Assessing Bankruptcy Probability with Alternative Structural Models and an Enhanced Empirical Model

Zenon Taoushianis[1]* — University of Cyprus

Chris Charalambous[2] — University of Cyprus

Spiros H. Martzoukos[3] — University of Cyprus

April 2016

[1] PhD Candidate. Contact address: Department of Accounting and Finance, School of Economics and Management, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus: taoushianis.zenon@ucy.ac.cy (+357 22 893640)

[2] Professor (Emeritus) of Management Science. Contact address: Department of Business and Public Administration, School of Economics and Management, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus: bachris@ucy.ac.cy (+357 22 892466)

[3] Associate Professor of Finance. Contact address: Department of Accounting and Finance, School of Economics and Management, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus: baspiros@ucy.ac.cy (+357 22 893615)

*Corresponding Author

Assessing Bankruptcy Probability with Alternative Structural Models and an Enhanced Empirical Model

April 2016

Abstract

The purpose of this paper is to examine the ability of two structural credit risk models to forecast firms' bankruptcy; Leland (1994) and Leland and Toft (1996). These models have received much less attention by researchers in the literature of corporate credit risk modeling relative to others and their empirical assessment in this study show that they can be a powerful alternative option for those concerned to forecast bankruptcy. Furthermore, when we extend the empirical accounting-based measure of bankruptcy, Z-score (Altman, 1968), by incorporating bankruptcy probabilities produced by our structural models as additional explanatory variables, its performance improves significantly. These two models which we call market-based Z-scores yield the most powerful models at in-sample and out-of-sample forecasts amongst several alternative specifications.

JEL classification codes: C52, G13, G33

Keywords: Bankruptcy Probability, Structural Models, Empirical Models, Leland, Leland-Toft, Z-score

1. Introduction

1.1. Background and Motivation

The corporate environment provides ample information to assess the risk of a firm to fall in bankruptcy, especially when the firm is publicly traded. On one hand, accounting information obtained from financial statements provides information about the past performance of the firm that resulted from its activities over previous periods. On the other hand, market information observed in equity markets provides an assessment of the prospects of the firm as perceived by market participants on aggregate. Therefore over the years researchers have developed several approaches and models to forecast bankruptcy that take into consideration these types of information. Two popular and widely used models belonging to each category and aim to assess firms' bankruptcy risk are Z-score and Black-Scholes-Merton (BSM) models. The first is an empirical model developed by Altman (1968) and relates bankruptcy with a set of accounting ratios using statistical analysis (Multivariate Discriminant Analysis in its original form). The second is a structural model that estimates bankruptcy risk based on options pricing theory developed by Black and Scholes (1973) and extended by Merton (1974) for the valuation of corporate debt.

Yet, there is a class of structural models that has been largely left unexplored in the literature and their ability to forecast bankruptcy accurately is questionable. These are structural models that extend the framework of BSM model from its restrictive assumptions to incorporate for instance other types of debt, such as a coupon-paying debt, allow interest rates being stochastic, allow bankruptcy to occur prior to the maturity of debt etc. We call these models alternative structural models. These structural models have mainly been examined on their ability to predict bond prices or spreads (see for instance Ogden (1987), Lyden and Saraniti (2001) and Eom et.al (2004) among others) and to predict default rates (see for instance Leland (2004), Suo and Wang (2006) and Tarashev (2008) among others). However, for sound risk management purposes and in line with the Basel Accord (2005), the validation of credit risk models consists of more thorough procedures and tests, which we address in this paper. Thus, there is a need to shed light on the performance of such alternative structural models and examine whether they can potentially serve as an alternative option for the assessment of bankruptcy risk.

In this study, we examine the ability of two alternative structural models to forecast firms' bankruptcy one year ahead; Leland (1994) and Leland and Toft (1996). Firstly, Leland (1994) extends Merton (1974) model to incorporate the effects of taxes and bankruptcy costs to the valuation of a corporate coupon-paying debt with infinite maturity when bankruptcy is determined endogenously or exogenously. Leland and Toft (1996) relax the assumption of infinite maturity and consider the case when the debt has a finite maturity but it is continuously rolled-over. While there are several other structural models to consider (see for instance Imerman, 2013), our analysis focuses only on these two because we don't aim to make a comprehensive comparison between the models. In contrast, we want to emphasize on whether these alternative structural models worth to consider and if they can be a powerful option for those concerned to forecast bankruptcy. In that perspective, Leland and Leland-Toft models is a good starting point because they are direct extensions of BSM model, while the formula to estimate the probability of bankruptcy is similar to that of BSM, adjusted such that the probability of bankruptcy at any time prior to maturity be positive.

1.2. Objectives and Findings

We perform several tests in order to examine the forecasting ability of Leland and Leland-Toft models on a sample of 5460 publicly traded U.S firms with total of 39830 firm-year observations over the period 1995-2014 from all non-financial industries. We compare the performance of the structural models with the accounting-based measure of bankruptcy, Z-score. We refer to Z-score as the model that uses the same financial information as the original Z-score does, but updating its coefficients by applying the logistic regression approach in our sample. We choose Z-score for several reasons. Firstly, it provides an alternative way to forecast bankruptcy since it is an empirical model constructed using financial ratios and the performance of our structural models relative to an empirical model is questionable. Secondly, Z-score is usually used as a benchmark when comparing the performance of bankruptcy prediction models. Finally we want to examine the effect that market-based bankruptcy probabilities obtained from our structural models (L_{prob} and LT_{prob}) have on the performance of an empirical model when they are incorporated as additional predictors.

Using financial statements and market data for each firm to construct all the parameters needed for Leland and Leland-Toft models, we employ several tests. We firstly measure the discriminatory power of Leland, Leland-Toft and Z-score and we test if the difference between the two structural models with Z-score is significant at in-sample (1995-2005) and out-of-sample (2006-2014) forecasts. This test will highlight the ability of each model to distinguish bankrupt from healthy firms one year prior to bankruptcy. Secondly, we perform tests on the predictive accuracy of Leland, Leland-Toft and Z-score or equivalently on their ability to empirically fit the data. Next, we go a step further and we examine the effect of bankruptcy probabilities produced by Leland, Leland-Toft and Z-score (which we call L_{prob}, LT_{prob} and Z_{prob} respectively) when incorporated in logit models as predictors. These tests will demonstrate the explanatory power of our bankruptcy probability measures and if they are significant predictors of bankruptcy. Last but not least, we construct two market-based versions of Z-score that incorporate L_{prob} and LT_{prob} as predictors, along with the financial ratios of Z-score in order to examine the degree to which the forecasting ability of Z-score can be improved. Thus, we examine the performance of Leland and Leland-Toft when we keep their functional form and additionally when we use their outputs as explanatory variables in logit models. Finally, we provide additional evidence about the performance of Z-score, Leland, Leland-Toft and of the two market-based Z-scores when we consider two alternative validation approaches. The first is a direct extension of the initial classification which is based on re-estimating each model by moving forward one year (the rolling window approach). For the second we divide the whole sample in five sub-samples with equal observations (the five-fold validation approach). Each time four of them are used as the in-sample period and the left one is the out-of-sample period in a way that all sub-samples to be included in the in-sample period and out-of-sample period in turn. From these tests we concentrate and discuss results on aggregate level but we provide detailed/per-period results in tables A1, A2 and A3 in the Appendix.

Results obtained from our analysis are indicative as to the performance of the models under investigation. Beginning from the baseline results (i.e. when the sample is divided into in-sample (1995-2005) and out-of-sample periods (2006-2014)), we find that at in-sample and out-of-sample tests, Leland-Toft model significantly outperforms Z-score in terms of their ability to distinguish bankrupt from healthy firms as it is shown from their AUROCs using DeLong et.al (1988) tests (DeLong test hereafter). Similarly, Leland model outperforms Z-score at in-sample forecasts with the difference though not being statistically significant. In contrast, Z-score slightly outperforms Leland model in out-of-sample forecasts with the difference being negligible and not statistically significant. Thus, from this perspective we report a superiority of Leland-Toft model to identify bankrupt from healthy firms one year prior to bankruptcy as opposed to Z-score. In contrast, the ability of Leland and Z-score seems to be equivalent.

Results from predictive accuracy tests show that Z-score exhibits better predictive accuracy and ability to fit the data than Leland and Leland-Toft, as indicated by their Log-Likelihoods at in-sample and out-of-sample periods. Further investigation shows that this result is due to the fact that the two structural models severely overestimate bankruptcy risk for 900 firms and this induces Log-Likelihoods to decrease substantially. Thus, from that view, structural models lack of empirical fit but as we will show later on, incorporating bankruptcy probabilities from structural models as predictors in logit models solves this problem.

Next, including Z_{prob}, L_{prob} and LT_{prob} in logit models, results show that they are significant predictors of bankruptcy (at significance level α=1%). AUROCs of these models are similar with Z-score, Leland and Leland-Toft (at in-sample and out-of-sample periods), suggesting that it is irrelevant whether we measure discriminatory power by keeping the functional form of the models or whether we use their output as predictors in logit models. However, logit models that include Z_{prob} with L_{prob} and Z_{prob} with LT_{prob}

perform better than models that include Z_{prob}, L_{prob} and LT_{prob} separately. This finding suggests that Z_{prob}, L_{prob} and LT_{prob} are insufficient to forecast bankruptcy when considered alone. In other words market-based information (reflected in L_{prob} and LT_{prob}) provide complementary information about bankruptcy as Z_{prob} does and vice-versa.

Last but not least, we find that the performance of Z-score significantly improves upon the inclusion of L_{prob} and LT_{prob} as additional predictors in Z-score and in fact these models which we call market-based Z-scores yield the most powerful models from all models constructed in the paper.

Finally, results from our additional tests support previous findings. For example, we find that Leland-Toft has higher AUROC than Z-score with their difference being statistically significant with significance α=1% as opposed to Leland which seems to have equal discriminating ability with Z-score. In addition, our two market-based bankruptcy measures, improve significantly the discriminating ability of Z-score. From the perspective of model fitness, we find that Leland and Leland-Toft lack of empirical fit as opposed to Z-score but for our market-based Z-scores, empirical fit is significantly better as indicated by their Log-Likelihoods. Finally, according to Pseudo-R^2 results, the two market-based Z-scores explain bankruptcy better than Z-score and thus confirming the idea that models that include both financial as well as market-based information are better in forecasting bankruptcy.

Overall our results suggest that these alternative structural models are powerful in terms of forecasting bankruptcy either when their functional form is kept or when they are included as predictors in logit models and thus we motivate researchers to consider these models in their future studies. The remainder of the paper proceeds as follows: In section 2 we discuss several papers that are close to ours, section 3 describes the models in more detail and presents the formulas for assessing bankruptcy probability. In section 4 we discuss the procedure to collect the data and construct the variables of interest, section 5 explains the methodological design of the paper, section 6 presents the results and section 7 concludes. Lastly, we provide detailed-analytical results for the five-fold validation approach.

2. Related Literature

Despite the fact that the literature has generated numerous models that aim to forecast firms' bankruptcy risk, the focus of this study is on structural models and specifically on a class of structural models that builds upon the work of Merton (1974). In this section we collect prior work on structural models and discuss those that are relevant to our paper.

It was not until the beginning of 2000's when researchers started to study BSM model more thoroughly after practitioners from Moody's and KMV had provided the key insights to implement the model from a practical point of view through a series of papers. From these papers Crosbie and Bohn (2003) is one of the most comprehensive papers that is dedicated on the description of the methodology to construct the model for the assessment of bankruptcy risk (or equivalently default risk). Beyond that, academic literature provides adequate empirical evidence about the performance of the model. Some of them are Hillegeist et.al (2004), Du and Suo (2007), Reisz and Perlich (2007), Agarwal and Taffler (2008), Bharath and Shumway (2008), Campbell et.al (2008), Wu et.al (2010), Afik et.al (2012) and Charitou et.al (2013), with Hillegeist et.al (2004), Agarwal and Taffler (2008) and Wu et.al (2010) being the closest studies to ours with respect to their objectives. For example Hillegeist et.al (2004) find that bankruptcy probabilities produced by BSM, Z-score and O-score [i.e. Ohlson (1980)] are significant predictors of bankruptcy when included in hazard rate models and that BSM model provides more information about bankruptcy relative to the accounting-based measures of bankruptcy Z-score and O-score. In contrast, Agarwal and Taffler (2008) find that Z-score provides more information relative to BSM model[4] and explains bankruptcy better when included in hazard rate models. In terms of discriminating ability, Reisz and Perlich (2007) and Agarwal and Taffler (2008) find that Z-score does better than BSM in forecasting bankruptcy one year ahead. Wu et.al (2010), in a comprehensive study comparing the performance of BSM[5] with several empirical models[6] find that the performance of BSM is adequate but not the best among the models as evident by their ability to explain bankruptcy and to classify firms as bankrupt and healthy. Thus, evidence on the performance of structural relative to empirical models is mixed.

