International Journal of Forecasting 38 (2022) 1100–1115

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The recurrence of financial distress: A survival analysis

Fanyin Zhou [a], Lijun Fu [b], Zhiyong Li [b,c,*], Jiawei Xu [b]

[a] School of Statistics and Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu 611130, China
[b] School of Finance and Fintech Innovation Center, Southwestern University of Finance and Economics, Chengdu 611130, China
[c] Collaborative Innovation Center of Financial Security, Southwestern University of Finance and Economics, 555 Liutai Avenue, Chengdu 611130, China

ARTICLE INFO

Keywords:
Recurrence of financial distress
Special treatment
Survival analysis
Stratified hazard model
Recurrent event prediction

ABSTRACT

Companies often suffer periods of financial distress before filing for bankruptcy. Unlike one-off bankruptcies, financial distress can occur repeatedly within the same individual firm. This paper is focused on the recurrence of financial distress and studies the Chinese stock market, where Special Treatment – an official indicator of financial distress – can be repeatedly applied to a listed company. We employ a stratified hazard model to predict the probability of subsequent distress with variables, including duration dependency, event-based factors, institutional variables, financial ratios, market-based variables and macroeconomic conditions. Our empirical results show that accounting and market-based variables have limited power in predicting the recurrence of distress, whereas the duration of recovery, restructuring events and their interaction terms with the accounting and macroeconomic factors affect the recurrent risk significantly. Tested on out-of-time samples, our proposed hazard models show a robust performance in the prediction of recurrent risk over time.

1. Introduction

Corporate bankruptcy is one of the events which has received much attention in financial studies, as it is likely to bring huge losses to creditors. Bankruptcy prediction has a long history, dating back to the works of Beaver (1966) and Altman (1968). However, most companies do not suddenly go bankrupt, but rather go through a period of financial distress for a considerable length of time beforehand. Financial distress is, therefore, a useful sign of impending bankruptcy or corporate default in credit risk management, and the success of financial distress prediction can give an early warning of potential losses to investors and creditors.

In Chinese financial distress studies (*e.g.*, Hu & Zheng, 2015; Sun, Jia, & Li, 2011; Wu, Liang, & Yang, 2008), it is common to apply 'Special Treatment' (ST) as the indicator of financial distress, a measure which was introduced by the regulator, the China Securities Regulatory Commission (CSRC) in 1998. It is designed to initiate a warning signal to investors on the stocks of the listed firms with abnormal financial conditions or suffering severe financial conditions. More detailed background on the ST regulation can be found in Kim, Ma, & Zhou (2016). In general, ST is characterised by negative net profit over two consecutive fiscal years. The financial condition of a firm can either improve or deteriorate further, ending up by being delisted.

Nearly all literature in corporate credit management is only focused on the first occurrence of financial distress, leaving future behaviours unobserved or censored. In fact, many distressed firms can in fact recover from the brink of financial difficulties and return to normal. Empirical research shows that 52% of sampled firms with a negative financial situation go through management

* Correspondence to: Room 521, School of Finance, SWUFE, 555 Liutai Avenue, Wenjiang, Chengdu, Sichuan, China.
E-mail addresses: zfy@swufe.edu.cn (F. Zhou), fulijun@smail.swufe.edu.cn (L. Fu), liz@swufe.edu.cn (Z. Li), xujiawei@smail.swufe.edu.cn (J. Xu).

https://doi.org/10.1016/j.ijforecast.2021.12.005
0169-2070/

turnover (Gilson, 1989) as well as comprehensive organisational changes in terms of governance and structure (Wruck, 1990). O'Neill (1986) examined the relationship of contextual factors to the effectiveness of four primary turnaround strategies which are management, cutback, growth and restructuring. The management of these companies attempts to cope with financial difficulty via resolution methods such as asset liquidation, which is accompanied by value losses (John, 1993). A recent study suggests that self-efficacy plays a positive role in avoiding financial distress and mitigating adverse financial shocks (Kuhnen & Melzer, 2018).

The nature of financial distress is such that some distressed firms can either fail or eventually find themselves delisted, some may recover and stay healthy for a while, while some may fall into distress once again. As evidenced by Kahl (2002), many distressed firms emerging from a debt restructuring remain highly leveraged, continue to invest little, perform poorly and often re-enter financial distress. However, there is a gap in the literature related to corporate credit management in that so far, no attention has been paid to distress relapse. This paper, therefore, aims to bridge the gap by proposing a stratified hazard model to investigate the recurrence risk of financial distress.

This study makes several contributions to the literature. First, it sets a precedent by predicting the recurrence of financial distress in a survival analysis framework. Second, we take account of the history of distress into prediction, where the behaviours of the firms during the first episode of financial distress and the recovering process are observed. They show significant impacts on the probability of future distress. Third, we use a stratified hazard model to accommodate the interactions between repeated distress and provide out-of-sample predictions on the distress recurrence of Chinese distressed firms.

The remainder of this paper is organised as follows. Section 2 reviews the relevant literature of financial distress prediction and the recurrent event analysis. Section 3 shows the descriptive statistics of distress recurrence under the multi-state framework and the comparative analysis between recovered and re-distressed companies. In Section 4, we introduce the data, model specification and explanatory variables for the analysis of recurrent distress. The empirical results and measures of predictive performance are presented in Section 5. Lastly, some concluding remarks can be found in Section 6.

2. Literature review

The signals reflecting the operating situation of a company can reveal signs of financial distress, which can then be integrated into predictive models. Beaver (1966) was the first to introduce financial ratios into bankruptcy prediction, and since then financial ratios have been the key piece of information in financial distress prediction for decades (Altman, 1968; Ohlson, 1980). Market-based information can give us a timely prediction, that is, under the assumption of efficient markets, the market price integrates all views into the future (Bharath & Shumway, 2008; Merton, 1974). Other information to be considered includes corporate governance (Li, Crook, Andreeva, & Tang, 2021; Platt & Platt, 2002), corporate efficiency (Li, Crook, & Andreeva, 2017; Paradi, Asmild, & Simak, 2004), external resource factors (Hu & Ansell, 2007) and macroeconomic factors (Duffie, Saita, & Wang, 2007; Tinoco & Wilson, 2013). Additionally, in recent years, unstructured data have drawn considerable attention in corporate studies. By utilising convolutional neural networks in information extraction, Mai, Tian, Lee, & Ma (2019) used textual data and Hosaka (2019) used image data constructed from financial statements to predict corporate bankruptcy.

Bankruptcy or financial distress prediction studies have employed statistical analysis and data mining techniques to enhance the decision support tools and improve decision-making (Yang, You, & Ji, 2011). Altman (1968) pioneered the use of multiple discriminant analysis (MDA) which was further developed by Deakin (1972), Edmister (1972) and others. Later, logistic regression (or Logit) supplanted the Z-score, as it can generate probabilistic results (Martin, 1977; Ohlson, 1980), which is later a requirement under Basel II. Since the last decade of the twentieth century, machine learning algorithms have started to appear in the literature. Tam & Kiang (1992) and Lacher, Coats, Sharma, & Fant (1995) used neural networks in classifying both bankrupted and normal listed firms. There are other innovative algorithms including genetic algorithm (Back, Laitinen, & Sere, 1996), Rough Sets (Dimitras, Slowinski, Susmaga, & Zopounidis, 1999; Li, Wang, & Deng, 2008; Wang & Li, 2007), decision trees (Geng, Bose, & Chen, 2015), support vector machines (Hua, Wang, Xu, Zhang, & Liang, 2007; Min & Lee, 2005) and many hybrid and ensemble models such as du Jardin (2017), Choi, Son, & Kim (2018) and Sun et al. (2011). Mathematical programming is yet another stream of algorithms. Data envelopment analysis (DEA) is a nonparametric method that can compare company peers and calculate their relative efficiencies measured by the distance to the efficient frontier. The applications of DEA to bankruptcy and financial distress predictions can be found in Paradi et al. (2004), Cielen, Peeters, & Vanhoof (2004), Li, Crook, & Andreeva (2014), Li et al. (2017), etc. Machine learning and mathematical programming algorithms are often criticised for their lack of transparency and explanation. The discussions of various models were reviewed by Altman, Marco, & Varetto (1994), Balcaen & Ooghe (2006), Kumar & Ravi (2007) and Verikas, Kalsyte, Bacauskiene, & Gelzinis (2010).

While the above techniques model financial distress as a classification problem, a survival analysis technique is concerned with both the timing and occurrence of the event. Survival analysis can also take advantages of incorporating time-varying covariates and censoring into modelling, which is superior to static classification algorithms. Lane, Looney, & Wansley (1986) pioneered the use of the Cox proportional hazard for predicting bank failure. Luoma & Laitinen (1991) used Cox proportional hazard models to predict the failure of Finnish industrial and retailing companies but found it is slightly inferior to both discriminant and logit analysis. Shumway (2001) develops a discrete-time hazard model for bankruptcy that combines both accounting and market data. Due to

Table 1
Summary of literature on bankruptcy and financial distress prediction.

