Quantitative Finance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

# A neuro-structural framework for bankruptcy prediction

**Christakis Charalambous, Spiros H. Martzoukos & Zenon Taoushianis**

*Quantitative Finance*, 2023
https://doi.org/10.1080/14697688.2023.2230241



# A neuro-structural framework for bankruptcy prediction

CHRISTAKIS CHARALAMBOUS†, SPIROS H. MARTZOUKOS‡ and ZENON TAOUSHIANIS §*

†Department of Business and Public Administration, School of Economics and Management, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus
‡Department of Accounting and Finance, School of Economics and Management, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus
§Department of Banking and Finance, Southampton Business School, University of Southampton, SO17 IBJ Southampton, UK

(*Received 6 September 2022; accepted 22 June 2023; published online 21 July 2023*)

We develop a framework to simultaneously compute the unobservable parameters underlying the structural-parametric models for bankruptcy prediction. More specifically, we compute the unobservable parameters such as, asset value and asset volatility, through learning by embedding in the structural models a neural network that maps the neural network's input space (e.g. companies' observable financial and market data) to the unobservable parameter space. Within such a 'neuro-structural' framework, the neural network and the structural model work together as a one unit during the learning phase by providing to each other forward and backward information, respectively, until the weights of the neural network are optimized according to a merit function. Empirical results show that structural models, like the Black-Scholes-Merton and the Down-and-Out option models, with parameters computed with our approach, perform better than alternative specifications of the structural models, out of sample, in terms of discriminatory power, information content and economic impact. Importantly, they also perform better than a standard neural network, suggesting that the co-joint dynamics between the neural network and the structural model are useful during the learning phase and can improve the prediction performance (and the training efficiency) of neural networks. Finally, our approach provides methodological (and empirical) enhancements over logit specifications such as, Campbell *et al.* [In search of distress risk. *J Finance*, 2008, **63**, 2899–2939]. There, financial and market data are the inputs, and the output is the probability of bankruptcy whereas our approach includes an intermediary step to obtain the unobservable parameters and subsequently the probability of bankruptcy.

*Keywords*: Parameters estimation; Bankruptcy prediction; Neuro-structural approach; Economic impact; Discriminatory power

*JEL Classifications*: G33, C45, C53, C61, D87

## 1. Introduction

### *1.1. Background and motivation*

In this paper, we develop a framework where an enhancement of the unobservable parameters underlying the structural-parametric models is obtained and used in the structural models for bankruptcy prediction. More specifically, we simultaneously compute the unobservable parameters of structural bankruptcy prediction models such as, asset value and asset volatility, through learning, by embedding in the structural models a neural network. In the innovative framework we propose, the neural network learns the unknown relationships between the observable inputs to the neural network and the unobservable parameters which are the outputs from the neural network (and subsequently the inputs to the structural model), while working co-jointly with the structural model during the learning phase in a forward and backward way, respectively, until the weights of the neural network are optimized. The resulting 'neuro-structural' model refers to the enhanced structural-parametric model, such as the Black-Scholes-Merton (BSM) and the Down-and-Out option models, supplied by the parameters computed within the proposed approach. Such models outperform alternative specifications

*Corresponding author. Email: z.taoushianis@soton.ac.uk

of the structural models (i.e. of the structural models which include parameters computed with alternative approaches), out of sample, in terms of discriminatory power, information content and economic impact. Importantly, the neuro-structural models outperform the standard neural network, suggesting that the co-joint dynamics between the neural network and the structural model add value during the learning phase as the neural network embeds knowledge from the structural model (and vice-versa).

While several studies have attempted, in the prior years, to compute the unobservable parameters of the structural models in the context of bankruptcy prediction, especially of the BSM, these are subject to certain well-known limitations. Hillegeist *et al.* (2004), Vassalou and Xing (2004), Campbell *et al.* (2008) and Charalambous *et al.* (2020) compute the unobservable parameters (asset value and/or asset volatility) by solving the BSM deterministic equations imposed by the options pricing framework. However, as argued by Crosbie and Bohn (2003), the BSM equations may hold instantaneously thus may be violated when market conditions change too fast, consequently yielding inaccurate parameter computations. In addition, as argued by Charitou *et al.* (2013) more recently, procedures typically used to compute the BSM equations may be subject to convergence errors, thus adding 'noise' during the estimations. Bharath and Shumway (2008) propose a naïve approach to compute the unobservable parameters, arguing that it is the functional form of the BSM equation that matters. However, the approach they use to compute the unobservable parameters is based on ad hoc deterministic relationships which may not hold when different samples are employed (Charitou *et al.* 2013). In the more recent years, Charitou *et al.* (2013) and Afik *et al.* (2016) have computed the unobservable parameters directly from historical equity data. However, by 'exogenously' computing the unobservable parameters using historical data, the structural model, which is the model under scrutiny, is barely involved in the estimation which may affect the quality of the estimations while the probability of bankruptcy may lack calibration.

The main contribution of this paper is the development of a more sophisticated, flexible, yet robust methodology to compute the unobservable parameters of structural bankruptcy prediction models which overcomes the limitations discussed above. To the extent that the parameter-estimation methodologies discussed above have specific functional forms, our approach estimates the unobservable parameters without the need to specify a-priori any functional form, thus offering flexibility to optimally estimate the unobservable parameters within our neuro-structural framework. We assume that the value of the unobservable parameters depends on several exogenous but observable variables that are elements of the vector $x$ (i.e. accounting and market variables), through some unknown relationships. We estimate these unknown relationships through learning, by embedding in the structural-parametric model a neural network that maps the exogenous input space (i.e. the exogenous observable variables of the vector x) to the unobservable parameter space, thus yielding a neuro-structural model for the estimation of the probability of bankruptcy. In Figure 1 we provide a sequence of our methodological framework. The inputs to the neural network are the exogenous variables in the vector $x$, and the outputs from the neural network are the unobservable parameter estimates which together with the observable parameters to the structural model are the inputs to the structural-parametric model.† For each set of weights of the neural network the estimated probability of bankruptcy for each company and the associated target value (which is equal to 1 (0) in the case the company goes bankrupt (survives)), enter the merit function (i.e. the log-likelihood). In this setting, in each iteration, the weights of the neural network are adjusted to provide improved unobservable parameter estimates, until the log-likelihood function is optimized. Moreover, since the neural network and the structural model are part of the same optimization structure (see Figure 1), they work in conjunction in each iteration to provide the optimal results. Figure 2 shows how the neural network and the structural model work together during the learning phase. In each iteration, the neural network provides forward information to the structural model (the unobservable parameter estimates, updated) while the structural model provides backward information to the neural network which is used to adjust the weights of the neural network. In this context, the learning process is enhanced because the neural network embeds knowledge from the structural model (and vice versa), highlighting the uniqueness of the proposed methodology.

The approach we propose in this paper, overcomes the limitations discussed above in the following ways. First, our methodology is flexible enough to optimally estimate the unknown relationships between the exogenous inputs, $x$, and the unobservable parameters. In this context, one does not need to make assumptions about the data structure or the dynamics of the unobservable parameters, for example, to impose any deterministic relationships or to specify functional forms, a priori, to compute them. Instead, by letting the unobservable parameters to depend on some exogenous inputs, $x$, through some unknown relationships, the neural network is optimized accordingly to freely learn the unknown relationships, providing in that way improved parameter values, while preserving the theoretical properties of the parametric model as they are part of the same structure (as Figure 1 shows). To this end, the neural network is, effectively, used as a nonparametric estimation tool for the unobservable parameters as it is used directly to estimate the underlying dynamics of the observable inputs and unobservable parameters, without specifying any functional forms. Such nonparametric estimation tools have been used extensively in options pricing (Ruf and Wang, 2020, Ruf and Wang 2021). In our paper, for the first time, we use neural networks as a nonparametric estimation tool for the unobservable parameters of structural bankruptcy prediction models.

Second, our approach enables both the neural network and the structural model to work together as a one unit during the learning phase thus the structural model is involved during the estimation. We show that such joint dynamics during the learning phase are important because the weights of the

† The set of unobservable parameters and the set of observable parameters are subsets of the full parameter set of the structural model. The former is estimated from the neural network while the latter passes directly to the structural model.

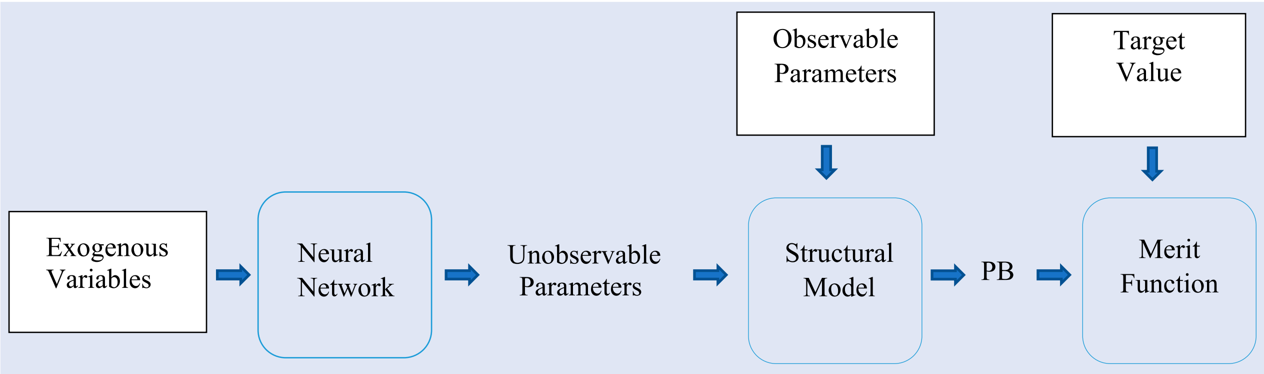


Figure 1. This figure shows the sequence of the proposed methodological framework. The exogenous (observable) variables in the vector x enter the neural network of which its outputs are the unobservable parameter estimates which together with the observable parameters enter the structural model, yielding a neuro-structural model for the estimation of the probability of bankruptcy. The probability of bankruptcy (PB) enters a merit function along with the target value of each company. In every iteration, the weights of the neural network are adjusted to provide improved unobservable parameter estimates, until the merit function is optimized.

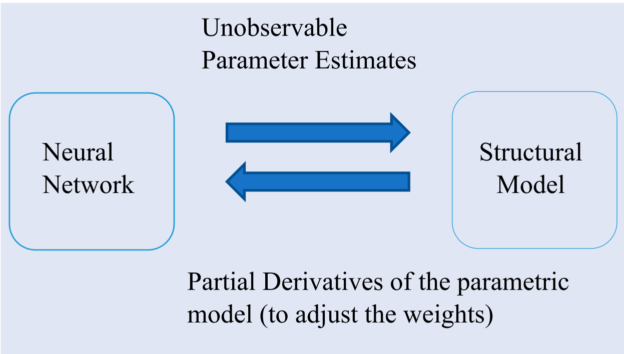


Figure 2. This figure shows how the neural network and the structural model work together as a one unit in each iteration to provide optimal results. Because they are part of the same optimization structure (see Figure 1), in each iteration, the neural network and the structural model pass forward and backward information to each other, respectively. Specifically, the neural network passes forward information to the structural model (the unobservable parameter estimates updated) while the structural model passes backward information to the neural network (partial derivatives of the structural model to adjust the weights of the neural network)

neural network are optimized by embedding knowledge from the structural model. Once the optimization of the weights is finished, (the outputs from) the neural network gives the estimated parameters (asset value and volatility) which enter the structural model, yielding a neuro-structural model for the estimation of the probability of bankruptcy which is an enhanced version of the traditional structural model.

The concept of using more sophisticated computational methods to estimate the parameters of parametric models is also explained by Aït-Sahalia and Lo (1998) and Aït-Sahalia and Duarte (2003) in the context of options pricing, arguing that economic theory imposes restrictions on the relationships between economic variables. This view is perfectly exhibited in the case of structural-parametric bankruptcy prediction models, such as the BSM which is a product of the restrictive assumptions underlying the options pricing theory where also key-structural parameters, for instance, the value of assets, depend on certain parametric assumptions which in reality are violated. Our proposed methodological framework, which is used for the first time in the context of bankruptcy prediction, overcomes these limitations as it is flexible enough to estimate the optimal relationships between the set of inputs to the neural network and the unobservable parameters of the structural-parametric bankruptcy prediction model, by adjusting the weights of the neural network until a merit function is optimized (see also Bandler *et al.* 1999, Andreou *et al.* 2010, Escanciano *et al.* 2016, Escanciano *et al.* 2017, Chernozhukov *et al.* 2020, Li *et al.* 2022 for further motivation on using semi-parametric models). Our methodological framework using neural networks is also justified by Gu *et al.* (2020). The authors aim to estimate the expected return as unknown function of some predictors and show that neural network is the best methodology to estimate the unknown function. Hence, our paper is not just an application of neural networks, but we use them as a mechanism to estimate the unobservable parameters of structural models, in the context of bankruptcy prediction, as unknown function of some observable variables.

