Estimating corporate bankruptcy forecasting models by maximizing discriminatory power

Chris Charalambous[1] · Spiros H. Martzoukos[2] · Zenon Taoushianis[3]

Abstract

In this paper, we estimate coefficients of bankruptcy forecasting models, such as logistic and neural network models, by maximizing their discriminatory power as measured by the Area Under Receiver Operating Characteristics (AUROC) curve. A method is introduced and compared with traditional logistic and neural network models, using out-of-sample analysis, in terms of discriminatory power, information content and economic impact while we forecast bankruptcy one year ahead, two years ahead but also financial distress, which is a situation that precedes firm bankruptcy. Using US public firms over the period 1990–2015, in all, we find that training models to maximize AUROC, provides more accurate out-of-sample forecasts relative to training them with traditional methods, such as maximizing the log-likelihood function, highlighting the benefits arising by using models with maximized AUROC. Among all models, however, a neural network trained with our method is the best performing one, even when we compare it with other methods proposed in the literature to maximize AUROC. Finally, our results are more pronounced when we increase the forecasting difficulty, such as forecasting financial distress. The implementation of our method to train bankruptcy models is robust in various settings and therefore well-justified.

✉ Zenon Taoushianis
z.taoushianis@soton.ac.uk

Chris Charalambous
bachris@ucy.ac.cy

Spiros H. Martzoukos
baspiros@ucy.ac.cy

[1] Department of Business and Public Administration, School of Economics and Management, University of Cyprus, P.O. Box 20537, 1678 Nicosia, CY, Cyprus

[2] Department of Accounting and Finance, School of Economics and Management, University of Cyprus, P.O. Box 20537, 1678 Nicosia, CY, Cyprus

[3] Department of Banking and Finance, Southampton Business School, University of Southampton, Southampton SO17 1BJ, UK

1 Introduction

1.1 Background and motivation

Increased attention has been paid in recent years for the development of powerful bankruptcy forecasting models, mainly for two reasons. First, the recent global financial crisis in 2007–2009 has left banks to experience huge losses from their credit portfolios and consequently their lending policies and decision-making processes have been seriously criticized from regulators, investors and other stakeholders. Second, since the reform of Basel Accord in 2006, banks can develop their own internal models to assess credit risks and protect themselves through the capital reserves that should withhold to face potential losses. Thus, for a matter of bank viability,[1] financial stability and investor protection, it would be of great interest to develop powerful bankruptcy forecasting models, which is the aim of this paper.

One of the most significant measures to evaluate the performance of bankruptcy forecasting models is their ability to discriminate bankrupt from healthy firms. It has been shown that models with higher discriminatory power are associated with higher economic benefits for a bank (Bloechlinger et al. 2006; Agarwal et al. 2008). Furthermore, Bauer et al. (2014) show that even small differences in the discriminatory power among bankruptcy forecasting models yield superior bank economic performance. In addition, commercial vendors and industry experts, such as Moody's KMV, extensively use discriminatory power as an integral part of their validation processes, especially when comparing their newly developed models with existing ones (see for instance the RiskCalc 3.1 model in Dwyer et.al. 2004). As it is stated in their paper:

> *"The greatest contribution to profitability, efficiency and reduced losses comes from the models' powerful ability to rank-order firms by riskiness so that the bank can eliminate high risk prospects."*

Beyond that, Moody's KMV provides ample explanatory documentation on how to use various discriminatory power measures in practice (see for instance Keenan et al. 1999 and Sobehart et al. 2000) and it is also extensively used in academic research to compare various bankruptcy forecasting models. This extensive use from practitioners and academics alike, in fact, highlights the importance of using discriminatory power as a leading measure to evaluate the performance of bankruptcy forecasting models.

Despite the empirical evidence on the economic benefits arising by using models with higher discriminatory power, it is somewhat surprising that a common practice in bankruptcy forecasting is to use discriminatory power only ex-post as an indication of model performance, rather than obtaining model coefficients directly by maximizing discriminatory power. Exceptions include Miura et al. (2010) and Kraus et al. (2014) in the related area of credit scoring which we discuss and compare with our method. We contribute to this limited literature by introducing a method that we use to train bankruptcy forecasting models, such as logistic and neural network models and comparing these models with traditional logistic and neural network models, such as models which maximize the log-likelihood function. Ultimately, our goal in this study is to highlight the importance of using models which are trained to maximize discriminatory power.

---

[1] See for instance Papakyriakou et al. (2019) for the consequences of bank failures around the globe.

To measure discriminatory power, we use the Area Under Receiver Operating Characteristics curve (AUROC or AUC). This is a widely used statistic that has been employed by many studies recently to compare discriminatory power of various bankruptcy forecasting models (including Chava et al. 2004; Campbell et al. 2008; Tinoco et al. 2013; Filipe et al. 2016 and many others). Furthermore, it has been used in related areas, such as mortgage default prediction (Fitzpatrick et al. 2016) and generally when assessing the performance of credit scoring models (see for instance Lessmann et al. 2015). Moreover, the AUROC is an appealing measure because it is easy to interpret and compute empirically and importantly, it does not depend on cut-off values, such as those needed when constructing the standard confusion matrices. Instead, the AUROC simply summarizes discriminatory power in a single number, thus it is easy to compare across various models, without using such cut-off values, which is the main reason it is has received considerable attention in bankruptcy studies. Due to these reasons, we select AUROC as the optimization criterion and we develop a method which seeks to maximize AUROC.

For our main analysis we collect annual financial data and daily equity prices for a large sample of U.S. public bankrupt and healthy firms and construct variables to make one-year forecasts, two-year forecasts and finally, we forecast financial distress which is a situation prior to the formal bankruptcy filing, over the period 1990–2015. We keep approximately 70% of the whole sample as a training set and evaluate the performance of the models in the testing set using three distinct type of tests, following Bauer et al. (2014); 1) AUROC analysis 2) Information content tests 3) Economic performance, when banks use various bankruptcy forecasting models in a competitive loan market.

1.2 Main findings

First, we employ standard statistical analysis to select few predictive variables, from a pool of variables, which individually exhibit high discriminatory power, have low correlation from each other and are statistically significant. This is to eliminate insignificant variables that may add noise and helps us constructing parsimonious models. When we consider only financial variables, we find that several financial variables related to firm leverage, profitability, liquidity and coverage, are significant predictors of bankruptcy. When we also consider market-based variables in the analysis, however, the model with both financial and market variables outperforms the model with only financial variables, consistent with prior research (Shumway 2001; Chava et al. 2004; Campbell et al. 2008; Wu et al. 2010 and Tinoco et al. 2013). These two selected sets of variables (financial variables and financial with market variables) are the inputs to all models (logistic and neural networks trained to maximize AUROC and the log-likelihood function).

We begin our analysis by evaluating and comparing the out-of-sample performance of logistic and neural network models trained with our method, to those trained to maximize the log-likelihood function, one year ahead, two years ahead and finally, when we forecast financial distress. Overall, we find that our proposed method yields logistic and neural network models which outperform, out-of-sample, traditional logistic and neural network models. The results with respect to the three testing approaches suggest that models with maximized AUROC 1) Significantly outperform the traditional models in terms of their ability to discriminate bankrupt from healthy firms, 2) They provide significantly more information about future bankruptcy-financial distress relative to traditional methods 3) Banks using models with maximized AUROC earn superior returns on a risk-adjusted basis relative to banks that use traditional models to forecast bankruptcy-financial distress. From all models, however, our

neural network is the best performing one. In addition, the results are more pronounced in the case of financial distress.

Next, we compare our method with other methods proposed in the literature to maximize AUROC. Using our proposed neural network model as a representative since it is the best performing model in all tests, we find that it outperforms the alternative AUROC maximization methods proposed by Miura et al. (2010) and Kraus et al. (2014). This result is more pronounced when forecasts are performed two years ahead and to the case of financial distress.

Finally, we compare, out-of-sample, the discriminating ability of logistic and neural networks trained by maximizing the AUROC, to the models trained with the traditional approach but this time, the input variables are constructed using quarterly data. In this way, we update the models as new information becomes available with higher frequency. In all, findings advocate the implementation of our estimation method since it provides better prediction performance, in terms of discriminatory power, relative to traditional estimation methods.

Our paper has implications in the way bankruptcy analysis is conducted and aims towards better decision-making through more accurate bankruptcy forecasts. First, our study can be viewed as a way towards improving the general practice of bankruptcy forecasting by providing an alternative estimation technique to obtain model coefficients relative to obtaining them with traditional methods. To this end, our proposed estimation method significantly improves performance, out-of-sample, especially when we increase the forecasting difficulty, such as forecasting bankruptcy two years in the future but also forecasting financial distress, which is a situation before the formal firm bankruptcy. In addition, our paper provides an extended methodological framework to commonly used traditional bankruptcy models such as Altman (1968); Ohlson (1980) but also more recent ones, such as Campbell et al. (2008) and many other similar models, by introducing a new optimization method to obtain their coefficients and increase forecasting accuracy. Finally, the advantage of our method is that it works well using any modelling approach where the output is a probability, thus it retains the same interpretability with the outputs of traditional estimation methods. This is also in contrast with the methodologies proposed by Miura et al. (2010) and Kraus et al. (2014) as these methods cannot be used by logistic and neural network models which are two of the most popular bankruptcy forecasting models (se for instance Kumar et al. 2007 and references therein). To the best of our knowledge, this is the first time such extensive work is performed to compare maximizing AUROC with the traditional maximizing the log-likelihood function for bankruptcy forecasting models and highlighting the benefits arising by AUROC-maximized models.

The remainder of the paper proceeds as follows: In Sect. 2 we discuss data collection, in Sect. 3 we present the methodology to maximize AUROC as well as the three distinct type of tests we use to evaluate performance, in Sect. 4 we discuss the results and Sect. 5 concludes.

2 Data

2.1 Sample

Our sample consists of 11,096 non-financial U.S. firms from which 422 filed for bankruptcy under Chapter 7 or Chapter 11 between 1990 and 2015. We have a total of

97,133[2] firm-year observations with non-missing data to forecast bankruptcies using the corresponding data which are lagged by one or two years for our one or two year ahead forecasts respectively but also, to forecast financial distress.[3] Bankrupt firms and the date of their bankruptcy filing were identified from BankruptcyData, which is a comprehensive database containing corporate bankruptcy and distressed information for firms in the US. Table 1 reports the frequency of bankrupt and healthy firms (i.e. non-bankrupt firms) collected each year over the sample period spanning the years 1990–2015.

