Forecasting Bankruptcy via Cross-Sectional Earnings Forecasts

MARTIN HUETTEMANN

DIETER HESS*

* For their insightful discussions and suggestions, we are grateful to Alexander Kempf, Tobias Lorsbach and Martin Meuter. The paper has also greatly benefitted from comments by seminar participants at the University of Cologne.

Contact: Martin Huettemann (presenting author), Cologne Graduate School in Management, Economics and Social Sciences (CGS), Universitaetsstr. 22a, 50937 Koeln, Germany, e-mail: huettemann@wiso.uni-koeln.de. Dieter Hess, University of Cologne, Albertus Magnus Platz, 50923 Koeln, Germany, and Centre for Financial Research (CFR), e-mail: hess@wiso.uni-koeln.de.

Forecasting Bankruptcy via Cross-Sectional Earnings Forecasts

Abstract

We develop a structural model to predict bankruptcies, exploiting that a firm's over-indebtedness (negative book equity) is a state of immediate financial distress. Accordingly, our key predictor of bankruptcy is the probability that future losses deplete the book equity. To calculate this probability, we use earnings forecasts and their standard deviations that we obtain from cross-sectional models. However, not all over-indebted firms finally turn bankrupt. Thus, in an expanded model, we add accounting variables that we find to discriminate between bankrupt and non-bankrupt firms with negative book equity. As these models solely require accounting data, we can provide bankruptcy predictions for a wide range of firms, including firms that have no access to capital markets. In strictly out-of-sample tests we show that our accounting model performs substantially better than alternatives of corporate failure risk that solely use accounting information. If we allow market information, we significantly outperform all leading alternatives, including those that use market information.

Keywords: probability of insolvency --- negative book equity --- structural model

Data Availability: Data used in this study are available from public sources identified in the study.

JEL Classification: G17, G33, M41, C25

EFM Codes: 130, 140

1. Introduction

General Motors, Lehman Brothers and WorldCom are only a few examples of bankruptcies with a huge impact on capital markets. Predicting corporate failures is critical for investors, managers, regulators and banks. For example, it enables investors to avoid specific securities, managers and regulators to take corrective actions or banks to decide whom to grant loans. The goal of this paper is to develop a structural model for bankruptcy prediction. Our overall assumption is that a firm's over-indebtedness is a good measure of bankruptcy. A firm is over-indebted if its book equity turns negative, i.e., if the value of its assets falls below the value of obligations it must service. In the U.S. a firm is not required to voluntarily file for bankruptcy when it becomes over-indebted. However, an over-indebtedness is a strong indication of financial distress, making it less likely for the firm to obtain further credit and ultimately to pay its debts when they come due.

We calculate next year's book equity as the sum of the current book equity and the earnings forecast (change in retained earnings). Thus, our key predictor is the probability that a firm's current book equity is smaller than its expected losses. To calculate this probability, we look at the distribution of a firm's losses. Our approach shares the use of the loss distribution with the value at risk concept. The value at risk is defined as that loss value which is not exceeded by a pre-specified probability. We, however, start with a specific value (book equity) and then measure the probability that a firm's loss exceeds this specific value. We use cross-sectional linear regressions as proposed by Hou et al. (2012) and Li and Mohanram (2014) to predict the range of possible earnings (or losses). Thus, we use lagged accounting ratios of all firms to calculate earnings forecasts of an individual firm. These linear regressions provide us with earnings forecasts and the corresponding standard errors. By this, we can derive a closed formula for the probability that the book equity becomes negative. To our best knowledge, this is the first study to incorporate cross-sectional earnings forecasts into bankruptcy prediction.

Our approach has a major advantage over non-structural models that use accounting or market ratios to discriminate between bankrupt and non-bankrupt firms (e.g., Altman (1968), Ohlson (1980)) as indicated by Vassalou and Xing (2004). We include the volatility of earnings forecasts. In non-structural models, firms with similar ratios inevitably have a similar distress risk. In our model, however, firms with similar levels of book equity and similar earnings forecasts might have a completely different distress risk if their volatility differs.

We present three bankruptcy prediction models: an over-indebtedness model, an accounting model and a market model. Our over-indebtedness model exploits the fact that over-indebtedness (negative book equity) is a good measure of bankruptcy. Thus, it consists of one single variable: the probability that forecasted losses exceed the book equity. This probability is calculated by the current book equity, the earnings forecast and the volatility of the earnings forecast. However, over-indebted firms do not inevitably turn bankrupt. Firms might intentionally operate on negative book equity, for example due to tax avoidance. Thus, our accounting model adds further accounting-based independent variables to discriminate between negative book equity firms that keep operating and those that turn bankrupt. Haowen (2015) finds, for example, that non-bankrupt negative book equity firms tend to have a higher book leverage ratio, have more capital expenditures, pay less tax, have a lower profitability and a smaller size. In addition, our market model replaces book leverage ratio by market leverage ratio. It further adds two common market-based variables: the excess return and its standard deviation. For example, Shumway (2001) and Campbell, Hilscher and Szilagyi (2008) show that market variables raise performance. In contrast, Reisz and Perlich (2007) and Agarwal and Taffler (2008) demonstrate that accounting models show a similar performance as market models. We analyze the out-of-sample performance of our three models and compare them to leading alternatives: We estimate the models of Altman (1968), Ohlson (1980) and Shumway (2001) and the best model version of Merton's distance-to-default approach as found by Bharath and

Shumway (2008). To eliminate the effects of different statistical methods we embed all these models into logistic regressions that exploit the full firm history.

Our empirical results are manifold. First, we find justification for our overall approach to estimate a firm as bankrupt, if it becomes over-indebted. We find that book equity and earnings diminish in the years before bankruptcy. The most dramatic fall happens in the year directly before bankruptcy. Second, we show that the probability of negative book equity is a good measure of bankruptcy by its own. Our one-variable over-indebtedness model produces reasonably better results than Altman and Ohlson. At the same time, we find that this probability is not a sufficient measure. Thus, our accounting model and our market model add covariates. Third, we find differences in the means of certain variables for bankrupt and non-bankrupt firms with negative book equity, respectively. We validate Haowen's (2014) results for the market leverage ratio, the profitability and the size. Fourth, we demonstrate that our purely accounting model significantly outperforms all other models that solely require accounting information. Fifth, we further improve performance, if we allow market information in our models: Our market model shows significantly better results than our accounting model. Importantly, it outperforms all leading alternatives of bankruptcy prediction, including those that use market information. By this, we support Shumway (2001), Beaver, McNichols and Rhie (2005) and Campbell, Hilscher and Szilagyi (2008) who demonstrate that market variables add explanatory power. Sixth, we provide evidence for the fact that the non-linear functional form that we use for the probability of negative book equity is a meaningful construct for predicting bankruptcies. We show that the functional value of our three inputs has explanatory power that is not covered when we use the inputs used to calculate the probability as individual predictors of bankruptcy.

Major improvements in the performance of bankruptcy prediction models have been achieved by the inclusion of market information (e.g., Shumway (2001), Hillegeist et al. (2004)). This performance boost, however, comes with a cost. These models are limited to firms with access to the capital market. However, Altman, Iwanicz-Drozdowska, Laitinen and Suvas (2017) point out the importance of predicting bankruptcies of private firms as well. Importantly, our purely accounting model achieves better out-of-sample performances than leading alternatives that solely use accounting information. Thus, we improve performance without limiting the scope of application. We can still provide bankruptcy predictions for a wide range of firms. This includes firms without access to capital markets, for example large start-up companies such as Facebook before their IPO.

