University of Cyprus

DEPARTMENT OF ACCOUNTING AND FINANCE

ESSAYS ON BANKRUPTCY PREDICTION

DOCTOR OF PHILOSOPHY DISSERTATION

ZENON TAOUSHIANIS

2019

University of Cyprus

DEPARTMENT OF ACCOUNTING AND FINANCE

ESSAYS ON BANKRUPTCY PREDICTION

ZENON TAOUSHIANIS

A Dissertation submitted to the University of Cyprus in partial fulfillment of the requirements for the degree of Doctor of Philosophy

May 2019

VALIDATION PAGE

Doctoral Candidate: Zenon Taoushianis

Doctoral Thesis Title: Essays on Bankruptcy Prediction

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Department of Accounting and Finance and was approved on the …………………………. by the members of the Examination Committee

Examination Committee:

Chair of the Committee: ______________________________________

Zenios Stavros, Professor of Finance, Department of Accounting and Finance, University of Cyprus

Research Supervisor: ______________________________________

Charalambous Chris, Emeritus Professor of Management Science, Department of Business and Public Administration, University of Cyprus

Research Supervisor: ______________________________________

Martzoukos Spiros, Associate Professor of Finance, Department of Accounting and Finance, University of Cyprus

Committee Member: ______________________________________

Topaloglou Nikolaos, Associate Professor of Finance, Department of International and European Economic Studies, Athens University of Economics and Business

Committee Member: ______________________________________

Hassapis Christis, Associate Professor of Economics, Department of Economics, University of Cyprus

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Zenon Taoushianis

ΠΕΡΊΛΗΨΗ ΔΙΑΤΡΙΒΗΣ

Η διατριβή εξετάζει θέματα σχετικά με την πρόβλεψη χρεωκοπίας δημόσιων εταιριών στην Αμερική και αποτελείται από τρία κεφάλαια. Στο πρώτο κεφάλαιο εξετάζεται μια εναλλακτική μέθοδος πρόβλεψης χρεωκοπίας που απευθύνεται σε εταιρίες που έχουν δάνεια και πληρώνουν κουπόνια (coupon-paying debts). Η μέθοδος βασίζεται στο μοντέλο Leland-Toft (1996) το οποίο αποτελεί επέκταση ενός ευρέως διαδεδομένου μοντέλου; του Black-Scholes-Merton. Παρ 'όλα αυτά, το μοντέλο Leland-Toft (1996) δεν έχει λάβει την απαραίτητη προσοχή στην βιβλιογραφία και γι' αυτό, εξετάζουμε την εμπειρική του απόδοση. Τέλος αποσκοπούμε στην βελτίωση της απόδοσης παραδοσιακών μοντέλων πρόβλεψης χρεωκοπίας, όπως τα μοντέλα του Altman (1968) και Ohlson (1980) αλλά και πιο σύγχρονων μοντέλων, όπως το μοντέλο των Campbell et al. (2008), χρησιμοποιώντας την πιθανότητα χρεωκοπίας που προκύπτει από το μοντέλο Leland-Toft, σαν επιπρόσθετο παράγοντα στα μοντέλα τους. Συνοπτικά, το κεφάλαιο αποδεικνύει τη χρησιμότητα του μοντέλου Leland-Toft για την πρόβλεψη χρεωκοπίας, καθώς παρέχει βελτιώσεις στην εμπειρική απόδοση των μοντέλων που εξετάζονται. Το δεύτερο κεφάλαιο ασχολείται με την μεγιστοποίηση του κριτήριου απόδοσης AUROC (Area Under ROC curve). Συγκεκριμένα, το AUROC είναι από τα πιο γνωστά κριτήρια απόδοσης μοντέλων χρεοκοπίας και έχει δειχθεί πως τράπεζες που χρησιμοποιούν μοντέλα με μεγαλύτερο δείκτη AUROC, πετυχαίνουν μεγαλύτερη κερδοφορία εν συγκρίσει με άλλες τράπεζες. Παρ 'όλα αυτά, δεν είναι κοινή πρακτική να εκπαιδεύονται μοντέλα τα οποία μεγιστοποιούν το κριτήριο AUROC. Στο κεφάλαιο αυτό, προτείνονται και συγκρίνονται διάφορες μέθοδοι για μεγιστοποίηση του κριτηρίου απόδοσης AUROC, με τον σκοπό να βρούμε την καλύτερη μέθοδο. Συνοπτικά, το κεφάλαιο δείχνει ότι οι προτεινόμενες μέθοδοι βελτιώνουν την εμπειρική απόδοση παραδοσιακών μοντέλων. Επίσης, αναδεικνύονται τα οικονομικά οφέλη που προκύπτουν όταν το κριτήριο AUROC χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης μοντέλων πρόβλεψης χρεωκοπίας. Στο τρίτο κεφάλαιο δίνεται έμφαση στην κατηγορία των δομικών (παραμετρικών) μοντέλων , όπου προτείνουμε μια τεχνική για να υπολογίσουμε τις πιο σημαντικές τους παραμέτρους, οι οποίες δεν μπορούν να παρατηρηθούν στην αγορά; την αγοραία αξία της εταιρίας ή αντίστοιχα την αξία των περιουσιακών της στοιχείων (asset value) και την τυπική απόκλιση τους (asset volatility). Εναλλακτικές τεχνικές υπολογισμού που προτάθηκαν στην βιβλιογραφία βασίζονται σε προσεγγίσεις που μπορεί να οδηγήσουν σε ανακριβείς υπολογισμούς (προσεγγίσεις που είναι noisy or simplified). Στο κεφάλαιο αυτό, υποθέτουμε πως οι παράμετροι που θέλουμε να υπολογίσουμε, εξαρτώνται από κάποιες

εξωγενής μεταβλητές μέσω αγνώστων σχέσεων-συναρτήσεων και χρησιμοποιώντας μια μη-παραμετρική μέθοδο, για παράδειγμα νευρωνικά δίκτυα, στοχεύουμε στην εκμάθηση αυτών των συναρτήσεων όπου θα δώσουν βελτιωμένες παραμέτρους. Αυτές οι παράμετροι, όταν ενσωματωθούν στο παραμετρικό μοντέλο, δημιουργούν ένα ημι-παραμετρικό μοντέλο. Χρησιμοποιώντας το παραμετρικό μοντέλο Black-Scholes-Merton ως παράδειγμα, το κεφάλαιο καταλήγει ότι η προτεινόμενη μέθοδος παρέχει παραμέτρους οι οποίες όταν ενσωματωθούν στο Black-Scholes-Merton, βελτιώνει σημαντικά την απόδοση του συγκριτικά με τις παραδοσιακές μεθόδους υπολογισμού τους.

SUMMARY OF THE DISSERTATION

The dissertation examines topics in bankruptcy prediction using public firms from U.S. and consists of three chapters. Chapter 1 is dedicated to investigating an alternative approach for bankruptcy prediction that measures the financial healthiness of firms with coupon-paying debts. The approach is based on the framework of Leland-Toft (1996) which is a structural model that extends a widely-used corporate bankruptcy model; the Black-Scholes-Merton model. Despite that, Leland-Toft (1996) has received limited attention in the bankruptcy literature and thus, we aim to examine its empirical performance. Finally, we are interested to improve the performance of well-established bankruptcy models, like Altman (1968) and Ohlson (1980) but also more recent ones, like Campbell et al. (2008) by incorporating the probability of bankruptcy derived from Leland-Toft as additional predictor in their models. Overall, the chapter demonstrates the usefulness of Leland-Toft in predicting bankruptcy, since it provides enhancements in the empirical performance of the examined models. Chapter 2 is dedicated on the maximization of the discriminatory power of bankruptcy prediction models, measured by the Area Under ROC curve (AUROC). Specifically, AUROC is a widely-used performance measure and it has been shown that models with higher AUROC are associated with higher economic benefits for banks. Yet, it is not a common practice to training bankruptcy models by maximizing AUROC. In this chapter, several methodologies to maximize AUROC are introduced and compared, with the objective to find the best one. Overall, the chapter shows that the proposed approaches provide enhancements in the empirical performance of the traditional bankruptcy models, highlighting also the economic benefits arising by using models where the AUROC is used as the optimization criterion during the training phase of the models. In Chapter 3, the focus is on structural (parametric) models where we propose an estimation technique to estimate their most important parameters which are not observed in the market; the value of assets and the volatility of assets. Alternative estimation techniques proposed in the literature, are based on "noisy" techniques or "simplified" approximations that may result to inaccurate estimation of the unobserved parameters. In this chapter, we assume that these parameters depend on some exogenous variables through some unknown relationships and by using a nonparametric approach, like neural networks, we seek to estimate these relationships, obtaining in that way improved parameter values that when enter the parametric model, yields a semiparametric method for the estimation of the probability of default. Using the Black-Scholes-Merton structural model as a paradigm, the chapter concludes that the proposed

methodology provides parameters which when enter the structural model, significantly improve the predictive performance of the model relative to the traditional methods.

FOREWORD

Prediction of corporate bankruptcy is an area of research that is active for the last 50 years and specifically, since the paper of Altman (1968) who shows how to construct bankruptcy scores for firms using readily available information from financial statements. Since then, there has been a massive interest for the development of powerful bankruptcy prediction models, resulting to significant methodological advancements in the field of corporate bankruptcy prediction. The effort to predict corporate bankruptcy in the most accurate way continues until today and we argue that the topic will always be of interest. This is because, it is a topic that concerns a lot of parties including banks, investors, regulators, auditors, employees, the management of the firm and generally the stakeholders of the firm.

The topic has regenerated increased attention recently, mainly for two reasons. The first one, is that the Basel Committee on Banking Supervision in a consultation document in 2006 (known as Basel II which has been substituted by Basel III later), reports that banks are now allowed to develop internally, models to provide probability of default estimates on their credit exposures to estimate capital requirements that must set aside in order to absorb possible losses stemming from potential customer (i.e. firm) bankruptcies. This has been a huge motivation for banks to devote resources and sophistication for the development of such models. As it is also shown in the subsequent chapters, banks which use models with higher predictive accuracy, earn higher returns relative to the competition. Much of the attention that has been given to the development of bankruptcy models, also comes from the global financial crisis that hit the markets internationally between 2007-2008 and resulted to many corporations filing for bankruptcy and left banks with huge losses from their credit portfolios. Several economists have also characterized the 2007-2008 financial crisis equivalent to the Great Depression back to 1930's. These facts remind us the importance to develop models that provide early warning signals related to the financial condition of the corporations, especially during the crises. As a response to this great need, each chapter dedicates a section for testing the performance of the proposed models during the recent global financial crisis

Understanding the importance and the challenges of developing bankruptcy prediction models, the main objective of this dissertation is to propose methodologies to enhance the accuracy of bankruptcy prediction models with the objective to provide powerful risk management tools for those interested in the prediction of corporate bankruptcy. It must also be stated that, the methodologies proposed in this dissertation are extensively tested for robustness using a battery of tests and most importantly, are tested out-of-sample. This is important because the extensive tests ensure the stability of the performance of the resulted models and their reliability before their actual implementation, something that has been highlighted by regulatory authorities.

This dissertation focuses on the two most common methodological approaches for bankruptcy prediction. The first one refers to the development of empirical models. In such approach, one is interested to find empirical relationships between a set of predictor variables, such as accounting and market variables, and the likelihood of bankruptcy. This is achieved by training a model subject to an optimization criterion. This approach has been pioneered by Altman (1968) with the development of Z-score using discriminant analysis. Since then, other researchers have provided methodological enhancements, such as predicting bankruptcy with logistic regression (Ohlson, 1980), neural networks etc. The second approach is the structural approach where bankruptcy depends on the evolution of the capital structure of the firm and bankruptcy occurs when the value of firm's assets falls below a threshold, for instance, the liabilities of the firm. This approach has been pioneered by Merton (1974) who has used the options pricing framework of Black and Scholes (1973) to show that equity is equivalent to a European call option on the assets of the firm. In this dissertation, we propose methodologies which provide enhancements to both approaches described in this paragraph.

With that respect, the dissertation is divided in three chapters. In all chapters, we use data from a large number of non-financial U.S. bankrupt and healthy firms over the recent period. The main source of data comes from the database BankruptcyData which provides the name and date of bankruptcy filing, while financial and market data were obtained from Compustat and CRSP respectively at the year before bankruptcy filing. In the first chapter, we investigate the empirical performance of the Leland-Toft (1996) structural model. The specific model has received limited attention in the literature of bankruptcy prediction in the sense that it has not tested by the prior literature, despite that it is an extended version of a very widely-used structural bankruptcy model; the Black-Scholes-Merton model. Thus, in the first chapter, we are interested to examine whether such extensions offered by Leland-Toft that may be useful for bankruptcy prediction, such as the interest or coupon payment of the firm, provide improvements in the empirical performance of the structural models. Another objective of the first chapter, is to enhance the performance of traditional empirical bankruptcy models that are widely used in the literature, like Altman (1968) and Ohlson (1980), but also of more recent ones, such as Campbell et al. (2008), by including the probability of bankruptcy derived from Leland-Toft as additional predictor in those models, seeking to improve the estimation of bankruptcy risk relative to using the empirical models in isolation. Consistent with our expectations, the structural model of Leland-Toft, outperforms the Black-Scholes-Merton model in all tests. For instance, we find out-of-sample that Leland-Toft exhibits higher discriminatory power than the Black-Scholes-Merton but also it is associated with higher economic benefits for banks using Leland-Toft than Black-Scholes-Merton. However, it is found that none of the structural models can stand alone in predicting bankruptcy, since they are outperformed by other models (such as Ohlson, 1980; Campbell et al., 2008 etc). Most importantly, it is demonstrated that augmenting the empirical models of Ohlson (1980) and Campbell et al. (2008) with the probability derived by Leland-Toft as additional predictor, yield models with improved performance relative to the original models. In fact, these models, which we call them hybrid models, exhibit the highest predictive accuracy among all models considered in this chapter.

The second chapter is partly motivated by the first chapter but also on several other facts that we explain subsequently. Specifically, a result that is not explicitly discussed in chapter one is that, the higher the discriminatory power of the bankruptcy model used by a bank is, as measured by the Area Under ROC curve (AUROC), the higher the economic benefits for the bank. Furthermore, evidence in the extant literature suggests that banks using models with higher discriminatory power (i.e. higher AUROC) relative to other banks, have higher economic performance because they reject loans to "bad" firms and hence, they manage a healthier and more profitable credit portfolio. This result is evident even when there are minor differences in model performance measured by AUROC. Overall, there is evidence which shows the importance of using AUROC as performance measure and generally, it has been well-established as a performance statistic in academic studies but in industry as well (for instance, Moody's KMV extensively use AUROC before bringing their models into commercial practice).

Yet, it is not a common practice to train bankruptcy models to maximize AUROC (i.e. to obtain model coefficients by maximizing AUROC) but rather, it is used ex-post (after training the models using another maximization criterion, such as the log-likelihood function). As a response to this limited literature, this chapter contributes to the literature by introducing and comparing several methodologies to maximize the discriminatory power of bankruptcy prediction models, as measured by the widely-used AUROC statistic. In particular, by using models with probabilistic and linear response functions (i.e. when the output is a probability and a linear score respectively), we introduce several merit functions seeking to optimize AUROC. In other words, we obtain model coefficients when the AUROC is used as the optimization criterion during the training phase and contrast performance when the model is trained using the log-likelihood merit function. The key finding from this chapter is that, models trained to optimize AUROC, exhibit higher discriminating ability (higher AUROC), out-of-sample, relative to traditional approaches, like logistic regression models. From the models with linear response functions, a merit function which takes care of the outliers which often characterizes financial data, has the highest performance. However, from all models, a neural network model with a probabilistic response function is the best performing one. Consistent with expectations, in a simulated paradigm it is shown that banks which use the models with the highest AUROC, earn the highest profitability relative to other banks.

The third chapter is somewhat more independent in the sense that it is not a follow-up study from the previous chapters. However, it addresses a well-known issue underlying the structural approach for estimating firm default risk, which has attracted a lot of research the last decade. In particular, in the third chapter we focus exclusively on structural (parametric) models and we contribute to the literature by proposing a novel methodology to estimate the value of assets and the volatility of assets, which are the most important input parameters to the structural models for the estimation of the probability of default. These inputs, however, are not observed in the market thus making the estimation of the probability of default a challenging task to accomplish. In the literature, there are two main approaches to estimate the two parameters. The first approach is based on iteratively solving equations derived from options theory, which we call a "noisy" estimation approach, since convergence errors may affect the final outputs but also the relationships imposed to these unobserved parameters, are based on the restrictive assumptions from options theory. The second approach is based on "simplified" or ad-hoc approximations. Our methodology assumes the value of assets and the volatility to depend on some exogenous variables, x, through some unknown relationships. We use a nonparametric approach, such as neural networks, to learn the unknown relationships, aiming to obtain improved parameter values which enter the structural model, yielding a semiparametric model. With this respect, the Black-Scholes-Merton structural model is used a paradigm. Results in this chapter demonstrate that the out-of-sample performance of the semiparametric approach is significantly better relative to the alternative Black-Scholes-Merton specifications based on several tests, giving in that way validity to our proposed approach. Moreover, in this chapter we were motivated to augment the sample of bankruptcies, with financially distressed firms, given that financially distressed firms are more difficult to predict since it much more difficult to predict the beginning of the crisis. It is shown that the semiparametric approach shows an impressive performance relative to the competing Black-Scholes-Merton specifications. Interestingly, the semiparametric approach outperforms alternative approaches for default prediction, like the logistic regression approach as well as neural networks.

ACKNOWLEDGEMENTS

A long journey has come to the end and it is time to acknowledge those who contributed to my dissertation in their own way.

First from all, I am indebted to my research supervisors, Professors Chris Charalambous and Spiros Martzoukos who along the years of our collaboration, have been supporting me, guiding me and helped me a lot in finishing the dissertation. I would like to express my gratefulness to you for all the advices you gave to me that made me a better person, for your useful ideas and the challenging discussions we had which made my dissertation better. Your kindness, your integrity and the attention you gave to me on whatever issues I had, will be a bright example for me which I will always remember and will carry forward.

I would also like to thank the faculty and administrative members of the Department of Accounting and Finance who have created a very stimulating and pleasant environment during my studies. Special thanks to Evgenia and Tasoula for their excellent administrative support. Also, I would like to thank Professors Stavros Zenios, Nikolas Topaloglou and Christis Hassapis who accepted to be part of my examination committee.

I would like to express my appreciation to my classmates who shared this journey with me and made my studies more pleasant and interesting. I would like to thank you for finding some time to provide feedback to a couple of chapters of my dissertation during the departmental seminars.

Finally, I would like to express my love to my family for always standing by my side during my studies and giving me the strength to reach my goals. Thank you for teaching me not to give up and to believe in myself. I owe whatever I am today to you.

DEDICATION

To my parents, Andreas and Panayiota

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

CHAPTER 1

Predicting corporate bankruptcy using the framework of Leland-Toft: Evidence from U.S.

Abstract

In this paper, we evaluate an alternative approach for bankruptcy prediction that measures the financial healthiness of firms that have coupon-paying debts. The approach is based on the framework of Leland and Toft (1996), which is an extension of a widely-used model; the Black-Scholes-Merton model. Using U.S. public firms between 1995 and 2014, we show that the Leland-Toft approach is more powerful than Black-Scholes-Merton in a variety of tests. Moreover, extending popular but also contemporary corporate bankruptcy models with the probability of bankruptcy derived from the Leland-Toft model, such as Altman (1968), Ohlson (1980) and Campbell et al. (2008), yields models with improved performance. One of our tests, for example, shows that banks using these extended models, achieve superior economic performance relative to other banks. Our results are consistent under a comprehensive out-of-sample framework.

1 Introduction

Corporate bankruptcy prediction models are valuable risk management tools to assist bank managers in the decision-making process of identifying firms which are likely to fail and therefore would not be able to pay their obligations. This is because, the consequences arising from bankruptcy are enormous and include, for instance, economic ones such as the loss of the amount lent, impaired profitability for the bank which in certain cases may harm the viability of the bank, the financial system and the economy as well[1]. From the perspective of an investor, economic consequences include the loss of the wealth invested in bankrupt firms but also include non-economic ones, such as the loss of investors' confidence towards the financial markets. For these reasons, among others, it is important for the interested parties to develop and apply reliable corporate bankruptcy prediction models.

Much of the attention that has been given to the development of bankruptcy prediction models recently, is attributed to the global financial crisis period that hit the markets internationally in 2007, mainly due the consequences that the crisis left behind. Several economists even characterized the recent financial crisis at least as severe as the Great Depression period back to 1930's. As argued by Switzer et al. (2018), the 2007-2008 financial crisis, engendered huge losses to many firms, especially firms to the financial sector and its impact on financial stability, has attracted the interest of practitioners, scholars and policy-makers. Furthermore, another strand of the literature proposes mechanisms to reduce or at least control the risk of firms prior or during the crisis (Caprio et al., 2007; Gupta et al., 2013). Our study, is related to this strand of the literature, aiming to enhance the estimation of bankruptcy risk for firms and providing in that way proper risk management tools that serve as a companion to those interested to predict bankruptcy. We dedicate a separate section with results from the credit crunch in a subsequent section.

While various models have been proposed in the literature, two of the most frequently used by academics and practitioners are Z-score (Altman, 1968) and O-score (Ohlson, 1980). These models mainly use information from the financial statements of the firm to relate past performance with bankruptcy risk. More recently, models with both accounting and market

[1] For instance, Papakyriakou et al. (2019) show that the failure of financial institutions from U.S, negatively affect the international stock markets.

information have been developed. These models have the advantage to incorporate timely market information and thus the likelihood of bankruptcy can be updated more frequently[2]. Studies such as Shumway (2001), Chava and Jarrow (2004), Campbell et al. (2008) and Tinoco and Wilson (2013) show that accounting and market information yield models with improved performance. Another approach is the contingent claims-based approach which is based on the framework of Black and Scholes (1973) and Merton (1974). There, in an options pricing framework, the probability of bankruptcy is the probability that the assets value of the firm will be worth less than its debts, at maturity[3]. Such models are frequently referred to, as structural models.

In this paper, we evaluate an alternative approach for bankruptcy prediction, and we construct with it powerful bankruptcy models, seeking to improve the performance of existing models. Specifically, we evaluate an approach that measures the financial healthiness of firms with coupon-paying debts, using the framework of Leland and Toft (1996). Leland and Toft (1996) belongs to a class of models that extends Merton (1974), to incorporate the effects of taxes and bankruptcy costs to the valuation of equity and a corporate coupon-paying debt with finite maturity. Other significant features of their framework are that, bankruptcy can occur prior to the maturity of the debt but also, they consider the case when the bankruptcy point is determined endogenously. Thus, Leland and Toft is a more appropriate corporate model than Black-Scholes-Merton because it includes a richer information set about the firm which can be useful for bankruptcy prediction[4].

Several models are considered in this paper. Firstly, we compare the performance of two structural models; Leland-Toft and Black-Scholes-Merton. We believe that the former would outperform the latter since it is an extended version, containing more information for bankruptcy prediction. Next, we compare the performance of three reduced-form models with three hybrid models (i.e reduced-form models augmented with structural models). The first reduced-form model is Ohlson (1980) which is a comprehensive model since it includes various accounting variables such as profitability, liquidity, leverage, cash flows etc. Next, we augment Ohlson

[2] Refer to Agarwal and Taffler (2008) for a discussion between accounting and market information in bankruptcy prediction models.
[3] See for instance Bharath and Shumway (2008) and Afik et al. (2016) for related literature regarding this approach.
[4] A strand of the literature also examines empirically the performance of the structural models in predicting corporate bond prices and spreads and find that they do not accurately predict them (see for instance Lyden and Saraniti, 2000 and Eom et al., 2004 and references therein).

model with the probability of bankruptcy derived from Leland-Toft, yielding a hybrid model. We believe that augmenting this comprehensive accounting model with Leland-Toft will improve its performance, yielding a powerful bankruptcy prediction model. Another reduced-form model we examine is an extension of Altman (1968) model, which includes an additional cash flow variable and was suggested by Almamy et al. (2016). The authors find that augmenting Altman's model with a cash flow variable yields improved predictive ability for U.K. firms. We believe that augmenting Altman's model with Leland-Toft will further increase its predictive ability. Next, we use a competent reduced-form model that includes accounting and market variables suggested by Campbell et al. (2008). This model has been examined by Bauer and Agarwal (2014) and was found to outperform other approaches, such as reduced-form models with accounting information as well as the Black-Scholes-Meton structural model. Finally, we seek to improve Campbell et al. (2008), by including Leland-Toft in their model.

For our analysis we use 5460 U.S. public firms with coupon-paying debts between 1995 and 2014. The performance of the models is compared on three dimensions and our results are based on an extensive out-of-sample framework: 1) On their ability to discriminate bankrupt from healthy firms using Receiver Operating Characteristics (ROC) analysis 2) On their ability to predict bankruptcy probabilities close to actual or equivalently on their ability to empirically fit the data using log-likelihood statistics and 3) By measuring the economic performance of banks when they are competing to grant loans to individual firms and each bank uses a corresponding model to evaluate prospective firm-customers. For this last test, we employ the setting of Agarwal and Taffler (2008).

The key findings of the paper are that 1) Leland-Toft approach is more powerful than Black-Scholes-Merton. Sensitivity analysis tests for Leland-Toft shows that its forecasting power is not affected by the choice of parameter values underlying the model. However, none of the structural models can stand alone in forecasting bankruptcies since they are outperformed by reduced-form (and also hybrid) models, 2) Further increase in predictive ability is achieved when augmenting Altman's model with Leland-Toft rather than a cash-flow variable, 3) Augmenting the comprehensive models of Ohlson (1980) and Campbell et al. (2008) with Leland-Toft yields models with improved performance 4) Reduced-form models augmented with Leland-Toft, outperform reduced-form models augmented with BSM. In fact, the hybrid models which include Leland-Toft are the best performing models in all tests. Most importantly, in our experiment with the competitive loan market we find that banks using these extended models achieve superior economic performance relative to their competitors.