We use accounting and market data between 1989 and 2014 for non-financial U.S. public firms to estimate the probability of bankruptcy using the various BSM specifications, over a one-year prediction horizon (for bankrupt firms between 1990 and 2015). For the estimation of the neuro-structural model, we divide the whole sample into two sub-samples; the first period includes bankruptcies in the period 1990 - 2006 and is used to train (and cross-validate) our proposed network structure. Once the model is optimized, we implement it to the out of sample dataset which includes bankruptcies in the period 2007–2015 and its performance is tested according to several criteria (see section 5 for model performance criteria).

### 1.2. *Main findings*

First, we compare the out of sample performance of our neuro-structural model with alternative specifications of the BSM parametric model; when asset value and volatility are estimated based on solving the BSM equations (i.e. Hillegeist *et al.* 2004, Vassalou and Xing 2004, Charalambous *et al.* 2020) and when estimated without solving the BSM equations (Bharath and Shumway 2008, Charitou *et al.* 2013). Specifically, we use three distinct types of tests, as suggested by Bauer and Agarwal (2014). In the first test, we compare the discriminatory power of the models based on the widely used Area Under Receiver Operating Characteristic curve (AUROC). Results indicate that the discriminating ability of the neuro-structural model is substantially better than the competing approaches. In the second test, we compare the information content of the neuro-structural model to alternative specifications of the BSM parametric model. Results show that bankruptcy probabilities produced by the neuro-structural model contain significantly more information than bankruptcy probabilities produced by the alternative BSM specifications. In the final test, we compare the economic impact arising when banks use the BSM with different specifications (one of which uses our approach) in the decision-making process of granting loans to individual firms. We find that the bank which uses the neuro-structural model earns superior risk-adjusted returns relative to the banks which use the alternative methodologies. Overall, results from our tests suggest that our approach yields more accurate asset values and volatilities which are reflected in the performance of the BSM model.

Second, several additional tests are conducted for robustness, including comparing the neuro-structural model against other widely used prediction methodologies such as logistic regression and standard neural networks. Our approach compares favorably to these methodologies. This finding is important because it suggests that the co-joint dynamics between the neural network and the structural model are useful during the learning phase and help to improve the prediction performance of the standard neural network.

Next, we apply our neuro-structural methodology on the Down-and-Out call option parametric model, which is an extension of the BSM model, as it allows bankruptcy to occur prior to the maturity of debt and used by other studies in the context of bankruptcy prediction, such as Brockman and Turtle (2003), Afik *et al.* (2016), among others. We find that the Down-and-Out neuro-structural model performs very well, which strongly supports the implementation of our methodology for default prediction.

We conclude by comparing the performance of our proposed neuro-structural model with the hybrid model used in Charalambous *et al.* (2020) and find that, our approach outperforms the hybrid model.

Our work has implications in the empirical application of bankruptcy prediction as it provides a more sophisticated†, yet more accurate and flexible method to compute the unobservable parameters of the structural-parametric models that improves their out of sample performance. Moreover, our findings have implications in the efficiency of the training of neural networks as their (bankruptcy) prediction performance can be improved when involving a theoretical model (i.e. the structural model) during the estimation of the neural network. As neural networks are frequently characterized as 'black boxes', our neuro-structural framework involves both the neural network and the structural (theoretical options pricing model) during the training phase thus adjusting the weights of the neural network by embedding knowledge from the structural model. Our work can also be viewed as a (neuro) extension of the powerful Campbell *et al.* (2008) model. In Campbell *et al.* (2008), the financial inputs enter the model to directly obtain the probability of bankruptcy. Our framework includes an intermediary step to obtain the unobservable structural parameters and then we supply the estimated parameters to the structural model to compute the probability of bankruptcy.

The remainder of the paper is organized as follows: Section 2 describes the alternative BSM specifications, which are used as benchmark; Section 3 describes our methodology to obtain improved parameter values for the parametric models; Section 4 discusses the data; Section 5 describes the three distinct type of tests we employ in order to test the performance of the models; Section 6 discusses the results, including robustness and finally, Section 7 concludes.

## 2. BSM model and estimation of asset value and volatility

### *2.1. Black-Scholes-Merton model*

Because the equity of the firm can be viewed as a European call option, the standard options pricing formula can be applied to value the equity of the firm as follows (see for example the seminal papers of Black and Scholes 1973 and Merton 1974):

$$E = VN\left(d_1\right) - Fe^{-rT} N\left(d_2\right), \tag{1}$$

where

$$d_1 = \frac{ln(^{V}\!/_{F}) + (r + 0.5\sigma_V^2)\mathrm{T}}{\sigma_V\sqrt{\mathrm{T}}}, \tag{2}$$

$$d_2 = d_1 - \sigma_V\sqrt{\mathrm{T}}. \tag{3}$$

Here, $V$ is the value of assets, $F$ the liabilities of the firm, $\sigma_V$ the volatility of assets value returns, $r$ is the riskless rate of return, *N(d)* is the standard normal distribution function and $T$ is the liabilities time to maturity. In equation (1), $N(d_2)$ represents the probability of solvency (i.e. the probability that the firm will not default on its liabilities). Therefore, the probability of bankruptcy is $1\text{-}N(d_2)$ or $N(-d_2)$. In the context of Black-Scholes-Merton, however, $N(-d_2)$ is the risk-neutral probability of bankruptcy, since $d_2$ is estimated using the riskless rate of return, $r$. We estimate the real-world probability of bankruptcy, by substituting $r$ with the real growth of assets, $\mu$. Hence, it is straightforward to show that the probability of bankruptcy, *PB*, is given by the following

† To the extent that the model is cross-validated and evaluated on the testing dataset, minimizes the risk of overfitting while ensuring its generalizability.

formula.†

$$PB = N(-d_2) = N\left(-\frac{ln(^V/_F) + (\mu - 0.5\sigma_V^2)\mathrm{T}}{\sigma_V\sqrt{\mathrm{T}}}\right). \quad (4)$$

The ratio inside of equation (4), gives the number of standard deviations the value of assets must drop in order the firm to default on its liabilities (i.e. how far the firm is away from default). For this reason, $d_2$ is commonly referred to as distance-to-default (*DD*).

However, the two most critical inputs in equation (4), $V$ and $\sigma_V$, are not observed in the market which makes the estimation of the probability of bankruptcy a challenging issue. Due to this, there was a burgeoning academic literature since the early 2000s regarding the estimation of $V$ and $\sigma_V$. We identify three main approaches for the estimation of these inputs, which we discuss in the following section.

### 2.2. Alternative approaches to estimate assets value and volatility

In this section, we present the various approaches used in the literature to estimate asset value and volatility, which we use as benchmark for our proposed approach.

**2.2.1. Two equations approach (2-Eqs. Approach).** One of the earliest and probably the most common estimation approach for $V$ and $\sigma_V$ was given by Jones *et al.* (1984) in the context of corporate debt valuation and by Ronn and Verma (1986) in the context of the empirical estimation of deposit insurance premiums. In the context of estimating the probability of bankruptcy, this approach has been used, for instance, by Hillegeist *et al.* (2004), Campbell *et al.* (2008) and Charalambous *et al.* (2020). In the framework of options pricing, there are two equations that can be solved iteratively to obtain the value of assets and the volatility. For the first equation, the standard options pricing formula given by equation (1) is solved with respect to $V$, yielding the following equation‡:

$$V = \frac{E + Fe^{-rT}N(d_2)}{N(d_1)}. \quad (5)$$

The second equation relates the (annualized) volatility of equity returns, $\sigma_E$, which is obtained from historical equity data, with the volatility of asset value returns, $\sigma_V$, through the equation $\sigma_E = \left(\frac{V}{E}\right)\frac{\partial E}{\partial V}\sigma_V$. Given that $\frac{\partial E}{\partial V} = N(d_1)$ and re-arranging the terms, $\sigma_V$ is calculated as follows:

$$\sigma_V = \frac{E\sigma_{\mathrm{E}}}{VN(d_1)}. \quad (6)$$

Starting from some initial values, for instance setting $V = E + F$ and $\sigma_V = \sigma_E$ on the RHS in equations (5) and (6), we obtain a new set of $V$ and $\sigma_V$ which are used in the next iteration in order to update the values of the two variables. The process is repeated until the changes of $V$ and $\sigma_V$ between two consecutive iterations are very small. When we obtain the two values, we can easily estimate $\mu$ as the return on asset values between two consecutive years i.e. $\ln(V_t/V_{t-1})$.

There are various studies acknowledging that simultaneously estimating the unobservable parameters, although theoretically sound, it is noisy practically (see for instance Crosbie and Bohn 2003, Charitou *et al.* 2013, etc.). Our approach estimates the unobservable parameters simultaneously but has two major differences. First, our approach estimates the unobservable parameters by maximizing the log-likelihood of the data. In other words, the values we obtain best fit the data, hence providing more accurate estimations. In such setting, the estimation of the unobservable parameters is optimal which is not the case for the 2-Eqs. Approach (thus they may add noise to the estimation procedure). Second, and probably the most important, our approach is more flexible because it does not impose specific functional forms on the unobservable parameters, so the neural network freely learns the unknown relationships between the inputs to the neural network and the unobservable parameters. Contrary, the 2-Eqs. Approach imposes specific functional forms on the unobservable parameters, according to equations (5) and (6), but such relationships may not hold in reality. We argue that the differences discussed above, between our approach and the standard approach, justify the (statistically significant) superior performance of our approach that we will document in the subsequent sections.

**2.2.2. Single equation approach (1-Eq. Approach).** A related approach with the 2-Eqs. Approach, is the 1-Eq. Approach used by Vassalou and Xing (2004) in their study on how firm default risk affects equity returns. In this case, given the observable daily time-series of equity for the entire year, we use equation (5) to obtain daily time-series for the value of assets.§ Once we obtain the time-series of asset values, we estimate the annualized volatility, $\sigma_V$, from the daily asset value returns. Using the new estimate of $\sigma_V$, in the next iteration we obtain a new series of asset values and returns, and we estimate a new value for $\sigma_V$. This process is repeated until the change in volatility is very small, i.e. 0.0001. This approach also requires setting initial values for $V$ and $\sigma_V$. We set $V = E + F$ and $\sigma_V = \sigma_E$.

Once we obtain the final daily series of *V's*, we calculate the annualized growth of assets, $\mu$, from the logarithmic changes of *V's*. The advantage of this approach is that it requires the solution of just one equation, possibly reducing convergence errors relative to using the two equations approach,

† In the context of bankruptcy prediction, Equations (1) and (4) come with variations. For instance, Hillegeist et al. (2004) include a dividend yield and use liabilities for $F$, while Vassalou and Xing (2004) do not include a dividend yield but use short-term debt plus half of long-term debt for $F$. For the purposes of our study, it is important to keep a common specification, like the standard formulas in Eqs. (1) and (4) and change only the methodology for asset value and volatility estimation in order to ensure that the source of improvement in model performance, comes from the methodology itself and not from the formula specification.

‡ We re-run by considering a dividend component as in Hillegeist et al. (2004), but we haven't found any differences in the performance.

§ We re-run by setting $F$ as short-term debt plus half of long-term debt as in Vassalou and Xing (2004). Again, we haven't found any differences in the performance.

but it is computationally intensive. Nevertheless, it still relies on convergence criteria that may affect the final outputs and consequently the accuracy of the probability of bankruptcy. Furthermore, it heavily depends on using the deterministic equation (5).

**2.2.3. Other estimation approaches.** Under this category, the estimation of $V$ and $\sigma_V$ is not based on iterative procedures but rather, the estimation relies on ad hoc approximations using observable data. Prominent among the studies that use such approximations is Bharath and Shumway (2008), which we denote as BS (2008). In their study, $V$ is approximated by the sum of market value of equity ($E$) plus the debt ($F$). Next, they calculate $\sigma_V$ as a weighted average of the volatility of equity and the volatility of debt:

$$\sigma_V = \frac{E}{E+F}\sigma_E + \frac{F}{E+F}\sigma_F, \tag{7}$$

where $\sigma_F = 0.05 + 0.25\sigma_E$. Finally, for the growth rate of $V$, they use the stock market return over the previous year ($\mu = r_{E,t-1}$). The authors show empirically that the BSM model performs better when $V$ and $\sigma_V$ are estimated with the simplified approximations, as opposed to solving the two BSM equations. They conclude that the accuracy stemming from BSM is due to its functional form and iterative procedures used to obtain $V$ and $\sigma_V$ are not useful.