There are two types of structural bankruptcy prediction models that are related to our approach as they use equity as an indication of bankruptcy. First, e.g., Vassalou and Xing (2004), Hillegeist, Keating, Cram and Lundstedt (2004), Bharat and Shumway (2008) and Charitou, Dionysiou, Lambertides and Trigeorgis (2013) use Merton's (1974) option pricing theory (OPT) to compute default probabilities. They view the market value of equity as a call option on the market value of assets where the strike price is the book value of liabilities. Thus, if market assets fall below liabilities, market equity goes to zero and the firm goes bankrupt. The value of assets is assumed to follow a Geometric Brownian Motion (GBM). OPT models do not contain any earnings variable. Furthermore, they require restrictive assumptions, e.g., that firm's assets follow a GBM and that a firm has one single zero-coupon bond. Second, Feller (1968) develops the Gambler's Ruin Theory (GRT) which says that a firm fails if its book equity turns negative. It assumes that a firm does not have access to the capital market and thus can meet losses solely by selling assets. An extension to GRT is Scott's (1976) perfect-access model (PAM) which assumes that a firm goes bankrupt due to investors' negative expectations. Accordingly, PAM uses market equity instead of book equity. For example, Wilcox (1973, 1976), Santomero and Vinso (1977) and Vinso (1979) apply GRT or PAM by using equity information along with the mean and the volatility of an earnings variable. However, earnings are modelled via time series that only exploit past earnings of the firm itself.

To model future equity, we do not require a closed theory such as OPT that comes along with assumptions and restrictions. Instead, we say that a firm's future book equity can be directly calculated as the current book equity plus the change in retained earnings. To forecast these earnings, we use cross-sectional models instead of time series. Thus, we exploit the history of all firms and a broader dataset. In contrast to time series models, cross-sectional models can provide forecasts for firms with no firm history. We further use accounting information to produce more accurate earnings forecasts. As we use book equity instead of market equity, we can also provide bankruptcy predictions for firms without access to the capital market.

The paper proceeds as follows: The second section describes and motivates the variables that we use in our bankruptcy prediction models. In the third section we describe our sample selection, report descriptive statistics and explain our methodology. In the fourth section we present and discuss our results. The fifth section concludes.

2. Constructing measures of bankruptcy

2.1 Probability of negative book equity – Over-indebtedness Model

Our overall assumption is that a firm's over-indebtedness is a good measure of bankruptcy. The rationale behind is the following: A firm is over-indebted if its book equity turns negative, i.e., if the value of its assets falls below the value of obligations it must service. A firm is not required to voluntarily file for bankruptcy protection when it becomes over-indebted. Likewise, a creditor cannot make an involuntary petition for bankruptcy filing in case of an over-indebtedness. However, over-indebtedness is a strong indication of financial distress, making it less likely for the firm to obtain further credit and ultimately to pay its debts when they come due.

Let $BkEq_{i,t}$ denote the current book equity of a firm i and let $Earn_{i,t+1}$ denote the future earnings for this firm for the subsequent period, t+1. We calculate next year's book equity as the sum of the current book equity and the earnings (change in retained earnings). Then, over-indebtedness occurs if the following sum is negative:

$$\mathrm{BkEq_{i,t}} + Earn_{i,t+1} < 0. \quad (1)$$

Accordingly, the main question is whether the firm might incur future losses that exceed its book equity. Thus, our key predictor is the probability that a firm's expected earnings are smaller than its negative current book equity:

$$\mathrm{PNBE_{i,t}} = \mathrm{Prob}\left(Earn_{i,t+1} < -\mathrm{BkEq_{i,t}}\right). \quad (2)$$

To calculate this probability, we use expected earnings, $\widehat{Earn_{\iota,t+1}}$ derived from rolling cross-sectional earnings forecast models in the spirit of Hou et al. (2012) and Li and Mohanram (2014) (for a description see Section 2.2). These linear regressions also provide us with a measure of uncertainty of these forecasts, i.e., the corresponding volatility of the individual earnings forecast, $\sigma\left(\widehat{Earn_{\iota,t+1}}\right)$.[1] Using these two estimates we can construct the prediction interval for the actual future earnings $Earn_{i,t+1}$. The lower one-sided prediction interval of firm i with level α contains all values that the actual future earnings value does not exceed with probability α. It is specified as

$$(-\infty, \widehat{Earn_{\iota,t+1}} - u_{1-\alpha} \cdot \sigma\left(\widehat{Earn_{\iota,t+1}}\right)],$$

that is

$$P\left(Earn_{i,t+1} \leq \widehat{Earn_{\iota,t+1}} - u_{1-\alpha} \cdot \sigma\left(\widehat{Earn_{\iota,t+1}}\right)\right) = \alpha, \quad (3)$$

[1] $\sigma\left(\widehat{Earn_{\iota,t+1}}\right)$ denotes the standard deviation of a predicted response of an individual firm for given data rather than the standard deviation of the estimated conditional mean. Thus, it yields the prediction interval rather than the confidence interval. In addition to the uncertainty in estimating the conditional mean, $\sigma\left(\widehat{Earn_{\iota,t+1}}\right)$ also reflects the variability of an individual observation in this conditional distribution: $\sigma\left(\widehat{Earn_{\iota,t+1}}\right) = \hat{\sigma}\sqrt{(1 + \widehat{x_\iota(X'X)^{-1}x_\iota})}$, where $\hat{\sigma} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y_\iota})}$ is the standard deviation of the residuals, x_i is the explanatory vector of firm i, X is the data matrix, y_i is the outcome of firm i and $\widehat{y_\iota}$ is the predicted outcome of firm i.

where $u_{1-\alpha}$ is the $(1-\alpha)$-quantile of a standard normal distribution.[2]

The prediction interval in formula (3) gives us an upper bound for the actual earnings value. This range is dependent on an exogenously given probability level α. Our value of interest is the probability in formula (2) that the upper bound for actual earnings is $-\mathrm{BkEq}_{\mathrm{i,t}}$. Then $PNBE_{i,t}$ is equal to that probability level $\hat{\alpha}$ in the prediction interval that equates both upper bounds:

$$-\mathrm{BkEq}_{\mathrm{i,t}} = \widehat{Earn_{\iota,t+1}} - u_{1-\hat{\alpha}} \cdot \sigma\left(\widehat{Earn_{\iota,t+1}}\right) \quad (4)$$

$$\Rightarrow \; PNBE_{i,t} \coloneqq \hat{\alpha} = 1 - \Phi\left(\frac{\widehat{Earn_{\iota,t+1}} + \mathrm{BkEq}_{\mathrm{i,t}}}{\sigma\left(\widehat{Earn_{\iota,t+1}}\right)}\right) \quad (5)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The probability of default depends on the sum of expected earnings and the current book equity relative to the volatility of the earnings forecast. OPT models view the market equity as a call option on the market value of assets where the strike price is the market value of liabilities. By this, they use a different setting than our model. Nevertheless, Vassalou and Xing (2004) and Hillegeist et al. (2004) find a formula for the probability of default with a similar structure. It depends on the ratio of expected future market equity and the asset volatility. The future market equity in the numerator is calculated as current market assets minus current liabilities plus the expected asset changes extracted from the Geometric Brownian Motion. Our PNBE and the probability extracted from OPT have in common that a lower volatility and a larger market equity reduce the default probability.

Our probability of default does not only consist of the expected firm's book equity. By using the volatility of the earnings forecast we provide substantially different inference about the risk one firm faces. In non-structural models, firms

[2] Prediction intervals of linear regressions are determined by the quantile of a t-distribution with $n-p$ degrees of freedom, $t_{1-\alpha,n-p}$. The amount of observations in the regression equals n and the number of independent variables in the regression equals p. Due to the large amount of observations n-p is consistently far above 40. Thus, the quantiles of the t-distribution are well approximated by quantiles of a standard normal distribution.

with similar ratios inevitably have a similar distress risk. In our model, however, firms with similar levels of market equity and similar earnings forecasts might have a completely different distress risk if their volatility differs.