The paper proceeds as follows; Section 2 describes the bankruptcy models and the research hypotheses, section 3 discusses our data, section 4 discusses the methodology, section 5 reports the results and section 6 concludes.

2 Bankruptcy Models and Research Hypotheses

2.1 Structural Models

2.1.1 Black-Scholes-Merton

Merton (1974) shows that the equity value of the firm (*E*) can be viewed as a European call option underlying the assets of the firm (*V*) and with strike price the zero-coupon debt of the firm (*D*). The Black and Scholes (1973) options pricing formula can therefore be used to price the equity of the firm:

$$E = VN(d_1) - De^{-rT}N(d_2) \tag{1}$$

where r is the riskless rate of return, $N(.)$ is the standard normal distribution function, T is the maturity of the debt and d_1, d_2 are defined as follows:

$$d_1 = \frac{ln\left({}^{V}/_{D}\right) + (r + 0.5\sigma_V{}^2)T}{\sigma_V\sqrt{T}} \tag{2}$$

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

and σ_V is the volatility of assets value returns. In the framework of Black-Scholes-Merton, the firm goes bankrupt when $V < D$ and thus the probability of bankruptcy, $prob(V<D)$, is the probability that at debt maturity, the assets value is lower than the debt. The probability of bankruptcy is then given by the Black-Scholes-Merton (BSM hereafter) formula:

$$prob = N\left(-\frac{ln\left({}^{V}/_{D}\right) + (\mu - 0.5\sigma_V{}^2)T}{\sigma_V\sqrt{T}}\right) \tag{4}$$

where μ is the return of assets.

2.1.2 Leland and Toft (1996)

Leland and Toft (1996) extends the framework of Merton (1974) to incorporate the effects of taxes and bankruptcy costs in the valuation of a corporate risky debt with finite maturity. Their framework considers the valuation of debt that pays coupons as opposed to the framework of Merton where the firm issues a zero-coupon debt. In this context, Leland-Toft derive closed-form solutions for the market value of equity, debt and total firm value. Most importantly, they consider the case where bankruptcy is determined endogenously as opposed to Merton (1974) where bankruptcy is determined exogenously. This consideration allows the calculation of an optimal bankruptcy point which is chosen by the management of the firm in favor of shareholders such that the equity value is maximized. When assets value reaches that point, it is optimal, from shareholders´ perspective, for the firm to file for bankruptcy. In contrast, when bankruptcy is determined exogenously, the bankruptcy point is chosen arbitrarily[5]. However, this consideration is suboptimal because firms usually continue operations even when assets value falls below firm's debt and practically there is not an agreed value to use. Eq. (5) shows the calculation of the bankruptcy point, VB_{LT}, underlying Leland-Toft model which is a key determinant of the bankruptcy probability[6]:

$$VB_{LT} = \frac{\left(\frac{C}{r}\right)\left(\frac{A}{rT} - B\right) - A\frac{P}{rT} - \tau\frac{Cx}{r}}{1 + cx - (1-c)B} \tag{5}$$

where

$$A = 2\alpha e^{-rT} N\left(a\sigma_V\sqrt{T}\right) - 2zN\left(z\sigma_V\sqrt{T}\right) - \frac{2}{\sigma_V\sqrt{T}} n\left(z\sigma_V\sqrt{T}\right) + \frac{2e^{-rT}}{\sigma_V\sqrt{T}} n\left(\alpha\sigma_V\sqrt{T}\right) + (z - a)$$

$$B = -\left(2z + \frac{2}{z{\sigma_V}^2 T}\right) N\left(z\sigma_V\sqrt{T}\right) - \frac{2}{\sigma_V\sqrt{T}} n\left(z\sigma_V\sqrt{T}\right) + (z - a) + \frac{1}{z{\sigma_V}^2 T}$$

$$a = \frac{(r - \delta - 0.5{\sigma_V}^2)}{{\sigma_V}^2}, \qquad z = \frac{\sqrt{a^2{\sigma_V}^4 + 2r{\sigma_V}^2}}{{\sigma_V}^2}, \qquad x = a + z$$

with N(·) and n(·) denoting the cumulative standard normal distribution and standard normal density functions respectively. A closer examination shows that Eq. (5) is a function of eight

---

[5] For example, in the Merton's model the bankruptcy point is the debt of the firm and thus, is determined exogenously.

[6] Hilberink and Rogers (2002), extend Leland-Toft (1996) to allow for sudden jumps in the asset value, *V*, and derive a new optimal bankruptcy point. However, the solutions are not explicit and some of the parameters are not straightforward to compute (see Eq. (3.16) and Eq. (3.23) in their paper).

parameters: risk-free rate (*r*), tax rate (τ), coupon payments (*C*), bankruptcy costs (*c*), volatility of assets (σ_V), debt principal (*P*), payout yield (δ), and debt maturity (*T*).

To evaluate bankruptcy risk in discrete points of time t, where t ≤ T, we need to define a cumulative distribution function. In the framework of Leland-Toft (LT hereafter), the probability that the current value of firm's assets, *V*, will fall to the bankruptcy point, VB_{LT}, for the first time at time t is obtained from Leland (2004) and defined as:

$$prob(t) = N(X) + e^{Y}N(Z) \tag{6}$$

where

$$X = \frac{-ln\left(\frac{V}{VB_{LT}}\right) - (\mu - \delta - 0.5{\sigma_V}^2)t}{\sigma_V\sqrt{t}}, Y = \frac{-2ln\left(\frac{V}{VB_{LT}}\right)(\mu - \delta - 0.5{\sigma_V}^2)}{{\sigma_V}^2}$$

$$Z = \frac{-\ln\left(\frac{V}{VB_{LT}}\right) + (\mu - \delta - 0.5{\sigma_V}^2)t}{\sigma_V\sqrt{t}}$$

Finally, *t* is the forecasting horizon, which in our case is one year.

2.1.3 Estimating Asset Values and Volatilities

The most important inputs to LT and BSM models are the value of assets and the volatility of assets returns which are not observed. In the context of options pricing, however, the following two non-linear equations can be solved iteratively to obtain the two variables of interest:

$$V = \frac{E - De^{-rT}N(d_2)}{N(d_1)} \tag{7}$$

$$\sigma_V = \frac{E\sigma_E}{VN(d_1)} \tag{8}$$

where σ_E is the volatility of equity returns that is directly estimated from daily equity data. The above iterative procedure, which we use to estimate the two unobserved inputs, is the standard approach for the estimation of asset value and volatility and has also been used by Eom et al. (2004), Hillegeist et al. (2004), Campbell et al. (2008), while Vassalou and Xing (2004) use a variation of the above iterative process[7].

[7] For other approaches, see Bharath and Shumway (2008), Charitou et al. (2013) and Afik et al. (2016)

2.2 *Reduced-Form Models*

Several reduced-form models are also considered. Ohlson (1980) is a model which relates bankruptcy with a set of accounting-based variables, defined as follows:

$$Ohlson = f\,(SIZE, TLTA, WCTA, CLCA, D(TL>TA), NITA, CFOTL, D\,(NI_t+NI_{t-1}<0), CHINI) \quad (9)$$

Next, we consider Almamy, Aston and Ngwa (2016), which we refer as *AAN*. This model is an extension of Altman's model which incorporates a cash flow variable as additional predictor and it is defined as follows:

$$AAN = f\,(WCTA, RETA, EBITTA, MVTL, SLTA, CFOTA) \quad (10)$$

Moreover, we consider the model proposed by Campbell, Hilscher and Szilagyi (2008), which we refer to as *CHS*. This model is a mixture of accounting ratios, scaled by the market value of assets, and other market information as predictors, defined as follows:

$$CHS = f\,(NIMTA, TLMTA, EXRET, SIGMA, RSIZE, CASHMTA, MB, PRICE) \quad (11)$$

The definition of the variables is in table 1.

[Insert Table 1 here]

2.3 *Hybrid Models*

Finally, we incorporate the probability of bankruptcy derived from LT as additional predictor in Ohlson, Altman and CHS models, yielding the following hybrid models which we refer to as *OLT*, *ALT* and *CHSLT* respectively:

$$OLT = f\,(SIZE, TLTA, WCTA, CLCA, D(TL>TA), NITA, CFOTL, D\,(NI_t+NI_{t-1}<0), CHINI, LT) \quad (12)$$

$$ALT = f\,(WCTA, RETA, EBITTA, MVTL, SLTA, LT) \quad (13)$$

$$CHSLT = f\,(NIMTA, TLMTA, EXRET, SIGMA, RSIZE, CASHMTA, MB, PRICE, LT) \quad (14)$$

2.4 Research Hypotheses

LT is an extended version of Merton's model with less restrictive assumptions and a richer information set about the firm. Therefore:

Hypothesis 1: LT is a better alternative approach than BSM

Prior research suggests that accounting and market information should be included in corporate bankruptcy prediction models since they provide complementary information. For instance, variables such as the volatility of equity and excess equity returns improve the performance of accounting-based models (Chava and Jarrow, 2004; Hillegeist et al., 2004; Agarwal and Taffler, 2008; Tinoco and Wilson, 2013 etc.). Therefore, we expect that including LT in Ohlson model will enhance its performance:

Hypothesis 2: Incorporating LT as additional predictor in Ohlson, yields a model with improved performance.

Hence, the model in Eq. (12) should outperform the model in Eq. (9). An extension of Hypothesis 2, is as follows:

Hypothesis 2a: Ohlson model augmented with LT, will outperform Ohlson model augmented with BSM.

Almamy et al. (2016) suggest that augmenting Altman's model with a cash-flow variable, increases its predictive ability. However, further increase in predictive ability could be obtained when augmenting Altman's model with a predictor that measures the financial healthiness of firms with coupon-paying debts. Hence, the model in Eq. (13) should outperform the model in Eq. (10).

Therefore:

Hypothesis 3: Augmenting Altman's model with LT will further increase predictive ability than a cash-flow variable.

An extension of Hypothesis 3, is as follows:

Hypothesis 3a: Altman's model augmented with LT, will outperform Altman's model augmented with BSM.

Campbell et al. (2008) find that augmenting their model with BSM, doesn't yield improved performance, arguing that all the information incorporated in BSM, such as returns and volatilities, are already included in their model. Since LT is an extension of BSM that includes additional information, we want to investigate if augmenting Campbell et al. (2008) with LT, would improve its performance. This leads to the fourth hypothesis:

Hypothesis 4: Incorporating LT as additional predictor in Campbell et al. (2008), yields a model with improved performance.

Finally, an extension of the fourth hypothesis, is the following:

Hypothesis 4a: Augmenting Campbell et al. (2008) with LT, will outperform Campbell et al. (2008) augmented with BSM

3 Data

3.1 Sample

We analyze a sample of 5460[8] U.S. public firms from which 333 filed for bankruptcy in a specific year between the 20-year period 1995-2014. Bankruptcy filings were identified from BankruptcyData[9] and include firms which filed for bankruptcy under Chapter 7 or Chapter 11. To avoid problems related to sample selection bias and increase efficiency of regression estimates, we collect all available observations in the selected period for each bankrupt and healthy firm. This practice increases our sample to 39830 firm-year observations. Furthermore, once a firm files for bankruptcy, future observations for that firm are excluded but past observations for all bankrupt firms are included in our sample (i.e. before a firm file for bankruptcy, it is considered as healthy).

Table 2 presents the distribution of observations across the years.

[Insert Table 2 here]

In general, bankruptcy rate in all years is less than 1% except for years 1999 (1.493%) and the mid-crisis years 2008 and 2009 with the bankruptcy rate being at its peak (1.190% and 2.133%

[8] The framework of Leland and Toft (1996) applies for firms with coupon-paying debt. Thus, we keep only firms which have interest payments in their income statements

[9] Available at http://www.bankruptcydata.com/findabrtop.asp

respectively). The average bankruptcy rate in the sample is 0.836% indicating the fact that bankruptcy is a rare event.

Similar with Bharath and Shumway (2008), Afik et al. (2016) and others, we exclude financial firms (SIC 6000-6799) due to the different nature of their operations and structure of their financial statements relative to other industrial firms. Firms are classified into a specific industry according to the Standard Industrial Classification (SIC) code provided by the United States Department of Labor. Table 3 shows the distribution of observations across industries.

[Insert Table 3 here]

Most of observations (53%) comes from the Manufacturing sector and then from Services, Transportation, Retail and Mining sectors, accounting for 16.42%, 10.36%, 8.41% and 5.87% of the sample respectively, whereas the Wholesale, Construction, Public Administration and Agriculture sectors account for the smallest proportions of the sample (4.03%, 0.95%, 0.62% and 0.35% respectively).

3.2 Variables Construction

To construct the relevant variables used in the structural and reduced-form models, we collect annual financial data from Compustat and daily equity data from CRSP. All variables are constructed at the fiscal year-end, prior to the year of bankruptcy.

To estimate the value of assets and the volatility with the iterative process described earlier, we need the market value of equity and the (annualized) volatility of equity return. For the first, we take the stock price at the fiscal year-end and multiply it with the number of shares outstanding. For the latter, we follow Campbell et al. (2008) by taking the standard deviation of stock returns for the last 30 days, prior to fiscal year-end. For the face value of debt (D), we follow convention in the literature, and we set it equal with short-term debt plus half of long-term debt (also used in Vassalou and Xing, 2004 and Campbell et al., 2008). The prediction horizon is one, thus T for the BSM and t for the Leland-Toft models equal 1. Another input to the structural models, is the assets value returns (μ). Campbell et al. (2008) use an equity risk-premium equal to the riskless rate plus 6% for all firms. However, we believe that using a common return for all firms, would undermine the predictive ability of the structural models. A better alternative would be to use a firm-specific return. A reasonable proxy, which we use in our study, is the annualized return of equity, also used by Bharath and Shumway (2008). Afik et al. (2016), instead, suggest using the maximum between the riskless rate of return and equity return. However, this specification would overstate the asset drift for firms with negative prospects (i.e. the bankrupt firms), undermining again the predictive ability of the structural models.

For the risk-free rate (r), the one-year Treasury Constant Maturity rate is used, obtained from Federal Reserve[10]. For the coupon payments (C) and debt principal (P), the interest expense and the short-term debt plus half of long-term debt are used as proxies respectively and the payout yield (δ) is defined as the sum of coupon payments plus dividends (ordinary and preferred) divided by the market value of assets. For corporate tax rate (τ), bankruptcy costs (c) and maturity of debt (T) we follow Leland (2004) who sets these parameters equal to 15%, 30% and 10 years respectively. However, as shown later and specifically in Appendix A, results are not sensitive with respect to different parameters choices and thus Leland-Toft is stable as far as its performance is concerned (refer to section 4 about model performance measures). All inputs of the models are summarized in Table 1.

4 Methodology

This section describes the methodology that is used to assess the performance of the bankruptcy prediction models.

4.1 Discriminatory Power

Discriminatory power refers to the ability of a model to discriminate bankrupt from healthy firms. The ROC curve is a graphical representation of the discriminatory power of a bankruptcy prediction model. It plots the true predictions on the vertical axis (the percentage of bankrupt firms correctly classified as bankrupt) against the false predictions on the horizontal axis (the percentage of healthy firms incorrectly classified as bankrupt) according to a pre-determined cut-off value. By performing this classification procedure for multiple cut-off values, we create a set of points which together constitute the ROC curve. Ideally, a perfect model will never make false predictions and will always correctly classify the bankrupt firms, for any level of cut-off point. Hence the ROC curve of a perfect model will pass through the point (0, 1) and in

[10] Available at http://www.federalreserve.gov/releases/h15/data.htm

general, the closer the ROC curve towards the top-left corner of the graph, the better the discriminatory power is.

A quantitative assessment of the discriminatory power of a bankruptcy prediction model is the Area Under ROC (AUROC) curve (see for example Hanley and McNeil, 1982 and Sobehart and Keenan, 2001) and calculated as follows:

$$\widehat{AUROC} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\psi\left(p_B^i, p_H^j\right) \quad (15)$$

where

$$\psi\left(p_B^i, p_H^j\right) = \begin{cases} 1, & p_B^i > p_H^j \\ 0.5, & p_B^i = p_H^j \\ 0, & p_B^i < p_H^j \end{cases}$$

and p_B^i is the bankruptcy probability of the *i-th* bankrupt firm, p_H^j is the bankruptcy probability of the *j-th* healthy firm, *n* is the number of bankrupt firms and *m* is the number of healthy firms in our sample.

We test for statistically significant differences between the AUROCs of two models. The hypothesis is as follows:

$$H0\colon AUROC\ 1 - AUROC\ 2 = 0 \quad Vs \quad H1\colon AUROC\ 1 - AUROC\ 2 \neq 0$$

We use the non-parametric approach of DeLong et al. (1988), which accounts for the correlation of the AUROCs produced by the two models. The construction of the test statistic is described in Appendix B.

4.2 Logit Models

The logistic regression approach is used to estimate the models in Eqs. (9)-(14). Specifically, we estimate the following logit model:

$$P\left(Y_{i,t+1} = 1 | X_{i,t}\right) = p_{i,t} = \frac{e^{a+\beta' X_{i,t}}}{1 + e^{a+\beta' X_{i,t}}} \quad (16)$$

where $p_{i,t}$ is the probability of bankruptcy at time "*t*", that the "*i-th*" firm will go bankrupt the next year, $Y_{i,t+1}$ denotes the status of the *i-th* firm (1 if it goes bankrupt at time *t+1*, 0 otherwise),

$X_{i,t}$ is the vector of covariates of the *i-th* firm at time *t*, β is the vector of coefficient estimates and a is the constant term which expresses the bankruptcy risk in the absence of the covariates.

The model in Eq. (16) represents a multi-period logit model because it includes observations for each firm across time. However, the inclusion of multiple firm-year observations per firm yields understated standard errors because the log-likelihood objective function, which is maximized to estimate the multi-period logit model, assumes that each observation is independent from each other. This is a wrong assumption since financial information of a firm at time *t+1* cannot be independent from the financial information of the same firm at *t*. Failing to address this econometric issue, leads to wrong inference regarding the significance of the individual coefficients. Similar with Filipe et al. (2016), we use clustered-robust standard errors to adjust for the number of firms in the sample but also for heteroskedasticity (Huber,1967 and White, 1980).

To compare the predictive accuracy of various logit models, we test the difference between their log-likelihoods. Hence, the hypothesis takes the following general form:

$$H_0: L_1(k_1) - L_2(k_2) = 0 \quad V_s \quad H_1: L_1(k_1) - L_2(k_2) \neq 0$$

where $L_1(k_1)$ is the log-likelihood of the first model with k_1 parameters and $L_2(k_2)$, is the log-likelihood of the second model, with k_2 parameters and $k_1 > k_2$. The construction of the test statistic for different types of logit models can be found in Appendix C.

4.3 Economic Analysis of Bankruptcy Models

The analysis so far addressed the accuracy of our bankruptcy models. But how accuracy is economically beneficial for banks? Here, we follow the approach of Agarwal and Taffler (2008) to examine it by assuming a loan market worth $100 billion and banks competing for granting loans to individual firms. Each bank uses one of our bankruptcy models to evaluate the credit worthiness of their customers.

4.3.1 Calculating Credit Spreads

We estimate the models using data spanning the years 1995-2005 (70% of the sample). We sort firm-customers from this sample in 10 groups of equal size and a credit spread is calculated according to the following rule; Firms in the first group, which are firms with the lowest bankruptcy risk, are given a credit spread, *k*, and firms in the remaining groups are given a credit spread, CS_i, obtained from Stein (2005) and Blochlinger and Leippold (2006) and it is defined as follows:

$$CS_i = \frac{p(Y=1|S=i)}{p(Y=0|S=i)} LGD + k \tag{17}$$

where $p(Y=1|S=i)$ and $p(Y=0|S=i)$, is the average probability of bankruptcy and non-bankruptcy 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 of firms that went bankrupt the following year divided by the number of firms in the group. Furthermore, *k*=0.3% and *LGD*=45%.

4.3.2 Granting Loans and Measuring Economic Performance

To evaluate economic performance, we assume that banks compete to grant loans to prospective firm-customers between the period 2006-2014. Each bank uses one of our bankruptcy models which have been estimated in the period 1995-2005. The bank sorts those customers according to their riskiness and rejects the bottom 5% with highest risk. The remaining firms are classified in 10 groups of equal size and firms from each group are charged a credit spread that has been obtained from the period 1995-2005. Finally, for each customer the bank that charges the lowest credit spread is granting the loan. Two measures of profitability are used. The first one, Return on Assets (ROA) is defined as Profits/Assets lent and the second one, Return on Risk-Weighted Assets (RORWA) takes into consideration the riskiness of the assets, defined as Profits/Risk-Weighted Assets. Risk-Weighted Assets are obtained from formulas provided by the Basel Committee on Banking Supervision (2006)[11].

5 Results

5.1 Descriptive Statistics

Table 4 reports mean values of the explanatory variables for the entire sample 1995-2014. As expected, the performance of bankrupt firms is worse than the performance of healthy firms, one year prior to bankruptcy with the differences in mean values being statistically significant

[11] See for instance the Appendix in Bauer and Agarwal (2014)

in most cases. For example, bankrupt firms; are less profitable (*EBITTA* is lower), have more leverage (*TLTA* is higher), their liquidity is more constrained (*WCTA* is lower) etc. In terms of the market performance, stock price of bankrupt firms is significantly more volatile than healthy firms (*SIGMA* is higher), they underperform the market (*EXRET* is lower) as well stock price is lower (*PRICE* is lower). Our variable of interest, *LT*, is higher for bankrupt firms relative to healthy firms

[Insert Table 4 here]

5.2 Reduced-Form and Hybrid Models Estimation

Table 5 reports estimation results for our models when applying the logistic regression approach on our data.

[Insert Table 5 here]

Here, the estimation sample spans the years between 1995 and 2005 which accounts for approximately 70% of our sample. Firstly, most of Ohlson and CHS variables are significant and with the correct sign. Noticeable exception is the case of *SIGMA* where in previous studies (Campbell et al., 2008 and Bauer and Agarwal, 2014) was found significant. Based on an analysis we have performed, we conclude that in our case the interaction of *SIGMA* with other market variables in the CHS model is the main cause for this kind of behavior. For example, when we include *SIGMA* individually or in the Ohlson model, is statistically significant and with the correct (positive) sign. Furthermore, in the estimation sample, average *SIGMA* (not tabulated) for bankrupt firms is 1.21 while for healthy firms is 0.60, which excludes possible data collection error. Another relatively odd estimation result is the positive coefficient for the *SIZE* variable[12]. According to Galil and Gilat (2018), a positive sign of this variable may hint on a selection bias in the bankruptcy sample toward larger corporates. However, this is not the case with our sample. In the estimation sample (whole sample), average total assets for bankrupt firms are 679.1 million (662.5 million), while for the healthy group, average total assets are 2074.5 million (3455.3 million). We believe that its interaction with other variables is the main cause for this result (including *SIZE* only in a logistic regression, yields a statistically significant coefficient with the correct (negative) sign.

[12] Hillegeist et al. (2004) also find a positive *SIZE* coefficient in the Ohlson model

Secondly, most of Altman variables included in AAN are not significant, consistent with Hillegeist et al. (2004). The cash flow ratio enters significantly and with the correct sign. Finally, the predictor of interest, LT, is highly statistically significant (α=1%) when incorporated in Altman (ALT), Ohlson (OLT), and Campbell (CHSLT) models.

5.3 Evaluating Leland-Toft and BSM (Hypothesis 1)

First, we compare the performance of the two structural models. For consistency, we estimate two logit models; The first includes the probability of bankruptcy derived from Leland-Toft in the period 1995-2005 as predictor and the second includes the probability of bankruptcy derived from BSM. Using these models, we forecast bankruptcies in the out-of-sample period, 2006-2014[13]. The performance is reported in Table 6.

[Insert Table 6 here]

We find that AUROC of LT is 0.8941 while for BSM is 0.8659, indicating that LT has better out-of-sample discriminating ability. Using the Delong test we find that the difference is statistically significant at α=1% (test statistic=2.74). Moreover, in Appendix A we perform a sensitivity analysis test to examine whether the AUROC of LT is affected by deviations in the choice of parameter values. In all, results suggest that LT is not sensitive as far as the ordinal ranking (AUROC) is concerned. Further, LT model explains bankruptcy variation better than BSM, according to pseudo-R^2 (19.72% and 17.90% respectively), although differences in their log-likelihoods are not significant (test statistic is 1.38).

However, as it is evident from Table 6, neither LT, nor BSM are sufficient statistics to forecast bankruptcies, since they are outperformed by other models such as Ohlson and CHS (differences in AUROCs and log-likelihoods are significant)[14]. Thus, none of the structural models can stand alone.

Finally, we perform an analysis of the economic benefits when banks use either LT or BSM in evaluating the credit worthiness of prospective customers. We make the paradigm more challenging by using Altman's model as a benchmark. Hence, we assume there are three banks

[13] This adjustment in the bankruptcy probability derived from structural models through a logit regression, is usually referred to as calibration.

[14] This result is also evident by the regression results in table 5, since LT enters significantly in Ohlson, Altman and CHS along with other variables, suggesting that individually, it doesn't capture all the bankruptcy-related information.

competing for loans; bank 1 uses LT, bank 2 uses BSM and bank 3 uses Altman's model. The results are reported in table 7:

[Insert Table 7 here]

As can be inferred, the quality of the loan portfolio for bank 1 which uses LT is the best among the three banks, since there are only 10 bankruptcies (0.25%), whereas there are 37 and 38 bankruptcies in the portfolios of banks 2 and 3 respectively, corresponding to 0.79% and 0.76% bankruptcy rate. Most importantly however, is that bank 1 generates superior economic performance relative to its competitors. For instance, on a risk-adjusted basis, bank 1 yields 1.09% return on the capital it has invested while bank 2 generates 0.64% and the return for bank 3 is 0.54%[15].