Methodologies	Examples	Advantages and limitations	References
Parametric Statistical method	Discriminant analysis Logistic regression Bayesian method	High transparency and strong interpretability	Altman (1968); Martin (1977); Ohlson (1980); Tam & Kiang (1992); Hu & Ansell (2007); Yang et al. (2011); Platt & Platt (2002); Tinoco & Wilson (2013).
Nonparametric Method	K-nearest neighbour DEA Malmquist DEA	Easy to apply and accommodate rules, but lack of explanation.	Paradi et al. (2004); Cielen et al. (2004); Li et al. (2014, 2017).
Machine learning, optimisation algorithm, hybrid and ensemble models	Neural networks, Support vector machines Random forests Genetic algorithm Ensemble learning Deep learning	Better accuracy, but lack of transparency and explanation	Tam & Kiang (1992); Lacher et al. (1995); Dimitras et al. (1999); Min & Lee (2005); Hua et al. (2007); Verikas et al. (2010); Sun et al. (2011); Geng et al. (2015); du Jardin (2017); Choi et al. (2018); Mai et al. (2019); Hosaka (2019).
Survival analysis	Cox proportional hazard model, Discrete hazard model	Can answer the question 'Not only whether or not default but also when'.	Lane et al. (1986); Luoma & Laitinen (1991); Shumway (2001); Gepp & Kumar (2008); Leonardis & Rocci (2008); Bonfim (2009); Kim et al. (2016); Kristanti & Herwany (2017).

the advantages in parameter computation and the nature of covariates observed periodically for firms, the discrete hazard model was followed by Chava & Jarrow (2004), Carling, Pan, Ariyan, Narayan, & Truini (2007), Leong, Nguyen, Meredith, et al. (2008) and Leonardis & Rocci (2008). In terms of the predictive accuracy, Gepp & Kumar (2008) found the Cox model to be comparable at equal misclassification costs but inferior in adapting to higher Type I error costs when compared with discriminant analysis and logistic regression. Kristanti & Herwany (2017) found satisfactory results by employing the survival analysis on distressed firms in Indonesian. We summarise the relevant work in Table 1.

Recurrent event data arise widely in medical studies, most notably in the study of epilepsy, asthma, heart attack and hospital stay (Clayton, 1994). One of the main characteristics of recurrent event data is within-subject correlation, where an event itself increases or decreases the likelihood of subsequent events (Box-Steffensmeier & Boef, 2006). Traditional statistical methods such as logistic regression and Cox proportional hazards regression either ignore the existence of recurrent events or do not account for within-subject correlation, resulting in an inappropriate estimation of standard errors and also a deviation from the original research question (Twisk, Smidt, & de Vente, 2005). Many alternatives have been proposed for the analysis of recurrent events which use full available information and within-subject correlations. Based on various definitions of risk sets, marginal intensity approaches allow all cases to be at risk for each repeated event (Wei, Lin, & Weissfeld, 1989), whereas conditional intensity models are estimated in the elapsed time or in the gap time, and cases are designated at risk for the kth repeated event only after experiencing the $(k-1)$th event (Andersen & Gill, 1982; Chang & Wang, 1999; Prentice, Williams, & Peterson, 1981). In the Andersen–Gill (AG) model (Andersen & Gill, 1982), it is assumed that the repeated events are ordered but have an equal risk of occurring. However, in the Prentice, Williams and Peterson (PWP) model (Prentice et al., 1981), an individual is assumed not at risk for a subsequent event until the previous event has occurred. Even though there is considerable literature on modelling of recurrent events using the PWP model in the fields of medical (*e.g.*, Ejoku, Odhiambo, & Chaba, 2020; Moulton & Dibley, 1997; Peña, Slate, & González, 2007; Pfennig et al., 2010), consumer behaviour (Bijwaard, Franses, & Paap, 2006) and product or equipment reliability (*e.g.*, Ascher, 1983; Jiang, Landers, & Reed Rhoads, 2006). Only a few studies can be found in corporate finance. For example, Parker, Peters, & Turetsky (2005) used the Cox and PWP models to analyse the impact of corporate governance factors on the repeated going concern assessments implemented by auditors on the distressed firms. Wang & Carson (2010) employed the PWP model to analyse the recurrent rating transitions of insurers. In the context of corporate loans, Godlewski (2015) used the PWP model to investigate the determinants of debt contract renegotiations between the bank and European companies. To the best of our knowledge, no one yet has used the PWP model for analysing the recurrent financial distress in corporate studies. Furthermore, we are focused on the recurrent events of distress, where the distress turnaround is also concerned. Distinct from the risk set defined in the previous studies mentioned above, we extend the methodology and study the dynamics of recurrent distress in Section 4.

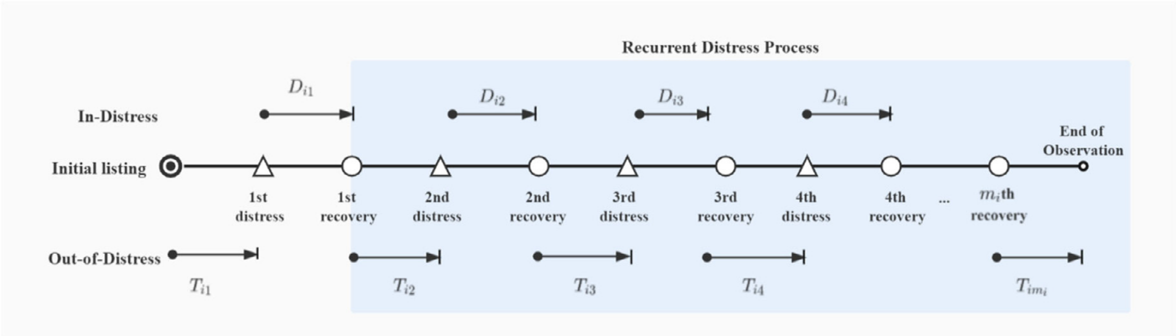


Fig. 1. Illustration of the recurrent distress process.

3. Exploratory analysis of distressed firms in China

In China, a publicly listed firm that has been identified as experiencing financial distress has an 'ST' prefix placed in front of its stock ticker (Wang & Deng, 2006), which is also called an ST label. Upon being specially treated, relevant restrictions and requirements will be applied to the firm on share trading, financing and reporting, and hence ST firms endeavour to recover their financial health. For the ST firms that successfully resolve their financial distress, their ST labels will be removed. On the other hand, a deteriorated financial condition can lead the ST firms to being delisted permanently from the stock exchange. However, recovery does not necessarily imply that the firms are able to remain financially healthy forever, because the firms may cycle back and forth between periods of in-distress and out-of-distress episodes.

A typical recurrent distress process is illustrated in Fig. 1. The jth duration of in-distress episode is denoted by D_j, whereas the duration of the out-of-distress episode is denoted by T_j. In fact, the investigation of T_1 falls into the classical study of the distressed or ST firms (Jiang & Jones, 2018; Lee, 2014). However, in this paper, we are more interested in the recurrent episodes $\{T_j\}_{j\geq 2}$, which characterise the recurring risk of distress for the recovered firms. It is also important to note that, the existing recurrent event studies focus on the gap time between successive events, which is essentially the sum of our two adjacent episodes $D_j + T_{j+1}$, without differentiating the composition of the two episodes.

We focus on the firms listed at the Shanghai and Shenzhen exchanges up to the date of June 2020. The following three criteria are applied to the preliminary sample: (1) The sample is restricted to firms issuing A-shares only; (2) firms are specially treated for financial reasons; (3) firms in the financial sectors are excluded.

Based on the criteria, there are 706 firms that have experienced ST by the end of June 2020. The first four cases of recurrent distress were observed as early as 2002, and 8 firms have experienced as many as four times of distress while still remained publicly listed at the exchange. As reported in Fig. 2, there are 200 firms that have experienced subsequent distress, representing 28.33% of the total distressed firms. These figures are expected to grow as time spans. Moreover, the average time from recovery to reoccurrence accelerates when firms repeatedly experience distress. This indeed is an illustration of the event dependence presented in the financial distress study, where the first distress itself increases the likelihood of subsequent distress. As there are an insufficient number of firms for the fourth or fifth event by now, we restrain our study to the fourth event from thereafter.

It might not be convincing by only analysing on the firms that have observed recurrent events, as most listed firms have no sufficient time to be fully observed. By taking account of the right censored cases, we can compare the survival rates of the events (first to fourth distress) by employing the Kaplan–Meier (KM) method (Kaplan & Meier, 1958), which gives a nonparametric estimate for the survival function, $S_j(t) = P\left(T_j > t\right), j \geq 1$, i.e., the probability of not having experienced the jth distress by time t.

The KM curves for the four events are shown in Fig. 3. As can be seen, the survival probability curve of the first distress is very different from that of the other three. It decreases much more slowly than others over time, whereas for the subsequent events, the survival rates drop rapidly in the first five years since the time of recovery from the previous distress. Therefore, it would be worth investigating the distress recurrence as a separate topic, unlike what have been done in previous literature.

In addition, the event dependence among the repeated events suggests that exploration of the first episodes of distress and recovery may provide valuable insights into the study of subsequent distress. The associations are explored in Fig. 4. Fig. 4(a) illustrates the association between the duration of the time to the first distress (T_1) and the subsequent distress episodes $\{T_j\}_{j\geq 2}$, where no pronounced relationship is present. More remarkably, a nonlinear correlation is shown in Fig. 4(b). The curve is clearly piecewise in the ranges of 1–2, 2–4 and 4–10 years from first distress to the first recovery. Inspired by this, we introduced two dummies to denote the first recovery within 2–4 years and more than 4 years, and we can observe two duration intervals from IPO to the first distress and from the first distress to the first recovery, respectively.