In a similar notion, Charitou *et al.* (2013), which we denote as CDLT (2013), suggest the estimation of $V$ and $\sigma_V$ directly from equity data. In their study, they use the sum of equity and liabilities as an approximation of $V$. Using monthly equity data over the previous 60 months, they calculate a time-series of $V$'s from which the annualized return ($\mu$) and volatility ($\sigma_V$) are obtained. We slightly modify CDLT (2013) by estimating the variables using daily equity data over the prior year, in order to be consistent with the standards of our study, since we use equity data over a one-year period. CDLT (2013) demonstrate that such specifications improve the performance of the BSM model compared with the ad-hoc specifications of BS (2008).

Although the estimation approaches discussed in this section avoid the problems associated with solving the BSM equations, they are subject to some important drawbacks. First, they are still based on deterministic relationships (i.e. Bharath and Shumway 2008) which may not hold when different samples are employed and second, they barely involve the model under scrutiny into the estimation process (Charitou *et al.* 2013).

In the following section we introduce a neural network-based methodology to obtain improved parameter values, avoiding in that way the limitations underlying the alternative estimation techniques described above.

## 3. Methodology: a neuro-structural framework

### *3.1. The general case*

In this section we demonstrate how our methodology works to any parametric-structural model and later, we implement our method to the widely used BSM parametric model, but we also consider an extension and specifically, the Down-and-Out option model.

Consider that we have a parametric model, $f_{PM}$, which requires the parameters $p$ to estimate the probability of bankruptcy, PB:

$$PB = f_{PM}(p), \tag{8}$$

where $p = [p_1, p_2, \ldots, p_L]$ is the $L$ dimensional vector with the $L$ parameters of the model and $f_{PM}$ refers to the functional form of the parametric model. Suppose that some parameters of $f_{PM}$, say $M$, where $M \leq L$, are not observable and thus:

$$PB = f_{PM}(p^-, p^+). \tag{9}$$

In equation (9), $p^- = [p_1^-, p_2^-, \ldots, p_M^-]$ is the vector which corresponds to the unobservable parameters, and $p^+ = [p_{M+1}^+, p_{M+2}^+, \ldots, p_L^+]$ is the vector which corresponds to the observable parameters. Note that the vector $p$ consists of the two subsets $p^-$ and $p^+$.

Suppose now that the unobservable parameters, $p^-$, depend on some exogenous variables that are elements of the vector $x$, through some unknown relationships:

$$\begin{aligned} p_1^- &= f_1(x), \\ &\vdots \\ p_M^- &= f_M(x), \end{aligned} \tag{10}$$

where $f_i(x)$ is some unknown function of $p_i^-$ with respect to the exogenous vector of variables, $x$, which we aim to estimate through learning, using neural networks, for $i = 1,2,\ldots,M$. In this context, the probability of bankruptcy is estimated as:

$$PB = f_{PM}(z, p^+), \tag{11}$$

where $z = [f_1(x), f_2(x), \ldots, f_M(x)]$ refers to the vector with the variables that are determined through the estimation of the unknown functions, using neural networks. In fact, the vector $z$ provides improved parameter values to the model in equation (11), which we refer to as neuro-structural model. Figure 3 provides a schematic representation of the proposed approach.

As can be seen from the figure the probability of bankruptcy is estimated using the functional form of the parametric model but using two sets of inputs: (1) the inputs that are observable and enter directly to the parametric model, $p^+$, and (2) the variables $z$, which depend on the exogenous variables $x$ through some unknown relationships that we aim to estimate using neural networks (i.e. $x$ are the inputs to the neural network that produce the outputs $z$). In this context, $z$ and consequently $PB$, depend on the weights of the neural network, $w$, imposed by the neural network. The next step is to estimate the weights by training the model. Consider that we have $N$ 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 prediction, the input sample $x_n$ can be information characterizing the $n$-*th* firm, such as financial or market information, whereas $t_n$ is an indicator variable which

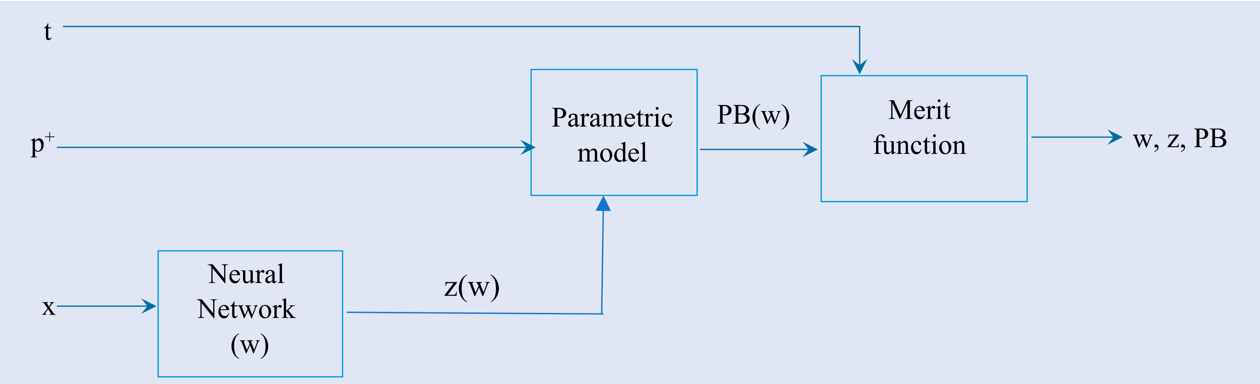


Figure 3. Schematic representation of our approach. Improved parameter values, z, are obtained in each iteration from the neural network and enter as inputs to the parametric model along with other parameters, $p^+$, that enter directly, yielding a neuro-structural model. Here, $x$, represents the vector of some exogenous inputs to the neural network. The proposed structure is optimized according to a merit function, to give the weights, w, of the neural network and subsequently the unobservable parameter estimates, $z$, and finally the probability of bankruptcy, PB. Note that in the merit function, the targets, $t$, are supplemented directly. In our case, $t = 1$ if the firm goes bankrupt and $t = 0$ otherwise and the merit function is the log-likelihood function.

is equal to 1 if the corresponding firm-observation is bankrupt and 0 otherwise. The output of the parametric model, *PB(w),* with the associated targets, $t$, are used in the merit function which is optimized in order to obtain the weights, $w$, of the neural network and consequently the final output, which is the probability of bankruptcy, *PB*. The neural network here serves as an auxiliary mechanism which adjusts the parameters of the parametric model during the training phase, until the weights of the neural network are optimized according to a merit function. As can be seen from the figure, both the neural network and parametric models belong to the same structure and work in conjunction. This is important because in this setting, the neural network embeds knowledge from the structural-parametric model which is useful during the training phase. This implicit knowledge from the parametric model reduces the complexity of the neural network and improves its generalization performance.

In this study, we exploit the strong learning capabilities of feedforward neural networks, which is the most commonly used neural network architecture and it has been widely used to approximate any unknown function. Cybenko (1989) proved that a feedforward neural network with a single hidden layer with enough neurons in the hidden layer, with monotonic increasing activation functions and linear outputs, can approximate any continuous function to any degree of accuracy. Similarly, Hornik *et al.* (1989) concludes that such network architectures are universal function approximators. In addition, from an empirical point of view, neural networks have been used extensively as nonparametric estimation tools in the context of options pricing (refer to Ruf and Wang 2020 for a comprehensive review). Therefore, neural networks are an appropriate methodology for our framework.†

A typical feedforward neural network is a system with interconnected units (neurons) organized into layers where information flow from the previous layers to the next layers, aiming to learn the unknown relationships between the inputs and outputs. The first layer in our network, presented in Figure 4, consists of $H$ units, with the *i-th* unit connected with the input features, $x$, through the K-dimensional weight vector $w_i^{(1)} = [w_{i1}^{(1)} \dots w_{iK}^{(1)}]$ and the biases $w_{i0}^{(1)}$. The *i-th* unit produces a weighted sum, $\psi_i^{(1)}$, which enters an activation function, $f_i^{(1)}$, to produce an output, $y_i^{(1)}$, where $i = 1,2,\dots,H$. The outputs from the first layer are forwarded to the second layer which is consisted with $M$ units, corresponding to the outputs of the network. The *j-th* unit in this layer is connected with the outputs from the previous layer through the H-dimensional weight vector $w_j^{(2)} = [w_{j1}^{(2)} \dots w_{jH}^{(2)}]$ and the biases $w_{j0}^{(2)}$. The *j-th* unit produces a weighted sum, $\psi_j^{(2)}$, which enters an activation function, $f_j^{(2)}$, to produce the final output, $y_j^{(2)}$, with $j = 1,2,\dots,M$.

The set of equations below shows the explicit derivation of the outputs from the neural network, $y_1^{(2)} \dots y_M^{(2)}$:

$$\begin{aligned} y_1^{(2)} &= f_1^{(2)}\left[w_{10}^{(2)} + \sum_{i=1}^{H} w_{1i}^{(2)} f_i^{(1)}\left(w_{i0}^{(1)} + \sum_{j=1}^{K} w_{ij}^{(1)} x_j\right)\right], \\ &\vdots \\ y_M^{(2)} &= f_M^{(2)}\left[w_{M0}^{(2)} + \sum_{i=1}^{H} w_{Mi}^{(2)} f_i^{(1)}\left(w_{i0}^{(1)} + \sum_{j=1}^{K} w_{ij}^{(1)} x_j\right)\right]. \end{aligned} \tag{12}$$

The RHS of the set of equations above, are the generalized (unknown) functions that we seek to estimate by optimizing the weights of the network, according to a merit function. The LHS of the set of equations, correspond to the improved parameters that enter the parametric model, yielding the neuro-structural model.

Overall, there are several advantages by using our proposed approach. First, we do not need to impose any ad-hoc or deterministic relationships for the parameters of the parametric model. Instead, by treating (some of) the parameters as generalized unknown functions, the network structure optimizes the weights accordingly, to determine the relationships between the observable input features to the neural network and the parameters under consideration, yielding improved parameters that enter the parametric model. Second, we utilize the

† In a subsequent section, we compare the neuro-structural model with the standard neural network approach.

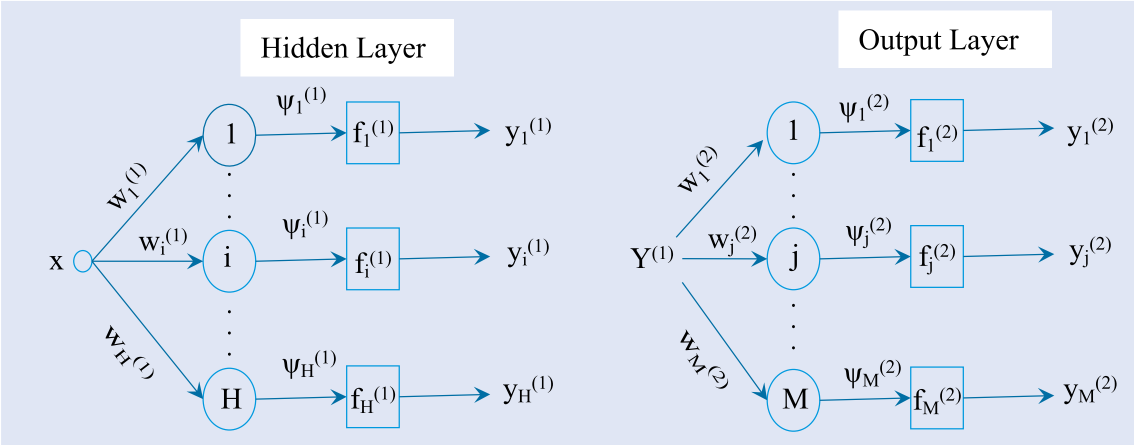


Figure 4. General structure of a two-layer feedforward neural network, with H neurons in the hidden layer and M neurons in the output layer.

strong learning capabilities of neural networks while preserving the theoretical properties of the parametric model. That is, the probability of bankruptcy is estimated using the underlying theory of the parametric model, while the neural network embeds knowledge from the parametric model which is useful during the training phase.

Once the network structure is optimized in the training phase and we obtain the weights, the out of sample probability of bankruptcy, which allows testing the performance of our method, is simply calculated as follows:

Step 1: Use the out of sample vector $x$ as input to the neural network. The outputs from the neural network correspond to the parameters of the parametric model (for instance asset value and volatility).

Step 2: Use the outputs of the neural network as inputs to the parametric model (i.e, the neuro-structural model) along with other inputs which enter directly, to obtain the out of sample probability of bankruptcy.

### 3.2. *The case of BSM Model*

In this section, we show how our method works to the case of BSM and in a later section, we consider the Down-and-Out option model but generally, the method applies to any parametric model.