From the analysis in this section, we conclude that a bank has more gain by using LT rather than BSM which in fact, lends support to our first hypothesis, indicating that LT is a better approach than BSM, due to the richer information set incorporated in LT.

5.4 Reduced-Form and Hybrid Model Performance (Hypotheses 2-4)

In this section, we test the performance of the models using three out-of-sample approaches, as well as the economic benefits when banks adopt the models in a competitive loan market, as outlined below.

5.4.1 Baseline Approach

Here, we use the models (as estimated in table 5) to forecast bankruptcies in the out-of-sample period which spans the years between 2006 and 2014. Results are reported in table 6.

Panel A reports the performance of the models while panels B and C test for differences in their discriminating ability and predictive accuracy respectively. Firstly, OLT performs better than Ohlson (AUROCs are 0.9449 vs 0.9252 and log-likelihoods are -483.43 vs -535.57). The differences are statistically significant (test statistics are 4.73 and 104.28 respectively) which lends support to our second hypothesis, that extending Ohlson with LT yields improved performance. Secondly, ALT outperforms AAN (AUROCs are 0.9207 vs 0.8597 and log-

[15] Results are robust with respect to various specifications of LGD (0.4-0.7) but k as well (0.002-0.004)

likelihoods are -519.56 vs -603.81). The differences are statistically significant (test statistics are 4.78 and 6.51 respectively) which is in line with our third hypothesis. That is, augmenting Altman's model with LT, further improves performance relative to a cash flow variable as suggested by Almamy et al. (2016). Including LT in CHS slightly improves discriminating ability (AUROCs are 0.9395 for CHSLT and 0.9332 for CHS), though their differences are not statistically significant (test statistic is 1.47). On the other hand, LT carries incremental information as indicated by their log-likelihoods, meaning that it is a missing variable for the model (log-likelihoods for CHSLT and CHS are -491.41 and -498.85 respectively), with the difference being statistically significant (test statistic is 14.88). Thus, we provide evidence regarding the fourth hypothesis, that augmenting CHS with LT yields improved performance. We complement the aforementioned results, with ROC curves provided in figure 1.

[Insert Figure 1 here]

A related performance statistic with AUROC is the partial AUROC (pAUROC)[16], which is based on a specified region of the area under ROC curve that might be of practical interest (see for instance Dodd and Pepe, 2003)[17]. Panel A in table 6 reports pAUROCs for the models. We have also tested for differences in pAUROCs, but we do not report the results to save space. Overall, the conclusions are similar with the case of AUROC, giving validity to hypotheses 1-4.

5.4.2 Rolling Window Approach

As a second way to test the models, we re-estimate them yearly based on a rolling window. For instance, we estimate the models using firms between 1995 and 2005 and apply them on firms in 2006. Then we re-estimate the models using firms between 1996 and 2006 and apply them on firms in 2007. This process is repeated until 2014 and we aggregate bankruptcy probabilities obtained from each year to measure the performance of the models. It should be noted that this approach should be used in practice because the models are updated more frequently as new information becomes available. Results are reported in table 8.

[16] We thank an anonymous referee for this suggestion.
[17] The selection of the partial region under the ROC curve, however, is subjective. We rely on STATA's default specifications to estimate pAUROCs and to test for significant differences.

[Insert Table 8 here]

We obtain similar results as before and summarized as follows: 1) Augmenting Ohlson with LT yields a model with improved performance as indicated by AUROC statistics (0.9469 for OLT vs 0.9289 for Ohlson) as well as log-likelihood statistics (-470.37 for OLT vs -522.87 for Ohlson). Differences are statistically significant (test statistics are 4.33 and 104.99 for the two performance tests respectively) which is consistent with hypothesis 2, 2) Incorporating LT in Altman's model further improves its performance as opposed to a cash flow variable (AUROC for ALT is 0.9253 vs 0.8673 for AAN, while log-likelihoods are -508.25 and -593.86 respectively). Differences in their performance are significant (test statistics for the two tests are 4.62 and 6.28 respectively) which provides evidence to support our third hypothesis and 3) LT incorporates information not included in CHS (log-likelihood for CHSLT is -479.66 while for CHS is -485.94, and test statistic is 12.57) which is in line with the fourth hypothesis. AUROC improvement is not significant.

5.4.3 Five Folds Approach

Here, we divide the whole sample period in five approximately equal-sized sub-samples. We use any four of them to estimate the models and apply them on firms in the left-out sample. This is to break the chronological order of the data, and to consider different periods as well, such as periods before the financial crisis period. Then we aggregate bankruptcy probabilities from each left-out sample to obtain single performance measures. Results are reported in table 9.

[Insert Table 9 here]

As expected, performance according to this approach is lower since we use data from different periods to make predictions, missing therefore potential trends. Despite this fact, we obtain similar insights as with the two previous approaches. Discriminating ability is improved when we include LT as additional predictor in Ohlson (AUROCs are 0.9091 for OLT and 0.8939 for Ohlson) while predictive accuracy is also better (log-likelihoods are -1379.48 for OLT and -1477.06 for Ohlson). Differences are statistically significant (test statistics are 4.56 and 195.16 for the two tests respectively) which is consistent with our second hypothesis regarding the superiority of this extended version of Ohlson's model. Next, ALT outperforms AAN as indicated by AUROC statistics (0.8826 vs 0.8438 respectively) as well as log-likelihood statistics (-1471.35 vs -1630.45 respectively), meaning that LT further improves performance when included in Altman's model as opposed to a cash flow variable, supporting therefore our third hypothesis. Differences in performance are significant (test statistics are 5.63 and 6.62 for the two tests respectively). Finally, as with the previous tests we find evidence that LT improves the performance of CHS. Specifically, log-likelihood for CHSLT is -1403.72 while for CHS is -1417.16. Difference is statistically significant (test statistic is 26.87) while discriminating ability, measured by AUROC, is only slightly improved.

5.4.4 Economic Benefits

So far, we have considered aspects of model performance such as discriminating ability, measured by AUROC, and empirical fit, measured by log-likelihood statistics. However, a bank is more interested in the economic benefits when using bankruptcy models in the decision-making process of granting loans to individual firms. Here, we show the case of five banks, where bank 1 uses OLT, bank 2 uses CHSLT, banks 3 and 4 use CHS and Ohlson respectively, whereas bank 5 uses a benchmark model such as Altman's model. Table 10 reports information regarding the economic results of these banks.

[Insert Table 10 here]

Clearly, the quality of the loan portfolio for banks 1 and 2 which use OLT and CHSLT respectively, is superior relative to that of banks 3 and 4 which use CHS and Ohlson respectively. This is evident by the lower concentration of bankruptcies in their portfolios (0.11% for bank 1 and 0.16% for bank 2) relative to other banks (0.44% for bank 3 and 0.58% for bank 4).

The most important result, however, is that banks 1 and 2 earn higher returns than the other banks on a risk-adjusted basis (i.e. after adjusting for the riskiness of the assets lent). For instance, for each dollar invested, banks 1 and 2 earn 1.74% and 1.54% respectively on a risk-adjusted basis, whereas the competing banks earn lower returns (1.12% for bank 3 and 1.02% for bank 4). Bank 5 which uses a generic bankruptcy model earns negative returns. It is also worth noting that differences in discriminating ability that we have not found to be statistically significant are reflected in the economic results[18]. For instance, both banks 1 and 2 that use OLT

[18] Bauer and Agarwal (2014) also reported that very small differences in AUROCs are shown up in the economic results

and CHSLT respectively, are more profitable than bank 3 which uses CHS, although there are no significant differences in their AUROCs, as reported in table 6. Thus, banks should take into consideration the economic benefits when judging what bankruptcy model to use.

Based on these results[19], we conclude that banks using OLT and CHSLT, can achieve superior economic performance relative to other banks that use, for instance, CHS or Ohlson.

5.5 Augmenting CHS, Ohlson and Altman with LT and BSM (Hypotheses 2a-4a)

Campbell et al. (2008) show that augmenting their model with BSM doesn't yield improved performance based on pseudo-R^2, indicating that all the information incorporated in BSM is already included in their model. Here, we re-examine this insight and compare it with the case of LT. Table 11 reports the results.

[Insert Table 11 here]

Indeed, augmenting CHS with BSM doesn't improve performance, since volatility and return measures are already included in the model (log-likelihood and pseudo-R^2 are the same when compared to CHS. Differences in model performance are not statistically significant according to test-statistics reported in panel A and B. However, this is not the case when we include LT. Specifically, pseudo-R^2 increases to 34.35% as well as log-likelihood (difference with CHS is statistically significant at α=1%), indicating that LT provides additional information not included in CHS. As expected though, we document increase in performance when augmenting Ohlson and Altman with LT and BSM, since the two reduced-form models do not incorporate market information (all test statistics for performance difference are significant at α=1%).

Finally, we find that hybrid models with LT outperform hybrid models with BSM. For example, log-likelihood for CHS augmented with LT is -491.41 while for CHS augmented with BSM is -497.96 (Vuong's test statistic is significant at α=5%). AUROC for the first, is slightly higher (test statistic is not significant). Similar is the case with Ohlson and Altman model. Augmenting these models with LT yields models with better performance relative to augmenting them with BSM (Delong's test for AUROCs and Vuong's test for log-likelihoods

[19] Results are robust with respect to different parameter specifications, for example, setting LGD = 0.4-0.7 and k=0.002-0.004

are significant at α=5%). In all, we find evidence in favor of hypotheses 2a-4a; hybrid models that incorporate LT, outperform those with BSM.

5.6 Time Robustness

In this test, we estimate the reduced-form and hybrid models in the period 2006-2014 and forecast bankruptcies in the period 1995-2005. This is to test the performance of the models in a completely new sample, since the previous tests, included data from the recent period to measure performance i.e. 2006-2014. Results are reported in table 12.

[Insert Table 12 here]

In all, the results support the superiority of the hybrid models over the reduced form models, suggesting that the LT is a significant addition to the models[20]. For instance, OLT outperforms Ohlson (differences in AUROCs and log-likelihoods are both significant at α=5% and α=1% respectively), CHSLT outperforms CHS (differences in log-likelihoods are significant at α=1%) and finally, ALT outperforms AAN (differences in AUROCs and log-likelihoods are significant at α=1%).

5.7 Focusing on the crisis period 2007-2009

The purpose of this section is to shed light on the performance of our bankruptcy models during the financial crisis period 2007-2009[21] (results not tabulated). With respect to this test we find qualitatively similar results with our previous tests, suggesting that the hybrid models performed better relative to the reduced-formed models and they could have been more valuable in terms of measuring bankruptcy risk more accurately during the financial crisis period. The evidence in this section confirms that LT is a missing predictor in bankruptcy models and its addition would be beneficial.

[20] We have also performed our test for the economic benefits, and we find that the banks using the hybrid models achieve higher returns relative to banks using the reduced-form models.

[21] Almamy et al. (2016) consider the period 2007-2008 as the financial crisis period. While this is true, we consider that the crisis may take some time to affect company performance and as such, we also include the year 2009 in our financial crisis period.

6 Summary and Conclusions

In this paper, we examine an alternative approach for bankruptcy prediction that is based on Leland and Toft (1996), which is a model that measures the financial healthiness of firms with coupon-paying debts. This model is an extension of a model widely used in the literature; the BSM model. The Leland-Toft (LT), however, incorporates information not captured by BSM and thus it should be a better one. Based on several tests, we find evidence suggesting that it is a better approach in terms of discriminatory power, predictive accuracy but also in terms of economic performance when a bank implements LT relative to BSM.

Next, we use the probability of bankruptcy derived from LT as additional predictor to extend two widely-used corporate bankruptcy models (Altman, 1968 and Ohlson, 1980) but also, a contemporary model which was found to outperform other approaches for bankruptcy prediction (Campbell et al., 2008). Our objective is to develop powerful models that are practical and easy to implement. Under a comprehensive out-of-sample analysis, we find that augmenting Altman's model with LT further improves its performance as opposed to a cash flow ratio, as suggested by Almamy et al. (2016). The most powerful models, however, are obtained when we augment Ohlson and CHS, with LT. Further, the models augmented with LT outperform the models augmented with BSM.

However, banks are more interested in the economic performance of such models. Based on our final test we find that banks using OLT and CHSLT earn higher returns than banks which implement other models to evaluate firm-customers in a competitive loan market. We therefore recommend the use of those augmented models as an appropriate risk management tool, that could be economically beneficial for banks.

Future work should emphasize the estimation of bankruptcy costs and debt maturity separately for each firm, rather than using average values, which we think will increase the accuracy of LT and its contribution to OLT and CHSLT.

Appendix A: Sensitivity Analysis of Leland-Toft

Consider the following three vectors with different parameter values for tax rate, τ, debt maturity, T and bankruptcy costs, c:

τ= {0.15, 0.20, 0.25, 0.30, 0.35, 0.40}

T= {8, 10, 12, 14, 16, 18, 20}

c= {0.15, 0.20, 0.25, 0.30, 0.35, 0.40}

Each scenario has as input the triplet $\{\tau_i, T_j, c_k\}$ where i=k=1,…,6 and j=1,…,7. For each input scenario, AUROC of Leland-Toft is obtained and a histogram is constructed for the 252 scenarios, as shown in the following figure.

[Insert Figure A.1 here]

As can be inferred, discriminating ability measured by AUROC, is not sensitive at all with respect to the different scenarios since it ranges between 0.8828 to 0.8848, with an average value of 0.8840. Thus, performance is not affected significantly by deviations in the choice of parameter values.

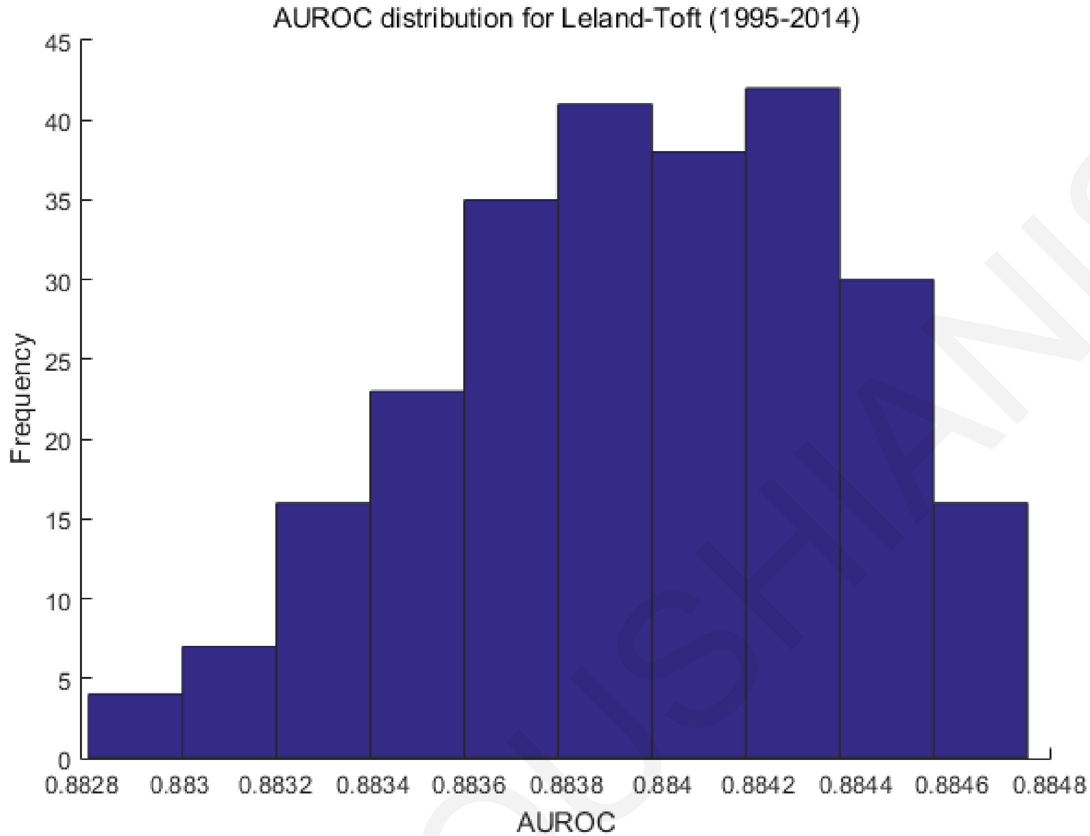

Figure A.1: This figure shows the distribution of AUROC produced by Leland-Toft model under different scenarios of its input parameters τ (the tax rate), T (debt maturity) and c (bankruptcy costs) during the period 1995-2014.

Appendix B: Discriminatory Power Test Statistic

The key element for the estimation of the test statistic is the covariance matrix of the AUROCs produced by the two models. Following DeLong et al. (1988), the covariance matrix is estimated as follows:

1) For each bankrupt firm calculate the AUROC:

$$\widehat{AUROC}\left(p_B^i\right) = \frac{1}{m}\sum_{j=1}^{m} \psi(p_B^i, p_H^j), \quad (i = 1,2,\dots,n) \tag{B.1}$$

2) For each healthy firm calculate the AUROC:

$$\widehat{AUROC}\left(p_H^j\right) = \frac{1}{n}\sum_{i=1}^{n}\psi(p_B^i, p_H^j), \quad (j = 1,2, \dots, m) \tag{B.2}$$

3) Define the 2x2 symmetric matrix S_{10} with *(k,r)th* element defined as:

$$s_{10}^{k,r} = \frac{1}{n-1}\sum_{i=1}^{n}[\widehat{AUROC}_k\left(p_B^i\right) - \widehat{AUROC}_k][\widehat{AUROC}_r\left(p_B^i\right) - \widehat{AUROC}_r] \tag{B.3}$$

4) Define the 2x2 symmetric matrix S_{01} with *(k,r)th* element defined as:

$$s_{01}^{k,r} = \frac{1}{m-1}\sum_{j=1}^{m}[\widehat{AUROC}_k\left(p_H^j\right) - \widehat{AUROC}_k][\widehat{AUROC}_r\left(p_H^j\right) - \widehat{AUROC}_r] \tag{B.4}$$

5) Then the covariance matrix of the two AUROCs is defined as:

$$S = \frac{1}{n}S_{10} + \frac{1}{m}S_{01} \tag{B.5}$$

Finally, the z-statistic which is standard-normally distributed is calculated as follows:

$$z = \frac{\widehat{AUROC}_1 - \widehat{AUROC}_2}{\left(s^{1,1} - 2s^{1,2} + s^{2,2}\right)^{1/2}} \tag{B.6}$$

with $s^{1,1}$ and $s^{2,2}$ being the variances of AUROCs of the two models under comparison and $s^{1,2}$ their covariance, all obtained from Eq. (B.5).

Appendix C: Predictive Accuracy Test Statistic

There are two distinct types of logit models; non-nested and nested models. In the case of non-nested models where the k_2 parameters in model 2 are not subset of the k_1 parameters in model 1, the Vuong (1989) test is used. The z-statistic in this case is standard-normally distributed and it is defined as follows:

$$z = \frac{2(L_1 - L_2) - (k_1 - k_2)\ln(N)}{2\sqrt{N}\omega_N} \tag{C.1}$$

Here, *N* is the number of observations and ω_N is the sample standard deviation of the individual log-likelihoods produced by each model, l_i , which is defined as follows:

$$l_i = ln\left[\frac{p_{1,i}{}^{y_i}(1-p_{1,i})^{(1-y_i)}}{p_{2,i}{}^{y_i}(1-p_{2,i})^{(1-y_i)}}\right] \tag{C.2}$$

where $p_{1,i}$ and $p_{2,i}$ are the bankruptcy probabilities for the *i-th* firm produced by models 1 and 2 respectively (time index "t" is dropped for simplicity). Furthermore, y_i indicates whether the firm is bankrupt ($y_i = 1$) or healthy ($y_i = 0$). Rejection of the null hypothesis indicates significant difference between the predictive accuracy of the two models.

On the other hand, to compare predictive accuracy between nested models where the k_2 parameters in model 2 are subset of the k_1 parameters in model 1, the standard Likelihood Ratio (LR) test is used. The statistic in that case is the following:

$$LR = -2[L(k_2) - L(k_1)] \tag{C.3}$$

and follows a chi-squared distribution with $k_1 - k_2$ degrees of freedom. Rejection of the null hypothesis indicates that predictive accuracy of the two models is not equivalent. Therefore, at least one of the extra $k_1 - k_2$ parameters in model 1 carry additional explanatory power about bankruptcy risk.

References

Afik, Z., Arad, O., & Galil, K. (2016). Using Merton model for default prediction: An empirical assessment of selected alternatives. *Journal of Empirical Finance, 35*, 43-67.

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance, 32*, 1541-1551.

Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance, 36*, 278-285.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*, 589-609.

Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance, 40*, 432-442.

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

Bharath, S. T., & Shumway, T. (2008). Forecasting Default with Merton Distance to Default Model. *The Review of Financial Studies, 21*, 1339-1369.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*, 637-654.

Blochlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking and Finance, 30*, 851-873.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance, 63*, 2899-2939.

Caprio, G., Laeven, L., & Levine, R. (2007). Governance and bank valuation. *Journal of Financial Intermediation, 16*, 584-617.

Charitou, A., Dionysiou, D., Lambertides, N., & Trigeorgis, L. (2013). Alternative Bankruptcy Prediction Models Using Option-Pricing Theory. *Journal of Banking and Finance, 37*, 2329-2341.

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

Delong, R., Delong, M., & Clarke-Pearson, L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A non-parametric approach. *Biometrics, 44*, 837-845.

Dodd, L. E., & Pepe, M. S. (2003). Partial AUC estimation and regression. *Biometrics, 59*, 614-623.

Eom, Y.H., Helwege, J., & Huang, Z.J. (2004) .Structural models of corporate bond pricing: An empirical analysis. *The Review of Financial Studies,* 17, 499-544.

Filipe, S. F., Grammatikos, T., & Michala, D. (2016). Forecasting distress in European SME portfolios. *Journal of Banking and Finance, 64*, 112-135.

Galil, K., & Gilat, N. (2008). Predicting default more accurately: To proxy or not to proxy for default? Available at SSRN: https://ssrn.com/abstract=2618190.

Gupta, K., Krishnamurti, C., & Tourani-Rad, A. (2013). Is corporate governance relevant during the financial crisis? *Journal of International Financial Markets, Institutions and Money*, *23*, 85-110.

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

Hilberink, B., & Rogers, L.C.G. (2002). Optimal capital structure and endogenous default. *Finance and Stochastics,* 6, 237-263.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *The Review of Financial Studies, 9*, 5-34.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221-233.

Leland, H. (2004). Predictions of default probabilities in structural models of debt. *Journal of Investment Management*, *2*, 5-20.

Leland, H., & Toft, K. B. (1996). Optimal capital structure, endogenous bankruptcy and the term structure of credit spreads. *Journal of Finance, 51*, 987-1019.

Lyden, S., & Saraniti, D. (2001). *An Empirical Examination of the Classical Theory of Corporate Security Valuation.* Availabe at SSRN: http://papers.ssrn.com/sol3/papers.cfm?19.

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

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

Papakyriakou, P., Sakkas, A., & Taoushianis, Z. (2019). Financial firm bankruptcies, international stock markets, and investor sentiment. *International Journal of Finance and Economics, 24*, 461-473.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, *74*, 101-124.

Sobehart, J., & Keenan, S. (2001). Measuring default accurately. *Risk* , 31-33.

Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing. *Journal of Banking and Finance, 29*, 1213-1236.

Switzer, L. N., Tu, Q., & Wang, J. (2018). Corporate governance and default risk in financial firms over the post-financial crisis period: International evidence. *Journal of International Financial Markets, Institutions and Money, 52*, 196-210.

Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394-419.

Vassalou, M., & Xing,Y. (2004). Default risk in equity returns. *Journal of Finance, 59,* 831-868.

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

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

Tables

Table 1: Model definition and estimation of the variables

	Variable	Estimation
AAN (2016)	WCTA	Working capital/total assets
	RETA	Retained earnings/total assets
	EBITTA	Earnings before interests and taxes/total assets
	MVTL	Market value of equity/total liabilities
	SLTA	Net sales/total assets
	CFOTA	Cash flows from operations/total assets
Ohlson (1980)	SIZE	Log (Total assets/GNP price level index)
	TLTA	Total liabilities/total assets
	CLCA	Current liabilities/current assets
	D(TL>TA)	1 if TL>TA and 0 otherwise, where TL are total liabilities and TA are total assets
	NITA	Net income/total assets
	CFOTL	Cash flows from operations/total liabilities
	$D(NI_t+NI_{t-1}<0)$	1 if the cumulative net income in two consecutive years is negative and 0 otherwise
	WCTA	Working Capital/total assets
	CHNI	$(NI_t - NI_{t-1})/(\lvert NI_t \rvert + \lvert NI_{t-1} \rvert)$, is the change in net income (takes values between -1 and 1)
CHS (2008)	NIMTA	Net income/market value of assets, where market value of assets is the sum of market value of equity and total liabilities
	TLMTA	Total liabilities/market values of assets
	EXRET	Annualized return of each firm's equity minus the annualized return of the S&P 500 index, over the previous 12 months
	SIGMA	Annualized standard deviation of daily stock returns, over the previous 3 months
	RSIZE	Relative size, defined as log (Market value of equity/market value of S&P 500 index)
	CASHMTA	Cash and short-term investments/market value of assets
	MB	Market value of equity/book value of equity
	PRICE	Log (stock price)

Leland and Toft (1996)	σ	Annualized volatility of asset returns, obtained by solving Eqs. (7) and (8)
	μ	Annualized return on assets, proxied by the annualized return on equity
	V	Market value of assets, obtained by solving Eqs. (7) and (8)
	r	Riskless rate of return, proxied by the one-year treasury constant maturity rate
	C	Coupon payments, proxied by the interest expenses in the income statement
	P	Principal value of debt, proxied by short-term debt plus half of long-term debt
	δ	Payout yield, which is the sum of coupons and dividends (ordinary and preferred) divided by the market value of assets
	τ	Corporate tax rate, 15% as in Leland (2004)
	c	Bankruptcy costs, 30% as in Leland (2004)
	T	Maturity, 10 years as in Leland (2004)

This table describes the input variables of four models: Almamy et al. (2016), denoted as AAN, Ohlson (19980), Campbell et al. (2008), denoted as CHS and Leland and Toft (1996). All variables are constructed using financial and market information one year prior to bankruptcy filing. The second column shows how variables are entered in the models and the third column shows how they are calculated.

