7. A Comparison of Static, Dynamic and Machine Learning Models in Predicting the Financial Distress of Chinese Firms

Umair Bin YOUSAF[1]
Khalil JEBRAN[2*]
Man WANG[3]

Abstract

As recent studies have begun to pay increasing attention to financial distress prediction (FDP), this study compares the performance of static, dynamic and machine learning (ML) models in predicting the financial distress of firms. Balanced and imbalanced datasets of Chinese listed firms that span the years 1992 to 2018 (27 years) and contain 29,000 firm-year observations are used for the comparison. Results show an incremental improvement in the FDP accuracy of these models, with the random forest (RF) ML model demonstrating the best performance, followed by the Dynamic Hazard model and the static model. Including the growth variable improved the predictive ability of these models. Whilst all these models demonstrated a decreasing prediction accuracy in the imbalanced dataset, a weaker trend was observed for ML models. These results support the debate regarding the FDP superiority of computational methods. The findings provide new evidence that ML models can predict financial distress more accurately as compared to conventional models in the Chinese context.

Keywords: financial distress prediction, static, dynamic, machine learning, growth, China
JEL Classification: C52, C63, G17, G33

1. Introduction

Many studies have recently examined the topic of financial distress prediction (FDP) or forecasting the tendency for a firm to enter bankruptcy or be liquidated (Barboza *et al*., 2017;

1 *Doctoral Student at School of Accounting, Dongbei University of Finance and Economics, China Internal Control Research Center, Dalian, China.*

2* *Associate Professor at School of Business Administration, Dongbei University of Finance and Economics, Dalian, China. Corresponding author. E-mail: khaliljebranuom@gmail.com, khaliljebran@dufe.edu.cn.*

3 *Professor at School of Accounting, Dongbei University of Finance and Economics, China Internal Control Research Center, Dalian, China.*

Sun *et al.*, 2014). Long-term investors are particularly concerned about the ability of firms to pay for their loans. In his pioneering study on FDP, Fitzpatric (1932) measured the predictive capacity of defaulted firms using economic indices. However, Altman (1968) in his seminal paper first time used the statistical method to predict financial distress. Research on FDP can be divided into three eras. The first era mainly involved the use of *logit* models, probit models, and multiple discriminant analysis (MDA), which, despite showing success in FDP, may not be applicable in certain situations (Beaver *et al.*, 2012) and ignore important explanatory variables of time. The second era began with Shumway (2001), who introduced the simple hazard model that considers time co-variates in FDP. Starting from the introduction of the first machine learning (ML) model for FDP by Tam (1991), the third era witnessed the continuous growth of ML models in FDP (Korol & Korodi, 2011; Le *et al.*, 2018; Sun *et al.*, 2014). However, Wang, Ma, and Yang (2014) argued that *'there is no mature or definite theory of corporate failure'* and that the performance of FDP models depends on the details and characteristics of the classification problem, the structure of the data, and the possibility of segregating classes based on these characteristics.

Whilst static and dynamic models (Shumway, 2001; Wu, Gaunt, & Gray, 2010) and traditional and machine learning methods have been compared in the literature (Barboza *et al.*, 2017), previous studies have mostly focused on developed economies, such as the US (Altman, 1968; Barboza *et al.*, 2017; Hillegeist *et al.*, 2004; Korol & Korodi, 2011; Ohlson, 1980; Taffler, 1983; Zmijweski, 1984), with only few focusing on emerging countries (*e.g.*, Wang & Campbell, 2010), including China (Li & Wang, 2018; Zhou, 2015; Wang & Campbell, 2010; Zhoua, Kim, & Ma, 2013). However, most of these studies do not follow any variable selection procedure and have limited sample sizes. In these studies, FDP is measured using time, market, liquidity, leverage, and profitability as variables (Bellovary *et al.*, 2007). However, only one study (Barboza *et al.*, 2017) used growth as a variable. To fill such a gap, we consider growth as a variable when comparing the performance of static, dynamic, and machine learning models in FDP.

Imbalanced datasets are common in real-life applications, and only a small proportion of firms operating in the market are considered distressed. Previous studies (*e.g*. Le *et al.*, 2018; Sun *et al.*, 2014) have proposed various solutions to problems emerging from imbalanced data. One proposed solution is to use learning algorithms, but these algorithms maximize accuracy without taking the minority class into consideration, thereby explaining their inability to generate meaningful classifiers (Le *et al.*, 2018). To prevent the type of data from influencing the results, both balanced and imbalanced datasets are used in this study to compare the performance of FDP models outside the context where these models were tested and developed (*i.e.*, China).

This study contributes to the literature in four ways. Firstly, this study compares the performance of static, dynamic, and machine learning models in forecasting the financial distress of firms. Whilst comparisons have also been conducted in previous studies, they are based on various contexts and datasets, thereby resulting in a lack of consensus regarding the performance of these models. Therefore, in this study, we compare different FDP models and find that ML models outperforms the other models. Secondly, the results of the comparison performed in this work can be replicated in both research and practice as such comparison is performed in a standardized computer setting. On the one hand, researchers may refer to these results to improve the comparability of their findings with those of other studies where similar models have been utilized. Moreover, industry practitioners can use the evaluated models in various situations, such as when assessing credit decisions. Thirdly, the comparison is performed using 27 years' worth of data on

Chinese listed firms. Such a large sample that covers more than 29,000 firm-year observations can improve the efficiency of models in forecasting the financial distress of firms. Moreover, to the best of our knowledge, this study is the first to consider a sample period of as long as 27 years to predict financial distress in China. In this way, the study underlines the overall financial distress situation in the Chinese stock market. Fourthly, this study highlights the role of growth in predicting the financial distress of firms, especially in the context of an emerging economy such as China.

The rest of this paper is organized as follows. Section 2 reviews the extant FDP models including those used in the current study and how financial distress is being handled in China. Section 3 describes the research methodology and the utilized data. Section 4 presents the empirical results. Section 5 concludes the paper, discusses the limitations of this work and proposes some directions for future research.

2. Literature review

Forecasting financial distress requires an accurate selection of estimation methods, financial ratios and samples. In terms of methods, researchers tend to construct FDP models, which are generally classified into static, dynamic, and ML models.