Since our main "distress" event is bankruptcy, we treat exits unrelated to bankruptcy as non-bankrupt observations[4] (i.e. healthy firms) and we report them in Table 1. In particular, Table 1 breaks down the healthy firms, each year, into three categories; 1) active firms are firms that survived during the year, 2) firms that stopped filing information due being merged or acquired (M&As) and 3) firms that stopped filing information for other reasons including conversion to private company, engaged in a levarage buyout etc. The delisting reasons was found in COMPUSTAT using the *DLRSN* variable which provides codes for each delisting reason.

Next, Fig. 1 presents graphically the yearly number of bankruptcies, to visualize the variation of the number of bankruptcies over our sample period. Figure 1 shows that bankruptcies peak in three major time-periods: 1) during the 1990–1991 US crisis, 2) during the dot-com bubble occurred around 2000 and 3) during the financial crisis period when bankruptcies peaked in 2009. Overall, the plot shows that the sample period we use captures the prevailing market conditions with higher (lower) number of bankruptcies during crisis (normal) periods.

2.2 Variables construction

We collect annual financial data and market (equity) data from Compustat and CRSP respectively and we construct several variables based on related studies in the literature. For example, in our analysis we consider variables used in traditional corporate bankruptcy studies, such as Altman (1968), Ohlson (1980), Zmijewski (1984) but also in more recent studies, such as Shumway (2001), Chava et al. (2004), Campbell et al. (2008) etc.

First, we construct financial ratios capturing aspects of a firm's financial performance, such as leverage, profitability, liquidity, coverage, activity, cash flows, as presented in panel A of Table 2. A limitation of financial variables is that by their nature look backwards and the quality of information they carry depends on accounting practices (Hillegeist et al. 2004; Agarwal et al. 2008). Market variables, instead, constructed from equity prices, are forward-looking since they carry market perceptions about the prospects of the firm. For publicly traded firms it would be more appropriate to incorporate market variables in the

---

[2] Possibly, there are two main reasons that make our sample to differ from other studies, such as Campbell et al. (2008). First, we exclude financial firms with SIC codes in the range of 6000 which is something commonly done in the literature (corresponds to more than 8,000 unique firms over our sample period). Second, we delete observations with missing data whereas Campbell et al. (2008) replace missing data with sample averages.

[3] We present more details about the financial distress case in Sect. 4.5.

[4] In other words, unhealthy firms that did not file for bankruptcy but eventually exited the sample, are considered as healthy observation. Failure to account for this, it will overestimate the predictive performance.

Table 1 Yearly distribution of bankrupt and non-bankrupt (i.e. healthy) firms

Year	Bankrupt Firms	Non-Bankrupt Firms			Total
		Active Firms	M&A's	Others	
1990	22	3215	85	12	3334
1991	25	3208	57	6	3296
1992	17	3239	46	6	3308
1993	20	3295	62	2	3379
1994	10	3487	112	4	3613
1995	14	3873	144	4	4035
1996	14	4186	167	5	4372
1997	13	4520	233	4	4770
1998	19	4675	286	2	4982
1999	27	4554	351	2	4934
2000	20	4336	318	6	4680
2001	21	4317	236	28	4602
2002	14	4115	118	30	4277
2003	15	3790	135	38	3978
2004	13	3500	115	12	3640
2005	15	3433	148	18	3614
2006	10	3432	163	4	3609
2007	14	3298	215	6	3533
2008	20	3288	136	5	3449
2009	31	3196	94	6	3327
2010	6	3046	139	2	3193
2011	9	2985	128	8	3130
2012	12	2920	118	6	3056
2013	12	2873	106	7	2998
2014	12	2872	85	14	2983
2015	17	2900	114	10	3041

This table shows the yearly distribution of bankrupt and healthy firms (i.e. non-bankrupt firms) over the sample period 1990–2015. Bankrupt firms were identified in the BankruptcyData database. The remaining firms which did not file for bankruptcy and have the relevant data in COMPUSTAT and CRSP databases, are considered as healthy firms. Healthy firms are those who survived each year (active firms) or exited for reasons other than bankruptcy such as, mergers and acquisitions (M&A's) or Other reasons (Others). Reasons for company deletion are obtained from COMPUSTAT using the *DLRSN* variable (01 is the code for M&A's and Other reasons include, for instance, leverage buyouts, conversion to private company with codes 06 and 09 respectively, etc.)

models. To this end, we collect daily equity prices from CRSP for the entire fiscal year and several market-based variables are constructed, as reported in panel B of Table 2. Annualized volatility of daily equity returns (*VOLE*) refers to the fluctuations of firm's equity value returns, expecting to be higher for bankrupt firms. Next, excess return (*EXRET*) refers to the difference between firm's annualized equity return and the annualized value-weighted return of a portfolio with NYSE, AMEX, NASDAQ stocks, expecting to be lower
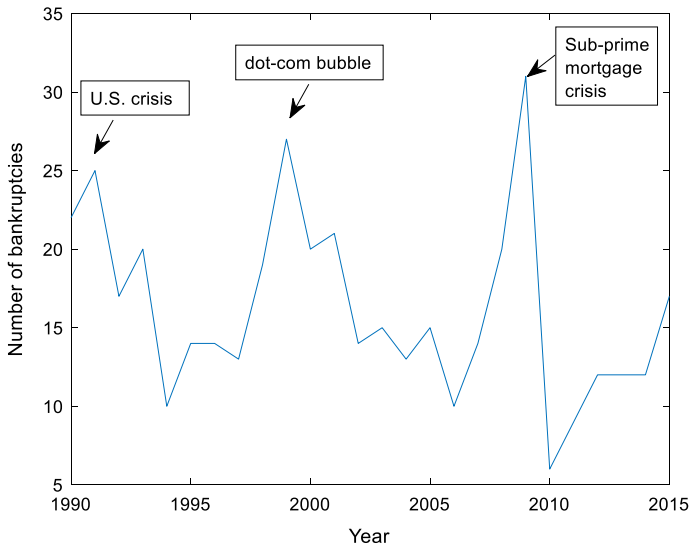
Fig. 1 This figure shows the yearly distribution of bankruptcies over the sample period 1990–2015

for bankrupt firms. Further, we consider the relative size of the firm (*RSIZE)*, the logarithm of stock price (*LOGPRICE*) and the Market-to-Book ratio (*MB*), expecting a negative association with bankruptcy risk. Finally, we include three financial variables scaled by firm's market value. More precisely, Campbell et al. (2008) show that scaling financial variables with a market-based measure of firm's value i.e. market equity + liabilities (*MTA*), compared to total assets as reported in the balance sheet, increases the predictive accuracy of bankruptcy forecasting models. These variables are cash over *MTA* (*CASHMTA),* net income over *MTA* (*NIMTA*), expecting a negative association with bankruptcy risk and lastly, total liabilities over *MTA* (*TLMTA*). Following common practice, we winsorize the variables between 1st and 99th percentile to avoid problems induced by outliers.

2.3 Variables selection

Table 2 presents an extensive list of variables that previous studies found to be significant predictors of bankruptcy risk. Out of these variables, a smaller set should be selected in order to construct parsimonious models with few variables but with high forecasting power. We establish a three-step approach to select the most powerful variables (see for instance Altman et al. 2007 and Filipe et al. 2016) and summarized in the following three steps:

Step 1: Removing variables with low discriminating ability (as a cut-off, we use AUROC equal to 0.60). The idea of this step is to qualify the variables that individually exhibit a satisfactory ability to discriminate bankrupt from healthy firms.

Step 2: Removing highly correlated variables using the Variance Inflation Factor (VIF) criterion. The idea of this step is to remove the variables that are highly correlated with others, since multicollinearity may yield misleading results regarding the significance of the variables in the final model. Beyond that, we end up with variables that provide

Table 2 List of financial and market variables

Panel A: Financial Ratios (Compustat)

Variable	Detailed Description	Compustat Item
NITA	Net Income/Total Assets	NI/AT
EBITTA	Earnings Before Interests and Taxes/Total Assets	EBIT/AT
RETA	Retained Earnings/Total Assets	RE/AT
CASHTA	Cash and Short-Term Investments/Total Assets	CHE/AT
WCTA	Working Capital/Total Assets	WCAP/TA
STDTA	Debt in Current Liabilities/Total Assets	DLC/AT
TLTA	Total Liabilities/Total Assets	LT/AT
CLCA	Current Liabilities/Current Assets	LCT/ACT
EBITCL	Earnings Before Interests and Taxes/Current Liabilities	EBIT/LCT
NICL	Net Income/Current Liabilities	NI/LCT
CFOTA	Operating Cash Flows/Total Assets	OANCF/AT
CFOTL	Operating Cash Flows/Total Liabilities	OANCF/LT
SLTA	Sales/Total Assets	SALE/AT
LOGASSETS	Natural logarithm of Total Assets	LOG(AT)

Panel B: Market Variables (CRSP)

Variable	Detailed Description	Compustat Item
VOLE	Annualized volatility of daily equity returns	
EXRET	Annualized equity return minus the value-weighted return of NYSE, AMEX, NASDAQ stocks	
LOGPRICE	Natural logarithm of the stock price, at the fiscal-year end	
RSIZE	Natural logarithm of firm's market capitalization over the total market capitalization of NYSE, AMEX, NASDAQ stocks	
MB	Firm's market capitalization over book value of equity (Market-to-Book ratio)	
TLMTA	Total Liabilities/ (Market Capitalization + Total Liabilities)	
NIMTA	Net Income/ (Market Capitalization + Total Liabilities)	
CASHMTA	Cash and Short-Term Investments/ (Market Capitalization + Total Liabilities)	

This table shows all financial ratios and market variables that we consider to construct the bankruptcy prediction models. From these, only a set of variables are selected according to a three-step procedure described in the text

Table 3 Descriptive statistics for the selected variables

	TLTA	STDTA	NITA	CASHTA	EBITCL	LOG-PRICE	EXRET	CASH-MTA	NIMTA
Bankrupt Firms									
Mean	0.854	0.181	− 0.382	0.107	− 0.578	0.528	− 0.224	0.0684	− 0.250
Median	0.825	0.099	− 0.211	0.045	− 0.153	0.560	− 0.349	0.032	− 0.172
St.Dev	0.327	0.183	0.450	0.157	1.289	1.146	0.876	0.100	0.257
Healthy Firms									
Mean	0.486	0.048	− 0.042	0.190	0.07	2.293	0.205	0.121	− 0.022
Median	0.478	0.014	0.031	0.101	0.153	2.474	0.106	0.063	0.022
St.Dev	0.253	0.083	0.254	0.217	1.322	1.266	0.684	0.162	0.149

This table reports descriptive statistics for the financial and market variables which enter in the final models, one year prior to bankruptcy, for both bankrupt and healthy firm observations. The definition of the variables can be found in Table 2

different information and explain bankruptcy uniquely. We use 5 as cut-off (variables with VIF ≥ 5 are removed).