First, it would be useful to rewrite equation (4) as follows:

$$PB = N(-DD) = N\left(-\frac{ln\left(\frac{Ve^{\mu \mathrm{T}}}{F}\right) - 0.5\sigma_V^2 \mathrm{T}}{\sigma_V \sqrt{\mathrm{T}}}\right). \tag{13}$$

Note that the numerator inside the logarithm in equation (13), $Ve^{\mu \mathrm{T}}$, is the expected value of assets which when scaled by the liabilities of the firm, $F$, gives the expected leverage, denoted by $E_L$. Thus, the probability of bankruptcy is given by the following formula:

$$PB = N(-DD) = N\left(-\frac{ln(E_L) - 0.5\sigma_V^2 \mathrm{T}}{\sigma_V \sqrt{\mathrm{T}}}\right). \tag{14}$$

Consider now that there are two outputs from the neural network; the expected value of assets (divided by liabilities, for scaling considerations), $y_1^{(2)} = E_L(w)$, and the volatility of asset value, $y_2^{(2)} = \sigma_V(w)$:

$$E_L(w) = f_L(x, w), \tag{15}$$

$$\sigma_V(w) = f_\sigma(x, w). \tag{16}$$

The RHS of equations (15) and (16) are the generalized (unknown) functions between the neural network input features, $x$, and $E_L$ and $\sigma_V$, that the neural network seeks to learn by optimizing the weights of the network structure. These two outputs enter as inputs to the BSM model and thus obtaining the probability of bankruptcy:

$$PB(w) = N[-DD(w)] = N\left(-\frac{\ln[E_L(w)] - 0.5\sigma_V^2(w)\mathrm{T}}{\sigma_V(w)\sqrt{\mathrm{T}}}\right). \tag{17}$$

Notice that the difference between equations (14) and (17) is that the latter depends on the weights imposed to $E_L$ and $\sigma_V$ through the neural network and consequently, the probability of bankruptcy, $PB$, is a function of the weights, yielding a neuro-structural model. For a sample of $N$ observations, the weights of the neural network are obtained by maximizing the log-likelihood function, LL, defined as follows:

$$LL(w) = \sum_{n=1}^{N} l_n(w), \tag{18}$$

where

$$l_n(w) = t_n \ln[PB_n(w)] + (1 - t_n)\ln[1 - PB_n(w)]. \tag{19}$$

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 weights of the neural network. In each iteration, the optimization algorithm updates the weights according to the partial derivatives that we provide. The gradient vector of $l_n(w)$ with respect to the weights is given by†

† See Charalambous (1992) for the efficient training of neural networks.

(for simplicity we drop the subscript $n$):

$$\frac{\partial l(w)}{\partial w} = c(w)\frac{\partial PB(w)}{\partial w}, \tag{20}$$

where $c(w) = \frac{t-PB(w)}{PB(w)[1-PB(w)]}$ and

$$\frac{\partial PB(w)}{\partial w} = \sum_{j=1}^{M} \frac{\partial PB(Y^{(2)})}{\partial y_j^{(2)}} \frac{\partial y_j^{(2)}}{\partial w}. \tag{21}$$

The quantity $\frac{\partial PB(Y^{(2)})}{\partial y_j^{(2)}} \equiv \frac{\partial f_{PM}}{\partial p_j}$ represents the partial derivative of the parametric model with respect to the *j-th* output of the neural network (i.e. the input to the parametric model) and $\frac{\partial y_j^{(2)}}{\partial w}$ represents the partial derivative of the *j-th* output with respect to the weights.

When $w \equiv w_j^{(2)}$,

$$\frac{\partial PB(w)}{\partial w_j^{(2)}} = \delta_j^{(2)} Y^{(1)}, j = 1, 2, \ldots, M, \tag{22}$$

where $\delta_j^{(2)} = \frac{\partial PB(Y^{(2)})}{\partial y_j^{(2)}} f_j^{(2)'}(\psi_j^{(2)})$. Here, the term $f_j^{(2)'}(\psi_j^{(2)})$ is the partial derivative of the activation function of the *j-th* output, valued at $\psi_j^{(2)}$.

When $w \equiv w_i^{(1)}$,

$$\frac{\partial PB(w)}{\partial w_i^{(1)}} = \delta_i^{(1)} x, i = 1, 2, \ldots, H, \tag{23}$$

where $\delta_i^{(1)} = pd_i^{(1)} f_i^{(1)'}(\psi_i^{(1)})$ and $pd_i^{(1)} = \sum_{j=1}^{M} w_{ji}^{(2)} \delta_j^{(2)}$. Here, $f_i^{(1)'}(\psi_i^{(1)})$ is the partial derivative of the activation function of the *i-th* output from the first layer, valued at $\psi_i^{(1)}$.

Once the weights of the neural network are optimized in the training sample, the out of sample probability of bankruptcy can be computed; 1) Use the out of sample vector $x$ as input to the neural network. The outputs of the neural network are the $E_L$ and $\sigma_V$ and 2) use the outputs of the neural network, $E_L$ and $\sigma_V$, as inputs to the BSM to obtain the out of sample probability of bankruptcy.

We also provide a representation of how the neural network is modified according to the partial derivatives:

$\delta_j^{(l)}, l = 1, 2$, can be considered as the propagation of the partial derivatives of PB w.r.t. the unobservable parameters back to the input of the *j-th* neuron of the *l-th* layer. equations (22) and (23) can be put in words as follows:

$$\begin{bmatrix} \textit{Partial derivatives of PB} \\ \textit{w.r.t the input weight vector} \\ \textit{to a given neuron} \end{bmatrix} = \begin{bmatrix} \textit{Propagation of the} \\ \textit{partial derivatives of PB w.r.t.} \\ \textit{unobservable parameters} \\ \textit{back to the input to that unit} \end{bmatrix} * \begin{bmatrix} \textit{Input vector} \\ \textit{to that unit} \end{bmatrix}.$$

The given neural network can be used to compute the values of the backpropagated partial derivatives but with the following modifications. Figure 5 shows the modified neural network (MNN).

(a) The output of a given neuron will become the input of the corresponding neuron in the MNN.
(b) (i) Each activation function is changed to multiplier with value equal to the partial derivative of the activation function w.r.t. the input,
 (ii) The position of the multiplier is interchanged with that of the adder.

During the learning phase, the neural network passes the input information to the structural model in a forward way to obtain the PB, while the structural model passes the output information (the partial derivatives of PB w.r.t. the unobservable parameters) to the neural network in a backwards way to obtain the partial derivatives of PB w.r.t. the weights that will be used to change the weights.

### 3.3. *Specifications of the neural network*

Several features of the neural network need to be specified such as the input variables, the number of neurons used in the hidden layer, as well as the activation functions in the input and output layers.

First, note that the bankruptcy process in the BSM model is based on the future distribution of assets value i.e. the expected value of assets and the volatility of asset value returns. We aim to forecast the future distribution by using data that captures the current performance of the firm. With respect to that, prior studies have identified firm-specific characteristics related to the bankruptcy process of the firm (see for instance Altman 1968, Ohlson 1980, Almamy *et al.* 2016, etc.). We use data from a more comprehensive model. In particular, Campbell *et al.* (2008) find that several accounting-based and market-based variables are significant predictors of bankruptcy. We use the variables of their study as inputs to the neural network that might affect the outputs. It should be noted that, the inputs include information about the leverage of the firm (liabilities divided by assets) and equity return data which might have an association with the expected value of assets divided by liabilities (i.e. market leverage), $E_L$, but also, it includes the volatility of equity, which might have an association with the volatility of assets, $\sigma_V$. Thus, using the variables from Campbell *et al.* (2008) as inputs to the neural network is a reasonable choice. In this setting, our work becomes a neuro-extension of Campbell *et al.* (2008); there, the authors use the variables to directly obtain the probability of bankruptcy. In our work, we use their variables to obtain improved option parameters (asset value and volatility) in an intermediary step which are the outputs from the neural networks, and then are used as inputs to the structural-parametric model to obtain the probability of bankruptcy.

The selection of the optimal number of neurons is done empirically, based on a validation process-a straightforward and easy to implement approach, which makes use only the in-sample data to determine the optimal number of neurons to minimize the risk of overfitting while we leave the testing sample intact to assess the generalizability of the model (see

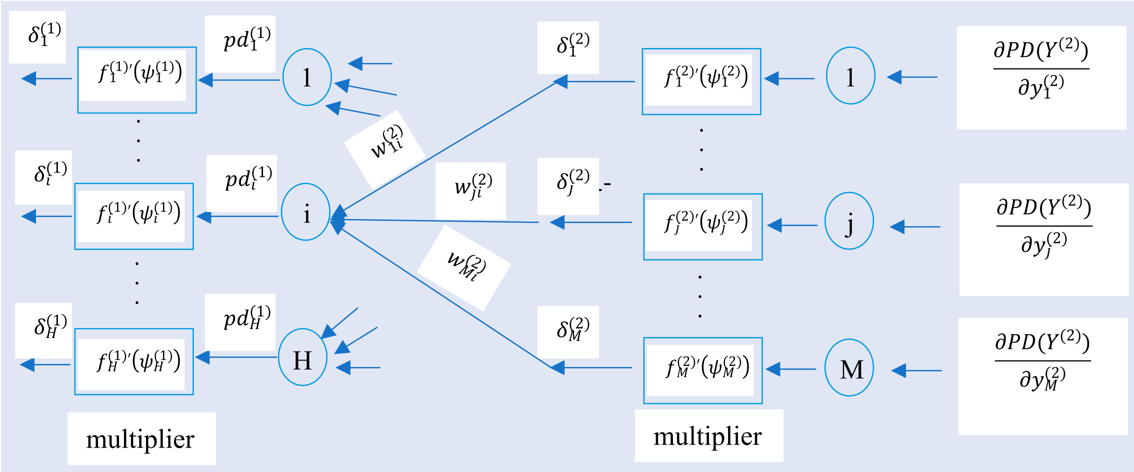


Figure 5. Schematic representation of the modified neural network which shows how the partial derivatives are transmitted from the structural model to the neural network in a backwards way during the learning face to change the weights of the neural network.

Andreou *et al.* 2010). Initially, we divide the whole sample into two sets: the training set (70%) and the testing set (30%). We further divide the training sample into training and validation while we leave the testing set intact to be used once the model is optimized. Using this training set, we estimate the network structure using one to five neurons. The optimal number of neurons is the one which performs the best on the validation set, according to AUROC. This process is repeated 20 times for each number of neurons in order to account for different initialization points (i.e. we train the model 100 times and select the one which performs the best in the validation set). Then we use the whole training set to estimate the network, using the optimal number of neurons and as starting point, we use the weights of the model that performed the best on the validation set. We find that three neurons perform the best in this setting ($H = 3$).

As for the activation functions, the hyperbolic tangent sigmoid function is used in the hidden layer, $f_H(.) = \frac{1-\exp(-2\psi_i^{(1)})}{1+\exp(-2\psi_i^{(1)})}$, which bounds the outputs from the hidden layer between $[-1, 1]$. A challenging task is the format of the transfer functions to be used in the output layer. This is because, $E_L$ and $\sigma_V$ must be non-negative and within reasonable values. In this case, we use a modification of the log-sigmoid function as follows; $f_M(.) = a + \frac{b-a}{(1+\exp[-\psi_j^{(2)}])}$, which bounds the outputs in the range [a, b]. In our case, j = 1,2, represent the two outputs; the expected value of assets (scaled by liabilities) and the volatility of assets. When $E_L$ is to be estimated, a = min [(E + F)/F] and b = max [(E + F)/F]. When $\sigma_V$ is to be estimated, a = min ($\sigma_E$) and b = max ($\sigma_E$). Notice that when $\psi_j^{(2)} \rightarrow \infty$, then $E_L$ and $\sigma_V \rightarrow b$. When $\psi_j^{(2)} \rightarrow -\infty$, then $E_L$ and $\sigma_V \rightarrow a$.

### *3.4. Summarizing our approach*

In this section, we summarize our approach and explain how it improves the BSM. As it is shown in Figure 3, there is a connection between the neural network and the structural model and works iteratively in the following way:

**3.4.1. Forward information.** For a set of weights (w) of the neural network, and using financial inputs to the neural network, we obtain the output of the neural network, z, which are the estimated asset value and volatilities. These enter the BSM model (parametric model in the Figure 3) along with p + which are the inputs observable to the BSM (i.e. firm liabilities). The parametric model gives the probability of bankruptcy, PB, which along with the target values (1 or 0, depending on the status of the firm), enter the merit function (the log-likelihood function) which we seek to maximize.

**3.4.2. Backward information.** The coefficients of the neural network are updated according to the partial derivatives we provide, as shown through equations (20)–(23) and Figure 5. For example, equation (22) includes the quantity $\frac{\partial f_{PM}}{\partial p_j}$, which is the partial derivative of the parametric model with respect to the j-th unobservable parameter (i.e. asset value or volatility). This quantity is very important for updating the weights of the neural network because it incorporates the theoretical properties of the parametric model.