Table 2: Distribution of observations per year

Bankruptcy year	Bankrupt firms	Healthy firms	Bankruptcy rate
1995	15	2749	0.543
1996	16	2804	0.567
1997	13	2933	0.441
1998	21	2186	0.952
1999	31	2045	1.493
2000	21	2572	0.810
2001	23	2425	0.940
2002	15	2206	0.675
2003	18	2045	0.873
2004	13	1919	0.673
2005	15	1865	0.798
2006	10	1796	0.554
2007	15	1738	0.856
2008	20	1661	1.190
2009	34	1560	2.133
2010	7	1508	0.462
2011	9	1431	0.625
2012	13	1388	0.928
2013	12	1353	0.879
2014	12	1313	0.906
Total	**333**	**39497**	**0.836**

This table reports the number of observations across the years 1995-2014. The first column shows the year of bankruptcy, the second and third columns show the number of bankrupt and healthy firms respectively and the last column shows the annual bankruptcy rate defined as bankrupt firms /(bankrupt firms + healthy firms).

Table 3: Distribution of observations per industry

Industry	Number of observations	Percentage
Agriculture, Forestry and Fishing	138	0.35
Mining	2340	5.87
Construction	379	0.95
Manufacturing	21109	53.00
Transportation, Communications, Electric, Gas, and Sanitary Services	4126	10.36
Wholesale Trade	1604	4.03
Retail Trade	3349	8.41
Services	6539	16.42
Public Administration	246	0.62

This table shows the distribution of observations per industry. Each observation is classified to an industry, according to SIC codes. Column 2 shows the number of observations that belong to each industry and column 3 shows the percentage of sample belonging to each industry calculated as industry observations / total observations.

Table 4: Summary statistics

Variable	Mean values		t-statistic for differences
	Bankruptcies	Non-Bankruptcies	
SIZE	0.410	1.391	8.56
TLTA	0.888	0.523	26.00
WCTA	-0.009	0.227	18.77
CLCA	1.389	0.658	25.75
D(TL>TA)	0.312	0.034	27.32
NITA	-0.414	-0.024	31.99
CFOTL	-0.302	0.102	13.08
D ($NI_t+NI_{t-1}<0$)	0.913	0.322	23.06
CHINI	-0.260	0.013	8.83
RETA	-1.736	-0.188	22.37
EBITTA	-0.238	0.029	25.03
MVTL	3.103	38.257	3.99
SLTA	1.250	1.172	1.72
CFOTA	-0.147	0.047	21.67
NIMTA	-0.270	-0.008	36.02
TLMTA	0.694	0.412	19.12
EXRET	-0.909	-0.056	26.79
SIGMA	1.219	0.609	26.01
RSIZE	-12.933	-10.820	18.46
CASHMTA	0.075	0.097	2.85
MB	1.459	1.416	0.24
PRICE	0.496	2.422	28.93
LT	0.441	0.042	41.46

This table reports mean value differences for the explanatory variables, between the bankrupt and non-bankrupt firms and t-tests for the statistical significance of the differences. The definition of variables can be found in Table 1.

Table 5: Reduced-form and hybrid models estimation, 1995-2005

Ohlson

Constant	SIZE	TLTA	WCTA	CLCA	D(TL>TA)	NITA	CFOTL	D (NI_t+NI_{t-1}<0)	CHINI
-7.280***	0.093***	1.461***	-1.997***	0.040	-0.047	-0.451*	-0.416***	2.120***	-0.545***
(0.368)	(0.035)	(0.391)	(0.553)	(0.155)	(0.361)	(0.275)	(0.130)	(0.253)	(0.124)

AAN

Constant	WCTA	RETA	EBITTA	MVTL	SLTA	CFOTA
-4.470***	-3.628***	0.084	-1.420***	-0.015	0.119*	-1.390***
(0.136)	(0.363)	(0.057)	(0.474)	(0.012)	(0.066)	(0.563)

CHS

Constant	NIMTA	TLMTA	EXRET	SIGMA	RSIZE	CASHMTA	MB	PRICE
-3.916***	-3.206***	2.566***	-0.545***	-0.073	0.140***	-4.186***	-0.00	-0.531***
(0.604)	(0.331)	(0.333)	(0.135)	(0.220)	(0.047)	(0.917)	(0.015)	(0.105)

ALT

Constant	WCTA	RETA	EBITTA	MVTL	SLTA	LT
-5.222***	-2.485***	0.004	-1.955***	-0.008	0.157**	2.520***
(0.148)	(0.331)	(0.055)	(0.316)	(0.007)	(0.067)	(0.181)

CHSLT

Constant	NIMTA	TLMTA	EXRET	SIGMA	RSIZE	CASHMTA	MB	PRICE	LT
-3.528	-3.129***	1.955***	-0.30***	-0.150	0.146***	-4.195***	-0.002	-0.529***	1.174***
(0.607)	(0.342)	(0.391)	(0.143)	(0.212)	(0.047)	(0.890)	(0.015)	(0.104)	(0.298)

OLT

Constant	SIZE	TLTA	WCTA	CLCA	D(TL>TA)	NITA	CFOTL	D (NI_t+NI_{t-1}<0)	CHINI	LT
-7.09***	0.030	1.202***	-1.809***	-0.102	-0.016	-0.447*	-0.341***	1.844***	-0.369***	2.034***
(0.357)	(0.038)	(0.375)	(0.552)	(0.159)	(0.345)	(0.267)	(0.127)	(0.252)	(0.132)	(0.194)

This table reports estimation results for six models; Ohlson (1980), Almamy et al., (2016), referred to as AAN, Campbell et al. (2008), referred to as CHS and three extended versions of Altman, Ohlson and Campbell et al. (2008) models, which include LT as additional predictor (referred to as ALT, OLT and CHSLT respectively). The sample includes 25950 firm-year observations, from which 201 went bankrupt in a year between 1995 and 2005. The predictor variables are constructed one year prior to bankruptcy. For the definition of variables refer to table 1. ***, ** and * indicate statistical significance at α=1%, α=5% and α=10% respectively. In parentheses clustered robust standard errors are reported, that take into consideration the panel character of our data.

Table 6: Model performance and test for differences, baseline approach

Panel A: Out-of-sample performance, baseline approach (2006-2014)

Model	AUROC	pAUROC	Log-Likelihood	Pseudo-R^2 (%)
Structural				
Leland-Toft	0.8941	0.3953	-600.93	19.72
BSM	0.8659	0.3750	-614.55	17.90
Hybrid				
OLT	0.9449	0.4477	-483.43	35.41
CHSLT	0.9395	0.4426	-491.41	34.35
ALT	0.9207	0.4288	-519.56	30.59
Reduced-Form				
CHS (2008)	0.9332	0.4392	-498.85	33.35
AAN (2016)	0.8597	0.3754	-603.81	19.33
Ohlson (1980)	0.9252	0.4280	-535.57	28.45

Panel B: Test-statistics for differences in AUROCs

Model	OLT	CHSLT	ALT	CHS	AAN	Ohlson	Leland-Toft
CHSLT	0.64						
ALT	2.58	1.63					
CHS	1.12	1.47	-0.92				
AAN	5.36	5.30	4.78	4.51			
Ohlson	4.73	1.65	-0.41	0.8127	-4.41		
Leland-Toft	3.59	3.15	1.67	2.46	-1.64	1.96	
BSM	6.02	4.14	3.03	3.58	-0.28	3.28	2.74

Panel C: Test-statistics for differences in log-likelihoods

Model	OLT	CHSLT	ALT	CHS	AAN	Ohlson	Leland-Toft
CHSLT	0.75						
ALT	4.38	2.38					
CHS	1.23	14.88	-1.48				
AAN	7.51	6.16	6.51	5.86			
Ohlson	104.28	2.98	1.22	2.58	-6.21		
Leland-Toft	8.07	6.93	6.73	5.76	-0.16	3.61	
BSM	8.01	7.21	6.56	6.51	0.63	4.47	1.38

This table reports out-of-sample performance for the two structural models (Leland-Toft and BSM), the three hybrid models (OLT, CHSLT and ALT) as well as the three reduced-form models (Ohlson,1980; Campbell et al., 2008, referred to as CHS and Almamy et al., 2016, referred to as AAN). For the definition of the models, refer to table 1. Panel A reports discriminating ability, measured by AUROC as well as predictive accuracy, measured by log-likelihood (and pseudo-R^2). Panel B reports test statistics for differences in the discriminating ability between various models, using Delong et al. (1988). Panel C reports test statistics for differences in predictive accuracy between various models using likelihood ratio tests or Vuong (1989) test. The results are based on a baseline approach, where the models are estimated on the period 1995-2005 and applied on the period 2006-2014. In the case of structural models, for consistency, we estimate two logistic regression models where the first contains the probability of bankruptcy derived from Leland-Toft and the second the probability of bankruptcy derived from BS as predictors.

Table 7: Economic performance of banks using different bankruptcy models (LT vs BSM)

	Bank 1	Bank 2	Bank 3
	LT	BSM	Altman
Credits	4037	4667	5022
Market Share (%)	29.09	33.62	36.18
Bankruptcies	10	37	38
Bankruptcies/Credits (%)	0.25	0.79	0.76
Average Spread (%)	0.38	0.49	0.45
Revenues ($M)	110.71	163.22	163.89
Loss($M)	26.41	97.70	100.34
Profit($M)	84.30	65.52	63.55
Return on Assets (%)	0.29	0.19	0.018
Return on RWA (%)	1.09	0.64	0.54

This table reports economic results for three banks in a competitive loan market worth $100 billion. Bank 1 uses LT, bank 2 uses BSM and bank 3 uses Altman. For the definition of the models, see table 1. The banks sort prospective customers (2006-2014) and reject the 5% of firms with the highest risk. The remaining firms are classified in 10 groups of equal size and for each group, a credit spread is calculated, as described in the main text (section 4.3). The bank that classifies the firm to the group with the lowest spread is finally granting the loan. Market share is the number of loans given divided by the number of firm-years, Revenues = market size*market share*average spread, Loss=market size*prior probability of bankruptcy*share of bankruptcies*loss given default. Profit=Revenues-Loss. Return on Assets is profits divided by market size*market share and Return on Risk-Weighted-Assets is profits divided by Risk-Weighted Assets, obtained from formulas provided by the Basel Accord (2006). The prior probability of bankruptcy is the bankruptcy rate for firms between 1995-2005 and equals 0.77%. Loss given default is 45%.

Table 8: Model performance and test for differences, rolling window approach

Panel A: Out-of-sample performance, rolling approach (2006-2014)

Model	AUROC	Log-Likelihood	Pseudo-R^2 (%)
Hybrid			
OLT	0.9469	-470.37	37.00
CHSLT	0.9438	-479.66	35.76
ALT	0.9253	-508.25	31.93
Reduced-Form			
CHS (2008)	0.9372	-485.94	34.92
AAN (2016)	0.8673	-593.86	20.46
Ohlson (1980)	0.9289	-522.87	29.97

Panel B: Test-statistics for differences in AUROCs

Model	OLT	CHSLT	ALT	CHS	AAN
CHSLT	0.41				
ALT	2.31	1.58			
CHS	1.00	1.80	-0.88		
AAN	5.24	5.12	4.62	4.42	
Ohlson	4.33	1.94	-0.34	0.91	-4.38

Panel C: Test-statistics for differences in log-likelihoods

Model	OLT	CHSLT	ALT	CHS	AAN
CHSLT	0.79				
ALT	4.57	2.20			
CHS	1.13	12.57	-1.46		
AAN	7.61	6.17	6.28	5.91	
Ohlson	104.99	2.85	1.06	2.49	-6.48

This table reports out-of-sample performance for the three hybrid models (OLT, CHSLT and ALT) as well as the three reduced-form models (Ohlson,1980; Campbell et al., 2008, referred to as CHS and Almamy et al., 2016, referred to as AAN). For the definition of the models refer to table 1. Panel A reports discriminating ability, measured by AUROC as well as predictive accuracy, measured by log-likelihood (and pseudo-R^2). Panel B reports test statistics for differences in the discriminating ability between various models, using Delong et al. (1988). Panel C reports test statistics for differences in the predictive accuracy between various models using likelihood ratio tests or Vuong (1989) test. The results are based on a rolling window approach, where the models are updated yearly and used to predict bankruptcies next year. For instance, the models are estimated between 1995 and 2005 and apply them on firms in 2006. Then we re-estimate the models between 1996 and 2006 and apply them on firms in 2007. This process is repeated up to 2014. Bankruptcy probabilities for each year are aggregated to obtain single performance measures.

Table 9: Model performance and test for differences, five folds approach

Panel A: Out-of-sample performance, five folds approach (1995-2014)

Model	AUROC	Log-Likelihood	Pseudo-R^2 (%)
Hybrid			
OLT	0.9091	-1379.48	28.49
CHSLT	0.9057	-1403.72	27.23
ALT	0.8826	-1471.35	23.73
Reduced-Form			
CHS (2008)	0.9014	-1417.16	26.54
AAN (2016)	0.8438	-1630.45	15.48
Ohlson (1980)	0.8939	-1477.06	23.43

Panel B: Test-statistics for differences in AUROCs

Model	OLT	CHSLT	ALT	CHS	AAN
CHSLT	0.50				
ALT	3.95	2.86			
CHS	1.02	1.96	-2.05		
AAN	7.09	6.14	5.63	5.42	
Ohlson	4.56	1.54	-1.38	0.94	-5.75

Panel C: Test-statistics for differences in log-likelihoods

Model	OLT	CHSLT	ALT	CHS	AAN
CHSLT	1.25				
ALT	5.20	3.07			
CHS	1.68	26.87	-2.11		
AAN	7.63	6.41	6.62	6.07	
Ohlson	195.16	2.79	0.22	2.34	-5.53

This table reports out-of-sample performance for the three hybrid models (OLT, CHSLT and ALT) as well as the three reduced-form models (Ohlson,1980; Campbell et al., 2008, referred to as CHS and Almamy et al., 2016, referred to as AAN). For the definition of the models refer to table 1. Panel A reports discriminating ability, measured by AUROC as well as predictive accuracy, measured by log-likelihood (and pseudo-R^2). Panel B reports test statistics for differences in the discriminating ability between various models, using Delong et al. (1988). Panel C reports test statistics for differences in the predictive accuracy between various models using likelihood ratio tests or Vuong (1989) test. The results are based on a five-fold cross-validation approach, where we divide the whole sample into five equal sub-samples. Any four of them are used to estimate the models and apply them on firms in the left-out sub-sample. Bankruptcy probabilities for each left-out sub-sample are aggregated to obtain single out-of-sample performance measures

Table 10: Economic performance for five banks when using different bankruptcy models

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5
	OLT	CHSLT	CHS	Ohlson	Altman
Credits	3571	2561	2758	1886	2935
Market Share (%)	25.73	18.45	19.87	13.59	21.15
Bankruptcies	4	4	12	11	48
Bankruptcies/Credits (%)	0.11	0.16	0.44	0.58	1.64
Average Spread (%)	0.35	0.35	0.43	0.51	0.59
Revenues ($M)	90.14	65.35	85.42	69.41	125.65
Loss($M)	10.56	10.56	31.69	29.05	126.75
Profit($M)	79.58	54.79	53.73	40.36	-1.10
Return on Assets (%)	0.31	0.30	0.27	0.30	-0.00
Return on RWA (%)	1.74	1.54	1.12	1.02	-0.00

This table reports economic results for five banks in a competitive loan market worth $100 billion. Bank 1 uses OLT, bank 2 uses CHSLT, bank 3 uses CHS, bank 4 uses Ohlson and bank 5 uses Altman. For the definition of the models, see table 1. The models are estimated using data from 1995-2005. The banks sort prospective customers (2006-2014) and reject the 5% of firms with the highest risk. The remaining firms are classified in 10 groups of equal size and for each group, a credit spread is calculated, as described in the main text (section 4.3). The bank that classifies the firm to the group with the lowest spread is finally granting the loan. Market share is the number of loans given divided by the number of firm-years, Revenues = market size*market share*average spread, Loss=market size*prior probability of bankruptcy*share of bankruptcies*loss given default. Profit=Revenues-Loss. Return on Assets is profits divided by market size*market share and Return on Risk-Weighted-Assets is profits divided by Risk-Weighted Assets, obtained from formulas provided by the Basel Accord (2006). The prior probability of bankruptcy is the bankruptcy rate for firms between 1995-2005 and equals 0.77%. Loss given default is 45%.

Table 11: Extending Campbell et al. (2008), Ohlson (1980) and Altman (1968) with BSM and LT

Panel A: Out-of-sample performance, 2006-2014

Model	AUROC	LL	Pseudo-R^2 (%)
CHS models			
CHS (2008)	0.9332	-498.85	33.35
CHS with BSM	0.9343	-497.96	33.47
CHS with LT	0.9395	-491.41	34.35
Ohlson models			
Ohlson (1980)	0.9252	-535.57	28.45
Ohlson with BSM	0.9383	-497.71	33.51
Ohlson with LT	0.9449	-483.43	35.41
Altman models			
AAN (2016)	0.8597	-603.81	19.33
Altman with BSM	0.9109	-538.87	28.01
Altman with LT	0.9207	-519.56	30.59

Panel B: Test statistics for differences in AUROC's

CHS vs CHS with BSM	0.82	Ohlson vs Ohlson with BSM	3.54	AAN vs Altman with BSM	4.13
CHS vs CHS with LT	1.49	Ohlson vs Ohlson with LT	4.71	AAN vs Altman with LT	4.78
CHS with LT vs CHS with BSM	1.47	Ohlson with LT vs Ohlson with BSM	2.37	Altman with LT vs Altman with BSM	1.97

Panel C: Test statistics for differences in log-likelihoods

CHS vs CHS with BSM	1.78	Ohlson vs Ohlson with BSM	75.72	AAN vs Altman with BSM	5.74
CHS vs CHS with LT	14.88	Ohlson vs Ohlson with LT	104.28	AAN vs Altman with LT	6.51
CHS with LT vs CHS with BSM	2.09	Ohlson with LT vs Ohlson with BSM	2.36	Altman with LT vs Altman with BSM	2.52

This table reports out-of-sample performance for the three reduced-form models (Campbell et al., 2008, referred to as CHS, Ohlson, 1980 and Almamy et al., 2016, referred to as AAN) and several hybrid models that augment the reduced-form models with the LT or BSM. The models are estimated using data from 1995-2005 and the results are based on the out-of-sample period, 2006-2014. Panel A reports discriminating ability and predictive accuracy as measured by AUROC and log-likelihood respectively. Panel B reports test statistics for differences in the discriminating ability between various models, using Delong et al. (1988). Panel C reports test statistics for differences in the predictive accuracy between various models using likelihood ratio tests or Vuong (1989) test.

Table 12: Model performance and test for differences, baseline approach (time-robustness)

Panel A: Out-of-sample performance, baseline approach (1995-2005)

Model	AUROC	Log-Likelihood	Pseudo-R^2 (%)
Hybrid			
OLT	0.8854	-960.51	18.72
CHSLT	0.8838	-937.03	20.71
ALT	0.8716	-1029.06	12.92
Reduced-Form			
CHS (2008)	0.8811	-951.09	19.52
AAN (2016)	0.8344	-1116.46	5.53
Ohlson (1980)	0.8741	-1014.27	14.17

Panel B: Test-statistics for differences in AUROCs

Model	OLT	CHSLT	ALT	CHS	AAN
CHSLT	0.16				
ALT	1.47	1.14			
CHS	0.42	1.28	-0.81		
AAN	3.88	3.49	3.98	3.25	
Ohlson	2.41	0.87	-0.22	0.62	-3.18

Panel C: Test-statistics for differences in log-likelihoods

Model	OLT	CHSLT	ALT	CHS	AAN
CHSLT	-1.20				
ALT	2.36	3.02			
CHS	-0.43	28.12	-2.41		
AAN	3.13	3.62	3.07	3.33	
Ohlson	107.52	3.14	-0.43	2.56	-2.10

This table reports out-of-sample performance for the three hybrid models (OLT, CHSLT and ALT) as well as the three reduced-form models (Ohlson,1980; Campbell et al., 2008, referred to as CHS and Almamy et al., 2016, referred to as AAN). For the definition of the models, refer to table 1. Panel A reports discriminating ability, measured by AUROC as well as predictive accuracy, measured by log-likelihood (and pseudo-R^2). Panel B reports test statistics for differences in the discriminating ability between various models, using Delong et al. (1988). Panel C reports test statistics for differences in the predictive accuracy between various models using likelihood ratio tests or Vuong (1989) test. The results are based on a baseline approach, where the models are estimated on the period 2006-2014 and applied on the period 1995-2005, in order to test the time robustness of our models.

Figures

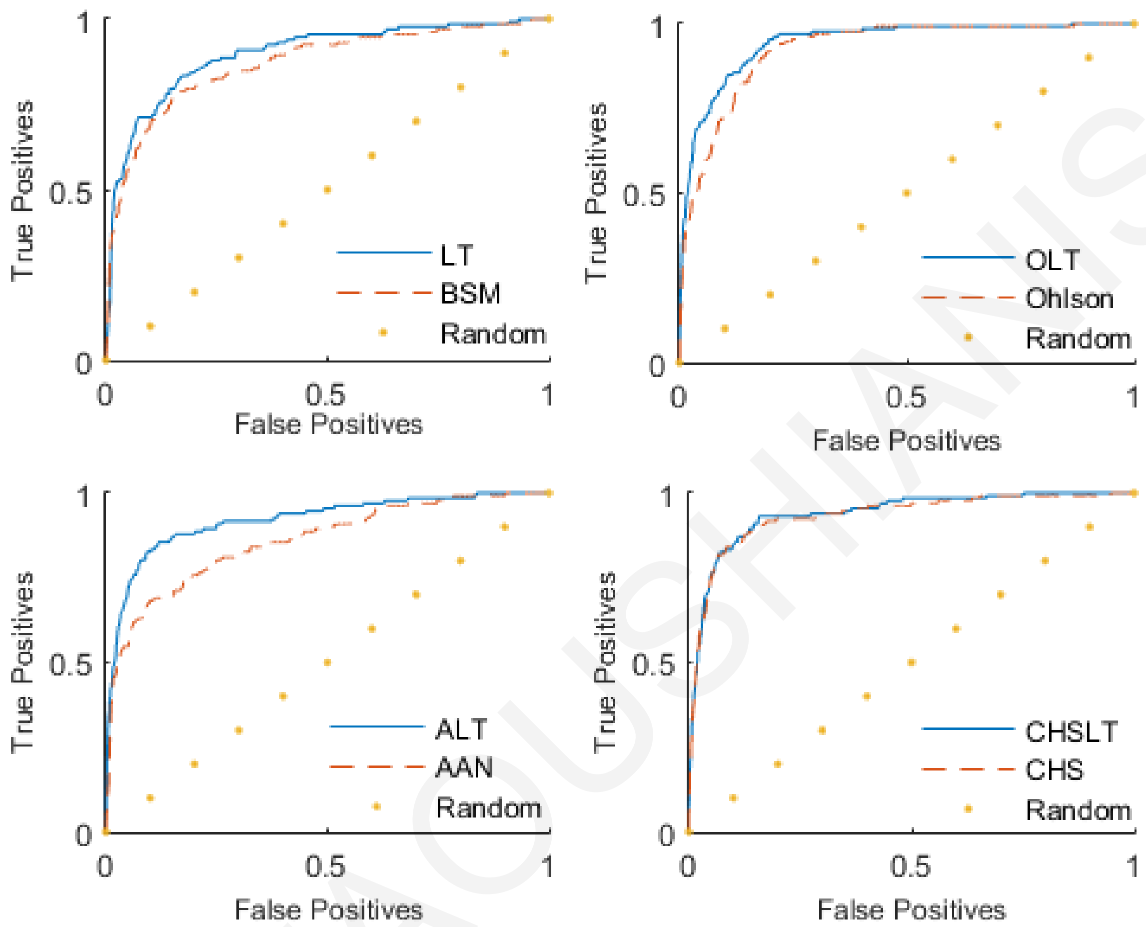

Figure 1: This figure provides graphical representation of the discriminatory power of various bankruptcy prediction models through the ROC curves. The ROC curves are constructed for the out-of-sample period 2006-2014.

CHAPTER 2

Maximizing discriminatory power of bankruptcy prediction models: Empirical evidence from U.S.

Abstract

In this paper several methodologies for maximizing the discriminatory power of bankruptcy prediction models, as measured by the Area Under Receiver Operating Characteristics (AUROC) curve, are introduced and compared. We consider linear and probabilistic response functions for the output of the models as well as different merit functions, used to obtain model coefficients. For our analysis we use accounting and market information for U.S. public bankrupt and healthy firms between 1990 and 2015. Results show higher discriminatory power when we implement our approaches as compared with traditional approaches, such as logistic regression models. We also find that using models with a merit function that accounts for outliers yields the best performance among the linear response functions. Among all models, however, a neural network with probabilistic response function is the best performing model. More importantly results hold under different tests. One of our tests for instance, shows that banks using models with maximized AUROC earn higher risk-adjusted returns relative to banks using traditional approaches, highlighting the benefits of using models with maximized AUROC.

1 Introduction

1.1 Background and Motivation

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

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

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

Beyond that, Moody's KMV provides ample explanatory documentation on how to use various discriminatory power measures in practice (see for instance Keenan and Sobehart, 1999 and Sobehart et al., 2000). This extensive use in fact highlights the importance of using discriminatory power as a leading measure to evaluate the performance of bankruptcy prediction models.