2.1 Static models

Static FDP models incorporate single-period predictor variables, with x_i representing the features for a certain period and y_i representing the predicted variable. In this case, samples are denoted by (x_i, y_i) in a static FDP model.

Altman (1968) adopted a balanced dataset of 66 US firms (of which 33 were bankrupt) covering the years 1946 to 1965 and where the bankrupt and non-bankrupt firms were segregated using multiple discriminant analysis. Ohlson (1980) used a sample of 2,163 US firms (of which 105 were bankrupt) covering the years 1970 to 1976 and where the financially distressed (FDF) and healthy firms (HF) were classified via logistic regression. Taffler (1983) used a balanced dataset of 46 bankrupt UK firms that covered the years 1969 to 1976 and where each bankrupt firm was matched with a non-bankrupt firm via linear discriminant analysis. Zmijweski (1984) divided a sample of 81 bankrupt firms into an estimation sample (including 40 firms) and a prediction sample (including 41 firms) and then applied probit regression to compare each of these samples with different combinations of 40 to 800 non-bankrupt firms.

Regardless of their wide application in previous research, static models have some limitations. Firstly, these models are fitted using the financial data of firms the year before the bankruptcy, thereby leading to a selection bias (Shumway, 2001). This is because financial distress condition changes throughout the life of a firm. Secondly, some variables used in these models are not time-covariates, and highlighting the deteriorating financial condition of a firm prior to its declaration of bankruptcy requires correct selection of variables (Shumway, 2001).

To represent static models, we used the logistic regression function in *R* software with some slight variations. The probability *p*(y =1| x_i) for firms to be financially distressed is estimated in the regression as:

$$p(y_i=1/x_i)=\frac{1}{1+e^{-(\alpha+\beta x_i)}}$$

where: *α* represents the intercept (scalar), and *β* represents the vector parameter.

2.2 Dynamic models

Unlike static models, dynamic models incorporate the 'time' aspect. The time lag between the action of causal variables and the effects on the dependent variable should be considered when assessing causal relationships. Incidents of manipulating accounting numbers in the short term through adjusting reserves, deferring payments, investing in future products and implementing revenue recognition and capitalization policies have also been frequently reported. Meanwhile, long-term performance measures are relatively difficult to manipulate as the true performance of firms becomes increasingly apparent over time (Carton & Hofer, 2006). Short-term measures have been frequently criticized for their questionable reliability. To address these problems, time-varying covariates or explanatory variables are incorporated into dynamic models.

The first dynamic model was developed by Shumway (2001) using variables that have never been used in earlier studies. Shumway argued that static models incorporate variables that are not associated with bankruptcy; instead, models for FDP should incorporate time-covariates derived from market or accounting variables. The *simple hazard* dynamic model of Shumway was built using data on 300 bankrupt and 2,882 non-bankrupt firms in the US from 1962 to 1992. Hillegeist *et al.* (2004) used the discrete hazard method to compare their proposed *BSM-Prob* model, which is based on market variables, with the models developed by Altman (1968) and Ohlson (1980). Their empirical results reveal that *BSM-Prob* outperforms the latter models. The current study uses dynamic hazard and generalized additive models as second era models.

2.2.1 Dynamic hazard model

Developed based on survival analysis, the *Dynamic Hazard* model is used to evaluate time to event data, that is, the remaining amount of time before the occurrence of an event, which is labelled either as a 'death' or 'failure'. Continuous (*i.e.*, cox proportional) and discrete time (*i.e.*, logistic) models are used in survival analysis. The *Dynamic Hazard* model is one example of a logistic model that utilizes survival analysis data. Right censored companies refer to those firms that have been excluded from the sample before the study period and have not experienced any financial distress at the end of the study period.

2.2.2 Generalized additive model

Developed by Hastie and Tibshirani (1986), the generalized additive model (*GAM*) utilizes a linear predictor that shows linear dependence on the unknown smooth functions of predictor variables. *GAM* is generally used to combine the characteristics of additive and generalized linear models and to make inferences about the aforementioned unknown smooth functions. A flexible generalization of ordinary linear regression, the generalized linear model (*GLM*) allows for dependent variables to not necessarily follow a normal distribution. Meanwhile, the additive model (AM) is a restricted non-parametric regression model that uses a one dimensional smoother. *GAM* can use either single- or multi-year data. Seeing that using multi-year data can enhance the predictive ability of *GAM* (Berg, 2007), a multi-year *GAM* is used as a dynamic model in this study.

2.3 ML models

ML models are amongst the most valued breakthrough solutions to classification problems *(Tian et al.*, 2012). These models effectively identify patterns in data and distinguish the observations for one group from those for another based on certain features (Barboza *et al*., 2017). As such, these models have attracted research attention across different fields, such as computing (Petersen & Ostendorf, 2009), medicine (Venkatesh *et al*., 2017) and engineering (Foster *et al*., 2014). ML can also effectively differentiate FDFs from HFs given that both groups tend to share similar features, such as their growth, market, size, liquidity, leverage and profitability. The existing research has employed different ML techniques in FDP including bagging, boosting, random forest (RF), support vector machines, decision trees, autoencoder and artificial neural networks (Barboza *et al*., 2017; Li & Wang, 2018; Wang, 2017). Amongst these techniques, given their simplicity, we only focus on *RF* and *Decision Tree* in the current study.

2.3.1 Decision Tree

As a non-parametric machine learning model designed for solving classification and regression problems, a decision tree is constructed based on predictor data and has been widely used in the literature to derive sequential and hierarchical decisions regarding the outcome variable. In FDP, *Decision Tree* models use a set of values called classification trees to locate the target variable. Each leaf or node in these trees represents class labels, each branch represents those feature combinations that lead to these labels and those paths leading from the root to leaf represent the classification rules. *Decision Tree* models are easy to comprehend given their excellent visual representation and their loose data requirements. Apart from their low computational cost, these models can accommodate large sizes of quantitative or qualitative data and are validated using statistical sets.

