN approaches

In the next stage of empirical analysis, we also consider the SVM and ANN for distress prediction along with the logistic model. For ANN estimation, we use multilayer perceptron method involving ten hidden layers.[7]The overall data, data source and the number of firms in the sample are the same as used in the logistic regression. However, we made some changes in the sample to deal with the problems created by data imbalance. It is noteworthy that the number of healthy firms far exceeds the number of distressed firms on year to year basis in our sample. Hence it is causing a serious data imbalance problem which is observable in Table 4. In the presence of severe data imbalance there is a high possibility that the forecasting accuracy of models may not be comparable and valid (see Kim et al., 2015 and Mslemi et al., 2017).

Keeping in mind the data imbalance problem, we created sub samples of financially healthy firms (i.e. majority class) in such a way that number of firms in each subsample for each

[7]The ANN and SVM models were estimated using the MATLAB (version 2015). The MATLAB allows us to choose the number of hidden layers in the ANN model, but the numbers of neurons (nodes) in the hidden layers are chosen automatically. For more details of ANN (Multi-layer perceptron) as well as SVM procedure see Mselmi et al. (2017) for further discussion.

year is individually close to the number of firms in the minority class (or distressed firms). For example, in 2014 (as per 2-factor classification for one-year-ahead predictions), 82 firms are classified as financially distressed and 828 firms are classified as healthy. To overcome this problem, we divided the larger sample containing healthy firms into subsamples by following a two-step process. First, we randomly divided the number of healthy firms by number of financially distressed firms without replacement (828/82) to calculate the number of subsamples (in our case approximately 10 subsamples). Second, subsamples are made as follows: If we create 10 subsamples, each one will contain 82.8 firms which is practically not possible. So, by rounding up this number, 9 subsamples of 83 firms are made while the remaining firms formed the 10^{th}subsamples. Similar process is adopted for each year in the study. Each year we estimate the models *N* times, where *N* is the number of sub-samples of healthy firms in that year. Each estimation involves a fixed sample of distressed firms and one matching sub-sample of healthy firm. The *N* prediction rates achieved from a model in a given year are then averaged to obtain the mean annual prediction rate for that model. As a part of the training process for both SVM and ANN, in each of *N* estimations; we use the 50% data for training and the remaining 50% for testing purpose. We then reverse the process by making our 50% test data as training data and 50% training data as test data. The estimation process ensures that our full data is used for both training as well as testing purpose. The mean prediction rate is then calculated for each sub-sample by taking the average of the two 50% - 50% training-testing samples. We further estimate the mean prediction rate for each year by taking the average of prediction rates for the *N* samples which is reported in tables. The feature selection method for selecting suitable features for classification of firms in healthy or distressed is based on Fisher discriminant ratio (FDR). We calculated FDR ratio for all the given parameters and selected six parameters with highest FDR ratios. The selected parameters are ROCE, ROA, ATR, GR, RE/TA and log(Closing). For comparing the predictive accuracy of all the three models, we re-estimated the logit model for each year with the balanced subsamples using the two-step estimation process mentioned above and finally computed its average accuracy (see Table 8). Both SVM and ANN have been estimated in two ways; one, using the factors identified by our logit model (pre-specified factor approach) and two, identified factors within the framework of SVM and ANN using the feature selection method with FDR (unspecified factor approach).[8]

5. Empirical Results

In this section, we discuss the empirical results of the study. The overall results are discussed in two stages. The first stage of analysis is related to the estimation results using the logistic model. The second stage involves estimation using the SVM and ANN model and comparison of prediction accuracies.

5.1 Binomial Logit based Distress Prediction Model

The year wise estimated coefficients using the logistic model for final model involving six highly significant determinants of distress are reported in Table 7. Our six-factor model is somewhat different from the Altman's distress prediction model which is popularly cited in prior research. Altman (1968) included measures of solvency, liquidity, operating efficiency, profitability and investment rates in his five-factor model. Although we find that the first four attributes of Altman's model also play an important role in distress prediction in the Indian context, but our optimal measures seem to differ from those suggested in Altman's work. For example, we find debt to equity ratio as a good measure of leverage while Altman measured solvency indirectly by

---

[8]Keeping the space constraint in mind, we do not discuss the SVM and ANN model in detail. See Mselmi et al. (2017) for further discussion.

employing price to book value. Further, Altman used Return on Assets and working capital to total assets as measures of profitability and liquidity, respectively. But our results suggest that ROCE and CFO/TL are better measures of profitability and liquidity for Indian corporate sector. Operational efficiency is a critical variable in both the studies and ATR seems to be its good measure. Altman (1968) additionally used an investment factor proxied by change in firm size. But it is noteworthy that recent studies have used asset growth as an important measure of investment rate of the firms. In addition, we also confirmed the role of asset structure (measured by fixed assets to total assets) in distress prediction. Finally, we included the sixth determinant of financial distress that measures firm size. We constructed this variable by taking natural log of total assets. Overall findings of the study clearly suggest that the critical determinants of financial distress and their measures may vary across developed and emerging markets like India. Hence, it is more suitable to use a country specific model for financial distress prediction and the commonly used Altman (1968) and other models may not be the best models for emerging markets.

For empirical analysis we used the determinants of all three model which are described in table 6 for predicting the status of firms (distressed or healthy) with one as well as two years lagged values of independent variables and the prediction accuracies are reported in Table 8. Our definitions helped us in firm classification and provide us the status for each firm that is healthy or distressed. The expected status of each firm is provided by the alternative models using endogenously determined threshold values. We refer a firm to be correctly classified, if its actual status matches with the expected status. If the actual and expected status does not match, we refer to it as an error. The prediction rates for any model are defined as the ratio of correctly classified cases to total cases. All the sample firms have been classified by using 2-factor and 3-

factor based definition as described in the paper. This provides us the actual number of healthy and distressed cases. The expected status of sampled firm (whether it is healthy or distressed) is provided by each predictive model. A company is classified correctly, if there is congruence in its actual and expected status. It is noteworthy that, in logistic regression if the output value is greater than a threshold value (0.5) we assign it a class 1 (i.e. distressed); else we assign it a class 0 (not distressed). The analysis using the neural network also has a threshold of 0.5. The output above or equal to 0.5 is identified with class 1 and output less than 0.5 is identified with class 0. In SVM, we take the output of the linear function and if that output is greater than 1, we classify it into class 1 and if the output is -1, we classify it into other class 0 (i.e. not distressed).

[Insert Table 8 about here]

The results, as reported in table 8, clearly indicate that average accuracy of Model–3 (based on 2-factor definition) is 88% for 1-year ahead prediction. The average accuracy declines marginally to 85% for 2-year ahead prediction using the same model. Similarly, the prediction based on Model–3 (based on 3-factor definition) for 1-year and 2-year ahead forecasting horizons yield average accuracy rates of 90% and 87%, respectively. Though the predictive accuracies of Model–1 and Model–2 are relatively high compared to Model–3, but given the fact model–1 includes 24 regressors whereas model–3 includes only six regressors, the resulting loss of predictive power is negligible. Therefore, we consider Model–3 to be the best model. Moreover, the predictive power of models based on the3-factor criteria of measuring distress is also better in the Indian context. Further, one year ahead predictions are better than two year ahead predictions which are similar to the findings of prior work (see Bhattacharjee and Han, 2014).

The results regarding average prediction accuracy for balanced sample are also presented in Table 8. Now in the light of the fact that Model–3 performed nearly as well as model-1 & 2, while still maintain the parsimonious nature, we decided to focus on Model–3 for further analysis. The prediction accuracies for both the time horizons are presented in the table. The average prediction accuracy, when the values of dependent variable are calculated using the 2-factor criteria, for one and two year ahead predictions are83 and 80, respectively. Similarly, average prediction accuracy, when values of dependent variable are computed on the basis of 3-factor criteria, for one and two years ahead is 86 and 79, respectively. Overall results suggest that there is a marginal decline in average accuracy after accounting for data imbalance problem. Hence, it may be concluded that the presence of data imbalance introduces distortions in forecasting results and erroneously leads to high forecasting accuracy. Further, in this study, we could not observe any significant industry patterns in financial distress as the industry dummies were insignificant in model estimation. Hence, we dropped the industry dummies while estimating Model–2 and3. On the close examination of misclassified cases, we observed that one third of such cases belong to the consumer discretionary sector which is experiencing high rate of disruptions globally as well as in India.[9]Hence, there is a need to identify a different set of accounting ratios and market variables that can correctly classify and accurately predict the financial distress of firms in the consumer discretionary industry. This is important in the sense that six critical determinants identified in this study and by other leading studies such as Altman (1986), Ohlson (1980) and Shumway (2001) will not be fully effective in predicting financial

[9]The sectors in the Indian economy currently facing massive disruptions in consumer discretionary sectors in India include financial services, information technology, communication and media, energy etc.

distress in consumer discretionary industry. This industry is experiencing large technological disruptions particularly with the advent of digital era.

5.2 Distress prediction using SVM and ANN and comparison of prediction accuracies

Now after clearly establishing that the six factors based parsimonious (model – 3) is relatively better in terms of prediction accuracy and data imbalance leads to overestimation of prediction accuracies, we used model – 3 and balanced data for distress prediction using the ANN and SVM models. In this sub section, we present and compare predictive efficiency of alternative forecasting techniques used in the study. For this purpose, we attempt to compare the predictive ability of the binomial logit model (with balanced sub-samples, see Table no. 8 – panel C) with two other machine learning based forecasting models namely SVM and ANN. For the estimation process using the SVM and ANN, we used 80 percent of the sample data for training and remaining 20 percent for testing.[10]The calculated prediction accuracy rates for all models are reported in Table 9. It is observable that machine learning based models perform better compared to the binomial logit model on both the forecasting horizons in all the cases. In the case of 2-factors based definition, the SVM technique achieved the highest prediction accuracy of 79.61 percent (for 1 year ahead predication) using the FDR based inputs and 77.77 percent (for 2 years ahead) using the pre-specified inputs. In the case of 3-factor based definition of financial distress and prediction over 1 year ahead horizon, again the SVM technique delivered the highest accuracy of 83.30 percent (using FDR based inputs) and ANN technique delivered the highest accuracy of 76.67 percent for 2 years ahead prediction (using FDR based inputs). The prediction superiority of SVM technique is clearly established in three out of four

[10] We are grateful the anonymous referee for suggesting this methodological improvement in splitting the data for training and testing purpose.

different empirical specifications as presented in Table 12. In short, the results suggest that the machine learning based models outperformed the binomial logit model. While the logit model based predictions delivered the accuracy of 81.44 percent between two different time horizons, the machine learning based models delivered the highest accuracy (i.e. SVM – 83.60 percent). Based on these findings, it can be concluded that the machine learning based models (i.e. ANN and SVM) have the predictive superiority over the binomial model. The superior performance of machine learning models in distress prediction is consistent with the findings of previous studies (see Mselmi et al., 2017 and references therein). These models can be used for financial distress prediction in an emerging market economy like India.

[insert Table 9 about here]

6. Summary and Conclusions

In this study, we aimed to examine the critical microeconomic (or firm specific) determinants of financial distress and attempt to develop a parsimonious distress prediction model based on some easily observable micro indicators of distress. Although there is a huge literature covering different theoretical and empirical aspects of financial distress, but very little is known about what determines the probability of corporate financial distress, especially in an emerging economy like India. Therefore, in this study, we attempt to bridge this gap by examining the probability of financial distress for a relatively large sample of listed firms from the Indian corporate sector. Further, we also attempt to compare the forecasting accuracies of competing distress prediction techniques to identify the most suitable technique in terms of predictive power. In order to identify a more appropriate measure of financial distress, we used two measures to classify firms in distressed and healthy category.

The main findings of the study could be summarized in the following point. First, out of the initial list of 34 firm-specific factors, the results suggest that six variables play statistically significant role in determining the probability of financial distress. These six critical determinants of corporate distress include ROCE, CFO/TL, ATR, DE, FA/TA and log (TA). Second, our three-factor based measure of financial distress appears to be more suitable way of defining distress as prediction accuracies of three factor-based definition are higher than the two factor-based definition. Three, our findings suggest that machine learning based models namely SVM and ANN are superior in terms of their prediction accuracy compared to the simple binomial logit model even in a relatively not so large time series data set. On average, the SVM technique achieved the highest prediction accuracy in three out of four empirical specifications and ANN model performed better in one specification. This result is in line with the findings of Mselmi et al. (2017). Four, the prediction accuracies of SVM and ANN models are better when inputs are selected automatically using the FDR. Five, as expected, the predictive accuracies of the all models declined with increase in forecasting horizon which is similar to the findings of Charalambakis and Grarrett, (2019).