The above forward and backward process works iteratively until the log-likelihood function is maximized. In every iteration, the weights of the neural networks are updated in order to provide more accurate asset values and volatilities.

As one can clearly see in the above estimation process that we propose for the estimation of asset values and volatilities, there is no restriction on their functional form as it happens with the BSM. Instead, the neural network is updated in every iteration to freely learn the relationships between the financial inputs to the neural network and the outputs (i.e. asset values and volatilities). In addition, the estimation of the asset values and volatilities, in our setting, is optimal as the log-likelihood of the data is optimized as opposed to the case of BSM where their estimation is sub-optimal (no function is optimized to get the asset value and volatility). Due to such improvements, we will later report a significant improvement in the empirical performance of the BSM when we estimate asset values and volatilities within our neuro-structural approach but also from the standard neural networks as the latter includes no knowledge from the theoretical (structural) model.

## 4. Data

### *4.1. Sample*

In the literature there are several different definitions for negative events of firms. For example, Pindado *et al.* (2008) define a firm to be in financial distress when certain financial criteria are met (i.e. profitability is lower than financial expenses for two consecutive years, the market value falls for two consecutive years, etc.). However, we argue that this definition is quite subjective. Other studies, for instance Bharath and Shumway (2008), use corporate defaults. Default is a state where the firm misses an interest (or principal) payment for more than 90 days. Such information can be found from publicly available sources, such as from Moody's reports. However, a standard database is non-existent in general for default events and such events are relatively small (Galil and Gilat 2019). In this paper, we consider firms that file for bankruptcy under Chapter 7 (liquidation) or Chapter 11 (re-organization) in the U.S. The bankruptcy state usually succeeds financial distress and default events and is considered to be more serious than the other two. Our sample of bankrupt firms consists of 420 non-financial U.S. public firms that filed for bankruptcy over the 26-year period 1990–2015 and have all data available in Compustat and CRSP one year prior to bankruptcy. Bankrupt firms are sourced from the database BankruptcyData. The final sample contains about 94,000 bankrupt and healthy firm-year observations. The distribution of observations across the years is shown in Table 1.

Table 1. Distribution of observations.

| Year | Bankrupt firms | Healthy firms | Bankruptcy rate (%) |
|---|---|---|---|
| 1990 | 22 | 3292 | 0.66 |
| 1991 | 25 | 3241 | 0.77 |
| 1992 | 17 | 3258 | 0.52 |
| 1993 | 20 | 3318 | 0.60 |
| 1994 | 10 | 3543 | 0.28 |
| 1995 | 14 | 3861 | 0.36 |
| 1996 | 14 | 4138 | 0.34 |
| 1997 | 13 | 4379 | 0.30 |
| 1998 | 19 | 4698 | 0.40 |
| 1999 | 26 | 4664 | 0.55 |
| 2000 | 20 | 4435 | 0.45 |
| 2001 | 21 | 4286 | 0.49 |
| 2002 | 14 | 4182 | 0.33 |
| 2003 | 15 | 3913 | 0.38 |
| 2004 | 13 | 3601 | 0.36 |
| 2005 | 14 | 3510 | 0.40 |
| 2006 | 10 | 3503 | 0.28 |
| 2007 | 14 | 3439 | 0.41 |
| 2008 | 20 | 3320 | 0.60 |
| 2009 | 31 | 3244 | 0.95 |
| 2010 | 6 | 3153 | 0.19 |
| 2011 | 9 | 3037 | 0.30 |
| 2012 | 12 | 2963 | 0.40 |
| 2013 | 12 | 2920 | 0.41 |
| 2014 | 12 | 2884 | 0.41 |
| 2015 | 17 | 2897 | 0.58 |
| **Total** | **420** | **93679** | |

Note: This table shows the distribution of bankrupt and healthy-firm observations across the sample period 1990–2015 and the annual bankruptcy rate, defined as the annual number of bankruptcies divided by the annual number of observations.

### *4.2. Variables construction*

To construct asset value, $V$, and the volatility, $\sigma_V$, for the alternative approaches described in section 2, we obtain data from three sources. From Compustat, we get total liabilities and from CRSP we get daily equity prices and shares outstanding to calculate; the equity value of the firm, $E$, as the closing stock price $*$ shares outstanding and the annualized volatility of daily equity returns, $\sigma_E$, over a one-year period. Using daily equity prices, we also calculate the annualized equity return, $r_{E,t-1}$, which is used in BS (2008) as proxy for assets growth, $\mu$. Finally, for the risk-free rate we use the 1-year Treasury bill rate, obtained from Federal Reserve Board Statistics.

Regarding the variables from Campbell *et al.* (2008) which we use as inputs to the neural network, we further get financial information from Compustat such as net income, cash and short-term investments and shareholders' equity value, to construct the following ratios; total liabilities divided by equity market value + total liabilities (*TLMTA*), net income divided by equity market value + total liabilities (*NIMTA)*, cash and short-term investments divided by equity market value + total liabilities (*CASHMTA*) and shareholders' equity value divided by equity market value i.e. book-to-market ratio (*BM*). Other variables used are the following; annualized volatility of stock returns (SIGMA) using daily returns in the previous three months, excess returns (*EXRET*), which is the difference between firm's annualized equity return and the annualized value-weighted return of a portfolio with NYSE, AMEX, NASDAQ stocks, the relative size of the firm (*RSIZE*), defined as the (log of) equity market value divided by the total market capitalization of NYSE, AMEX, NASDAQ stocks and finally, the natural logarithm of stock price† (*LOGPRICE*). The variables are lagged by one year such that at the beginning of every year in the bankruptcy period 1990–2015 the data are available in the market for predictions within the year.

Table 2 provides descriptive statistics for bankrupt and healthy firm observations.

As can be seen from Table 2, there are several differences in the financial performance between bankrupt and healthy firms. Specifically, bankrupt firms are less profitable (*NIMTA* is lower), have less liquidity (*CASHMTA* is lower) and have higher levels of leverage (*TLMTA* is higher). Furthermore, book-to-market ratios of bankrupt firms are smaller (*BM* is lower) and tend to be smaller in size (*RSIZE* is lower), have lower stock prices (*LOGPRICE* is lower) and perform worse than the market (*EXRET* is negative for bankrupt firms and positive for healthy firms). Finally, equity returns for bankrupt firms are more volatile relative to healthy firms (*SIGMA* is higher). In the last column of the table, t-statistics for mean

† We follow Bauer and Agarwal (2014) and use the logarithm of unadjusted stock price.

Table 2. Descriptive statistics for the entire sample.

| Variables | Bankrupt firms Mean | Bankrupt firms Median | Bankrupt firms St.Dev | Healthy firms Mean | Healthy firms Median | Healthy firms St.Dev. | *t*-statistics |
|---|---|---|---|---|---|---|---|
| NIMTA | – 0.249 | – 0.1722 | 0.254 | – 0.021 | 0.023 | 0.148 | – 18.38 |
| CASHMTA | 0.068 | 0.032 | 0.098 | 0.119 | 0.062 | 0.160 | – 10.60 |
| TLMTA | 0.697 | 0.784 | 0.264 | 0.378 | 0.332 | 0.258 | 24.71 |
| BM | 1.036 | 0.481 | 2.551 | 1.461 | 0.537 | 4.807 | – 3.39 |
| RSIZE | – 12.774 | – 12.765 | 1.490 | – 10.919 | – 10.983 | 2.075 | – 25.40 |
| LOGPRICE | 0.523 | 0.560 | 1.145 | 2.291 | 2.473 | 1.274 | – 31.56 |
| EXRET | – 0.213 | – 0.340 | 0.864 | 0.207 | 0.106 | 0.670 | – 9.95 |
| SIGMA | 1.070 | 0.959 | 0.486 | 0.657 | 0.551 | 0.418 | 17.39 |

Note: This table reports descriptive statistics for the entire sample period 1990–2015 of the inputs, *x*, which enter the neural network model, as used in Campbell *et al.* (2008). The construction of the variables is described in section 4.2. The last column reports t-statistics for mean differences between bankrupt and healthy firms.

differences are reported. Overall, all mean differences are statistically significant.

## 5. Model performance

The aim is to examine whether our proposed methodology for the estimation of asset value and volatility outperforms the commonly used approaches which we discussed in section 2. With respect to that, we employ three distinct tests to compare the performance of the models, following Bauer and Agarwal (2014); (1) A discriminatory power test based on AUROC, (2) Information content test and (3) Economic benefits arising from using the different bankruptcy models.

### *5.1. Discriminatory power*

We evaluate the ability of the models to discriminate the bankrupt firms from the healthy firms. For a given cut-off probability, firms whose bankruptcy probability is higher than the cut-off, are classified as bankrupt and healthy otherwise. A way to measure discriminatory power is by counting the true predictions (percentage of bankrupt firms correctly classified as bankrupt) and the false predictions (percentage of healthy firms incorrectly classified as bankrupt). Doing this classification process for multiple cut-offs, we obtain a set of true and false predictions and when we plot them (true predictions on the y-axis and false predictions on the x-axis), we get the Receiver Operating Characteristics (ROC) curve. The higher the ROC curve towards the top-left corner, the more powerful the model is (since it will hit more true predictions and less false predictions). A quantitative assessment of the discriminatory power is given by the Area Under ROC (AUROC) curve which is widely used to measure the performance of binary response models and suitable in our case. We compute the AUROC following common practice (see for instance Hanley and McNeil 1982, Sobehart and Keenan 2001, Fitzpatrick and Mues 2016, Gupta *et al.* 2018, etc.).

### *5.2. Information content tests*

We evaluate the information content of the various BSM specifications by including the out of sample scores they produce in discrete hazard models. The score is: ln [PB / (1-PB)], where PB is the bankruptcy probability produced by the various structural models. In particular, we follow related studies such as, Hillegeist *et al.* (2004) and Agarwal and Taffler (2008), to estimate the following discrete logit model:

$$p(Y_{i,t+1} = 1|Score_{i,t}) = p_{i,t} = \frac{e^{a*Rate_t + \beta * Score_{i,t}}}{1 + e^{a*Rate_t + \beta * Score_{i,t}}}, \qquad (24)$$

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 $Score_{i,t}$, which is the out of sample score of the *i-th* firm at time $t$. Finally, $\beta$ is the coefficient estimate and, like prior studies, we proxy the baseline hazard rate using the actual bankruptcy rate at time $t$, which is equal to the number of bankruptcies in year $t$ divided by the number of observations in the year in our sample (denoted as *Rate*). We are interested in the statistical significance of the '$\beta$' coefficients which indicate whether the out of sample scores contain significant information about future bankruptcies. Finally, to test whether the out of sample scores of a model contains significantly more information than the other models (i.e. of our neuro-structural model versus the alternative BSM structural specifications), we compare the log-likelihoods of the corresponding logit regressions using the Vuong (1989)† test as well as their AUROCs using the DeLong *et al.* (1988) test. It should be noted here that, both the DeLong *et al.* (1988) and Vuong (1989) tests are fully applicable for our proposed methodology. When we back out the asset values and volatilities, these are used in the BSM model (referred to as NS approach, since these are obtained through our neuro-structural approach) to compute the probability of bankruptcy, as happens with the competing BSM models (that estimate asset values and volatilities differently). These bankruptcy probabilities are then used to derive the DeLong *et al.* (1988) and Vuong (1989) test statistics as shown later in Tables 4 and 5.

† As Vuong (1989) explains, the tests are derived for the cases of non-nested models (like our models 1–5 in table 5), nested models or overlapping models, thus the use of Vuong's test is appropriate for the standards of our study. In addition, the methodology has been employed in related studies such as, Charalambous et al. (2020).

Shumway (2001) argues that a panel logit model like the one in equation (24) should be estimated based on standard log-likelihood maximization programs, but with a minor adjustment. The number of independent observations is the number of firms in the estimation sample and not the number of firm-year observations. Failing to address this issue could yield understated standard errors, leading to wrong inference about the coefficient estimates. 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, White 1980).

Finally, although some studies have employed Cox proportional hazard models (Bharath and Shumway 2008, Charitou *et al.* 2013), the choice of logit hazard models is appropriate for our study. We are interested in the likelihood of an event to occur, in our case bankruptcy, but Cox hazard models address the time to event, which is not a primary concern of our study. In addition, logit hazard models have been widely used in related studies recently (see for instance Tian *et al.* 2015, Alzugaiby *et al.* 2021, Charalambous *et al.* 2022, etc.). Moreover, Gupta *et al.* (2018) empirically showed that logit hazard models are superior to Cox proportional hazard models.