In contrast to the previous studies, the empirical validation of structural models other than BSM and specifically their ability to forecast firms' bankruptcy is not common. In the literature several studies assess the performance of some models that belong to this alternative class in various contexts. Eom et.al (2004) examine five structural models on their ability to predict corporate bond spreads.[7] In general they find that spreads produced by these models deviate significantly from true spreads. Leland (2004) examines two structural models on their ability to predict average default probabilities of corporate bonds belonging to certain credit ratings[8]. Assuming common parameters for all firms i.e. common asset return (μ=12%), asset volatility (σ=23%) etc, he finds that the models predict the general shape and level of default probabilities but underestimate them at

[4] The authors compare the performance of Z-score with two versions of BSM model; with that of Hillegeist et.al (2004) and Bharath and Shumway (2008).
[5] The authors build the BSM model as done in Hillegeist et.al (2004)
[6] These are models developed in Altman (1968), Ohlson (1980), Zmijewski (1984) and Shumway (2001)
[7] These models are Merton (1974), Geske (1977), Longstaff and Schwartz (1995), Leland and Toft (1996) and Collin-Dufresne and Goldstein (2001).
[8] These models are Longstaff and Schwartz (1995) and Leland and Toft (1996).

shorter horizons. Suo and Wang (2006) compare the ability of four structural models to predict default rates one and four years ahead for firms on certain credit ratings[9]. They find that Longstaff and Schwartz (1995) and Leland and Toft (1996) provide reasonable default rates compared to historical averages provided by Moody's and Standard and Poors. Patel and Pereira (2007) compare six structural models on their ability to produce expected default probabilities for failed and non-failed UK real estate firms[10]. Using a cut-off point of 20% above from which the firms are classified as failed and non-failed otherwise, they find that Merton and Leland-Toft models have the worst performance as they have the highest type I error whereas Ericsson-Reneby and Collin-Dufresne and Goldstein models have the best performance as they misclassify 8% of the firms. In univariate logistic regressions, they show that these measures are statistically significant, meaning that they are significant predictors of default. Tarashev (2008) examine five structural models on their ability to accurately forecast actual default rates of firms belonging to certain credit ratings.[11] Results show that default rates produced by the models accurately reflect actual default rates one and five years ahead. Finally, Wong et.al (2010) examine the discriminatory power and calibration quality of three structural models[12]. In terms of the first test, they find that all models exhibit adequate discriminatory power and their differences are not material.

3. Models and Bankruptcy Probability

This section analyzes in more detail the theoretical underpinnings and features of the three models under examination with special emphasis on the two structural models discussed in the next sub-section and shows how bankruptcy probability is estimated.

3.1. Leland and Leland-Toft Models

Leland (1994) extends the work of Merton (1974) to incorporate the effects of taxes and bankruptcy costs in the valuation of corporate risky debt with infinite maturity. The advantage of his framework is that it enables the valuation of debt (or of a bond) that pays coupons as opposed to the framework of Merton where the firm issues only one zero-coupon bond. In this context, Leland derives closed-form solutions for the market value of equity, debt and total firm value. More importantly, he also considers the case where bankruptcy is determined endogenously as opposed to Merton (1974) where

[9] These are Merton (1974) with and, Longstaff and Schwartz (1995), Leland and Toft (1996) and Collin-Dufresne and Goldstein (2001). All models (except Leland and Toft (1996) include stochastic and non-stochastic interest rates.
[10] These models are Merton (1974), Black and Cox (1976), Longstaff and Schwartz (1995), Leland and Toft (1996), Ericsson and Reneby (1998), and Collin-Dufresne and Goldstein (2001).
[11] These are Longstaff and Schwartz (1995), Anderson et.al (1996), Leland and Toft (1996), Collin-Dufresne and Goldstein (2001) and Huang and Huang (2012).
[12] These are Longstaff and Schwartz (1995), Leland and Toft (1996) and Collin-Dufresne and Goldstein (2001).

bankruptcy is determined exogenously. This consideration enables the calculation of an optimal bankruptcy point which is chosen by the management in favor of shareholders such that the equity value is maximized. When assets value hits that point, it is optimally, from shareholders´ perspective, for the firm to bankrupt. In contrast, when bankruptcy is determined exogenously, the bankruptcy barrier is chosen arbitrarily[13]. The assumption of exogenous bankruptcy barrier is unrealistic because usually firms still operate even when the assets value falls below from firm's liabilities in which case the firm is likely to enter in a re-organization process. Thus, the framework created by Leland provides a more realistic approach for valuing corporate debt than the Merton framework. Equation (1) shows the calculation of the bankruptcy point underlying the Leland model which is a key determinant of the bankruptcy probability:

$$VB_L = \frac{(1-\tau)C}{r + 0.5\sigma^2} \tag{1}$$

where C is the coupon payment, τ the corporate tax rate, r the risk-free rate and σ^2 the variance of asset returns.

Leland and Toft (1996) extend the framework of Leland (1994) to the case where corporate debt has a finite maturity but it is rolled-over on a continuous basis when it matures with the same terms (i.e. same maturity and same coupon payments). In this context, they again derive closed-form solutions for the market value of equity, debt and total firm value as well as for the endogenously-determined bankruptcy point which now depends on debt maturity, T. Equation (2) shows the calculation of the bankruptcy barrier underlying the Leland-Toft model:

$$VB_{LT} = \frac{\left(\frac{C}{r}\right)\left(\frac{A}{rT} - B\right) - A\frac{P}{rT} - \tau\frac{Cx}{r}}{1 + \alpha x - (1-\alpha)B} \tag{2}$$

where

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

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

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

[13] For example, in the Merton model the bankruptcy barrier is the liabilities of the firm and thus, it is determined exogenously.

with N(·) and n(·) denoting the cumulative standard normal distribution function and standard normal density function respectively. Note that "α" in equation 2 is the parameter for bankruptcy costs and it is different from "α" included in "A" and "B". Furthermore, a closer examination of equation (2) shows that it is a function of six parameters which are observable and this in fact allows for direct estimation: the risk free rate (*r*), the coupon payments (*C*), the bankruptcy costs (α), the volatility of assets (σ), the debt principal (*P*) and the payout yield (*d*). Also when $T \rightarrow \infty$, the Leland-Toft bankruptcy barrier converges to that of Leland and as a consequence bankruptcy probability too.

Unlike in the case of Merton when bankruptcy occurs only at debt maturity, *T*, this is not the case in Leland and Leland-Toft framework which bankruptcy can occur at any time. That is, in order to assess the probability of bankruptcy, in this context we need to define a cumulative distribution function which allows the evaluation of bankruptcy risk in discrete points of time, t where t ≤ T. The probability that the current value of firm's assets will fall to the bankruptcy barrier for the first time at time t conditionally that V > VB is given by equation (3):

$$prob(t) = N(X) + e^{Y}N(Z) \tag{3}$$

where

$$X = \frac{-ln\left(\frac{V}{VB}\right) - (\mu - d - 0.5\sigma^2)t}{\sigma\sqrt{t}}, Y = \frac{-2ln\left(\frac{V}{VB}\right)(\mu - d - 0.5\sigma^2)}{\sigma^2}$$

$$Z = \frac{-ln\left(\frac{V}{VB}\right) + (\mu - d - 0.5\sigma^2)t}{\sigma\sqrt{t}}$$

The term "VB" is the bankruptcy point as defined by Leland and Leland-Toft in equations (1) and (2) respectively and μ is the return of asset value, V. When VB=P, at t=T the first term of equation (3) is the same as the bankruptcy probability produced by BSM model[14]. However, since (3) has an additional non-negative element in the RHS, it turns out that bankruptcy probability produced by (3) will always be higher than that produced by BSM model.

3.2. Z-score Model

According to Altman's (1968)[15] analysis on 33 bankrupt firms match-paired with 33 healthy firms from the manufacturing industry of U.S, five financial ratios were found to

---

[14] $BSMprob = N(-DD)$ with $DD = \frac{ln\left(\frac{V}{VB}\right) + (\mu - d - 0.5\sigma^2)T}{\sigma\sqrt{T}}$

[15] We choose Altman's empirical model as a comparison with Leland and Leland-Toft models because prior work on bankruptcy prediction was based on the analysis of single financial ratios such that of Beaver (1966). Altman's model was the first multivariate model and despite the fact that it was developed more

be significant predictors of bankruptcy. These were *Earnings before Interests and Taxes /Total Assets (EBITTA), Retained Earnings/Total Assets (RETA), Working Capital/Total Assets (WCTA), Sales/Total Assets (SLTA)* and *Equity Value/Total Liabilities (EVF).* The original model was established based on Multivariate Discriminant Analysis (MDA) and produced scores about the financial healthiness of the firm; the higher the Z-score, the healthier the firm is. In this study we re-estimate the Z-score by applying a logistic regression model on firms collected between 1995-2005 accounting for approximately 65% of the total sample and we left firms collected in the years 2006-2014 for out-of-sample forecasting. Logistic regression allows estimation of coefficients through the maximum likelihood approach and bankruptcy probability is estimated using the logistic distribution function shown in equation (4):

$$Z_{prob} = \frac{e^y}{1 + e^y} \tag{4}$$

where

$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_5 x_5$ with $b_1 \ldots b_5$ and $x_1 \ldots x_5$ being the coefficient estimates and accounting ratios respectively.

4. Data

This section discusses the sample selection, the procedure that is followed in the study to construct the variables as well as descriptive statistics of the variables included in the Z-score, Leland and Leland-Toft models.

4.1. Sample Selection

We analyze a sample of 5460[16] U.S public firms from which 333 filed for bankruptcy in a specific year between the recent 20-year period of 1995-2014; 5127 firms constitute the healthy sample (firms that did not file for bankruptcy in any of the years under consideration). Bankruptcy filings were identified from BankruptcyData.com[17] and include firms that filed for bankruptcy under Chapter 11 and Chapter 7. To avoid problems related to sample selection bias and to 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 filed for bankruptcy, future observations for that firm were excluded (if any). Though, past

---

than 45 years ago, it is still used as benchmark when researchers compare their own bankruptcy prediction models (see for instance Falkenstein et.al (2000), Fernandez (2005), Altman and Sabato (2007) among others and Altman et.al (2014) for a discussion of studies that have employed Z-score after 2000).

[16] One of our requirements is that each firm must have a non-zero interest expense because this variable is our proxy for coupon payments (*C*). Thus, we lost many firms who had no interest expenses.

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

observations for all bankrupt firms were included in our sample i.e. before a firm files for bankruptcy, it is considered as healthy (with the definition of healthiness defined above).

Table 1 presents the distribution of observations in the sample.

[Insert Table 1 here]

In general, the bankruptcy rate in all years is less than 1% except from years 1999 (1.493%) and the mid-crisis years 2008 and 2009 with the bankruptcy rate being 1.190% and 2.133% respectively. The average bankruptcy rate in the sample is 0.836% indicating the fact that bankruptcy is a rare event.

Firms from all sectors were collected except from Finance, Insurance and Real Estates sectors[18] due to the different nature of their operations and structure of their financial statements relative to 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 2 shows the industry distribution of our sample.

[Insert Table 2 here]

The majority 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).

4.2. Variables Construction

For this study accounting-based and market-based information are collected from WRDS COMPUSTAT and CRSP respectively in order to construct the Z-score, Leland and Leland-Toft models. Since the interest is the forecasting of bankruptcy, year-end information from financial statements such as Earnings before Interests and Taxes (EBIT), Total Assets (TA), Total Liabilities (F), Sales (SL), Working Capital (WC), Retained Earnings (RE), Interest Expense (IE) and Dividends (Ordinary and Preferred, denoted as D) are collected from WRDS COMPUSTAT at the year before a firm files for bankruptcy[19]. In this manner our variables do not coincide with the year of bankruptcy filing, in which case accounting but especially market variables would have been affected significantly. This also allows for one-year bankruptcy prediction since we have financial and market information at the year prior to bankruptcy filing.