The findings of the study have some important practical implications for creditors, policymakers, regulators other stakeholders. First, rather than monitoring and collecting information on a list of predictor variables, only six most important accounting ratios maybe monitored to track the transition of a healthy firm into financial distress. Second, our six-factor model can be used to devise a sound early warning system for corporate financial distress. Three, machine learning based distress prediction models have prediction accuracy superiority over the commonly used time series model in the available literature for distress prediction involving a binary dependent variable. Four, our findings suggested that a large part of misclassified cases are concentrated in consumer discretionary sector. Hence, it can be argued that our models and other similar models, generally used in the available literature, may be not be efficient in predicting financial distress of firms in the consumer discretionary industry. Therefore a different set of explanatory variables needs to be identified for understanding the distress dynamics of this sector. Finally, we used the most recent available data but restricted our sample to cover the post global financial crisis till the implementation of insolvency and bankruptcy code (IBC) in India. Once the bankruptcy code is implemented effectively and the numbers of pending cases are reduced to minimum, studies can take a longer data set and can reexamine the performance of prediction techniques in future. Also, a detelied sectoral or industry wise study will help in uncovering any industry specific pattern in financial distress in the Indian economy.

References

Agarwal, V., &Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*.

Allen, L., & Saunders, A. (2003). A survey of cyclical effects in credit risk measurement models. Technical report, BIS Working Paper No. 126, the Bank for International Settlements.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction corporatebankruptcy. *Journal of Finance*.

Altman, E. I. and Edith Hotchkiss (2006), Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt, John Wiley & Sons, Inc., Hoboken, New Jersey

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

Bernard, A.B., and J.B. Jensen. 1999. Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics* 47: 1–25.

Bhattacharjee, A. and J. Han, 2014. Financial distress of Chinese firms: Microeconomic, macroeconomic and institutional influences. *China Economic Review*, 30: 244-262.

Blum, M. (1974). Failing company discriminant analysis. *Journal of Accounting Research*, 12(10), 1-25.

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

Carlos *López-Gutiérrez,* Sergio Sanfilippo-Azofra and BegoñaTorre-Olmo (2015), Investment decisions of companies in financial distress, *Business Research Quarterly*, Volume 18, Issue 3, July–September 2015, Pages 174-187

Carpenter, R., Fazzari, S. and Petersen, B. (1994). 'Inventory investment, internal finance fluctuation, and the business cycles', *Brookings Papers on Economic Activity*, 2, pp. 75–122.

Charalambakis, E. C. and Ian Garrett (2019), On corporate financial distress prediction: What can we learn from private firms in a developing economy? Evidence from Greece, *Rev Quant Finan Acc* 52:467–491.

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

Chen, M. Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38(9), 11261–11272

Christidis, A., & Gregory, A. (2010). Some new models for financial distress prediction in the UK. Xficentre for finance and investment discussion paper no. 10

Dambolena, I. G., & Khoury, S. J. (1980). Ratio stability and corporate failure. *Journal of Finance*, 35(4), 1017-1026.

Deakin, E. B. "A Discriminant Analysis of Predictors of Failure," *Journal of Accounting Research*, 10 (Spring 1972): 167-79.

Economic Survey (2017-18), Government of India, Ministry of Finance, Department of Economic Affairs, Economic Division, New Delhi.

Economic Survey (2018-19), Government of India, Ministry of Finance, Department of Economic Affairs, Economic Division, New Delhi.

Edminster, R. (1972). "An empirical test of financial ratio analysis for small business failure prediction", *Journal of Financial and Quantitative Analysis*, pp. 1477-1493.

Fazzari, S. R., Hubbard, G. and Peterson, B. (1988). 'Financing constraints and corporate investment', *Brookings Papers on Economic Activity*, 1, pp. 141–95.

Gilchrist, S. and Himmelberg, C. P. (1995). 'Evidence on the role of cash flow for investment', *Journal of Monetary Economics*, 36(3), pp. 541–72.

Harada, N., &Kageyama, N. (2011). Bankruptcy dynamics in Japan. *Japan and the World Economy*, 23(2), 119-128.

Haralambous, C, Spiros H. Martzoukos & Zenon Taoushianis (2020): Predicting corporate bankruptcy using the framework of Leland-Toft: evidence from U.S.,QuantitativeFinance,Volume 20 (2), pp. 329-346.

Hotchkiss (2006), Corporate Financial Distress & Bankruptcy, 3rd edition, John Wiley.

Ismael G. Dambolenaand Sarkis J. Khoury (1980), Ratio Stability and Corporate Failure, *The Journal of Finance*, Vol. 35, No. 4 (Sep., 1980), pp. 1017-1026.

Jackendoff, N. (1962). A study of published industry financial and operating ratios. Philadelphia: Temple University, *Bureau of Economic and Business Research*.

Kam, A., Citron, D., &Muradoglu, G. (2008). Distress and restructuring in China: Does ownership matter? *China Economic Review*, 19, 567–579.

Keasy K. and P. McGuiness, "The Failure of UK Industrial Firms 1976-1984, Logistic Analysis and Entropy Measures", *Journal of Business Finance and Accounting* (vol. 17 1990), pp. 119- 136.

Kim, M. J., Kang, D. K., & Kim, H. B. (2015). Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications*, 42(3), 1074–1082

Liu, J. and Wilson, N. (2002). 'Corporate failure rates and the impact of the 1986 Insolvency Act: an econometric analysis', *Managerial Finance*, 28(6), pp. 61–71.

Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603–614.

Mslemi N., Lahiani A., Hamza T. (2017), 'Financial Distress prediction: The case of French small and medium-sized firms', *International Review of Financial Analysis*.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*.

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

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

Rees, W. (1995), *Financial Analysis*, 2nd edition, Prentice Hall.

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

Tinoco M.H. and Wilson N. (2013), Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *InternationalReview of Financial Analysis*.

Vlieghe, G. W. (2001). ‘Indicators of fragility in the UK corporate sector’, *Bank of England Working Paper No. 146*.

Wadhwani, S. B. (1986). ‘Inflation, bankruptcy, default premia and the stock market’, *Economic Journal*, 96(381), pp. 120–38.

Wasim Ahmad, Sanjay Sehgal and N.R. Bhanumurthy (2013), Eurozone crisis and BRIICKS stock markets: Contagion or market interdependence? *Economic Modelling* 33 (2013) 209–225.

Wilson, R. L. & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision support systems*, Volume 11, Issue 5, June 1994, Pages 545-557.

Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research, Vol. 22, Studies on Current Econometric Issues in Accounting Research (1984), pp. 59-82

Appendix – 1

Table of Results

Table 1: Year wise information of total firms in the sample

	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Total firms listed on NSE	1957	1957	1957	1957	1957	1957
(Less)financial firms	215	215	215	215	215	215
Total non-financial firms listed on NSE	1742	1742	1742	1742	1742	1742
(Less) suspended firms	40	37	38	39	49	61
Number of firms after suspension	1702	1705	1704	1703	1693	1681

Notes: (i) This table presents information about the total number of firms arrived at after elimination of financial firms and suspended firms in each year from 2011 to 2016, (ii) NSE stands for National Stock Exchange.

Table 2: Summary of accounting ratios and market variables initially selected as determinants of financial distress

Category	Code Name	Ratio	Formula	Category	Code Name	Ratio	Formula
Profitability	P1	NP margin	Net Profit/Net Sales	**Activity**	A18	Creditor days	Creditors/Operating Revenue*360
	P2	GP margin	Gross Profit/Net Sales		A19	Debtor days	Debtors/Operating Revenue * 360
	P3	RNW	Net profit/Shareholders Funds		A20	ATR	Sales Revenue/Total Assets
	P4	ROCE	Net Profit/Capital Employed		A21	STR	Operating Revenue/Stock
	P5	ROA	Net Profit/Total Assets	**Solvency**	SL22	Financial Leverage	Long-term Liabilities/Total Assets
Liquidity	L6	QR	Quick Assets/ Current Liabilities		SL23	GR	Long-Term Liabilities/Capital Employed
	L7	CR	Current Assets/Current Liabilities		SL24	debt ratio	Total Debt/Total Assets
	L8	NCI	(Quick Assets-Current Liabilities)/Daily Operating Expenses		SL25	SR	PAT + Depreciation/ (Long-Term Liabilities + Short-Term Liabilities)
	L9	CFO/TL	Total Cash from Operations/Total Liabilities		SL26	Repayment capacity	Financial Debt/cash Flow
	L10	CF/TA	Cash Flow/Total Assets		SL27	DE	Debt/Equity
	L11	CF/OR	Cash Flow/Operating Revenue		SL28	RE/TA	Retained Earnings/Total Asset
	L12	WC/TA	Working Capital/Total Assets	**Size**	SZ29	ln(TA)	ln(Total Assets)
Structure	S13	Eq/TA	Equity/Total Asset		MV30	ln(Closing Price)	ln(Closing Price)
	S14	CA/TA	Current Assets/Total Assets		MV31	ln(Market Capitalisation)	ln(Market Capitalisation)
	S15	CL/TL	Current Liabilities/Total Liabilities		MV32	P/E	Price/Earnings Per Share
	S16	FA/TA	Fixed Assets/Total Assets		MV33	P/B	Price/Book Value Per Share
	S17	SF/NCL	Shareholders' Funds/Non-Current Liabilities		MV34	RESIDUAL RET'10	Cumulative monthly security return minus cumulative monthly NSE500 index return

Notes:(i) This table presents details of 34 financial variables (accounting ratios and market variables) initially selected for empirical analysis. (ii) where the code names of ratios indicate category and the number of variables. For example, code name P1 and MV34 indicates 'category – Profitability (P) and variable no. 1, and category – Market Variable (MV) and variable no. 34, respectively, and so on.

Table 3: Calculation of total annual observations considered for prediction

Panel – A (Narrow definition or 2-factor based definition of financial distress)						
Years	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Number of firms after suspension (see raw 6 of table no. 1)	1702	1705	1704	1703	1693	1681
Missing values of dependent variable	582	508	488	484	474	374
Unclassified firms	886	319	675	269	404	603
Total firms classified as distressed/healthy	**234**	**878**	**541**	**950**	**815**	**704**
Less Missing values of independent variable (for 1-year ahead)	38	36	44	40	74	71
Total firms (For 1 year ahead forecasting)	196	842	497	910	741	633
Total firms classified as distressed/healthy	**234**	**878**	**541**	**950**	**815**	**704**
Less Missing values of independent variable (for 2-year ahead)	63	126	41	41	65	89
Total firms (for 2 year ahead forecasting)	171	752	500	909	750	615
Panel – B (Broad definition or 3-factor based definition of financial distress)						
Number of firms after suspension	1702	1705	1704	1703	1693	1681
Missing values of dependent variable	582	508	488	484	474	450
Unclassified firms	948	493	797	556	663	707
Total firms classified as distressed/healthy	**172**	**704**	**419**	**663**	**556**	**524**
Less Missing values of independent variable (for 1-year ahead)	27	25	26	28	47	40
Total firms (For 1 year ahead forecasting)	145	679	393	635	509	484
Total firms classified as distressed/healthy	**172**	**704**	**419**	**663**	**556**	**524**
Less Missing values of independent variable (for 2-year ahead)	43	104	29	33	41	60
Total firms (for 2 year ahead forecasting)	129	600	390	630	515	464

Note: (i) Panel A presents the details and computation of total number of firms considered for prediction when values of dependent variable are computed using the narrow measure of financial distress or the 2-factor criteria; (ii) Panel B presents the details and computation of total number of firms considered for prediction when values of dependent variable are computed using the broad measure of financial distress or the 3-factor criteria.