2.3.2 Random Forest

Although based on decision trees, *RF* prevents the common problem associated with such models. Introduced by Breiman (2001), *RF* can achieve a similar level of accuracy as AdaBoost. This model also achieves robustness by allowing the training sample to have outliers (Yeh *et al*., 2014). By recognizing the importance of each variable in its classification outcome, *RF* provides information about those factors that differentiate one group from another.

2.4 Model implementations

Different researchers have examined the three era models predominantly in developed country contexts. Begley *et al*. (1996) evaluated the use of static models in the US manufacturing industry during the 1980s and highlighted differences between the original and re-estimated models. Boritz *et al*. (2007) found that local FDP models outperform traditional ones in Canada. Barboza *et al*. (2017) validated the FDP superiority of ML models in North America.

Although previous studies on FDP have been mostly performed in the developed countries, few researchers have also explored such a topic in the developing countries. For instance, Sandin and Porporato (2011) tested and validated the acceptability of static models in Argentina by using a sample of 11 FDFs and 11 HFs. Wang and Campbell (2010a) used the static FDP model of Altman on a small sample of 84 publicly listed firms in China from 2000 to 2008 and validated the acceptable performance of both the original and re-estimated models. Wang and Campbell (2010b) validated the high predictive accuracy of Ohlson's (1980) model in the Chinese context by using five of its original ratios and a relatively small sample. Using the cox proportional hazard model, Zhoua *et al*. (2012) investigated publicly listed firms in China going through special treatment (ST) procedures. However, their findings have limited generalizability given the econometric shortcomings of their model (Li *et al*., 2021) and their small sample size. Using the same model, Bhattacharjee and Han

(2014) examined those macro and microeconomic factors that influence publicly listed firms in China from 1996 to 2006 and found that economic stability and microeconomic factors both have significant effects on the financial distress of these firms. Li *et a*l. 2021) employed the discrete time hazard model of Shumway (2001) to determine the corporate governance predictors of large Chinese firms and did not report any improvements in the predictive ability of this model. Ding *et al.* (2008) proposed an SVM-based model for FDP in the Chinese context and highlighted the superior performance of this model compared with conventional ones. Whilst Li and Wang (2018) compared the performance of statistical and ML methods in predicting the bankruptcy of Chinese listed companies, they only considered accounting measures as predictors.

2.5 Feature Selection

Feature selection is critical to FDP (Tsai, 2009). In their review of FDP models proposed between 1932 and 2007, Bellovary *et al.* (2007) identified some of the most frequently utilized features in the FDP literature. The importance of feature selection to FDP has also been supported by other authors (*e.g.* Farooq & Qamar, 2019; Habib *et al.*, 2018). These features generally include accounting indicators, such as leverage, profitability, market and liquidity, macroeconomic indicators, and corporate governance indicators (Habib *et al.*, 2018).

Growth plays an indispensable role in evaluating the performance of a firm (Hill *et al.*, 1996; Kohtamäki *et al.*, 2019). Accounting-based measures of growth often include the changes in the total assets, operating assets, total expenses, operating expenses and sales of a firm. Capon, Farley & Hoenig (1990) in their meta-analysis suggested that sales growth rate was a generally accepted performance indicator. Apart from growth in sales, empirical studies have also used the growth in the payroll and R&D expenses, number of employees and profits of a firm. However, whilst several studies have used various macroeconomic and accounting variables to measure organizational performance (*e.g.* Altman, 1968; Liu, 2004; Ohlson, 1980; Shumway, 2001; Taffler, 1983; Zmijweski, 1984), the importance of growth as a predictor has been ignored in these studies, except for Barboza, Kimura and Altman (2017), who predicted firm bankruptcy whilst incorporating growth into their model in the US. However, the findings of Barboza *et al*. (2017) were mainly based on a comparison of static and ML models and limited to developed country context.

To address the aforementioned gaps, we consider growth rate in sales and other commonly used variables in the literature, selected through stepwise regression, in comparing the FDP performance of static, dynamic and ML models. We attempt to identify which of these models achieves the best prediction accuracy in balanced and imbalanced datasets including when an additional variable is considered.

2.6 China's Financial Distress Handling System

The majority of the enterprises in China are owned by the state (*i.e.*, state-owned enterprises or SOEs). The economy of China greatly differs from that of other countries (Rashid *et al.*, 2019). Accordingly, China has a different way of handling financial distress. Specifically, affected firms in China need to undergo a special treatment (ST) (Zhou, 2017) procedure, which was introduced in 1998 by the China Securities Regulatory Commission. Firms satisfying three criteria are named ST: (1) those firms reporting negative profits for two consecutive years or earning a net capital per share that is below the per share value; (2) firms that purposefully distort any of the information in their financial statements; and (3) those firms demonstrating any of the abnormal behavior described in the Chinese Stock

Listing Exchange Rules. Those FDFs that fail to alleviate their distress after the ST status either delisted from the public exchange or have their operations suspended for a certain duration. Some of these companies even enter mergers or restructure themselves completely to escape their FDF status (Zhou, 2017; Zhoua *et al.*, 2013). An HF may also be tagged as an ST firm because of reasons other than experiencing bankruptcy. Similarly, the ST may ignore certain firms that are actually experiencing distress. In this study, we focus on those firms that have been suspended or delisted from the exchange as a result of their failure to cope with financial distress regardless of their participation or non-participation in ST.

3. Data and Methodology

3.1 Data

Data from 1992 to 2018 were extracted from the China Stock market and Accounting Research (CSMAR) database. Stock information of financial firms was ignored as financial firms have a capital structure different from that of non-financial firms (Shumway, 2001). The final sample was selected based on the following criteria:

a) A- and B-share firms are included;
b) Only non-financial firms (in line with the recommendations of Zhoua *et al.,* 2013) are considered;
c) firms suspended or delisted from the stock market after earning negative profits for two consecutive years regardless of their participation or non-participation in ST are included. These firms were tagged as FDFs;
d) firms suspended or delisted from the stock market for reasons other than net losses are excluded from sample; and
e) HFs are the firms that neither went through the ST procedure nor earned negative profits for two consecutive years during the research period (1992 to 2018). HFs were randomly selected from the available list of healthy firms.