2.2 Earnings Forecasts

We use earnings forecasts for the subsequent year to calculate PNBE. Following Hess, Meuter and Kaul (2017), who compare the performance of several cross-sectional models, we implement the RI model of Li and Mohanram (2014) on a per-share basis.[3] Note that all predictor variables are lagged by three months to ensure that they are publicly observable at the time we use them. That is, we make one-year estimations three months after the fiscal year end. This approach ensures that the estimation is made promptly as soon as all information are in hand. By this, we differ from Hou et al. (2012) and Li and Mohanram (2014) who only make predictions at the end of next June. We use a rolling regression technique by using the past 10 years of accounting data to estimate parameters that we use for forecasting. Hereby, we only use data that is publicly available by the estimation date. That is, once a month we run the following cross-sectional OLS:

$$E_{i,t+\tau} = \alpha_0 + \alpha_1 E_{i,t} + \alpha_2 NegE_{i,t} + \alpha_3 NegE_{i,t} \cdot E_{i,t}$$
$$+ \alpha_4 BkEq_{i,t} + \alpha_5 AC_{i,t} + \varepsilon_{i,t+\tau}, \qquad (6)$$

where $E_{i,t}$ denotes the change in retained earnings per share of firm i at time t, $NegE_{i,t}$ is a dummy that takes the value of 1 if firm i shows negative earnings at time t and $NegE_{i,t} \cdot E_{i,t}$ is an interaction term. $BkEq_{i,t}$ is the book value of equity per share and $AC_{i,t}$ are accruals per share. Using the coefficients from this regression we can easily calculate out-of-sample predictions, i.e., earnings forecasts for the subsequent year. These regressions provide us with the forecasted values of the

[3] We also tested the cross-sectional earnings forecast model of Hou et al. (2012) and the EP model of Li and Mohanram (2014) and all models with level earnings instead of per-share earnings. All results of this paper are robust regarding this method.

future earnings $\widehat{Earn_{\iota,t+1}}$ along with the corresponding individual standard deviations $\sigma\left(\widehat{Earn_{\iota,t+1}}\right)$.

2.3 Accounting Model

Our over-indebtedness model implicitly assumes that an over-indebtedness directly leads to bankruptcy. However, firms with negative equity do not inevitably turn bankrupt. Firms might intentionally operate on negative book equity, for example due to tax avoidance. To discriminate between negative book equity firms that are healthy and those that turn bankrupt, our accounting model adds further independent variables. Haowen (2015) finds that non-bankrupt negative book equity firms tend to have a lower leverage ratio, more capital expenditures, pay less tax, have a lower profitability and a smaller size than bankrupt negative book equity firms.

Our accounting model consists of the following accounting-based measures: We again use the probability that book equity turns negative (PNBE) as constructed above. We also use a dummy that takes the value 1 if the book equity is negative, and the value 0, otherwise (NegBkEq). Similarly, we add a dummy that is 1, if the earnings forecast is negative, and that is 0, otherwise (NegEarnFrc). Following Haowen (2014), we use the book leverage ratio (BLR) calculated as the sum of long-term debt and current debt divided by total assets, capital expenditures divided by total assets (CAPXTA), paid taxes (TXT), profitability measured by earnings before interest and taxes over total assets (EBITTA) and size measured by the logarithmic sales (SIZE). EBITTA is also used by Altman (1968), SIZE is also used by Ohlson (1980).

2.4 Market Model

Our over-indebtedness and our accounting model solely require accounting ratios and can thus be applied to a wide range of firms. There is an ongoing debate whether market variables consistently add explanatory power compared to accounting variables. For example, Shumway (2001), Hillegeist et al. (2004), Beaver, McNichols and Rhie (2005) and Campbell et al. (2008) demonstrate that market variables improve accuracy. In contrast, for example Reisz and Perlich (2007) and Agarwal and Taffler (2008) show that accounting-based models have a similar performance. To test these hypotheses, we add two market variables that are taken from Shumway (2001) in our market model: the stock's past excess return (ER) which is the last year's stock's return minus last year's value-weighted index return and the standard deviation of the stock's return (STDER). We further replace the book leverage ratio by the market leverage ratio (MLR) calculated as the sum of long-term debt and current debt divided by the sum of long-term debt, current debt and market equity.

3. Data and Method

3.1 Sample Dataset

We use bankruptcy information taken from Chava and Jarrow (2004) which is updated in Chava (2014) and Alanis, Chava and Kumar (2016).[4] This data includes 2,804 bankruptcy events of firms traded on NYSE, AMEX or NASDAQ and spans from January 1964 to December 2014. Bankruptcy is defined as a petition for filing for Chapter 7 or Chapter 11. We make one-year earnings forecasts three months after the fiscal year end to ensure public availability of the information that we use. That is, we estimate the book equity of 15 months after the fiscal year end. Accordingly, we declare a firm to turn bankrupt during the subsequent fiscal year if the bankruptcy date lies between the last fiscal year end plus three months and the last fiscal year end plus 15 months. Thus, the dependent variable equals one if the firm turns bankrupt during this period and zero otherwise. As we have bankruptcies

[4] We are grateful to Sudheer Chava for kindly providing us with his bankruptcy data.

till the end of 2014, our sample only includes observations with a fiscal year end before or equal to the end of September 2013.

Table 1 summarizes information about our bankruptcy dummies. The first column shows the number of active firms in each year. The second column shows the number of firms that have a bankruptcy dummy equal to one and the third column the corresponding percentage of active firms that turn bankrupt. Figures that aggregate the numbers of 1968 to 2013 are given in the last row. In our final sample the amount of bankruptcy events is reduced to 1490. The overall bankruptcy rate is 0.79% with a strong fluctuation over the years. Chava and Jarrow (2004) use a total of 464 for their sample period from 1963 till 1998 and Shumway (2001) uses 300 bankruptcies between 1962 and 1992. It is apparent that bankruptcies were rare until the late 1970s. The bankruptcy rate rose in the 1980s with a highest rate of 1.20% in 1985. Until mid-1990s the bankruptcy fell to a level similar to the 1970s. From the mid-1990s to the mid-2000s the bankruptcy rate rose dramatically to a peak of 2.47% in 2001. After 2008, bankruptcies were rare, again.

[TABLE 1]

Our initial sample includes all firms in the intersection of the annual Compustat North America fundamentals files and daily and monthly data from CRSP between 1958 and 2013 that are listed on NYSE, AMEX or NASDAQ. We obtain earnings forecasts by using a rolling regression technique. Requiring 10 years of data for the cross-sectional earnings regressions, in a first step, we obtain one-year ahead earnings forecasts for the years 1969 to 2014. The first earnings forecasts are made in 1968 for the year 1969 using accounting data from 1958 to 1967, and the last forecasts are made in 2013 for the year 2014 based on data from 2003 to 2012. Note that these forecasts are look-ahead bias free as we only use information up to the point in time when the forecast is made. We then use the resulting earnings forecasts to predict bankruptcies in a second step. To produce strictly out-of-sample forecasts, we estimate the parameters using only data from 1968 to 2002 and use the coefficients for predicting bankruptcies for the period from 2003 to 2013. Just like in the earnings regression, all our measures of bankruptcy are lagged to ensure that they are observable at the time we use them for estimation. We assume that the accounting and market information are available three months after the fiscal year end. For bankruptcy predictions we use the earnings forecasts that are made three months after the fiscal year end. Accordingly, we make our one-year bankruptcy predictions three months after the fiscal year end.

We delete observations with any missing variable that is used in the earnings forecast model or in any of the bankruptcy prediction models. These include the variable sets of our over-indebtedness model, our accounting model, our market model, Altman (1968), Ohlson (1980) and Shumway (2001) and Bharath and Shumway's (2008) distance-to-default (DD) model.[5] The appendix of this paper describes the variable construction for these bankruptcy prediction models. To reduce the effect of outliers, we winsorize all variables (except for the indicator variables and probabilities) annually at the 1st and 99th percentile.

Table 2 provides summary statistics for all variables described above. Panel A shows the measures used to forecast bankruptcy and Panel B shows the measures used to forecast earnings. We report means, medians, standard deviations and certain percentiles of 189,251 firm years with complete data availability for the period from 1968 to 2013. Most importantly, the overall firm year average of the probability that losses exceed the current book equity (PNBE) is 10.6%. At the same time only 25% of all firm years have a PNBE which is greater than 8.0%. For 1% of all firm years PNBE is greater than 97.6%. Half of firms have a PNBE that is zero. By this, PNBE might be a good proxy for the probability of default, although the overall bankruptcy rate is only 0.79%. For 34.4% or all firm years, the cross-sectional earnings models forecast negative earnings.