Step 3: Performing a stepwise multivariate logistic regression to the remaining variables in order to obtain the most significant variables from a statistical point of view (we use a significance level of $\alpha = 5\%$). The logistic regression program estimates coefficients assuming independent observations, which is an invalid assumption, since the data contains information for firms over multiple periods. In such case, an appropriate correction measure which we adopt in our study, is to use clustered robust standard errors (also used by Filipe et al. 2016).

Using the three-step approach, we develop two types of models. The first one is a "private firm" type of model, including only financial variables. We further develop a "public firm" type of model, including both financial and market variables. For example, the private firm model includes five financial variables (*TLTA, STDTA, NITA, CASHTA, EBITCL*), while the public firm model includes six variables (*TLTA, STDTA, LOGPRICE, CASHMTA, NIMTA, EXRET*). Notice that two financial-based variables (*CASHTA* and *NITA*) are replaced with *CASHMTA* and *NIMTA*. Generally, the majority of variables that are found to be significant for the public firm model are market variables, which is consistent with the perception that market-based variables are better bankruptcy risk measures, due to their forward-looking nature. These two sets of variables are the inputs to all models (i.e. used in the models which we train to maximize AUROC and in the models trained to maximize the log-likelihood function). For simplicity, these two sets of variables are also used in the models when forecasting bankruptcy two years ahead and when forecasting financial distress.

2.4 Descriptive statistics

Table 3 reports descriptive statistics for the accounting and market variables that we find to be significant predictors of bankruptcy. As expected, bankrupt firms are more levered on average relative to healthy firms (*TLTA* and *STDTA* for bankrupt firms are higher), they are also less profitable (*NITA* and *NIMTA* are lower for bankrupt firms). Furthermore, bankrupt

firms are more constrained in terms of cash available (*CASHTA* and *CASHMTA* are lower) as opposed to healthy firms. Going to the market variables, it is evident that the stock price of bankrupt firms (*LOGPRICE)* on average is lower than healthy firms, possibly due to their deteriorating financial position that is priced by investors, leading to a depreciation of their stock prices at the year prior to bankruptcy. Finally, bankrupt firms exhibit lower and negative market performance relative to the market (*EXRET* is lower one year prior to bankruptcy), as opposed to healthy firms.

3 Methodology

3.1 Measuring discriminatory power

Discriminatory power refers to the ability of a model to discriminate bankrupt from healthy firms. According to a cut-off score, firms whose bankruptcy score exceeds the cut-off are classified as bankrupt and healthy otherwise. Therefore, a way to measure the discriminating ability of a model is, for a given cut-off score, to count the true forecasts (percentage of bankrupt firms correctly classified as bankrupt) and the false forecasts (percentage of healthy firms incorrectly classified as bankrupt). Doing this process with multiple cut-offs, we get a set of true and false forecasts. A graph made from this set is the ROC curve with false forecasts on the x-axis and true forecasts on the y-axis. A perfect model would always (never) make true (false) forecasts and thus its ROC curve would pass through the point (0,1). Generally, the closer the ROC curve to the top-left corner, the better the discriminatory power of the model.

The ROC curve provides a graphical way to visualize discriminatory power. A quantitative assessment of the discriminatory power is given by the Area under ROC curve (AUROC) which is calculated as follows[5]:

$$\widehat{AUROC} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} I\left(s_B^i - s_H^j > 0\right) \tag{1}$$

where $I(x)$ is an indicator function, defined to be 1 if x is true and 0 otherwise, s_B^i and s_H^j denote the bankruptcy scores of a model for the *i-th* bankrupt firm and for the *j-th* healthy firm observation respectively. Finally, n is the number of bankrupt firms and m is the number of healthy firm observations. Note that Eq. (1) is discontinuous and non-differentiable.

3.2 Maximizing discriminatory power

In this section, we present a methodology to maximize the discriminatory power (AUROC) when the bankruptcy score, *s*, is a probability, meaning that the model has a probabilistic

---

[5] For further explanation, refer to Hanley et al. (1982) and Sobehart et al. (2001).

response function[6] which is the case of popular bankruptcy forecasting models such as logistic and neural network models.

Ideally, we should have used Eq. (1) directly as the objective function in the optimization. However, traditional gradient-based optimization methods cannot be used to maximize Eq. (1) directly because it is discontinuous and non-differentiable. For this reason, we introduce a surrogate function that seeks to maximize the discriminatory power. We define:

$$d_{i,j}(\beta) = s_B^i(X_i, \beta) - s_H^j(X_j, \beta) = p_B^i(X_i, \beta) - p_H^j(X_j, \beta) \qquad (2)$$

as the difference between the probability of bankruptcy for the *i-th* bankrupt firm, $p_B^i(X_i, \beta)$ and the probability of bankruptcy for the *j-th* healthy firm observation, $p_H^j(X_j, \beta)$, conditional on the predictor variables in X which could be a set of financial and market variables. From Eq. (1), to obtain the coefficients, β, that maximize the discriminatory power of a model we would like as many as possible $d_{i,j}$'s to be positive because AUROC increases in this way. A way to achieve this is through the minimization of the following surrogate merit function:

$$F(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \max\left(0, \gamma - d_{i,j}(\beta)\right) \qquad (3)$$

where $0 \leq \gamma \leq 1$. The above merit function ignores the terms where $d_{i,j}(\beta) > \gamma$ (meaning that the difference in bankruptcy probabilities between the *i-th* bankrupt firm and *j-th* healthy firm observation is relatively high, as specified by the parameter γ) and penalizes the terms where $d_{i,j}(\beta) \leq \gamma$. In other words, the parameter γ can be considered as a parameter which controls the magnitude of the $d_{i,j}$'s that are to be penalized. For instance, if $\gamma = 0$, we penalize only the negative $d_{i,j}$'s (i.e. only the cases where the model assigned a higher probability of bankruptcy for a healthy firm than a bankrupt firm) while if $\gamma = 1$, we penalize all $d_{i,j}$'s.

Based on the optimality conditions of minimizing $F(\beta)$, at the optimal solution, a number of $d_{i,j}$'s must satisfy the condition $d_{i,j} = \gamma$.[7] Hence, by selecting γ (close) to zero, we force a number of $d_{i,j}$'s to be close to zero in absolute terms. In that case, a small change of the input data can easily induce $d_{i,j}$'s to change signs which in turn will cause a change in the AUROC. This may be particularly evident in the case of out-of-sample data. That is, by training a model to produce $d_{i,j}$'s close to zero, may yield a model with poor generalization ability and consequently the out-of-sample AUROC will be very sensitive. On the other hand, selecting γ (close) to one, coefficient estimates can blow up and provide unreasonable results. Thus, theoretically, the parameter value should be in between 0 and 1 (we explain later in this section how we compute the parameter empirically).

However, the surrogate function in Eq. (3) is non-differentiable when $z = \gamma - d_{i,j}(\beta) = 0$. To overcome this problem and thus being able to use traditional gradient-based optimization algorithms, we replace the term $\max(0, z)$ with a differentiable function. Note that, we can minimize $F(\beta)$ given by Eq. (3) using linear programming provided that the response function is linear with respect to the coefficients, β. Here, the probability is a non-linear function and as such we should use non-linear optimization algorithms to obtain the coefficients. We replace the term $\max(0, z)$ by the following ε-smoothed differentiable approximation, $h_\varepsilon(z)$:

---

[6] A good choice for the probabilistic response function that is usually used in bankruptcy-related studies and also adopted in our study, is the logistic function; $p = 1/(1 + \exp(\beta*X))$.

[7] This draws on results from Charalambous (1979).
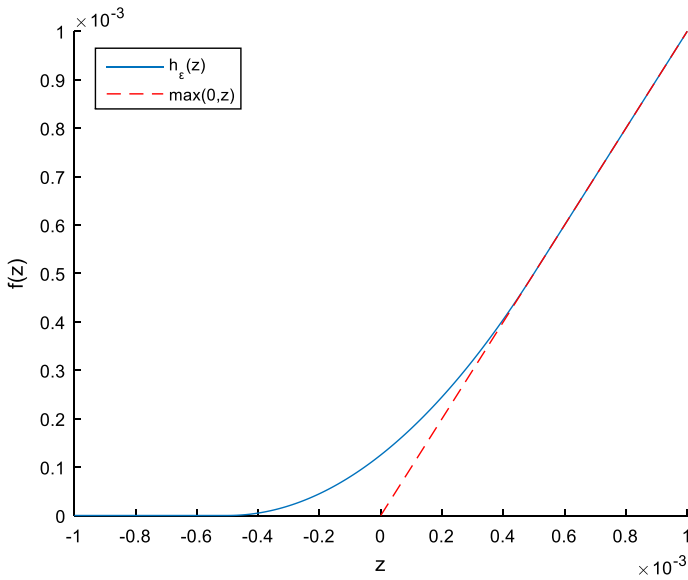
Fig. 2 The function max(0,z) is a surrogate function aiming to maximize the AUROC. However, this function is non-differentiable when z=0. Thus, we replace it with the differentiable ε-smoothed function, $h_\varepsilon(z)$
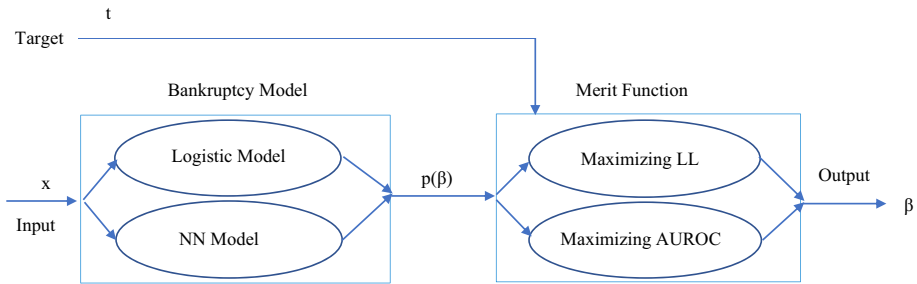
Fig. 3 This figure summarizes the work in our study and specifically how we train the models. The input vector, x, which can be financial and market data, enter the bankruptcy model (logistic or neural network model). The output of the model is the probability of bankruptcy, $p(\beta)$, which depends on the coefficients imposed by the model and enters the merit function along with the target, t. The merit function can be the log-likelihood function (LL) or the AUROC, which is the one we propose in this study to use as the optimization criterion to obtain model coefficients. At each iteration, the optimization algorithm updates the coefficients until the merit function is optimized. For training we use data from the period 1990–2006

$$h_\varepsilon(z) = \begin{cases} 0, & z \leq -\varepsilon/2 \\ \frac{1}{2\varepsilon}(z + \varepsilon/2)^2, & -\varepsilon/2 < z \leq \varepsilon/2 \\ z, & z > \varepsilon/2 \end{cases} \qquad (4)$$

where ε is a small positive number close to zero. Here we set $\varepsilon = 0.001$. The ε-smoothed function $h_\varepsilon(z)$, which we graphically present in Fig. 2, is a shifted version of the smoothed function used previously by Charalambous et al. (2007) to value call options. As can be

seen from the graph, the $h_\varepsilon(z)$ function has similar properties with the max $(0, z)$ except that $h_\varepsilon(z)$ is differentiable when z = 0.