Yet, it is somewhat surprising that a common practice in bankruptcy prediction studies is to use discriminatory power as an indication of model performance, rather than obtaining model coefficients directly by maximizing discriminatory power. Exceptions include Miura et al. (2010) and Kraus and Kuchenhoff (2014) in the related area of credit scoring which we also discuss and compare. We contribute to this limited literature by introducing and comparing several new methodologies that maximize the discriminatory power of bankruptcy prediction models and through a battery of tests, we highlight the importance of using such models. To measure discriminatory power, we use the Area Under Receiver Operating Characteristics curve (AUROC or AUC). This is a widely-used statistic to measure the discriminatory power of bankruptcy prediction models and it has also been used in related areas, such as mortgage default prediction (Fitzpatrick and Mues, 2016) and generally when assessing the performance of credit scoring models (see for instance Lessmann et al., 2015, and references therein for recent studies that use AUROC as a performance measure).

For our analysis we collect annual financial data and daily equity prices for a large sample of U.S. public bankrupt and healthy firms between 1989 and 2014 and construct variables to make one-year forecasts (for bankruptcies between 1990 and 2015). We keep approximately 70% of the whole sample as a training set (1990-2006) and evaluate the performance of the models in the testing set (2007-2015) using three distinct type of tests, following Bauer and Agarwal (2014); 1) AUROC analysis 2) Information content tests 3) Economic performance, when banks use various bankruptcy prediction models in a competitive loan market.

1.2 Main Findings

Firstly, we employ the logistic regression approach[22] which is our benchmark against models with maximized AUROC and find that several financial variables related to firm leverage, profitability, liquidity and coverage, are significant predictors of bankruptcy. When we also consider market-based variables in the analysis, however, the model with both financial and market variables outperforms the model with only financial variables consistent with prior research (Shumway, 2001; Chava and Jarrow, 2004; Campbell et al., 2008; Wu et al., 2010 and Tinoco and Wilson, 2013).

[22] We use logistic regression, since it is the most common approach of deriving a classification rule (Crook et al., 2007) and an approach adopted by many researchers in recent bankruptcy prediction studies (i.e. Westgaard and Wijst, 2001; Chava and Jarrow, 2004; Altman and Sabato, 2007; Campbell et al., 2008; Tinoco and Wilson, 2013, etc.).

Next, we develop models with maximized AUROC using the set of variables that we find to be significant (with financial only data and financial plus market data) according to the logistic regression approach. We consider models with probabilistic and linear response functions. That is, the output of the models is a probability in the first case and a linear score in the second case. We also consider various merit functions used in the optimization to obtain the coefficient estimates and we seek to find the specification which yields the best performance. We find that our proposed approaches outperform out-of-sample logistic regression models. Among the linear response functions, we find that a merit function that takes care of the outliers which often characterize financial data, is the best performing one. However, a neural network model with a probabilistic response function is the best performing among all. The results are consistent with respect to the three testing approaches and summarized as follows: 1) Models with maximized AUROC outperform logistic regression models in terms of AUROC, out-of-sample, 2) The models with the highest AUROC are selected for the remaining tests and we find that they contain significantly more information relative to a logistic regression model according to information content tests and 3) Banks using models with maximized AUROC, earn superior returns on a risk-adjusted basis, relative to banks that use traditional models to predict bankruptcy. Therefore, we recommend the implementation of models with maximized AUROC in bankruptcy prediction since, according to our findings, such models are more valuable and appropriate risk management tools relative to traditional bankruptcy prediction models.

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

2 Data

2.1 Sample

Our sample consists of 11,096 non-financial U.S. firms from which 422 filed for bankruptcy under Chapter 7 or Chapter 11, between 1990-2015. We have a total of 97,133 firm-year observations with non-missing data, collected between 1989-2014 to forecast bankruptcies one year ahead. The date of bankruptcy filing was identified from the database BankruptcyData.

2.2 *Variables Construction*

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

First, we construct financial ratios capturing aspects of a firm's financial performance, such as leverage, profitability, liquidity, coverage, activity, cash flows, as presented in panel A of Table 1. A limitation of financial variables is that by their nature look backwards and the quality of information they carry depends on accounting practices (Hillegeist et al., 2004; Agarwal and Taffler, 2008). Market variables instead, constructed from equity prices, look forward since they carry market perceptions about the prospects of the firm. For publicly traded firms it would be more appropriate to incorporate market variables in the models. To this end, we collect daily equity prices from CRSP for the entire fiscal year and several market-based variables are constructed, as reported in panel B of Table 1. Annualized volatility of daily equity returns (*VOLE*) refers to the fluctuations of firm's equity value returns, expecting to be higher for bankrupt firms. Next, excess returns (*EXRET*) refer to the difference between firm's annualized equity return and the annualized value-weighted return of a portfolio with NYSE, AMEX, NASDAQ stocks, expecting to be lower for bankrupt firms. Further, we consider the relative size of the firm (*RSIZE)*, the logarithm of stock price at fiscal year-end (*LOGPRICE*) and the Market-to-Book ratio (*MB*), expecting a negative association with bankruptcy risk. Finally, we include three financial variables scaled by firm's market value. More precisely, Campbell et al. (2008) show that scaling financial variables with a market-based measure of firm's value i.e. market equity + liabilities (*MTA*), compared to total assets as reported in the balance sheet, increases the predictive accuracy of bankruptcy prediction models. These variables are cash over *MTA* (*CASHMTA),* net income over *MTA* (*NIMTA*), expecting a negative association with bankruptcy risk and lastly, total liabilities over *MTA* (*TLMTA*). Following common practice, we winsorize the variables between 1st and 99th percentile to avoid problems induced by outliers.

[Insert Table 1 here]

2.3 *Variables Selection*

Table 1 presents an extensive list of variables that previous studies find to be significant predictors of bankruptcy risk. Out of these variables, a smaller set should be selected in order to construct parsimonious models with few variables but with high forecasting power. We establish a three-step approach to select the most powerful variables (see for instance Altman and Sabato, 2007 and Filipe et al., 2016) and summarized in the following three steps:

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

Step 2: Remove highly correlated variables using the Variance Inflation Factor (VIF) criterion. The idea of this step is to remove the variables that are highly correlated with others, since multicollinearity may yield misleading results regarding the significance of the variables in the final model. Beyond that, we end up with variables that provide different information and explain bankruptcy uniquely. We use 5 as cut-off (variables with VIF ≥ 5 are removed).

Step 3: Perform a stepwise multivariate logistic regression to the remaining variables in order to obtain the most significant variables from a statistical point of view (we use a significance level of $\alpha = 5\%$).

The logistic regression program estimates coefficients assuming independent observations, which is an invalid assumption, since the data contains information for firms over multiple periods. In such case, an appropriate correction measure which we adopt in our study, is to use clustered robust standard errors (also used by Filipe et al.,2016).

Using the three-step approach, we develop two types of models. The first one is a "private firm" type of model, including only financial variables. We further develop a "public firm" type of model, including both financial and market variables. For example, the private firm model includes five financial variables (*TLTA*, *STDTA*, *NITA*, *CASHTA*, *EBITCL*), while the public firm model includes six variables (*TLTA*, *STDTA*, *LOGPRICE, CASHMTA, NIMTA, EXRET*). Notice that two financial-based variables (*CASHTA* and *NITA*) are replaced with *CASHMTA* and *NIMTA.* Generally, the majority of variables that are found to be significant for the public firm model are market variables, which is consistent with the perception that market-based variables are better bankruptcy risk measures, due to their forward-looking nature.

2.4 Descriptive Statistics

Table 2 reports descriptive statistics for the accounting and market variables that we find to be significant predictors of bankruptcy. As expected, bankrupt firms are more levered on average relative to healthy firms (*TLTA* and *STDTA* for bankrupt firms are higher), they are also less profitable (*NITA* and *NIMTA* are lower for bankrupt firms). Furthermore, bankrupt firms are more constrained in terms of cash available (*CASHTA* and *CASHMTA* are lower) as opposed to healthy firms. Going to the market variables, it is evident that the stock price of bankrupt firms (*LOGPRICE)* on average is lower than healthy firms, possibly due to their deteriorating financial position that is priced by investors, leading to a depreciation of their stock prices at the year prior to bankruptcy. Finally, bankrupt firms exhibit lower and negative market performance relative to the market (*EXRET* is lower one year prior to bankruptcy), as opposed to healthy firms.

[Insert Table 2 here]

3 Methodology

3.1 Measuring Discriminatory Power

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

The ROC curve provides a graphical way to visualize discriminatory power. A quantitative assessment of the discriminatory power is given by the Area under ROC curve (AUROC) which is calculated as follows[23]:

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

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

3.2 *Maximizing Discriminatory Power*

In this section we present different methodologies for maximizing the discriminatory power. First, we consider the case of a probabilistic response function and second, we consider the case of a linear response function. Finally, several merit functions used in the optimization to obtain model coefficients, are introduced

3.2.1 *Probabilistic Response Function*

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

We define $d_{i,j}(\beta) = s_B^i(X_i,\beta) - s_H^j(X_j,\beta) = p_B^i(X_i,\beta) - p_H^j(X_j,\beta)$ as the difference between the probability of bankruptcy[24] for the *i-th* bankrupt firm and the probability of bankruptcy for the *j-th* healthy firm observation, conditional on the predictor variables in *X* which could be a set of financial and market variables. From Eq. (1), to obtain the

[23] For further explanation, refer to Hanley and McNeil (1982) and Sobehart and Keenan (2001).

[24] A good choice for the probabilistic response function that is usually used in bankruptcy prediction studies and also adopted in our study, is the logistic function.

coefficients, β, that maximize the discriminatory power of a model we would like as many as possible $d_{i,j}'s$ to be positive. A way to achieve this is through the minimization of the following merit function:

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} max\left(0, \gamma - d_{i,j}(\beta)\right) \tag{2}$$

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

Based on the optimality conditions of minimizing *F(β)*, at the optimal solution, a number of $d_{i,j}'s$ must satisfy the condition $d_{i,j} = \gamma$ [25]. Hence, by selecting γ (close) to zero, we force a number of $d_{i,j}'s$ to be close to zero in absolute terms. In that case, a small change of the input data can easily induce $d_{i,j}'s$ to change signs which in turn will cause a change in the AUROC. This may be particularly evident in the case of out-of-sample data. That is, by training a model to produce $d_{i,j}'s$ close to zero, may yield a model with poor generalized ability and consequently the out-of-sample AUROC will be very sensitive. On the other hand, selecting γ (close) to one, coefficient estimates can blow up and provide unreasonable results. We suggest using a validation procedure to select the parameter value. In this study, we further divide our training sample into training (70%) and validation (30%) sets. We train the models by choosing from the set of parameter values $\gamma = \{0, 0.1, 0.2, \ldots, 1\}$ and keep the value that gives the highest AUROC on the validation set. For instance, using our private and public firm models we find that γ equals 0.3 and 0.1 respectively. Then we merge the training and validation sets, to train the models and test their performance on the testing set.

However, the surrogate function in Eq. (2) is non-differentiable when $\gamma - d_{i,j}(\beta) = 0$. To overcome this problem and thus being able to use traditional gradient-based optimization algorithms, we should replace the term $max(0, z)$ with a differentiable function. Note that,

[25] This draws on results from Charalambous (1979).

we can solve Eq. (2) using linear programming provided that the response function is linear with respect to the coefficients, β. We examine that in a subsequent section. Here, the probability is a non-linear function and as such we should use non-linear optimization algorithms to obtain the coefficients. We replace the term $max(0, z)$ by the following ε-smoothed differentiable approximation, $h_\varepsilon(z)$:

$$h_\varepsilon(z) = \begin{cases} 0, & z \le -\varepsilon/2 \\ \frac{1}{2\varepsilon}(z + \varepsilon/2)^2, & -\varepsilon/2 < z \le \varepsilon/2 \\ z, & z > \varepsilon/2 \end{cases} \tag{3}$$

where ε is a small positive number close to zero. Here we set $\varepsilon = 0.001$. The ε-smoothed function $h_\varepsilon(z)$, which we graphically present in the top plot of Figure 1, is a shifted version of the smoothed function used previously by Charalambous et al. (2007) to value call options[26].

[Insert Figure 1 here]

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

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} h_\varepsilon\left(\gamma - d_{i,j}(\beta)\right) \tag{4}$$

We further illustrate the role of γ by providing an example using our data. We estimate the private firm model using Eq. (4) as the objective function to obtain the coefficients and we calculate the $d_{i,j}'s$. We further estimate a logistic regression model but in that case the log-likelihood function is used in the optimization to obtain the coefficients and we also calculate the $d_{i,j}'s$. Figure 2 shows a sample of those $d_{i,j}'s$[27], produced by logistic regression (top plot) and by maximizing AUROC with the ε-smoothed function, setting γ =0 (middle plot) and γ =0.3 (bottom plot).

[Insert Figure 2 here]

Recall that we would like as many as possible of $d_{i,j}'s$ to be greater than zero. Hence, they should lie above the solid straight line. For the logistic regression, some lie above and some below. Using the ε-smoothed function, we want to make as many as possible negative $d_{i,j}'s$ to move above the straight line. Setting γ=0, we observe that all $d_{i,j}'s$ are close to zero. Some

[26] See also Pinar and Zenios (1994) for similar ε-smoothed functions.
[27] We chose a sample consisting of 289 $d_{i,j}'s$ that were calculated for a randomly selected healthy firm against all bankrupt firms.

cases, 21 in particular, that were negative according to the logistic regression became positive (denoted with green crosses) and one case that was positive became negative (denoted with a red star), highlighting the limitation of producing $d_{i,j}'s$ that are close to zero. Setting γ=0.3, not only more $d_{i,j}'s$ that were negative became positive (59 in particular), but now the majority lie well above the solid straight line, several also passing the γ parameter which are the points that lie above the dashed line. Notice now that none of the $d_{i,j}'s$ that were positive became negative because the higher value of γ, causes $d_{i,j}'s$ to be well above zero and as a consequence, AUROC will not be sensitive.

3.2.2 Linear Response Function

In this section we examine several specifications for Eq. (2) but now we consider a linear response function. That is, the bankruptcy score, s, is given by $\beta^T X$ and thus, $d_{i,j}(\beta) = s_B^i(X_i, \beta) - s_H^j(X_j, \beta) = \beta^T(X_B^i - X_{NB}^j)$. We firstly show how to solve the problem by linear programming and finally we compare various merit functions that are continuous and differentiable, accounting also for outliers that frequently characterize financial data. Our aim is to find the specification that yields the best performance in terms of AUROC. It would be useful to say here that the choice of parameter γ in the case of linear response function, affects only the scaling of coefficient estimates, β, and will not affect the ranking of firms, meaning that the AUROC will not be affected by the choice of γ. It is a common practice to set γ=1[28].

3.2.2.1 L1-max Merit Function

Under the specifications introduced in this section, the non-differentiable function, F(β), becomes:

$$F(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} max\left(0, 1 - d_{i,j}(\beta)\right) \quad (5)$$

[28] We borrow the idea of setting γ=1 from the literature of classification, based on support vector machines (i.e. Vapnik, 1995; Vapnik, 1998). Again, provided that the parameter γ is a (non-zero) positive number, it will only affect the scale of the coefficients, but the ranking of firms will remain the same and therefore AUROC will not change.

The above function is the normalized linear sum of max(0,z) and it is similar to the one proposed by Vapnik (1995) in the context of classification. Minimizing F(β), is equivalent to the following linear programming problem:

$$\min_{z,\beta} F(z,\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} z_{i,j} \tag{6}$$

s.t.

$$z_{i,j} \geq 0 \qquad i=1,2,\ldots,n \;\; and\; j=1,2,\ldots,m$$
$$z_{i,j} \geq 1\text{-}d_{i,j}(\beta)$$

In the following sub-sections, we introduce other merit functions that are continuous and differentiable in order to use gradient-based optimization algorithms.

3.2.2.2 Function 1: L2-max Merit Function

A straightforward function that is continuous and differentiable, is the squared function $\{\max(0,z)\}^2$, as shown in the bottom plot of Figure 1. In that case, the merit function to be minimized is the following:

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[\max\left(0, 1 - d_{i,j}(\beta)\right)\right]^2 \tag{7}$$

3.2.2.3 Function 2: Exponential Square Merit Function

A drawback of the function in Eq. (7) and to less extend the function in Eq. (5) is that, both are sensitive to outliers that eventually can affect the optimization and consequently the coefficient estimates. To this end, we use an exponential square function similar with the one used by Feng et al. (2016) in the context of classification, as shown in Eq. (8).

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\sigma^2\left(1 - \exp\{-\left[\max\left(0, 1 - d_{i,j}(\beta)\right)\right]^2 /\sigma^2\}\right) \tag{8}$$

Note that the following holds true for the function $\sigma^2\{1\text{-}\exp[-(.)\,]\,\}$:

$$\sigma^2\{1 - \exp[-(.)\,]\,\} = \begin{cases} 0, & d_{i,j}(\beta) = 1 \\ \sigma^2, & d_{i,j}(\beta) \rightarrow -\infty \end{cases} \tag{9}$$

The role of σ^2 and the way it affects the optimization, however, is not clear. Here for simplicity we set σ^2=1. Finally, we plot the exponential square function in the bottom plot of Figure 1. As can be seen from the plot, while the L2-max function sharply increases for large values of the variable *z*, the exponential square function is restricted up to σ^2, which in our case equals 1.

3.2.2.4 Function 3: ε-Smoothed Merit Function

Another function that can be used is the ε-smoothed function that we have introduced in an earlier section, yielding the following merit function:

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} h_\varepsilon\left(1 - d_{i,j}(\beta)\right) \tag{10}$$

3.2.2.5 Function 4: Exponential ε-Smoothed Merit Function

We consider the exponential ε-smoothed function to avoid problems induced by outliers. In this case, the function to be minimized, is the following:

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} \sigma^2\{1 - \exp[\,-h_\varepsilon\left(1 - d_{i,j}(\beta)\right)/\sigma^2\,]\,\} \tag{11}$$

As with Eq. (8), we set σ^2=1.

3.2.2.6 Other Approaches

We also consider two other approaches proposed by Miura et al. (2010) and Kraus and Kuchenhoff (2014), to maximize AUROC of credit scoring models[29]. Miura et al. (2010) suggest a sigmoid function as an approximation of Eq. (1). Specifically, they maximize the following objective function[30]:

$$F(\beta) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1}{1 + \exp\left[-d_{i,j}(\beta)/\sigma\right]} \tag{12}$$

---

[29] Several related approaches for maximizing AUROC have been proposed and applied in other domains, such as in computer science (Tayal et al., 2015). In our study, we focus on approaches suggested in credit scoring studies.

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

where $d_{i,j}(\beta) = \beta^T(X_B^i - X_{NB}^j)$. However, unlike the functions that we previously introduced, it treats all $d_{i,j}'s$ in the same way, whereas our functions, give more emphasis on the "bad" cases, for example, when a healthy firm has higher bankruptcy score than a bankrupt firm. Further, the authors consider only a linear response function. In this paper, we also consider a probabilistic response function which can be used by any modeling approach, such as neural networks.

Finally, Kraus and Kuchenhoff (2014) suggest using directly Eq. (1) as the objective function and implementing derivative-free methods (such as Nelder and Mead, 1965) to optimize the coefficients. The optimization algorithm that is used, however, assumes that the objective function is continuous, which is not the case for Eq. (1). Also, this approach while is easy to implement, ignores information provided by the gradient which could increase the accuracy of the coefficients after the optimization process and thus we believe that using specifications with differentiable functions is a better choice[31].

3.2.3 Outline of the Methodologies Used to Maximize AUROC

In Figure 3 we outline the models along with their response functions and the various merit functions that we use to maximize AUROC.

[Insert Figure 3]

The advantage of using a probabilistic response function (panel A), is that it can be applied regardless of the approach used to model the probability of bankruptcy. In this study, we consider a two-layer feed-forward neural network, since it is a widely-used modeling approach in bankruptcy prediction studies (Kumar and Ravi, 2007). The specifications of the neural network are as follows: 1) In the hidden layer we use two neurons, selected based on a validation procedure[32], 2) We use a logistic transfer function in the hidden layer and 3) We use one neuron in the output layer, to produce the probability of bankruptcy, using the

[31] We use the optimization toolbox in Matlab. For the linear programming, we use the *linprog* command with the dual-simplex algorithm. For Kraus and Kuchenhoff (2014) we use the *fminsearch* command while for the rest problems with continuous and differentiable functions, we use the *fminunc* command with the trust-region algorithm.

[32] We divide the training sample (1990-2006) into training (70%) and validation (30%). We train the neural network using one, two, three and four neurons, starting also from various initial coefficient values, and we select the number of neurons that performs the best (in terms of AUROC) in the validation set. Then we merge the two samples to train the neural network and measure the performance on the testing set (2007-2015).

logistic function. Then the probability of bankruptcy is used in the merit function and by using standard optimization methods, we obtain the coefficients.

We also consider a simpler model where the linear score, ψ= $\beta^T X$, is converted directly to a probability using the logistic function, p(β)=1/[1+exp{-ψ(β)}] and by using the merit function, such as the one in Eq. (4) we obtain the coefficients. Notice that, this simpler model would be equivalent to a logistic regression if the merit function was the log-likelihood function.

In the case of a linear response function (panel B) the linear score, ψ= $\beta^T X$=s(β), is directly entered to the merit function. To this end, we consider four different merit functions as shown in panel B of Figure 3.

3.3 Information Content Tests

We further consider information content tests, also done in related studies (see for instance Hillegeist et al., 2004; Agarwal and Taffler, 2008; Charitou et al., 2013; Bauer and Agarwal, 2014). In such tests the out-of-sample bankruptcy probabilities produced by various models, such as by models with maximized AUROC, are entered as inputs to logistic regression models and we are interested to assess their explanatory power. Recall that in the case of linear response functions, the output is not a probability. For consistency, we use the logistic function to convert the linear score into a probability. Note that the logistic function provides a monotonic transformation and thus will not change the ranking of firms (and consequently the AUROC will not be affected). In particular, we estimate the following panel logit specification:

$$p(Y_{i,t+1} = 1 | \, prob_{i,t}) = p_{i,t} = \frac{e^{a_t + \beta * prob_{i,t}}}{1 + e^{a_t + \beta * prob_{i,t}}} = \frac{e^{a * Rate_t + \beta * prob_{i,t}}}{1 + e^{a * Rate_t + \beta * prob_{i,t}}} \tag{13}$$

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

The specification in Eq. (13) is equivalent with the hazard model specifications used in related bankruptcy prediction studies, such as Hillegeist et al. (2004), Agarwal and Taffler (2008), Bauer and Agarwal (2014) etc. Specifically, Shumway (2001) argues that a panel logit model, like the one in Eq. (13), is equivalent with a hazard rate model and therefore standard log-likelihood procedures can be used to estimate the logit model in Eq. (13), with a minor adjustment that we explain below.

The model in Eq. (13) represents a multi-period logit model as it includes observations for each firm across time. However, the inclusion of multiple firm-year observations per firm yields understated standard errors because the log-likelihood objective function, which is maximized to estimate the multi-period logit model, assumes that each observation is independent from each other. This is a wrong assumption since firm observations at time $t+1$ cannot be independent from firm observation at time t. Failing to address this econometric issue, could lead to wrong inference regarding the significance of the individual coefficients. Similar with Filipe et al. (2016), we use clustered-robust standard errors to adjust for the number of firms in the sample but also for heteroskedasticity (Huber,1967 and White, 1980).

3.4 Economic Analysis of Bankruptcy Models

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

3.4.1 Calculating Credit Spreads

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

$$CS_i = \frac{p(Y=1|S=i)}{p(Y=0|S=i)} LGD + k \tag{14}$$

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 of firms that went bankrupt the following year divided by the number of firms in the group. Furthermore, *k*=0.3% and *LGD*=45%.

3.4.2 Granting Loans and Measuring Economic Performance

To evaluate economic performance, we assume that banks compete to grant loans to prospective firm-customers between the period 2007-2015. Each bank uses a bankruptcy model that has been estimated in the period 1990-2006. The bank sorts those customers according to their riskiness and rejects the bottom 1% with highest risk. The remaining firms are classified in 10 groups of equal size and firms from each group are charged a credit spread that has been obtained from the period 1990-2006. Finally, the bank that charges the lowest credit spread for the customer (i.e for the firm-year observation) is granting the loan. Two measures of profitability are used. The first one, Return on Assets (ROA) is defined as Profits/Assets lent and the second one, Return on Risk-Weighted Assets (RORWA) takes into consideration the riskiness of the assets, defined as Profits/Risk-Weighted Assets. Risk-Weighted Assets are obtained from formulas provided by the Basel Committee on Banking Supervision (2006).

4 Results

In this section we present the results of our tests. We start the analysis by evaluating the AUROCs of the different specifications. Next, we assess the information content of models with maximized AUROC and finally, we examine the economic effects of using models with maximized AUROC.

4.1 AUROC Results

Table 3 shows the out-of-sample performance (2007-2015) of the various specifications for maximizing AUROC and also includes a logistic regression model, trained by maximizing the log-likelihood function.

[Insert Table 3 here]

Overall, models that are trained to maximize AUROC perform better out-of-sample as compared to the logistic regression model that is trained to maximize the log-likelihood function, indicating that the functions we examine perform well out-of-sample, in discriminating firms that will go bankrupt the next year. Notice that when using the exponential square and exponential ε-smoothed merit functions, we obtain higher AUROCs compared to using the L2-max and ε-smoothed merit functions. These functions prevent the outliers, which are usually included in the financial data, to influence the optimization process, allowing for a smoother calculation of the coefficient estimates.