### *5.3. Economic impact*

Earlier research suggested that the accurate prediction of corporate bankruptcies is important for bank profitability (Atiya 2001). In this section, we examine whether the accuracy of the models is economically beneficial for banks. Following Agarwal and Taffler (2008), we assume a competitive loan market worth \$100 billion, and each bank uses a different bankruptcy model to evaluate the creditworthiness of prospective clients.

**5.3.1. Calculating credit spreads.** We use the period 1990–2006 (70% of the sample) to calculate credit spreads. We sort the customers (i.e. firm-year observations) in 10 groups of equal size. The first and tenth group include the observations with the lowest and highest bankruptcy risk respectively, and a credit spread is calculated according to the following rule; Observations classified in the first group receive a credit spread $k$ and observations in the remaining groups receive a credit spread $CS_i$, which is obtained from Blochlinger and Leippold (2006) and it is defined as follows:

$$CS_i = \frac{p(Y=1|S=i)}{p(Y=0|S=i)} LGD + k, \tag{25}$$

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 and Taffler (2008), the average probability of bankruptcy for the *i-th* group is the actual bankruptcy rate for that group, defined as the number bankrupt observations divided by the number of observations in the group. Furthermore, $k = 0.3\%$ and $LGD = 45\%$.

**5.3.2. Measuring economic performance.** Banks compete to grant loans to prospective customers (i.e. firm-year observations) in the period 2007–2015. Using the different bankruptcy models, each bank sorts the customers according to their riskiness and denies credit to the bottom 5% with the highest risk. The remaining customers are divided in 10 groups and a credit spread is charged to each group, that was obtained from the period 1990–2006. Finally, the bank that charges the lowest credit spread for the customer 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 (RRWA), 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, pp. 64).

## 6. Results

This section discusses the results of the paper. It is important to mention here that the network structure, which we discussed in the methodology section, is optimized using data from the period 1990–2006 and it is implemented in the out of sample period 2007–2015 to test its performance. We begin by reporting the out of sample estimation of asset values and

Table 3. Mean asset values and volatilities in the out of sample period 2007–2015.

| | Mean-bankrupt firms | | Mean-healthy firms | | *t*-statistics | |
|---|---|---|---|---|---|---|
| Estimation approaches | V/F | $\sigma_V$ | V/F | $\sigma_V$ | V/F | $\sigma_V$ |
| 2-Eqs. Approach | 1.958 | 0.565 | 6.380 | 0.383 | – 2.69 | 4.53 |
| 1-Eq. Approach | 1.887 | 0.400 | 6.378 | 0.353 | – 2.74 | 2.09 |
| Direct Estimation | | | | | | |
| 1) BS (2008) | 2.004 | 0.489 | 6.406 | 0.417 | – 2.68 | 2.81 |
| 2) CDLT (2013) | 2.004 | 0.328 | 6.406 | 0.335 | – 2.68 | – 0.32 |
| NS Approach | 3.676 | 0.72 | 6.398 | 0.572 | – 24.12 | 25.73 |

Note: This table reports mean asset and volatility values obtained with respect to the various estimation approaches, in the out of sample period 2007–2015. The 2-Eqs. Approach refers to estimating asset values and volatilities by simultaneously solving equations (5) and (6). The 1-Eq. Approach refers to estimating the time-series of asset values over the previous year by solving equation (5) and estimating the volatility of asset values until convergence (see sections 2.2.1 and 2.2.2, respectively). BS (2008) and CDLT (2013) refer to the direct estimation approach as done in Bharath and Shumway (2008) and Charitou *et al.* (2013) respectively (see section 2.2.3). Finally, the NS approach refers to estimating expected asset value and the volatility based on the neuro-structural approach (see sections 3.1 and 3.2). The last column reports t-statistics for mean differences between bankrupt and healthy firms.

volatility with respect to the different estimation approaches and then, we report the performance of the various models in the out of sample period, based on AUROC, information content and economic impact. Several robustness tests are also conducted in this section, including a) comparing our neuro-structural model to logistic regression and standard neural networks and b) implementing our methodology to an extension of BSM model and specifically to a Down-and-Out option model which relaxes the assumption of default happening only at debt maturity.

In Table 3, we report asset values (expected leverage in the case of the neuro-structural approach, $E_L$) and volatility values with respect to the different estimation approaches, in the out of sample period 2007–2015. As expected, the ratio *V/F* is lower for bankrupt firms in all cases. Differences in the mean values between bankrupt and healthy firms are statistically significant. Similarly, $\sigma_V$ is higher for bankrupt firms, except in the case of CDLT (2013). Differences in the mean values of the remaining approaches are statistically significant. Overall, results are indicative of the impaired financial condition of bankrupt firms relative to healthy firms one year prior to bankruptcy. We conclude that our neuro-structural (NS) approach produces reasonable expected asset and volatility values.

### *6.1. AUROC results*

Table 4 presents the out of sample discriminatory power of the various BSM specifications based on the AUROC.†

Table 4. AUROC results in the out of sample period 2007–2015.

| | AUROC | Delong *et al.* (1988) test |
|---|---|---|
| 2-Eqs. Approach | 0.8964 | 5.64 |
| 1-Eq. Approach | 0.9026 | 5.40 |
| Direct estimation | | |
| (1) BS (2008) | 0.8791 | 6.45 |
| (2) CDLT (2013) | 0.9044 | 5.08 |
| NS approach | 0.9387 | – |

Note: This table reports AUROC results for the various BSM specifications in the out of sample period spanning the years 2007–2015. The 2-Eqs. Approach refers to estimating asset values and volatilities by simultaneously solving equations (5) and (6). The 1-Eq. Approach refers to estimating the time-series of asset values over the previous year by solving equation (5) and estimating the volatility of asset values until convergence (see sections 2.2.1 and 2.2.2 respectively). BS (2008) and CDLT (2013) refer to the direct estimation approach as done in Bharath and Shumway (2008) and Charitou *et al.* (2013) respectively (see section 2.2.3). Finally, the NS approach refers to estimating expected asset value and volatility based on our neuro-structural approach (see sections 3.1 and 3.2). The last column reports the Delong *et al.* (1988) test statistic, to test for statistically significant differences in the AUROCs between the neuro-structural model with the alternative BSM specifications.

The key finding is that the neuro-structural model substantially outperforms the competing approaches, suggesting that it is more powerful in discriminating the bankrupt firms from the healthy firms. Specifically, the AUROC of the model is 0.9387 whereas for the two and single equations approach, AUROCs are 0.8964 and 0.9026, respectively. In the last column, we report test statistics to examine for statistically significant differences in AUROCs when we compare our neuro-structural model against the competing approaches (the corresponding test statistic next to the NS approach is missing as the model cannot be compared by itself). According to Delong *et al.* (1988) test, differences in AUROCs between the neuro-structural model and the two and single equations approaches are statistically significant at the 1% level (test statistics are 5.64 and 5.40, respectively). The neuro-structural model is also superior from the direct estimation approaches, since the AUROCs of BS (2008)‡ and CDLT (2013) are 0.8791 and 0.9044 respectively. According to the Delong *et al.* (1988) test, differences in AUROCs between the neuro-structural model and the two direct estimation approaches are statistically significant at the 1% level (test statistics are 6.45 and 5.08, respectively).

Results from this test clearly shows the superiority of the neuro-structural approach in discriminating bankrupt from healthy firms relative to the alternative parametric BSM specifications.

### *6.2. Information content results*

Table 5 reports the results from information content tests. Models 1–5 are logit models that include, as predictors, the out of sample scores from the various BSM specifications as shown in Table 4. Models 1–2 include the scores by estimating the asset values and volatilities of the BSM model according to the 2-Eqs. and 1-Eq. Approaches respectively (denoted as Score 1 and Score 2, respectively). Next, Models 3–4 include the scores by estimating asset values and volatilities of the BSM based on BS (2008) and CDLT (2013), respectively, and are denoted as Score 3 and Score 4, respectively. Finally, Model 5 includes the score from our neuro-structural model (Score 5).

According to the results, out of sample scores produced by the various BSM specifications are highly statistically significant at the 1% level, indicating that they carry significant information in predicting bankruptcies one year ahead. More importantly, out of sample scores produced by the neuro-structural model contains significantly more information compared to the alternative approaches. In the last two rows, we report test statistics to examine for statistically significant differences in log-likelihoods and AUROCs when we compare Model 5 (which is the model of interest) against Models 1–4 (test statistics for Model 5 are missing as the model cannot be compared by itself). Using the Vuong (1989) test to compare the log-likelihoods, we find that the log-likelihood of Model 5 is significantly different from Models 1-4. Differences are statistically significant at the 1% level. This result also holds when we compare the AUROCs of the various

† The AUROCs in the table were estimated by using the BSM structural Equation (4) for the competing approaches and by using the structural Equation (14) for our approach. In a subsequent analysis, for information content test, we use the outputs of these structural models in logit regressions and run again AUROC and information content tests.

‡ BS (2008), however, use corporate defaults to tailor their model, hence their model may not be as accurate in predicting bankruptcies.

Table 5. Information content results in the out of sample period 2007–2015.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Score 1 | 0.022*** | | | | |
| | (13.14) | | | | |
| Score 2 | | 0.387*** | | | |
| | | (6.80) | | | |
| Score 3 | | | 0.281*** | | |
| | | | (9.78) | | |
| Score 4 | | | | 0.420*** | |
| | | | | (5.23) | |
| Score 5 | | | | | 1.248*** |
| | | | | | (20.01) |
| Rate | − 0.602*** | − 0.965*** | − 0.645*** | − 0.894*** | − 0.872*** |
| | ( − 3.65) | ( − 5.03) | ( − 3.41) | ( − 4.18) | ( − 4.53) |
| Constant | − 7.599*** | − 9.332*** | − 7.864*** | − 8.537*** | − 2.844*** |
| | ( − 8.35) | ( − 8.76) | ( − 7.50) | ( − 7.50) | ( − 2.87) |
| Log-Likelihood | − 693.60 | − 657.50 | − 690.23 | − 659.65 | − 573.61 |
| Pseudo-$R^2$ (%) | 17.83 | 22.11 | 16.43 | 21.85 | 32.05 |
| AUROC | 0.9040 | 0.9101 | 0.8837 | 0.9114 | 0.9388 |
| Vuong (1989) | 6.41 | 4.36 | 6.11 | 4.25 | – |
| DeLong *et al.* (1988) | 4.50 | 4.06 | 5.80 | 3.80 | – |

Note: This table reports information content results. We estimate five logit models as follows: First, we estimate the out of sample bankruptcy probabilities (in the period 2007–2015) using the various BSM specifications as shown in table 4. The probabilities are transformed to a score = ln [PB / (1-PB)], where PB is the probability of bankruptcy. Models 1 and 2 include scores from the 2-Eqs. and 1-Eq. Approaches (denoted with Score 1 and Score 2, respectively). Models 3 and 4 include scores produced by the direct estimation approach (Score 3 and Score 4 are scores from BS 2008 and CDLT 2013, respectively). Finally, Model 5 includes the score from our neuro-structural approach. In all logit models, we include *Rate*, defined as the annual number of bankruptcies in our sample divided by the number of observations in the year (in our sample), as a proxy for the baseline hazard rate (for consistency, it is transformed to a score as explained above). The last two rows of the table report the Vuong (1989) and DeLong *et al.* (1988) test statistics, to test for statistically significant differences in the log-likelihoods and AUROCs, respectively, between Model 5 and Models 1-4.
*** denote statistical significance at the 1% level. *t*-tests are reported in parentheses.

logit regressions. Surprisingly, the *Rate* control is negative and statistically significant. However, when we run the regressions by including only the *Rate* as independent variable, the coefficient is positive but insignificant. We believe that the interaction of *Rate* with *Score* causes this unexpected result.

From the tests we conclude that the scores obtained from the neuro-structural model contain significantly more information about future bankruptcies relative to the scores from other BSM specifications. This finding confirms that our approach yields more accurate asset value and volatilities that improve the performance of the parametric model.

### 6.3. Economic impact results

So far, we have assessed the performance of various BSM specifications based on discriminatory power and information content. However, banks are interested in the economic benefits arising by using the bankruptcy models in the decision-making process of giving loans to firms. Thus, does the improved performance using the neuro-structural model yields superior returns? We test this conjecture using the framework of Agarwal and Taffler (2008), by assuming a competitive loan market worth \$100 billion and five banks use the different bankruptcy models in their credit decisions.

Table 6 reports economic results for five banks. Banks 1 and 2 use the 2-Eqs. and 1-Eq. Approaches respectively for the estimation of asset values and volatilities. Banks 3 and 4 use the direct estimation approaches based on BS (2008) and CDLT (2013) respectively. Finally, Bank 5 uses our neuro-structural model.