Hence, the merit function to be minimized is replaced by:

$$F(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} h_\varepsilon\left(\gamma - d_{i,j}(\beta)\right) \tag{5}$$

The next step is to estimate the coefficients, β, by training the model to minimize F(β) given by Eq. (5). Figure 3 summarizes the work in our study.

Consider that we have N training input samples (i.e. observations). Each input sample, $x_n = [x_{1n}, x_{2n}, \ldots, x_{kn}]$, is associated with a known target,t_n, where $n = 1,2,\ldots, N$ and k is the number of variables. In the context of bankruptcy forecasting, the input sample x_n can be information characterizing the *n-th* firm, such as financial and market information, whereas t_n is an indicator variable which equals 1 if the corresponding firm-observation goes bankrupt and 0 otherwise. The inputs enter the bankruptcy model (logistic or neural network model) to produce the probability of bankruptcy which is a function of the coefficients imposed by the model. The output of the bankruptcy model, *p(β),* with the associated target, *t*, are used in the merit function. Traditionally, the log-likelihood function is used to obtain the coefficients. In this study, we propose another way to obtain coefficients and specifically we use the merit function given by Eq. (5) which is optimized in order to obtain the coefficients of the bankruptcy models and consequently the probability of bankruptcy. Note that the target, *t*, is indirectly used in the merit function given by Eq. (5) in order to identify the bankrupt and healthy firms and to estimate their probability of bankruptcy. In this paper, the training sample spans the period 1990–2006. To solve the problem, we formulate a nonlinear unconstrained optimization process using MATLAB. Specifically, we use the *fminunc* command and the *trust-region* optimization algorithm to obtain the coefficients of the logistic and neural network models. At each iteration, the optimization algorithm updates the coefficients and the probability of bankruptcy (as shown in Fig. 3) until the merit function we propose is optimized.

As far as the parameter γ is concerned, we compute it empirically based on validation-a straightforward and easy to implement approach, which makes use only the training data to determine the parameter γ while the testing data remain intact. Also, validation is a frequently used method implemented by many studies to determine parameters underlying the models. We further divide our training sample into training (70%) and validation (30%) sets.[8] We train the models by choosing from the set of parameter values $\gamma = \{0, 0.1, 0.2, \ldots, 1\}$ and keep the value that gives the highest AUROC on the validation set. For instance, using our private and public firm models we find that γ equals 0.3 and 0.1 respectively, consistent with our conjecture that the γ parameter should be between 0 and 1. Then we merge the training and validation sets, to train the models as explained before and test their performance on the testing set 2007–2015.

We further illustrate the role of γ by providing an example using our data to provide an idea of how our method works and why it increases AUROC. First, we estimate the coefficients of a logistic model by maximizing the log-likelihood function and we calculate the $d_{i,j}$'s. Second, we estimate the coefficients of a logistic model by minimizing

---

[8] We also use these two sets of data to compute the optimal number of neurons for the neural networks. We consider one, two, three and four neurons, starting also from various initial coefficient values and we select the number of neurons that performs the best (in terms of AUROC) in the validation set. We find that the optimal number of neurons is two. We also use a logistic transfer function in the hidden and output layer.
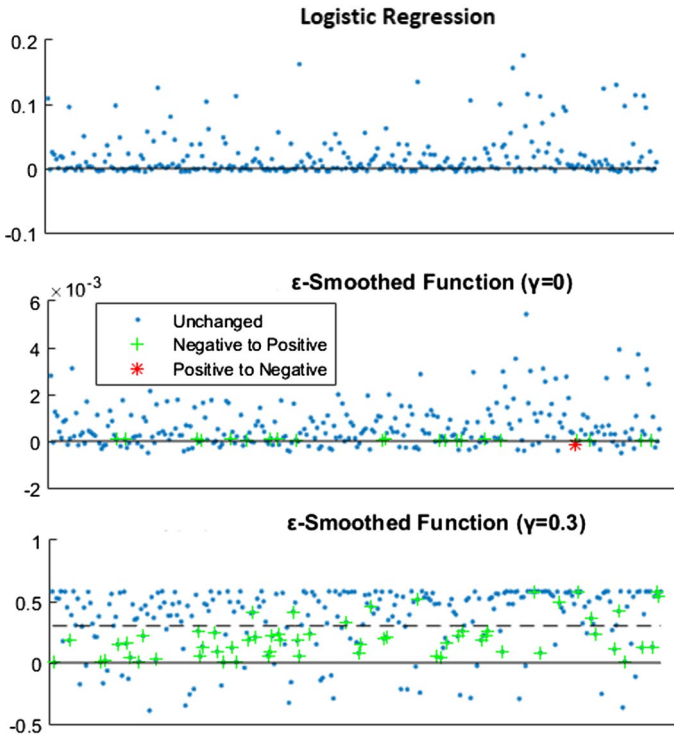
Fig. 4 This figure presents a sample of dij's of three models. The top plot presents the dij's generated by a logistic model trained to maximize the log-likelihood function, given by Eq. (6). The middle plot presents the same dij's generated by a logistic model but maximizng AUROC using our proposed ε-smoothed function given by Eq. (5) and setting the parameter γ = 0. The bottom plot presents the same dij's generated from a logistic model but trained to maximize AUROC using our proposed ε-smoothed function given by Eq. (5) and setting the parameter γ = 0.3

$F(\beta)$ given by Eq. (5) and we calculate the $d_{i,j}$'s. Figure 4 shows a sample of those $d_{i,j}$'s, produced by logistic regression i.e. by the model trained to maximize the log-likelihood function (top plot) and by maximizing AUROC with the ε-smoothed function, setting $\gamma = 0$ (middle plot) and $\gamma = 0.3$ (bottom plot). Recall that we would like as many as possible of $d_{i,j}$'s to be greater than zero. Hence, they should lie above the solid straight line. For the logistic regression, some lie above and some below. Using the ε-smoothed function, we want to make as many as possible negative $d_{i,j}$'s to move above the straight line. Setting $\gamma = 0$, we observe that all $d_{i,j}$'s are close to zero. Some cases, 21 in particular, that were negative according to the logistic regression became positive (denoted with green crosses) and one case that was positive became negative (denoted with a red star), highlighting the limitation of producing $d_{i,j}$'s that are close to zero. Setting $\gamma = 0.3$, not only more $d_{i,j}$'s that were negative according to logistic regression became positive (59 in particular), but now the majority lie well above the solid straight line, several also passing the γ parameter which are the points that lie above the dashed line. Notice now that none of the $d_{i,j}$'s that were positive according to logistic regression became negative because the higher value of γ, causes $d_{i,j}$'s to be well above zero and as a consequence, AUROC will not be sensitive.

Finally, as a benchmark, we obtain the coefficients of the logistic and neural networks models by maximizing the log-likelihood function, *LL*. Assuming that we have *N* training samples, *LL* is defined as follows:

$$LL(\beta) = \sum_{n=1}^{N} t_n \ln\left(p_n(x_n, \beta)\right) + \left(1 - t_n\right) \ln\left(1 - p_n(x_n, \beta)\right) \tag{6}$$

where $p_n(x_n, \beta)$ is the bankruptcy probability of the *n-th* observation, given the input vector of variables, x_n and the coefficients, β.

3.3 Information content tests

We further consider information content tests, also employed by related studies (see for instance Hillegeist et al. 2004; Agarwal et al. 2008; Charitou et al. 2013; Bauer et al. 2014). In such tests the out-of-sample bankruptcy probabilities produced by various models, such as by models with maximized AUROC, enter as inputs to logistic regression models and we are interested to assess their explanatory power. In particular, we estimate the following panel logit specification:

$$p\left(Y_{i,t+1} = 1 | prob_{i,t}\right) = p_{i,t} = \frac{e^{a_t + \beta * prob_{i,t}}}{1 + e^{a_t + \beta * prob_{i,t}}} = \frac{e^{a * Rate_t + \beta * prob_{i,t}}}{1 + e^{a * Rate_t + \beta * prob_{i,t}}} \tag{7}$$

where $p_{i,t}$ is the probability of bankruptcy at time *t*, that the *i-th* firm will go bankrupt the next year and $Y_{i,\,t+1}$ is the status of the *i-th* firm the next year (1 if it goes bankrupt and 0 if it is solvent). The variable of interest is $prob_{i,t}$, which is the out-of-sample bankruptcy probability of the *i-th* firm at time *t*, produced by a model, for instance with maximized AUROC. Finally, β is the coefficient estimate and a_t is the baseline hazard rate that is only time-dependent and it is common to all firms at time *t*. Similar with prior studies, we proxy the baseline hazard rate with the actual bankruptcy rate at time *t*.

The specification in Eq. (7) is equivalent with the hazard model specifications used in related bankruptcy studies, such as Hillegeist et al. (2004); Agarwal et al. (2008); Bauer et al. (2014) etc. Specifically, Shumway (2001) argues that a panel logit model, like the one in Eq. (7), is equivalent with a hazard rate model and therefore standard log-likelihood maximization procedures can be used to estimate the logit model in Eq. (7), with a minor adjustment that we explain below.