[18] These are firms with SIC codes between 6000-6799

[19] For example if a firm files for bankruptcy in 15/03/2006, we collect its financial statements that concern the financial performance of the firm over the entire fiscal year ending in 2005.

Once information from WRDS COMPUSTAT was collected, we obtain monthly equity prices from CRSP. The collection of observations starts from the fiscal year-end month prior to bankruptcy and we go 13 months backwards. For each month we calculate the asset value (V_t) as the sum of equity value plus the face value of total liabilities ($V_t = E_t + F$), with E_t being the equity value which is defined as the end-of-month stock price x shares outstanding. Since F is the total liabilities taken from annual financial statements, it remains constant when calculating the monthly value of assets. Then we calculate the asset value log-returns for each month [$i.e. \ln\left(\frac{V_t}{V_{t-1}}\right)$]. This procedure generates a time series of 12 observations of asset log-returns and we calculate the annualized standard deviation as the monthly standard deviation x $\sqrt{12}$ and annualized return as the (average) monthly return x 12 which are our proxies for asset volatility and asset return respectively.

Other information needed for the construction of Leland and Leland-Toft models are the risk-free rate, the coupon payments, the debt principal, the payout yield, the tax rate, the bankruptcy costs and the maturity of liabilities. For the risk-free rate (r), the one-year Treasury Constant Maturity rate is used for all years under examination, obtained from Federal Reserve[20]. For the coupon payments (C) and debt principal (P), the interest expense and total liabilities are used as proxies respectively and the payout yield (d) is defined as the sum of coupon payments plus dividends (ordinary and preferred) divided to market value of assets. For corporate tax rate (τ), bankruptcy costs (α) and maturity of liabilities (T) we follow Leland (2004) which sets these parameters equal to 15%, 30% and 10 years respectively.

Table 3 depicts the construction of variables and the parameters needed for the estimation of Z-score, Leland and Leland-Toft models.

[Insert Table 3 here]

Finally, to avoid problems induced by outliers, we follow the literature and we winsorize all accounting-based and market-based variables by setting all values lower than the 1st and higher than the 99th percentiles equal to the values corresponding to 1st and 99th percentiles.

4.3. Descriptive Statistics

Descriptive statistics in table 4 present the main features of bankrupt and healthy firms in a univariate context that includes both differences in financial variables (after winsorization) as well as in bankruptcy probabilities produced by our models.

[Insert Table 4 here]

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

Table 4 reveals several characteristics about the financial condition of bankrupt and healthy firms one year before bankruptcy. Regarding the results about the financial ratios, bankrupt firms are less liquid than healthy firms as can be inferred from WCTA, with the difference being statistically different from zero (at significance level α=1%). In addition, they are less profitable (they actually have losses on average) relative to healthy firms as can be inferred from EBITTA, with the difference being statistically significant at significance level α=1%. Furthermore, the two measures of leverage (EVF and P/V) on average it is higher for bankrupt firms as opposed to healthy firms, with the difference being statistically significant at significance level α=1%. A slight surprising result is the fact that bankrupt firms are more active than healthy firms as can be inferred from SLTA which is higher for bankrupt firms but the difference is not statistically significant (neither at mean nor in median). Finally, bankrupt firms pay relatively more coupons as opposed to healthy firms as can be inferred from C/V (with the difference being statistically significant at significance level α=1%) which also drives the payout ratio, d, for bankrupt firms to be significantly higher than healthy firms. Overall it is evident that the financial performance of bankrupt firms is worse as compared to that of healthy firms one year prior to bankruptcy.

The two variables that play an important role in determining bankruptcy risk in Leland and Leland-Toft models are σ_V and μ. Since these variables are constructed using equity information, they capture firms' performance in the market. From the table it can be inferred that the value of assets of bankrupt firms is more volatile than that of healthy firms (with difference being statistically significant at significance level α=1% only for mean), whereas the return of assets for bankrupt firms is lower (and negative) relative to healthy firms (with difference being statistically significant at significance level α=1%) who earn positive asset value returns on average. Therefore, market performance of bankrupt firms is worse as compared to that of healthy firms one year before bankruptcy.

Finally, the table provides some preliminary results about bankruptcy probability produced by our models. Firstly, it seems that all models show some ability to distinguish bankrupt from healthy firms, as the average bankruptcy probability produced by the models for bankrupt firms is higher relative to healthy firms, with differences being statistically significant at significance level α=1%. Secondly, from the three bankruptcy probability measures, Leland-Toft model produces the highest bankruptcy probabilities for both bankrupt and healthy firms relative to the other two models. On one hand this is an indication that the model is able to assign high bankruptcy probabilities to bankrupt firms and hence the model seems to forecast bankruptcy risk successfully. On the other hand, it seems that the model overestimates bankruptcy risk for healthy firms. Finally, results show that both Leland and Leland-Toft models overestimate bankruptcy risk. For example unreported calculations show that the average bankruptcy probability produced by these models is 4.68% and 5.02% respectively in contrast to Z-score[21] which produces an average bankruptcy probability equal to 0.840% which is similar to the bankruptcy rate in our sample.

5. Methodology

This section describes the methodology that is used in this study to assess the performance of our bankruptcy risk models. We firstly describe the methodology to measure and test the discriminating ability of each model and finally we explain how to estimate our logit models and test them in terms of their ability to fit the data using Log-Likelihood-based tests.

5.1. Discriminatory Power

Discriminatory power refers to the ability of a particular model to discriminate the bankrupt firms from healthy firms. The Receiver Operating Characteristics (ROC) curve is a graphical representation of the discriminatory power of a bankruptcy risk model. It plots the true predictions on the vertical axis (the percentage of the bankrupt firms which are correctly classified as bankrupt) against the false predictions on the horizontal axis (the percentage of healthy firms which are incorrectly classified as bankrupt) according to a pre-determined cut-off value. If we perform this classification procedure for multiple cut-off values, we create as many 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. That is, the perfect model will pass through the point (0, 1) and in 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 risk model is the Area Under ROC (AUROC) curve (see for example Soberhart and Keenan, 2001). Following Hanley and McNeil (1982), AUROC measures the probability that when two firms are selected randomly one from the bankrupt population and the other from the healthy population, their scores will be correctly ranked (i.e. the bankruptcy probability of the bankrupt firm will be higher than that of the healthy firm). The AUROC is calculated as:

$$A\widehat{URO}C = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\psi(p_B^i, p_H^j) \tag{5}$$

where

[21] Here the Z-score is fitted in the whole sample.

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

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.

To assess whether our two theoretically-driven models (i.e. Leland and Leland-Toft) outperform the Z-score model in terms of discriminatory power, we test if the difference between their AUROCs are significantly different. Thus, we test the following two hypotheses:

$$H_0: AUROC_L - AUROC_{Z\text{-}score} = 0 \quad V_s \quad H_1: AUROC_L - AUROC_{Z\text{-}score} \neq 0$$

$$H_0: AUROC_{LT} - AUROC_{Z\text{-}score} = 0 \quad V_s \quad H_1: AUROC_{LT} - AUROC_{Z\text{-}score} \neq 0$$

We use the non-parametric approach of DeLong et.al (1988) which accounts for the correlation of the AUROCs produced by any two models. The key element for the estimation of the test statistic is the covariance matrix of the AUROCs produced by our models. Following DeLong et.al (1988), the covariance matrix is estimated as follows:

1) For each bankrupt firm calculate the AUROC:

$$A\widehat{URO}C\left(p_B^i\right) = \frac{1}{m}\sum_{j=1}^{m} \psi(p_B^i, p_H^j), \quad (i = 1,2, \dots, n) \tag{6}$$

2) For each healthy firm calculate the AUROC:

$$A\widehat{URO}C\left(p_H^j\right) = \frac{1}{n}\sum_{i=1}^{n} \psi(p_B^i, p_H^j), \quad (j = 1,2, \dots, m) \tag{7}$$

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

$$s_{10}^{k,r} = \frac{1}{n-1}\sum_{i=1}^{n} [A\widehat{URO}C_k\left(p_B^i\right) - A\widehat{URO}C_k][A\widehat{URO}C_r\left(p_B^i\right) - A\widehat{URO}C_r] \tag{8}$$

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

$$s_{01}^{k,r} = \frac{1}{m-1}\sum_{j=1}^{m} [A\widehat{URO}C_k\left(p_H^j\right) - A\widehat{URO}C_k][A\widehat{URO}C_r\left(p_H^j\right) - A\widehat{URO}C_r] \tag{9}$$

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

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

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

$$z = \frac{A\widehat{URO}C_1 - A\widehat{URO}C_2}{\left(s^{1,1} - 2s^{1,2} + s^{2,2}\right)^{1/2}} \tag{11}$$

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 (10).

5.3. Logit Models

Logit models are constructed to test several hypotheses such as whether L_{prob} and LT_{prob} are significant predictors of bankruptcy, whether L_{prob} and LT_{prob} improve Z-score etc. We follow Hillegeist et.al (2004) in the construction and estimation of the following logit model:

$$p_{i,t} = \frac{e^{a + X_{i,t}\beta}}{1 + e^{a + X_{i,t}\beta}} \tag{12}$$

where $p_{i,t}$ is the probability of bankruptcy of firm "*i*" at time "*t*", $X_{i,t}$ 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 logit model (12) represents a multi-period logit model because it includes multiple observations (when available) for each firm across time. However, the inclusion of multiple-year observations per firm can result to understated standard errors because the Log-Likelihood objective function which is used for estimation of the multi-period logit model assumes that each observation is independent to each other. This is a wrong assumption since financial information of a particular firm at time *t* cannot be independent from the financial information of the same firm at time *t-1.* To fix this econometric issue we estimate robust standard errors using the Huber-White covariance matrix [Huber (1967), White (1980)].

To compare the fitness of the models (i.e. which model has better predictive accuracy than the other) we use the Vuong (1989) test which is appropriate for non-nested models (i.e. none of the models can be expressed as a reduced-form version of the other)[22] and is

[22] For example, this test is employed to compare the fitness between the two market-based Z-scores and the univariate models that include Z_{prob}, L_{prob}, LT_{prob}.

based on the comparison of Log-Likelihoods between the two models. Thus, the hypothesis is the following:

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

where

L_1 and L_2 are the Log-Likelihoods of the two models under comparison and k_1 and k_2 are the number of parameters of each model.

The z-statistic in this case is standard-normally distributed and it is defined as follows:

$$z = \frac{2(L_1 - L_2) - (k_1 - k_2)\ln(N)}{2\sqrt{N}\omega_N} \tag{13}$$

where

N 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:

$$l_i = ln\left[\frac{{p_{1,i}}^{y_i}(1-p_{1,i})^{(1-y_i)}}{{p_{2,i}}^{y_i}(1-p_{2,i})^{(1-y_i)}}\right] \tag{14}$$

where $p_{1,i}$ and $p_{2,i}$ are the bankruptcy probabilities for the *i-th* firm produced by models 1 and 2 respectively and y_i indicates whether the firm is bankrupt ($y_i = 1$) or healthy ($y_i = 0$). Rejection of the null hypothesis means that predictive accuracy of the two models is not the same.

On the other hand, to compare predictive accuracy between nested-models we use standard Likelihood Ratio (LR) tests; the one model is the full model with k_1 parameters and the other model is expressed as a reduced-form version of the full model (i.e. the variables in the reduced model are contained in the full model)[23] with k_2 parameters where k_1>k_2. The Log-Likelihoods of the two models are tested indirectly by testing whether the extra $k_1 - k_2$ parameters in the full model provide any information about bankruptcy risk and improve the fit as opposed to the reduced model. Thus, the hypothesis is the following:

$$H_0\text{: } \beta_{k_2+1} = \beta_{k_2+2} = \cdots = \beta_{k_1} = 0 \quad V_s \quad \text{At least one beta} \neq 0$$

The statistic in that case is the following:

[23] For example, this test is employed to compare the fitness between each market-based Z-score and Z-score and the univariate models that include $Z_{prob}, L_{prob}, LT_{prob}$ and the bivariate models that include Z_{prob} with L_{prob} and Z_{prob} with LT_{prob}.

$$LR = -2[L(k_2) - L(k_1)] \tag{15}$$

Finally critical values for (15) are obtained from chi-square distribution with $k_1 - k_2$ degrees of freedom. Rejection of the null hypothesis means that at least one of the extra $k_1 - k_2$ variables is important and therefore predictive accuracy of the full model is better than that of the reduced model.