Both balanced and imbalanced datasets were used to prevent the type of observations from affecting the results. Table 1 lists the number of firms in the balanced and imbalanced sets, whereas Table 2 presents their industry-wide distribution.

Table 1. Sample distribution of financially distressed (FDF) and healthy firms (HF)

	Total Sample	BALANCED SET		IMBALANCED SET	
		Train	Test	Train	Test
Financially Distressed Firms (FDF)	124	114	10	114	10
Healthy Firms (HF)	2893	114	10	1000	1893
Total sample	3017	228	20	1114	1903
Percentage of FDF in Total	4.11	50.00	50.00	10.23	0.53

Table 2. Industry wide distribution

Industry name	BALANCED SET				IMBALANCED SET				Total
	FDF		HF		FDF		HF		
	test	train	test	train	test	train	train	test	
Commerce	1	6	1	6	1	6	57	115	172
Conglomerates	2	16	2	16	2	16	35	70	105
Industrials	4	60	4	60	4	60	677	1366	2043
Properties	1	11	1	11	1	11	63	127	190
Public Utility	2	21	2	21	2	21	168	339	507
Total	10	114	10	114	10	114	1000	2017	3017

Firms with a minimum of 3 firm-year observations are included in the sample. As the research period covered 27 years, a sample firm can provide a maximum of 27 firm-year observations. Firms that went out of the study period without being financially distressed are considered censored and assigned the value of 0 in the last year of observation. Table 4 presents the descriptive statistics of the FDFs and HFs in the sample. Proportion of each industry was calculated prior to the formation of train and test sets.

Table 3. Variables' definitions

Variables	Measurement	Sources
Accounting ratios		
Return on assets (ROA)	Net income divided by total assets	Bellovary *et al*. (2007); Ohlson 1980)
Current ratio	Current assets divided by current liabilities	Bellovary *et al*. (2007); Zmijweski (1984)
Net working capital ratio	current assets minus current liabilities and then divided by total assets	Altman (1968); Bellovary *et al*. (2007); Ohlson (1980); Wu *et al*. (2010)
EBIT to Total assets ratio	Earnings before interest and taxes (EBIT) divided by total assets	Altman (1968); Bellovary *et al*. (2007); Wu *et al*. (2010)
Operating return on assets (OROA)	Operating profit divided by total assets	Bellovary *et al*. (2007)
Debt ratio (TLTA)	Total liabilities divided by total assets	Bellovary *et al*. (2007); Ohlson 1980); Shumway, (2001); Zmijweski (1984)
Interest coverage ratio	Earnings before interest and taxes over interest expense	Bellovary *et al*. (2007); Tinoco and Wilson (2013); Zhou (2019)
Retained earnings to total assets ratio	Retained earnings divided by total assets	Altman (1968); Altman *et al*. (2016)
Asset turnover ratio	Sales divided by total assets	Altman (1968); Wu *et al*. (2010)
Net income to total liabilities ratio	Net income divided by total liabilities	Zmijweski (1984)
Market ratios		
Relative size (RSIZE)	Log of the number of outstanding shares multiplied by year-end share price then divided by total market value	Campbell *et al*. (2008); Chan *et al*. (2016); Darrat *et al*. (2014); Oz and Mugan (2018); Shumway (2001); Wu *et al*. (2010)

Variables	Measurement	Sources
Sigma	The standard deviation of each firm's daily stock return over the past 3 months	Campbell *et al*. (2008); Chan *et al*. (2016); Darrat *et al*. (2014); Shumway (2001); Wu *et al*. (2010)
Ex return (EX_RETURN)	Cumulative annual return in year t-1 minus the value-weighted index return in year t-1.	Campbell *et al*. (2008); Chan *et al*. (2016); Oz and Mugan (2018); Shumway (2001); Wu *et al*. (2010)
Growth ratios		
Growth in sales (GS)	Sales in current year minus sales in prior year, then divided by sales in prior year	Altman *et al*. (2016); Barboza *et al*. (2017); Carton and Hofer (2006); Lohmann and Ohliger (2019)
Growth in assets (GA)	Total Assets (TA) in Year t minus TA in year t-1 then divided by TA in year t-1	Altman *et al*. (2016); Barboza *et al*. (2017); Carton and Hofer (2006)
Growth in return on equity (GROE)	Growth in ROE= ROE in Year t minus ROE in year t-1 then divided by ROE in year t-1	Barboza *et al*. (2017); Carton and Hofer (2006)
Macro-economic variables		
Growth in GDP	Annual percentage growth rate of GDP at market prices based on constant local currency	Bhattacharjee and Han, (2014); Cole, (2009); Li *et al*. (2021)
Unemployment rate	The share of the labor force that is without work but available for and seeking employment.	Bhattacharjee and Han (2014); Cybinski (2001); Li *et al*. (2021)
Inflation rate	Percentage price change in the economy as a whole in a year	Li *et al*. (2021); Liu (2004); Mensah (1984); Tinoco and Wilson (2013)
Real Interest rate	Interest rate adjusted for inflation as measured by the GDP deflator in a year	Bhattacharjee and Han (2014); Cole (2009); Cybinski (2001); Mensah (1984); Tinoco and Wilson (2013)
Corporate governance variables		
Board Size	Calculated as the total number of directors on the board	Adams and Ferreira (2009); Chaganti *et al*. (1985); Daily and Dalton (1994); Darrat *et al*. (2014)
CEO Duality	It equals 1 if the CEO is also the board chairman, otherwise 0	Chaganti *et al*. (1985); Daily and Dalton (1994); Li *et al*. (2021)
State ownership	Proportion of shares held by the state	Bhattacharjee and Han (2014); Li *et al*. (2008); Wang and Deng (2006)

Different methods are available for feature selection (Tsai, 2009). We employed stepwise regression to construct our baseline model, into which we incorporated the growth, accounting and market variables we identified from the literature. We chose to exclude macroeconomic and corporate governance variables as they had no effects on the predictive power of FDP models. Table 3 lists those variables used in the stepwise regression. Those variables whose VIF values exceeded 5 were excluded to avoid multicollinearity. The same winsorization procedure used in previous studies (*e.g.* Beaver *et al*., 2012; Shumway, 2001) was applied for all variables at the 1% and 99% levels.