Table 4: Summary statistics for annual observations in each category (Financially distressed & Financially healthy)

Panel A: Classification according to 2-factor definition												
	t-1						t-2					
Years	2011	2012	2013	2014	2015	2016	2011	2012	2013	2014	2015	2016
Financially Distressed	104	75	159	82	79	122	85	73	175	84	97	132
Financially Healthy	92	767	338	828	662	511	86	679	325	825	653	483
TOTAL	196	842	497	910	741	633	171	752	500	909	750	615
% of financially distressed firms	53%	9%	32%	9%	11%	19%	50%	10%	35%	9%	13%	21%
Panel B: Classification according to 3-factor definition												
Financially Distressed	56	44	100	57	47	78	50	42	109	59	60	87
Financially Healthy	89	635	293	578	462	406	79	558	281	571	455	377
TOTAL	145	679	393	635	509	484	129	600	390	630	515	464
% of financially distressed firms	39%	6%	25%	9%	9%	16%	39%	7%	28%	9%	12%	19%

Notes: The table reports classification of firms into financially distressed and healthy. The last row of each table shows the percentage of financially distressed firm out of total firms.

Table 5: Correlation Matrices and Multicollinearity Diagnosis

Panel A: Correlation matrix for Profitability Ratios							
Variables (Code name)	**P2**	**P1**	**P5**	**P4**			
P2	1						
P1	0.5643	1					
P5	0.1969	0.1887	1				
P4	0.1809	0.1652	0.9474	1			
Panel B: Correlation Matrix for Liquidity Ratios							
Variable	**L11**	**L10**	**L7**	**L9**	**L6**	**L12**	
L11	1						
L10	0.0048	1					
L7	-0.0097	0.1551	1				
L9	0.0491	0.0597	-0.0389	1			
L6	-0.0159	0.2094	0.9208	-0.0284	1		
L12	-0.0197	0.2193	0.3125	-0.0362	0.3051	1	
Panel C: Correlation Matrix for Solvency ratios							
Variable	**SL27**	**SL22**	**SL24**	**SL23**	**SL26**	**SL28**	**SL25**
SL27	1						
SL22	-0.0137	1					
SL24	0.0117	0.5576	1				
SL23	0.0017	-0.0299	-0.0878	1			
SL26	0.0002	0.0623	0.0216	0.0011	1		
SL28	-0.0006	-0.0158	-0.0243	0.0009	0.0002	1	
SL25	-0.0012	-0.0645	-0.1113	0.0015	0.0157	0.0042	1
Panel D: Correlation Matrix for Activity Ratios							
Variable	**A20**	**A18**	**A19**				
A20	1						
A18	-0.0575	1					
A19	-0.0701	0.4792	1				
Panel E: Correlation Matrix for Structural Ratios							
Variable	**S14**	**S15**	**S13**	**S16**	**S17**		
S14	1						
S15	0.2543	1					
S13	-0.0114	-0.8448	1				
S16	-0.3721	0.2091	-0.3567	1			
S17	0.0071	-0.0359	0.0539	-0.0238	1		
Panel F: Correlation Matrix for Market Variables							
Variable	**MV29**	**MV30**	**MV34**	**MV32**			
MV29	1						
MV30	0.6817	1					
MV34	0.3715	0.8139	1				
MV32	0.4445	0.4362	0.0931	1			

Note: See Table 2 (column no. 2 & 6) for the category wise names of variables for their respective code names.

Table 6: List of variables included in models

	Model -1		Model – 2		Model – 3	
	(21 accounting ratios, 3 market variables and industry dummy)		(12 accounting ratios and 2 market variables)		(6-factor model)	
Category	Code Name	predictor variable	Code Name	predictor variable	Code Name	predictor variable
Profitability	P1	NP margin	P4	ROCE	P4	ROCE
	P4	ROCE	P5	ROA		
	P5	ROA				
Liquidity	L7	CR	L9	CFO/TL	L9	CFO/TL
	L9	CFO/TL	L10	Cash/TA		
	L10	Cash/TA	L11	CFO/OR		
	L11	CFO/OR				
	L12	WC/TA				
Activity	A18	Creditor days	A18	Creditor Days	A20	ATR
	A19	Debtor days	A19	Debtor Days		
	A20	ATR	A20	ATR		
Solvency	SL22	Financial Leverage	SL27	DE	SL27	DE
	SL23	GR				
	SL25	SR				
	SL26	Repayment capacity				
	SL27	DE				
	SL28	RE/TA				
Structure	S14	CA/TA	S14	CA/TA	S16	FA/TA
	S15	CL/TL	S15	CL/TL		
	S16	FA/TA	S16	FA/TA		
	S17	SF/NCL				
Size	MV30	ln(Closing Price)	MV33	P/B	SZ29	ln(TA)
	MV33	P/B	SZ29	ln(TA)		
	SZ29	ln(TA)				

Table 7: Estimation results of logistic regression

Panel A (2-factor based definition of financial distress)												
Years	**2011**		**2012**		**2013**		**2014**		**2015**		**2016**	
Variables	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*
ROCE	-0.167***	-0.148***	-0.169***	-0.096***	-0.181***	-0.147***	-0.104***	-0.031***	-0.19***	-0.068***	-0.176***	-0.119***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.041)	(0.00)	(0.00)	(0.00)	(0.00)
CFO/TL	-5.853***	-5.972***	-2.729***	-2.528**	-0.535	-2.964***	-2.201***	-5.148***	0.954	-1.531***	-2.488*	0.498
	(0.009)	(0.010)	(0.042)	(0.097)	(0.718)	(0.019)	(0.00)	(0.001)	(0.664)	(0.002)	(0.180)	(0.759)
ATR	-0.432	-0.057	-0.119	-0.032	-0.434**	0.068	-1.447***	-1.079***	-0.921***	-0.884***	-0.89***	-0.715***
	(0.235)	(0.879)	(0.634)	(0.891)	(0.087)	(0.645)	(0.00)	(0.00)	(0.007)	(0.001)	(0.005)	(0.002)
DE	0.467***	0.0006	-0.004	0.166***	0.084**	-0.0009	0.058**	0.143***	0.057***	0.029**	0.0005	0.067***
	(0.00)	(0.968)	(0.614)	(0.005)	(0.053)	(0.878)	(0.081)	(0.005)	(0.011)	(0.097)	(0.820)	(0.034)
FA/TA	0.267	0.267	-1.012***	-1.048***	-2.184***	-2.076***	-1.494***	-1.429***	-1.207***	-1.627***	-0.951***	-1.927***
	(0.733)	(0.714)	(0.048)	(0.039)	(0.00)	(0.00)	(0.004)	(0.007)	(0.023)	(0.00)	(0.025)	(0.00)
ln(TA)	-0.138	0.031	-0.107	-0.213***	0.08	0.106*	0.097	0.099	0.318***	0.251***	0.165***	0.127**
	(0.396)	(0.845)	(0.317)	(0.042)	(0.366)	(0.182)	(0.271)	(0.233)	(0.00)	(0.001)	(0.039)	(0.074)
Panel B (3-factor based definition of financial distress)												
ROCE	-0.165***	-0.199***	-0.240***	-0.095***	-0.226***	-0.178***	-0.152***	-0.022*	-0.265***	-0.114***	-0.225***	-0.132***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.148)	(0.000)	(0.000)	(0.000)	(0.000)
CFO/TL	-8.743***	-2.510	-4.093***	-3.030*	0.165	-2.942***	-2.048***	-5.082***	6.809***	-5.580***	-1.648	0.771
	(0.005)	(0.403)	(0.016)	(0.124)	(0.922)	(0.041)	(0.004)	(0.008)	(0.022)	(0.023)	(0.482)	(0.697)
ATR	0.419***	0.102	0.441*	0.302	-0.195	0.209*	-1.397***	-0.987***	-0.631*	-0.675***	-0.636**	-0.663***
	(0.007)	(0.804)	(0.194)	(0.32)	(0.453)	(0.165)	(0.000)	(0.002)	(0.15)	(0.047)	(0.095)	(0.014)
DE	2.271***	-0.001	-0.003	0.221***	0.103***	0.019	0.152**	0.262***	0.037	0.040	-0.000	0.129***
	(0.043)	(0.95)	(0.748)	(0.005)	(0.050)	(0.232)	(0.054)	(0.001)	(0.355)	(0.4)	(0.915)	(0.006)
FA/TA	-0.591*	0.898	-1.504**	-1.125*	-2.105***	-2.003***	-1.588***	-1.085**	-1.879***	-1.197**	-0.357	-1.606***
	(0.196)	(0.288)	(0.065)	(0.116)	(0.000)	(0.000)	(0.013)	(0.099)	(0.009)	(0.061)	(0.524)	(0.002)
ln(TA)	0.279*	0.061	0.239*	0.414***	-0.014	-0.037	0.085	0.062	-0.133	-0.190**	-0.121	-0.105
	(0.185)	(0.752)	(0.121)	(0.006)	(0.896)	(0.7)	(0.472)	(0.542)	(0.29)	(0.064)	(0.228)	(0.223)

Notes:(i) This table reports the results of logistic regression of the binary dependent variables on predictor variables. Models were computed for two-time frames, one in which the predictor variables assume a year prior values (from the event of financial distress) and the other in which predictor variables assume two-year prior values. (ii) * denotes significant at 20%, ** denotes significant at 10% and *** denotes significant at 5%, (iii) Values in (#) are *p*-values.

Table 8: Prediction accuracy of logit model with unbalanced and balanced subsamples

Years	1-year-ahead						Average	2-year-ahead						Average
	2011	2012	2013	2014	2015	2016		2011	2012	2013	2014	2015	2016	
Panel A (2-factor based definition of financial distress)														
Model – 1	82%	91%	84%	92%	91%	87%	88%	82%	91%	84%	92%	91%	87%	88%
Model – 2	85%	93%	87%	93%	93%	88%	90%	77%	91%	81%	91%	91%	83%	86%
Model – 3	82%	93%	84%	93%	91%	87%	88%	75%	91%	78%	91%	88%	84%	85%
Panel B (3-factor based definition of financial distress)														
Model – 1	92%	97%	90%	95%	95%	92%	94%	83%	94%	85%	92%	93%	89%	89%
Model – 2	85%	95%	89%	93%	94%	90%	91%	81%	94%	84%	92%	92%	87%	88%
Model – 3	85%	95%	87%	93%	94%	89%	90%	81%	94%	82%	90%	90%	86%	87%
Panel C (Prediction accuracy of logit model with balanced subsamples)														
Model – 3(A)	82%	81%	84%	85%	85%	84%	83%	75%	76%	88%	76%	87%	87%	80%
Model – 3(B)	87 %	86%	85%	86%	87%	87%	86%	81%	77%	78%	77%	81%	82%	79%

Notes: This table presents financial distress prediction accuracy of logistic regression one and two years before financial distress in each year. Panel A shows the results obtained for Model-1 (full model – 24 factor), Model-2 (14-factor model) and Model-3 (6-factor model) using 2-factor or narrow definition of financial distress; (ii) Panel B shows the results obtained using 3-factor or broad definition of financial distress; (iii) Model – 3(A) and Model – 3(B) indicate the final parsimonious model with 6 regressors estimated for two factor (‘A’) and three factor (‘B’) based definition of financial distress (i.e. the binary dependent variable).