The model in Eq. (7) represents a multi-period logit model as it includes observations for each firm across time. However, the inclusion of multiple firm-year observations per firm yields understated standard errors because the log-likelihood objective function, which is maximized to estimate the multi-period logit model, assumes that each observation is independent from each other. This is a wrong assumption since firm observations at time *t + 1* cannot be independent from firm observation at time *t*. Failing to address this econometric issue, could lead to wrong inference regarding the significance of the individual coefficients. Similar with Filipe et al. (2016), we use clustered-robust standard errors to adjust for the number of firms in the sample but also for heteroskedasticity (Huber 1967 and White 1980).

3.4 Economic analysis of bankruptcy models

The analysis so far addressed the forecasting accuracy of the bankruptcy models. But how accuracy is economically beneficial for banks? In particular, Bauer et al. (2014) show that even small differences in the AUROCs between the models affect the profitability of a bank. Similar findings are found in Charalambous et al. (2020). Therefore, it would be interesting to investigate the effect of using models with maximized AUROC, on bank economic performance. Here, we follow the approach of Agarwal et al. (2008) and Bauer et al. (2014) to examine it by assuming a loan market worth $100 billion and banks compete to grant loans to individual firms. Each bank uses a bankruptcy model to evaluate the credit worthiness of their customers.

3.4.1 Calculating credit spreads

We estimate the models using data spanning the years 1990–2006 (70% of the sample). We sort firm-customers from this sample in 10 groups of equal size and a credit spread is calculated according to the following rule; Firms in the first group, which are firms with the lowest bankruptcy risk, are given a credit spread, k and firms in the remaining groups are given a credit spread, CS_i, obtained from Blochlinger et al. (2006) and it is defined as follows:

$$CS_i = \frac{p(Y = 1|S = i)}{p(Y = 0|S = i)} LGD + k \tag{8}$$

where $p(Y=1|S=i)$ and $p(Y=0|S=i)$ is the average probability of bankruptcy and non-bankruptcy respectively, for the i-th group, with i=2, 3, …,10 and LGD is the loan loss upon default. Following Agarwal et al. (2008), the average probability of bankruptcy for the i-th group is the actual bankruptcy rate for that group, defined as the number of firms that went bankrupt the following year divided by the number of firms in the group. Furthermore, $k=0.3\%$ and $LGD=45\%$.

3.4.2 Granting loans and measuring economic performance

To evaluate economic performance, we assume that banks compete to grant loans to prospective firm-customers between the period 2007–2015. Each bank uses a bankruptcy model that has been estimated in the period 1990–2006. The bank sorts those customers according to their riskiness and rejects the bottom 5% with highest risk. The remaining firms are classified in 10 groups of equal size and firms from each group are charged a credit spread that has been obtained from the period 1990–2006. Finally, the bank that charges the lowest credit spread for the customer (i.e. for the firm-year observation) is granting the loan. Two measures of profitability are used. The first one, Return on Assets (ROA) is defined as Profits/Assets lent and the second one, Return on Risk-Weighted Assets (RORWA) takes into consideration the riskiness of the assets, defined as Profits/Risk-Weighted Assets. Risk-Weighted Assets are obtained from formulas provided by the Basel Committee on Banking Supervision (2006).

Table 4 AUROC results-Out of sample (2007–2015)

Panel A: AUROC estimation

Models	Private firms model	Public firms models
Models with maximized AUROC		
Neural Network	0.9332	0.9508
Logistic	0.9221	0.9470
Models with maximized log-likelihood		
Neural Network	0.9138	0.9440
Logistic	0.8991	0.9425

Panel B: DeLong (1988) test statistic for differences in AUROCs

	Private Firms Model	Public Firms Model
Neural Network with max. AUROC vs Neural Network with max. LL	2.04	1.72
Logistic model with max. AUROC vs Logistic model with max. LL	2.43	1.68

This table reports AUROC results for logistic and neural network models in two cases; First, when their coefficients are estimated by maximizing the AUROC and second, when their coefficients are estimated by maximizing the log-likelihood function (denoted as LL). The models are trained in the period 1990–2006 and the table reports results in the out-of-sample period 2007–2015. In the models, we use either financial variables (Private Firms Models) or financial and market data (Public Firms Models)

4 Results

In this section, we present the out-of-sample comparisons between models with maximized AUROC and models with maximized log-likelihood (traditional models). We use the bankruptcy years 1990–2006 for training and keep the bankruptcy years 2007–2015 as the testing set. We start our analysis by comparing their performance in terms of discriminatory power, information content and economic benefits, when forecasting bankruptcy one year ahead.[9] Next, using the same tests, we compare their performance by forecasting bankruptcy two years ahead and then when our sample consists of financially distressed firms. An additional analysis is performed in this section where we compare our methodology with other methods proposed in the literature to maximize AUROC using the same analysis as before. Finally, we provide out-of-sample discriminatory power comparisons, when we use quarterly data.

---

[9] For instance, we train the models in the bankruptcy years 1990–2006 using the corresponding variables lagged by one year (1989–2005). To forecast bankruptcy in the bankruptcy years 2007–2015, we apply the models using as inputs, the variables lagged by one year (2006–2014). For example, if a firm goes bankrupt within 2010, we use its latest financial information in 2009 as inputs to the models so that by the beginning of 2010, investors have the information available to predict bankruptcy within 2010. For the two year ahead forecasts, we use its financial information in 2008 (but the firm is reported as bankrupt in 2010).

4.1 AUROC results

Table 4 shows the out-of-sample performance (2007–2015) of the models with maximized AUROC versus the models with maximized log-likelihood, in terms of discriminatory power.

Overall, models that are trained to maximize AUROC perform better out-of-sample compared to models trained to maximize the log-likelihood function, indicating that the function we introduced performs well out-of-sample in discriminating firms that will go bankrupt the next year. The effect by maximizing the AUROC, as expected, is more pronounced in the case of "private firms model" where only limited information is available (i.e. financial information), hence there is more space to improve the performance. For instance, the AUROC of the logistic and neural network model trained to maximize the log-likelihood function (LL) are 0.8991 and 0.9138 respectively where those trained to maximize AUROC are 0.9221 and 0.9332 respectively. DeLong tests indicate that AUROC differences are statistically significant at the 5% level (2.43 and 2.04 respectively). In contrast, the effect is less pronounced in the case of "public firms model", since the inclusion of market data in addition to financial data, further increases the forecasting power of the models. In fact, the AUROCs of models trained to maximize the log-likelihood function are quite high and specifically, 0.9425 for the logistic[10] model and 0.9440 for the neural network. This improves to 0.9470 and 0.9508 respectively when maximizing AUROC. Differences are statistically significant at the 10% level.

From the results in this section, we suggest using the neural network model trained to maximize AUROC since it is the best-performing model which is consistent with the notion that neural networks outperform simpler modeling approaches (Zhang et al. 1999; Kumar et al. 2007; Lessmann et al. 2015). For the user interested in simpler models, we suggest the implementation of the logistic model but trained to maximize AUROC.

4.2 Information content results

In this section we report the results from information contest tests. We compare the information contained in out-of-sample bankruptcy probabilities produced by models where the AUROC is maximized versus where the log-likelihood is maximized. Models 1 and 2 include the out-of-sample (2007–2015) bankruptcy probabilities produced by a neural network (Prob1) and a logistic model (Prob2) respectively, obtained by maximizing the AUROC. Models 3 and 4 include the bankruptcy probabilities produced by a neural network (Prob3) and a logistic model (Prob4) respectively, obtained by maximizing the log-likelihood function.

Table 5 reports the results of logit models that include the out-of-sample bankruptcy probabilities as explanatory variables but also the annual bankruptcy rate (Rate) as the baseline hazard rate.

---

[10] We acknowledge that such AUROC values are relatively high but can happen, even rarely, as they have been observed in prior research. Chava and Jarrow use Shumway's (2001) model with five accounting and market variables and report that AUROC can reach up to 0.9421. More recently, Charalambous et al. (2020) show that incorporating market-based information in accounting models, yields an AUROC that can reach up to 0.9449. Moreover, Afik et al. (2016) use the simple Black–Scholes-Merton default prediction model and show that its AUROC can reach up to 0.9636.

Table 5 Information content test results-Out of sample (2007–2015)

Variable	Private firms model				Public firms model			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Panel A: Logit models estimation								
Prob1	0.069				0.195			
	(0.004)				(0.012)			
Prob2		0.066				0.182		
		(0.004)				(0.008)		
Prob3			0.695				0.511	
			(0.028)				(0.055)	
Prob4				0.296				0.271
				(0.041)				(0.026)
Rate	-1.120	-1.061	-0.879	-0.218	-1.070	-1.431	-0.789	-0.370
	(0.387)	(0.387)	(0.387)	(0.371)	(0.470)	(0.454)	(0.469)	(0.402)
Constant	-8.018	-8.553	-5.716	-5.521	-20.324	-17.456	-5.673	-5.540
	(0.306)	(0.341)	(0.183)	(0.181)	(1.085)	(0.688)	(0.220)	(0.192)
LL	-601.35	-624.16	-687.86	-774.96	-554.91	-561.34	-680.25	-728.26
Pseudo-R^2	29.05%	26.36%	18.84%	8.56%	34.53%	33.77%	19.74%	14.07%
Panel B: Vuong test statistics for differences in log-likelihoods								
					Vuong test stat			
	Private Firms Model							
	Model 1 vs Model 3				5.75			
	Model 2 vs Model 4				7.37			
	Public Firms Model							
	Model 1 vs Model 3				4.37			
	Model 2 vs Model 4				8.07			

This table reports results from information content tests. Panel A shows estimation of four logit models. Models 1 and 2 include out-of-sample (2007–2015) bankruptcy probabilities produced by a neural network and a logistic model respectively, whose coefficients are obtained by maximizing AUROC, with financial data as inputs. Models 3 and 4 include bankruptcy probabilities produced by a neural network and a logistic model respectively, whose coefficients are obtained by maximizing the log-likelihood function, with financial data as inputs (private firm models). We re-estimate the four logit models 1–4 as explained before and report results in the next four columns but this time, the out-of-sample bankruptcy probabilities are generated with financial and market data as inputs to the neural network and logistic models (public firm models)

All logit regression models include the Rate as proxy for the baseline hazard rate, which is the prior year bankruptcy rate in our sample. The last two rows of the panel reports log-likelihood and pseudo-R^2 for each model. Panel B reports Vuong test statistics for differences in the log-likelihoods between the models

Panel A reports results from four logit regression models. Models 1–4 in the first four columns refer to the models where the corresponding bankruptcy probability which is included as predictor (Prob1-Prob4) is generated with financial data only. We re-estimate the four logit regressions, of which their results are presented in the next four columns of Table 4 but this time, the corresponding bankruptcy probability is generated with financial and market data.