Interestingly, among the models with a linear response function, the one that uses the exponential ε-smoothed merit function performs the best when using both the private and public firm models (AUROCs equal to 0.9247 and 0.9480 respectively). We believe that this result bears a possible explanation. As can be seen in the bottom plot of Figure 1, the exponential ε-smoothed function is the only function that gives emphasis on small "z" values (whereas the L2-max and exponential square functions give less or no emphasis), but also accounts for outliers. The L1-max function also emphasizes on the small values but does not account for the outliers.

However, from all the models we consider, the neural network model with a probabilistic response function is the best performing model, which is consistent with the notion that neural networks outperform simpler modeling approaches (Zhang et al., 1999; Kumar and Ravi, 2007; Lessmann et al., 2015).

The effect by maximizing the AUROC, as expected, is more pronounced in the case of "private firms model" where only limited information is available (i.e. financial information), hence there is more space to improve the performance. In contrast, the effect is less pronounced in the case of "public firms model", since the inclusion of market data in addition to financial data, further increases the forecasting power of the models. In fact, a logistic regression model achieves high discriminatory power, measured by AUROC (0.9425), but models whose coefficients are estimated by maximizing the AUROC outperform the former.

From the results in this section, we suggest using the neural network model trained to maximize AUROC. For the user interested in simpler models, such as models with linear response functions, we suggest the implementation of the exponential ε-smoothed function, suitable for data with outliers. For the remaining tests, we use these two approaches and compare them with a model estimated without maximizing AUROC, for example, a logistic regression model, in order to test the conjecture that models with maximized AUROC are more valuable risk management tools in bankruptcy prediction.

4.2 Information Content Results

In this section we report the results from information contest tests. We compare the information contained in out-of-sample bankruptcy probabilities produced by models without maximized AUROC, for example a logistic regression (Prob 1), and by models with maximized AUROC. We consider the best model in the linear response family that accounts for the outliers (Prob 2) and by the neural network model which we find to be the best performing model (Prob 3). Table 4 reports the results of logit models that include the out-of-sample bankruptcy probabilities as explanatory variables but also the annual bankruptcy rate (Rate) as the baseline hazard rate.

[Insert Table 4 here]

Panel A reports results from six models. Model 1, 2 and 3 include out-of-sample bankruptcy probabilities produced by the logistic regression model, by the best model among the linear response functions and by the neural network respectively, using only financial data (private firms model). Models 4-6 correspond to models 1-3 but include financial and market data (public firms model) for the estimation of the probability of bankruptcy.

According to the results, bankruptcy probabilities in models 1-6 are highly statistically significant, indicating that they carry significant information in predicting bankruptcy one year ahead, (coefficient estimates are significant at the 1% significance level). More importantly, bankruptcy probabilities produced by models with maximized AUROC contain significantly more information than bankruptcy probabilities produced by models without maximizing AUROC. In particular, in panel B, we use the Vuong (1989) test-statistic to test for differences in the log-likelihoods between various (non-nested) models. Results show that the log-likelihoods of models 2 and 3 are significantly different than model 1 (test-statistics are 5.38 and 8.21 respectively). We also document higher explanatory power of model 3 over model 2 (Vuong test-statistic is 3.40).

As far as models 4-6 is concerned, evidence confirms that models with maximized AUROC capture more bankruptcy-related information, according to log-likelihood comparisons. For instance, differences in the log-likelihoods of model 6 over model 4 and model 5 over model 4, are statistically significant (Vuong test-statistics are 7.71 and 3.84 respectively). Significant difference is also documented between the log-likelihoods of models 6 and 5, as the Vuong test-statistic is 5.61.

Overall, our results suggest that models with maximized AUROC provide probability estimates that contain significantly more information about bankruptcies over the next year compared to a logistic regression model, even when the increase in AUROC is relatively small (as in the case of our public firm models).

4.3 Economic Performance Results

So far we have considered discriminatory power and information contest tests to assess model performance. However, a bank is generally interested in the economic benefits arising by using bankruptcy prediction models in the decision-making process of granting loans to individual firms. Following Agarwal and Taffler (2008) and Bauer and Agarwal (2014), we consider a loan market worth $100 billion and four banks are competing to grant loans to prospective firm customers. We hypothesize that bank 1 is a “naïve” bank, using a generic corporate model such as Altman’s *Z*-score for its credit decisions. Further, bank 2 uses a statistical approach to develop a bankruptcy prediction model such as the logistic regression model developed in this study. Finally, banks 3 and 4 are more sophisticated in the sense that they use models with maximized AUROC. To this end, bank 3 uses the best model with the linear response function and bank 4 uses the neural network model. In Table 5, we report the results, for both private and public firm models.

[Insert Table 5 here]

Clearly, banks 3 and 4 which use models with maximized AUROC manage loan portfolios with higher quality relative to banks 1 and 2 which use Altman’s Z-score model and a logistic regression model respectively. This is evident by the lower concentration of bankruptcies they attract. In particular, the bankruptcy rate of bank's 4 portfolio is 0.060% and 0.053% when using the private and public firms model respectively and the bankruptcy rate for bank 3 is 0.18% and 0.12% using the private and public firms model respectively. In contrast, 0.50% and 0.25% of the loans provided by bank 2, using the private and public firms model respectively, file for bankruptcy the next year while for bank 1 the bankruptcy rates are 1.20% and 1.06% for the private and public firms model respectively. Notice that the higher the AUROC for the model (reported in Table 3), the better the quality of loans granted by the bank. Among the four banks, bank 4 which uses a neural network models manages the credit portfolio with the highest quality.

More importantly, banks 3 and 4 achieve superior economic performance compared to banks 1 and 2 on a risk-adjusted basis. For example, considering the private firms model, banks 3 and 4 which use models with maximized AUROC, earn 2.37% and 1.60% per dollar invested respectively, while banks 1 and 2 earn 0.16% and 0.84% respectively. Similarly, considering the public firms model, banks 3 and 4 earn 2.24% and 2.21% respectively, whereas banks 1 and 2 earns a lower return (0.25% and 1.80% respectively)[33]. Notice that the small differences between the AUROC of models (especially for bank 3 and 4), are depicted in the economic performance consistent with the findings of Bauer and Agarwal (2014).

4.4 Focusing on the financial crisis period 2007-2009

We perform an additional test by measuring the performance of the models during the financial crisis period 2007-2009. For this test, we compare the models with maximized AUROC (the neural network and the best linear response model) with models which are trained without maximizing AUROC, such as logistic regression (results are not tabulated). Overall, we find qualitatively similar results with the previous tests, suggesting that under stressed conditions, models which use AUROC as the optimization criterion, outperform logistic regression models.

5 Conclusions

In this paper, we develop bankruptcy prediction models where the discriminatory power as measured by the Area Under ROC curve (AUROC), is used as the optimization criterion

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

to obtain their coefficients and we highlight the benefits of using such models. First, we consider variables used in well-established bankruptcy prediction studies, such as Campbell et al. (2008) to construct a traditional model based on logistic regression and using these variables, we introduce and compare several methodologies to maximize AUROC. We consider linear and probabilistic response functions for the output of the models and we examine several merit functions used to obtain the coefficients. We find that the proposed approaches outperform, out-of-sample, the logistic regression models according to different tests. For the users interested in simpler models with a linear response function, we recommend the use of the model whose coefficients are estimated with a merit function that takes care of the outliers. For users interested in more advanced models, we recommend a neural network model since according to our findings is the best performing model. The results hold under various tests such as AUROC analysis, information content tests and in terms of economic benefits when banks use different bankruptcy prediction models in a competitive environment. We therefore suggest the consideration of AUROC as an optimization criterion when developing bankruptcy prediction models.

References

Agarwal, V., & Taffler, R. (2008). Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models. *Journal of Banking and Finance, 32*, 1541-1551.

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance, 23*, 589-609.

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

Basel Committee on Banking Supervision. (2006). International convergence of capital measurement and capital standards: A revised frameework.

Bauer, J., & Agarwal, V. (2014). Are Hazard Models Superior to Traditional Bankruptcy Prediction Approaches? A Comprehensive Test. *Journal of Banking and Finance, 40,* 432-442.

Blochlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking and Finance, 30,* 851-873.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In Search of Distress Risk. *The Journal of Finance, 63*, 2899-2939.

Charalambous, C. (1979). On conditions for optimality of the non-linear l1 problem. *Mathematical Programming, 17*, 123-135.

Charalambous, C., Christofides, N., Constantinide, E. D., & Martzoukos, S. H. (2007). Implied non-recombining trees and calibration for the volatility smile. *Quantitative Finance*, *7*, 459-472.

Charitou, A., Dionysiou, D., Lambertides, N., & Trigeorgis, L. (2013). Alternative Bankruptcy Prediction Models Using Option-Pricing Theory. *Journal of Banking and Finance, 37,* 2329-2341.

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

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*, 1447-1465.

Dwyer, D. W., Kocagil, A. E., & Stein, R. M. (2004). Moody's KMV RiskCalc v3.1 model. Moody's KMV.

Feng, Y., Yang, Y., Huang, X., & Mehrkanoon, S. (2016). Robust support vector machines for classification with nonconvex and smooth losses. *Neural Computation, 28,* 1217-1247.

Filipe, S. F., Grammatikos, T., & Michala, D. (2016). Forecasting distress in European SME portfolios. *Journal of Banking and Finance, 64,* 112-135.

Fitzpatrick , T., & Mues, C. (2016). An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. *European Journal of Operational Research, 249*, 427-439.

Hanley , J. A., & McNeil, B. J. (1982). The Meaning and Use of the Area Under a Receiver Operating Characteristics (ROC) Curve. *Radiology, 143*, 29-36.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the Probability of Bankruptcy. *The Review of Financial Studies, 9,* 5-34.

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

Keenan, S. C., & Sobehart, J. R. (1999). Performance measures for credit risk models. Moody's Risk Management Services.

Kraus, A., & Kuchenhoff, H. (2014). Credit scoring optimization using the area under the curve. *The Journal of Risk Model Validation, 8*, 31-67.

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

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*, 124-136.

Miura, K., Yamashita, S., & Eguchi, S. (2010). Area under the curve maximization method in credit scoring. *The Journal of Risk Model Validation, 4,* 3-25.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7,* 308-313.

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

Pinar, M. C., & Zenios, S. A. (1994). On smoothing exact penalty functions for convex constrained optimization. *SIAM Journal of Optimization, 4,* 486-511.

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

Sobehart, J. R., Keenan, S. C., & Stein, R. M. (2000). Benchmarking quantitative default risk models: A validation methodology. Moody's Investors Service.

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

Tayal, A., Coleman, T. F., & Li, Y. (2015). RankRC: Large-scale nonlinear rare class ranking. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 27*, 3347-3359.

Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394-419.

Vapnik, V. N. (1995). Support-vector networks. *Machine Learning, 20*, 273-297.

Vapnik, V. N. (1998). Statistical Learning Theory. New York: Wiley.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica, 57*, 307-333.

Westgaard, S., & van der Wijst, N. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research, 135*, 338-349.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*, 817-838.

Wu, Y., Gaunt, C., & Gray, S. (2010). A Comparison of Alternative Bankruptcy Prediction Models. *Journal of Contemporary Accounting & Economics, 6*, 34-45.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*, 16-32.

Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research, 22*, 59-82.

Tables

Table 1: List of financial and market variables

Variable	Detailed Description	Compustat Item
Panel A: Financial Ratios (Compustat)		
NITA	Net Income/Total Assets	NI/AT
EBITTA	Earnings Before Interests and Taxes/Total Assets	EBIT/AT
RETA	Retained Earnings/Total Assets	RE/AT
CASHTA	Cash and Short-Term Investments/Total Assets	CHE/AT
WCTA	Working Capital/Total Assets	WCAP/TA
STDTA	Debt in Current Liabilities/Total Assets	DLC/AT
TLTA	Total Liabilities/Total Assets	LT/AT
CLCA	Current Liabilities/Current Assets	LCT/ACT
EBITCL	Earnings Before Interests and Taxes/Current Liabilities	EBIT/LCT
NICL	Net Income/Current Liabilities	NI/LCT
CFOTA	Operating Cash Flows/Total Assets	OANCF/AT
CFOTL	Operating Cash Flows/Total Liabilities	OANCF/LT
SLTA	Sales/Total Assets	SALE/AT
LOGASSETS	Natural logarithm of Total Assets	LOG(AT)
Panel B: Market Variables (CRSP)		
VOLE	Annualized volatility of daily equity returns	
EXRET	Annualized equity return minus the value-weighted return of NYSE, AMEX, NASDAQ stocks	
LOGPRICE	Natural logarithm of the stock price, at the fiscal-year end	
RSIZE	Natural logarithm of firm's market capitalization over the total market capitalization of NYSE, AMEX, NASDAQ stocks	
MB	Firm's market capitalization over book value of equity (Market-to-Book ratio)	
TLMTA	Total Liabilities/ (Market Capitalization + Total Liabilities)	
NIMTA	Net Income/ (Market Capitalization + Total Liabilities)	
CASHMTA	Cash and Short-Term Investments/ (Market Capitalization + Total Liabilities)	

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

Table 2: Descriptive statistics for the selected variables

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

This table reports descriptive statistics for several financial and market variables, one year prior to bankruptcy for both bankrupt and healthy firm observations. The definition of the variables is given in Table 1.

Table 3: AUROC results

Methodology		Private Firms Model	Public Firms Models
Logistic Regression		0.8991	0.9425
Probabilistic Response Function			
$h_\varepsilon[\gamma-(P_B-P_{NB})]$	Logistic	0.9221	0.9470
	Neural Network	**0.9331**	**0.9508**
Linear Response Function			
$\max[0, 1-\beta(X_B-X_{NB})]$		0.9147	0.9456
$h_\varepsilon[1-\beta(X_B-X_{NB})]$		0.9151	0.9468
$\{\max[0, 1-\beta(X_B-X_{NB})]\}^2$		0.9121	0.9462
$1-\exp(-\{\max[0, 1-\beta(X_B-X_{NB})]\}^2)$		0.9129	0.9473
$1-\exp(-h_\varepsilon[1-\beta(X_B-X_{NB})])$		**0.9247**	**0.9480**
Other Approaches			
Miura et al. (2010)		0.9188	0.9471
Kraus and Kuchenhoff (2014)		0.9046	0.9473

This table reports AUROC results for a logistic regression model as well as for models with maximized AUROC, when we consider probabilistic and linear response functions, as well as various merit functions used to obtain the coefficients. The models are trained in the period 1990-2006 and the table reports results in the out-of-sample period 2007-2015. For the logistic regression, the log-likelihood function is used as a merit function to obtain the coefficients.

Table 4: Information content tests results

Panel A: Logit models estimation

	Private Firms Model			Public Firms Model		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Prob1	0.296			0.271		
	(7.27)			(10.33)		
Prob2		0.103			0.065	
		(5.63)			(6.11)	
Prob3			0.069			0.195
			(19.25)			(16.60)
Rate	-0.218	-0.711	-1.128	-0.370	-0.747	-1.07
	(-0.59)	(-1.93)	(-2.91)	(-0.92)	(-1.95)	(-2.28)
Constant	-5.52	-13.24	-7.98	-5.54	-9.64	-20.32
	(-30.42)	(-7.78)	(-26.33)	(-28.84)	(-9.44)	(-18.72)
Log-Likelihood	-774.96	-656.72	-601.53	-728.26	-639.96	-554.91
Pseudo-R^2	8.56%	22.51	29.03%	14.07%	24.49%	34.53%

Panel B: Vuong test statistics for differences in log-likelihoods

Models	1	2	3	Models	4	5	6
3	8.21	3.40	-	6	7.71	5.61	-
2	5.38	-		5	3.84	-	
1	-			4	-		

This table reports results from information content tests. Panel A shows estimation of six logit models. Model 1, 2 and 3 include out-of-sample (2007-2015) bankruptcy probabilities produced by logistic regression, and by models with maximized AUROC (the best model with linear response function and by a neural network) respectively, using financial data only. Models 4-6 correspond to models 1-3 but using financial and market data. All the models include the Rate as proxy for the baseline hazard rate, which is the prior year bankruptcy rate in our sample. The last two rows of the panel reports log-likelihood and pseudo-R^2 for each model. Panel B reports Vuong test statistics for differences in the log-likelihoods between the six models.

Table 5: Economic performance

	Private Firms Model				Public Firms Model			
	Bank1	Bank2	Bank3	Bank4	Bank1	Bank2	Bank3	Bank4
Credits	5649	6253	3410	13278	8083	3228	7785	9491
Market Share (%)	19.68	21.78	11.88	46.25	28.15	11.24	27.12	33.06
Bankruptcies	68	31	6	8	86	7	9	5
Bankruptcies/Credits (%)	1.20	0.50	0.18	0.060	1.06	0.25	0.12	0.053
Average Spread (%)	0.56	0.37	0.35	0.33	0.53	0.36	0.34	0.34
Revenues ($M)	109.25	81.41	41.41	153.15	149.67	40.08	91.45	111.79
Loss($M)	97.18	44.30	8.57	11.43	122.90	11.43	12.86	7.15
Profit($M)	12.07	37.11	32.84	141.72	26.77	28.65	78.59	104.64
Return on Assets (%)	0.061	0.17	0.28	0.31	0.095	0.25	0.29	0.32
Return on RWA (%)	0.16	0.84	**1.60**	**2.37**	0.25	1.80	**2.21**	**2.24**

This table reports economic results for four banks in a competitive loan market worth $100 billion. Bank 1 is a bank using simply the Altman's Z-score model for estimating the bankruptcy score. Bank 2 uses a statistical approach, such as the logistic regression model developed in this study. Banks 3 and 4 are more sophisticated, using models with maximized AUROC. Bank 3 uses the best model with a linear response function and bank 4 uses the neural network model.

The banks sort prospective customers (2007-2015) and reject the 1% of firms with the highest risk. The remaining firms are classified in 10 groups of equal size and for each group, a credit spread is calculated, as described in the main text (section 5.3). The bank that classifies the firm to the group with the lowest spread is finally granting the loan. Market share is the number of loans given divided by the number of firm-years, Revenues = (market size)*(market share)*(average spread), Loss=(market size)*(prior probability of bankruptcy)*(share of bankruptcies)*(loss given default). Profit=Revenues-Loss. Return on Assets is profits divided by market size*market share and Return on Risk-Weighted-Assets is profits divided by Risk-Weighted Assets, obtained from formulas provided by the Basel Accord (2006). The prior probability of bankruptcy is the bankruptcy rate for firms between 1990-2006 and equals 0.42%. Loss given default is 45%.

Figures

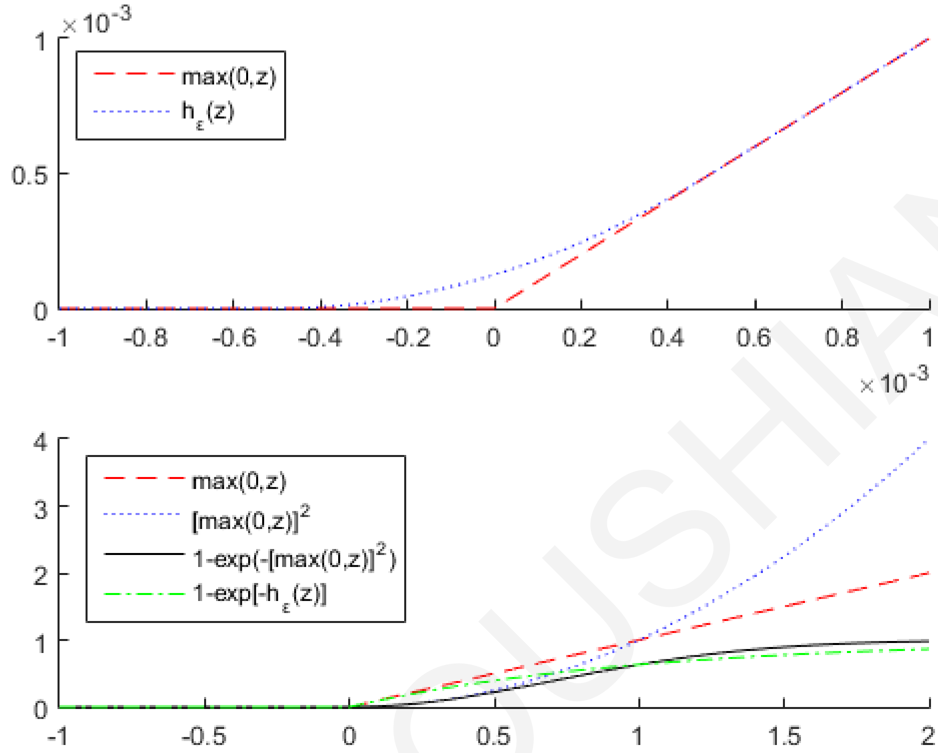

Figure 1: Plotting various merit functions. For the exponential-type functions, we set σ=1.

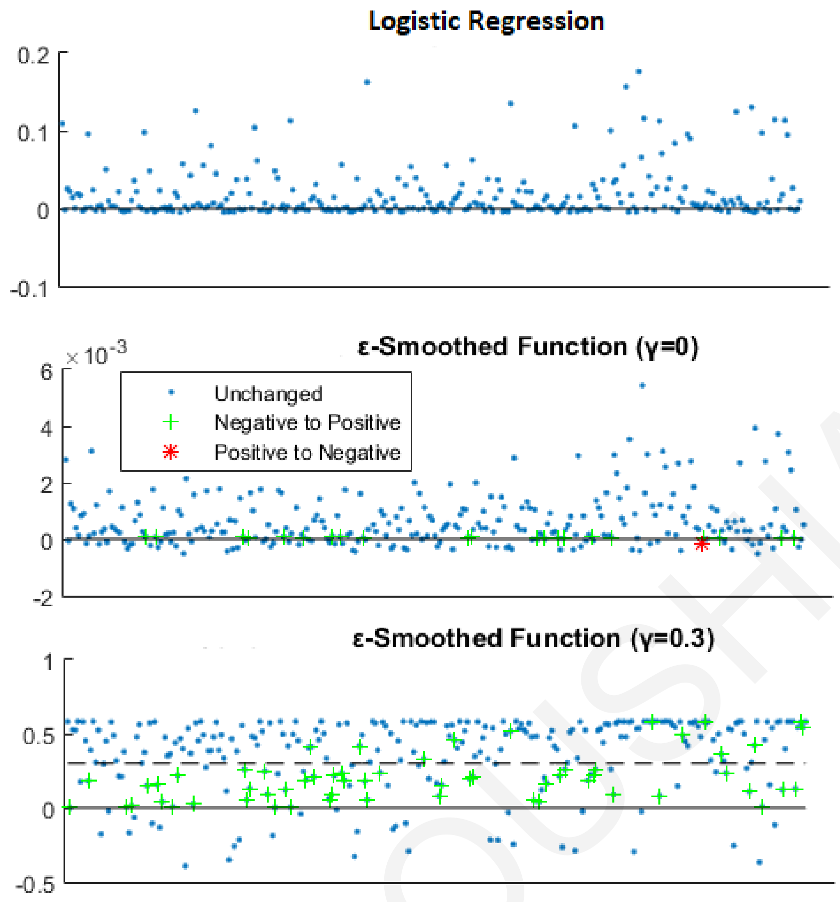

Figure 2: Plotting a sample of dij's, estimated using logistic regression and with models based on AUROC maximization, using the ε-smoothed merit function

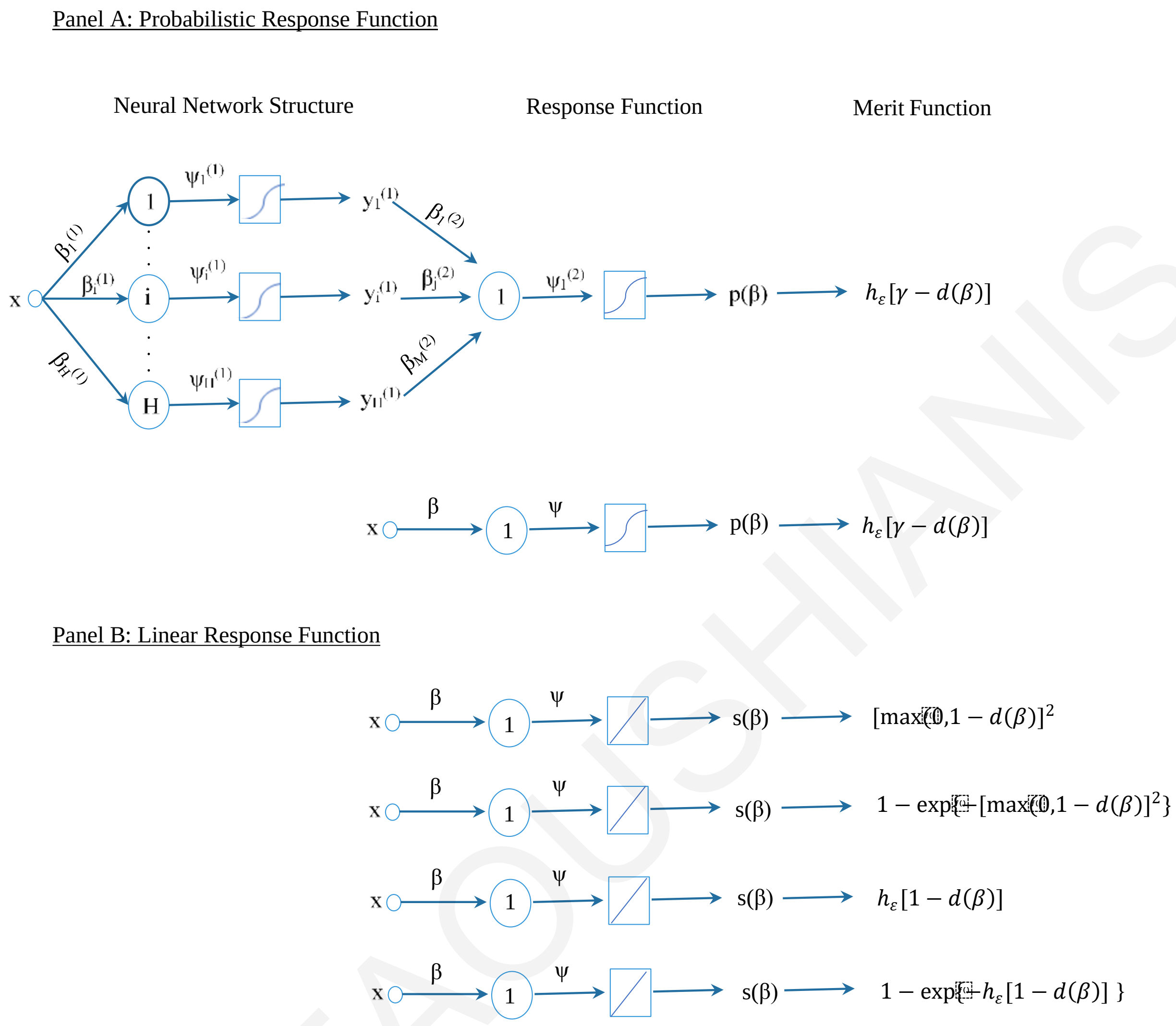

Figure 3: Outline of the modeling approaches, response functions and merit functions

CHAPTER 3

A semiparametric default forecasting model

Abstract

A fundamental limitation of structural (parametric) models for the estimation of the probability of default is that their most important parameters, the value of assets and volatility, are not observed in the market. In this paper we develop a methodology where the unobserved parameters are viewed as generalized functions. Using a nonparametric approach for their estimation, we obtain improved parameter values which enter a parametric model, yielding a semiparametric model. In this context, the Black-Scholes-Merton model is used as a paradigm. Results show substantial improvement in the out-of-sample performance when comparing our semiparametric model with other alternative specifications of the Black-Scholes-Merton model in terms of discriminatory power, information content and economic impact.