6. Results

This section reports and discusses the results of the paper. We start from the estimation of Z-score and then we examine the performance of the models. We firstly report results on the discriminatory power and secondly results about the fitness (i.e. predictive accuracy) of the models. Next we examine the impact of bankruptcy probabilities produced by the models when entered as predictors (of bankruptcy) in logit models. Finally we estimate two logit models that include the accounting ratios of Z-score along with L_{prob} and LT_{prob} as additional predictors and we examine the extent to which Z-score is improved. We call these two models market-based Z-scores. Beyond these results which are based on the separation of the sample into one in-sample period (1995-2005) and one out-of-sample period (2006-2014), we examine the forecasting ability of all models based on two additional approaches; the rolling window approach according to which the models are continuously estimated by moving forward one year and the five-fold validation approach that generates five in-sample periods and five out-of-sample periods. The two approaches allow us to test the models and measure performance multiple times and thus a safer conclusion about their forecasting ability can be drawn. We discuss the out-of-sample results on aggregate level while we record the per-period performance and report the results in the Appendix for the latter approach.

6.1. Estimation of Z-score

Table 5 presents the estimation of Z-score.

[Insert Table 5 here]

From the logistic regression results presented in table 5, *WCTA, EBITTA* and *EVF* have the correct sign (corresponding coefficient values in column 2 are all negative) suggesting that an increase of the value of each ratio induce a reduction of firm's bankruptcy risk. However, *RETA* and *SLTA* have a positive coefficient which is counter-intuitive. Furthermore, *WCTA, EBITTA* and *SLTA* are statistically significant at significance levels α=1% for the first two and α=10% for the last one. This is in consistent with Hillegeist et.al (2004) who find that not all ratios of Z-score are statistically significant. Finally, unreported tests showed that multicollinearity does not affect our results based both on bivariate correlations and the VIF criterion[24]. Nevertheless, we keep the form of Z-score as it is.

6.2. Discriminatory Power Tests

To examine the degree to which each model is able to discriminate the bankrupt firms from healthy firms, we measure the AUROC of each model in two periods. The first period is the in-sample period 1995-2005 which accounts for about 65% of total firm-year observations of our sample. Then, we estimate AUROC in our out-of-sample period 2006-2014. Obviously, for our two structural models this does not matter whereas for the Z-score this does matter since coefficients are optimized by using data from the in-sample period and the model is applied in the out-of-sample period. Finally, we test the two hypotheses described in section 5.1.

Table 6 presents discriminatory power results of the three models in the two periods of interest.

[Insert Table 6 here]

From the results reported in table 6, it is evident that all models show a significant ability to distinguish bankrupt from healthy firms at both in-sample and out-of-sample periods. A random model which cannot distinguish bankrupt from healthy firms has an AUROC equal to 50%. In contrast, Z-score, Leland and Leland-Toft according to table 6 have an AUROC equal to 81.77%, 83.29% and 86.97% respectively at the in-sample period and 85.69%, 84.18% and 90.31% respectively at the out-of-sample period. These results show that all models are not random models but instead they exhibit a significant ability to discriminate bankrupt from healthy firms confirming our earlier discussion in section 4.3 about the discriminating ability of each model.

Going to the comparison of the models, Leland-Toft model seems to have the highest ability to distinguish bankrupt from healthy firms as it has the highest AUROC amongst all models at both in-sample (86.97%) and out-of sample periods (90.31%). Furthermore, in-sample and out-of-sample results suggest that Leland-Toft significantly outperforms Z-score. This is evident by the fact that DeLong tests reject the hypothesis of equal AUROCs at significance level $\alpha=1\%$ (z-statistic equal to 3.135 at the in-sample period and 2.444 at the out-of-sample period). Next, comparing Leland model versus Z-score at the in-sample period, it seems that Leland model slightly outperforms Z-score in terms of discriminatory power. The difference though of their AUROCs is not statistically significant (z-statistic equal to 0.841). Finally at the out-of-sample period, Z-score slightly outperforms Leland model, with the difference in AUROCs not being statistically significant (z-statistic=-0.621).

[24] For example at the worst case, VIF=2.2361

Thus, results show that Z-score and Leland models perform equally well with respect to their ability to distinguish bankrupt from healthy firms.

Results in table 6 exhibit some interesting features. We expected that at in-sample forecasts, Z-score would have outperformed Leland and Leland-Toft models but instead, the opposite result has occurred with both structural models to outperform Z-score as they both have higher AUROCs. Furthermore, the out-of-sample discriminatory power of all models is higher than that of in-sample discriminatory power. However, this could be due to the fact that the out-of-sample period contains the years which coincides with the 2007 financial crisis in which case bankrupt firms would have easily been more detectible than healthy firms. Another explanation which affects the results in the two periods is that the out-of-sample period contains a different number of firms and different number of bankrupt and healthy firms.

Finally, plot 1 of figure 1 depicts the graphical representation of the discriminating ability of the three models.

[Insert Plot 1 of Figure 1 here]

As it can be seen from the plot, the ROCs of Z-score and Leland are close and therefore indicating similar discriminating ability. In contrast, the ROC curve of Leland-Toft demonstrates higher discriminating ability since it is above the curve of Z-score (and Leland).

Overall, findings in this section show that all models have a significant discriminating ability. Furthermore, Leland-Toft is the most powerful model with AUROC being significantly different than that of Z-score. Though, this is not the case for Leland model which performs equally well with Z-score based on AUROCs at both periods.

6.3. Predictive Accuracy Tests

While discriminatory power provides information about the ability of each model to distinguish bankrupt form healthy firms, fitness tests provide information about the predictive accuracy of the models (i.e. their ability to generate accurate bankruptcy probabilities). In this study, the Log-Likelihood statistic is our indicator of the ability of each model to empirically fit the data. Table 7 reports the results.

[Insert Table 7 here]

Based on the Log-Likelihood results reported in table 7, it is evident that the empirical model, Z-score exhibits much better predictive accuracy than the two structural models at both in-sample and out-of-sample periods. Specifically at the in-sample period, the Log-Likelihood of Z-score is about five and seven times higher than the Log-Likelihoods of Leland-Toft and Leland models respectively. Similar results are also obtained for the out-of-sample period where the Log-Likelihood of Z-score is about four and seven times higher than the Log-Likelihoods of Leland-Toft and Leland models respectively. These results are driven by the following two reasons. Firstly, Z-score is designed such that the Log-Likelihood is maximized and therefore it reasonably has the highest Log-Likelihood from all models. Secondly our structural models severely overestimate bankruptcy risk for some healthy firms and this effect causes the Log-Likelihood function to decrease substantially. For example, for the 900 observations with the highest bankruptcy risk within the healthy group at the in-sample period, Leland-Toft model assigns bankruptcy probabilities higher than 50%, with the average being 80%. More severely, the Log-Likelihood of these 900 observations is -3390 (whereas the Log-Likelihood of the in-sample period is -4885.68). Thus, the overestimation of bankruptcy risk for these 900 observations is the primary cause of the large differences in Log-Likelihoods between the Z-score and the two structural models. Finally, from the two structural models, Leland-Toft model provides better predictive accuracy than that of Leland since it has higher Log-Likelihood at in-sample and out-of-sample periods.

To sum up, results in this section show that the predictive accuracy and fitness of Z-score is much better than that of the two structural models. This result motivates us to use the outputs of Leland and Leland-Toft models (L_{prob} and LT_{prob}) as explanatory variables in logit models since on one hand, market-based bankruptcy measures produced by our structural models have high discriminating ability (i.e. Leland-Toft) and on the other hand empirical models improve the accuracy and fit.

6.4. Bankruptcy Probability as Predictor

This section examines whether the three bankruptcy probability measures Z_{prob}, L_{prob} and LT_{prob} are significant predictors of bankruptcy when included in logit models as explanatory variables. The idea is to examine the degree to which single variables that reflect accounting-based information (i.e. Z_{prob}) and market-based information (L_{prob} and LT_{prob}) are significant predictors of bankruptcy. The logit models are estimated using the in-sample period 1995-2005 and their performance is evaluated using both the in-sample and out-of-sample periods. Table 8 reports the results of these tests.

[Insert Table 8 here]

Panel A reports results on the estimation of five logit models with bankruptcy probability (Z_{prob}, L_{prob} and LT_{prob}) as predictor. Models 1-3 are univariate logit models and models 4 and 5 include accounting-based information captured by Z_{prob} in combination with market-based information captured by L_{prob} and LT_{prob} respectively. According to the results in table 8, panel A, our bankruptcy probability measures are statistically significant (with significance level α=1%) in models 1-3 indicating that they are significant predictors of bankruptcy. In addition, our predictors remain statistically significant in models 4 and 5 suggesting that Z_{prob}, L_{prob} and LT_{prob} are not sufficient predictors of bankruptcy when included individually in the logit models. Thus, market-based information provided by L_{prob} and LT_{prob} complement information provided by accounting-based information reflected by Z_{prob} and vice-versa. This result is consistent with Hillegeist et.al (2004), Agarwal and Taffler (2008) and Bharath and Shumway (2008), who find that accounting-based or market-based information (captured by BSM_{prob}) are not sufficient predictors when included individually in hazard rate models.

Panel B in table 8 provides results about the performance of the five logit models at the in-sample period. Results from the univariate logit models 1-3 suggest that they perform equally well with Z-score, Leland and Leland-Toft respectively with respect to their ability to discriminate bankrupt from healthy firms. This is shown by the fact that the AUROCs of these models are similar (i.e. AUROCs in table 6 are similar with AUROCs in table 8, panel B). This is due to the fact that the logit function provides a monotonic transformation of the structural probabilities and therefore ranking of the firms is unaffected. Thus, from this perspective it is irrelevant if we measure bankruptcy probability by keeping the functional form of Leland and Leland-Toft or if we use these bankruptcy probabilities as predictors in logit models, since their discriminating ability is equivalent. However, Log-Likelihood results of logit models 2 and 3 suggest that predictive accuracy and fit are improved when bankruptcy probabilities produced by our structural models are included in logit models as predictors. This is evident by the fact that the Log-Likelihoods of models 2 and 3 are substantially higher than those of Leland and Leland-Toft in table 7. Thus, from this perspective, it is preferably to include these bankruptcy probability measures as predictors in logit models in order to improve predictive accuracy and fit. Next, a comparison between models 1-3 shows that models that include market-based bankruptcy measures as predictors (models 2 and 3) perform better than models that include bankruptcy measures that summarize financial performance (model 1). For example, the AUROCs of models 2 and 3 are higher than that of model 1 but the difference is only statistically significant in the case of model 3 that includes LT_{prob}. Unreported DeLong tests show that this difference is significantly different from zero (with significance level α=1%). Furthermore, predictive accuracy and fitness of models 2 and 3 is significantly better than model 1, as their Log-Likelihoods are significantly different (with significance level α=1%, based on Vuong tests). Finally, pseudo-R^2 of models 2 and 3 is higher than model 1, meaning that they are able to explain bankruptcy better than model 1.

Going to the extended models 4 and 5 which include both accounting-based and market-based variables, evidence shows that they perform better than when they are included individually (models 1-3). For example, model 4 has higher discriminating ability than models 1 and 2 (unreported DeLong tests show that their AUROCs are significantly different at significance levels α=1% and α=5% respectively). Furthermore it has better predictive accuracy and fit (unreported LR tests show that its Log-Likelihood is significantly different from models 1 and 2, at significance levels α=1%). Also, it better explains bankruptcy as it is shown from its higher pseudo-R^2. Similar results are obtained for model 5 versus its counterparts (i.e. model 5 versus model 1 and 3), except from the fact that the AUROCs of model 5 and model 3 are not significantly different. These results confirm earlier discussion about the insufficiency of accounting-based and market-based information to forecast bankruptcy when included alone as predictors. Thus, past performance measured by financial ratios and summarized in a single variable of bankruptcy, Z_{prob}, provide additional information that are not captured by market-based information and summarized in L_{prob} and LT_{prob} and vice-versa. Therefore, a complete forecasting model should include both accounting-based and market-based variables.