[5] DD models use Merton's (1974) option pricing theory. They have been shown to be a good predictor of bankruptcy by e.g., Hillegeist et al. (2004) and Vassalou and Xing (2004). We use the DD version model that Bharath and Shumway (2008) find to perform best. They call this best model 'Model 7'. It comprises their naïve version of Merton's DD probability, the inputs of this probability as individual measures and the ratio of net income and total assets.

[TABLE 2]

Table 3 provides a profile analysis for the bankruptcy measures described above of 189,251 firm years with complete data availability for the period from 1968 to 2013. We report the mean and the standard deviation of those measures for the group of non-bankrupt firm years and the group of bankrupt firm years, respectively. The column labeled 'Diff' shows the mean difference between healthy firm years and bankrupt firm years. We further report results of Welch's t-test on mean equality which is a two-sample test for the hypothesis that two populations have the same mean. Unlike the more common Student's t-test, Welch's t-test does not assume equal variances and equal sample sizes.

[TABLE 3]

The test is significant for all bankruptcy measures, that is the hypothesis that bankrupt and non-bankrupt firm years have the same mean is rejected for all variables. Firms that are about to turn bankrupt differ from non-bankrupt firms in ways that we expect for most of our bankruptcy measures: Bankrupt firms show an average probability of negative book equity (PNBE) of 42.3% which is significantly higher than 10.3% for the non-bankrupt group. Bankrupt firms have more often negative earnings forecasts, a higher leverage ratio, a lower profitability measured by EBITTA, a smaller size, a lower excess return and a higher standard deviation of the return. Unexpectedly, bankrupt firms have higher capital expenditures relative to their total assets and pay less tax. In the next chapter, we investigate if those variables are significant predictors for bankruptcy.

3.2 Logistic regressions

Following Shumway (2001), Chava and Jarrow (2004) and Campbell et al. (2008), we estimate the probability that a firm turns bankrupt in the subsequent fiscal year by a logistic regression. Thus, this probability follows a logistic distribution with parameters (α, β) and is equal to

$$P_t(y_{it} = 1) = \frac{1}{1+\exp(-\alpha-\beta x_{i,t})}, \quad (7)$$

where y_{it} is the bankruptcy dummy that equals one if the firm fails in the following fiscal year and zero, otherwise, and $x_{i,t}$ is the vector of explanatory variables that are known at year t. The higher $\alpha + \beta x_{i,t}$, the higher is the estimated probability of bankruptcy. To produce strictly out-of-sample forecasts, we estimate the parameters using only data from 1968 to 2002 and use the coefficients for predicting bankruptcies for the period from 2003 to 2013.

Static models (for example Altman (1968), Ohlson (1980)) use one single observation per firm and thus make a sample selection bias. In contrast, we use all available firm observations to estimate the logistic regression. Hence, our estimation technique exploits more information and eliminates the sample selection bias. Note that applying such a technique to the static models of Altman (1968) and Ohlson (1980) already improves their performance as compared to using the estimation techniques suggested originally.

4. Results of Empirical Analysis

4.1 Properties of variables related to equity or earnings before bankruptcy

For our structural model we say that an over-indebtedness is a good indication of bankruptcy. That is, we assume that the book equity of bankrupt firms diminishes in the years before bankruptcy and is finally depleted by losses in the year of bankruptcy. We analyze the evolution of variables that are related to book equity or earnings of bankrupt firms. Table 4 reports means and medians of these variables in the last five years before bankruptcy and in the year of bankruptcy. For example, year -5 denotes the fifth year before bankruptcy. In this analysis, we only include firms that appear in each year of our analysis. That is, we compare the same set of firms over time.

On average, PNBE is monotonously rising in the years before bankruptcy. In year -5, the mean of PNBE is 10.4% which is close to the average PNBE of non-bankrupt firms (10.3%). In year -1 it is 21.9% and in the year of bankruptcy it makes a big jump to 42.0%. Year -2 is the first year in which 50% of firm years have a PNBE which is higher than zero. The median rises to 2.2% in year -1, before it finds its peak of 36.5% in the year of bankruptcy. The average book equity consistently declines from year -3 on. It experiences the most severe fall from year -1 with 175.299 to the year of bankruptcy with 84.151. The median book equity has similar properties and has its most significant fall from year -1 with 24.962 to year 0 with 8.057. Accordingly, the ratio of firms with negative book equity rises from 11.4% in year -1 to 29.1% in year 0. For the earnings there is a similar pattern. The average earnings have a downward trend from year -4 on and experience the most significant fall from -46.486 in year -1 to -101.702 in year 0. The median earnings are monotonously falling from 0.729 in year -5 to -15.150 in year 0. Again, the ratio of firm years with negative earnings rises from 39.8% in year -5 to 89.3% the year of bankruptcy. Importantly, the average losses in the year of bankruptcy (101.702) deplete the average book equity in the year of bankruptcy (84.151). Accordingly, the median losses in year 0 (15.150) exceed the median book equity in year 0 (8.057).

[TABLE 4]

We find evidence for our overall assumption that book equity and earnings diminish in the years before bankruptcy. Especially in the year directly before bankruptcy these variables experience a dramatic fall. We further find evidence for the fact that losses exceed book equity in the year of bankruptcy. Looking only at bare median and mean values of book equity and earnings, we neglect the volatility of earnings. PNBE which incorporates this volatility shows good properties in the years before bankruptcy as well.

4.2 Profile analysis of bankrupt and non-bankrupt negative book equity firms

Our over-indebtedness model only consists of the probability of negative book equity. However, over-indebted firms do not inevitably turn bankrupt. Thus, our accounting model adds variables that are supposed to differentiate between bankrupt and non-bankrupt negative book equity firms. Table 5 provides a profile analysis for these bankruptcy measures of the 6,166 firm years with negative book equity for the period from 1968 to 2013. Among these negative book equity firm years there are 5,752 non-bankrupt and 414 bankrupt years. We report the mean and the standard deviation of those measures for the group of non-bankrupt negative book equity firm years and the group of bankrupt negative book equity firm years, respectively. The column labeled 'Diff' reports the mean difference between healthy firm years and bankrupt firm years. We again report results of Welch's t-test on mean equality.

Welch's t-test is significant for all bankruptcy measures except for taxes (TXT), that is all variables except for TXT have a significantly different mean for non-bankrupt and bankrupt firm years with negative book equity. For most of our measures, these differences are in line with the profile analysis of the full sample: Bankrupt firms with negative book equity have an average PNBE of 88.5% which is significantly higher than 79.3% for the non-bankrupt group. Since PNBE focuses on a decreasing book equity by construction, these probability values, especially for non-bankrupt firms, are higher compared to the full sample. There are 96.9% of bankrupt firm years with a negative earnings forecasts and only 81.1% of non-bankrupt firm years. The market and the book leverage ratios of non-bankrupt firm years with negative book equity are lower than those of the bankrupt group which is in line with the profile analysis of the full sample. This also supports Haowen's (2014) results. The mean of CAPXTA for bankrupt firm years is higher, which is consistent with the full sample, but not with Haowen's results. The mean difference for TXT is nonsignificant, which contradicts both Haowen's and the full sample results. Interestingly, the profitability measured by EBITTA and the SIZE are higher for bankrupt firms. This contradicts the results for the full sample but is in line with Haowen's findings.

[TABLE 5]

We find variables that have different means for bankrupt and non-bankrupt firms with negative book equity, respectively. They might complement the probability of over-indebtedness in bankruptcy prediction models by discriminating between bankrupt and non-bankrupt over-indebted firms. The results for EBITTA and SIZE for negative book equity firms differ from the results for the full sample. In the next section, we investigate in which way these two measures are incorporated in our prediction model.

4.3 Logistic regression results

Table 6 reports the estimation results of several logistic regressions. It shows the results of our over-indebtedness model, our accounting model and our market model. Given are the parameter estimates, their standard deviation, their significance and the likelihood ratio test for each model. Our over-indebtedness model is univariate and only comprises PNBE. Our accounting model is multivariate by adding further bankruptcy measures to discriminate between bankrupt and non-bankrupt negative book equity firms. Our market model adds two common market-based measures and replaces book leverage ratio by market leverage ratio.