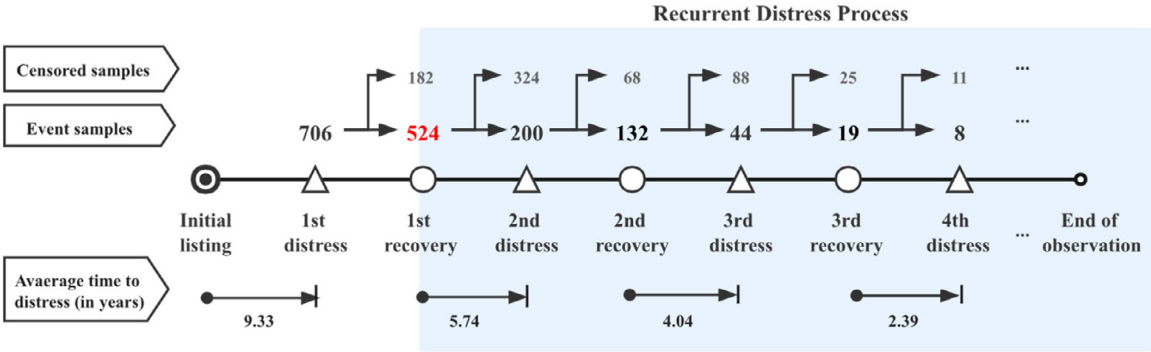


Fig. 2. : Summary of the samples across the recurrent process.

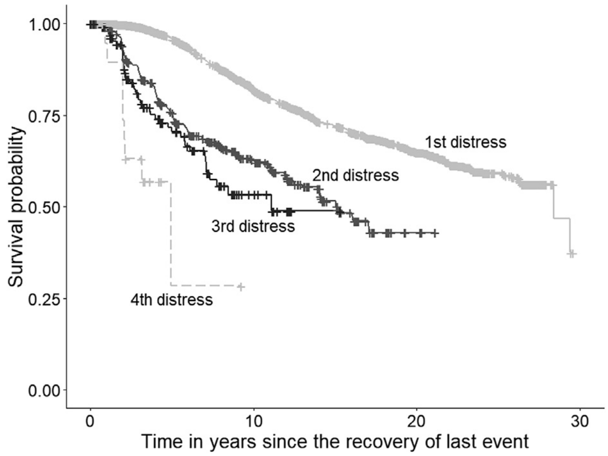
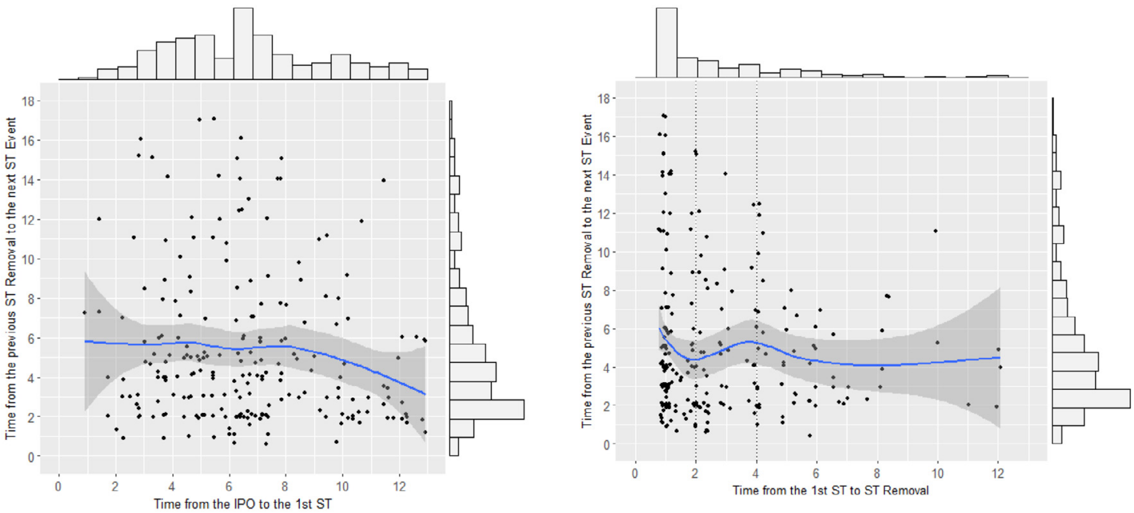


Fig. 3. Survival rates for recurrent distress.



(a) Correlation between the time to the first distress (T_1) and the subsequent distress episodes $\{T_j\}_{j\geq 2}$

(b) Correlation between the time from the first distress to recovery (D_1) and the subsequent distress episodes $\{T_j\}_{j\geq 2}$

Fig. 4. Exploration of the historical durations on the subsequent distress episodes.

4. Methodology

4.1. Data and model specifications

The sample in our study, therefore, only includes the distressed listed firms in China that have resolved their first financial distress in the period from January 1998 to June 2020, which in total are 524 firms. In the sample, 324 firms are censored cases that remained healthy until the end of observation. The remaining 200 firms contribute to 252 observations of the event, among which 200 observations have experienced the second distress, 44 have experienced the third distress, and 8 have experienced the fourth distress.

We assume that subject i is observed over the $[0, C_i]$ interval, where C_i denotes the censoring time and is observed falling into distress at times $R^*_{i,1}, R^*_{i,2}, \ldots, R^*_{i,m_i}$. Hence, we can re-express the durations of out-of-distress episodes in Fig. 1 as $T_{i,k} = \min(R^*_{i,k}, C_i) - (R^*_{i,k-1} + D_{i,k-1}), k \geq 2$, which is subject to right censoring. The corresponding observed event indicators are given by $Y_{ik}(t) = I(T_{i,k} < t)$. The distribution of the kth event time can be given by the hazard function:

$$h_{ik}(t) = \lim_{\Delta t \to 0} \frac{P\left(t < T_{i,k} \leq t + \Delta t | T_{i,k} > t\right)}{\Delta t} \quad (1)$$

which is the instantaneous risk for a firm to have the kth distress at time t, given that it has survived till time t.

In our research, events (financial distress) are in order, as a firm cannot experience a subsequent distress without the termination of the prior episode of distress. Furthermore, it is reasonable to assume that there is a dependence of events for a given firm, as evidenced by the accelerating time of event gaps in Fig. 2. Lastly, we are interested in how various factors are associated with the risk of recurrent distress. A model that can incorporate both event-specific and overall effects of each explanatory variable is preferred. As suggested by Cai & Schaubel (2003), the conditional gap time approach proposed by Prentice et al. (1981), i.e. the PWP model, is most suitable for an analysis of ordered events, where event dependence is accommodated by stratification and event-specific regression parameters are permitted. However, the traditional PWP gap time model implies a restricted risk set in the sense that firms which have not experienced the $(k-1)$th distress are not included in the analysis of the kth distress. However, in our case, the risk set should be modified so that the kth risk set only includes the firms that not only have experienced the $(k-1)$th distress but also have been successfully recovered. This is different from the risk-set construction used in the previous studies using the PWP model (see (Godlewski, 2015), for example). Therefore, by incorporating a modified risk-set indicator, we propose a stratified conditional intensity model in the following form:

$$h_{ik}(t|X) = \delta_{ik}(t)\, h_{0k}(t) \exp\left(\boldsymbol{\beta}^T \boldsymbol{X}_i + \boldsymbol{\alpha}^T \boldsymbol{X}_i^{Int} + \boldsymbol{\gamma}_k^T \boldsymbol{Z}_i\right) \quad (2)$$

where

$\delta_{ik}(t) = I(R_{i,k-1} + D_{i,k-1} + t \leq R_{i,k})$ is the modified risk-set indicator for the kth event, so that analysis of the kth distress is necessarily restricted to subjects who have resolved the $(k-1)$th distress;

$h_{0k}(\cdot)$ is the unspecified baseline hazard function for the kth recurrence of financial distress;

$\boldsymbol{X} = (X_1, X_2, \ldots, X_{p_1})$ represents the characteristic vector of the ith firm, with a common regression parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_{p_1})$ across all the events;

$\boldsymbol{X}^{Int} = (X_1X_2, X_1X_3, \ldots, X_{p_1}X_{(p_1-1)})$ represents the interaction terms among the variables in $\boldsymbol{X}$, with the associated regression parameter $\boldsymbol{\alpha}$;

$\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_{p_2})$ represents the characteristic vector that may have distinct effects over different distress episodes, and the associated episode-specific parameter is denoted as $\boldsymbol{\gamma_k}$.

The estimated coefficients show the effects of firm characteristics on the hazard rate, which can be estimated by the partial likelihood function (Chang & Wang, 1999). If the coefficient is statistically large than 0, the covariate is a risk factor. If it is statistically smaller than 0, the hazard rate decreases as the covariate increases. It is thus a protective factor. If the coefficient is insignificant, the covariate is an irrelevant factor.

4.2. Explanatory variables

As suggested by the exploratory results in Section 3, the durations of the first two episodes, measured by the length of time a firm took to get into the first financial distress (noted as T_1 in Table 2) and the time it took to enact a successful turnaround (noted as D_1 in Table 2), are included. The observed behaviours and activities of the firms associated with these two intervals are also considered. They are characterised by the event-based variables from the two following aspects. Firstly, the litigation risk of a firm is captured by the number of lawsuits that it holds regarding contract breaches, loan defaults and other cases, plus the number of times it has been listed as a dishonest debtor by the Supreme People's Court (SPC) and the number of executions enforced by the court. The latter two variables are useful in revealing the deterioration of a company's governance and creditworthiness. Secondly, inspired by the case studies in Kam, Citron, & Muradoglu (2010), restructuring efforts are commonly made by distressed firms to improve their financial prospects, but whether these actions have a subsequent influence on the recurrence of financial distress is yet to be discovered. For this reason, we consider a full range of restructuring actions that have been taken by a listed firm up to its first episode of distress and recovery, respectively. These include (1) asset restructure, such as asset acquisition, replacement and divestments (denoted by variables REST_ASRES_T1, REST_DTRES_T1 in Table 2); (2) debt restructure (denoted by variables REST_ASRES_D1 and REST_DTRES_D1 in Table 2); (3) share transferring or acquisition, which is measured by the equity transfer or changes of shareholders (i.e., variables REST_SHRES_T1 and REST_SHRES_D1 in Table 2); (4) ownership restructure defined by changes of the largest shareholder ownership (variables REST_OWNRES_T1 and REST_OWNRES_D1 in Table 2); (5) management restructure measured by

Table 2
Description of duration and event-based variables.