Panel C in table 8 reports results about the out-of-sample performance of the five logit models. We obtain similar insights as of that of the in-sample performance and hence our results are not affected when the models are applied to a different dataset other than the dataset used to estimate them. Specifically, out-of-sample AUROCs of models 1-3 are equivalent to out-of-sample AUROCs of Z-score, Leland and Leland-Toft, reported in table 6. Log-Likelihoods though, are better for models 1-3. An out-of-sample comparison between models 1-3 confirms the superiority of models with market-based bankruptcy measures as predictors from models that include bankruptcy measures and reflect information from financial statements. This can be inferred by the fact that the AUROC of model 3 is higher and significantly different from model 1 (at significance level α=1%), though the AUROC of model 2 is lower than model 1, with the difference not being significantly different from zero according to DeLong test (p-value =39.67%). Furthermore, models 2 and 3 have better predictive accuracy and fit and exhibit higher ability to explain bankruptcy variation than model 1 since Log-Likelihood differences are significantly different from zero (at significance level α=1%) according to Vuong tests (unreported), and pseudo-R^2 of models 2 and 3 are higher than model 1. Finally, extended models 4 and 5 improve discriminatory power as compared with their counterpart univariate models. Though, differences in AUROCs between models 3 and 5 are not statistically significant. In contrast, predictive accuracy and the ability to explain bankruptcy variation are improved as differences in Log-Likelihoods are significantly different from zero (according to unreported LR tests) and pseudo-R^2 is higher for models 4 and 5.

Overall results in this section suggest that bankruptcy probability measures produced by Z-score, Leland and Leland-Toft are significant predictors of bankruptcy, with the last two measures being better forecasting variables relative to the first. This is evident by the fact that models that include these variables as predictors perform better relative to models that include Z_{prob}. However none of these measures is sufficient to forecast bankruptcy alone and thus both accounting-based and market-based information are required in the forecasting model. Most importantly, all results hold even when the models are applied to the out-of-sample period 2006-2014. Thus, our models of interest which incorporate information from Leland and Leland-Toft perform better than models which incorporate information from Z-score.

6.5. Market-Based Z-score

In this section we estimate two market-based versions of Z-score that include L_{prob} and LT_{prob} respectively as predictors in addition to the five financial ratios of Z-score. The idea is to examine the degree to which Z-score can be improved when we include market-based information extracted from our structural models (captured by L_{prob} and LT_{prob}). Table 9 reports the results of the two market-based Z-score models.

[Insert Table 9 here]

Panel A in table 9 shows the estimation of the market-based Z-score in two versions; model 6 is the Z-score with L_{prob} as additional predictor and similarly model 7 include LT_{prob} as predictor. In both models, these additional predictors are highly statistically significant (at significance levels α=1%) confirming the necessity of market-based information to forecast bankruptcy especially when the model contains only financial information. From the financial ratios, only profitability (EBITTA) and liquidity (WCTA) are statistically significant (at significance level α=1%) in the two specifications and have the correct sign. The other ratios are either insignificant (except SLTA in model 6 at significance level α=10%) and/or have the opposite signs. Unreported correlations show that multicollinearity is not a problem.

Panel B in table 9 presents the performance of our market-based Z-scores at the in-sample period. Results show that all market-based Z-scores perform significantly better than the Z-score in all aspects. Beginning from discriminatory power, models 6 and 7 have higher AUROCs than Z-score (85.92% and 86.73% respectively versus 81.77%), with the differences being significantly different from zero (at significance level α=1%) using DeLong tests, suggesting that information provided by our bankruptcy probability measures improve significantly the discriminating ability of Z-score. Furthermore, predictive accuracy and ability to explain bankruptcy variation is significantly improved, as it is evident from the higher Log-Likelihoods of models 6 and 7 than Z-score (unreported LR tests show that the differences in Log-Likelihoods are significantly different from zero at significance level α=1%) and pseudo-R^2 (for Z-score this is 13.96% and reported in table 7).

Finally, Panel C in table 9 reports results about the performance of the market-based Z-scores at the out-of-sample period. Again we find similar insights as of that obtained from the in-sample period. Discriminatory power is high for models 6 and 7 and substantially different from Z-score (91.66% and 92.64% respectively versus 85.69% at significance level α=1%) according to DeLong tests. Additionally, according to LR tests, out-of-sample Log-Likelihoods of models 6 and 7 are higher than those of Z-score (differences are significantly different from zero at significance level α=1%) confirming that predictive accuracy is improved when L_{prob} and LT_{prob} are included as additional predictors in Z-score. Finally, pseudo-R^2 of models 6 and 7 is high (substantially higher than Z-score which equals to 19.02% and reported in table 7).

Finally, figure plot 2 of figure 1 depicts the ROC curves of the five models Z-score, model 6 and model 7).

[Insert Plot 2 of Figure 1 here]

As it can be see, the ROC curves for model 6 and model 7 are well above of the ROC curve of Z-score, indicating higher discriminating ability with that of model 7 being slightly better than that of model 6.

Overall results in this section suggest that a market-based form of Z-score that includes firms' bankruptcy risk measured by structural models and summarize market performance in a single variable (i.e. L_{prob} or LT_{prob}) improve Z-score significantly in terms of discriminatory power and predictive accuracy. In fact, the performance of these two market-based Z-scores yields the most powerful models according to our baseline results.

6.6. The rolling window approach

Consider again our initial in-sample period 1995-2005. According to the rolling window approach, the models are estimated in that time period and applied on 2006 firms. Then, we exclude firms from the oldest year and include the newest firms (thus the second in-sample period is the period 1996-2006 and the out-of-sample period is 2007). This procedure is repeated until the models are applied on the last out-of-sample year which is 2014. We then combine out-of-sample bankruptcy probabilities and report aggregate performance[25]. Therefore, with this approach the models are estimated in a dynamic setting and takes into consideration timely information as opposed to our baseline classification where the model is applied once on 2006-2014 firms. Thus, we expect that the rolling window approach will give better results regarding the models performance since the models are continuously updated. Table 10 reports the performance of Z-score, model 6 and model 7.

[Insert Table 10 here]

[25] Note that for the two structural models, Leland and Leland-Toft, this approach is equivalent by using firms' information from 2006-2014 and estimate bankruptcy probabilities. Thus results will be the same as of those reported in tables 6 and 7 and ROC curves are similar to the plot 1 of figure 1

Results from table 10 confirm earlier results about the significant superiority of model 6 and model 7 as compared to Z-score. Tests based on DeLong (unreported) show that the AUROCs between model 6 and Z-score, and model 7 and Z-score differ significantly (with significance level α=1%). Furthermore, LR tests show that model 6 and model 7 fit the data better relative to Z-score. Finally Pseudo-R^2 results suggest that models 6 and 7 explain bankruptcy better than Z-score. Furthermore, consistent with our expectations, the performance of these models based on the rolling window approach is slightly improved as compared to the case where the models are estimated and applied once on 1995-2005 and 2006-2014 respectively. This can be inferred from the corresponding out-of-sample indicators reported on tables 6, 7 (for Z-score) and 9 (for models 6 and 7).

Finally, figure 2 depicts the ROC curves of the three empirical models (Z-score, model 6 and model 7).

[Insert Figure 2 here]

As it can be seen from figure 2, the ROCs of model 6 and model 7 are well above the ROC curve of Z-score, with model 7 being slightly steeper and therefore better than model 6.

6.7. Five-Fold Validation Approach

This section provides further evidence about the overall performance of Z-score, Leland, Leland-Toft and of the two market-based Z-score models (model 6 and model 7 that include L_{prob} and LT_{prob} respectively as additional predictors in Z-score). The analytical-per period results can be found in tables A1, A2 and A3 of the Appendix. The procedure to obtain these additional results is as follows: We split the whole sample in five consecutive sub-samples with 7966 observations each[26]. Then, we take four out of the five sub-samples as the in-sample period and the one left out is the out-of-sample period in a way that all sub-samples to be included in the in-sample and out-of-sample period in turn. This procedure generates five in-sample periods with 31864 observations each, with five corresponding out-of-sample periods with 7966 observations each. The empirical models (Z-score, model 6 and model 7) are estimated for each in-sample period (for instance table A2, Panel A shows the estimation of model 6 and model 7 for each period) and applied on the corresponding out-of-sample period to obtain bankruptcy probabilities for out-of-sample firms. Then we combine the out-of-sample bankruptcy probabilities in order to obtain aggregate-single indicators about the performance of our models. Note that the combined out-of-sample periods give the whole sample. Therefore, for the structural models bankruptcy probabilities are estimated using the necessary information from all firms. These results are reported in table 11. Meanwhile we record the performance of all

[26] For example, starting from 1995 we take the first 7966 observations which we call it period 1. Then we take the next 7966 observations which we call it period 2 and so on.

models on both in-sample and out-of-sample periods which are reported in the appendix, but in this section we emphasize on the aggregate results.

Table 11 reports the overall results of Z-score, Leland, Leland-Toft, model 6 and model 7.

[Insert Table 11]

Beginning from the structural models, when their functional form is kept, then Leland-Toft has higher AUROC than Z-score, with the difference being statistically significant according to DeLong test. This finding confirms previous results about the superiority of Leland-Toft relative to Z-score on their ability to discriminate bankrupt from healthy firms. In contrast, on that dimension, Z-score and Leland model perform equally well according to DeLong test (z-statistic=-0.223). In addition, it is evident that the two structural models lack of empirical fit since their Log-Likelihoods are substantially lower relative to Z-score.

Next, going to the two market-based Z-scores (model 6 and model 7), results show that they perform significantly better than Z-score in all aspects. For example AUROCs of model 6 and model 7 are higher than Z-score, with the differences being statistically significant according to DeLong tests (z-statistics equal to 6.71 and 7.09 respectively). In addition, their Log-Likelihoods are substantially higher (LR tests show that differences in Log-Likelihoods are significantly different from zero with significance level α=1%) and finally they explain bankruptcy better as indicated by their higher pseudo-R^2 relative to Z-score.

Finally, figure 3 depicts the ROC curves of the five models.

[Insert Figure 3 here]

As it can be seen from plot 1 of figure 3, the ROCs of Z-sore and Leland are not different whereas the ROC curve of Leland-Toft is well above the other two. Finally, from plot 2, the ROCs of model 6 and model 7 are well above than that of Z-score with the ROC curve of model 7 being slightly steeper which means slightly better discriminating ability.

We overlook the discussion for the detailed results but what worth to mention is the fact that our primary variables of interest, L_{prob} and LT_{prob} remain highly statistically significant (at significance level α=1%) in all periods suggesting that they are robust predictors of bankruptcy (table A2). From the financial ratios, *WCTA* and *EBITTA* are statistically significant (at significance level α=1%) in all periods suggesting that liquidity and profitability are significant indicators of bankruptcy. Finally, another important aspect of the models is that coefficient estimates do not change significantly in magnitude from period to period (as well as their standard errors) suggesting that they remain stable over time.

7. Summary and Conclusion

We examine the performance of two structural models for the assessment of corporate bankruptcy risk; Leland (1994) and an extended version of this developed by Leland and Toft (1996) and compare them with an empirical model; Z-score developed by Altman (1968). The models are examined with respect to their ability to distinguish bankrupt firms from healthy firms, to empirically fit the data and to their ability to forecast bankruptcy when included in logit models as predictors.

Overall results show that our alternative structural models have high discriminating ability, with Leland-Toft model being better than Z-score whereas Leland performs equally well with Z-score. Though, our structural models lack of empirical fit. When we incorporate bankruptcy probabilities produced by our models, L_{prob} and LT_{prob}, in logit models, we obtain more insights about their performance as explanatory variables. Specifically, we show that they have significant explanatory power, better than Z_{prob}, but when considered alone are insufficient predictors of bankruptcy since it is possible to construct models with better performance that incorporate Z_{prob} with L_{prob} and Z_{prob} with LT_{prob}. Thus accounting-based information summarized in Z_{prob} provides additional information of bankruptcy, not captured by market-based information obtained by our structural models and vice-versa.