The results of our over-indebtedness model confirm that PNBE is an extremely significant bankruptcy predictor by its own. In our accounting model all variables are statistically significant as well. Thus, we can conclude that PNBE is not a sufficient measure of bankruptcy probability, though. The measures that we add in our market model are significant as well. This gives evidence for the fact that these market-based variables have additional explanatory power. The signs of most coefficients in the accounting and the market model are consistent with economic intuition and our profile analyses. Firms with a higher PNBE are more likely to fail and firms with negative earnings forecasts (NegEarnFrc) are more likely to fail. The higher the book leverage ratio (BLR), the higher the market leverage ratio (MLR), the higher the capital expenditures (CAPXTA), the larger the size (SIZE) and the higher the volatility of the return (STDER), the higher is the estimated probability of bankruptcy. The lower the tax (TXT), the lower the profitability (EBITTA) and the lower the excess return (ER), the higher is the estimated probability of bankruptcy.

[TABLE 6]

For the full sample and for negative book equity firms we have found variables with different means for bankrupt and non-bankrupt firms. All those measures are significant in our bankruptcy models. The coefficients of NegEarnFrc, MLR, BLR and CAPXTA are consistent with both our profile analyses. For the return-based variables, ER and STDER, and for TXT the coefficients are in line with our profile analysis for full sample. For EBITTA, the results validate our profile analysis for the full sample rather than the profile analysis for negative book equity firms. Interestingly, the parameter sign of SIZE is consistent with the profile analysis for negative book equity firms that says that bankrupt firms are larger than non-bankrupt firms. The sign contradicts the profile analysis of our full sample that find a smaller size for bankrupt firms. By this, SIZE helps to discriminate between bankrupt and non-bankrupt firms that are already over-indebted.

4.4 Out-of-sample results

Table 7 assesses the out-of-sample predictive ability of different variable sets. To create this table, we rank firms into deciles based on their fitted bankruptcy probability values for every year of our validation sample (2002 to 2013). That is, the firms that will most likely default in the subsequent year are sorted into the first decile and the firms with the lowest estimated default probabilities are assigned into the tenth decile. We report the percentages of bankrupt firms that fall into each of the ten probability deciles. A model is accurate if it estimates a high default probability for bankrupt firm years and thus assigns many bankrupt firms into low deciles.

We compare our over-indebtedness, our accounting model and our market model to leading and well-known alternatives: We use the variable sets of Altman (1968), Ohlson (1980) and of Shumway (2001). Furthermore, we estimate the best model version of Merton's DD approach as found by Bharath and Shumway (2008). Our over-indebtedness model is univariate and only comprises PNBE. Thus, there is no difference if you rank the firms by the probability estimated by the logistic regression or directly by PNBE. Differences in the out-of-sample results compared to other studies like for example Shumway (2001) and Bharath and Shumway (2008) are due to our augmented database.

[TABLE 7]

Our accounting model classifies 63.01% of all bankrupt firm years into the highest default probability decile (decile one). That is, a bank can exclude 63.01% of all bankruptcies if it does not lend money to the 10% of firms with the highest expected default measures. By this, it significantly outperforms the models by Altman (51.03%), Ohlson (55.82%) that use accounting information as well. Even the univariate over-indebtedness model identifies 56.85% of bankruptcies correctly (in the first decile) and thus outperforms Altman and Ohlson. For the top two deciles (in aggregate) the correct predictions are 76.71% in our accounting model and 67.81% in our over-indebtedness model. That is, if a bank does not lend money to the 20% of firms with the highest default probability forecasted by the accounting model it excludes 76.71% of all bankruptcies. Again, our accounting model performs better than Altman (60.96%), Ohlson (74.66%). Our over-indebtedness model performs better than Altman. Given that our over-indebtedness model has one variable, it performs surprisingly well.

Note that our accounting model solely requires accounting information. Major improvements in the performance of bankruptcy prediction models have been achieved by the inclusion of market information (e.g., Shumway (2001), Hillegeist et al. (2004)). This performance boost, however, comes with a cost. These models are limited to firms with access to the capital market. Importantly, our purely accounting model achieves better out-of-sample performances than leading alternatives that solely use accounting information. Thus, we improve performance without limiting the scope of application. We can still provide bankruptcy predictions for a wide range of firms.

If we allow market information to be added into our model, we further significantly improve performance. Our market model classifies 75.34% of all bankrupt firm years into the highest default probability decile (decile one). That is, a bank can exclude 75.34% of all bankruptcies if it does not lend money to the 10% of firms with the highest expected default measures. Importantly, our market model significantly outperforms all leading alternatives including those that use market information. Bharath and Shumway (2008) only classify 66.44% of bankrupt firms into the first decile and Shumway (2001) assigns 72.6% correctly. For the top two deciles (in aggregate), the correct predictions made by our market models are 89.04% and thus higher than for Shumway (84.93%) and Bharath and Shumway (82.19%). Remarkably, Shumway has a better out-of-sample performance than Bharath and Shumway. This stands in contrast to the results by Campbell, Hilscher and Szilagyi (2011). Furthermore, our market model performs significantly better than our accounting model. Thus, we support for example Shumway (2001), Hillegeist et al. (2004) and Campbell et al. (2008) who demonstrate that market variables add explanatory power.

4.5 Functional Form of PNBE

We use a non-linear functional form of three inputs for calculating the probability of negative book equity. We analyze if this rigid functional form is important for predicting bankruptcy. For this, we construct two models. Model 1 uses the same inputs as PNBE, but does not squeeze these variables into one variable: the current book equity, the earnings forecast and the inverse of its standard deviation. Model 2 comprises all covariates of model 1, but adds the functional form of these variables, that is PNBE. Table 8 reports the results of these two models. Panel A contains the estimation results of the logistic regressions and Panel B assesses their out-of-sample predictive ability by reporting the goodness-of-fit deciles.

[TABLE 8]

In Model 2, we see that PNBE is a significant predictor of bankruptcy even if we add all covariates that are used to construct PNBE into the logistic regression. This gives evidence for the fact that the functional form of PNBE that we use is meaningful for predicting bankruptcies. Vice versa, we see that the inverse of the earnings forecast volatility is significant as an individual input. This implies that PNBE is not a sufficient measure of bankruptcy prediction. Looking at the out-of-sample assessment, we see that model 2 classifies 56.85% of all bankruptcies into the highest probability decile and thus outperforms model 1 (47.95%). This gives further evidence that the functional form of PNBE is an important construct.

5. Conclusion

We develop a new structural framework to predict bankruptcies, focusing on the probability that future losses might deplete the firm's current book equity. To our best knowledge, this study is the first to incorporate cross-sectional earnings forecasts as proposed by Hou et al. (2012) into bankruptcy prediction models. By this, we propose a new way of calculating the probability that book equity turns negative (PNBE) which does not require the option pricing framework (e.g., Hillegeist et al. (2004), Bharath and Shumway (2008)).

We propose two bankruptcy prediction models. First, our accounting model outperforms alternative models that use accounting information. Until now, major improvements have been achieved by the inclusion of market measures. This performance boost, however, comes along with a strict limitation to firms with access to the capital market. We demonstrate that these performance improvements can be achieved without limiting the scope of application. Second, our market model outperforms all leading alternatives of corporate risk failure, including those that use market measures. If you want to predict a bankruptcy of a non-public firm, our accounting model is the method of choice. If you want to predict a bankruptcy of a public firm, our market model produces best results.

The described models focus on one-year predictions. To create multi-period bankruptcy prediction models, one must simply use multi-period earnings forecast models which are described in Hou et al. (2012). Additionally, one can use analysts' earnings forecasts instead of mechanical forecasts to model the changes of book equity. Further research can aim at further grasping the imperfect relation between firms with a negative book equity and bankrupt firms. On the one hand, one could add further variables that help to discriminate between bankrupt and non-bankrupt firms with negative book equity. On the other hand, one could re-define which components are to be included in the book equity. For example, one could delete those components that belong to the typical definition of book equity but do not have an influence on bankruptcy.

Appendix: Construction of variables of earlier bankruptcy prediction models

In this appendix, we discuss the construction of variables that are used by Altman (1968), Ohlson (1980) and Shumway (2001) and the best model of Bharath and Shumway (2008). The variables of our models are constructed in the second chapter.