According to the results, the bankruptcy probabilities in all cases are highly statistically significant, indicating that they carry significant information in predicting bankruptcy one year ahead, (coefficient estimates are significant at the 1% significance

Table 6 Economic performance results

	Private firms model				Public firms model			
	Bank1	Bank2	Bank3	Bank4	Bank1	Bank2	Bank3	Bank4
Credits	12,689	3,723	4,095	7,459	12,136	6,049	4,111	5,311
Market Share (%)	44.20	12.97	14.26	25.98	42.27	21.07	14.32	18.50
Bankruptcies	6	14	7	42	9	4	7	28
Bankruptcies/Credits (%)	0.047	0.38	0.17	0.56	0.074	0.066	0.17	0.53
Average Spread (%)	0.34	0.46	0.36	0.54	0.35	0.42	0.38	0.81
Revenues ($M)	151.30	59.77	51.06	139.28	148.64	88.01	54.52	150.70
Loss($M)	8.57	20.01	10.00	60.02	12.86	5.72	10.00	40.01
Profit($M)	142.73	39.76	41.06	79.26	135.78	82.29	44.52	110.69
Return on Assets (%)	0.32	0.31	0.29	0.31	0.32	0.39	0.31	0.60
Return on RWA (%)	2.24	1.27	1.52	0.91	1.98	1.53	1.68	1.20

This table reports economic results for four banks in a competitive loan market worth $100 billion. Bank 1 uses a neural network model with maximized AUROC for estimating the bankruptcy score of its customers. Bank 2 uses a logistic model with maximized AUROC. Banks 3 and 4 are competitors to banks 1 and 2 respectively, using a neural network and a logistic model respectively, but by maximizing the log-likelihood function. The table reports results in two cases; When the inputs to the models include financial data (private firm models) and when the inputs include financial and market data (public firm models)

The banks sort prospective customers (2007–2015) and reject the 5% of firms with the highest risk. The remaining firms are classified in 10 groups of equal size and for each group, a credit spread is calculated, as described in the main text. The bank that classifies the firm to the group with the lowest spread is finally granting the loan. Market share is the number of loans given divided by the number of firm-years, Revenues = (market size)*(market share)*(average spread), Loss = (market size)*(prior probability of bankruptcy)*(share of bankruptcies)*(loss given default). Profit = Revenues-Loss. Return on Assets is profits divided by market size*market share and Return on Risk-Weighted-Assets is profits divided by Risk-Weighted Assets, obtained from formulas provided by the Basel Accord (2006). The prior probability of bankruptcy is the bankruptcy rate for firms between 1990–2006 and equals 0.42%. Loss given default is 45%

level). More importantly, bankruptcy probabilities produced by models with maximized AUROC (Prob1 and Prob2) contain significantly more information than bankruptcy probabilities produced by models with the log-likelihood being maximized (Prob3 and Prob4). This is especially evident by the substantially higher pseudo-R^2 of models 1 and 2 (29.05% and 26.36% respectively) compared to models 3 and 4 (18.84% and 8.56% respectively) for the private firms case. Similarly, pseudo-R^2 of models 1 and 2 (34.53% and 33.77% respectively) is substantially higher than models 3 and 4 (19.74% and 14.07% respectively) in the case of public firms models.

In panel B, we use the Vuong (1989) test-statistic to test for differences in model log-likelihood values between various (non-nested) models. Results show that, in the case of private firms, the log-likelihoods of models 1 and 2 are significantly different than models 3 and 4 (test-statistics are 5.75 and 7.37 respectively) whereas Vuong test-statistics are 4.37 and 8.07 respectively for public firm models.

Overall, our results suggest that models with maximized AUROC provide probability estimates that contain significantly more information about bankruptcies over the next year compared to models which are trained to maximize traditional functions, such as the log-likelihood function, even when the increase in AUROC is relatively small (as in the case

of our public firm models as shown in Table 4). From all models, however, our proposed neural network model is the best performing one.

4.3 Economic performance results

So far, we have considered discriminatory power and information contest tests to assess model performance. However, a bank is generally interested in the economic benefits arising by using bankruptcy forecasting models in the decision-making process of granting loans to individual firms. Following Agarwal et al. (2008) and Bauer et al. (2014), we consider a loan market worth $100 billion and four banks are competing to grant loans to prospective firm customers. We hypothesize that banks 1 and 2 are more sophisticated banks, training models by maximizing AUROC (a neural network and a logistic model respectively) whereas banks 3 and 4 are more "naïve", training models by maximizing the log-likelihood function (a neural network and a logistic model respectively). In Table 6 we report the results, for both private and public firm models.

Clearly, banks 1 and 2 which use models with maximized AUROC, respectively, manage loan portfolios with higher quality relative to banks 3 and 4. This is evident by the lower concentration of bankruptcies they attract. In particular, the bankruptcy rate of bank's 1 portfolio, which uses a neural network trained to maximize AUROC is 0.047% and 0.074% when using the private and public firms model respectively. In contrast, the bankruptcy rate of bank's 3 portfolio, which uses a neural network trained to maximize the log-likelihood is 0.17% when using the private and public firms model respectively. Similarly, bank 2 which uses a logistic model trained to maximize AUROC manages a credit portfolio with bankruptcy rate equal to 0.38% and 0.066% when using the private and public firms model whereas for the bank using models to maximize the log-likelihood function (bank 4) the rates are 0.56% and 0.53% respectively.

Consequently, banks 1 and 2 achieve superior economic performance[11] compared to banks 3 and 4 respectively, on a risk-adjusted basis. For example, considering the private firms model, bank 1 which uses a neural network model with maximized AUROC, earns 2.24% relative to bank 3 which uses a traditional neural network model (1.52%). Also, bank 2 which uses a logistic model trained to maximize AUROC, earns 1.27% relative to bank 4 which uses a traditional logistic model (0.91%). In the case of public firms model, similar insights are obtained; bank 1 earns higher risk-adjusted returns relative to bank 3 (1.98% and 1.68% respectively). This is also the case for banks 2 and 4 (1.53% and 1.20% respectively). Again, the neural network trained to maximize AUROC, provides the higher economic benefits for the bank using it (bank 1).

4.4 Forecasting bankruptcy two years ahead

In this section, we evaluate and compare the performance of the models by increasing the forecasting horizon to two years. This is a more challenging problem because the characteristics of bankrupt firms are less pronounced relative to one year prior to bankruptcy and therefore is more difficult to forecast bankruptcy. Also, identifying the signs of the crisis

---

[11] Results are robust with respect to different specifications for LGD (0.4–0.7) and k (0.002–0.004), suggesting that models with maximized AUROC, outperform the traditional approaches.

Table 7 Forecasting bankruptcy two years ahead; Maximizing AUROC vs maximizing LL

Panel A: Performance, out-of-sample, 2007–2015

Model	Private Firms Model			Public Firms Model		
	AUROC	Info. Cont	Econ. Ben	AUROC	Info. Cont	Econ. Ben
Models with maximized AUROC						
Neural Network	0.8678	16.20%	1.36%	0.8864	18.27%	1.57%
Logistic	0.8503	13.83%	0.82%	0.8664	15.24%	0.56%
Models with maximized log-likelihood						
Neural Network	0.8441	13.66%	0.81%	0.8571	9.53%	0.90%
Logistic	0.8113	2.10%	0.50%	0.8558	4.33%	0.97%

Panel B: DeLong (1988) test statistic for differences in AUROC

	Private Firms Model	Public Firms Model
Neural Network with max. AUROC vs Neural Network with max. LL	2.18	3.17
Logistic model with max. AUROC vs Logistic model with max. LL	7.15	1.28

Panel C: Vuong (1989) test statistic for differences in information content

	Private Firms Model	Public Firms Model
Neural Network with max. AUROC vs Neural Network with max. LL	2.34	6.13
Logistic model with max. AUROC vs Logistic model with max. LL	7.15	7.33

This table reports out-of-sample performance in terms of discriminatory power, information content and economic impact, of models trained to maximize AUROC versus models trained to maximize the log-likelihood function, when forecasting bankruptcy two years ahead. Panel A reports the performance when financial variables are used in the models (private firms models) and when financial and market variables are used in the models (public firms models). Panel B reports DeLong (1988) test statistics for differences in AUROCs between the models of interest whereas panel C reports Vuong (1989) test statistics for differences information content between the models of interest

earlier, although difficult, can help the management of the firm to take remedial actions to correct the adverse situation and avoid bankruptcy. We perform the same analysis as before and we provide the results in Table 7.

As expected, performance has dropped relative to when forecasting bankruptcy one year ahead due to the increased difficulty of the problem. However, models which are trained to maximize AUROC outperform those trained to maximize the log-likelihood function. Starting from the AUROC, which is the focus of this paper to improve, we document out-of-sample that it is significantly higher, especially for the neural networks case. Two years prior to bankruptcy, the neural network with maximized AUROC achieves an AUROC equal to 0.8678 and 0.8864 for the private and public firms models respectively whereas for the neural network trained to maximize the log-likelihood function, these are 0.8441 and 0.8571 respectively. DeLong tests indicate that the differences are significant at the 5% and 1% level respectively (test statistics are 2.18 and 3.17 respectively). A logistic model with accounting data as predictors (private firms model) trained to maximize AUROC, has significantly higher discriminatory power than a competing logistic model trained to maximize the log-likelihood function (0.8503 vs 0.8113 respectively). Differences in AUROCs are statistically significant at the 1% level according to the DeLong test (test statistic is 3.54). For the public firms case, an enhancement is achieved (0.8664 vs 0.8558) but the difference is not statistically significant (DeLong test statistic is 1.28).

The summary of the remaining tests is that the neural network with maximized AUROC performs significantly better in terms of information content and provides higher economic benefits relative to a competing neural network which is trained by maximizing the log-likelihood function. The former is also the best performing model in all tests. Finally, a logistic model with maximized AUROC provides significantly more information than a logistic with maximized log-likelihood function, albeit no economic benefits are achieved in this case.