Table 9: Comparison of prediction accuracies of alternative forecasting techniques**

			Panel A (1 year ahead)							Panel B (2 year ahead)						
Definition of financial distress	Techniques		2011	2012	2013	2014	2015	2016	Average Accuracy	2011	2012	2013	2014	2015	2016	Average Accuracy
2-factor	Logit (Model – 3)	**Six**	71.011	72.00	76.56	86.18	71.09	81.5	76.39	55.88	71.48	70.00	67.65	73.75	72.22	68.49
	SVM	**Six**	73.68	79.00	78.47	83.53	73.99	75.61	77.38	76.67	77.40	78.46	79.00	77.30	77.81	77.77
		FDR	76.32	81.00	80.47	81.47	73.44	85.00	79.61	64.70	74.07	75.41	72.22	78.33	78.40	73.85
	ANN	**Six**	50.00	79.00	77.34	83.53	70.31	78.5-	73.11	64.71	68.89	72.86	70.59	72.50	74.69	70.70
		FDR	50.00	79.67	79.69	81.47	73.44	85.50	74.96	76.46	76.30	77.14	69.60	76.67	76.54	75.45
3-factor	Logit (Model – 3)	**Six**	75.00	83.33	83.75	86.25	77.22	83.13	81.44	75.00	71.70	70.45	73.61	67.26	79.17	72.86
	SVM	**Six**	74.91	83.73	85.16	82.94	77.4	78.25	80.39	78.5	73.41	73.48	75.9	73.51	75.20	75.00
		FDR	79.17	86.9	81.25	81.67	83.89	88.75	83.60	65.00	82.05	73.86	67.59	79.76	81.25	74.91
	ANN	**Six**	79.17	79.76	83.75	80.83	75.56	81.87	80.16	75.00	68.80	70.45	71.76	67.26	75.69	71.49
		FDR	79.17	82.54	85.00	79.17	83.89	86.25	82.67	75.00	75.64	75.00	71.76	78.57	84.03	76.67

Notes:(i) This table reports financial distress prediction accuracy (for prediction of 1 and 2-years ahead status of a firm) of all the models considered for final comparison; (ii) Prediction accuracy (in percent) obtained for classification as per 2-factory and 3-factor definition is displayed for each of these models. (iii) ** indicates models estimated using 80% training and 20% testing sample data.

On the Determinants and Prediction of Corporate Financial Distress in India

Sanjay Sehgal[1], Ritesh Kumar Mishra[2], and Rupali Vashisht[3]

Abstract

Purpose – The main aim of the study is to identify some critical microeconomic determinants of financial distress and to design a parsimonious distress prediction model for an emerging economy like India. In doing so, we also attempt to compare the forecasting accuracy of alternative distress prediction techniques.

Design/methodology/approach – In this study, we use two alternatives accounting information-based definitions of financial distress to construct a measure of financial distress. We then use the binomial logit model and two other popular machine learning based models, namely Artificial Neural Network and Support Vector Machine, to compare the distress prediction accuracy rate of these alternative techniques for the Indian corporate sector.

Findings – Our empirical results suggest that five financial ratios, namely return on capital employed, cash flows to total liability, asset turnover ratio, fixed assets to total assets, debt to equity ratio and a measure of firm size (log total assets) play highly significant role in distress prediction. Our findings suggest that machine learning based models namely SVM and ANN are superior in terms of their prediction accuracy compared to the simple binomial logit model. Results also suggest that one year ahead forecasts are relatively better than the two year ahead forecasts.

[1]Professor, Department of Finance and Business Economics, University of Delhi, India. Email: sanjayfin15@yahoo.co.in

[2]Assistant Professor (Economics), Department of Finance and Business Economics, University of Delhi, India. Email: ritesh.mishra@mfc.edu (Corresponding Author).

[3] Research Scholar, Department of Financial Studies, University of Delhi, Delhi, India, 110021

Acknowledgement – We gratefully acknowledge the financial support provided by the Ministry of Corporate Affair, Government of India for conducting this study. The authors also thank Dr Ajay Jaiswal, Assistant Professor, SSCBS, and University of Delhi for his technical assistance in the estimation of SVM and ANN models.

Originality/value – This study is one of the first comprehensive attempts to investigate and design a parsimonious distress prediction model for the emerging Indian economy which is currently facing high levels of corporate financial distress. Unlike the previous studies, we use two different accounting information-based measures of financial distress in order to identify an effective way of measuring financial distress. Some of the determinants of financial distress identified in this study are different from the popular distress prediction models used in the literature. Our distress prediction model can be useful for the other emerging markets for distress prediction.

Keywords: Financial distress prediction, Logit Model, Support Vector Machine, Artificial Neural Networks, Corporate Profitability

JEL Classification: G32, G33, C45

1. Introduction

The main aim of the study is to identify some critical microeconomic determinants of financial distress and to design a parsimonious distress prediction model for an emerging economy like India. The issue of financial distress and the case of Indian corporate sector are important and interesting for the fact that Indian economy is facing a somewhat perplexing situation best described as the coexistence of a relatively high growth rate and considerably high financial distress in the corporate sector. The present study is an attempt to provide some new evidence on financial distress and contribute to the existing literature by using a comprehensive dataset of publicly listed non-financial companies in India.

Corporate financial distress has now become a worrying economic reality for policy makers both in advanced as well as emerging economies. The level of corporate distress in the post global financial crisis period has increased to the extent that sometimes investors and lenders look suspicious about the old adage "too big to fail" (Altman and Hotchkiss, 2006). This is also appearing to be true for India in the sense that many big business firms have expressed their inability to repay their debts in the recent past. In short, financial distress refers to a situation in which a firm's cash flows are not sufficient enough to meet contractually required payment obligations. There are large direct and indirect costs of financial distress and relatively high levels of distress can destabilize the overall financial system by gradually impairing the balance sheet of lending institutions (Economic Survey, 2017-2018).

In normal times, the birth and death of firms are in fact not a completely unnatural phenomenon which demands immediate policy intervention. It is rather considered to be a part of the overall economic process governing the production, distribution and consumption activities during different phases of the business cycles. During the normal progression of economic activities, firms take birth, gradually grow and become unicorns and many a times they exit the production process without significantly affecting the overall production. But sometimes the unusual exit of firms due to financial distress imposes huge direct and indirect cost on the economy in terms of output, employment, demand and revenue (see Altman and Hotchkiss, 2006, pp. 93). Further, it also leads to under investment and misallocation of resources as distressed firms have the tendency to under-invest by only focusing on some investment project that will only help them in avoiding bankruptcy (López-Gutiérrez et al. (2015). Hence, it is of great policy importance to investigate and understand the dynamics of financial distress by focusing on an emerging economy like India.

At present, the Indian economy, one of the fastest growing large economies in the world, is passing through a tough business environment. Unlike many advanced and emerging countries, the Indian economy has witnessed impressive growth with the gross domestic product (GDP) consistently above 7 percent per annum between 2011 and 2017 with some moderation in 2018–2019. However, notwithstanding with this impressive growth rates, many firms in the corporate sector have expressed their inability to service their debt and revealed severe financial distress in their respective balance sheets. The level of financial distress has increased considerably and now it has starting impacting the balance sheet of lending institutions.[4] In response to the growing financial distress, the government of India implemented the Insolvency and Bankruptcy Code, 2016 (IBC). After its implementation around 14,000 applications had been filed within first 27 months for initiation of Corporate Insolvency Resolution Process (CIRPs) by February 2019 (see Economic Survey, 2017-18). This clearly indicates the seriousness of financial distress in the Indian corporate sector where a large number of firms are waiting to exit. Therefore, aclear understanding of distress dynamics of the Indian corporate sector and identification of key determinants of financial distress can be useful in developing a sound distress prediction system.

Against this background, the present study attempts to contribute to the existing literature in multiple ways. First, building on the findings of previous studies and moving a step further, we considered two different measures of financial distress to classify firms in distressed and healthy categories. This exercise will help in identifying a more effective way of measuring financial distress for an emerging market like India. Second, while some of the previous studies have focused on establishing a relationship between financial distress and accounting ratios (see for example, Altman 1968, Mselmi, 2017, Charalambakis and Garrett, 2019), some other studies have mainly focused only on the market factors (Merton 1974, Rees 2005). In this study, we attempt to examine the usefulness of a combined model by using the both accounting and market factors to evaluate their usefulness in predicting financial distress (see Campbell et al., 2008 and Tinoco and Wilson, 2013). Three, we attempt to identify some critical determinants of financial

[4] It is noteworthy that by the year 2013 nearly one-third of corporate debt was owed by firms with an interest coverage ratio (ICR) less than 1.

distress from a list of 34 initial factors to develop a parsimonious distress prediction model. Four, we attempt to estimate our empirical distress prediction models on two-time horizons, one year ahead as well as two year ahead, to compare the predictive accuracy of models on different time horizons. Five, we attempt to compare the predictive abilities of three forecasting techniques namely, binomial Logit, Artificial Neural Network (ANN) and Support Vector Machine (SVM). Finally, to the best of our knowledge, this is one of the first extensive efforts to develop a parsimonious distress prediction model for publicly listed non-financial companies in India. Hence, we attempt to contribute to the literature by providing some new evidence from an emerging economy.

The remainder of the paper is organized as follows: the next section provides a brief review of literature. Section three provides description of data and database along with the selection and construction procedure of variables used in the empirical analysis. Section four presents a detailed discussion on the empirical methodology adopted in the study. The empirical results are presented in the fifth section. And in the last section we provide summary and conclusions.

2. Review of Literature

Although the issue of financial distress has been studied extensively, empirical studies so far have only focused on the advanced economies and the empirical evidences from emerging economies are very limited in number and scope.[5] In this regard, important early studies on prediction of corporate financial distress include Beaver (1966), Atlman (1968) and Deakin (1972) that focused on the estimation of univariate or multivariate discriminative functions for a sample of distressed and healthy firms. Empirical findings of early studies collectively suggested

[5] To conserve space, we only provide a brief discussion and review of literature. See Bhattacharjee and Han (2014), Tinoco and Wilson (2013), Mselmi et al. (2017) and Charalambakis and Garrett (2019) for further discussion.

that accounting ratios of failing firms are significantly different from those of healthy firms and accounting ratios can be useful in investigation and identification of financial distress. The financial ratios of distressed firms were found to be very poor compared to the healthy firms and all were facing unstable financial situations. For example, Beaver (1966) used univariate analysis to analyze the ability of accounting data for distress and bankruptcy prediction. This approach is based on the comparison of a financial ratio of interest with a benchmark ratio to distinguish between a failed and non-failed firm. Altman (1968) used the multiple discriminant analysis to constructed Z-score which is now widely used for predicting financial distress. Dambolena and Khoury (1980) used Multivariate Discriminate Analysis (MDA) to predict bankruptcy with prediction accuracy of 87%, 85% and 78% from one, three and five years prior to bankruptcy, respectively.

Rees (1995) argued that information regarding market prices might be helpful in prediction of bankruptcy because they reflect ability of firms to generate cash flows. Shumway (2001) showed that accounting ratios employed in earlier work on bankruptcy were statistically insignificant whereas market variables like security returns are found to be highly correlated with bankruptcy. Christidis and Gregory (2010) tested the usefulness of market variables in predicting financial distress for quoted companies of UK and found that inclusion of market variables enhances the predictive ability of their model. Similarly, Chava and Jarrow (2004) showed that the power of predictive model can be enhanced by accounting for industry classification.

Tinoco and Wilson (2013) used ex-ante models for distress identification and prediction one and two years prior to the distress event for United Kingdomand highlighted the importance of a combined model in distress prediction. Most recently, Mslemi et al. (2017) examined the issue for a sample of French firms using recently developed machine learning based techniques.

The results of the study indicated that for one-year prior to financial distress, SVM is the best classifier with an overall accuracy of 88.57%. Charalambakis et al. (2019) investigated the determinants of corporate financial distress by using a multi-period logit model and concluded that profitability, leverage, size and output growth rate have significant prediction power of financial distress for Greek firms.

In short, previous empirical studies have identified several determinants of financial distress with significant explanatory power and a large part of the available empirical evidences are for advanced countries. To the best of our knowledge, empirical studies for the Indian corporate sector are negligible. Therefore, in this study we attempt to develop an efficient distress prediction model for the Indian corporate sector by using most recent information on publicly listed Indian firms. Lastly, most of the previous studies have focused on the industry specific approach. But in the present study, we aim to perform an extensive empirical analysis of financial distress by covering companies across many industries covered in cooperate sector.