	Category	Abbreviation	Name and description
Duration variables			
		T_1	Duration from IPO to 1st distress, grouped into (1) <2 years, (2) 2–4 years and (3) 4+ years.
		D_1	Duration from 1st distress to 1st recovery, grouped into (1) <2 years, (2) 2–4 years and (3) 4+ years.
Event-based variables			
From IPO to 1st distress	Litigation risk	LR_DL_T1	Indicator of lawsuits as defendant
		LR_PL_T1	Indicator of lawsuits as plaintiff
		LR_DLBC_T1	Indicator of lawsuits regarding breaching of contract (as defendant)
		LR_PLBC_T1	Indicator of lawsuits regarding breaching of contract (as plaintiff)
		LR_DLCD_T1	Indicator of lawsuits regarding corporate lending (as defendant)
		LR_PLCD_T1	Indicator of lawsuits regarding corporate lending (as plaintiff)
		LR_DLO_T1	If the average number of other lawsuits (as defendant) per year is greater than 0.5, it is 0; otherwise, 0.
		LR_PLO_T1	If the average number of other lawsuits (as plaintiff) per year is greater than 0.5, it is 1; otherwise, 0.
		LR_DIS_T1	Indicator of dishonest debtors published by the Supreme People's Court (SPC)
		LR_EXE_T1	Indicator of executions enforced by the court
	Restructure actions	REST_ASRES_T1	Indicator of asset restructuring
		REST_ASREP_T1	Indicator of asset replacement
		REST_EQTS_T1	Indicator of equity transfer
		REST_SHCH_T1	Indicator of shareholder changes
		REST_DTRES_T1	Indicator of debt restructuring
		REST_OWNRES_T1	Indicator of the largest shareholder ownership changed
		REST_MANRES_T1	Indicator of the chairman, CEO or legal representative changed
		REST_OPRES_T1	Indicator of changes in operation information
From 1st distress to recovery	Litigation risk	LR_DL_D1	Indicator of lawsuits as defendant
		LR_PL_D1	Indicator of lawsuits as plaintiff
		LR_DLBC_D1	Indicator of lawsuits regarding breaching of contract (as defendant)
		LR_PLBC_D1	Indicator of lawsuits regarding breaching of contract (as plaintiff)
		LR_DLCD_D1	Indicator of lawsuits regarding corporate lending (as defendant)
		LR_PLCD_D1	Indicator of lawsuits regarding corporate lending (as plaintiff)
		LR_DLO_D1	If the average number of other lawsuits (as defendant) per year is greater than 0.5, it is 1; otherwise, 0
		LR_PLO_D1	If the average number of other lawsuits (as plaintiff) per year is greater than 0.5, it is 1; otherwise, 0
		LR_DIS_D1	Indicator of dishonest debtors published by the Supreme People's Court (SPC)
		LR_EXE_D1	Indicator of executions enforced by the court
	Restructure actions	REST_ASRES_D1	Indicator of asset restructuring
		REST_ASREP_D1	Indicator of asset replacement
		REST_EQTS_D1	Indicator of equity transfer
		REST_SHCH_D1	Indicator of shareholder changes
		REST_DTRES_D1	Indicator of debt restructuring
		REST_OWNRES_D1	Indicator of the largest shareholder ownership changed
		REST_MANRES_D1	Indicator of the chairman, CEO or legal representative changed
		REST_OPRES_D1	Indicator of changes in operation information

changes of chairman, CEO or legal representative (i.e., variables REST_MANRES_T1 and REST_MANRES_D1 in Table 2); (6) operation restructure indicated by changes regarding company name, address and operating field of industry (denoted by variables REST_OPRES_T1 and REST_OPRES_D1 in Table 2).

To statistically investigate the effects of the event-based variables on the survival rate, a log-rank test (Mantel, 1966) is employed. Table 3 reports the test statistics and p-values for each event-based variable in the two stages. Initial observation shows that the firms with distress recurrence seem to have significantly different characteristics regarding those lawsuits in which they

Table 3
Log-rank tests for event-based variables.

	Variables	Statistics	p-value	Variables	Statistics	p-value
	LR_DL_T1	1	0.3	LR_DL_D1	3.6	0.06
	LR_PL_T1	9.5	0.002	LR_PL_D1	10.7	0.001
	LR_DLBC_T1	4.5	0.03	LR_DLBC_D1	0.3	0.6
	LR_PLBC_T1	4.3	0.04	LR_PLBC_D1	2.6	0.1
Litigation risk	LR_DLCD_T1	0.3	0.6	LR_DLCD_D1	0.1	0.8
	LR_PLCD_T1	5	0.02	LR_PLCD_D1	0.2	0.6
	LR_DLO_T1	10.4	0.001	LR_DLO_D1	0.3	0.6
	LR_PLO_T1	15	<0.0001	LR_PLO_D1	16.5	<0.0001
	LR_DIS_T1	0.9	0.4	LR_DIS_D1	1.7	0.2
	LR_EXE_T1	6.1	0.01	LR_EXE_D1	8.2	0.004
	REST_ASRES_T1	0.6	0.4	REST_ASRES_D1	1.6	0.2
	REST_ASREP_T1	0.3	0.6	REST_ASREP_D1	0.3	0.6
	REST_EQTS_T1	4.9	0.03	REST_EQTS_D1	0.7	0.4
Restructure actions	REST_DTRES_T1	9.2	0.002	REST_DTRES_D1	0.2	0.7
	REST_SHCH_ _T1	0.2	0.7	REST_SHCH _D1	0.5	0.5
	REST_OWNRES_T1	2.3	0.1	REST_OWNRES_D1	1.7	0.2
	REST_MANRES_T1	0.6	0.4	REST_MANRES_D1	0.1	0.7
	REST_OPRES_T1	1.7	0.2	REST_OPRES_D1	0.3	0.6

were involved as the plaintiff, and the event of execution enforcement. In terms of restructure efforts, debt restructuring actions prior to the first distress are likely to affect the risk of repeated distress.

The variables that have been widely adopted in the conventional studies of first distress are also considered here. There are three categories, namely institutional variables, accounting-based variables (Altman, 1968) and market-based variables (Merton, 1974), as summarised in Table 4. In particular, we incorporate ownership concentration, government ownership and board composition as relevant factors in the institutional background. All the institutional, accounting and market information was collected from the last annual reports prior to the recovery date.

As mentioned, the distress recurrence spans over a period of 17 years since 2004, so that macroeconomic conditions at the time of recovery are also taken into consideration, as shown in Table 5. Table 6 reports the 81 variables and their descriptive statistics. The data are sourced from the China Stock Market Accounting Research (CSMAR), WIND and Big Business Data (BBD) databases.

4.3. Performance measures for survival models

When validating a prediction model, the predictive performance of the model is commonly addressed by the overall performance based on predictive accuracy, i.e. the 'distance' between the observed and predicted outcomes, and the discrimination power based on concordance, which is the ability to distinguish between low and high-risk samples. Several performance measures have been suggested for use with survival prediction models (see (Choodari-Oskooei, Royston, & Parmar, 2012; Schemper & Stare, 1996)). Following the study of Rahman, Ambler, Choodari-Oskooei, & Omar (2017), we have selected the following four measures below to assess the model performance.

(1) **Predictive accuracy measure: Brier score**

Brier score (Brier, 1950) is expressed as the squared error between the observed status and the predicted probability, which does not fully capture the time series of the survival curves. Hence, a time-dependent extension of the Brier score is considered reasonable in this case (Graf & Schumacher, 1995). Following Graf, Schmoor, Sauerbrei, & Schumacher (1999), we use the modified Brier score to accommodate for the recurrent events, which is given by:

$$BS\left(\hat{S},t\right) = \sum_k BS_k\left(\hat{S},t\right) = \frac{1}{n}\sum_k \sum_{i=1}^{n} I(m_i \le k)\left(Y_{ik}(t) - \hat{S}_{ik}(t|X_i)\right)^2 \times W_i(\widehat{G}_k, t) \quad (3)$$

where the weight function W_i depends on $\widehat{G}_k$, the estimation for the distribution of censoring times for the kth event. By assuming the weighting scheme to be independent of the survival model, the Kaplan–Meier estimator is used for the censoring distribution. The integrated Brier score can thus be defined as $IBS\left(\hat{S}\right) = \sum_k IBS_k\left(\hat{S}\right) = \sum_k \int BS_k\left(\hat{S},t\right)$. The lower the Brier score, the higher is the predictive accuracy of the model.

(2) **Concordance Measures**

Concordance measures usually take values between 0.5 and 1, where a value of 0.5 indicates no discrimination and a value of 1 indicates perfect discrimination. Three concordance measures are selected: the Harrell's Concordance (Harrell Jr, Lee, & Mark, 1996), the GH's Concordance (Gönen & Heller, 2005) and the Uno's Concordance (Uno, Cai, Pencina, D'Agostino, & Wei, 2011).

Harrell's Concordance ($\mathbf{C_H}$). The concordance probability is the probability that of a randomly selected pair of subjects, and the subject with a shorter survival time has a higher predicted risk. The Harrell's estimator considers all usable pairs of subjects for which shorter time corresponds to an event. Therefore, the estimated C_H can be interpreted as the proportion of observed pairs for which the subject of shorter survival time has a higher predicted risk.