4.5 Forecasting financial distress

In this section, we change the event and instead of forecasting bankruptcy, we forecast financial distress. There are several reasons as to why firm stakeholders should be interested in models forecasting financial distress more accurately. First, financial distress is a state prior to bankruptcy filing and therefore forecasting the early signs of the crisis may help, for instance the management of the firm, to take corrective measures in order to avoid further deterioration that may ultimately lead to bankruptcy in which case the firm loses most of its value (see for instance Asquith et al. 1994; Glover 2016). Second, forecasting the early signs of the crisis is more challenging to accomplish not only because it is the starting point of the crisis but also because financial distress is not a formal event like bankruptcy, thus we need to construct a financial distress indicator. In this study, we follow Keasey et al. (2015) and Gupta et al. (2018) and we consider a firm as financially distressed if the following conditions are satisfied; 1) Earnings Before Interest, Tax and Depreciation and Amortization (EBITDA) is less than financial expenses (i.e. interest payments) for two consecutive years 2) Total Debt is higher than the Net Worth of the firm for two consecutive years and 3) The firm experiences negative Net Worth growth between two consecutive years. The firm is classified as financially distressed in the year immediately following these three events. For forecasting purposes, we use the data two years before financial distress. For example, when the conditions are satisfied for the years *t* and *t-1*, then the firm is considered as financially distressed in the year *t* and we construct the variables at *t-2* to

Table 8 Forecasting financial distress; Maximizing AUROC vs maximizing LL

Panel A: Performance, out-of-sample, 2007–2015

Model	Private firms model			Public firms model		
	AUROC	Info. Cont	Econ. Ben	AUROC	Info. Cont	Econ. Ben
Models with maximized AUROC						
Neural Network	0.9175	32.82%	1.14%	0.9000	28.21%	1.05%
Logistic	0.8982	26.84%	0.36%	0.8824	24.28%	0.56%
Models with maximized log-likelihood						
Neural Network	0.8956	25.89%	0.44%	0.8870	22.23%	0.97%
Logistic	0.8897	12.70%	-0.34%	0.8753	12.60%	0.28%

Panel B: DeLong (1988) test statistic for differences in AUROC

	Private Firms Model	Public Firms Model
Neural Network with max. AUROC vs Neural Network with max. LL	5.42	5.99
Logistic model with max. AUROC vs Logistic model with max. LL	3.23	2.24

Panel C: Vuong (1989) test statistic for differences in information content

	Private Firms Model	Public Firms Model
Neural Network with max. AUROC vs Neural Network with max. LL	7.62	7.32
Logistic model with max. AUROC vs Logistic model with max. LL	10.27	9.31

This table reports out-of-sample performance in terms of discriminatory power, information content and economic impact, of models trained to maximize AUROC versus models trained to maximize the log-likelihood function, when forecasting financial distress. Panel A reports the performance when financial variables are used in the models (private firms models) and when financial and market variables are used in the models (public firms models). Panel B reports DeLong (1988) test statistics for differences in AUROCs between the models of interest whereas panel C reports Vuong (1989) test statistics for differences information content between the models of interest

predict financial distress. Following these conditions, we generate an extensive database with 1,929 financially distressed firms. In Table 8 we report the out-of-sample results from this exercise.

Starting from the private firms model, logistic and neural network models trained to maximize AUROC, significantly outperform models trained to maximize the log-likelihood function (0.9175 vs 0.8956 for neural networks and 0.8982 vs 0.8897 for logistic models). Differences in AUROCs are statistically significant at the 1% level (DeLong test statistics are 5.42 and 3.29 for neural networks and logistic models respectively). Similar results are found with respect to the public firms model (0.9000 vs 0.8870 for neural networks and 0.8824 vs 0.8753 for logistic models). Differences in AUROCs are statistically significant at the 1% and 5% level respectively (DeLong test statistics are 5.99 and 2.24 for neural network and logistic models respectively).

Regarding the remaining tests, we find that the both models which are trained to maximize AUROC provide significantly more information and there is more gain by banks using them relative to using models trained to maximize the log-likelihood function. Overall, from this test we conclude that our methodology can help the interested parties to improve bankruptcy forecasts, considering the harder nature of the problem, either using a neural network or a logistic model. Once again, the neural network constructed to maximize AUROC is the best performing one.

4.6 Comparing our methodology with other methodologies

In this section we use the same analysis as before to compare our proposed methodology with other methods of AUROC maximization proposed in the bankruptcy literature and the advantages (shortcomings) of our method (other methods) are discussed.

We consider two other approaches proposed by Miura et al. (2010) and Kraus et al. (2014), to maximize AUROC of credit scoring models. Miura et al. (2010) suggest a sigmoid function as an approximation of Eq. (1). Specifically, they maximize the following objective function[12]:

$$F(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{1 + \exp\left[-d_{i,j}(\beta)/\sigma\right]} \tag{9}$$

where $d_{i,j}(\beta) = \beta^T \left(X_B^i - X_{NB}^j \right)$. However, unlike the function that we introduced previously, it treats all $d_{i,j}$'s in the same way, whereas our function, give more emphasis on the "bad" cases, for example, when a healthy firm has higher bankruptcy score than a bankrupt firm. Further, the authors consider only a linear response function (the output is a linear score) and unlike our method, it cannot be used by models which employ probabilistic response functions such as logistic models and highly non-linear models such as neural networks. Instead, our methodology works well with logistic and neural networks, which are the among the most popular bankruptcy models, because they allow for a probabilistic response function. Finally, Kraus et al. (2014) suggest using directly Eq. (1) as the objective function and implementing derivative-free methods (such as Nelder et al. 1965) to optimize the coefficients. The optimization algorithm that is used, however, assumes that

---

[12] The authors, in the original specification, set the tuning parameter $\sigma=0.01$ or 0.1. Here, we use $\sigma=1$ because the original specifications performed poorly. Further, they constrain the norm of coefficients to be 1. Again, we find that this specification performs poorly.

Table 9 Forecasting bankruptcy one year ahead; Our methodology vs alternative methodologies

Panel A: Performance, out-of-sample, 2007–2015

Model	Private firms model			Public firms model		
	AUROC	Info. Cont	Econ. Ben	AUROC	Info. Cont	Econ. Ben
Our methodology						
Neural Network	0.9332	29.05%	2.21%	0.9508	34.53%	1.92%
Alternative methodologies						
Miura et al. (2010)	0.9188	17.59%	1.55%	0.9471	26.76%	1.80%
KK (2014)	0.9136	8.57%	1.23%	0.9473	25.44%	1.64%

Panel B: DeLong (1988) test statistic for differences in AUROC

	Private Firms Model	Public Firms Model
Our methodology vs Miura et al. (2010)	1.55	1.15
Our methodology vs KK (2014)	1.99	1.18

Panel C: Vuong (1989) test statistic for differences in information content

	Private Firms Model	Public Firms Model
Our methodology vs Miura et al. (2010)	5.72	5.24
Our methodology vs KK (2014)	8.32	5.50

This table reports out-of-sample performance in terms of discriminatory power, information content and economic impact, using our methodology to maximize AUROC on a neural network versus alternative methods to maximize AUROC as proposed by Miura et al. (2010) and Kraus et al. (2014), denoted as KK (2014), when forecasting bankruptcy one year ahead. Panel A reports the performance when financial variables are used in the models (private firms models) and when financial and market variables are used in the models (public firms models). Panel B reports DeLong (1988) test statistics for differences in AUROCs between the models of interest whereas panel C reports Vuong (1989) test statistics for differences information content between the models of interest

Table 10 Forecasting bankruptcy two years ahead; Our methodology vs alternative methodologies

Panel A: Performance, out-of-sample, 2007–2015

Model	Private Firms Model			Public Firms Model		
	AUROC	Info. Cont	Econ. Ben	AUROC	Info. Cont	Econ. Ben
Our methodology						
Neural Network	0.8678	16.76%	1.42%	0.8864	18.92%	1.47%
Alternative methodologies						
Miura et al. (2010)	0.8479	10.97%	1.39%	0.8605	12.85%	0.97%
KK (2014)	0.8488	5.08%	0.87%	0.8595	7.47%	0.84%

Panel B: DeLong (1988) test statistic for differences in AUROC

	Private Firms Model	Public Firms Model
Our methodology vs Miura et al. (2010)	1.83	2.90
Our methodology vs KK (2014)	1.75	3.01

Panel C: Vuong (1989) test statistic for differences in information content

	Private Firms Model	Public Firms Model
Our methodology vs Miura et al. (2010)	3.63	5.02
Our methodology vs KK (2014)	6.69	7.27

This table reports out-of-sample performance in terms of discriminatory power, information content and economic impact, using our method to maximize AUROC on a neural network versus alternative methods to maximize AUROC as proposed by Miura et al. (2010) and Kraus and Kuchenhoff (2014), denoted as KK (2014), when forecasting bankruptcy two years ahead. Panel A reports the performance when financial variables are used in the models (private firms models) and when financial and market variables are used in the models (public firms models). Panel B reports DeLong (1988) test statistics for differences in AUROCs between the models of interest whereas panel C reports Vuong (1989) test statistics for differences information content between the models of interest

Table 11 Forecasting financial distress; Our methodology vs alternative methodologies

Panel A: Performance, out-of-sample, 2007–2015

Model	Private Firms Model			Public Firms Model		
	AUROC	Info. Cont	Econ. Ben	AUROC	Info. Cont	Econ. Ben
Our methodology						
Neural Network	0.9175	32.82%	1.12%	0.9000	28.21%	1.07%
Alternative methodologies						
Miura et al. (2010)	0.8982	3.97%	0.70%	0.8825	10.53%	0.71%
KK (2014)	0.8964	7.75%	0.47%	0.8827	9.89%	0.71%

Panel B: DeLong (1988) test statistic for differences in AUROC

	Private Firms Model	Public Firms Model
Our methodology vs Miura et al. (2010)	6.14	6.33
Our methodology vs KK (2014)	7.69	6.19

Panel C: Vuong (1989) test statistic for differences in information content

	Private Firms Model	Public Firms Model
Our methodology vs Miura et al. (2010)	15.55	13.93
Our methodology vs KK (2014)	14.39	12.65

This table reports out-of-sample performance in terms of discriminatory power, information content and economic impact, using our method to maximize AUROC on a neural network versus alternative methods to maximize AUROC as proposed by Miura et al. (2010) and Kraus and Kuchenhoff (2014), denoted as KK (2014), when forecasting financial distress. Panel A reports the performance when financial variables are used in the models (private firms models) and when financial and market variables are used in the models (public firms models). Panel B reports DeLong (1988) test statistics for differences in AUROCs between the models of interest whereas panel C reports Vuong (1989) test statistics for differences information content between the models of interest

the objective function is continuous, which is not the case for Eq. (1). Also, this approach while is easy to implement, ignores information provided by the gradient which could increase the accuracy of the coefficients after the optimization process and thus we believe that using specifications with differentiable functions is a better choice.[13] Table 9 presents the out-of-sample results by forecasting bankruptcy one year ahead (we use our neural network model which consistently outperformed the competing models).