Having found that our structural models have high discriminating ability (especially Leland-Toft) and that included in logit models as predictors retain their discriminating ability and improve accuracy and fit, we motivate to extend Z-score by including L_{prob} and LT_{prob} as additional predictors. Indeed, these market-based Z-scores significantly improve Z-score in terms of discriminatory power and predictive accuracy. Therefore, taking a benchmark empirical model we have managed to extend it into a powerful forecasting model by just including bankruptcy probabilities obtained from Leland and Leland-Toft as additional predictors. Hence, when accounting ratios along with market information obtained from structural models are combined together yield powerful models for predicting corporate bankruptcy. Most importantly these results are consistent with results from our additional tests and thus we overall have strong supportive evidence about the performance of our models. The conclusion of this study is that alternative structural models should be considered by researchers when studying bankruptcy and shift from traditional models.

Appendix-Per Period Performance (Five-Fold Validation Approach)

In this appendix, reported is the per-period performance of our models according to the five-fold validation approach. Table A1 reports the in-sample and out-of-sample performance of Z-score and of the two structural models in panels A and B respectively.

[Insert Table A1 here]

Differences in AUROCs between Z-score and Leland are not statistically significant according to DeLong tests (at both in-sample and out-of-sample periods), while for Z-score and Leland-Toft differences are statistically significant except in periods 3 and 4 (out-of-sample). Finally, predictive accuracy for Z-score is better since Log-Likelihood is much higher in all periods.

Table A2 reports in-sample and out-of-sample performance of models 6 and 7 in panels B and C respectively.

[Insert Table A2 here]

As can be inferred, per-period performance of model 6 and model 7 is better over the performance of Z-score. AUROCs for models 6 and 7 are significantly different than that of Z-score in all in-sample and out-of-sample periods according to DeLong tests. This is also the case for Log-Likelihoods according to LR tests. This means that for every period, models 6 and 7 are more accurate relative to Z-score. Finally, Pseudo-R^2 results suggest that models 6 and 7 explain bankruptcy better than Z-score during each period.

Another interesting result that arises is that discriminatory power of model 6 is much better than Leland model in all periods, whereas this is not the case for the discriminatory power of model 7 over Leland-Toft. In the out-of-sample periods, sometimes Leland-Toft is better but in periods where model 7 is better (i.e. periods 3 and 5) the effect is stronger and this possibly drives the aggregate AUROC (see table 11) to be slightly better for model 7 than Leland-Toft.

Panel A: In-sample performance of Z-score, Leland and Leland-Toft models							
	Z-score			Leland		Leland-Toft	
Period	AUROC	Log-Likelihood	Pseudo-R^2	AUROC	Log-Likelihood	AUROC	Log-Likelihood
2-5	84.72%	-1371.70	17.64%	83.94%	-9361.10	88.50%	-6739.70
1 & 3-5	85.78%	-1212.50	18.09%	83.52%	-9441.70	88.81%	-6267.10
1-2 & 4-5	83.75%	-1299.10	15.53%	83.62%	-8243.50	88.76%	-4821.80
1-3 & 5	84.20%	-1282.10	16.63%	83.71%	-8673.40	88.52%	-5951.20
1-4	83.51%	-1255.70	14.89%	83.45%	-8425.90	87.19%	-5723.40
Panel B: Out-of-sample performance of Z-score, Leland and Leland-Toft models							
1	79.53%	-238.36	8.60%	79.81%	-1637.90	86.25%	-636.99
2	81.03%	-397.86	10.56%	84.52%	-1607.90	86.72%	-1109.50
3	87.40%	-310.19	19.82%	84.05%	-2863.80	87.56%	-2595.30
4	84.47%	-334.48	13.54%	84.03%	-2392.90	88.34%	-1425.40
5	87.10%	-354.52	21.17%	84.78%	-2621.40	91.98%	-1682.90

Table A1: This table reports detailed-per period results of the performance of Z-score, Leland and Leland-Toft models when the total sample is divided chronologically in five equal sub-samples of 7966 observations each. Each time, four of them are used as the in-sample period and the left one is the out-of-sample period, in a way that all sub-samples to be in the in-sample period and out-of-sample period in turn. Panel A reports the results for each of our five in-sample periods (each period contains 31864 observations) and Panel B reports the results for each of our five out-of-sample periods (each period contains 7966 observations). Performance is measured by statistics such as AUROC, Log-Likelihood and Pseudo-R^2 (in this table Pseudo-R^2 is valid only for Z-score).

Panel A: Estimation of the two market-based Z-scores for each of the five in-sample periods										
	Periods 2-5		Periods 1 & 3-5		Periods 1-2 & 4-5		Periods 1-3 & 5		Periods 1-4	
Variables	model 6	model 7	model 6	model 7	model 6	model 7	model 6	model 7	model 6	model 7
Constant	-5.228***	-5.318***	-5.510***	-5.629***	-5.252***	-5.360***	-5.102***	-5.184***	-5.204***	-5.247***
	(0.135)	(0.132)	(0.150)	(0.146)	(0.140)	(0.141)	(0.140)	(0.137)	(0.139)	(0.135)
Financial Ratios										
WCTA	-2.315***	-2.216***	-2.246***	-2.115***	-2.230***	-2.096***	-2.408***	-2.306***	-2.208***	-2.146***
	(0.290)	(0.290)	(0.311)	(0.311)	(0.309)	(0.306)	(0.306)	(0.307)	(0.306)	(0.306)
RETA	-0.048	-0.081*	-0.089*	-0.129***	-0.07	-0.092**	-0.012	-0.042	-0.038	-0.057
	(0.044)	(0.044)	(0.046)	(0.047)	(0.042)	(0.043)	(0.045)	(0.046)	(0.048)	(0.049)
EBITTA	-2.528***	-2.213***	-2.449***	-2.117***	-2.359***	-2.244***	-2.567***	-2.243***	-2.273***	-2.021***
	(0.284)	(0.288)	(0.315)	(0.317)	(0.276)	(0.277)	(0.280)	(0.287)	(0.291)	(0.293)
EVF	-0.014	-0.011	-0.013	-0.010	-0.014	-0.012	-0.035	-0.029	-0.014	-0.013
	(0.010)	(0.008)	(0.014)	(0.011)	(0.011)	(0.01)	(0.022)	(0.018)	(0.011)	(0.009)
SLTA	0.172***	0.131**	0.249***	0.207***	0.217***	0.189***	0.109*	0.066	0.209***	0.176***
	(0.059)	(0.059)	(0.059)	(0.059)	(0.067)	(0.068)	(0.062)	(0.062)	(0.059)	(0.059)
Market-Based Bankruptcy Measures										
L_{prob}	2.765***		2.812***		2.953***		2.702***		2.550***	
	(0.158)		(0.171)		(0.164)		(0.164)		(0.173)	
LT_{prob}		3.080***		3.192***		3.365***		3.012***		2.7848***
		(0.162)		(0.176)		(0.169)		(0.168)		(0.175)
Panel B: Performance of the two market-based Z-scores for each of the five in-sample periods										
AUROC	89.15%	90.03%	90.02%	91.19%	88.44%	89.35%	88.63%	89.71%	87.36%	88.03%
Log-Likelihood	-1236.20	-1216.30	-1091.30	-1066.60	-1168.10	-1143.00	-1166.10	-1147.60	-1165.50	-1156.30
Pseudo-R^2	25.78%	26.97%	26.28%	27.94%	24.05%	25.68%	24.18%	25.38%	21.01%	21.63%

Panel C: Performance of the two market-based Z-scores for each of the five out-of-sample periods										
	Period 1		Period 2		Period 3		Period 4		Period 5	
AUROC	84.63%	85.52%	85.31%	85.70%	90.45%	92.12%	88.14%	88.17%	92.80%	94.23%
Log-Likelihood	-223.73	-220.01	-371.13	-372.93	-295.12	-300.31	-298.99	-293.62	-296.81	-284.23
Pseudo-R^2	14.21%	15.64%	16.57%	16.17%	23.71%	22.37%	22.71%	24.10%	34.00%	36.80%

Table A2: This table reports results on the performance of the two market-based Z-scores that include L_{prob}(model 6) and LT_{prob} (model 7) as additional explanatory variables in the Z-score when the total sample is divided chronologically in five equal sub-samples of 7966 observations each. Each time, four of them are used as the in-sample period and the left one is the out-of-sample period, in a way that all sub-samples to be in the in-sample period and out-of-sample period. WCTA is Working Capital to Total Assets, RETA is Retained Earnings to Total Assets, EBITTA is Earnings Before Interests and Taxes to Total Assets, EVF is Equity (Market) Value to Total Liabilities and SLTA is Sales to Total Assets. The dependent variable is a binary variable that equals one if the firm files for bankruptcy during the next year and zero otherwise. Panel A shows the estimation of the logit models for each of the five in-sample periods (each period contains 31864 observations). In parentheses Huber-White standard errors are reported. *, **, * indicate statistical significance at α=1%, α=5% and α=10%. Panel B reports the performance of the logit models for each of the five in-sample periods. Panel C reports the performance of the logit models for each of the five out-of-sample periods (each period contains 7966 observations). Performance is measured by statistics such as AUROC, Log-Likelihood and Pseudo-R^2.**

References

Afik, Z., Arad, O., & Galil, K. (2012). Using the Merton Model: An Empirical Assessment of Alternatives. *Discussion Paper No. 12-02*.

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.

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

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

Altman, E., Drozdowska, M. I., & Suvas, A. (2014). *Distress Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model*. Retrieved from http://people.stern.nyu.edu/ealtman/IRMC2014ZMODELpaper1.pdf

Anderson, R. W., Sundaresan, S., & Tychon, P. (1996). Strategic Analysis of Contigent Claims. *European Economic Review, 40*(3-5), 871-881.

Basel Committee on Banking Supervision (2005). *Studies on the Validation of Internal Rating Systems. Working Paper No. 14.*

Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research, 4*, 71-111.

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

Black, F., & Cox, J. C. (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance, 31*(2), 351-367.

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

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

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.

Collin-Dufresne, P., & Goldstein, R. S. (2001). Do Credit Spreads Reflect Stationary Leverage Ratio? *Journal of Finance, 56*(5), 1929-1957.

Crosbie, P., & Bohn, J. (2003). Modeling Default Risk, http://www.macs.hw.ac.uk/~mcneil/F79CR/Crosbie_Bohn.pdf. *Moody's KMV*.

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.

Du, Y., & Suo, W. (2007). Assessing Credit Quality from the Equity Market: Can a structural Approach Forecast Credit Ratings? *Canadian Journal of Administrative Sciences, 24*(3), 212-228.

Eom, Y. H., Helwege, J., & Huang, J. Z. (2004). Structural Models of Corporate Bond Pricing: An Empirical Analysis. *The Review of Financial Studies, 17*(2), 499-544.

Ericsson, J., & Reneby, J. (1998). A Framework for Valuing Corporate Securities. *Applied Mathematical Finance, 5*, 143-163.

Falkenstein, E., Boral, A., & Carty, L. V. (2000). RiskCalc for Private Companies: Moody's Default Model. *Moody's Investors Service Global Credit Research*.

Fernandes, J. E. (2005). *Corporate Credit Risk Modeling: Quantitative Rating System and Probability of Default Estimation*. Retrieved from http://www.defaultrisk.com/pp_score_47.htm

Geske, R. (1977). The Valuation of Corporate Liabilities as Compound Options. *Journal of Financial and Quantitative Analysis, 12*(4), 541-552.

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.

Huang, J. Z., & Huang, M. (2012). How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk? *The Review of Asset Pricing Studies, 2*(2), 153-202.

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.

Imerman, M. B. (2013). Structural Credit Risk Models: Endogenous Versus Exogenous Default. In C.-F. Lee, & A. C. Lee, *Encyclopedia of Finance* (pp. 645-657). New York: Springer US.

Leland, H. (1994). Corporate Debt Value, Bond Covenants, and Optimal Capital Structure. *Journal of Finance, 49*(4), 1213-1252.

Leland, H. (2004). Predictions of Default Probabilities in Structural Models of Debt. *Journal of Investment Management, 2*(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*(3), 987-1019.

Longstaff, F. A., & Schwartz, E. S. (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *Journal of Finance, 50*(3), 789-819.