GH's Concordance (C_{GH}). Gönen & Heller (2005) proposed an alternative estimator based on a reversed definition of concordance, which is the probability that of

Table 4
Description of financial ratios.

Category	Abbreviation	Name and description
Institutional variables		
Basic information	BAS_D_MANUF	Indicator of manufacturing firm
	BAS_D_IPOYEAR	IPO year before 1998 = 1, otherwise 0
	BAS_AGE	Firm age at time of recovery
Cooperate governance	COOP_TOP1SHR	Largest shareholder ownership
	COOP_TOP10SHR	Top 10 leading shareholding ratio
	COOP_PIS	Proportion of institutional shareholding
	COOP_STATESHR	The percentage of shares owned by the government
	COOP_D_MED	Indicator of dual appointment of board chairman and firm CEOs
	COOP_PID	Proportion of independent directors to board
	COOP_RTD	The total remuneration of the top three directors, in units of ten million
Accounting-based variables		
Profitability	PROF_ROE	Return on equity
	PROF_ROA	Return on assets
	PROF_OISR	Operating income to sales ratio
	PROF_RPC	Profit-to-cost ratio
	PROF_EPS	Earnings per share
	PROF_RE/TA	Retained earnings total assets ratio
Operational efficiency	OPER_TTC	Turnover of total capital
	OPER_TCA	Turnover of current assets
	OPER_R_CUR	Current ratio
	OPER_R_QUI	Quick ratio
	OPER_TAC	Average accounts receivable turnover ratio
Liabilities	LIAB_R_AL	Asset–liability ratio
	LIAB_R_CD	Current debt ratio
	LIAB_EBIT_IC	Interest coverage
	LIAB_R_CASH	Cash ratio
	LIAB_LTA	Log of total asset
Growth potential	GROW_GOP	Year-on-year growth in operating profits
	GROW_GTA	Total assets year-on-year growth rate
	GROW_GOI	Year-on-year growth in operating income
	GROW_P_E	P/E ratio (divided by 100)
	GROW_PCF	Price cash flow ratio (divided by 100)
Market-based variables		
	MARK_LMC	Log (market capitalization of the firm/total market capitalization)
	MARK_EX_RET	Excess return of equity
	MARK_TRDTURNR	Yearly turnover of total tradable shares, i.e., trading volume/number of tradable shares outstanding
	MARK_DEBT_M	Total debt/total market value of equity

Table 5
Description of macroeconomic variables.

Abbreviation	Name and description
Macro_GGDP	Yearly growth of GDP per head
Macro_CPI	CPI
Macro_RUUR	Registered urban unemployment rate (RUUR) (%)
Macro_RPI	Retail price index (RPI)
Macro_IR_LOAN	Weighted average interest rate on underlying loans
Macro_AY_CSI300	Annual yield of CSI300

a randomly selected pair of subjects. Higher predicted risk indicates a shorter survival time. To avoid the bias caused by censoring, their estimator is a function of parameters and the predictor distributions, which relies on the specification of the prediction model.

Uno's Concordance (C_U). In the presence of censoring, C_H is biased as they ignore the pairs where the shorter observed time is censored. Because of this deficiency, Uno et al. (2011) proposed a modified estimator C_U, where inverse probability weighting is used to compensate for censoring.

5. Results and discussion

5.1. Empirical results

To evaluate both the model's accuracy and prediction performance, we divide the entire sample into two independent sets: a training set of 465 observations that recovered prior to 2012 and a holdout test set of 210 observations that recovered between 2013 and 2020. It is thus an out-of-time validation. Four models are established to examine the respective effects of duration

Table 6
Summary statistics of variables.

Variables	Mean	Std	Min	Median	Max	Variables	Mean	Std	Min	Median	Max
BAS_D_MANUF	0.526	0.5	0	1	1	D_T1_2To4	0.104	0.306	0	0	1
BAS_D_IPOYEAR	0.576	0.495	0	1	1	D_T1_MT4	0.885	0.32	0	1	1
BAS_AGE	15.21	5.723	1	15	35	D_D1_2To4	0.212	0.409	0	0	1
COOP_TOP1SHR	35.13	16.64	2.197	29.96	89.41	D_D1_MT4	0.234	0.424	0	0	1
COOP_R_TOP10LS	54.81	16.17	4.404	55.6	97.39	LR_DL_T1	0.158	0.365	0	0	1
COOP_PIS	26.52	20.46	0	26.35	91.81	LR_PL_T1	0.177	0.382	0	0	1
COOP_STATE_SHR	16.46	23.68	0	0	97.12	LR_DLBC_T1	0.139	0.347	0	0	1
COOP_D_MED	0.199	0.4	0	0	1	LR_PLBC_T1	0.167	0.374	0	0	1
COOP_PID	32.42	12.09	0	33.33	66.67	LR_DLCT1	0.004	0.061	0	0	1
COOP_RTD	0.896	1.004	0.012	0.625	12.77	LR_PLCT1	0.019	0.135	0	0	1
PROF_ROE	0.101	0.287	−1.688	0.048	1.032	LR_DLO_T1	0.022	0.148	0	0	1
PROF_ROA	0.056	0.14	−1.052	0.025	0.647	LR_PLO_T1	0.046	0.211	0	0	1
PROF_OISR	−0.084	2.386	−52.95	0.041	1.406	LR_DIS_T1	0.011	0.105	0	0	1
PROF_RPC	0.125	0.31	−2.575	0.043	1.404	LR_EXE_T1	0.279	0.449	0	0	1
PROF_EPS	0.184	0.381	−1.78	0.09	2.051	REST_ASRES_T1	0.905	0.293	0	1	1
PROF_RE_TA	−1.166	8.45	−103.7	−0.106	0.45	REST_ASREP_T1	0.273	0.446	0	0	1
OPER_TTC	0.991	1.133	−1.685	0.719	7.226	REST_EQTS_T1	0.758	0.428	0	1	1
OPER_TCA	1.361	1.108	0.001	1.049	5.69	REST_SHCH_T1	0.171	0.377	0	0	1
OPER_R_CUR	1.363	1.382	0.022	1.039	11.83	REST_DTRES_T1	0.061	0.24	0	0	1
OPER_R_QUI	0.977	1.24	0.011	0.671	10.36	REST_OWNRES_T1	0.439	0.497	0	0	1
OPER_TAC	24.61	45.81	0.062	7.371	207.88	REST_MANRES_T1	0.939	0.24	0	1	1
LIAB_R_AL	0.614	0.52	0.106	0.612	9.664	REST_OPRES_T1	0.195	0.397	0	0	1
LIAB_R_CD	0.872	0.157	0.221	0.937	1	LR_DL_D1	0.151	0.358	0	0	1
LIAB_EBIT_IC	11.94	35.79	−60.50	3.382	319.3	LR_PL_D1	0.21	0.408	0	0	1
LIAB_R_CASH	0.493	0.837	0	0.247	7.047	LR_DLBC_D1	0.141	0.349	0	0	1
LIAB_LTA	21.01	1.415	15.60	20.91	25.81	LR_PLBC_D1	0.206	0.405	0	0	1
GROW_GOP	1.889	5.587	−40.47	1.114	22.55	LR_DLCD1	0.019	0.135	0	0	1
GROW_LTA	0.144	0.615	−0.767	0.011	3.456	LR_PLCD1	0.046	0.211	0	0	1
GROW_GOI	0.886	2.587	−0.958	0.142	13.12	LR_DLO_D1	0.091	0.288	0	0	1
GROW_P_E	0.275	3.59	−17.76	−0.058	22.35	LR_PLO_D1	0.134	0.341	0	0	1
GROW_PCF	−1.113	12.94	−94.26	−0.08	47.34	LR_DIS_D1	0.026	0.159	0	0	1
MARK_LMC	−9.151	0.83	−12.93	−9.204	−5.048	LR_EXE_D1	0.301	0.459	0	0	1
MARK_EX_RET	0.388	1.455	−0.815	0.042	14.90	REST_ASRES_D1	0.896	0.306	0	1	1
MARK_TRDTURNR	3.816	3.109	0	3.093	32.97	REST_ASREP_D1	0.234	0.424	0	0	1
MARK_DEBT_M	0.986	1.3	−0.001	0.479	10.51	REST_EQTS_D1	0.61	0.488	0	1	1
MACRO_GGDP	109.2	2.044	106	109.1	114.2	REST_SHCH_D1	0.078	0.269	0	0	1
MACRO_CPI	102.5	1.903	98.6	102.1	105.9	REST_DTRES_D1	0.19	0.392	0	0	1
MACRO_RUUR	3.992	0.298	3.1	4.1	4.3	REST_OWNRES_D1	0.353	0.478	0	0	1
MACRO_RPI	101.6	2.155	97	101.4	105.9	REST_MANRES_D1	0.736	0.441	0	1	1
MACRO_IR_LOAN	5.394	0.56	4.35	5.58	6.86	REST_OPRES_D1	0.199	0.4	0	0	1
MACRO_AY_CSI300	0.177	0.547	−0.66	0.046	1.575						

and event variables on the model's performance. As a benchmark model, Model 0 contains financial variables and macroeconomic variables only. Model 1 combines the duration variable and Model 2 further includes event-based variables. Interaction terms are also explored in Models 1 and 2 to investigate the possible correlated effects among the variables. Based on Model 2, episode-specific terms are investigated in Model 3. With a limited number of observations but a large number of candidate variables, we carry out the feature selection process as follows. First, we use the backward and forward stepwise methods to select the explanatory variables in $\boldsymbol{X}$. Once the effective variables are chosen, their interaction terms are added. Stepwise selection is applied again to seeking for significant interaction terms. The exploration of episode-specific terms is also constrained within the effective variables in $\boldsymbol{X}$ to avoid the curse of dimensionality. The regression results and the goodness-of-fit of the four models are presented in Table 7.