Overall, we find that the neural network trained with our method has higher discriminatory power than the other models but the differences is AUROCs are not statistically significant. Despite that, the neural network provides significantly more information relative to Miura et al. (2010) and KK (2014) which is also economically beneficial.

When we increase the complexity of the problem, however, by forecasting bankruptcy two years ahead and forecasting financial distress, our neural network model consistently outperforms the competing methods and differences in performance are statistically significant. Results regarding forecasting bankruptcy two years ahead, reported in Table 10, show that the neural network model more accurately discriminates bankrupt from healthy firms

Table 12 Forecasting bankruptcy and financial distress using quarterly data

	Maximizing AUROC		Maximizing LL	
	Logistic	Neural Net	Logistic	Neural Net
Panel A: 1 quarter ahead				
Private Firm Model	0.9255	0.9201	0.9150	0.9111
Public Firm Model	0.9589	0.9611	0.9513	0.9472
Panel B: 4 quarters ahead				
Private Firm Model	0.9040	0.9186	0.8927	0.9050
Public Firm Model	0.9444	0.9456	0.9431	0.9300
Panel C: 8 quarters ahead				
Private Firm Model	0.8128	0.8280	0.7825	0.7918
Public Firm Model	0.8485	0.8523	0.8398	0.8258
Panel D: Financial Distress Case				
Private Firm Model	0.8648	0.8871	0.8558	0.8797
Public Firm Model	0.8566	0.8733	0.8544	0.8568

This table reports AUROC results for logistic and neural network models, when their coefficients are estimated by maximizing the AUROC (reported in the second and third columns respectively) and when their coefficients are estimated by maximizing the log-likelihood function (denoted as LL and reported in the fourth and fifth columns respectively)

The models are trained in the period 1990–2006 and the table reports the prediction performance, in terms of AUROC, in the out-of-sample period when predicting bankruptcy one quarter ahead (panel A), four quarters ahead (panel B), eight quarters ahead (panel C) but also financial distress (panel D). In the models, we incorporate either financial variables (Private Firms Model) or financial and market data (Public Firms Model) constructed using quarterly data

---

[13] We use the optimization toolbox in MATLAB. For Kraus et al. (2014) we use the *fminsearch* command while for Miura et al. (2010) we use the *fminunc* command with the trust-region algorithm.

two years prior to bankruptcy according to AUROCs (0.8678 vs 0.8479 for the Miura et al. 2010 and 0.8488 for KK 2014 in the private firm model case and 0.8864 vs 0.8605 and 0.8595 in the public firm model case). Differences are statistically significant at the 10% level using the private firms model and at the 1% level using the public firms model.

In the remaining tests, we document that the neural network provides significantly more information about future bankruptcies than the Miura et al. (2010) and KK (2014) methods and the better performance is associated with higher economic benefits for the bank which uses our proposed method.

Finally, in Table 11 we report our results when forecasting financial distress. Consistent with previous results, our method achieves significantly higher discriminatory power than the other methods (0.9175 vs 0.8982 for the Miura et al. 2010 and 0.8964 for KK 2014 in the private firm model case and 0.9000 vs 0.8825 and 0.8827 in the public firm model case). Differences are statistically significant at the 1% level.

In the remaining tests we confirm previous findings; our method provides significantly more information and the better performance overall is economically beneficial for the bank using our method as opposed to the competition.

4.7 Forecasting using quarterly data

Public firms issue financial information each quarter, thus investors can update their risk assessments more frequently as new information becomes available. In this section, we perform the same analysis as before but this time, the input variables to the models are constructed using quarterly data and we make predictions one, four and eight quarters ahead. Overall, the AUROC results reported in Table 12, are qualitatively similar with the results reported in the case where yearly data are used. More specifically, training the bankruptcy models with our proposed method, in all cases, improves the out-of-sample performance in terms of discriminatory power as opposed to training the models to maximize the traditional log-likelihood function.

5 Conclusions

The goal of this paper is to propose an alternative method to estimate the coefficients of bankruptcy forecasting models and specifically logistic and neural network models which are the most popular bankruptcy models used in prior research. In particular, we suggest those interested in forecasting bankruptcy, to obtain the coefficients by maximizing the discriminatory power as measured by the Area Under ROC curve (AUROC). In this study, a method is introduced and we highlight the benefits arising, out-of-sample, by using models which are trained to maximize AUROC over models trained with traditional methods, such as optimizing the log-likelihood function. Overall, we find that models trained to maximize AUROC outperform traditional methods, out-of-sample, in terms of discriminatory power, information content and economic impact. Our results hold when we test the method in different settings, such as forecasting bankruptcy one year ahead which is the most common horizon, forecasting bankruptcy two years ahead and forecasting financial distress which make forecasts more difficult (using yearly and quarterly data). Thus, forecasting bankruptcy accurately well in advance, which would

be beneficial for the firm to take corrective measures, requires a more sophisticated estimation method, such as maximizing the AUROC function by using our method. From all models, the neural network trained with our method is the best performing one.

Next, we compare our method with alternative methods proposed in the literature and we provide both theoretical as well as empirical justifications as to why our method should be preferred. As expected, the results are more pronounced when we increase the forecasting difficulty, such as forecasting financial distress.

Our results have implications to the way bankruptcy forecasting is performed. Our proposed estimation approach provides, to those interested to forecast bankruptcy, a significant advancement over traditional methods that can be used by logistic and neural networks for better bankruptcy analysis and possibly can be extended to areas such as in credit risk analysis.

Declarations

Conflict of interest he authors declare no conflicts of interest.

References

Afik Z, Arad O, Galil K (2016) Using Merton model for default prediction: an empirical assessment of selected alternatives. J Empir Financ 35:43–67

Agarwal V, Taffler R (2008) Comparing the performance of market-based and accounting-based bankruptcy prediction models. J Bank Financ 32(8):1541–1551

Altman E (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Financ 23(4):589–609

Altman E, Sabato G (2007) Modeling Credit Risk for SMEs: Evidence from the US Market. Abacus 43(3):332–357

Asquith P, Gertner R, Scharfstein D (1994) Anatomy of financial distress: An examination of junk-bond issuers. Quart J Econ 109:625–628

Basel Committee on Banking Supervision (2006) International convergence of capital measurement and capital standards: a revised framework

Bauer J, Agarwal V (2014) Are hazard models superior to traditional bankruptcy prediction approaches? a comprehensive test. J Bank Financ 40:432–442

Blochlinger A, Leippold M (2006) Economic benefit of powerful credit scoring. J Bank Financ 30:851–873

Campbell JY, Hilscher J, Szilagyi J (2008) In search of distress risk. J Financ 63(6):2899–2939

Charalambous C (1979) On conditions for optimality of the non-linear l1 problem. Math Program 17:123–135

Charalambous C, Christofides N, Constantinide ED, Martzoukos SH (2007) Implied non-recombining trees and calibration for the volatility smile. Quant Financ 7:459–472

Charalambous C, Martzoukos SH, Taoushianis Z (2020) Predicting corporate bankruptcy using the framework of Leland-Toft: evidence from US. Quant Financ 20:329–346

Charitou A, Dionysiou D, Lambertides N, Trigeorgis L (2013) Alternative bankruptcy prediction models using option-pricing theory. J Bank Financ 37(7):2329–2341

Chava S, Jarrow RA (2004) Bankruptcy prediction with industry effects. Rev Financ 8:537–569

DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biom 44:837–845

Dwyer DW, Kocagil AE, & Stein RM (2004) Moody's KMV RiskCalc v3.1 model. Moody's KMV.

Filipe SF, Grammatikos T, Michala D (2016) Forecasting distress in European SME portfolios. J Bank Financ 64:112–135

Fitzpatrick T, Mues C (2016) An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. Eur J Oper Res 249:427–439

Glover B (2016) The expected costs of default. J Financ Econ 119:284–299

Gupta J, Gregoriou A, Ebrahimi T (2018) Empirical comparison of hazard models in predicting SMEs failure. Quant Financ 18:437–466

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristics (ROC) curve. Radiol 143(1):29–36

Hillegeist SA, Keating EK, Cram DP, Lundstedt KG (2004) Assessing the probability of bankruptcy. Rev Financ Stud 9(1):5–34

Huber PJ (1967) The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions., 221–233. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability

Keasey K, Pindado J, Rodrigues L (2015) The determinants of the costs of financial distress in SMEs. Int Small Bus J 33:862–881

Keenan SC & Sobehart JR (1999) Performance measures for credit risk models. Moody's Risk Management Services.

Kraus A, Kuchenhoff H (2014) Credit scoring optimization using the area under the curve. J Risk Model Valid 8:31–67

Kumar PR, Ravi V (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques-A review. Eur J Oper Res 180:1–28

Lessmann S, Baesens B, Seow H-V, Thomas LC (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. Eur J Oper Res 247:124–136

Miura K, Yamashita S, Eguchi S (2010) Area under the curve maximization method in credit scoring. J Risk Model Valid 4:3–25

Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313

Ohlson JA (1980) Financial ratios and the probabilistic prediciton of bankruptcy. J Account Res 18(1):109–131

Papakyriakou P, Sakkas A, Taoushianis Z (2019) Financial firm bankruptcies, international stock markets and investor sentiment. Int J Financ Econ 24:461–473

Shumway T (2001) Forecasting bankruptcy more accurately: a simple hazard model. J Bus 74(1):101–124

Sobehart JR, Keenan SC, & Stein RM (2000) Benchmarking quantitative default risk models: A validation methodology. Moody's Investors Service.

Soberhart J, & Keenan S (2001) Measuring default accurately. Risk. 31–33.

Tayal A, Coleman TF, Li Y (2015) RankRC: Large-scale nonlinear rare class ranking. IEEE Trans Knowl Data Eng 27:3347–3359

Tinoco MH, Wilson N (2013) Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. Int Rev Financ Anal 30:394–419

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57(2):307–333

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48(4):817–838

Wu Y, Gaunt C, Gray S (2010) A comparison of alternative bankruptcy prediction models. J Contemp Account Econ 6(1):34–45

Zhang G, Hu MY, Patuwo BE, Indro DC (1999) Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. Eur J Oper Res 116:16–32

Zmijewski ME (1984) Methodological issues related to the estimation of financial distress prediction models. J Account Res 22:59–82