Altman (1968) ends up with the Z-Score, a linear weighted sum of five ratios which best discriminates between failing and surviving firms on his sample:

$$Z = \beta_0 + \beta_1 \cdot \text{WCTA} + \beta_2 \cdot \text{RETA} + \beta_3 \cdot \text{EBITTA} + \beta_4 \cdot \text{METL} + \beta_5 \cdot \text{STA}, \quad \text{where}$$

- WCTA = Working capital / Total assets,
- RETA = Retained Earnings / Total assets,
- EBITTA = EBIT / Total assets,
- METL = Market value equity / Book value of total debt,
- STA = Sales / Total assets and
- Z = Z-Score (overall index).

X1 is a proxy for a firm's liquidity and X2 and X3 measure different aspects of profitability. X4 is a widely used measure of leverage and X5 describes the firm's efficiency to use assets in generating sales. The Z-score characterizes the financial strength of a firm by aggregating the above five accounting ratios into one figure via the estimated coefficients $\beta_1, \dots, \beta_5$.

Ohlson (1980) finds nine variables to be significant and defines his O-score model as:

$$\begin{aligned} O = \; & \beta_0 + \beta_1 \cdot SIZE + \beta_2 \cdot TLTA + \beta_3 \cdot WCTA \\ & + \beta_4 \cdot CLCA + \beta_5 \cdot OENEG + \beta_6 \cdot NITA \\ & + \beta_7 \cdot FUTL + \beta_8 \cdot INTWO + \beta_9 \cdot CHIN, \end{aligned}$$

where

- SIZE = log (Total assets),
- TLTA = Total liabilities over total assets,
- WCTA = Working capital over total assets,
- CLCA = Current liabilities over current assets,
- OENEG = Dummy, if total liabilities exceed total assets,
- NITA = Net income over total assets,
- FUTL[6] = Funds provided by operations over total liabilities,

[6] Funds provided by operations are not reported anymore. We use an approximation by summing pretax income and depreciations and amortization.

- INTWO = Dummy, if net income was negative for the last two years,
- CHIN = Change in net income and
- O = O-Score (overall index).

WCTA and CLCA measure liquidity. NITA, FUTL, INTWO and CHIN capture different aspects of profitability. TLTA and OENEG describe the capital structure. SIZE is a measure of firm size.

In addition to selected financial ratios already used by Ohlson, Shumway (2001) adds two market variables which are the excess return and its standard deviation:

$$S = \beta_0 + \beta_1 \cdot SIZE + \beta_2 \cdot TLTA + \beta_3 \cdot NITA + \beta_4 \cdot ER + \beta_5 \cdot STDER,$$

where

- RSIZE = logarithm of market equity divided by value-weighted market equity of index,
- TLTA = Total liabilities over total assets,
- NITA = Net income over total assets,
- ER = Excess return calculated as the difference of last year's return and last year's value-weighted index return
- STDER = Standard deviation of return and
- S = S-Score (overall index).

TLTA measures solvency and describes the capital structure. Profitability is captured by NITA. ER measures the profit of an investment, where STDER captures the variability of the excess return. RSIZE is a measure of the firm's size. Again, we re-estimate Shumway's model on our sample.

Bharath and Shumway (2008) expand on the distance-to-default models that e.g., Vassalou and Xing (2004) and Hillegeist et al. (2004) construct by applying Merton's (1974) option pricing theory. Merton's probability of bankruptcy is calculated as

$$PD - Merton = N\left(-\left(\frac{ln(V/F) + (\mu - 0.5{\sigma_V}^2)}{\sigma_V}\right)\right),$$

where V is the market value of a firm's assets, σ_V its standard deviation, μ is the expected return on assets, F is the market value of the firm's debt, and N(·) is the cumulative standard normal distribution function. Vassalou and Xing (2004) compute V and σ_V numerically by applying an iterative procedure. Bharath and Shumway, however, propose a naïve approach. They approximate the market value of debt by the book value of debt and thus calculate F as debt in current liabilities plus one half of long-term debt. Furthermore, the volatility of a firm's debt is approximated by

$$\sigma_F = 0.05 + 0.25 \cdot \sigma_E,$$

where σ_E is the volatility of market equity. Market equity is denoted by E and is calculated as the product of the share price at the end of the month and the number of shares outstanding. An approximation of the volatility of the firm's assets is then given by

$$\sigma_V = \frac{E}{E+F}\sigma_E + \frac{F}{E+F}\sigma_F.$$

The expected return on assets μ is approximated by last year's return on assets. And the market value of assets is approximated by the sum of the market value of equity and the book value of debt.

Besides PD-Merton, the best model in Bharath and Shumway consists of the following covariates: the logarithm of market equity E, the logarithm of the book value of debt F, the inverse of the volatility of market equity, the excess return calculated as the difference of last year's return and the risk-free rate measured by the 1-year Treasury Bill from the Board of Governors of the Federal Reserve system, and the ratio of net income and total assets (NITA).

References

Alanis, E., Chava, S., Kumar, P., 2016, Shareholder Bargaining Power, Debt Overhang and Investment, Working Paper, Georgia Institute of Technology.

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

Altman, E. I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23(4), 589-609.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., Suvas, A., 2017, Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model, *Journal of International Financial Management and Accounting*, 28, 131-171.

Beaver, W. H., McNichols, M. F., Rhie, J.-W., 2005, Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy, *Review of Accounting Studies*, 10, 93-122.

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

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

Campbell, J. Y., Hilscher, J., Szilagyi, J., 2011, Predicting Financial Distress and the Performance of Distressed Stocks, *Journal of Investment Management*, 9, 14-34.

Charitou, A., Dionysiou, D., Lambertides, N., Trigeorgis, L., 2013, Alternative bankruptcy prediction models using option-pricing theory, *Journal of Banking & Finance*, 37, 2329-2341.

Chava, S. and Jarrow, R. A., 2004, Bankruptcy Prediction with Industry Effects, *Review of Finance,* 8, 537-569.

Chava, S., 2014, Environmental Externalities and Cost of Capital, *Management Science,* 60(9), 2223-2247.

Feller, W., 1968, An introduction to probability theory and its applications, Wiley, New York.

Haowen, L., 2015, Aggressive Capital Structure and the Myth of Negative Book Equity Firms, Unpublished conference paper, Southwestern Finance Association, March 2015, Houston, Texas.

Hess, D., Meuter, M., Kaul, A., 2017, The Performance of Mechanical Earnings Forecasts, Working Paper.

Hillegeist, S. A., Keating, E. K., Cram, D. P., Lundstedt, K. G., 2004, Assessing the probability of bankruptcy. *Review of Accounting Studies*, *9*(1), 5-34.

Hou, K., Van Dijk, M.A., Zhang, Y., 2012, The Implied Cost of Capital: A new approach, *Journal of Accounting and Economics*, 53, 504-526.

Li, K., Mohanram, P., 2014, Evaluating cross-sectional forecasting models for implied cost of capital, *Review of Accounting Studies*, 19(3), 1152-1185.

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

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

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

Santomero, A. M., Vinso, J. D., 1977, Estimating the probability of failure for commercial banks and the banking system, *Journal of Banking & Finance*, 1(2), 185-205.

Scott, J., 1976, A theory of optimal capital structure, *Bell Journal of Economics*, 7(1): 33-54.

Shumway, T., 2001, Forecasting bankruptcy more accurately: A simple hazard model, *The Journal of Business*, 74(1), 101-124.

Vassalou, M., Xing, Y., 2004, Default Risk in Equity Returns, *The Journal of Finance*, 59: 831-868.

Vinso, J. D., 1979, A determination of the risk of ruin, *The Journal of Financial and Quantitative Analysis*, 14(1), 77-100.

Wilcox, J. W., 1973, A prediction of business failure using accounting data, *Journal of Accounting Research*, 11, 163-179.

Wilcox, J. W., 1976, The gambler's ruin approach to business risk, *Sloan Management Review*, fall, 33-46.