1 Introduction

1.1 Background and Motivation

The Black-Scholes-Merton model (i.e. Black and Scholes, 1973 and Merton, 1974) is one of the most widely used corporate default forecasting models, which become popular among academics and practitioners in the early 2000's. In particular, Crosbie and Bohn (2003) is one of the earliest papers that provides detailed explanation of the model. Since then, it has become state of the art in the academic literature, as several papers have compared its predictive power with other widely-used models[34] (Hillegeist et al., 2004; Reisz and Perlich, 2007; Agarwal and Taffler, 2008; Bauer and Agarwal, 2014). Another strand of the literature has also tried to extend and improve the performance of the model through alternative estimation of its input parameters (Bharath and Shumway, 2008; Charitou et al.,2013; Afik et al., 2016). The present study is related with the second strand of the literature. In particular, we develop a new estimation technique that provides improved parameter values, eventually improving the forecasting power of the Black-Scholes-Merton (BSM).

The intuition behind the BSM model is very simple. The equity of the firm is viewed as a European call option underlying the assets of the firm and with strike price being the liabilities of the firm. At maturity, the firm defaults if assets value falls below liabilities. In this case, equity holders receive nothing but walk free due to their limited liability. In the opposite scenario, equity holders are the residual claimants after all obligations are paid and the firm continues as a going concern. In this setting, the probability of default is the probability that at maturity, the assets value worth less than the liabilities.

The empirical application of the model requires several parameters, like for instance, the value of assets, the volatility of asset value changes, the expected growth of assets and the liabilities. However, two of the most important parameters, the value of assets and the volatility of asset value changes, are not observed, which makes the implementation of the model a challenging task. The literature provides two different estimation techniques to obtain the unobserved parameters. The first one, is based on iterative procedures (implemented by Hillegeist et al., 2004; Vasallou and Xing, 2004; Campbell et al., 2008). However, as argued by Crosbie and Bohn (2003), such estimation approaches might yield

[34] Evidence in the literature is conflicting. For instance, Hillegeist et al. (2004) find that the Black-Scholes-Merton model performs better than the Altman (1968) and Ohlson (1980) models, whereas Agarwal and Taffler (2008) find that Altman (1968) performs better. Nevertheless, Bharath and Shumway (2008) and Campbell et al. (2008) find that it is not a sufficient statistic, suggesting that other information not included in the model might be useful for default prediction.

inaccurate probability of default estimations when market leverage changes too fast. In addition, numerical errors during the iterations might affect the estimation of the two unobserved parameters (Charitou et al., 2013), which makes the specific approach quite "noisy". To avoid the iteration approach, several papers have proposed a second approach which is based on "simplified" approximations for the two unobserved parameters (Bharath and Shumway, 2008; Charitou et al., 2013; Afik et al., 2016). We review the major estimation approaches in the subsequent section.

It is inevitable, therefore, that the estimation technique affects the performance of the model and by improving the estimation of the unobserved parameters would improve the forecasting power of the BSM model. In this paper, we develop an approach to obtain improved parameter values that are used in the parametric model (i.e. the BSM), based on a nonparametric approach. Specifically, we assume the value of assets and the volatility to depend on several exogenous variables that are elements of the vector *x*, through some unknown relationships. We estimate these unknown relationships through learning, by embedding in the model a nonparametric structure, such as neural networks. The inputs to the neural network are the variables in the vector *x*, and the outputs are the unobserved parameters which are the inputs to the parametric model, thus yielding a semiparametric model for the estimation of the probability of default. In this setting, the weights of the neural network are adjusted in order to maximize a merit function. The proposed approach provides an alternative estimation method that outperforms the "noisy" iterative procedures and the "simplified" approximations in out-of-sample forecasts.

The basic advantage of the proposed approach is that one does not need to make assumptions about the structure of the unobserved parameters, for example to impose any deterministic relationships to calculate them. Instead, by letting the unobserved parameters to depend on some exogenous inputs, *x*, through some unknown functions, the network is optimized accordingly to learn the unknown relationships, providing improved parameter values, while preserving the theoretical properties of the parametric model.

Semiparametric methods have been used by Bandler et al. (1999) which show that such network structures can be used to adjust the parameters of imperfect models to get more accurate outputs. In the context of options pricing, semiparametric methods have been used by Aït-Sahalia and Lo (1998) and Aït-Sahalia and Duarte (2003). Furthermore, Andreou et al. (2008) used semiparametric methods to obtain improved parameters (implied volatility, skewness and kurtosis), in options pricing. The results justify the implementation of the semiparametric approach, since the option prices were more accurate relative to other estimations approaches.

In this paper, for the first time, we use semiparametric models in the context of default prediction. In our framework, the imperfect model is the BSM model and the network structure adjusts the parameters of the BSM model (such as the value of assets and the volatility) in order to obtain improved parameters and eventually, more accurate probability of default outputs.

We use accounting and market data between 1989 and 2014 for non-financial U.S. public firms to estimate the probability of default with the various BSM specifications, over a one-year forecasting horizon (for defaults between 1990-2015). For the estimation of the semiparametric model, we divide the whole sample into two sub-samples; the in-sample period includes defaults between 1990 and 2006 and the model is used to make forecasts in the out-of-sample period which includes defaults between 2007 and 2015.

1.2 Main Findings

We compare the performance of our semiparametric model with alternative BSM specifications; When asset value and volatility are estimated based on iterative procedures (i.e. Hillegeist et al., 2004 and Vassalou and Xing, 2004) and when estimated using direct estimation approaches (Bharath and Shumway, 2008 and Charitou et al., 2013). Specifically, we use three distinct type of tests. 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 semiparametric model is substantially better than the competing approaches. In the second test, we compare the information content of the various BSM specifications. Results show that default probabilities produced by the semiparametric model contain significantly more information than default probabilities produced by the alternative BSM specifications. In the final test, we compare the economic impact arising when banks use the different BSM models in the decision-making process of granting loans to individual firms. We find that the bank which uses the semiparametric 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.

Several additional tests are conducted for robustness, including augmenting the sample of events with financially distressed firms-a practice which makes prediction more challenging. We find that our semiparametric approach substantially outperforms the alternative BSM specifications. Interestingly, using the new dataset with financially distressed firms, the semiparametric approach is also better than other widely-used methodologies such as the logistic regression and the nonparametric approaches.

The remainder of the paper is organized as follows: Section 2 describes the alternative BSM specifications, which are used as benchmark, while section 3 describes our methodology to obtained improved parameter values for the BSM model. Section 4 discusses the data and section 5 describes the three distinct-type of tests we employ in order to test the performance of the models. Section 6 discusses the main results, and section 7 provides additional results for robustness and section 8 concludes.

2 BSM Model and Estimation of Asset Value and Volatility

2.1 Black-Scholes-Merton Model

Since equity 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:

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

where

$$d_1 = \frac{ln\left(V/_F\right) + (r + 0.5\sigma_V^2)T}{\sigma_V\sqrt{T}} \tag{2}$$

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

Here, *V* is the value of assets, *F* the liabilities of the firm, σ_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 Eq. (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 default 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 default, since d_2 is estimated using the riskless rate of return, *r*. We estimate the real-world probability of default, by substituting *r* with the real growth of assets, μ. Hence, it is straightforward to show that the probability of default, *PD*, is given by the following formula[35].

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

The ratio inside of Eq. (4), known as the distance-to-default, gives the number of standard deviations the value of assets must drop in order the firm to default (i.e. how far the firm is away from default).

However, the two most critical inputs in Eq. (4), *V* and σ_V, are not observed in the market which makes the estimation of the probability of default a challenging issue. Due to this, there was a burgeoning academic literature since the early 2000's regarding the estimation of *V* and σ_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 σ_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 default probability estimation, this approach has been used, for instance, by Hillegeist et al. (2004) and Campbell et al. (2008).

In particular, in the framework of options pricing there are two equations that we can solve iteratively to obtain the value of assets and the volatility. For the first equation, we solve the

[35] In the context of default prediction, Eqs (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.

standard options pricing formula given by Eq. (1), with respect to *V*, yielding the following equation[36]:

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

The second equation relates the (annualized) volatility of equity changes, σ_E, which is obtained from historical equity data, with the volatility of asset changes, σ_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, σ_V is calculated as follows:

$$\sigma_V = \frac{E\sigma_E}{VN(d_1)} \tag{6}$$

Starting from some initial values, for instance setting $V=E+F$ and $\sigma_V = \sigma_E$ on the RHS in Eqs. (5) and (6), we obtain a new set of *V* and σ_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 σ_V between two consecutive iterations are very small. When we obtain the two values, we can easily estimate μ as the return on asset values between two consecutive years i.e. $\ln(V_t / V_{t-1})$. Note that the two equations approach, is performed at a specific point in time, for instance, at the fiscal year-end prior to the year of default.

However, as argued by Charitou et al. (2013), convergence problems in the numerical procedures may yield numerical errors into the estimation of *V* and σ_V. Further, Crosbie and Bohn (2003) argue that depending on how quickly market leverage changes, solving Eqs. (5) and (6) biases the probability of default because Eq. (6) holds instantaneously. Moreover, because Eqs. (5) and (6) are derived from the basic assumptions underlying options theory, we believe that these assumptions pose restrictions in accurately estimating these unobserved parameters.

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 Eq. (5) to obtain daily

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

time-series for the value of assets[37]. Once we obtain the time-series of asset values, we estimate the annualized volatility, σ_V, from the logarithmic changes of asset values. Using this estimate of σ_V, in the next iteration we obtain a new series of asset values and estimate a new value for σ_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 σ_V. We set V=E+F and $\sigma_V = \sigma_E$.

Once we obtain the daily series of *V's,* we calculate the annualized growth of assets, μ, 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, 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 default.

2.2.3 Direct Estimation Approaches

Under this category, the estimation of V and σ_V is not based on iterative procedures at all but rather, the estimation relies on 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 σ_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 (μ=$r_{E,t-1}$). The authors show empirically that the BSM model performs better when V and σ_V are estimated with the simplified approximations, as opposed to more complex iterative procedures. They conclude that the accuracy stemming from BSM is due to its functional form and iterative procedures used to obtain V and σ_V are not useful.

In a similar notion, Charitou et al. (2013), which we denote as CDLT (2013), suggest the estimation of V and σ_V directly from equity data. In their study, they use the sum of equity

[37] 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.

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 (μ) and volatility (σ_V) are obtained. We slightly modify CDLT (2013), by estimating the aforementioned 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). The authors, however, do not compare their results using the two equations approach, or the single equation approach.

In the following section we present a nonparametric methodology where improved parameters enter the BSM parametric model, yielding a semiparametric model, avoiding in that way the estimation of the parameters by solving the equations or using simplified approximations as described above.

3 Methodology: A Semiparametric Model

3.1 The General Case

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

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

where $p = [p_1, p_2, \dots, p_L]$ 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:

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

In Eq. (9), $p^- = [p_1^-, p_2^-, \dots, p_M^-]$ is the vector which corresponds to the unobservable parameters, and $p^+ = [p_{M+1}^+, p_{M+2}^+, \dots, p_L^+]$ 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:

$$p_1^- = f_1(x) \tag{10}$$

$$\vdots$$
$$p_M^- = f_M(x)$$

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 nonparametrically through learning, for *i=1,2,…,M*. In this context, the probability of default is estimated as:

$$PD = f_{PM}(z, p^+) \quad (11)$$

where $z = [f_1(x), f_2(x), \dots, f_M(x)]$ refers to the vector with the variables that are determined through the nonparametric estimation of the unknown functions. In fact, the vector *z* provides improved parameter values to the model in Eq. (11), which we refer to as semiparametric model. Figure 1 provides a schematic representation of the proposed approach.

[Insert Figure 1 here]

As can be seen from the figure, the probability of default is estimated using the functional form of the parametric model but using two sets of inputs: 1) the inputs that 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 nonparametrically (i.e. *x* are the inputs to the nonparametric model that produce the outputs *z*). In this context, *z* and consequently *PD*, depend on the weights imposed by the nonparametric model. 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, $x_n = [x_{1n,} x_{2n}, \dots, x_{kn}]$, is associated with a known target, t_n, where *n=1,2,…, N* and *k* is the number of variables. In the context of default 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 equals 1 if the corresponding firm-observation defaults and 0 otherwise. The output of the parametric model, *PD(w),* with the associated targets, *t*, are used in the merit function which is optimized in order to obtain the weights of the nonparametric model and consequently the final output, which is the probability of default, *PD.* The nonparametric model here serves as an auxiliary mechanism which adjusts the parameters of the parametric model during the training phase, until the merit function is optimized. Note that both the nonparametric and parametric models belong to the network structure. This is important because in this setting, the nonparametric model embeds knowledge from the parametric model.

In this study, we use a feedforward neural network since it is the most common 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. Furthermore, neural networks have been successfully applied in the context of default prediction. For example, Kumar and Ravi (2007) in a comprehensive review for the work done during 1968-2005, report that in general, neural networks outperform other popular approaches for default prediction. Therefore, neural networks is 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 2, is consisted with *H* units, with the *i-th* unit connected with the input features, *x*, through the k-dimensional weight vector $w_i^{(1)}$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,…,H.* The outputs from the first layer are further processed in 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 $w_j^{(2)}$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,…,M.*

[Insert Figure 2 here]

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

$$y_1^{(2)} = f_1^{(2)}\left[w_{10}^{(2)} + \sum_{i=1}^{H} w_{1i}^{(2)}\, f_i^{(1)}\left(w_{i0}^{(1)} + \sum_{j=1}^{K} w_{ij}^{(1)}\, x_j\right)\right] \tag{12}$$

$$\vdots$$

$$y_M^{(2)} = f_M^{(2)}\left[w_{M0}^{(2)} + \sum_{i=1}^{H} w_{Mi}^{(2)}\, f_i^{(1)}\left(w_{i0}^{(1)} + \sum_{j=1}^{K} w_{ij}^{(1)}\, x_j\right)\right]$$

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 semiparametric model.

Overall, there are several advantages by using our proposed approach. First, we do not need to impose a priori ad-hoc or simplified approximations for the parameters of the parametric model. Instead, by treating (some of) the parameters as generalized functions, the network structure optimizes the weights accordingly, to determine the relationships between the input features and the parameters under consideration, yielding improved parameters that enter the parametric model. Second, we utilize the strong learning capabilities of the nonparametric model while preserving the theoretical properties of the parametric model. That is, the probability of default is estimated using the underlying theory of the parametric model, while the nonparametric model embeds knowledge from the parametric model which is useful during the training phase of the network.

3.2 The Case of BSM Model

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

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

Note that the numerator inside the logarithm in Eq. (13), $Ve^{\mu T}$, 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 default is given by the following formula:

$$PD = N(-DD) = N\left(-\frac{ln(E_L) - 0.5\sigma_V^2 T}{\sigma_V\sqrt{T}}\right) \qquad (14)$$

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

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

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

The RHS of Eqs. (15) and (16) are the generalized (unknown) functions between the input features, *x*, and E_L and σ_V, that the neural network seeks to learn by optimizing the weights of the network structure. These two outputs are entered as inputs to the BSM model and thus obtaining the probability of default:

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

Notice that the difference between Eqs. (14) and (17) is that the latter depends on the weights imposed to E_L and σ_V through the neural network and as a consequence, the probability of default, *PD*, is a function of the weights, yielding a semiparametric model. For a sample of *N* observations, the weights of the neural network are obtained by maximizing the Log-Likelihood, LL, defined as follows:

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

where

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

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. At 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[38] (for simplicity we drop the subscript *n*):

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

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

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

[38] The notation we use here for the gradient used for the adaptation of the weights of the neural network, is based on the principles from Charalambous (1992) for the efficient training of neural networks.

The quantity $\frac{\partial PD(Y^{(2)})}{\partial y_j^{(2)}} \equiv \frac{\partial f_{PM}}{\partial p_j}$ 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≡$w_j^{(2)}$,

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

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

When w≡$w_i^{(1)}$,

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

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

3.3 Specifications of the Nonparametric Model

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, notice that the default 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 default 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 default. 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, σ_V. Thus, using the variables from Campbell et al. (2008) as inputs to the neural network, is a reasonable choice.

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 (see Andreou et al., 2008). 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. 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 neuron, in order to account for different initialization points. 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, $f_H(.) = \frac{1-\exp\left(-2\psi_i^{(1)}\right)}{1+\exp\left(-2\psi_i^{(1)}\right)}$, 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 σ_V must be non-negative and within reasonable values. In this case, we use a modification of the log-sigmoid function as follows; $f_M(.) = a + \frac{b-a}{\left(1+\exp\left[-\psi_j^{(2)}\right]\right)}$, 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 σ_V is to be estimated, a=min (σ_E) and b= max (σ_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 $\sigma_V \rightarrow a$.

4 Data

4.1 Sample

Our sample of defaulted firms consists of 420 non-financial U.S. public firms that file for bankruptcy under Chapter 7 or Chapter 11 over the 26-year period 1990-2015 and have all data available in Compustat and CRSP one year prior to bankruptcy. Bankrupt firms were sourced from the database BankruptcyData. The final sample contains about 94,000 default and healthy firm-year observations. The distribution of observations across the years is shown in Table 1.

[Insert Table 1 here]

4.2 Variables Construction

To construct assets value, *V*, and the volatility, σ_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*, at fiscal year-end as the closing stock price * shares outstanding and the annualized volatility of daily equity returns, σ_E, for the entire fiscal year. 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, μ. 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 nonparametric model, 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 daily equity returns, 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 at fiscal year-end (*LOGPRICE*). Table 2 provides descriptive statistics for defaulted and healthy firm observations.

[Insert Table 2 here]

As can be seen from Table 2, there are several differences in the financial performance between defaulted and healthy firms. Specifically, defaulted 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 defaulted 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 defaulted firms and positive for healthy firms). Finally, equity returns for defaulted firms are more volatile relative to healthy firms (*SIGMA* is higher). In the last column of the table, t-tests for mean differences are reported. All mean differences are significant at the 1% level except from *BM* which is significant at the 10% level.

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 have 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) Discriminatory power based on AUROC, 2) Information content tests and 3) Economic benefits arising from using different default models.

5.1 Discriminatory Power

With this test we evaluate the ability of the models to discriminate the defaulted firms from the healthy firms. For a given cut-off probability, firms whose default probability is higher than the cut-off, are classified as defaulted and healthy otherwise. A way to measure discriminatory power is by counting the true predictions (percentage of defaulted firms correctly classified as defaulted) and the false predictions (percentage of healthy firms incorrectly classified as defaulted). 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 more the ROC curve approaches 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 (see for instance Hanley and McNeil, 1982 and Sobehart and Keenan, 2001), defined as follows:

$$\widehat{AUROC} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} I\left(PD_B^i > PD_H^j\right) \quad (24)$$

5.2 *Information Content Tests*

With this test, we evaluate the explanatory power of the models by including the out-of-sample default probabilities they produced in discrete hazard models. Following related studies, such as Hillegeist et al. (2004) and Agarwal and Taffler (2008), we estimate the following discrete logit model:

$$p\left(Y_{i,t+1} = 1 \middle| PD_{i,t}\right) = p_{i,t} = \frac{e^{a_t+\beta*PD_{i,t}}}{1+e^{a_t+\beta*PD_{i,t}}} = \frac{e^{a*Rate_t+\beta*PD_{i,t}}}{1+e^{a*Rate_t+\beta*PD_{i,t}}} \quad (25)$$

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

Shumway (2001) argues that a panel logit model like the one in Eq. (25) 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 to 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 and White, 1980).

5.3 *Economic Impact*

The previous tests measure the accuracy of the models. In this test, we examine how 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 default model to evaluate the credit-worthiness 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 firm-customer observations in 10 groups of equal size, with the first and tenth group being the firms with the lowest and highest default risk respectively, and a credit spread is calculated according to the following rule; Firms classified in the first group receive a credit spread k and firms in the remaining groups receive a credit spread CS_i, which is obtained from Blochlinger and Leippold (2006) and it is defined as follows:

$$CS_i = \frac{p(Y=1|S=i)}{p(Y=0|S=i)} LGD + k \tag{26}$$

where $p(Y=1|S=i)$ and $p(Y=0|S=i)$ is the average probability of default and non-default 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 default for the *i-th* group is the actual default rate for that group, defined as the number of firms that defaulted the divided by the number of firms in the group. Furthermore, k=0.3% and *LGD*=45%.

5.3.2 *Measuring Economic Performance*

Banks compete to grant loans to prospective firm-customers in the period 2007-2015. Using different default 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 (RORWA), takes into consideration the riskiness of the assets, defined as Profits/Risk-Weighted Assets. Risk-Weighted Assets are obtained from formulas provided by the Basel Committee on Banking Supervision (2006).

6 Results

This section discusses the results of the paper. We begin by reporting the estimation of asset and volatility values with respect to the different estimation approaches and finally, we report the performance of the various models, based on AUROC, information content and economic impact.

6.1 Asset Value and Volatility Estimation Results

In Table 3, we report asset values (expected leverage in the case of the semiparametric approach, E_L) and volatility values with respect to the different estimation approaches.

[Insert Table 3 here]

As expected, the ratio *V/F* is lower for defaulted firms in all cases. Differences in the mean values between defaulted and healthy firms are statistically significant. Similarly, σ_V is higher for defaulted 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 defaulted firms relative to healthy firms one year prior to bankruptcy. We conclude that our approach produces reasonable expected asset and volatility values.

6.2 AUROC Results

Table 4 presents the out-of-sample discriminatory power of the various approaches based on AUROC.

[Insert Table 4 here]

The key finding is that the semiparametric model substantially outperforms the competing approaches, suggesting that it is more powerful in discriminating the defaulted firms from the healthy firms. Specifically, the AUROC of the semiparametric model is 0.9387 whereas for the two and single equations approach, AUROCs are 0.8964 and 0.9026 respectively. According to Delong (1988) test, differences in AUROCs between the semiparametric 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 semiparametric 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 Delong (1988) test, differences in AUROCs between the semiparametric 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 semiparametric approach in discriminating defaulted from healthy firms relative to the alternative BSM specifications.

6.3 Information Content Results

Table 5 reports the results from information content tests. Models 1-5 are logit models that include as predictors, out-of-sample default probabilities produced by various BSM specifications. Models 1-2 include the default probabilities produced by estimating asset values and volatilities with the 2-Eqs. and 1-Eq. Approaches respectively (denoted as Prob 1 and Prob 2 respectively). Next, Models 3-4 include default probabilities produced by estimating asset values and volatilities based on BS (2008) and CDLT (2013) respectively and are denoted as Prob 3 and Prob 4 respectively. Finally, Model 5 includes out-of-sample default probabilities produced from our semi-parametric model (Prob 5).

[Insert Table 5 here]

According to the results, out-of-sample default probabilities produced by all BSM specifications are highly statistically significant at the 1% level, indicating that they carry significant information in predicting defaults one year ahead. More importantly, out-of-sample default probabilities produced by the semiparametric model contains significantly more information compared with the alternative approaches. 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 significant at the 1% level. The higher explanatory power of default probabilities produced by the semiparametric model, is also shown from the high pseudo-R^2 of Model 5 (28.60%) relative to the other Models which range from 16.43% to 21.21%.

From this test we conclude that the default probabilities obtained from the semi-parametric model contain significantly more information about future defaults as opposed to other BSM specifications. This finding confirms that our approach yields more accurate asset value and volatilities that improve the performance of the parametric model.

6.4 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 default models in the decision-making process of giving loans to individual firms. Thus, does the improved performance using the semi-parametric 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 default 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 semiparametric model.

[Insert Table 6 here]

As can be inferred from the table, Bank 5 manages a credit portfolio with the lowest concentration of defaults (0.08%) whereas for the remaining banks, concentration of defaults 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%[39].

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

6.5 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 increase the sample of events with firms which experienced financial distress during the sample period and we compare the semiparametric model with the various BSM specifications as well as with other widely-used methodologies.

[39] Results are robust with respect to different parameter specifications (k=0.002-0.004 and LGD=0.4-0.7).