3. Data and Measurement of Variables

The information on the sample firms used in the study is taken from Prowess Database.[6] The study is based on the annual financial data of 1,957 companies listed on the National Stock Exchange (NSE). The primary reason behind considering companies listed on the NSE is that it accounts for the largest part of trading activity in India as compared to the Bombay Stock Exchange (BSE). Following the standard practice in the related literature, we excluded companies related to financial services from the sample because of their unique operating, financial and risk characteristics (see Bhattacharjee and Han, 2014; Mslemi et al., 2017). Hence,

[6]Prowess database is provided by the Centre for Monitoring Indian Economy (CMIE).

we initially included observations for a total of 1,742 non-financial publicly listed companies. The sample period of the study spans from 2010 to 2016. We select 2010 as the starting point of study in order to eliminate the impact of global financial crisis on our results (see Ahmed et al., 2013 for further discussion). Further, we restrict the sample period by the end of 2016 to avoid impact of the implementation of the Insolvency and Bankruptcy code (IBC) by the Government of Indian on May 28, 2016.The aftermath period witnessed sudden increase in number of firms that applied for the corporate insolvency resolution process (see Economic Survey, 2018 – 2019).[7] In order to enhance the reliability of the data, in each year, we excluded those firms from the sample that were suspended by the stock exchange from trading. Thus, the actual sample size in each year is 1,742 less the number of companies suspended by the stock exchange in that particular year. This means that the number of companies varies (between 1681 to 1705) on year to year basis in our sample. Table 1 presents information on total number of companies in each year after eliminating financial firms and suspended companies.

[Insert Table 1 about here]

The selection of accounting ratios and market variables considered for empirical analysis is based on prior literature on financial distress (see Min and Lee, 2005; Chen, 2011; Mslemi et al., 2017). Initially we collected data on 34 variables (i.e. 29 financial ratios and 5 market variables)

[7] The first case under the IBC was admitted by the National Company Law Appellate Tribunal (NCLT) on January 17, 2017 and the first insolvency resolution plan was approved on August 2, 2017. It is noteworthy that by February 2019, that is, within 27 months of the implementation of the IBC, as many as 14,000 applications had been filed for initiation of corporate insolvency resolution process (CIRPs) (See Economic Survey, 2018-19).

for empirical analysis. Further, because of inconsistencies or missing values in the data, or unavailability of data points for some of the years or high correlation between variable, a multi-stage refinement and elimination process is adopted to identify most critical determinants of financial distress. A list of all financial ratios and market variables initially selected for analysis is presented in table no. 2. We classify companies into different industries on the basis of Global Industry Classification Standard (GICS), Thomson Reuters, to test if classification of firms on the basis of industry can enhance the predictive power of the model. Further, we also include an industry dummy in the empirical analysis.

[Insert Table 2 about here]

4. Empirical Methodology

4.1 Financial Distress Measures

The first issue in the empirical analysis of financial distress is the definition and measurement of financial distress. We need to clearly define financial distress using some easily observable financial measures or indicators that can be used for the classification of firms in financially distressed or healthy category. Previous studies have used different measures of financial distress for classification and distress prediction (see Altman, 1993; Allen and Saunders, 2003 for extensive reviews).

Given the fact that the government of India has recently implemented the IBC and many distressed firms are in the early stages of legal proceedings, we attempt to adopt a measure of distress that can be applied regardless of legal consequences of bankruptcy and liquidation process. Specifically, we aim to develop an accounting-based definition of financial distress by mainly focusing on the ability of firms to repay its debt and other financial obligations (see Asquith et al. 1994). Following Pindado et al. (2008), Hernandez and Wilson (2013) and Bhattacharjee and Han (2014), we considered the following two definitions of financial distress to classify firms in distressed and healthy category.

The first definition, hereafter called two factor based definition of financial distress, is based on two factors which focuses on the Interest Coverage Ratio and change in market capitalization.[8] Using this two factor definition a firm is classified as financially distressed: first, when its EBITDA is lower than its interest expenses in the year under consideration (i.e. ICR < 1) and second, when the firm experiences negative growth in market value in the same year. If the first condition is satisfied, it can be inferred that the profits generated by the firm from operations are not sufficient enough to cover interest expense. And if the second condition is satisfied, it implies that equity holders are likely to lose confidence in the firm which could be attributed to poor operational efficiency. Therefore, negative growth in market value can be interpreted as a sign that a firm is perceived negatively by its equity holders, and hence, a decline in the market value reflects that a firm is in the state of financial distress (see Tinoco and Wilson, 2013).

The second definition, hereafter called the three-factor based measure of financial distress, is formulated by using one additional criterion that is, change in total assets of a firm (see Bhattacharjee and Han, 2014). According to this definition, a firm is classified as financially distressed, if it satisfies the above mentioned two conditions as discussed under two-factor

[8]Interest Coverage Ratio (ICR) that is defined as the EBITDA (Earnings before Interest, Tax, Depreciation and Amortization) divided by the Interest Charge. In literature, interest cover is a frequently employed as a measure of financial distress and an important determinant of bankruptcy (seeKam et al., 2008).

definition, and it also suffers from a negative growth in assets. We include this additional criterion with the aim to design a more appropriate classification of firms and for the reason that a low interest coverage ratio may arise when debt is used as a major source of finance. In such a situation, the assets of a firm must increase. And if it does not increase, it can be inferred that the debt borne by a firm is not employed in productive asset building. Any firm that fails to satisfy all the criteria in a given definition is considered to be a part of the middle group that is, neither distressed nor healthy, and hence, it is excluded from the analysis. For simplicity, in the remainder of this paper, the binary dependent variable constructed using both the definitions of financial distress will be referred to as 'financial distress indicator' or dependent variable. Financially distressed firms are assigned value 1 and healthy firms are assigned value 0.The year wise details and process of elimination adopted to select the final sample of distressed and healthy firms classified using both the definitions of distress are presented in Table 3.

[Insert Table no. 3 & 4 about here]

The classification of firms according to both the definitions varies with different time horizons because of the fact that firms with missing data points are eliminated from the sample. Panel A of Table 3 presents the details of firms during the process of elimination at each stage when the dependent variable is defined using the 2-factor criteria. And panel B reports the same for the dependent variable defined using the 3-factor criteria. The actual number of firms classified as financially distressed and healthy (whose status is predicted using 1-year and 2-year lagged values of independent variables) is also reported in Table 3. The summary of firms identified and included in financially distressed and healthy category for one and two years ahead prediction are presented in table 4. It is clearly observable that the percentage of firms in distressed category is highest in the year 2011 using both the definitions of financial distress. For example, by using the 2-factor definition 53% firms are classified as financially distressed in the year 2011 whereas the 39% firms are classified as distress using the 3-factor definition during the same year.

4.2 Selection of independent variables

Previous studies have considered a range of potential independent variables for the analysis of financial distress. Based on their findings, we initially select 34 accounting ratios as potential explanatory variables to begin the empirical analysis (see Chen, 2011; Sun and Li, 2012; Min and Lee, 2005 and Mselmi et al., 2017). Out of the initial 34variables, there are 29 accounting ratios selected to cover six different aspects (or groups) of firm performance: profitability, liquidity, structure, activity, solvency and size. Each aspect of firm performance includes at least four or more financial ratios that provide some indication of firms' financial position. Previous studies have used different proxies for firm size. In this study, we consider an accounting-based variable for size namely log (total assets) and five market variables (MV 30 – MV 34) as discussed in table 2. The selection of final regressors is based on a stepwise cleaning and testing process. We first eliminate some of the variables mainly due to the non-availability of full data for a sample of firms. Out of 29 accounting ratios and 5 market variables chosen from prior studies (as shown in Table 2), 3 ratios (namely Stock Turnover ratio, No Credit Interval and Return on Net Worth) and 2 market variables (namely Residual Return and PE ratio) were eliminated due to numerous missing values. Second, given the possibility of high correlation among various ratios belonging to one group, we perform the pairwise correlation and the results are presented in Table 5. The accounting ratios that were found to be highly correlated with one or the other ratios in a group were eliminated. Out of the remaining 25 ratios and 4 market variables, we removed 4 ratios (Gross Profit Margin, Quick Ratio, Debt Ratio, and Equity to Total Assets Ratio) and one market variable (market capitalization) to avoid possibility of multicollinearity. Finally, we used the binary dependent variable and the remaining 24 predictor variables and industry dummy for further analysis. The details of all 24 predictor variables identified after the first stage of elimination and selection process are reported as Model – 1 in column 3 of Table 6.

[Insert Table 5 & 6 about here]

4.3 Distress Prediction Models

4.3.1 Binomial Logit Model

The next stage of empirical analysis involves further screening and elimination of variables using the logistic regression in order to identify the most critical determinants of financial distress. The starting point of the empirical analysis is the estimation of Model – 1 that takes into account all 24 variables (21 accounting ratios, 3 market variables and industry dummy) selected on the basis of correlation and missing values in the first stage of screening process (see Table 6, column 2 & 3). We now performed the logistic regression using the dependent binary variable constructed using the two-factor definition of financial distress. The logistic regression was estimated using the values of 1-year lagged regressors and industry dummies. Independent variables statistically insignificant at 20 percent level of significance were eliminated and variables that were significant at 20 percent level, in at least three or more years of the sample period, were considered for further investigation.[9] We call this relatively condensed model as Model – 2 which includes a total of 14 variables (i.e. 12 accounting ratios and 2 market variables) and the details are reported in column 4 & 5 of Table 6. This process helped us in identifying a more summarized model which now includes only the most important regressors from each category and has a relatively high explanatory power.

We again repeat the same process using the most significant variables within each group, as identified in model–2, and arrived at a more parsimonious final model (termed as Model–3) which includes only six highly significant regressors. It is noteworthy that none of the market variables, included as a proxy for size, is significant and hence we will focus only on log (total assets) as a proxy for size. This final model will be used for further analysis and distress prediction purpose. The estimated results of logistic regression for the final model are reported in Table 7. The signs of the estimated coefficients are consistent with theoretical expectations. Broadly, they suggest that a high ROCE (profitability), CFO/TL (liquidity), ATR (efficiency) and FA/TA (structure) lead to low probability of financial distress. Further, a high DE ratio (solvency) and large total assets (firm size)result in high probability of financial distress in the Indian corporate sector. While the sign of the solvency related coefficient is broadly on expected line indicating a high debt increases financial distress, the impact of firm size needs some discussion. The results suggest that a large firm size is associated with more financial distress. The positive coefficient of size variables is somewhat similar to the findings of Charalambous et al. (2020). This may be because of the fact that efficient management of large firms becomes difficult once the size of firms increases beyond some threshold level. It is also indicative of the fact that very large firms generally exhaust all profitable investment opportunities and new

[9] In a multiple regression model, some of the regressors may be weakly associated with the financial distress and make a small but important contribution in distress prediction. Keeping this in mind, we consider the 20 percent significance level.

incremental investments can only deliver fewer returns compared to initial investment. Hence, additional capital is less efficiently deployed. The results (positive sign of *size* coefficient) also make sense in the Indian case as recent crisis of NPA is more of a large firm problem. We confirm our argument by comparing the mean size of distressed and healthy firms over the study period. The mean size (log of total assets) of distressed firms (3-factor based definition) is 9.19 which are higher than that of healthy firms which is 8.97. Further, in order to rule out the interaction effect leading to positive size coefficients (see Charalambous et al., 2020), we estimate a univariate model involving size (log of total assets) as only explanatory variable. The estimated results suggest that the coefficient of *size* is significantly positive and hence re-confirmed our findings of multivariate analysis. The results for distressed and healthy firms based on two-factor based definition are broadly similar.

[Insert Table 7 about here]

4.3.2 Other Predictive Models – SVM and ANN approaches

In the next stage of empirical analysis, we also consider the SVM and ANN for distress prediction along with the logistic model. For ANN estimation, we use multilayer perceptron method involving ten hidden layers.[10]The overall data, data source and the number of firms in the sample are the same as used in the logistic regression. However, we made some changes in the sample to deal with the problems created by data imbalance. It is noteworthy that the number of healthy firms far exceeds the number of distressed firms on year to year basis in our sample.

[10]The ANN and SVM models were estimated using the MATLAB (version 2015). The MATLAB allows us to choose the number of hidden layers in the ANN model, but the numbers of neurons (nodes) in the hidden layers are chosen automatically. For more details of ANN (Multi-layer perceptron) as well as SVM procedure see Mselmi et al. (2017) for further discussion.

Hence it is causing a serious data imbalance problem which is observable in Table 4. In the presence of severe data imbalance there is a high possibility that the forecasting accuracy of models may not be comparable and valid (see Kim et al., 2015 and Mslemi et al., 2017).