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

Table 1 Number of Bankruptcies per Year

Year	Active Firms	Bankruptcies	(%)	Year	Active Firms	Bankruptcies	(%)
1968	1210	0	0.00	1991	4447	41	0.92
1969	1427	1	0.07	1992	4592	33	0.72
1970	1654	3	0.18	1993	4799	32	0.67
1971	1867	2	0.11	1994	5257	22	0.42
1972	1961	5	0.25	1995	5505	27	0.49
1973	2999	10	0.33	1996	5791	28	0.48
1974	3284	15	0.46	1997	6227	35	0.56
1975	3290	8	0.24	1998	6293	58	0.92
1976	3286	11	0.33	1999	5990	65	1.09
1977	3245	11	0.34	2000	5711	112	1.96
1978	3227	13	0.40	2001	5501	136	2.47
1979	3384	15	0.44	2002	5071	93	1.83
1980	3502	22	0.63	2003	4788	64	1.34
1981	3622	24	0.66	2004	4515	22	0.49
1982	3846	36	0.94	2005	4502	26	0.58
1983	4001	24	0.60	2006	4422	16	0.36
1984	4267	44	1.03	2007	4326	29	0.67
1985	4330	52	1.20	2008	4225	38	0.90
1986	4383	40	0.91	2009	3994	21	0.53
1987	4616	39	0.84	2010	3829	16	0.42
1988	4715	53	1.12	2011	3723	23	0.62
1989	4519	41	0.91	2012	3642	20	0.55
1990	4399	47	1.07	2013	3577	17	0.48
				Total	187761	1490	0.79

This table lists the number of active firms, the number of bankruptcy dummies and the percentage of bankruptcy dummies among active firms for every year of our sample period of 1968 to 2013. A bankruptcy dummy takes the value of one, if a firm turns bankrupt in the period of 3 months after the fiscal year end and 15 months after the fiscal year end.

Table 2 Summary statistics (N=189,251)

Panel A: Variables in Bankruptcy Prediction Models

Variable	Source	Mean	STD	1%	25%	Median	75%	99%
$PNBE_t$	*Insol / Acc / Mark*	0.106	0.210	0.000	0.000	0.000	0.080	0.976
$Neg\ EarnFrc_t$	*Acc / Mark*	0.344	0.475	0.000	0.000	0.000	1.000	1.000
BLR_t	*Acc*	0.332	0.360	0.000	0.069	0.295	0.499	1.545
$CAPXTA_t$	*Acc / Mark*	0.070	0.075	0.000	0.023	0.048	0.089	0.387
TXTt	*Acc / Mark*	34.469	145.888	-26.662	0.000	1.452	12.045	708.000
$EBITTA_t$	*Acc / Mark*	0.005	0.342	-1.274	-0.004	0.073	0.128	0.353
$Size_t$	*Acc / Mark*	4.828	2.225	0.368	3.230	4.638	6.298	10.352
MLR_t	*Mark*	0.257	0.247	0.000	0.032	0.190	0.422	0.904
ER_t	*Mark*	0.019	0.637	-0.955	-0.343	-0.068	0.218	2.348
$STDER_t$	*Mark*	0.121	0.082	0.000	0.066	0.100	0.151	0.427

Panel B: Variables in Earnings Forecast Models

Variable	Mean	STD	1%	25%	Median	75%	99%
E_t	37.000	274.703	-303.000	-1.654	1.547	13.814	995.000
$Neg\ E_t$	0.342	0.474	0.000	0.000	0.000	1.000	1.000
$Neg\ ExE_t$	-14.448	90.286	-303.000	-1.654	0.000	0.000	0.000
$BkEq_t$	579.256	2444.080	-38.613	10.943	47.933	233.478	9709.550
AC_t	-78.489	375.495	-1572.000	-23.332	-2.775	0.178	71.000

This table reports the summary statistics of the following forecast variables ($ millions for all values except dummy variables and probability values). For more details see the data construction: PNBE is the probability that losses deplete current book equity, NegBE is a dummy for a negative book equity, NegEarnFrc is a dummy for a negative earnings forecast, BLR is the book leverage ratio, CAPXTA are capital expenditures over total assets, TXT are taxes, EBITTA are earnings before interest and taxes over total assets, size is the logarithmic sales, MLR is the market leverage ratio, ER is the excess return, STDER is the standard deviation of the return, E is the change of retained earnings, NegE is a dummy for negative earnings, NegExE is an interaction term of the dummy for negative earnings and earnings, BkEq is the book equity and AC are accruals. Panel A shows those variables used to forecast bankruptcy and Panel B shows those variables used to forecast earnings. Each observation represents one particular firm in one particular year. The reported values are the time series averages of yearly cross sectional means, medians, standard deviations and respective percentiles. All variables (except indicator variables and probability values) are winsorized annually at the 1st and 99th percentile. The column labeled 'Source' indicates in which model the variable is used, with Insol meaning our insolvency model, Acc meaning our accounting model and Mark meaning our market model. The sample period is from 1968 to 2013. Summary statistics are reported for those observations for which all variables of any model are available.

Table 3 Profile Analysis and t-test for mean equality (N=189,251)

		Non-Bankrupt Firm Years		Bankrupt Firm Years				
Variable	Source	Mean	STD	Mean	STD	Diff	t-stat	
$PNBE_t$	*Insol / Acc / Mark*	0.103	0.207	0.423	0.371	-0.319	-33.19	***
$Neg\ EarnFrc_t$	*Acc / Mark*	0.340	0.474	0.878	0.328	-0.538	-62.84	***
BLR_t	*Acc*	0.328	0.353	0.789	0.729	-0.461	-24.38	***
$CAPXTA_t$	*Acc / Mark*	0.070	0.074	0.082	0.105	-0.012	-4.43	***
TXTt	*Acc / Mark*	34.707	146.336	4.389	62.279	30.318	18.39	***
$EBITTA_t$	*Acc / Mark*	0.007	0.338	-0.244	0.626	0.251	15.46	***
$Size_t$	*Acc / Mark*	4.831	2.227	4.453	1.962	0.378	7.40	***
MLR_t	*Mark*	0.254	0.245	0.581	0.309	-0.327	-40.73	***
ER_t	*Mark*	0.022	0.637	-0.448	0.517	0.470	34.90	***
$STDER_t$	*Mark*	0.120	0.082	0.195	0.106	-0.075	-27.18	***
N		187,761		1,490				

This table reports the summary statistics of the following forecast variables ($ millions for all values except dummy variables and probability values) for the bankrupt and the non-bankrupt group, respectively. For more details see the data construction: PNBE is the probability that losses deplete current book equity, NegBE is a dummy for a negative book equity, NegEarnFrc is a dummy for a negative earnings forecast, BLR is the book leverage ratio, CAPXTA are capital expenditures over total assets, TXT are taxes, EBITTA are earnings before interest and taxes over total assets, size is the logarithmic sales, MLR is the market leverage ratio, ER is the excess return and STDER is the standard deviation of the return. Each observation represents one particular firm in one particular year. The reported values are the time series averages of yearly cross sectional means and standard deviations. All variables (except indicator variables and probability values) are winsorized annually at the 1st and 99th percentile. The column labeled 'Source' indicates in which model the variable is used, with Insol meaning our insolvency model, Acc meaning our accounting model and Mark meaning our market model. The sample period is from 1968 to 2013. Quantities are reported for those observations for which all variables are available. The column Diff shows the difference of the means of the non-bankrupt group and the bankrupt group. The t-statistic of Welch's t-test on mean equality are reported where an independent two-sample and inequal variances are assumed. *** denotes significance at the 1% level.