Table 4. Descriptive Statistics

Total sample	ROA	TLTA	EX_RETURN	RSIZE	Sigma	GS	OPTA
Min.	-0.6032	0.0656	-0.8139	0.0000	0.0326	-0.7078	-0.3315
1st Qu.	0.0190	0.3179	-0.2063	0.0001	0.0465	-0.0317	0.0089
Median	0.0664	0.4744	-0.0264	0.0002	0.0654	0.1160	0.0341
Mean	0.1170	0.4753	0.0842	0.0004	0.0744	0.2088	0.0309
3rd Qu	0.1637	0.6219	0.2310	0.0005	0.0993	0.2979	0.0661
Max	1.2989	1.1535	2.3321	0.0040	0.1349	3.9244	0.2194
SD	0.2459	0.2141	0.5095	0.0006	0.0314	0.5688	0.0766
FDF Sample	ROA	TLTA	EX_RETURN	RSIZE	Sigma	GS	OPTA
Min.	-0.6032	0.1926	-0.8139	0.0000	0.0326	-0.7078	-0.3315
1st Qu.	-0.4788	0.8890	-0.3782	0.0000	0.0615	-0.6226	-0.3315
Median	-0.2431	1.1535	-0.1884	0.0001	0.0660	-0.2966	-0.2730
Mean	-0.2929	0.9803	-0.1354	0.0001	0.0756	-0.2089	-0.2298
3rd Qu	-0.1195	1.1535	-0.0009	0.0002	0.0776	-0.0436	-0.1356
Max	-0.0035	1.1535	2.3321	0.0034	0.1349	3.9244	0.0297
SD	0.1946	0.2642	0.4420	0.0003	0.0266	0.6658	0.1126
HF Sample	ROA	TLTA	EX_RETURN	RSIZE	Sigma	GS	OPTA
Min.	-0.6032	0.0656	-0.8139	0.0000	0.0326	-0.7078	-0.3315
1st Qu.	0.0193	0.3173	-0.2056	0.0001	0.0465	-0.0301	0.0091
Median	0.0669	0.4735	-0.0258	0.0002	0.0654	0.1169	0.0343
Mean	0.1185	0.4734	0.0850	0.0004	0.0744	0.2104	0.0319
3rd Qu	0.1644	0.6205	0.2313	0.0005	0.0993	0.2987	0.0662
Max	1.2989	1.1535	2.3321	0.0040	0.1349	3.9244	0.2194
SD	0.2448	0.2116	0.5096	0.0006	0.0314	0.5678	0.0747

Note: First; total dataset, Second; Financially Distress Firms (FDFs), Third; Healthy Firms (HF).

3.2 Methodology

The compared models were implemented using the *R* statistical software. For first era models, *logistic regression* and *static_glm* are used. The *glm( )* function with an attached binomial family link function was applied for the logistic regression. The model outputs a probability whose value ranges from 0 to 1. The *static_glm* function in the *dynamichazard* package was used to apply the *static glm* model. Unlike logistic regression, *static_glm* is a static version of the *ddhazard* model fit that uses data for *survival analysis* where firm-year observations are completely distinct from one another. Static_glm also consumes less memory by using weights.

For second era dynamic models, Generalized Additive and Dynamic Hazard models are employed. For *Generalised Additive and Dynamic Hazard,* the *GAM* function from the *mgcv* package, the *ddhazard* function from the *dynamichazard* package in R setting are used. Meanwhile, *Decision Trees* and *RF* models were implemented using the *ctree* function from the *party* package and the *rfsrc* function from the *randomforestSRC* package, respectively for ML models.

The usual start and stop times format observed in survival analysis were followed in the chosen datasets, and each firm considered in the trial included at least one observation. Table 5 presents some observations from these datasets. Column *code* presents the ID of the selected firms. Columns *tstart* and *tstop* indicate the range of valid rows. Values 1 and 0 in column *FD* indicate that the firm is an FDF or HF at *tstop*, respectively.

For comparing the models, the training and test sets are calculated using different packages and measures, including ROC curves, specificity, accuracy, types I and II errors and sensitivity. An inaccurate classification of an FDP model can lead to losses. Accordingly, lenders and owners focus on improvements in sensitivity than in specificity. Given the trade-off between specificity and sensitivity, we built an accuracy table by setting a high sensitivity threshold level. The types I and II errors indicate the proportion of FDFs and HFs predicted by FDP models as financially healthy and financially distressed, respectively. The number of accurate classifications was divided by the total number of classifications to determine the accuracy of an FDP model.

Table 5. Sample Dataset

Code	*tstart*	*tstop*	FD	NPTA	TLTA	Ex Return	Relative size	SIGMA	GRS	OPTA
2	0	1	0	0.1486	0.5645	0.1080	0.0040	0.1349	0.6391	0.0842
2	1	2	0	0.1198	0.5709	0.2030	0.0030	0.1349	0.1324	0.0788
2	2	3	0	0.0891	0.5788	-0.2646	0.0022	0.1070	0.2250	0.0587
2	3	4	0	0.0786	0.5658	-0.1780	0.0026	0.1287	-0.2168	0.0521
2	4	5	0	0.1024	0.4980	1.3976	0.0028	0.1118	0.6537	0.0536
2	5	6	0	0.1091	0.4795	0.5755	0.0017	0.0644	0.1533	0.0546
2	6	7	0	0.1023	0.5225	-0.2704	0.0016	0.1253	0.2790	0.0602
3	0	1	0	0.1609	0.5467	2.1452	0.0040	0.1349	0.3381	0.0885
3	1	2	0	0.0901	0.5796	0.1090	0.0025	0.1349	0.0173	0.0537
3	2	3	0	0.0176	0.6416	-0.2761	0.0024	0.1070	-0.1596	0.0144
3	3	4	0	-0.0142	0.6697	-0.0802	0.0023	0.1287	-0.2563	-0.0085
3	4	5	0	0.0120	0.7043	0.7104	0.0010	0.1118	-0.2150	0.0099
3	5	6	0	-0.1036	0.8586	-0.4759	0.0007	0.0644	-0.0036	-0.0875
3	6	7	0	-0.0631	0.9657	-0.1642	0.0005	0.1253	-0.0835	-0.0504
3	7	8	1	-0.2000	1.1535	-0.1080	0.0004	0.0604	-0.0327	-0.2410

Note: Code shows a company's code assigned by stock exchange, tstart column represents when the row is valid from and the tstop column indicates when the row is valid to.