The macroeconomic variables are lagged by 1 year. We find similar results for models with longer lag periods, except that the effects and statistical significance tend to be weaker.

5.2. Analysis and discussion

In Model 0, we note that listed manufacturing firms are subject to a significantly higher recurrent risk. Four variables of corporate governance are shown to be significant. As expected, the proportion of institutional shareholders is negatively related to the hazard rate. This finding is similar to that in Li et al. (2021), whose result shows that when the institutional investor has a stake in a listed company, the chances of distress are lower. The dual position of CEO and board chairman (COOP_D_MED) characterises the managerial ownership of a firm, which was found ineffective to distress prediction in China or other countries (Deng and Wang, 2006; Ma and Tian). However, in our study, this indicator significantly reduces the probability of future distress recurrence, maybe due to the improved decision-making speed. Measured by the largest shareholder ownership (COOP_TOP1SHR), a higher ownership concentration increases the likelihood of survival, which is in line with the findings in Kim et al. (2016). A high proportion of independent directors to the board has a positive impact on survival. This is consistent with the

Table 7
Results of the four models.

Independent variables		Model 0	Model 1	Model 2	Model 3
Duration	T1_MT4		−0.0481	−0.1768	−0.2206
	D1_2To4		−5.5287*	−2.0896	−2.4896
	D1_MT4		0.6231**	0.7027**	0.7502**
Event-based variables during T1	LR_EXE_T1			0.4563	0.4842
	REST_EQTS_ T1			0.6897**	0.6642***
	REST_DTRES_ T1			−19.2572*	−18.9569*
	REST_SHCH_ T1			−1.3797**	−1.4472**
	REST_OPRES_ T1			−0.5776	−0.5004
Event-based variables during D1	REST_MANRES_D1			4.8726*	4.9673*
	REST_ASRES_ D1			0.8584**	0.8356**
Basic information	BAS_D_MANUF	0.4617**	0.5372***	0.7025***	0.7443***
	BAS_D_IPOYEAR	−0.2915	−0.3513	−0.4948	−0.532
Corporate governance	COOP_TOP1SHR	−0.0183*	−0.0166*	−0.0185*	−0.0192*
	COOP_PIS	−0.01*	−0.0112*	−0.0114*	−0.0112*
	COOP_PID	−0.0141*	−0.0215*	−0.0338***	−0.0335***
	COOP_D_MED	−1.1409**	−1.2668**	−1.2584**	−1.2985**
Profitability	PROF_RPC		0.5455	0.6191	−0.1467
Operational efficiency	OPER_R_QUI	0.4092*	0.4247*	0.4986*	0.4876**
Liabilities	LIAB_LTA	0.4709**	0.3953**	0.5942***	0.5762***
	LIAB_R_CASH	−0.8135**	−0.8311**	−0.8218**	−0.8069**
Growth potential	GROW_GTA			−0.3042*	−0.3488*
Market-based	MARK_EX_RET	−0.2703*	−0.3202**	−0.2918*	−0.2925*
	MARK_LMC	−0.6833***	−0.5392**	−1.032***	−1.0376***
	MARK_DEBT_M	−0.0076	0.004	0.0307	0.0316
	MARK_TRDTURNR	0.0344	0.0574*	0.0481	0.0457
Macroeconomic factors	GGDP_lag1	−0.1493*	−0.1174*	−0.1037*	−0.0965*
	IR_LOAN_lag1		−0.3108	−0.3397	−0.3495
Interaction terms	BAS_D_IPOYEAR:MARK_DEBT_M	−0.3069*	−0.2562*	−0.285*	−0.2858*
	BAS_D_IPOYEAR:COOP_TOP1SHR			0.0231*	0.0236*
	MARK_TRDTURNR:COOP_D_MED	0.1695*	0.1912*	0.1969*	0.2029*
	T1_MT4:PROF_RPC		−0.6317*	−0.8073*	−0.2209*
	D1_2To4:PROF_RPC		−2.4228*	−2.963*	−3.2344*
	D1_2To4:IR_LOAN_lag1		1.0829*	0.9168*	0.9877*
	D1_2To4:REST_ASRES_ D1			−2.5247**	−2.4719*
	MARK_LMC:REST_DTRES_ T1			−2.1296	−2.0931*
	MARK_LMC:REST_MANRES_ D1			0.5356*	0.5493*
	BAS_D_MANUF:REST_OPRES_ T1			1.5968*	1.5453*
	BAS_D_MANUF:LR_EXE_ T1			−0.873*	−0.8886*
Episode-specific terms	BAS_D_MANUF for the 4th distress				−2.0528*
	PROF_RPC for the 3rd distress				1.1326*
Model goodness of fit					
	Integrated Brier score	0.106	0.104	0.097	0.095
	Harrell's concordance	0.6302	0.6446	0.6717	0.6655
	GH's concordance	0.6816	0.6977	0.7389	0.7454
	Uno's concordance	0.7063	0.7158	0.7409	0.7344

$^{***}p$-value < 0.001.
$^{**}p$-value < 0.01.
$^{*}p$-value < 0.05.

study of Bai, Liu, Lu, Song, & Zhang (2004) on the performance of Chinese firms, where the outside shareholders increase the firm value.

We observe that the firms with a high level of total assets (LIAB_LTA) are subject to higher risks, which is similar to the result obtained in Kim et al. (2016). The cash ratio, on the other side, reduces the risk of further distress. However, we observe that the firms with higher risks on distress recurrence seem to have a higher quick ratio (OPER_R_QUI). This may appear to be contradictory to the intuition, but a similar discovery can be found in Kim et al. (2016). As the current rules applied to the recovery from distress are heavily dependent on the firms' accounting conditions, this leads us to the question whether those repeatedly distressed firms have truly improved their financial situation by the time of recovery, or their accounting is 'decorated'. This is an issue that can be further investigated in future.

The market measures of firm size (MARK_LMC) and excess return (MARK_EX_RET) have a positive impact on reducing the risk of recurrent distress, which is consistent with the results reported in Kim et al. (2016)

Table 8
Coefficients of the profit-to-cost ratios for different scenarios in Model 1.

	Recovery in 2–4 years	Recovery < 2 years or > 4 years
First distress beyond 4 years after the IPO	**−2.5090**	−0.0862
First distress within 4 years after the IPO	−1.8773	0.5455

that these two market-driven variables also have positive effects on recovery. Similar results can also be found in Shumway (2001), Chava & Jarrow (2004), Tinoco & Wilson (2013) suggesting that increasing excess return and increasing relative size indicate a smaller chance of financial distress or bankruptcy. Another interesting finding is that the negative effect of the interaction term of the trading turnover and the duality of CEO-chairman (MARK_TRDTURNR: COOP_D_MED) is found. The result suggests that failing firms with the duality of CEO-chairman attract more investor trading demands.

One macroeconomic variable is also identified as effective in Model 0. Similar to Bonfim (2009), the yearly growth rate of GDP per capita at the time of recovery has a positive effect on the survival time of recovered firms.

By introducing the two duration variables and their interaction terms with financial and macroeconomic variables, Model 1 shows a significant improvement in model accuracy and concordance. Except for the profit-to-cost ratio (OPER_RPC), the coefficients of all predictors in Model 0 are of their consistent signs and statistically significant in Model 1. Interestingly, the variable of OPER_RPC is not significant at 5% level in Model 0 and Model 1, but its interaction terms with two duration variables (i.e., OPER_RPC: T1_MT4 (4+ years) and OPER_RPC: D1_2To4 (2–4 years)) are pronounced in Model 1. To better understand the duration dependence of the profit-to-cost ratio in Model 1, we recompute the coefficients of the variable in four scenarios, as shown in Table 8.

Notably, for firms having encountered their first period of distress beyond four years after the IPO and having taken 2–4 years for recovery, OPER_RPC is identified as an effective predictor, such that a 10% increase in the profit-to-cost ratio lowers the hazard risk by 28.52% (i.e., $1 - \exp(-2.5090 \times 10\%)$). In contrast, for firms with their first period of distress within four years from IPO and either less than 2 or longer than 4 years for recovery, the coefficient of OPER_RPC changes sign and a 10% increase in profit-to-cost ratio now rises the hazard risk by 5.60%. It is not difficult to see that the duration of recovery presents an interesting pattern in Table 8. Firms who recovered within 2 years are no different from the ones that need much longer time (>4 years) to restore their financial health. This observation leads us to conjecture that distressed firms with a quick turnaround in their financial situation may not be as promising as what is shown in their financial reports. A high interest rate of loans (IR_LOAN) implies a high financing cost for firms and hence increases the risk of further distress. However, this is only true for those firms which have taken 2–4 years for recovery, as suggested by the interactive effect between IR_LOAN and D1 _2To4.

By further combining the event-based variables in Model 2, the model goodness-of-fit improves further. A number of variables reflecting restructuring actions are presented as significantly effective. In particular, during the first episode of financial distress, actions in debt restructuring, shareholder changes and the operation information changes increase the hazard rate effectively. However, effective actions, such as the changes of the chairman, CEO or legal representative, and asset replacement after the first distress are subject to a lower risk of subsequent distress. However, firms having taken 2–4 years for recovery and asset replacement within the turnaround process have a lower risk on the recurrence of distress. More interaction terms are observed in Model 2. The market measure of firm size (MARK_LMC) is a protective factor, but its effectiveness can be influenced by the event of debt restructuring in the first distress episode and the event of chairman, CEO or legal representative changing within the recovery period. We also have observed that those listed manufacturing firms with changes in operation information are subject to a significantly higher recurrent risk.