As can be inferred from the table, Bank 5 manages a credit portfolio with the lowest concentration of bankruptcies (0.08%) whereas for the competitor banks, concentration of bankruptcies is higher, ranging from 0.10% to 0.90%. More importantly, Bank 5 earns higher risk-adjusted returns (i.e. accounting for the riskiness of the portfolio rather than the total profit earned). In particular, Bank 5, on a risk-adjusted basis, earns 2.06% per dollar invested while risk-adjusted returns for the competing banks range from 0.30% to 1.81%.†

Results from this test, overall, suggest that banks can have a competitive advantage using the neuro-structural approach relative to any of the alternative BSM specifications.

### 6.4. Robustness analysis

In this section, we perform several robustness tests. We begin the analysis by measuring the out of sample performance of the models using several other performance statistics. As a next test, we re-run and compare the models based on a five-fold validation approach. As an additional test, we compare the neuro-structural model with other widely used prediction methodologies such as, logistic regression and standard neural networks. Moreover, we implement our neuro-structural methodology to an advancement of BSM and specifically, to a Down-and-Out option model which sets a more realistic approach to bankruptcy, by allowing debt default to happen any time prior to debt maturity.

† Results are robust with respect to different parameter specifications ($k$ = 0.002–0.004 and LGD = 0.4–0.7).

Table 6. Economic results in the out of sample period 2007–2015.

| | Bank 1 | Bank 2 | Bank 3 | Bank 4 | Bank 5 |
|---|---|---|---|---|---|
| Credits | 5115 | 3546 | 2652 | 3254 | 12977 |
| Market Share (%) | 18.27 | 12.67 | 9.47 | 11.63 | 46.36 |
| Bankruptcies | 43 | 5 | 24 | 4 | 10 |
| Bankruptcies/Credits (%) | 0.84 | 0.14 | 0.90 | 0.10 | 0.08 |
| Average Spread (%) | 0.54 | 0.35 | 0.46 | 0.36 | 0.35 |
| Revenues ($M) | 98.90 | 44.57 | 43.17 | 41.33 | 162.92 |
| Loss($M) | 63.16 | 7.34 | 35.25 | 5.88 | 14.69 |
| Profit($M) | 35.74 | 37.23 | 7.92 | 35.45 | 148.23 |
| Return on Assets (%) | 0.20 | 0.29 | 0.08 | 0.30 | 0.32 |
| Return on RWA (%) | 0.54 | 1.81 | 0.30 | 1.73 | 2.06 |

Note: This table reports economic results for five banks in a competitive loan market worth $100 billion. Banks 1 and 2 use the BSM specification where asset values and volatilities are obtained using the 2-Eqs. and 1-Eq. Approaches respectively. Banks 3 and 4 use the direct estimation approach to obtain asset values and volatilities, based on BS (2008) and CDLT (2013) respectively. Finally, Bank 5 uses the neuro-structural approach. Banks sort prospective customers (firm-year observations between 2007 and 2015) and reject the 5% of observations with the highest risk. The remaining observations are classified in 10 groups of equal size and for each group, a credit spread is calculated as described in the main text (section 5.3). The bank that classifies the observation in 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 and 2006 and equals 0.43%. Loss given default is 45%.

**6.4.1. Other performance statistics.** Several other tests exist in the literature to evaluate the performance of bankruptcy prediction models. In this section, we use the Kolmogorov–Smirnov (KS) statistic, the Conditional Information Entropy Ratio (CIER) statistic and the H-measure. Our results (not tabulated for bervity) demonstrate that the neuro-structural model outperforms the alternative BSM specifications. Specifically, the KS statistic is 0.75 for the neuro-structural model whereas for the 2-Eqs. approach is 0.68, for the 1-Eq. approach is 0.67, for BS (2008) is 0.61 and for CDLT (2013) is 0.68. The CIER statistic for the neuro-structural model is 0.22 whereas for the other approaches CIER statistic is 0.19, 0.17, 0.15, 0.18 (we keep the same order of the models as with the KS). Finally, the H-measure for the neuro-structural model is 0.65 whereas for the competing models the H-measure is 0.51, 0.51, 0.45, 0.52.

**6.4.2. Five-fold validation.** For this test, we divide the full sample (1990–2015) into five approximately equal subsamples in chronological order. We use any four of them to train the neuro-structural model and use the left-out sample to measure its performance. Then, we compare its performance with the alternative specifications in each of the left-out subsample, using AUROC as a summary statistic. In each sub-sample, the neuro-structural model outperforms the alternative BSM specifications (not tabulated for brevity). Its average performance is 0.9102 where for the other models, performance is as follows: Using the 2-Eqs. and 1-Eq. approaches, average AUROC is 0.8431 and 0.8727 respectively. For BS (2008) and CDLT (2013), average AUROCs are 0.8507 and 0.8747, respectively. The performance though is lower than the performance reported in the earlier sections, because the five-fold validation approach breaks the chronological order of the data (i.e. we use subsequent periods to train the model and measuring performance on earlier periods). However, the key finding remains: Estimating asset value and volatility using our approach, outperforms the alterative BSM specifications.

**6.4.3. Comparison with alternative methodologies.** The excellent performance of the neuro-structural model motivates us to compare its performance to alternative bankruptcy prediction methodologies. Specifically, we compare the performance of the neuro-structural model with two widely used approaches; the logistic regression (LR) approach and the neural networks approach (NN). As explanatory variables for both approaches, we use the variables of Campbell *et al.* (2008), which are also used as inputs when estimating our neuro-structural model. Furthermore, in the case of the traditional neural network, we use the same specifications as was done for the neuro-structural model for consistency; In the hidden layer, we use three neurons ($H = 3$) as well as we use the tan-sigmoid activation function. In the output layer, we use one neuron ($M = 1$) using the log-sigmoid activation function to obtain a probability. Finally, the log-likelihood function is used to train the neural network to obtain its coefficients. Out of sample performance results are reported in Table 7.

As expected, the results now are more comparable since all approaches generally perform well in predicting bankruptcy. However, the neuro-structural model provides, overall, better predictive accuracy relative to the competing methodologies. This is evident by the higher out of sample AUROC it exhibits relative to the LR and NN approaches (which equal to 0.9283 and 0.9271, respectively) with the differences being statistically significant according to the DeLong *et al.* (1988) test. Next, the neuro-structural model is better in terms of information content from the standard neural network (difference in log-likelihoods is statistically significant according to the Vuong's test) but insignificant compared to the logistic regression model. Finally, a bank which uses the neuro-structural model is more profitable relative to banks that use either the

Table 7. Performance comparisons between the neuro-structural approach and alternative approaches in the out of sample period 2007–2015.

| | AUROC | LL-Info. Content | RRWA (%) | DeLong *et al.* (1988) Test | Vuong (1989) Test |
|---|---|---|---|---|---|
| LR Approach | 0.9283 | – 571.94 | 1.42 | 2.08 | – 1.29 |
| NN Approach | 0.9271 | – 585.13 | 1.26 | 2.45 | 3.04 |
| NS Approach | 0.9387 | – 573.61 | 2.45 | – | – |

Note: This table reports performance results of the alternative approaches for predicting bankruptcy such as the logistic regression (LR) approach, the neural network (NN) approach and finally, the neuro-structural (NS) approach. Performance is measured in the out of sample period spanning the years 2007–2015. The first column reports AUROC results (equivalent to Table 4), the second column reports log-likelihoods from information content tests (equivalent to Table 5) and the third column reports the return on risk weighted assets (RRWA) when banks compete to grant loans in a competitive economy (equivalent to the last row of Table 6). The last two columns report DeLong *et al.* (1988) and Vuong (1989) test statistics, to test for statistically significant differences in the AUROCs and log-likelihoods, between the neuro-structural approach and the alternative methodologies.

logistic model or the neural network. In particular, the bank which uses the neuro-structural model achieves a return of 2.45% on risk weighted assets, whereas banks which use the LR or NN approach achieve 1.42% and 1.26%, respectively. Note that the economic benefits (RRWA) for our model reported in Table 7 differ from those reported in Table 6 simply because in Table 6 the customers-observations are allocated to five banks (as the number of structural models compared) whereas in Table 7 the customers-observations are allocated to three banks (as the number of models compared).

Consequently, the profits are shared among three banks, which explains the slightly higher RRWA of our model reported in Table 7 relative to Table 6.

A note is worth mentioning here. The superiority of our neuro-structural model over standard neural networks is reasonable because our model retains the strong capabilities of neural networks, but their training is improved because of the 'co-operation' with the structural model. As we explained earlier in section 3.4, during the training of neural networks, the weights are updated according to the theoretical properties of the parametric model as opposed to the standard neural network which includes no knowledge from any theoretical (structural) model of bankruptcy.

Overall, the results in this section suggest that the neuro-structural model is a promising methodology since it outperforms other well-known prediction methodologies in the majority of tests we employ. In other words, the co-joint dynamics between the neural network and the structural model add value to the standard neural network and the standard structural model.

**6.4.4. Extensions of BSM: the case of down-and-out call option.** The BSM model which we previously analyzed and improved by estimating its unobservable parameters with our neuro-structural approach, is based on several assumptions as it was the first parametric model that was proposed in the context of bankruptcy prediction. In the options pricing literature, however, there are other models which relax some of the restrictive assumptions of BSM. One of such models is the one which views the firm as a Down-and-Out call option. In the Down-and-Out option (DaO), a barrier exists such that whenever it is breached up to the maturity time of the option, it causes the termination of the option contract (i.e. whenever the underlying asset value breaches the barrier, the option becomes worthless). This point of view can be applied in the case of a firm. Whenever the value of the firm falls below a barrier, for example the value of liabilities, up to the maturity time of liabilities, the firm goes bankrupt. Thus, according to the DaO framework, the firm can go bankrupt at any time prior to the maturity of liabilities whenever the value of the firm falls below liabilities. Unlike the case of the standard BSM that assumes bankruptcy can only happen at the maturity of liabilities, the DaO framework recognizes that bankruptcy can occur at any point prior to maturity (see for instance Brockman and Turtle 2003, Afik *et al.* 2016 for more details). In addition, as we will show subsequently, the probability of bankruptcy in the DaO framework is always higher than the BSM model because it accounts for the possibility of bankruptcy to occur prior to maturity (in BSM this probability is zero).

In this section, we use the DaO model and using our neuro-structural approach, we estimate its unobservable parameters. Following the DaO option studies, we measure the probability of bankruptcy as follows†:

$$PB = N(A) + \exp(-2P_1P_2)N(B), \tag{26}$$

where

$$A = -(P_1 + P_2), B = -(P_1 - P_2),$$

and

$$P_1 = \frac{\ln(V/F)}{\sigma_V\sqrt{T}}, P_2 = \frac{(\mu - 0.5\sigma_V^2)T}{\sigma_V\sqrt{T}}.$$

Unlike in the case of BSM where we can merge asset return (μ) and asset value (V) to produce the expected value of assets, in equation (26) this is not possible. Thus, we have three unobservable parameters: V, $\sigma_V$ and μ, which we estimate within our neuro-structural framework. More specifically, the three parameters can be viewed as functions of some exogenous variables in the vector x:

$$L(w) = f_L(x, w), \tag{27}$$

$$\sigma_V(w) = f_\sigma(x, w), \tag{28}$$

† Note that N(A) in Equation (27) gives the probability of bankruptcy based on the BSM. Since the equation contains another non-negative term, the probability of bankruptcy in the DaO framework is always equal to, or higher than BSM

Table 8. Performance comparisons between the neuro-structural approach (Down-and-Out model) and alternative approaches in the out of sample period 2007–2015.

| | AUROC | LL-Info. Content | RRWA (%) | DeLong *et al.* (1988) Test | Vuong (1989) Test |
|---|---|---|---|---|---|
| LR Approach | 0.9283 | − 571.94 | 1.43 | 2.12 | − 1.50 |
| NN Approach | 0.9271 | − 585.13 | 1.30 | 2.47 | 3.10 |
| NS Approach (DaO) | 0.9399 | − 574.84 | 2.11 | – | – |

Note: This table reports performance results of the alternative approaches for predicting bankruptcy such as the logistic regression (LR) approach, the neural network (NN) approach and finally, the neuro-structural (NS) approach when the DaO option framework is used. Performance is measured in the out of sample period spanning the years 2007–2015. The first column reports AUROC results (equivalent to Table 4), the second column reports log-likelihoods from information content tests (equivalent to Table 5) and the third column reports the return on risk weighted assets (RRWA) when banks compete to grant loans in a competitive economy (equivalent to the last row of Table 6). The last two columns report DeLong *et al.* (1988) and Vuong (1989) test statistics, to test for statistically significant differences in the AUROCs and log-likelihoods, between the neuro-structural approach and the alternative methodologies.

$$\mu(w) = f_\mu(x, w). \tag{29}$$

Equations (27)–(29) are estimated using the same methodology that we introduced for the case of BSM but adapted to the case of DaO model. Results, reported in Table 8 show that the neuro-structural model is again the best performing one compared to logistic regression and neural networks in terms of discriminatory power, information content and economic benefits. Specifically, the AUROC differences of the neuro-structural model against the standard neural network and logistic regression are statistically significant according to the DeLong *et al.* (1988) test. As for the information content, we find statistically significant differences between the log-likelihoods of the neuro-structural model and the standard neural network.