Lyden, S., & Saraniti, D. (2001). *An Empirical Examination of the Classical Theory of Corporate Security Valuation*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=271719

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

Ogden, J. (1987). Determinants of the Ratings and Yields on Corporate Bonds: Tests of the Contigent Claims Model. *The Journal of Financial Research, 10*(4), 329-339.

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

Patel, K., & Pereira, R. (2007). Expected Default Probabilities in Structural Models: Empirical Evidence. *Journal of Real Estate Finance and Economics, 34*(1), 107-133.

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

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

Suo, W., & Wang, W. (2006). Assessing Default Probabilities from Structural Credit Risk Models. *Working Paper, Queens University*.

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

Tarashev, N. (2008). An Empirical Evaluation of Structural Credit-Risk Models. *International Journal of Central Banking, 4*(1), 1-53.

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.

Wong, T. C., Hui, C. h., & Lo, C. f. (2010). Discriminatory Power and Predictions of Defaults of Structural Credit Risk Models. *The Journal of Risk Model Validation, 3*(4), 39-60.

Wu, Y., Gaunt, C., & Gray, S. (2010). A Comparison of Alternative Bankruptcy Prediction Models. *Journal of Contemporary Accounting & Economics, 6*(1), 34-45.

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

Tables

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

Table 1: Distribution 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).

Industry	# 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

Table 2: Industry distribution of observations. Each observation is classified in one of the above industries shown in column 1, 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.

Variable/Parameter	Symbol	Estimation
Working Capital to Total Assets	WCTA	WC/TA
Retained Earnings to Total Assets	RETA	RE/TA
Earnings before Interest and Taxes to Total Assets	EBITTA	EBIT/TA
Equity Value to Total Liabilities	EVF	EV/F
Sales to Total Assets	SLTA	SL/TA
Dividends (Ordinary and Preferred)	D	end-of-year ordinary + preferred dividends
Annualized Asset Volatility	σ_V	monthly volatility x $\sqrt{12}$
Annualized Asset Return	μ	monthly return x 12
Assets Value	V	$E + F$
Risk-free rate	r	one-year Treasury Constant Maturity rate
Coupon	C	end-of-year Interest Expense
Debt Principal	P	Total Liabilities, F
Payout-yield	d	(C+D)/V
Tax rate	τ	15% as in Leland (2004)
Bankruptcy Costs	α	30% as in Leland (2004)
Maturity	T	10 years as in Leland (2004)

Table 3: Description and estimation of the variables needed for the construction of Z-score, Leland and Leland-Toft models. All variables are constructed using financial information one year prior to bankruptcy filing. The first column of the table shows the name of the variable, the second column shows how it is entered in our models and column three shows how they are calculated.

Variable	Status	Mean	St.Dev	Min	Q1	Median	Q3	Max
WCTA	Healthy	0.2274***	0.2281	-0.3921	0.0595	0.2080***	0.3772	0.8015
	Bankrupt	-0.0089	0.29	-0.3921	-0.2865	-0.0053	0.1959	0.6518
RETA	Healthy	-0.1882***	1.2419	-7.8156	-0.1426	0.1347***	0.3425	0.9066
	Bankrupt	-1.7361	2.4758	-7.8156	-2.3729	-0.6037	-0.0996	0.9066
EBITTA	Healthy	0.0291***	0.1919	-1.0085	0.0120	0.0709***	0.1193	0.3114
	Bankrupt	-0.2380	0.3553	-1.0085	-0.4101	-0.0934	0.0085	0.3114
EVF	Healthy	38.2574***	160.92	0.0017	0.3235	1.65***	8.0511	1297.4
	Bankrupt	3.1027	20.70	0.0017	0.0144	0.1260	0.8257	322.88
SLTA	Healthy	1.1721	0.8231	0.0053	0.5991	1.0197	1.5194	4.4725
	Bankrupt	1.2504	0.9698	0.0053	0.5513	1.0893	1.6597	4.4725
σ_V	Healthy	0.2973	0.2316	0.0045	0.1390	0.2328	0.3881	1.2193
	Bankrupt	0.3361***	0.2733	0.0045	0.1252	0.2584	0.4810	1.2193
d	Healthy	0.0234	0.0253	0.00	0.0052	0.0167	0.0323	0.1447
	Bankrupt	0.0518***	0.04	0.00	0.0193	0.0463***	0.0745	0.1447
μ	Healthy	0.0267***	0.3730	-1.2508	-0.1299	0.0227***	0.1904	1.2336
	Bankrupt	-0.2802	0.4009	-1.2508	-0.4653	-0.1911	-0.0484	1.2336
C/V	Healthy	0.0167	0.0524	0.00[a]	0.0031	0.0098	0.0218	4.7936
	Bankrupt	0.0965***	0.4513	0.00	0.0167	0.0437***	0.0664	7.3520
P/V	Healthy	0.4043	0.2501	0.00	0.2006	0.3695	0.5799	1.00
	Bankrupt	0.6946***	0.2634	0.0188	0.5202	0.7853***	0.9052	0.9979
Z_{prob}	Healthy	0.0072	0.0151	0.00	0.0022	0.0044	0.0078	0.3620
	Bankrupt	0.0372***	0.0516	0.00	0.0081	0.0185***	0.0425	0.5065
L_{prob}	Healthy	0.044	0.1709	0.00	0.00	0.00	0.00	0.9999
	Bankrupt	0.3909***	0.4211	0.00	0.00	0.145***	0.8885	0.9999
LT_{prob}	Healthy	0.0471	0.1644	0.00	0.00	0.00	0.00	0.9999
	Bankrupt	0.4189***	0.3872	0.00	0.0189	0.3311***	0.8194	0.9999

Table 4: This table presents descriptive statistics about the financial ratios used in the Z-score model, the variables used in the Leland and Leland and Toft models and the bankruptcy probabilities estimated by the three models. *, **, * indicate significant statistical difference at α=1%, α=5%, α=10% respectively between mean and median values for bankrupt and healthy firms. For the mean, a two-tailed t-test is used while for median the Wilcoxon rank-sum test is used.**

[a] Variables with values less than 0.0001 are entered as 0.00 in the table.

Variable	β-Coefficient	p-value (%)
Constant	-4.520***	0.00
	(0.129)	
WCTA	-3.459 ***	0.00
	(0.360)	
RETA	0.046	38.50
	(0.052)	
EBITTA	-2.215***	0.00
	(0.301)	
EVF	-0.015	18.00
	(0.114)	
SLTA	0.107 *	9.80
	(0.065)	

Table 5: This table reports the results of Z-score estimated from logistic regression using financial information of firms collected in years 1995-2005. WCTA is the Working Capital to Total Assets, RETA is the Retained Earnings to Total Assets, EBITTA is the Earnings before Interests and Taxes to Total Assets, EVF is the Equity Value to Total Liabilities and SLTA is the Sales to Total Assets. The dependent variable is a binary variable that equals one if the firm files for bankruptcy during the next year and zero otherwise. Second column reports the coefficient estimates and third column of the table show the p-value of a two-sided test whether the coefficient estimates are zero (Null Hypothesis) or not (Alternative Hypothesis). In parentheses Huber-White standard errors are reported. *, **, * indicate statistical significance at α=1%, α=5% and α=10%.**

	In-sample period (1995-2005)		Out-of-sample period (2006-2014)	
Model	AUROC	z-statistic	AUROC	z-statistic
Z-score	81.77%	-	85.69%	-
Leland	83.29%	0.841	84.18%	-0.621
Leland-Toft	86.97%	3.135***	90.31%	2.444***

Table 6: This table reports results from tests on the discriminatory power of Leland and Leland-Toft relative to Z-score in two periods; in-sample period spans the years 1995-2005 and out-of-sample period spans the years 2006-2014. Under the null hypothesis, the AUROC of Leland and Leland-Toft are equal to that of Z-score whereas under the alternative hypothesis, the AUROC of Leland and Leland-Toft models differ from that of Z-score. *, **, * indicate statistical significance at α=1%, α=5% and α=10% respectively by performing a two-tailed test, where the test statistic (z-statistic) is constructed using the non-parametric method of DeLong et.al (1988).**

	In-sample period (1995-2005)		Out-of-sample period (2006-2014)	
Model	Log-Likelihood	Pseudo-R^2	Log-Likelihood	Pseudo-R^2
Z-score	-1012.83	13.96%	-606.17	19.02%
Leland	-6810.11	-	-4251.68	-
Leland-Toft	-4885.68	-	-2498.18	-

Table 7: This table reports fitness results of Z-score, Leland and Leland-Toft models for two periods; in-sample period spans the years 1995-2005 and out-of-sample period spans the years 2006-2014. For both periods, Log-Likelihood and Pseudo-R^2 (valid for Z-score) statistics are provided to measure the predictive accuracy of the three models.

Panel A: Model Estimation (period 1995-2005), 25950 observations					
Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-5.106*** (0.074)	-5.296*** (0.087)	-5.387*** (0.088)	-5.509*** (0.090)	-5.576*** (0.091)
Z_{prob}	16.977*** (1.077)			15.527*** (1.10)	14.278*** (1.098)
L_{prob}		3.124*** (0.158)		2.923*** (0.166)	
LT_{prob}			3.430*** (0.152)		3.200*** (0.162)
Panel B: In-sample performance (1995-2005), 25950 observations					
AUROC	81.77%	82.99%	87.01%	85.56%	87.75%
Log-Likelihood	-1111.33	-1061.65	-1039.82	-1012.60	-997.602
Pseudo-R^2	5.60%	9.82%	11.67%	13.98%	15.26%
Panel C: Out-of-sample performance (2006-2014), 13880 observations					
AUROC	85.69%	83.62%	90.27%	89.62%	91.34%
Log-Likelihood	-688.11	-620.07	-597.45	-576.25	-556.73
Pseudo-R^2	8.07%	17.16%	20.18%	23.01%	25.62%

Table 8: This table reports results on several logit models that include bankruptcy probability as explanatory variables; Z_{prob}, L_{prob} and LT_{prob} which are bankruptcy probabilities estimated by Z-score, Leland and Leland-Toft models respectively. The dependent variable is a binary variable that equals one if the firm files for bankruptcy during the next year and zero otherwise. Panel A shows the estimation of the logit models. The estimation is made using the in-sample period 1995-2005 (25950 observations). In parentheses Huber-White standard errors are reported. *, **, * indicate statistical significance at α=1%, α=5% and α=10%. Panel B reports the performance of the logit models on the in-sample period. Panel C reports the performance of the logit models at the out-of-sample period 2006-2014 (13880 observations). Performance is measured by statistics such as the AUROC, Log-Likelihood and Pseudo-R^2.**

Panel A: Model estimation (1995-2005), 25950 observations		
Variables	Model 6	Model 7
Constant	-5.076***	-5.125***
	(0.148)	(0.146)
Financial Ratios		
WCTA	-2.486***	-2.424***
	(0.339)	(0.341)
RETA	0.038	0.02
	(0.055)	(0.055)
EBITTA	-2.429***	-2.158***
	(0.307)	(0.310)
EVF	-0.01	-0.01
	(0.009)	(0.008)
SLTA	0.132*	0.104
	(0.068)	(0.067)
Market-Based Bankruptcy Measures		
L_{prob}	2.431***	
	(0.192)	
LT_{prob}		2.671***
		(0.195)
Panel B: In-sample performance (1995-2005), 25950 observations		
AUROC	85.92%	86.73%
Log-Likelihood	-947.142	-938.731
Pseudo-R^2	19.54%	20.26%
Panel C: Out-of-sample performance (2006-2014), 13880 observations		
AUROC	91.66%	92.64%
Log-Likelihood	-523.29	-510.76
Pseudo-R^2	30.09%	31.76%

Table 9: This table reports results of two market-based versions of Z-score model that include bankruptcy probability and financial ratios as explanatory variables; L_{prob} and LT_{prob} which are bankruptcy probabilities produced by Leland and Leland-Toft models. WCTA is Working Capital to Total Assets, RETA is Retained Earnings to Total Assets, EBITTA is Earnings Before Interests and Taxes to Total Assets, EVF is Equity (Market) Value to Total Liabilities and SLTA is Sales to Total Assets. The dependent variable is a binary variable that equals one if the firm files for bankruptcy during the next year and zero otherwise. Panel A shows the estimation of the logit models. The estimation is made using the in-sample period 1995-2005 (25950 observations). In parentheses Huber-White standard errors are reported. *, **, * indicate statistical significance at α=1%, α=5% and α=10%. Panel B reports the performance of the logit models on the in-sample period. Panel C reports the performance of the logit models at the out-of-sample period 2006-2014 (13880 observations). Performance is measured by statistics such as the AUROC Log-Likelihood and Pseudo-R^2.**