Keeping in mind the data imbalance problem, we created sub samples of financially healthy firms (i.e. majority class) in such a way that number of firms in each subsample for each year is individually close to the number of firms in the minority class (or distressed firms). For example, in 2014 (as per 2-factor classification for one-year-ahead predictions), 82 firms are classified as financially distressed and 828 firms are classified as healthy. To overcome this problem, we divided the larger sample containing healthy firms into subsamples by following a two-step process. First, we randomly divided the number of healthy firms by number of financially distressed firms without replacement (828/82) to calculate the number of subsamples (in our case approximately 10 subsamples). Second, subsamples are made as follows: If we create 10 subsamples, each one will contain 82.8 firms which is practically not possible. So, by rounding up this number, 9 subsamples of 83 firms are made while the remaining firms formed the 10thsubsamples. Similar process is adopted for each year in the study. Each year we estimate the models *N* times, where *N* is the number of sub-samples of healthy firms in that year. Each estimation involves a fixed sample of distressed firms and one matching sub-sample of healthy firm. The *N* prediction rates achieved from a model in a given year are then averaged to obtain the mean annual prediction rate for that model. As a part of the training process for both SVM and ANN, in each of *N* estimations; we use the 50% data for training and the remaining 50% for testing purpose. We then reverse the process by making our 50% test data as training data and 50% training data as test data. The estimation process ensures that our full data is used for both training as well as testing purpose. The mean prediction rate is then calculated for each sub-sample by taking the average of the two 50% - 50% training-testing samples. We further estimate the mean prediction rate for each year by taking the average of prediction rates for the *N* samples which is reported in tables. The feature selection method for selecting suitable features for classification of firms in healthy or distressed is based on Fisher discriminant ratio (FDR). We calculated FDR ratio for all the given parameters and selected six parameters with highest FDR ratios. The selected parameters are ROCE, ROA, ATR, GR, RE/TA and log(Closing). For comparing the predictive accuracy of all the three models, we re-estimated the logit model for each year with the balanced subsamples using the two-step estimation process mentioned above and finally computed its average accuracy (see Table 8). Both SVM and ANN have been estimated in two ways; one, using the factors identified by our logit model (pre-specified factor approach) and two, identified factors within the framework of SVM and ANN using the feature selection method with FDR (unspecified factor approach).[11]

5. Empirical Results

In this section, we discuss the empirical results of the study. The overall results are discussed in two stages. The first stage of analysis is related to the estimation results using the logistic model. The second stage involves estimation using the SVM and ANN model and comparison of prediction accuracies.

5.1 Binomial Logit based Distress Prediction Model

The year wise estimated coefficients using the logistic model for final model involving six highly significant determinants of distress are reported in Table 7. Our six-factor model is somewhat different from the Altman's distress prediction model which is popularly cited in prior research.

[11]Keeping the space constraint in mind, we do not discuss the SVM and ANN model in detail. See Mselmi et al. (2017) for further discussion.

Altman (1968) included measures of solvency, liquidity, operating efficiency, profitability and investment rates in his five-factor model. Although we find that the first four attributes of Altman's model also play an important role in distress prediction in the Indian context, but our optimal measures seem to differ from those suggested in Altman's work. For example, we find debt to equity ratio as a good measure of leverage while Altman measured solvency indirectly by employing price to book value. Further, Altman used Return on Assets and working capital to total assets as measures of profitability and liquidity, respectively. But our results suggest that ROCE and CFO/TL are better measures of profitability and liquidity for Indian corporate sector. Operational efficiency is a critical variable in both the studies and ATR seems to be its good measure. Altman (1968) additionally used an investment factor proxied by change in firm size. But it is noteworthy that recent studies have used asset growth as an important measure of investment rate of the firms. In addition, we also confirmed the role of asset structure (measured by fixed assets to total assets) in distress prediction. Finally, we included the sixth determinant of financial distress that measures firm size. We constructed this variable by taking natural log of total assets. Overall findings of the study clearly suggest that the critical determinants of financial distress and their measures may vary across developed and emerging markets like India. Hence, it is more suitable to use a country specific model for financial distress prediction and the commonly used Altman (1968) and other models may not be the best models for emerging markets.

For empirical analysis we used the determinants of all three model which are described in table 6 for predicting the status of firms (distressed or healthy) with one as well as two years lagged values of independent variables and the prediction accuracies are reported in Table 8. Our definitions helped us in firm classification and provide us the status for each firm that is healthy or distressed. The expected status of each firm is provided by the alternative models using endogenously determined threshold values. We refer a firm to be correctly classified, if its actual status matches with the expected status. If the actual and expected status does not match, we refer to it as an error. The prediction rates for any model are defined as the ratio of correctly classified cases to total cases. All the sample firms have been classified by using 2-factor and 3-factor based definition as described in the paper. This provides us the actual number of healthy and distressed cases. The expected status of sampled firm (whether it is healthy or distressed) is provided by each predictive model. A company is classified correctly, if there is congruence in its actual and expected status. It is noteworthy that, in logistic regression if the output value is greater than a threshold value (0.5) we assign it a class 1 (i.e. distressed); else we assign it a class 0 (not distressed). The analysis using the neural network also has a threshold of 0.5. The output above or equal to 0.5 is identified with class 1 and output less than 0.5 is identified with class 0. In SVM, we take the output of the linear function and if that output is greater than 1, we classify it into class 1 and if the output is -1, we classify it into other class 0 (i.e. not distressed).

[Insert Table 8 about here]

The results, as reported in table 8, clearly indicate that average accuracy of Model–3 (based on 2-factor definition) is 88% for 1-year ahead prediction. The average accuracy declines marginally to 85% for 2-year ahead prediction using the same model. Similarly, the prediction based on Model–3 (based on 3-factor definition) for 1-year and 2-year ahead forecasting horizons yield average accuracy rates of 90% and 87%, respectively. Though the predictive accuracies of Model–1 and Model–2 are relatively high compared to Model–3, but given the fact model–1 includes 24 regressors whereas model–3 includes only six regressors, the resulting loss of predictive power is negligible. Therefore, we consider Model–3 to be the best model. Moreover, the predictive power of models based on the3-factor criteria of measuring distress is also better in the Indian context. Further, one year ahead predictions are better than two year ahead predictions which are similar to the findings of prior work (see Bhattacharjee and Han, 2014).

The results regarding average prediction accuracy for balanced sample are also presented in Table 8. Now in the light of the fact that Model–3 performed nearly as well as model-1 & 2, while still maintain the parsimonious nature, we decided to focus on Model–3 for further analysis. The prediction accuracies for both the time horizons are presented in the table. The average prediction accuracy, when the values of dependent variable are calculated using the 2-factor criteria, for one and two year ahead predictions are83 and 80, respectively. Similarly, average prediction accuracy, when values of dependent variable are computed on the basis of 3-factor criteria, for one and two years ahead is 86 and 79, respectively. Overall results suggest that there is a marginal decline in average accuracy after accounting for data imbalance problem. Hence, it may be concluded that the presence of data imbalance introduces distortions in forecasting results and erroneously leads to high forecasting accuracy. Further, in this study, we could not observe any significant industry patterns in financial distress as the industry dummies were insignificant in model estimation. Hence, we dropped the industry dummies while estimating Model–2 and3. On the close examination of misclassified cases, we observed that one third of such cases belong to the consumer discretionary sector which is experiencing high rate of disruptions globally as well as in India.[12]Hence, there is a need to identify a different set of accounting ratios and market variables that can correctly classify and accurately predict the financial distress of firms in the consumer discretionary industry. This is important in the sense that six critical determinants identified in this study and by other leading studies such as Altman (1986), Ohlson (1980) and Shumway (2001) will not be fully effective in predicting financial distress in consumer discretionary industry. This industry is experiencing large technological disruptions particularly with the advent of digital era.

5.2 Distress prediction using SVM and ANN and comparison of prediction accuracies

Now after clearly establishing that the six factors based parsimonious (model – 3) is relatively better in terms of prediction accuracy and data imbalance leads to overestimation of prediction accuracies, we used model – 3 and balanced data for distress prediction using the ANN and SVM models. In this sub section, we present and compare predictive efficiency of alternative forecasting techniques used in the study. For this purpose, we attempt to compare the predictive ability of the binomial logit model (with balanced sub-samples, see Table no. 8 – panel C) with two other machine learning based forecasting models namely SVM and ANN. For the estimation process using the SVM and ANN, we used 80 percent of the sample data for training and remaining 20 percent for testing.[13]The calculated prediction accuracy rates for all models are reported in Table 9. It is observable that machine learning based models perform better compared to the binomial logit model on both the forecasting horizons in all the cases. In the case of 2-factors based definition, the SVM technique achieved the highest prediction

[12]The sectors in the Indian economy currently facing massive disruptions in consumer discretionary sectors in India include financial services, information technology, communication and media, energy etc.
[13] We are grateful the anonymous referee for suggesting this methodological improvement in splitting the data for training and testing purpose.

accuracy of 79.61 percent (for 1 year ahead predication) using the FDR based inputs and 77.77 percent (for 2 years ahead) using the pre-specified inputs. In the case of 3-factor based definition of financial distress and prediction over 1 year ahead horizon, again the SVM technique delivered the highest accuracy of 83.30 percent (using FDR based inputs) and ANN technique delivered the highest accuracy of 76.67 percent for 2 years ahead prediction (using FDR based inputs). The prediction superiority of SVM technique is clearly established in three out of four different empirical specifications as presented in Table 12. In short, the results suggest that the machine learning based models outperformed the binomial logit model. While the logit model based predictions delivered the accuracy of 81.44 percent between two different time horizons, the machine learning based models delivered the highest accuracy (i.e. SVM – 83.60 percent). Based on these findings, it can be concluded that the machine learning based models (i.e. ANN and SVM) have the predictive superiority over the binomial model. The superior performance of machine learning models in distress prediction is consistent with the findings of previous studies (see Mselmi et al., 2017 and references therein). These models can be used for financial distress prediction in an emerging market economy like India.

[insert Table 9 about here]

6. Summary and Conclusions

In this study, we aimed to examine the critical microeconomic (or firm specific) determinants of financial distress and attempt to develop a parsimonious distress prediction model based on some easily observable micro indicators of distress. Although there is a huge literature covering different theoretical and empirical aspects of financial distress, but very little is known about what determines the probability of corporate financial distress, especially in an emerging economy like India. Therefore, in this study, we attempt to bridge this gap by examining the probability of financial distress for a relatively large sample of listed firms from the Indian corporate sector. Further, we also attempt to compare the forecasting accuracies of competing distress prediction techniques to identify the most suitable technique in terms of predictive power. In order to identify a more appropriate measure of financial distress, we used two measures to classify firms in distressed and healthy category.

The main findings of the study could be summarized in the following point. First, out of the initial list of 34 firm-specific factors, the results suggest that six variables play statistically significant role in determining the probability of financial distress. These six critical determinants of corporate distress include ROCE, CFO/TL, ATR, DE, FA/TA and log (TA). Second, our three-factor based measure of financial distress appears to be more suitable way of defining distress as prediction accuracies of three factor-based definition are higher than the two factor-based definition. Three, our findings suggest that machine learning based models namely SVM and ANN are superior in terms of their prediction accuracy compared to the simple binomial logit model even in a relatively not so large time series data set. On average, the SVM technique achieved the highest prediction accuracy in three out of four empirical specifications and ANN model performed better in one specification. This result is in line with the findings of Mselmi et al. (2017). Four, the prediction accuracies of SVM and ANN models are better when inputs are selected automatically using the FDR. Five, as expected, the predictive accuracies of the all models declined with increase in forecasting horizon which is similar to the findings of Charalambakis and Grarrett, (2019).

The findings of the study have some important practical implications for creditors, policymakers, regulators other stakeholders. First, rather than monitoring and collecting information on a list of predictor variables, only six most important accounting ratios maybe monitored to track the transition of a healthy firm into financial distress. Second, our six-factor model can be used to devise a sound early warning system for corporate financial distress. Three, machine learning based distress prediction models have prediction accuracy superiority over the commonly used time series model in the available literature for distress prediction involving a binary dependent variable. Four, our findings suggested that a large part of misclassified cases are concentrated in consumer discretionary sector. Hence, it can be argued that our models and other similar models, generally used in the available literature, may be not be efficient in predicting financial distress of firms in the consumer discretionary industry. Therefore a different set of explanatory variables needs to be identified for understanding the distress dynamics of this sector. Finally, we used the most recent available data but restricted our sample to cover the post global financial crisis till the implementation of insolvency and bankruptcy code (IBC) in India. Once the bankruptcy code is implemented effectively and the numbers of pending cases are reduced to minimum, studies can take a longer data set and can reexamine the performance of prediction techniques in future. Also, a detailed sectoral or industry wise study will help in uncovering any industry specific pattern in financial distress in the Indian economy.