Finally, the neuro-structural DaO model (in Table 8) does not necessarily outperform the neuro-structural BSM model (in Table 7). It has a higher AUROC but not as good a log-likelihood and RRWA, suggesting that there is very little to differentiate between the two.

**6.4.5. A note on other hybrid models.** Some other studies obtain the distance to default or the probability of bankruptcy from the structural-parametric model and then use it as input, along with other predictors, in the bankruptcy prediction model thus creating a 'hybrid' model. For instance, Campbell *et al.* (2008) include the distance to default as additional predictor in their model, but they find no significant improvement. Charalambous *et al.* (2020) use the probability of bankruptcy obtained from the structural-parametric model of Leland-Toft (1996) as additional input to the Campbell *et al.* (2008) model and they find that the hybrid model outperforms the Campbell *et al.* (2008) model.

In all, these papers use a set of variables as inputs in the bankruptcy prediction model and the output is directly the probability of bankruptcy (in this context, the structural model equation is not directly used during the estimation). Our work, on the other hand, includes an intermediary step to obtain structural parameters (asset value and volatility) and then obtaining the probability of bankruptcy using the structural model. By adding this intermediary step, we exploit the joint dynamics of the neural network and the structural-option model, which are useful during the training phase because the neural network and the structural model work in conjunction by providing information forward and backward, respectively, during the learning phase (as shown in Figures 2–5). This is not possible with the hybrid methods used in Campbell *et al.* (2008) and Charalambous *et al.* (2020).

We compare† the discriminatory power of our neuro-structural model and the hybrid model in Charalambous *et al.* (2020) and find that the AUROC of the neuro-structural model, as shown previously in Table 4, is 0.9387 whereas for the hybrid model is 0.9291, indicating that our approach yields higher discriminatory power than the hybrid approach used by other studies.

## 7. Summary and conclusions

In this paper, we introduce an estimation method to obtain parameter values such as the asset value and volatility, which are used in the structural-parametric models for the estimation of the probability of bankruptcy. Specifically, we view the unobservable parameters of parametric models as unknown functions of some exogenous variables in the vector $x$, which we estimate through learning, by embedding in the parametric models a neural network, yielding a neuro-structural model for bankruptcy prediction. Our method provides significant advantages over traditional estimation methods which are well-supported by our empirical results. Specifically, embedding neural networks in the parametric models, allows both to work in conjunction which is useful during the training phase of the neural network.

Our empirical results suggest that our estimation method provides more accurate parameters estimation relative to other estimation approaches, as shown by the superior performance of the neuro-structural model in terms of discriminatory power, information content and economic impact. Of course, we should acknowledge the fact that some of the models we compare, such as BS (2008), tailor their model for defaults rather than bankruptcies and may be responsible for some of the observed superiorities of our approach. Moreover, the neuro-structural model outperforms the standalone neural network, suggesting that the co-joint dynamics of the neural network and the structural model are useful during the learning phase.

† Extensive results are not tabulated but can be provided upon request.

Our work has implications in the empirical application of bankruptcy prediction as it provides a more sophisticated†, yet more accurate method to compute the unobservable parameters of the structural-parametric models that improves their out of sample performance. Moreover, our findings have implications in the training of neural networks as their (bankruptcy) prediction performance improves when involving a theoretical model (i.e. the structural model) during the estimation of the weights of the neural network. As neural networks are frequently characterized as 'black boxes', our neuro-structural framework involves both the neural network and the structural (theoretical options pricing model) during the training phase thus adjusting the weights of the neural network by embedding knowledge from the structural model. In addition, our work can be considered as a neuro-extension of the methodology in Campbell *et al.* (2008). There, the input vector $x$ is used to directly obtain the probability of bankruptcy. Our work uses the same input vector $x$ but includes an intermediary step to obtain option parameters, such as asset value and volatility, through the neural network model. Once the parameters are obtained, they are used as inputs to the parametric model to estimate the probability of bankruptcy.

Future extensions could focus on extended parametric models for bankruptcy prediction such as, Leland (1994) or Leland and Toft (1996).

## Acknowledgements

We would like to thank Mathias Verreydt and John Finnerty for discussing the paper at the FMA 2021 European Conference in Limassol, Cyprus and at the FMA 2020 International Conference in New York, USA, respectively, the conference participants of the FMARC 2019 annual conference in Limassol, Cyprus and IFORS 2017 triennial conference at Quebec, Canada.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

*Zenon Taoushianis* http://orcid.org/0000-0003-2002-6040

## References

Afik, Z., Arad, O. and Galil, K., Using merton model for default prediction: An empirical assessment of selected alternatives. *J. Emp. Finance*, 2016, **35**, 43–67.

Agarwal, V. and Taffler, R., Comparing the performance of market-based and accounting-based bankruptcy prediction models. *J. Bank. Finance*, 2008, **32**, 1541–1551.

Ait-Sahalia, Y. and Duarte, J., Nonparametric option pricing under shape restrictions. *J. Econom.*, 2003, **116**, 9–47.

Ait-Sahalia, Y. and Lo, A. W., Nonparametric estimation of state-price densities implicit in financial asset prices. *J Finance*, 1998, **53**, 499–547.

Almamy, J., Aston, J. and Ngwa, L., An evaluation of altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *J. Corp. Finance*, 2016, **36**, 278–285.

Altman, E. I., Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Finance*, 1968, **23**, 589–609.

Alzugaiby, B., Gupta, J., Mullineux, A. and Ahmed, R., Relevance of size in predicting bank failures. *International J. Finance Econ.*, 2021, **26**, 3504–3543.

Andreou, P. C., Charalambous, C. and Martzoukos, S. H., Generalized parameter functions for option pricing. *J. Bank. Finance*, 2010, **34**, 633–646.

Atiya, A. F., Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Trans. Neural Netw.*, 2001, **12**, 929–935.

Bandler, J. W., Ismail, M. A., Rayas-Sanchez, J. E. and Zhang, Q.-J., Neuromodeling of microwave circuits exploiting space-mapping technology. *IEEE Trans. Micro. Theor. Techn.*, 1999, **47**, 2417–2427.

Basel Committee on Banking Supervision, *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, 2006 (Bank for International Settlements).

Bauer, J. and Agarwal, V., Are hazard models superior to traditional bankruptcy prediction models? A comprehensive test. *J. Bank. Finance*, 2014, **40**, 432–442.

Bharath, S. T. and Shumway, T., Forecasting default with the Merton distance to default model. *Rev. Finan. Studies*, 2008, **21**, 1339–1369.

Black, F. and Scholes, M., The pricing of options and corporate liabilities. *J. Political. Econ.*, 1973, **81**, 637–654.

Blochlinger, A. and Leippold, M., Economic benefit of powerful credit scoring. *J. Bank. Finance*, 2006, **30**, 851–873.

Brockman, P. and Turtle, H. J., A barrier option framework for corporate security valuation. *J. Financ. Econ.*, 2003, **67**, 511–529.

Campbell, J., Hilscher, J. and Szilagyi, J., In search of distress risk. *J Finance*, 2008, **63**, 2899–2939.

Charalambous, C., Conjugate gradient algorithm for efficient training of artificial neural networks. *IEE Proceed. G (Circuits, Devices and Systems)*, 1992, **139**, 301–310.

Charalambous, C., Martzoukos, H. S. and Taoushianis, Z., Estimating corporate bankruptcy forecasting models by maximizing discriminatory power. *Review Quant. Fin. Acc.*, 2022, **58**, 297–328.

Charalambous, C., Martzoukos, S. H. and Taoushianis, Z., Predicting corporate bankruptcy using the framework of leland-toft: Evidence from US. *Quant. Finance*, 2020, **20**, 329–346.

Charitou, A., Dionysiou, D., Lambertides, N. and Trigeorgis, L., Alternative bankruptcy prediction models using option-pricing theory. *J. Bank. Finance*, 2013, **37**, 2349–2341.

Chernozhukov, V., Fernandez-Val, I., Newey, W., Stouli, S. and Vella, F., Semiparametric estimation of structural functions in non-separable triangular models. *Quant. Econom.*, 2020, **11**, 503–533.

Crosbie, P. and Bohn, J., *Modeling Default Risk*, 2003 (Moody's KMV).

Cybenko, G., Approximation by superpositions of sigmoidal function. *Math. Cont. Signal Syst.*, 1989, **2**, 303–314.

DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L., Comparing the areas under two or more correlated receiver operatingcharacteristic curves: A nonparametric approach. *Biometrics*, 1988, **44**, 837–845.

Escanciano, J. C., Jacho-Chavez, D. and Lewbel, A., Identification and estimation of semiparametric two-step models. *Quant. Econom.*, 2016, **7**, 561–589.

† To the extent that the model is cross-validated and evaluated on the testing dataset, minimizes the risk of overfitting while ensuring its generalizability.

Escanciano, J. C., Pardo-Fernandez, J. C. and Keilegom, I. V., Semiparametric estimation of risk-return relationships. *J. Bus. Econ. Stat.*, 2017, **25**, 40–52.

Filipe, S. F., Grammatikos, T. and Michala, D., Forecasting distress in European SME portfolios. *J. Bank. Finance*, 2016, **64**, 112–135.

Fitzpatrick, T. and Mues, C., An empirical comparison of classification algortihms for mortgage default prediction: Evidence from distressed mortgage market. *Eur. J. Oper. Res.*, 2016, **249**, 427–439.

Galil, K. and Gilat, N., Predicting default more accurately: To proxy or not to proxy for default? *Int. Rev. Finance*, 2019, **19**, 731–758.

Gu, S., Kelly, B. and Dacheng, X., Empirical asset pricing via machine learning. *Review Finan. Studies*, 2020, **33**, 2223–2273.

Gupta, J., Gregoriou, A. and Ebrahimi, T., Empirical comparison of hazard models in predicting SMEs failure. *Quant. Finance*, 2018, **18**, 437–466.

Hanley, J. A. and McNeil, B. J., The meaning and use of of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982, **143**, 29–36.

Hillegeist, S. A., Keating, E. K., Cram, D. P. and Lundstedt, K. G., Assessing the probability of bankruptcy. *Review Account. Stud.*, 2004, **9**, 5–34.

Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators. *Neural Netw.*, 1989, **2**, 359–366.

Huber, P. J., The behavior of maximum likelihood estimates under nonstandard conditions. *Berkeley Symp. Math. Stat. Prob.*, 1967, 221–233.

Jones, E. P., Mason, S. P. and Rosenfeld, E., Contingent claims analysis of corporate capital structures: An empirical investigation. *J Finance*, 1984, **39**, 611–625.

Leland, H., Corporate debt value, bond covenants, and optimal capital structure. *J Finance*, 1994, **49**, 1213–1252.

Leland, H. E. and Toft, K. B., Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *J Finance*, 1996, **51**, 987–1019.

Li, R., Leng, C. and You, J., Semiparametric tail index regression. *J. Bus. Econ. Stat.*, 2022, **40**, 82–95.

Merton, R., On the pricing of corporate debt: The risk structure of interest rates. *J Finance*, 1974, **29**, 449–470.

Ohlson, J. A., Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.*, 1980, **18**, 109–131.

Pindado, J., Rodrigues, L. and de la Torre, C., Estimating financial distress likelihood. *J. Bus. Res.*, 2008, **61**, 995–1003.

Ronn, E. I. and Verma, A. K., Pricing risk-adjusted deposit insurance: An option-based model. *J Finance*, 1986, **41**, 871–895.

Ruf, J. and Wang, W., Neural networks for option pricing and hedging: A literature review. *J. Comput. Finance*, 2020, **24**, 1–46.

Ruf, J. and Wang, W., Hedging with linear regressions and neural networks. *J. Bus. Econ. Stat.*, 2021, forthcoming.

Shumway, T., Forecasting bankruptcy more accurately: A simple hazard model. *J. Bus.*, 2001, **74**, 101–124.

Sobehart, J. and Keenan, S., Measuring default accurately. *Risk*, 2001, 31–33.

Tian, S., Yu, Y. and Guo, H., Variable selection and corporate bankruptcy forecasts. *J. Bank. Finance*, 2015, **52**, 89–100.

Vasallou, M. and Xing, Y., Default risk in equity returns. *J Finance*, 2004, **59**, 831–868.

Vuong, Q. H., Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 1989, **57**, 307–333.

White, H., A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 1980, **48**, 817–838.