6.5.1 Other Performance Statistics

Several other tests exist in the literature to evaluate the performance of default prediction models. In this section, we use the Kolmogorov-Smirnov (KS) statistic, the Conditional Information Entropy Ratio (CIER) statistic (see for instance Russel et al., 2012 for information regarding these tests) and the H-measure (Hand, 2009). Our results (not tabulated) demonstrate that the semiparametric model outperforms the alternative BSM specifications. Specifically, the KS statistic is 0.75 for the semiparametric 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 semiparametric 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 semi-parametric model is 0.65 whereas for the competing models the H-measure is 0.51, 0.51, 0.45, 0.52.

6.5.2 Five-Fold Validation

For this test, we divide the full sample (1990-2015) into five approximately equal sub-samples in chronological order. We use any four of them to train the semiparametric model and use the left-out sample to measure its performance. We then compare its performance with the alternative specifications in each of the left-out subsample, using AUROC as a summary statistic. In each subsample, the semiparametric model outperforms the alternative BSM specifications (not tabulated). Its average performance is 0.9102 where for the remaining 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.

6.5.3 The Case of Financial Distress

We further explore the prediction performance of the semiparametric model by augmenting the event sample with financially distressed firms. Generally speaking, it is preferable to develop models to identify the early signs of the crisis (i.e. financial distress) rather than waiting until bankruptcy occurs, in which case the firm might already have lost most of its value and the firm faces additional costs arising from bankruptcy, such as liquidation and legal costs[40]. Another advantage from the prediction of financial distress relative to bankruptcy, is that the firm may have enough time to reverse the situation through corrective measures, such as assets sales, reductions in capital expenditures, debt restructurings etc (see Asquith et al., 1994 and references therein), preventing in that way further deterioration which may eventually lead to bankruptcy. Finally, we believe that predicting financial distress poses an interesting problem because predicting the early stages of the problem is a harder task to accomplish and therefore, will challenge the performance of all models.

Despite the benefits of the prediction of financial distress, only a handful of papers have addressed the issue. We believe that the main reason is the lack of a formal definition of financial distress and as such, it must be defined using subjective criteria based on financial performance. However, most of the studies agree that the key criterion should be a form of inability of the firm to cover its financial obligations, such as the inability to cover its interest payments (see for instance Pindado et al., 2008; Gupta et al., 2018).

In this study, we follow Keasey et al., (2015) to classify the firms as financially distressed (also used by Gupta et al., 2018). Specifically, we consider a firm to be in financial distress when all of the following conditions are satisfied; 1) Earnings Before Interest, Tax and Depreciation and Amortization (EBITDA) is less than financial expenses (i.e. interest payments) for two consecutive years 2) Total Debt is higher than the Net Worth of the firm for two consecutive years and 3) The firm experiences negative Net Worth growth between two consecutive years. The firm is classified as financially distressed in the year immediately following these three events. For prediction purposes, we use the data two years before financial distress. For example, when the conditions are satisfied for the years *t* and *t-1*, then the firm is considered as financially distressed in the year *t* and we construct the variables at *t-2* to predict financial distress. By doing so, we have classified a total of 2022 firms as financially distressed between 1991 and 2015 and the total sample amounts to 72042 firm-year observations.

[40] According to a recent study from Glover (2016), the direct and indirect costs arising from default amounts to 45% of firm value.

Results reported in table 7, clearly demonstrate the superior performance of the semiparametric model relative to the alternative BSM specifications.

[Insert Table 7 here]

The results from Table 7 also reveal that in states of financial distress, the simple BSM models are not useful in predicting the firms undergoing financial distress as indicated by AUROC results, which in all cases are substantially lower relative to when predicting bankruptcies (table 4). In contrast, in cases of financial distress, more advanced methodologies should be used. For instance, the performance of the semiparametric model is quite impressive, given the relatively difficult nature of the problem, although the performance according to AUROC has been dropped (as expected) relative to when predicting bankruptcies (0.8997 vs 0.9387 respectively). Information content and economic benefits results also demonstrate that the semiparametric model performs better than the alternative BSM models. For instance, in a competitive environment, the bank which uses the semiparametric model manages a portfolio of clients where only the 0.43% are financially distressed. In contrast, for the other banks which use various BSM specifications, the financially distressed ratio ranges between 2.34% - 6.25%.

6.5.4 Comparison with Alternative Methodologies-Financial Distress Case

The good performance of the semiparametric model motivates us to compare its performance with alternative methodologies. We compare the performance of the semiparametric model with two very widely-applied approaches; the logistic regression (LR) approach and the nonparametric (NP) approach, such as neural networks. As explanatory variables for both approaches, we use the variables of Campbell et al. (2008), which are also used as inputs when estimating our semiparametric model. Furthermore, in the case of the traditional neural network, we use the same specifications as was done for the semiparametric model; In the hidden layer, we use three neurons ($H=3$) as well as we use the tan-sigmoid activation function. This is done for consistency. In the output layer, we use one neuron ($M=1$) with the log-sigmoid activation function in order to obtain a probability. Finally, the log-likelihood function is used to train the neural network in order to obtain its coefficients. Performance results are reported in table 8.

[Insert Table 8 here]

As expected, the results are now more comparable since all approaches generally perform well in predicting financial distress. However, the semiparametric model is the best performing model, according to all tests. This is evident by the higher out-of-sample AUROC it exhibits relative to the LR and NP approaches (which equal to 0.8528 and 0.8802 respectively) with the differences being statistically significant according to the DeLong test (at significance level α=1%). Next, the semiparametric model is better in terms of information content (differences in log-likelihoods are statistically significant at α=1% according to the Vuong test). Finally, a bank which uses the semiparametric model, manages a better-quality portfolio of clients as indicated by the lowest fraction of financially distressed firms it attracts relative to other banks, which use the LR approach or the NP approach. In particular, for the bank which uses the semiparametric model, 0.80% of the firms it attracts are financially distressed, whereas for banks which use the LR or NP approach, the financial distress ratio amounts to 3.63% and 1.04% respectively. Overall, the results in this section suggest that the semiparametric model is a promising methodology since it outperforms other well-known financial distress prediction methodologies.

6.5.5 Focusing on the financial crisis period 2007-2009

In this section, we compare the performance of the models in the out-of-sample period which includes only the years 2007-2009, where the financial crisis has arrived and might have impacted the financial performance of the firms severely (results are not tabulated). The purpose of this section is to test the performance of the models under unfavorable conditions in the market. Overall, evidence from this test confirms the superior performance of the semiparametric approach relative to the alternative BSM models during the financial crisis period.

7 Summary and Conclusions

In this paper, we introduce and compare an estimation technique to obtain parameter values, such as the asset value and volatility, which are used in parametric models for the estimation of the probability of default. Specifically, we view asset value and volatility as generalized functions and by using a nonparametric technique, such as neural networks, we obtain improved asset values and volatilities which enter the parametric model, yielding a semiparametric model. Using the BSM model as a paradigm, we compare the performance of the semiparametric model with popular BSM alternative specifications with respect to 1)

AUROC, 2) Information content and 3) Economic benefits. Our results demonstrate the superiority of the semiparametric model since in all tests, the semiparametric model outperforms the competing BSM specifications. We further examine the performance of the models when the sample of events is augmented with financially distressed firms. In this respect, we find that the semiparametric model outperforms not only the BSM models but also, other methodologies as well, such as the logistic regression approach and the nonparametric approach, which in fact justifies the implementation of the semiparametric model in future research for default/financial distress prediction.

However, we believe that the semiparametric model may be subject to improvement. Future research may emphasize on the examination of other activation functions to be used in the output layer of the neural network, such that possibly more accurate asset values and volatilities can be obtained. Another promising avenue for future research, would be to increase the number of outputs from the nonparametric model, for instance, asset values, asset volatility and asset expected return, whereas in this study, the asset value with expected return were merged in order to obtain the expected value of assets. Also, it would have been useful to examine several other inputs to the nonparametric model, beyond the variables from Campbell et al. (2008), which could further increase the precision of outputs from the nonparametric model. Finally, the BSM which we improve its estimation, is the earliest parametric model. Given that several extensions have been proposed in the literature (see for instance Leland, 1994; Leland and Toft, 1996), one may possibly use the approach proposed in this study, to improve the estimation of such extended parametric models.

References

Afik, Z., Arad, O., & Galil, K. (2016). Using Merton Model for Default Prediction: An Empirical Assessment of Selected Alternatives. *Journal of Empirical Finance, 35*, 43-67.

Agarwal, V., & Taffler, R. (2008). Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models. *Journal of Banking and Finance, 32*(8), 1541-1551.

Aït-Sahalia, Y., & Duarte, J. (2003). Nonparametric option pricing under shape restricitons. *Journal of Econometrics, 116*, 9-47.

Aït-Sahalia, Y., & Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance, 53*, 499-547.

Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance, 36*, 278-285.

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance, 23*(4), 589-609.

Andreou, P. C., Charalambous, C., & Martzoukos, S. H. (2008). Generalized parameter functions for options pricing. *Journal of Banking and Finance, 34*, 633-646.

Asquith, P., Gertner, R., & Scharfstein, D. (1994). Anatomy of financial distress: An examination of junk-bond issuers. *Quarterly Journal of Economics*, *109*, 625-628.

Bandler, J. W., Ismail, M. A., Rayas-Sanchez, J. E., & Zhang, Q. J. (1999). Neuromodeling of microwave circuits exploiting space-mapping technology. *IEEE Transactions on Microwave Theory and Techniques, 47*, 2417-2427.

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

Bauer, J., & Agarwal, V. (2014). Are Hazard Models Superior to Traditional Bankruptcy Prediction Approaches? A Comprehensive Test. *Journal of Banking and Finance, 40*, 432-442.

Bharath, S. T., & Shumway, T. (2008). Forecasting Default with Merton Distance to Default Model. *The Review of Financial Studies, 21*(3), 1339-1369.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy, 81*(3), 637-654.

Blochlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking and Finance, 30*, 851-873.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In Search of Distress Risk. *The Journal of Finance, 63*(6), 2899-2939.

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

Charitou, A., Dionysiou, D., Lambertides, N., & Trigeorgis, L. (2013). Alternative Bankruptcy Prediction Models Using Option-Pricing Theory. *Journal of Banking and Finance, 37*(7), 2329-2341.

Crosbie, P., & Bohn, J. (2003). Modeling default risk. *Moody's KMV*

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems, 2*, 303-314.

Delong, R., Delong, M., & Clarke-Pearson, L. (1988). Comparing the Areas Under Two or More Correalted Receiver Operating Characteristic Curves: A Non-Parametric Approach. *Biometrics*, *44*(3), 837-845.

Filipe, S. F., Grammatikos, T., & Michala, D. (2016). Forecasting distress in European SME portfolios. *Journal of Banking and Finance*, *64*, 112-135.

Glover, B. (2016). The expected costs of default. *Journal of Financial Economics*, *119,* 284-299.

Gupta, J., Gregoriou, A., & Ebrahimi, T. (2018). Empirical comparison of hazard models in predicting SMEs failure. *Quantitative Finance*, *18*, 437-466.

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, *77*, 103-123.

Hanley , J. A., & McNeil, B. J. (1982). The Meaning and Use of the Area Under a Receiver Operating Characteristics (ROC) Curve. *Radiology, 143*(1), 29-36.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the Probability of Bankruptcy. *The Review of Financial Studies, 9*(1), 5-34.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*, 359-366.

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

Jones, E. P., Mason, S. P., & Rosenfeld, E. (1984). Contingent claims analysis of corporate capital strucrures: An empirical investigation. *Journal of Finance, 39*, 611-625.

Keasey, K., Pindado, J., & Rodrigues, L. (2015). The determinants of the costs of financial distress in SMEs. *International Small Business Joural, 33,* 862-881.

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

Leland, H. (1994). Corporate Debt Value, Bond Covenants, and Optimal Capital Structure. *Journal of Finance, 49*, 1213-1252.

Leland, H., & Toft, K. B. (1996). Optimal Capital Structure, Endogenous Bankruptcy and the Term Structure of Credit Spreads. *Journal of Finance, 51*, 987-1019.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance, 29*(2), 449-470.

Pindado, J., Rodrigues, L., & De la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research, 61*, 995-1003.

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

Reisz, A. S., & Perlich, C. (2007). A Market-Based Framework for Bankruptcy Prediction. *Journal of Financial Stability, 3*(2), 85-131.

Ronn, E. I., & Verma, A. K. (1986). Pricing risk-adjusted deposit insurance: An option-based model. *Journal of Finance, 41*, 871-895.

Russell, H., Tang, Q. K., & Dwyer, D. W. (2012). The effect of imperfect data on default prediction validation tests. *Journal of Risk Model Validation, 6*, 1-20.

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

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

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *Journal of Finance, 59*, 831-868.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica, 57*(2), 307-333.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*(4), 817-838.

Tables

Table 1: Distribution of observations

Year	Defaulted Firms	Healthy Firms	Default 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

This table shows the distribution of default and healthy-firm observations across the sample period 1990-2015 and the annual default rate, defined as the number of defaults divided by the annual number of observations

Table 2: Descriptive statistics

	Defaulted Firms			Healthy Firms			t-test
Variables	Mean	Median	St.Dev	Mean	Median	St.Dev.	
NIMTA	-0.249	-0.1722	0.254	-0.021	0.023	0.148	-31.40
CASHMTA	0.068	0.032	0.098	0.119	0.062	0.160	-6.51
TLMTA	0.697	0.784	0.264	0.378	0.332	0.258	25.35
BM	1.036	0.481	2.551	1.461	0.537	4.807	-1.81
RSIZE	-12.774	-12.765	1.490	-10.919	-10.983	2.075	-18.31
LOGPRICE	0.523	0.560	1.145	2.291	2.473	1.274	-28.37
EXRET	-0.213	-0.340	0.864	0.207	0.106	0.670	-12.79
SIGMA	1.070	0.959	0.486	0.657	0.551	0.418	20.14

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

Table 3: Mean asset values and volatilities from the different estimation approaches

Estimation Apporaches	Mean-Defaulted Firms		Mean-Healthy Firms		t-test	
	V/F	σ_V	V/F	σ_V	V/F	σ_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
<u>Direct Estimation</u>						
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
SP Approach	3.676	0.72	6.398	0.572	-24.12	25.73

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 Eqs. (5) and (6). The 1-Eq. Approach refers to estimating the time-series of asset values over the previous year by solving Eq. (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 SP approach refers to estimating expected asset value and the volatility based on the semiparametric approach (see sections 3.1 and 3.2). The last column reports t-tests for mean differences between defaulted and healthy firms.

Table 4: AUROC results

	AUROC	Delong 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
SP Approach	**0.9387**	-

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 Eqs. (5) and (6). The 1-Eq. Approach refers to estimating the time-series of asset values over the previous year by solving Eq. (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 SP Approach refers to estimating expected asset value and volatility based on our semiparametric approach (see sections 3.1 and 3.2). The last column reports the Delong (1988) test statistic, to test for statistically significant differences in the AUROCs between the semiparametric model with the alternative BSM specifications.

Table 5: Information content results

	Model 1	Model 2	Model 3	Model 4	Model 5
Prob 1	0.039 (20.93)				
Prob 2		0.045 (19.63)			
Prob 3			0.037 (17.99)		
Prob 4				0.051 (20.80)	
Prob 5					0.337 (26.45)
Rate	-1.562 (-3.60)	-2.329 (-4.61)	-1.360 (-2.75)	-2.458 (-4.34)	-2.073 (-4.20)
Constant	-5.810 (-25.66)	-5.741 (-24.10)	-5.870 (-24.89)	-5.459 (-22.92)	-5.692 (-26.32)
Log-Likelihood	-690.70	-665.81	-705.42	-671.69	**-607.70**
Pseudo-R^2 (%)	18.18	21.21	16.43	20.43	**28.01**
Vuong Test	4.78	4.50	5.41	4.46	-

This table reports information content results. We estimate five logit models, where the out-of-sample default probabilities (from the period 2007-2015) produced by the various BSM specifications are included in the logit estimation. Models 1 and 2 include default probabilities produced by the 2-Eqs. and 1-Eq. Approaches (denoted with Prob 1 and Prob 2 respectively). Models 3 and 4 include default probabilities produced by the direct estimation approach (Prob 3 and Prob 4 are default probabilities produced by BS, 2008 and CDLT, 2013 respectively). Finally, Model 5 includes default probabilities produced by our semiparametric approach. The last row of the table reports the Vuong (1989) test statistic, to test for statistically significant differences in the log-likelihoods between Model 5 with the Models 1-4. In all logit models, we include *Rate*, defined as the annual default rate of the previous year, as proxy for the baseline hazard rate.

Table 6: Economic impact results

	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
Defaults	43	5	24	4	10
Defaults/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**

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 with 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 semiparametric approach. Banks sort prospective customers (2007-2015) and reject the 5% of firms with the highest risk. The remaining firms are classified in 10 groups of equal size and for each group, a credit spread is calculated as described in the main text (section 5.3). The bank that classifies the firm to the group with the lowest spread is finally granting the loan. Market share is the number of loans given divided by the number of firm-years, Revenues = (market size)*(market share)*(average spread), Loss=(market size)*(prior probability of bankruptcy)*(share of bankruptcies)*(loss given default). Profit=Revenues-Loss. Return on Assets is profits divided by market size*market share and Return on Risk-Weighted-Assets is profits divided by Risk-Weighted Assets, obtained from formulas provided by the Basel Accord (2006). The prior probability of bankruptcy is the bankruptcy rate for firms between 1990-2006 and equals 0.43%. Loss given default is 45%.

Table 7: Performance comparisons between the semiparametric approach and BSM specifications-Financial distress case

	AUROC	LL-Info. Content	Portfolio Quality	DeLong Test	Vuong Test
2-Eqs. Approach	0.6884	-2235.00	5.53%	17.41	14.21
1-Eq. Approach	0.7235	-2202.61	2.34%	15.59	13.15
Direct Estimation					
1) BS (2008)	0.6846	-2257.86	6.25%	17.74	14.38
2) CDLT (2013)	0.7217	-2216.28	2.78%	15.63	13.46
SP Approach	**0.8997**	**-1720.45**	**0.43%**	-	-

This table reports performance results of the various BSM specifications in the out-of-sample period spanning the years 2007-2015, when the sample of events is augmented with financially distressed firms. The 2-Eqs. Approach refers to estimating asset values and volatilities by simultaneously solving Eqs. (5) and (6). The 1-Eq. Approach refers to estimating the time-series of asset values over the previous year by solving Eq. (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 SP Approach refers to estimating expected asset value and volatility, based on our semiparametric approach (see sections 3.1 and 3.2). 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 concentration of financially distressed firms, when banks compete to grant loans in a competitive economy (equivalent to the fourth row of table 6). The last two columns report DeLong (1988) and Vuong (1989) test statistics, to test for statistically significant differences in the AUROCs and log-likelihoods, between the semiparametric approach and the various BSM specifications.

Table 8: Performance comparisons between the semiparametric approach and alternative approaches-Financial distress case

	AUROC	LL-Info. Content	Portfolio Quality	DeLong Test	Vuong Test
LR Approach	0.8528	-2105.94	3.63%	7.19	12.43
NP Approach	0.8802	-1841.27	1.04%	3.96	5.02
SP Approach	**0.8997**	**-1720.45**	**0.80%**	-	-

This table reports performance results of the alternative approaches for financial distress prediction, such as the logistic regression (LR) approach, the nonparametric (NP) approach and specifically neural networks and finally, the semiparametric (SP) 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 concentration of financially distressed firms, when banks compete to grant loans in a competitive economy (equivalent to the fourth row of table 6). The last two columns report DeLong (1988) and Vuong (1989) test statistics, to test for statistically significant differences in the AUROCs and log-likelihoods, between the semiparametric approach and the alternative methodologies.

Figures

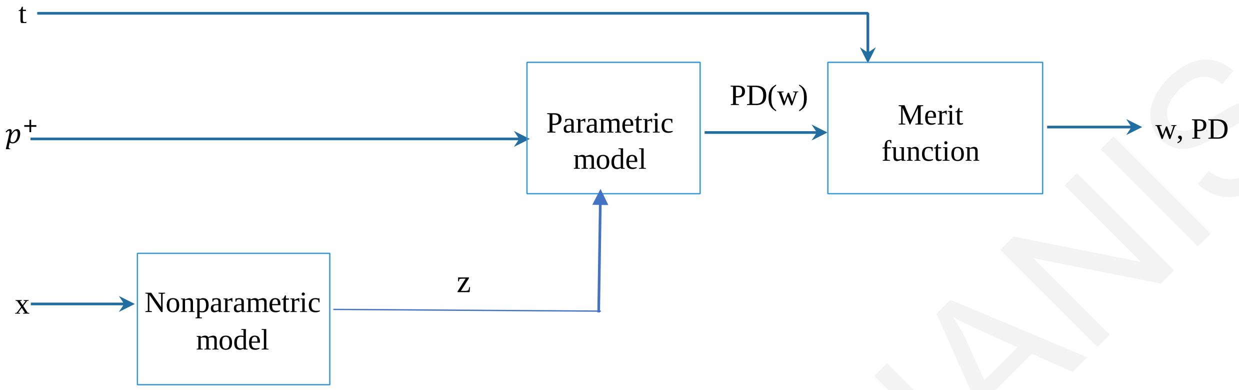

Figure 1: Schematic representation of our approach. Improved parameter values, z, are obtained from the nonparametric model and enter as inputs to the parametric model along with other parameters, p^+, that enter directly, yielding a semiparametric model. Here, x, represents some exogenous inputs to the nonparametric model. The proposed structure is optimized according to a merit function, to give the weights, w, and finally the probability of default, PD. Note that in the merit function, the targets t are supplemented directly. In our case, t=1 if the firm defaults and t=0 otherwise and the merit function is the log-likelihood function.

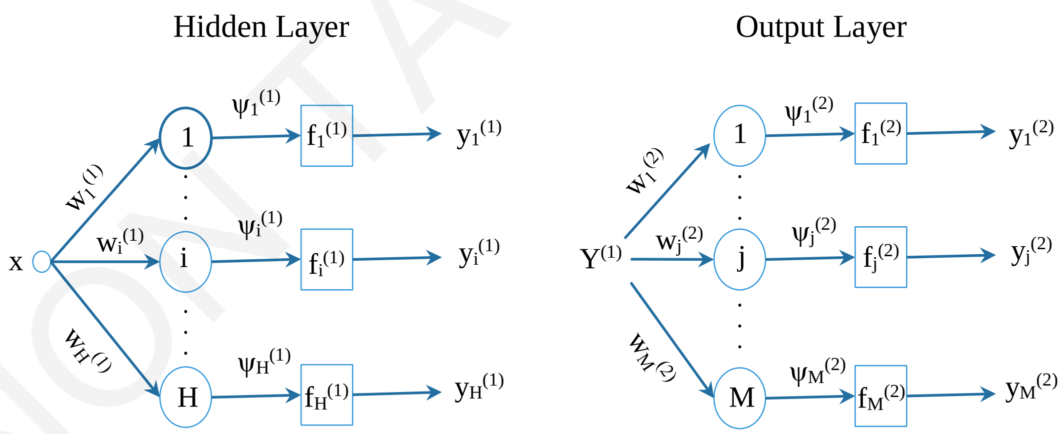

Figure 2: General structure of a two-layer feedforward neural network, with H neurons in the hidden layer and M neurons in the output layer.

CONCLUSIONS

Bankruptcy prediction of firms has been in the forefront of academic research over the past decades and the effort to identify the troubled firms early, will continue in the future. The lessons learned from the recent global financial crisis, where we have evidenced the bankruptcy of many firms which led banks to suffer huge losses from their loan portfolios, remind us the importance of developing bankruptcy prediction models. Besides the economic costs arising from bankruptcy, several others include social ones, such as the loss of investors' confidence towards the markets, lawsuits to the management of the firm but also, people losing their jobs due to the closure of the company.

This dissertation aimed to provide innovations to the most common bankruptcy prediction approaches; The structural approach and the empirical approach. Firstly, in the first chapter, it is found that the structural model from the framework of Leland-Toft (1996) is a better approach relative to the most widely-used structural model; The Black-Scholes-Merton model. Therefore, for those interested to forecast bankruptcy using the structural approach, the Leland-Toft model should be preferred. The chapter also found that including the probability of bankruptcy derived from Leland-Toft as additional predictor in models like Altman (1968), Ohlson (1980) and Campbell et al. (2008) yields models with improved out-of-sample performance and these models were the best performing in all tests. With that respect, evidence suggests that Leland-Toft probability is a missing predictor in empirical models and it is recommended to be considered in association with the original empirical models.

The second chapter of the dissertation focused on the empirical approach and proposed methodologies to maximize their ability to discriminate bankrupt from healthy firms as measured by AUROC. It is found that the proposed methodologies provide bankruptcy models with improved predictive ability relative to traditional approaches for bankruptcy prediction and the improvement in predictive ability is also evident economically when banks use such models. Therefore, for those interested using the empirical approach to estimate bankruptcy risk of firms, it is recommended to train their models using AUROC as the optimization criterion. More specifically, a merit function which takes care of the outliers should be used when the response function is linear and a neural network model when the response function is probabilistic.

The third chapter which is dedicated on the structural approach, proposed a nonparametric methodology to estimate unobserved parameters of the structural (parametric) models; the value of assets and volatility of asset value returns. When these parameters are viewed as generalized functions of some exogenous inputs, *x*, the nonparametric approach can be used to uncover these functions through learning. With that respect, the Black-Scholes-Merton model was used as paradigm and it is found that our approach provides improved parameter values which when enter the structural model, yields a semiparametric model with substantially improved performance relative to the alternative parameter estimation approaches widely-used in the literature. This chapter also considered the case of financially distress prediction, which is a state prior to bankruptcy, and it is found that while the traditional approaches did not perform well, the semiparametric approach exhibited impressive out-of-sample performance. In all, our semiparametric model is the best performing in all tests considered in this chapter and we conclude that when using the structural approach, the nonparametric methodology should be implemented in order to obtain the unobserved parameter values.