While Model 2 gives a comprehensive exploration of the effective covariates across the different distress episodes, episode-specific effects may also be present in our data. As shown in Model 3, the manufacturing firms are subject to a significantly lower risk for the 4th distress. Also, for the 3rd distress, the negative impacts of profit-to-cost ratio as shown in Table 7 have been significantly reduced.

5.3. Out-of-sample assessment of predictive performance

Prediction error curves for each of the four fitted models are computed based on Eq. (3) and shown in Fig. 5. Overall, the errors of four models are gradually increasing as time progresses. For the whole test sample, Models 2 and 3 outperform the other two models within the 4 to 7 years prediction window, which implies that the litigation risk and restructure actions are effective indicators for subsequent distress recurrence. A more detailed comparison is shown in Fig. 5(b) and (c), where separate predictions are carried out for the second and the third distress among the test samples. Again, Fig. 5(b) shows that the inclusion of event-based variables increases accuracy prediction of the first recurrence of financial distress within the 4- to 7-year prediction window. With a much shorter observation time for the third distress, the predictive performance of the models varies over time. However, Model 2 still provides a higher predictive accuracy within 2.5 to 4 years since the recovery from the second distress.

Further comparison of the predictive accuracy and the discriminant power between four models is shown in Table 9. The best performance for each scenario is highlighted in bold. Overall, Model 2 with the duration and event-based variables are superior in Brier score and most

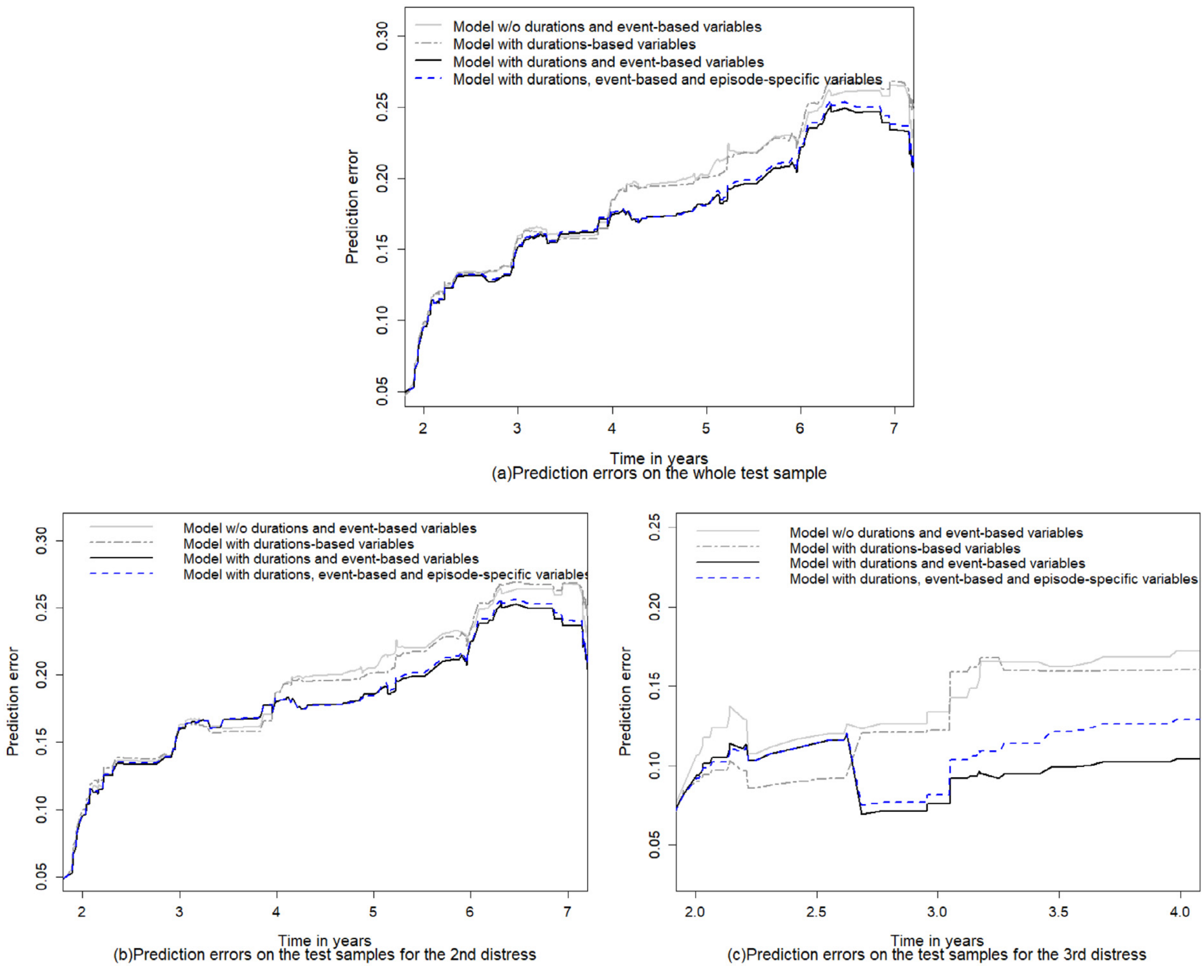


Fig. 5. Comparison of the out-of-time predictions on different panels.

concordance statistics. When breaking down to the prediction for the third distress only, Model 3 seems to have higher discriminant power than Model 2 does, showing the effectiveness of the episode-specific terms explored. Nonetheless, the durations and observable behaviours of the firms during the first episode of distress and recovery have demonstrated to be strong predictors for the recurrence risk.

6. Conclusions

In the same way that human beings can become sick and subsequently recover, corporations also experience periods of ‘sickness’, which do not necessarily automatically lead to immediate failure. However, financial distress studies in corporate credit research have so far only focused on the first event of corporate sickness, which we have referred to as ‘financial distress’ in this paper. This study provides the first evidence for the prediction of financial distress recurrence. Using a survival analysis framework, we empirically investigated the recurrent distress events for Chinese listed firms from 2001 to 2020. The key findings of the paper are as follows.

First, the recurrent risk for a recovered firm is duration dependent. In particular, those firms who either spent more than 4 years from IPO to their first occurrence of distress or else took between 2–4 years to recover their initial financial condition are subject to a lower risk of recurrence. Moreover, event-based variables, such as debt restructuring, shareholder changes, asset replacement and the operation information changes, are all found to substantially enhance the predictive power of the model, when the accounting-driven, market-based and macroeconomic variables are already included.

Second, our study stands out from previous research on financial distress because of its thorough investigation into the interactive effects on the different categories of variables. Our empirical findings have suggested that the effectiveness of several accounting-driven and market-driven variables is highly influenced by any historical litigation and restructuring events that may have been experienced by the distressed firms. This insight provides a new perspective for future researchers to investigate the field of financial distress.

Finally, this paper has evaluated the stratified hazard model both in terms of model adequacy and out-of-time

Table 9
Comparison of out-of-time predictions.

Measure	Models	Target (2nd distress)	Target (3rd distress)	Target (All events)
Integrated Brier score	Model 0	0.145	0.087	0.144
	Model 1	0.144	0.080	0.143
	Model 2	**0.137**	**0.066**	**0.135**
	Model 3	0.138	0.070	0.136
Harrell's concordance	Model 0	0.599	0.585	0.599
	Model 1	0.595	0.564	0.592
	Model 2	**0.622**	0.622	**0.622**
	Model 3	0.621	**0.624**	0.621
GH's concordance	Model 0	0.706	0.716	0.705
	Model 1	0.712	0.688	0.709
	Model 2	0.763	0.750	0.760
	Model 3	**0.765**	**0.753**	**0.763**
Uno's concordance	Model 0	0.619	0.705	0.626
	Model 1	0.626	0.668	0.628
	Model 2	**0.663**	0.739	**0.667**
	Model 3	0.662	**0.752**	0.666

prediction performance. This superior accuracy suggests that investors may utilise this model to improve their investment strategy for distressed firms. In addition, the empirical evidence revealed in the paper may also serve as a reference for market regulators to better monitor distressed firms and thus refine the delisting rules. This paper has some limitations. The risk of distress recurrence is not only related to the financial and macroeconomic conditions at the time of distress turnaround, but it is also closely related to changing operating conditions and the wider economic environment. Therefore, a time-dependent model may be worth considering for future studies. Additionally, we only took the sample from among cases of the first three distress for analysis though the framework is applicable to all recurrent risks. When the number of the fourth or fifth distress increases in future, it would be interesting to study the gaps between repeated distress.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank for the support from the funding of National Natural Science Foundation in China [No. 72010107001] and the Fundamental Research Funds for the Central Universities, China [JBK2103006] [JBK1806002] [JBK2103013]. This work was also supported by the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics, China. We thank Xiaomeng Zhai and Wanlin Qie for their preliminary analysis with lead to the foundation of this research.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609.
Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, *18*(3), 505–529.
Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, *10*, 1100–1120.
Ascher, H. (1983). Regression analysis of repairable systems reliability. In *Electronic systems effectiveness and life cycle costing* (pp. 119–133). Berlin, Heidelberg: Springer.
Back, B., Laitinen, T., & Sere, K. (1996). Neural networks and genetic algorithms for bankruptcy predictions. *Expert Systems with Applications*, *11*(4), 407–413.
Bai, C., Liu, Q., Lu, J., Song, F. M., & Zhang, J. (2004). Corporate governance and market valuation in China. *Journal of Comparative Economics*, *32*(4), 599–616.
Balcaen, S., & Ooghe, H. (2006). 35 Years of studies on business failure: an overview of the classical statistical methodologies and their related problems. *The British Accounting Review*, *38*(1), 63–93.
Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *4*, 71–111.
Bharath, S. T., & Shumway, T. (2008). Forecasting default with the merton distance to default model. *Review of Financial Studies*, *21*(3), 1339–1369.
Bijwaard, G. E., Franses, P. H., & Paap, R. (2006). Modeling purchases as repeated events. *Journal of Business & Economic Statistics*, *24*, 487–502.
Bonfim, D. (2009). Credit risk drivers: evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking & Finance*, *33*(2), 281–299.
Box-Steffensmeier, J. M., & Boef, S. D. (2006). Repeated events survival models: the conditional frailty model. *Statistics in Medicine*, *25*(20), 3518–3533.
Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.
Cai, J., & Schaubel, D. E. (2003). Analysis of recurrent event data. *Handbook of Statistics*, *23*, 603–623.
Carling, T., Pan, D., Ariyan, S., Narayan, D., & Truini, C. (2007). Diagnosis and treatment of interval sentinel lymph nodes in patients with cutaneous melanoma. *Plastic and Reconstructive Surgery*, *119*(3), 907–913.
Chang, S.-H., & Wang, M.-C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association*, *94*(448), 1221–1230.
Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, *8*(4), 537–569.