Model	AUROC	Log-Likelihood	Pseudo-R^2
Z-score	86.52%	-596.06	20.17%
model 6	92.52%	-514.10	31.15%
model 7	93.48%	-498.96	33.17%

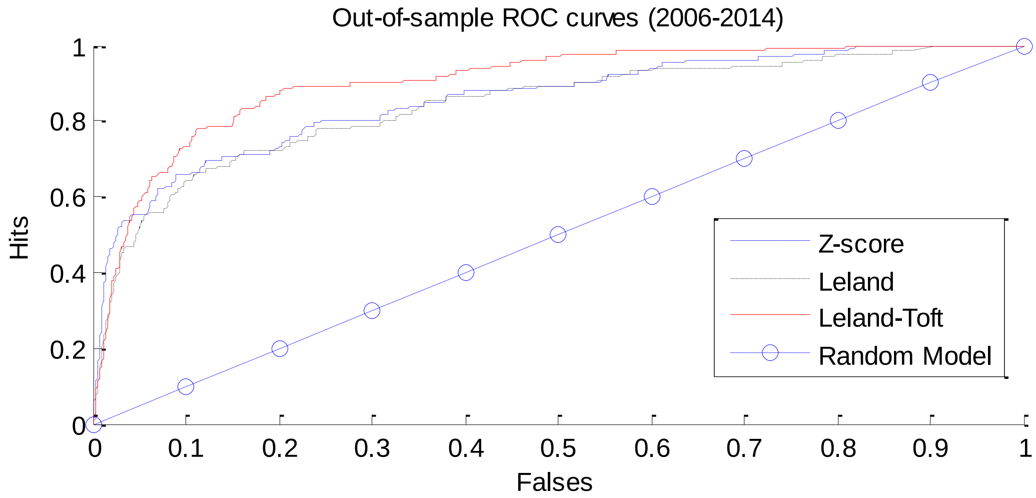
Table 10: This table reports aggregate out-of-sample performance for the three empirical models; Z-score, model 6 that includes L_{prob} as additional predictor in Z-score and model 7 that includes LT_{prob} as additional predictor in Z-score. The models are estimated on a specific time period (with 1995-2005 being the first one) and applied on next year firms. Then the oldest firms are excluded and the newest ones are included and repeat this procedure until we reach on the last year (2014). For each out-of-sample period, bankruptcy probabilities are kept and combined in order to measure overall performance, where performance is measured by AUROC, Log-Likelihood and Pseudo-R^2.

Panel A: Aggregate out-of-sample performance				
Model	# of observations	AUROC	Log-Likelihood	Pseudo-R^2
Z-score	39830	83.79%	-1635.41	15.22%
Leland	39830	83.46%	-11123.87	-
Leland-Toft	39830	88.13%	-7450.12	-
model 6	39830	87.92%	-1485.78	22.98%
model 7	39830	88.69%	-1471.10	23.74%

Table 11: This table reports the aggregate out-of-sample performance of Z-score, Leland, Leland-Toft and of the two market based Z-scores that include L_{prob} (model 6) and LT_{prob} (model 7) as additional explanatory variables in Z-score. The whole sample is divided in chronological order into five equal sub-samples with 7966 observations each. Each time, four of them are used as the in-sample period and the left one is the out-of-sample period, in a way that all sub-samples to be in the in-sample period and out-of-sample period in turn. This procedure generates five in-sample periods with five corresponding out-of-sample periods. The empirical models (Z-score, model 6 and model 7) are estimated for each in-sample period and they are applied to the corresponding out-of-sample period in order to estimate bankruptcy probabilities for out-of-sample firms. Panel A shows the performance of the five models when bankruptcy probabilities of the five out-of-sample periods are combined and thus obtaining aggregate-single indicators of performance measured by AUROC, Log-Likelihood and Pseudo-R^2 (see for instance table A2 in the appendix which reports the estimation of model 6 and model 7 for each in-sample period).

Figures

Plot 1



Plot 2

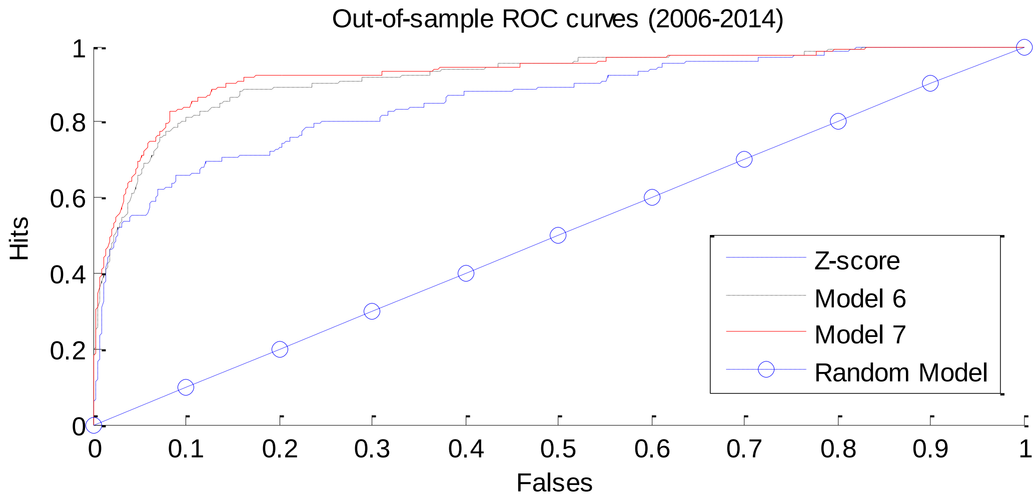


Figure 1: These set of plots show the ROC curves of five models; Z-score, Leland and Leland-Toft (plot 1) and Z-score, model 6 and model 7 (plot 2). The models are estimated using firms collected during period 1995-2005 and applied on firms collected during 2006-2014. Corresponding ROC curves are obtained for firms in the latter period.

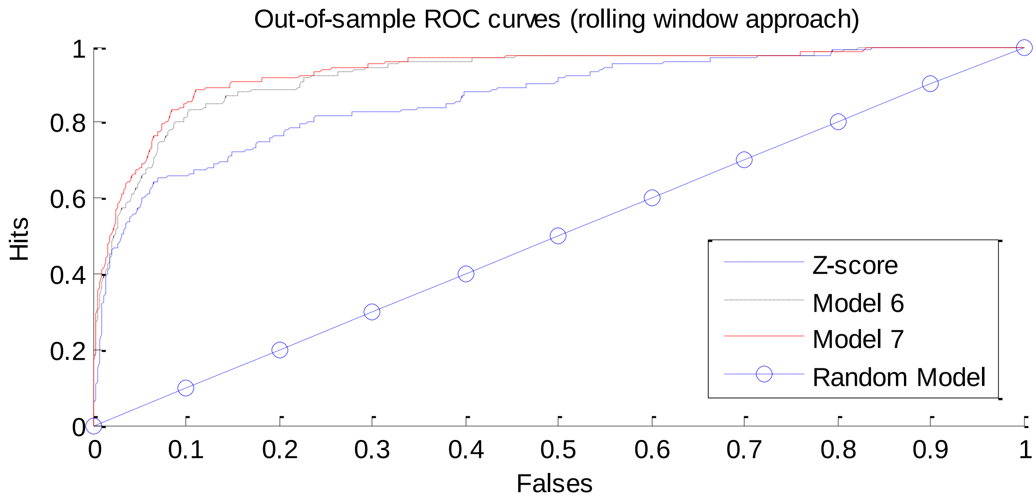
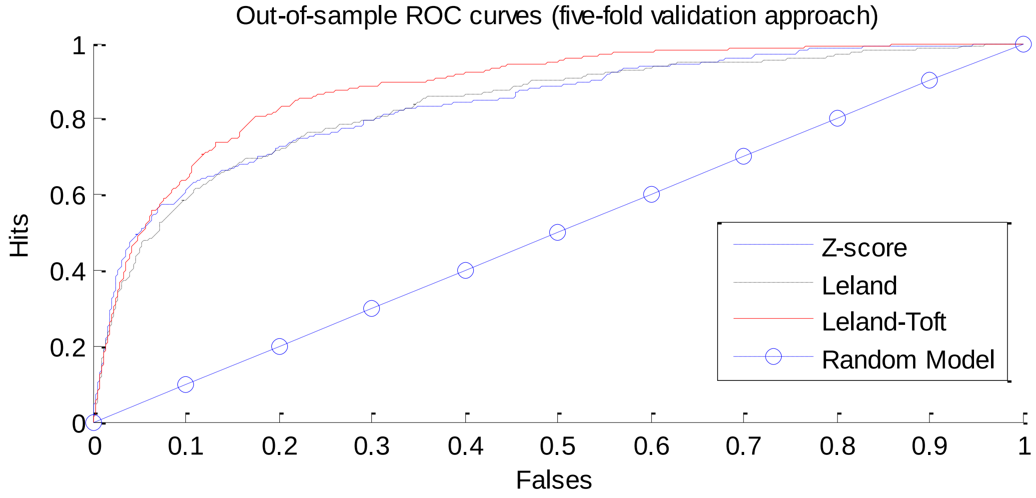


Figure 2: This plot shows the ROC curves of the three empirical models; Z-score, Leland and Leland-Toft. The models are estimated on a specific time period (with 1995-2005 being the first one) and applied on next year firms. Then the oldest firms are excluded and the newest ones are included and repeat this procedure until we reach on the last year (2014). For each out-of-sample year-period, bankruptcy probabilities are estimated and combined in order to construct the ROC curves.

Plot 1



Plot 2

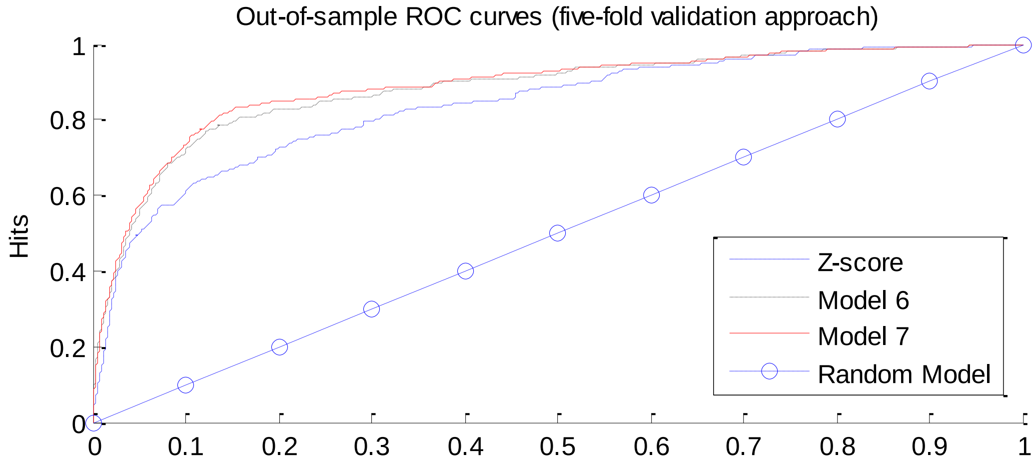


Figure 3: These set of plots show the ROC curves of five models; Z-score, Leland and Leland-Toft (plot 1) and Z-score, model 6 and model 6 (plot 2). The whole sample is divided in chronological order into five equal sub-samples with 7966 observations each. Each time, four of them are used as the in-sample period and the left one is the out-of-sample period, in a way that all sub-samples to be in the in-sample period and out-of-sample period in turn. This procedure generates five in-sample periods with 31864 observations each and five corresponding out-of-sample periods with 7966 observations each. The models are estimated for each in-sample period and they are applied to the corresponding out-of-sample period. Then bankruptcy probabilities obtained for each out-of-sample period are combined in order to obtain corresponding ROC curves.