3.2.1 Correlation

Both the balanced and imbalanced training datasets were subjected to a correlation analysis given that the nature of data influences the correlation. The correlation amongst variables is highlighted in Table 6. Given the presence of high correlations amongst some variables, the potential presence of multicollinearity was tested by conducting a VIF test. WCTA was removed after obtaining a VIF value of 9.18. After removing this variable, another VIF test was performed, and all variables obtained VIF values of lower than 5. Following previous studies (Altman, 1968; Barboza *et al*., 2017), the remaining seven variables with correlations were retained.

Table 6. Correlation analysis

Balanced set	NPTA	TLTA	Ex Return	Relative size	Sigma	GRS	WCTA	OPTA
NPTA	1.000							
TLTA	-0.676***	1.000						
$r_{it-1} - r_{mt1}$	0.138*	-0.119*	1.000					
Relative size	0.195**	-0.182**	0.384***	1.000				
Sigma	-0.639***	0.499***	0.088	-0.138*	1.000			
GRS	0.616***	-0.433***	0.246***	0.229***	-0.380***	1.000		
WCTA	0.630***	-0.873***	0.137*	0.252***	-0.503***	0.405***	1.000	
OPTA	0.945***	-0.779***	0.190**	0.250***	-0.627***	0.629***	0.746***	1.000
Imbalanced set	NPTA	TLTA	Ex Return	Relative size	Sigma	GRS	WCTA	OPTA
NPTA	1.000							
TLTA	-0.625***	1.000						
$r_{it-1} - r_{mt1}$	0.159***	-0.019***	1.000					
Relative size	0.207***	0.026***	0.151***	1.000				
Sigma	-0.023***	0.046***	0.073***	0.154***	1.000			
GRS	0.258***	-0.003	0.133***	0.141***	-0.060***	1.000		
WCTA	0.450***	-0.611***	0.039***	-0.027***	-0.072***	0.056***	1.000	
OPTA	0.887***	-0.384***	0.178***	0.309***	-0.032***	0.329***	0.317***	1.000

Note: ∗ $p < 0.05$; ∗∗ $p < 0.01$; and ∗∗∗ $p < 0.001$

4. Empirical Results

The performance of each FDP model was compared in both balanced and imbalanced datasets, and the results are presented in Table 7. Such performance was assessed by taking specificity, accuracy, sensitivity, AUC and types I and II errors as evaluation measures. The *RF* model obtained the highest accuracy and lowest error for both balanced and imbalanced datasets. The *RF* model was followed by the *Dynamic Hazard*, *Decision Tree*, *GAM* and *logit* models, with s*tatic GLM* demonstrating the worst performance.

At the training stage, both the *RF* and *Decision Tree* models outperform all the other models in terms of accuracy, which may be ascribed to the fact that these models can lead to overfitting. However, the fineness of these models cannot be guaranteed based on their accuracy alone. The *Decision Tree* model faces a more severe overfitting problem, especially with a high tree depth, relative to *RF* models. The *RF* models in turn effectively reduce overfitting by aggregating multiple decision trees.

However, these two models demonstrate a significant reduction in their accuracy at the testing stage. Amongst these models, *RF* demonstrates the best accuracy in the training and testing sets, followed by *Dynamic Hazard* and *Decision Tree*. The performance of these models deteriorates further as the number of non-FDF observations increase and as the data become increasingly imbalanced. These models obtain a very low type I error given the very small number of FDFs in the imbalanced test set. *RF* has consistently outperformed all models regardless of the nature of the data.

In both the balanced and imbalanced datasets, the *RF* model has also outperformed all the other models except for *Dynamic Hazard* in the imbalanced test set. Such result can be ascribed to the fact that all the available information for each sample firm can be utilized by *Dynamic Hazard*. These two models also demonstrate high robustness.

The above models also achieve a very high accuracy as they do not require a linear structure. Despite effectively classifying FDFs, these models obtain a high type II error, which indicates their high tendency of misclassifying HFs. Between these models; *RF* obtains the lowest type II errors in both the balanced (0%) and imbalanced datasets (5.62%).

Table 7. Comparison of FDP Models

Balanced Data Train Set	Type I Error	Type II Error	Sensitivity	Specificity	Accuracy%	AUC (%)
Logit	1.75	12.28	98.25	87.72	92.98	99.30
Static GLM	9.65	84.21	90.35	15.79	53.07	57.70
GAM	0.88	12.28	99.12	87.72	93.42	99.40
Dynamic Hazard	2.63	0.88	97.37	99.12	98.25	99.50
Decision tree	0.88	2.63	99.12	97.37	98.25	99.10
Random Forest	0.00	0.00	100.00	100.00	100.00	100.00
Test Set						
Logit	10.00	20.00	90.00	80.00	85.00	89.00
Static GLM	10.00	20.00	90.00	80.00	85.00	89.00
GAM	10.00	20.00	90.00	80.00	85.00	89.00
Dynamic Hazard	0.00	20.00	100.00	80.00	90.00	91.50
Decision tree	0.00	20.00	100.00	80.00	90.00	89.00
Random Forest	0.00	0.00	100.00	100.00	100.00	100.00
Imbalanced Data Train Set	Type I Error	Type II Error	Sensitivity	Specificity	Accuracy%	AUC (%)
Logit	5.26	23.00	94.74	77.00	77.66	93.79
Static GLM	9.65	89.45	90.35	10.55	13.52	51.30
GAM	6.14	22.39	94.74	77.61	78.25	94.60
Dynamic Hazard	16.67	13.40	89.47	86.60	86.71	98.41
Decision tree	16.67	7.73	83.33	92.27	91.93	95.20
Random Forest	0.00	0.00	100.00	100.00	100.00	100.00
Test Set						
Logit	0.00	23.74	100.00	76.26	76.27	98.89
Static GLM	0.00	90.28	100.00	9.72	9.75	98.90
GAM	0.00	18.85	100.00	81.15	81.15	99.00
Dynamic Hazard	0.00	9.14	100.00	90.86	90.87	99.40
Decision tree	0.00	11.37	100.00	88.63	88.63	98.60
Random Forest	0.00	5.28	100.00	94.72	94.72	99.06

Note: The total number of firms used in analysis is 3017.Type I error indicates the percentage of Financially Distressed Firms (FDFs) that were incorrectly classified as Healthy Firms (HF). Type II error shows the percentage of HF that were predicted to be FDFs. AUC is the area under the ROC curve.