Choi, H., Son, H., & Kim, C. (2018). Predicting financial distress of contractors in the construction industry using ensemble learning. *Expert Systems with Applications*, *110*, 1–10.

Choodari-Oskooei, B., Royston, P., & Parmar, M. K. (2012). A simulation study of predictive ability measures in a survival model II: explained randomness and predictive accuracy. *Statistics in Medicine*, *31*(23), 2644–2659.

Cielen, A., Peeters, L., & Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, *154*(2), 526–532.

Clayton, D. (1994). Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research*, *3*(3), 244–262.

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, *3*(3), 167–179.

Dimitras, A. I., Slowinski, R., Susmaga, R., & Zopounidis, C. (1999). Business failure prediction using rough sets. *European Journal of Operational Research*, *114*(2), 263–280.

du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, *75*, 25–43.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, *83*(3), 635–665.

Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, *7*(2), 1477–1493.

Ejoku, J., Odhiambo, C., & Chaba, L. (2020). Analysis of recurrent events with associated informative censoring: Application to HIV data. *International Journal of Statistics in Medical Research*, *9*(21).

Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: an empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236–247.

Gepp, A., & Kumar, K. (2008). The role of survival analysis in financial distress prediction. *International Research Journal of Finance and Economics*, *16*(16), 13–34.

Gilson, S. C. (1989). Management turnover and financial distress. *Journal of Financial Economics*, *25*(2), 241–262.

Godlewski, C. J. (2015). The dynamics of bank debt renegotiation in Europe: A survival analysis approach. *Economic Modelling*, *49*, 19–31.

Gönen, M., & Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, *92*(4), 965–970.

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, *18*(17–18), 2529–2545.

Graf, E., & Schumacher, M. (1995). An investigation on measures of explained variation in survival analysis. *Journal of the Royal Statistical Society. Series D*, *44*(4), 497–507.

Harrell Jr, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*(4), 361–387.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, *117*, 287–299.

Hu, Y. C., & Ansell, J. (2007). Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research*, *183*(3), 1595–1606.

Hu, D., & Zheng, H. (2015). Does ownership structure affect the degree of corporate financial distress in China? *Journal of Accounting in Emerging Economies*, *5*(1), 35–50.

Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, *33*(2), 434–440.

Jiang, Y., & Jones, S. (2018). Corporate distress prediction in China: a machine learning approach. *Account Finance*, *58*, 1063–1109.

Jiang, S. T., Landers, T. L., & Reed Rhoads, T. (2006). Proportional intensity models robustness with overhaul intervals. *Quality and Reliability Engineering International*, *22*(3), 251–263.

John, T. A. (1993). Accounting measures of corporate liquidity, leverage, and costs of financial distress. *Financial Management*, *22*(3), 91–100.

Kahl, M. (2002). Economic distress, financial distress, and dynamic liquidation. *The Journal of Finance*, *57*(1), 135–168.

Kam, A., Citron, D., & Muradoglu, G. (2010). Financial distress resolution in China – two case studies. *Qualitative Research in Financial Markets*, *2*(2), 46–79.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Kim, M., Ma, S., & Zhou, Y. (2016). Survival prediction of distressed firms: evidence from the Chinese special treatment firms. *Journal of the Asia Pacific Economy*, *21*(3), 418–443.

Kristanti, Farida Titik, & Herwany, Aldrin (2017). Corporate governance, financial ratios, political risk and financial distress: A survival analysis. *Accounting and Finance Review*, *2*(2), 26–34.

Kuhnen, C. M., & Melzer, B. T. (2018). Noncognitive abilities and financial delinquency: the role of self-efficacy in avoiding financial distress. *The Journal of Finance*, *73*(6), 2837–2869.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review. *European Journal of Operational Research*, *180*(1), 1–28.

Lacher, R. C., Coats, P. K., Sharma, S. C., & Fant, L. F. (1995). A neural network for classifying the financial health of a firm. *European Journal of Operational Research*, *85*(1), 53–65.

Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An application of the cox proportional hazards model to bank failure. *Journal of Banking & Finance*, *10*(4), 511–531.

Lee, M. C. (2014). Business bankruptcy prediction based on survival analysis approach. *International Journal of Computer Science & Information Technology*, *6*(2), 103.

Leonardis, D. D., & Rocci, R. (2008). Assessing the default risk by means of a discrete – time survival analysis approach. *Applied Stochastic Models in Business and Industry*, *24*(4), 291–306.

Leong, R. W., Nguyen, N., Meredith, C. G., et al. (2008). In vivo confocal endomicroscopy in the diagnosis and evaluation of celiac disease. *Gastroenterology*, *135*(6), 1870–1876.

Li, Z., Crook, J., & Andreeva, G. (2014). Chinese companies distress prediction: an application of data envelopment analysis. *Journal of the Operational Research Society*, *65*(3), 466–479.

Li, Z., Crook, J., & Andreeva, G. (2017). Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications*, *80*, 94–106.

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

Li, H., Wang, Z., & Deng, X. (2008). Ownership, independent directors, agency costs and financial distress: evidence from Chinese listed companies. *Corporate Governance: The International Journal of Business in Society*, *8*(5), 622–636.

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

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, *274*(2), 743–758.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, *50*(3), 163–170.

Martin, D. (1977). Early warning of bank failure: a logit regression approach. *Journal of Banking & Finance*, *1*(3), 249–276.

Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *The Journal of Finance*, *29*(2), 449–470.

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.

Moulton, L. H., & Dibley, M. J. (1997). Multivariate time-to-event models for studies of recurrent childhood diseases. *International Journal of Epidemiology*, *26*(6), 1334–1339.

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

O'Neill, H. M. (1986). Turnaround and recovery: What strategy do you need? *Long Range Planning*, *19*(1), 80–88.

Paradi, J. C., Asmild, M., & Simak, P. C. (2004). Using DEA and worst practice DEA in credit risk evaluation. *Journal of Productivity Analysis*, *21*(2), 153–165.

Parker, S., Peters, G. F., & Turetsky, H. F. (2005). Corporate governance factors and auditor going concern assessments. *Review of Accounting and Finance*, *4*(3), 5–29.

Peña, E. A., Slate, E. H., & González, J. R. (2007). Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, *137*(6), 1727–1747.
Pfennig, A., Schlattmann, P., Alda, M., Grof, P., Glenn, T., Müller-Oerlinghausen, B., et al. (2010). Influence of atypical features on the quality of prophylactic effectiveness of long-term lithium treatment in bipolar disorders. *Bipolar Disorders*, *12*(4), 390–396.
Platt, H. D., & Platt, M. B. (2002). Predicting corporate financial distress: Reflections on choice-based sample bias. *Journal of Economics and Finance*, *26*(2), 184–199.
Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, *68*, 373–389.
Rahman, M. S., Ambler, G., Choodari-Oskooei, B., & Omar, R. Z. (2017). Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Medical Research Methodology*, *17*(1), 60.
Schemper, M., & Stare, J. (1996). Explained variation in survival analysis. *Statistics in Medicine*, *15*(19), 1999–2012.
Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business*, *74*(1), 101–124.
Sun, J., Jia, M. Y., & Li, H. (2011). Adaboost ensemble for financial distress prediction: an empirical comparison with data from Chinese listed companies. *Expert Systems with Applications*, *38*(8), 9305–9312.
Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management Science*, *38*(7), 926–947.
Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, *30*, 394–419.
Twisk, J., Smidt, N., & de Vente, W. (2005). Applied analysis of recurrent events: a practical overview. *Journal of Epidemiology and Community Health*, *59*, 706–710.
Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, *30*(10), 1105–1117.
Verikas, A., Kalsyte, Z., Bacauskiene, M., & Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing*, *14*(9), 995–1010.
Wang, Y. L., & Carson, J. (2010). Macroeconomic factors and insurer rating transitions. *Forensic Economics EJournal*.
Wang, Z., & Deng, X. (2006). Corporate governance and financial distress. *The Chinese Economy*, *39*(5), 5–27.
Wang, Z., & Li, H. (2007). Financial distress prediction of Chinese listed companies: a rough set methodology. *Chinese Management Studies*, *1*(2), 93–110.
Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association*, *84*(408), 1065–1073.
Wruck, K. H. (1990). Financial distress, reorganization, and organizational efficiency. *Journal of Financial Economics*, *27*(2), 419–444.
Wu, D., Liang, L., & Yang, Z. (2008). Analysing the financial distress of Chinese public companies using probabilistic neural networks and multivariate discriminate analysis. *Socio-Economic Planning Sciences*, *42*(3), 206–220.
Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, *38*(7), 8336–8342.