References

Agarwal, V., &Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*.

Allen, L., & Saunders, A. (2003). A survey of cyclical effects in credit risk measurement models. Technical report, BIS Working Paper No. 126, the Bank for International Settlements.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction corporatebankruptcy. *Journal of Finance*.

Altman, E. I. and Edith Hotchkiss (2006), Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt, John Wiley & Sons, Inc., Hoboken, New Jersey

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

Bernard, A.B., and J.B. Jensen. 1999. Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics* 47: 1–25.

Bhattacharjee, A. and J. Han, 2014. Financial distress of Chinese firms: Microeconomic, macroeconomic and institutional influences. *China Economic Review*, 30: 244-262.

Blum, M. (1974). Failing company discriminant analysis. *Journal of Accounting Research*, 12(10), 1-25.

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

Carlos *López-Gutiérrez,* Sergio Sanfilippo-Azofra and BegoñaTorre-Olmo (2015), Investment decisions of companies in financial distress, *Business Research Quarterly*, Volume 18, Issue 3, July–September 2015, Pages 174-187

Carpenter, R., Fazzari, S. and Petersen, B. (1994). 'Inventory investment, internal finance fluctuation, and the business cycles', *Brookings Papers on Economic Activity*, 2, pp. 75–122.

Charalambakis, E. C. and Ian Garrett (2019), On corporate financial distress prediction: What can we learn from private firms in a developing economy? Evidence from Greece, *Rev Quant Finan Acc* 52:467–491.

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

Chen, M. Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38(9), 11261–11272

Christidis, A., & Gregory, A. (2010). Some new models for financial distress prediction in the UK. Xficentre for finance and investment discussion paper no. 10

Dambolena, I. G., & Khoury, S. J. (1980). Ratio stability and corporate failure. *Journal of Finance*, 35(4), 1017-1026.

Deakin, E. B. "A Discriminant Analysis of Predictors of Failure," *Journal of Accounting Research*, 10 (Spring 1972): 167-79.

Economic Survey (2017-18), Government of India, Ministry of Finance, Department of Economic Affairs, Economic Division, New Delhi.

Economic Survey (2018-19), Government of India, Ministry of Finance, Department of Economic Affairs, Economic Division, New Delhi.

Edminster, R. (1972). "An empirical test of financial ratio analysis for small business failure prediction", *Journal of Financial and Quantitative Analysis*, pp. 1477-1493.

Fazzari, S. R., Hubbard, G. and Peterson, B. (1988). 'Financing constraints and corporate investment', *Brookings Papers on Economic Activity*, 1, pp. 141–95.

Gilchrist, S. and Himmelberg, C. P. (1995). 'Evidence on the role of cash flow for investment', *Journal of Monetary Economics*, 36(3), pp. 541–72.

Harada, N., &Kageyama, N. (2011). Bankruptcy dynamics in Japan. *Japan and the World Economy*, 23(2), 119-128.

Haralambous, C, Spiros H. Martzoukos & Zenon Taoushianis (2020): Predicting corporate bankruptcy using the framework of Leland-Toft: evidence from U.S.,QuantitativeFinance,Volume 20 (2), pp. 329-346.

Hotchkiss (2006), Corporate Financial Distress & Bankruptcy, 3rd edition, John Wiley.

Ismael G. Dambolenaand Sarkis J. Khoury (1980), Ratio Stability and Corporate Failure, *The Journal of Finance*, Vol. 35, No. 4 (Sep., 1980), pp. 1017-1026.

Jackendoff, N. (1962). A study of published industry financial and operating ratios. Philadelphia: Temple University, *Bureau of Economic and Business Research*.

Kam, A., Citron, D., &Muradoglu, G. (2008). Distress and restructuring in China: Does ownership matter? *China Economic Review*, 19, 567–579.

Keasy K. and P. McGuiness, "The Failure of UK Industrial Firms 1976-1984, Logistic Analysis and Entropy Measures", *Journal of Business Finance and Accounting* (vol. 17 1990), pp. 119- 136.

Kim, M. J., Kang, D. K., & Kim, H. B. (2015). Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications*, 42(3), 1074–1082

Liu, J. and Wilson, N. (2002). 'Corporate failure rates and the impact of the 1986 Insolvency Act: an econometric analysis', *Managerial Finance*, 28(6), pp. 61–71.

Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603–614.

Mslemi N., Lahiani A., Hamza T. (2017), 'Financial Distress prediction: The case of French small and medium-sized firms', *International Review of Financial Analysis*.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*.

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

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

Rees, W. (1995), *Financial Analysis*, 2nd edition, Prentice Hall.

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

Tinoco M.H. and Wilson N. (2013), Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *InternationalReview of Financial Analysis*.

Vlieghe, G. W. (2001). 'Indicators of fragility in the UK corporate sector', *Bank of England Working Paper No. 146*.

Wadhwani, S. B. (1986). 'Inflation, bankruptcy, default premia and the stock market', *Economic Journal*, 96(381), pp. 120–38.

Wasim Ahmad, Sanjay Sehgal and N.R. Bhanumurthy (2013), Eurozone crisis and BRIICKS stock markets: Contagion or market interdependence? *Economic Modelling* 33 (2013) 209–225.

Wilson, R. L. & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision support systems*, Volume 11, Issue 5, June 1994, Pages 545-557.

Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research, Vol. 22, Studies on Current Econometric Issues in Accounting Research (1984), pp. 59-82

Appendix – 1

Table of Results

Table 1: Year wise information of total firms in the sample

	2011	2012	2013	2014	2015	2016
Total firms listed on NSE	1957	1957	1957	1957	1957	1957
(Less)financial firms	215	215	215	215	215	215
Total non-financial firms listed on NSE	1742	1742	1742	1742	1742	1742
(Less) suspended firms	40	37	38	39	49	61
Number of firms after suspension	1702	1705	1704	1703	1693	1681

Notes: (i) This table presents information about the total number of firms arrived at after elimination of financial firms and suspended firms in each year from 2011 to 2016, (ii) NSE stands for National Stock Exchange.

Table 2: Summary of accounting ratios and market variables initially selected as determinants of financial distress

Category	Code Name	Ratio	Formula	Category	Code Name	Ratio	Formula
Profitability	P1	NP margin	Net Profit/Net Sales	**Activity**	A18	Creditor days	Creditors/Operating Revenue*360
	P2	GP margin	Gross Profit/Net Sales		A19	Debtor days	Debtors/Operating Revenue * 360
	P3	RNW	Net profit/Shareholders Funds		A20	ATR	Sales Revenue/Total Assets
	P4	ROCE	Net Profit/Capital Employed		A21	STR	Operating Revenue/Stock
	P5	ROA	Net Profit/Total Assets	**Solvency**	SL22	Financial Leverage	Long-term Liabilities/Total Assets
Liquidity	L6	QR	Quick Assets/ Current Liabilities		SL23	GR	Long-Term Liabilities/Capital Employed
	L7	CR	Current Assets/Current Liabilities		SL24	debt ratio	Total Debt/Total Assets
	L8	NCI	(Quick Assets-Current Liabilities)/Daily Operating Expenses		SL25	SR	PAT + Depreciation/ (Long-Term Liabilities + Short-Term Liabilities)
	L9	CFO/TL	Total Cash from Operations/Total Liabilities		SL26	Repayment capacity	Financial Debt/cash Flow
	L10	CF/TA	Cash Flow/Total Assets		SL27	DE	Debt/Equity
	L11	CF/OR	Cash Flow/Operating Revenue		SL28	RE/TA	Retained Earnings/Total Asset
	L12	WC/TA	Working Capital/Total Assets	**Size**	SZ29	ln(TA)	ln(Total Assets)
Structure	S13	Eq/TA	Equity/Total Asset		MV30	ln(Closing Price)	ln(Closing Price)
	S14	CA/TA	Current Assets/Total Assets		MV31	ln(Market Capitalisation)	ln(Market Capitalisation)
	S15	CL/TL	Current Liabilities/Total Liabilities		MV32	P/E	Price/Earnings Per Share
	S16	FA/TA	Fixed Assets/Total Assets		MV33	P/B	Price/Book Value Per Share
	S17	SF/NCL	Shareholders'		MV34	RESIDUAL	Cumulative monthly security

			Funds/Non-Current Liabilities			RET'10	return minus cumulative monthly NSE500 index return

Notes:(i) This table presents details of 34 financial variables (accounting ratios and market variables) initially selected for empirical analysis. (ii) where the code names of ratios indicate category and the number of variables. For example, code name P1 and MV34 indicates ‘category – Profitability (P) and variable no. 1, and category – Market Variable (MV) and variable no. 34, respectively, and so on.

Table 3: Calculation of total annual observations considered for prediction

Panel – A (Narrow definition or 2-factor based definition of financial distress)						
Years	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Number of firms after suspension (see raw 6 of table no. 1)	1702	1705	1704	1703	1693	1681
Missing values of dependent variable	582	508	488	484	474	374
Unclassified firms	886	319	675	269	404	603
Total firms classified as distressed/healthy	**234**	**878**	**541**	**950**	**815**	**704**
Less Missing values of independent variable (for 1-year ahead)	38	36	44	40	74	71
Total firms (For 1 year ahead forecasting)	196	842	497	910	741	633
Total firms classified as distressed/healthy	**234**	**878**	**541**	**950**	**815**	**704**
Less Missing values of independent variable (for 2-year ahead)	63	126	41	41	65	89
Total firms (for 2 year ahead forecasting)	171	752	500	909	750	615
Panel – B (Broad definition or 3-factor based definition of financial distress)						
Number of firms after suspension	1702	1705	1704	1703	1693	1681
Missing values of dependent variable	582	508	488	484	474	450
Unclassified firms	948	493	797	556	663	707
Total firms classified as distressed/healthy	**172**	**704**	**419**	**663**	**556**	**524**
Less Missing values of independent variable (for 1-year ahead)	27	25	26	28	47	40
Total firms (For 1 year ahead forecasting)	145	679	393	635	509	484
Total firms classified as distressed/healthy	**172**	**704**	**419**	**663**	**556**	**524**
Less Missing values of independent variable (for 2-year ahead)	43	104	29	33	41	60
Total firms (for 2 year ahead forecasting)	129	600	390	630	515	464

Note: (i) Panel A presents the details and computation of total number of firms considered for prediction when values of dependent variable are computed using the narrow measure of financial distress or the 2-factor criteria; (ii) Panel B presents the details and computation of total number of firms considered for prediction when values of dependent variable are computed using the broad measure of financial distress or the 3-factor criteria.

Table 4: Summary statistics for annual observations in each category (Financially distressed & Financially healthy)

Panel A: Classification according to 2-factor definition												
	t-1						t-2					
Years	2011	2012	2013	2014	2015	2016	2011	2012	2013	2014	2015	2016
Financially Distressed	104	75	159	82	79	122	85	73	175	84	97	132
Financially Healthy	92	767	338	828	662	511	86	679	325	825	653	483
TOTAL	196	842	497	910	741	633	171	752	500	909	750	615
% of financially distressed firms	53%	9%	32%	9%	11%	19%	50%	10%	35%	9%	13%	21%
Panel B: Classification according to 3-factor definition												
Financially Distressed	56	44	100	57	47	78	50	42	109	59	60	87
Financially Healthy	89	635	293	578	462	406	79	558	281	571	455	377
TOTAL	145	679	393	635	509	484	129	600	390	630	515	464
% of financially distressed firms	39%	6%	25%	9%	9%	16%	39%	7%	28%	9%	12%	19%

Notes: The table reports classification of firms into financially distressed and healthy. The last row of each table shows the percentage of financially distressed firm out of total firms.