The ROC curves of each model compared in this study are presented in Figures 1(a), 1(b), 2(a) and 2(b). For brevity, all figures are made available as online supplementary files. The curves highlight *RF* as the best model. Given the very small proportion of FDFs in the test sets, the performances of the static *GLM* and *GAM* models have improved in the test set. Whilst the *GAM*, *Dynamic Hazard* and *RF* models have AUCs of 99% or above in the imbalanced test set, *RF* achieves a significantly higher accuracy compared with the other two models. The *Dynamic Hazard* model demonstrates an efficient performance in both the balanced and imbalanced datasets and even reports the second-lowest type II error of 9.14

after *RF*. With its efficiency, the *Dynamic Hazard* model can be used to replace complex ML techniques in FDP.

In this study, we also compared the performances of models with and without growth variables in the balanced and imbalanced datasets. For brevity, the results are made available as online supplementary files. The addition of the growth variable improves the accuracy of all three models. The *VIMP* function of the *randomforestSRC* package is used to test the explanatory power of each variable for financial distress. Results show that growth significantly contributes to explaining the dependent variable. Adding this variable to *Dynamic Hazard* increases its AUC from 97.59 to 99.40, and such increase highlights the time-varying effect of sales growth on FDP.

In sum, *RF* emerges as the best model for FDP regardless of the inclusion of the growth variable. Meanwhile, ML models achieve the best prediction accuracy, followed by dynamic models. These findings not only support the contentions of previous studies that have underscored the superiority of ML models over traditional ones (Barboza *et al.*, 2017; Li & Wang, 2018) but also emphasize the key role of growth as a predictor of financial distress.

5. Conclusion

This study compares the performances of static, dynamic, and ML models in predicting the financial distress of publicly listed companies in China. Whilst financial distress has received much attention in the literature, the financial distress of Chinese firms warrants further investigation given the unique environment of China. With their increasing applications in the field of finance, the potential of ML models in FDP has attracted academic attention. Whilst outperformed by the newer-generation ML models, both static and dynamic models remain useful in the FDP of Chinese firms. Moreover, given its simplicity, practicality and consistent framework, the conditional *logit* model has been used in FDP. Therefore, future studies should consider combining these three types of models. The model comparisons presented in this work can also benefit studies that focus on credit risks.

We highlight how our study contributes to the extant literature. First, the study is conducted in the context of China, which is unique because of capital market structure, corporate governance, and the government's role (Li & Wang, 2018; Liu, Qi, Qin, & Su, 2019). China has approximately five percent of total corporate financing funded by equity investors. Listed firms are mostly state-owned, and rely heavily on retained earnings and bank loans which are different from developed countries like US where companies depend more on equity financing. It can be inferred that an FDP study conducted in the Chinese context may reveal different results. Second, the study sample comprised of 27 years of data of Chinese financially distressed firms, which has not been done before in the literature. The large test set from imbalanced dataset helps in generalizing the results to overall financial distress prediction of Shanghai and Shenzhen stock exchanges. Further, the study thoroughly checks the impact of growth variables on financial distress prediction. Accuracy of all the models increases when growth variable is used in addition to other widely used variables in the literature.

We also acknowledge some limitations that may be explored in future research. Firstly, this work mostly relied on default algorithms in R, which prevent us from using the compared models in their full capacity. However, using simple algorithm settings can also highlight the superiority of the *RF* model over the other models. Secondly, many variables presented in the FDP literature were not considered in this work although a stepwise regression was applied for the features selection. Future studies should consider incorporating other factors, including corporate social responsibility and auditor characteristics, as variables. Thirdly, this study did not discuss the thresholds and threshold level selection even though a threshold level for reducing type I errors was applied.

Some implications can be derived from the results. Firstly, unusual in the FDP literature, a large sample of 29,000 firm-year observations was used in this study for comparing static, dynamic, and ML models. In this way, the comparison results are applicable across different credit portfolios and financial institutions and may even be used to examine corporate failures in the stock markets of Shenzhen and Shanghai. Secondly, the findings of this work can help future studies in highlighting the supremacy of ML models over static and dynamic models, which are frequently applied in the loan market. Thirdly, managers can take a hint from these findings and adopt both ML and dynamic models in their prediction. Whilst the superior performance of ML models over static and dynamic ones may be difficult to explain, the complex nature of FDP highlights the potential usefulness of these models in analyzing credit default. Given its key role in improving the predictive ability of FDP models, growth should also be regularly considered in credit risk measurements and financial distress studies.

References

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

Barboza, F., Kimura, H. and Altman, E., 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, pp. 405–417.

Beaver, W.H., Correia, M. and McNichols, M.F., 2012. Do differences in financial reporting attributes impair the predictive ability of financial ratios for bankruptcy? *Review of Accounting Studies*, 17(4), pp. 969–1010.

Begley, J., Ming, J. and Watts, S., 1996. Bankruptcy classification errors in the 1980s: an empirical analysis of Altman’s and Ohlson’s models. *Review of Accounting Studies*, 1(4), pp. 267–284.

Berg, D., 2007. Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry,* 23, pp. 129–143.

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

Boritz, J.E., Kennedy, D.B. and Sun, J.Y., 2007. Predicting Business Failures in Canada. *Accounting Perspective*, 6(2), pp. 141–165.

Breiman, L., 2001. Random Forests. *Machine Learning*, 1(45), pp. 5–32.

Capon, N., Farley, J.U. and Hoenig, S., 1990. Determinants of financial performance. A meta-analysis. *Management Science*, 36(10), pp. 1143–1158.

Carton, R.B. and Hofer, C.W., 2006. *Measuring organizational performance: Metrics for entrepreneurship and strategic management research*. Edward Elgar Publishing limited.

Ding, Y., Song, X. and Zen, Y., 2008. Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, 34(4), pp. 3081–3089.

Farooq, U. and Qamar, M.A.J., 2019. Predicting multistage financial distress: Reflections on sampling, feature and model selection criteria. *Journal of Forecasting*, 38(7), pp. 632–648.

FitzPatric, P., 1932. A comparison of ratios of successful industrial enterprises with those of failed companies. *The Certified Public Accountants*.

Foster, K. R., Koprowski, R. and Skufca, J.D., 2014. Machine learning, medical diagnosis, and biomedical engineering research - commentary. *BioMedical Engineering Online*, 13(1), pp. 1–9.

Habib, A., Costa, M.D., Huang, H.J., Bhuiyan, M.B.U. and Sun, L., 2018. Determinants and consequences of financial distress: review of the empirical literature. *Accounting and Finance*, 60, pp. 1023–1075.

Hastie, T. and Tibshirani, R., 1986. Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to Statistical Science. *Statistical Science*, 10(4), pp. 354–363.

Hill, R., Murphy, G. and Trailer, J., 1996. Measuring performance in entrepreneurship research. *Journal of Business Research*, 36(1), pp. 15–23.

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

Kohtamäki, M., Heimonen, J. and Parida, V., 2019. The nonlinear relationship between entrepreneurial orientation and sales growth: The moderating effects of slack resources and absorptive capacity. *Journal of Business Research*, 100, pp. 100–110.

Korol, T. and Korodi, A., 2011. An evaluation of effectiveness of fuzzy logic model in predicting the business bankruptcy. *Romanian Journal of Economic Forecasting,* 14(3), pp. 92–107.

L. Bellovary, J., Giacomino, D. and Akers, M., 2007. A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 33, pp. 1–42.

Le, T., Lee, M.Y., Park, J.R. and Baik, S.W., 2018. Oversampling techniques for bankruptcy prediction: Novel features from a transaction dataset. *Symmetry*, 10(4), pp. 1–13.

Li, Y. and Wang, Y. 2018. Machine Learning Methods of Bankruptcy Prediction Using Accounting Ratios. *Open Journal of Business and Management*, 6(1), pp. 1–20.

Li, Z., Crook, J., Andreeva, G. and Tang, Y., 2021. Predicting the risk of financial distress using corporate governance measures. *Pacific-Basin Finance Journal*, 68. pp. 101334.

Liu, J., 2004. Macroeconomic determinants of corporate failures: Evidence from the UK. *Applied Economics*, 36(9), pp. 939–945.

Liu, S., Qi, X.Z., Qin, M. and Su, C.W., 2019. Financial stability or instability? Impact from Chinese consumer confidence. *Romanian Journal of Economic Forecasting*, 22(4), pp. 25–43.

Ohlson, J.A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), pp. 109-131.

Petersen, S. E. and Ostendorf, M., 2009. A machine learning approach to reading level assessment. *Computer Speech and Language*, 23(1), pp. 89–106.

Rashid, A., Jehan, Z., Khan, M. and Saeed, M., 2019. The influence of capital flow bonanzas on real and financial sectors: a comparative study of South Asia and China. *Romanian Journal of Economic Forecasting*, 22(4), pp. 124–147.

Sandin, A. R. and Porporato, M., 2011. Corporate bankruptcy prediction models applied to emerging economies Evidence from Argentina in the years 1991-1998. *International Journal of Commerce and Management*, 17(4), pp. 295–311.

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

Sun, J., Li, H., Huang, Q.H. and He, K.Y., 2014. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, 57, pp. 41–56.

Taffler, R.J., 1983. The Assessment of Company Solvency and Performance Using a Statistical Model. *Accounting and Business Research*, 13(52), pp. 295-308.

Tam, K., 1991, Neural network models and the prediction of bank bankruptcy. *Omega*, 19(5), pp. 429–445.

Tian, Y., Shi, Y. and Liu, X., 2012. Recent advances on support vector machines research. *Technological and Economic Development of Economy*, 18(1), pp. 5–33.

Tsai, C.F., 2009. Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, 22(2), pp. 120–127.

Wang, G., Ma, J. and Yang, S. 2014. Expert Systems with Applications an improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, 41(5), pp. 2353–2361.

Wang, N., 2017. Bankruptcy Prediction Using Machine Learning. *Journal of Mathematical Finance*, 7(4), pp. 1-11.

Wang, Y. and Campbell, M., 2010a. Business failure prediction for publicly listed companies in China. *Journal of Business and Management*, 16(1), pp. 75–88.

Wang, Y. and Campbell, M., 2010b. Financial ratios and the prediction of bankruptcy: the Ohlson model applied to Chinese publicly traded companies. *ASBBS Annual Conference*, 17, pp. 334–338.

Wu, Y., Gaunt, C. and Gray, S., 2010.A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting and Economics*, 6(1), pp. 34–45.

Yeh, C.C., Chi, D.J. and Lin, Y.R., 2014. Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, 254, pp. 98–110.

Zhou, L., 2015. A comparison of dynamic hazard models and static models for predicting the special treatment of stocks in China with comprehensive variables. *Journal of the Operational Research Society*, 66(7), pp. 1077–1090.

Zhou, Y., 2017. A Comparative Study of China and the US Delisting System. *American Journal of Industrial and Business Management*, 7(7), pp. 855–863.

Zhoua, Y.A., Kim, M. H.-Y. and Ma, S. 2012, Survive or Die? An Empirical Study on Chinese ST Firms. *International Conference of the American Committee for Asian Economic Studies (ACAES) Melbourne, Australia: Deakin University*, 1–32.

Zmijweski, M.E., 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, pp. 59–82.