Table 5: Correlation Matrices and Multicollinearity Diagnosis

Panel A: Correlation matrix for Profitability Ratios							
Variables (Code name)	**P2**	**P1**	**P5**	**P4**			
P2	1						
P1	0.5643	1					
P5	0.1969	0.1887	1				
P4	0.1809	0.1652	0.9474	1			
Panel B: Correlation Matrix for Liquidity Ratios							
Variable	**L11**	**L10**	**L7**	**L9**	**L6**	**L12**	
L11	1						
L10	0.0048	1					
L7	-0.0097	0.1551	1				
L9	0.0491	0.0597	-0.0389	1			
L6	-0.0159	0.2094	0.9208	-0.0284	1		
L12	-0.0197	0.2193	0.3125	-0.0362	0.3051	1	
Panel C: Correlation Matrix for Solvency ratios							
Variable	**SL27**	**SL22**	**SL24**	**SL23**	**SL26**	**SL28**	**SL25**
SL27	1						
SL22	-0.0137	1					
SL24	0.0117	0.5576	1				
SL23	0.0017	-0.0299	-0.0878	1			
SL26	0.0002	0.0623	0.0216	0.0011	1		
SL28	-0.0006	-0.0158	-0.0243	0.0009	0.0002	1	
SL25	-0.0012	-0.0645	-0.1113	0.0015	0.0157	0.0042	1
Panel D: Correlation Matrix for Activity Ratios							
Variable	**A20**	**A18**	**A19**				
A20	1						
A18	-0.0575	1					
A19	-0.0701	0.4792	1				
Panel E: Correlation Matrix for Structural Ratios							
Variable	**S14**	**S15**	**S13**	**S16**	**S17**		
S14	1						
S15	0.2543	1					
S13	-0.0114	-0.8448	1				
S16	-0.3721	0.2091	-0.3567	1			
S17	0.0071	-0.0359	0.0539	-0.0238	1		
Panel F: Correlation Matrix for Market Variables							
Variable	**MV29**	**MV30**	**MV34**	**MV32**			
MV29	1						
MV30	0.6817	1					
MV34	0.3715	0.8139	1				
MV32	0.4445	0.4362	0.0931	1			

Note: See Table 2 (column no. 2 & 6) for the category wise names of variables for their respective code names.

Table 6: List of variables included in models

	Model -1		Model – 2		Model – 3	
	(21 accounting ratios, 3 market variables and industry dummy)		(12 accounting ratios and 2 market variables)		(6-factor model)	
Category	Code Name	predictor variable	Code Name	predictor variable	Code Name	predictor variable
Profitability	P1	NP margin	P4	ROCE	P4	ROCE
	P4	ROCE	P5	ROA		
	P5	ROA				
Liquidity	L7	CR	L9	CFO/TL	L9	CFO/TL
	L9	CFO/TL	L10	Cash/TA		
	L10	Cash/TA	L11	CFO/OR		
	L11	CFO/OR				
	L12	WC/TA				
Activity	A18	Creditor days	A18	Creditor Days	A20	ATR
	A19	Debtor days	A19	Debtor Days		
	A20	ATR	A20	ATR		
Solvency	SL22	Financial Leverage	SL27	DE	SL27	DE
	SL23	GR				
	SL25	SR				
	SL26	Repayment capacity				
	SL27	DE				
	SL28	RE/TA				
Structure	S14	CA/TA	S14	CA/TA	S16	FA/TA
	S15	CL/TL	S15	CL/TL		
	S16	FA/TA	S16	FA/TA		
	S17	SF/NCL				
Size	MV30	ln(Closing Price)	MV33	P/B	SZ29	ln(TA)
	MV33	P/B	SZ29	ln(TA)		
	SZ29	ln(TA)				

Table 7: Estimation results of logistic regression

Panel A (2-factor based definition of financial distress)												
Years	**2011**		**2012**		**2013**		**2014**		**2015**		**2016**	
Variables	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*	*t-1*	*t-2*
ROCE	-0.167***	-0.148***	-0.169***	-0.096***	-0.181***	-0.147***	-0.104***	-0.031***	-0.19***	-0.068***	-0.176***	-0.119***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.041)	(0.00)	(0.00)	(0.00)	(0.00)
CFO/TL	-5.853***	-5.972***	-2.729***	-2.528**	-0.535	-2.964***	-2.201***	-5.148***	0.954	-1.531***	-2.488*	0.498
	(0.009)	(0.010)	(0.042)	(0.097)	(0.718)	(0.019)	(0.00)	(0.001)	(0.664)	(0.002)	(0.180)	(0.759)
ATR	-0.432	-0.057	-0.119	-0.032	-0.434**	0.068	-1.447***	-1.079***	-0.921***	-0.884***	-0.89***	-0.715***
	(0.235)	(0.879)	(0.634)	(0.891)	(0.087)	(0.645)	(0.00)	(0.00)	(0.007)	(0.001)	(0.005)	(0.002)
DE	0.467***	0.0006	-0.004	0.166***	0.084**	-0.0009	0.058**	0.143***	0.057***	0.029**	0.0005	0.067***
	(0.00)	(0.968)	(0.614)	(0.005)	(0.053)	(0.878)	(0.081)	(0.005)	(0.011)	(0.097)	(0.820)	(0.034)
FA/TA	0.267	0.267	-1.012***	-1.048***	-2.184***	-2.076***	-1.494***	-1.429***	-1.207***	-1.627***	-0.951***	-1.927***
	(0.733)	(0.714)	(0.048)	(0.039)	(0.00)	(0.00)	(0.004)	(0.007)	(0.023)	(0.00)	(0.025)	(0.00)
ln(TA)	-0.138	0.031	-0.107	-0.213***	0.08	0.106*	0.097	0.099	0.318***	0.251***	0.165***	0.127**
	(0.396)	(0.845)	(0.317)	(0.042)	(0.366)	(0.182)	(0.271)	(0.233)	(0.00)	(0.001)	(0.039)	(0.074)
Panel B (3-factor based definition of financial distress)												
ROCE	-0.165***	-0.199***	-0.240***	-0.095***	-0.226***	-0.178***	-0.152***	-0.022*	-0.265***	-0.114***	-0.225***	-0.132***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.148)	(0.000)	(0.000)	(0.000)	(0.000)
CFO/TL	-8.743***	-2.510	-4.093***	-3.030*	0.165	-2.942***	-2.048***	-5.082***	6.809***	-5.580***	-1.648	0.771
	(0.005)	(0.403)	(0.016)	(0.124)	(0.922)	(0.041)	(0.004)	(0.008)	(0.022)	(0.023)	(0.482)	(0.697)
ATR	0.419***	0.102	0.441*	0.302	-0.195	0.209*	-1.397***	-0.987***	-0.631*	-0.675***	-0.636**	-0.663***
	(0.007)	(0.804)	(0.194)	(0.32)	(0.453)	(0.165)	(0.000)	(0.002)	(0.15)	(0.047)	(0.095)	(0.014)
DE	2.271***	-0.001	-0.003	0.221***	0.103***	0.019	0.152**	0.262***	0.037	0.040	-0.000	0.129***
	(0.043)	(0.95)	(0.748)	(0.005)	(0.050)	(0.232)	(0.054)	(0.001)	(0.355)	(0.4)	(0.915)	(0.006)
FA/TA	-0.591*	0.898	-1.504**	-1.125*	-2.105***	-2.003***	-1.588***	-1.085**	-1.879***	-1.197**	-0.357	-1.606***
	(0.196)	(0.288)	(0.065)	(0.116)	(0.000)	(0.000)	(0.013)	(0.099)	(0.009)	(0.061)	(0.524)	(0.002)
ln(TA)	0.279*	0.061	0.239*	0.414***	-0.014	-0.037	0.085	0.062	-0.133	-0.190**	-0.121	-0.105
	(0.185)	(0.752)	(0.121)	(0.006)	(0.896)	(0.7)	(0.472)	(0.542)	(0.29)	(0.064)	(0.228)	(0.223)

Notes:(i) This table reports the results of logistic regression of the binary dependent variables on predictor variables. Models were computed for two-time frames, one in which the predictor variables assume a year prior values (from the event of financial distress) and the other in which predictor variables assume two-year prior values. (ii) * denotes significant at 20%, ** denotes significant at 10% and *** denotes significant at 5%, (iii) Values in (#) are *p*-values.

Table 8: Prediction accuracy of logit model with unbalanced and balanced subsamples

Years	1-year-ahead						Average	2-year-ahead						Average
	2011	2012	2013	2014	2015	2016		2011	2012	2013	2014	2015	2016	
Panel A (2-factor based definition of financial distress)														
Model – 1	82%	91%	84%	92%	91%	87%	88%	82%	91%	84%	92%	91%	87%	88%
Model – 2	85%	93%	87%	93%	93%	88%	90%	77%	91%	81%	91%	91%	83%	86%
Model – 3	82%	93%	84%	93%	91%	87%	88%	75%	91%	78%	91%	88%	84%	85%
Panel B (3-factor based definition of financial distress)														
Model – 1	92%	97%	90%	95%	95%	92%	94%	83%	94%	85%	92%	93%	89%	89%
Model – 2	85%	95%	89%	93%	94%	90%	91%	81%	94%	84%	92%	92%	87%	88%
Model – 3	85%	95%	87%	93%	94%	89%	90%	81%	94%	82%	90%	90%	86%	87%
Panel C (Prediction accuracy of logit model with balanced subsamples)														
Model – 3(A)	82%	81%	84%	85%	85%	84%	83%	75%	76%	88%	76%	87%	87%	80%
Model – 3(B)	87 %	86%	85%	86%	87%	87%	86%	81%	77%	78%	77%	81%	82%	79%

Notes: This table presents financial distress prediction accuracy of logistic regression one and two years before financial distress in each year. Panel A shows the results obtained for Model-1 (full model – 24 factor), Model-2 (14-factor model) and Model-3 (6-factor model) using 2-factor or narrow definition of financial distress; (ii) Panel B shows the results obtained using 3-factor or broad definition of financial distress; (iii) Model – 3(A) and Model – 3(B) indicate the final parsimonious model with 6 regressors estimated for two factor ('A') and three factor ('B') based definition of financial distress (i.e. the binary dependent variable).

Table 9: Comparison of prediction accuracies of alternative forecasting techniques**

			Panel A (1 year ahead)							Panel B (2 year ahead)						
Definition of financial distress	Techniques		2011	2012	2013	2014	2015	2016	Average Accuracy	2011	2012	2013	2014	2015	2016	Average Accuracy
2-factor	Logit (Model – 3)	**Six**	71.011	72.00	76.56	86.18	71.09	81.5	76.39	55.88	71.48	70.00	67.65	73.75	72.22	68.49
	SVM	**Six**	73.68	79.00	78.47	83.53	73.99	75.61	77.38	76.67	77.40	78.46	79.00	77.30	77.81	77.77
		FDR	76.32	81.00	80.47	81.47	73.44	85.00	79.61	64.70	74.07	75.41	72.22	78.33	78.40	73.85
	ANN	**Six**	50.00	79.00	77.34	83.53	70.31	78.5-	73.11	64.71	68.89	72.86	70.59	72.50	74.69	70.70
		FDR	50.00	79.67	79.69	81.47	73.44	85.50	74.96	76.46	76.30	77.14	69.60	76.67	76.54	75.45
3-factor	Logit (Model – 3)	**Six**	75.00	83.33	83.75	86.25	77.22	83.13	81.44	75.00	71.70	70.45	73.61	67.26	79.17	72.86
	SVM	**Six**	74.91	83.73	85.16	82.94	77.4	78.25	80.39	78.5	73.41	73.48	75.9	73.51	75.20	75.00
		FDR	79.17	86.9	81.25	81.67	83.89	88.75	83.60	65.00	82.05	73.86	67.59	79.76	81.25	74.91
	ANN	**Six**	79.17	79.76	83.75	80.83	75.56	81.87	80.16	75.00	68.80	70.45	71.76	67.26	75.69	71.49
		FDR	79.17	82.54	85.00	79.17	83.89	86.25	82.67	75.00	75.64	75.00	71.76	78.57	84.03	76.67

Notes:(i) This table reports financial distress prediction accuracy (for prediction of 1 and 2-years ahead status of a firm) of all the models considered for final comparison; (ii) Prediction accuracy (in percent) obtained for classification as per 2-factory and 3-factor definition is displayed for each of these models. (iii) ** indicates models estimated using 80% training and 20% testing sample data.