Table 4: Variables related to equity and earnings in the years before failure (N=718)

Panel A: Mean

Variable	-5	-4	-3	-2	-1	0
$PNBE_t$	0.104	0.116	0.123	0.150	0.219	0.420
$BkEq_t$	206.520	219.917	228.495	202.585	175.299	84.151
$Neg\ BE_t$	0.039	0.040	0.047	0.071	0.114	0.291
$Earn_t$	-1.763	-0.777	-19.819	-23.255	-46.486	-101.702
$Neg\ E_t$	0.398	0.446	0.481	0.549	0.726	0.893

Panel B: Median

Variable	-5	-4	-3	-2	-1	0
$PNBE_t$	0.000	0.000	0.000	0.001	0.022	0.365
$BkEq_t$	35.165	27.162	36.276	31.325	24.962	8.057
$Neg\ BE_t$	0.000	0.000	0.000	0.000	0.000	0.000
$Earn_t$	0.729	0.299	0.176	-0.676	-5.278	-15.150
$Neg\ E_t$	0.000	0.000	0.000	1.000	1.000	1.000

This table reports summary statistics of the following variables related to equity and earnings ($ millions for all values except dummy variables and probability values) in the last five years before bankruptcy and the year of bankruptcy. For more details see the data construction: PNBE is the probability that losses deplete current book equity, BkEq is the book equity, NegBE is a dummy for a negative book equity, Earn is the change of retained earnings and NegE is a dummy for a negative earnings. Each observation represents one particular firm in one particular year of a bankrupt firm. Panel A shows the time series averages of yearly cross sectional means and Panel B shows the corresponding medians. All variables (except indicator variables and probability values) are winsorized annually at the 1st and 99th percentile. The sample period is from 1968 to 2013. Quantities are reported for those observations for which all variables are available.

Table 5 Profile Analysis for negative book equity firms (N=6,166)

		Non-Bankrupt Negative Book Equity Firm Years		Bankrupt Negative Book Equity Firm Years				
Variable	Source	Mean	STD	Mean	STD	Diff	t-stat	
$PNBE_t$	*Insol / Acc / Mark*	0.793	0.189	0.885	0.135	-0.092	-12.97	***
$Neg\ EarnFrc_t$	*Acc / Mark*	0.818	0.386	0.969	0.174	-0.151	-15.23	***
BLR_t	*Acc*	1.147	1.213	1.392	1.096	-0.244	-4.37	***
$CAPXTA_t$	*Acc / Mark*	0.064	0.086	0.083	0.106	-0.019	-3.61	***
TXTt	*Acc / Mark*	12.933	83.544	12.055	98.475	0.877	0.18	
$EBITTA_t$	*Acc / Mark*	-0.486	1.182	-0.369	0.856	-0.117	-2.61	***
$Size_t$	*Acc / Mark*	3.964	2.588	4.512	2.077	-0.548	-5.11	***
MLR_t	*Mark*	0.505	0.308	0.731	0.261	-0.226	-16.85	***
N		5,752		414				

This table reports the summary statistics of the following forecast variables ($ millions for all values except dummy variables and probability values) for firm years with negative book equity. Statistics are given for non-bankrupt and non-bankrupt negative book equity firms, respectively. For more details see the data construction: PNBE is the probability that losses deplete current book equity, NegBE is a dummy for a negative book equity, NegEarnFrc is a dummy for a negative earnings forecast, BLR is the book leverage ratio, CAPXTA are capital expenditures over total assets, TXT are taxes, EBITTA are earnings before interest and taxes over total assets, size is the logarithmic sales and MLR is the market leverage ratio. Each observation represents one particular firm in one particular year. The reported values are the time series averages of yearly cross sectional means and standard deviations. All variables (except indicator variables and probability values) are winsorized annually at the 1st and 99th percentile. The column labeled 'Source' indicates in which model the variable is used, with Insol meaning our insolvency model, Acc meaning our accounting model and Mark meaning our market model. The sample period is from 1968 to 2013. Quantities are reported for those observations for which all variables are available. The column Diff shows the difference of the means of the non-bankrupt group and the bankrupt group. The t-statistic of Welch's t-test on mean equality are reported where an independent two-sample and inequal variances are assumed. *** denotes significance at the 1% level.

Table 6 Parameter estimates of the bankruptcy models

Variable	Over-indebtedness Model		Accounting Model		Market Model	
Constant	-5.394	***	-7.609	***	-8.381	***
	0.043		0.113		0.126	
$PNBE_t$	3.508	***	1.673	***	1.167	***
	0.082		0.113		0.100	
$Neg\ EarnFrc_t$			1.984	***	1.553	***
			0.095		0.098	
BLR_t			0.608	***		
			0.049			
$CAPXTA_t$			1.573	***	2.114	***
			0.290		0.293	
TXTt			-0.001	***	-0.007	***
			0.002		0.002	
$EBITTA_t$			-0.200	***	-0.226	***
			0.059		0.055	
$Size_t$			0.206	***	0.127	***
			0.015		0.017	
MLR_t					2.688	***
					0.117	
ER_t					-0.607	***
					0.070	
$STDER_t$					3.118	***
					0.292	
N	143,416		143,416		143,416	
LRT	1,474.25	***	2,515.31	***	3,258.70	***

This table reports the results of multiperiod logistic regressions of the bankruptcy indicator for our over-indebtedness model, our accounting model and our market model. Parameter estimates for all variables in each model are reported along with their standard errors below. The logistic model is estimated for the sample period of 1968-2002 with 143,416 observations and 1,198 bankruptcies. The chi-square of the likelihood ratio test for each model is reported in the row labeled LRT. *** denotes significance at the 1% level.

Table 7 Goodness-of-fit Deciles

Decile	Over-indebtedness	Accounting	Market	Altman (1968)	Ohlson (1980)	Shumway (2001)	Bharath and Shumway (2008)
1	56.85	63.01	75.34	51.03	55.82	72.6	66.44
2	10.96	13.7	13.7	9.93	18.84	12.33	15.75
3	9.25	8.22	2.05	11.3	11.99	4.79	6.16
4	5.82	5.48	3.08	5.14	4.45	4.11	3.42
5	4.45	3.77	1.37	3.42	3.08	0.68	1.71
6	3.08	2.05	1.03	3.08	0.68	1.37	1.71
7	3.08	1.71	1.37	4.79	1.03	0.34	1.71
8	2.74	0	0	2.05	1.71	2.05	1.37
9	3.42	0.34	0.34	4.45	1.03	1.03	1.37
10	0.34	1.71	1.71	4.79	1.37	0.68	0.34

This table compares the out-of-sample accuracy of various bankruptcy prediction models. Parameter estimates from the training sample (1968 to 2002) are used to predict bankruptcies for the validation period of 2003 to 2013. This validation sample includes 45,835 firm years and 292 bankruptcies. All models are estimated with a multiperiod logistic regression. For every year we rank firms into deciles based on their fitted bankruptcy probability values, where the firms with the highest values fall into the first decile. We report the percentage of bankrupt firms that are classified into each probability deciles.

Table 7 Goodness-of-fit Deciles

Decile	Insolvency	Accounting	Market	Altman (1968)	Ohlson (1980)	Shumway (2001)	Bharath and Shumway (2008)
1	56.85	63.01	75.34	51.03	55.82	72.6	66.44
2	10.96	13.7	13.7	9.93	18.84	12.33	15.75
3	9.25	8.22	2.05	11.3	11.99	4.79	6.16
4	5.82	5.48	3.08	5.14	4.45	4.11	3.42
5	4.45	3.77	1.37	3.42	3.08	0.68	1.71
6	3.08	2.05	1.03	3.08	0.68	1.37	1.71
7	3.08	1.71	1.37	4.79	1.03	0.34	1.71
8	2.74	0	0	2.05	1.71	2.05	1.37
9	3.42	0.34	0.34	4.45	1.03	1.03	1.37
10	0.34	1.71	1.71	4.79	1.37	0.68	0.34

This table compares the out-of-sample accuracy of various bankruptcy prediction models. Parameter estimates from the training sample (1968 to 2002) are used to predict bankruptcies for the validation period of 2003 to 2013. This validation sample includes 45.835 firm years and 292 bankruptcies. All models are estimated with a multiperiod logistic regression. For every year we rank firms into deciles based on their fitted bankruptcy probability values, where the firms with the highest values fall into the first decile. We report the percentage of bankrupt firms that are classified into each probability deciles.

Table 8 Importance of functional form of PD

Panel A: Coefficient Estimates

Variable	Model 1		Model 2	
Constant	1.369	***	-4.297	***
	0.249		0.310	
PNBE $_t$			3.357	***
			0.089	
BkEq $_t$	-0.001	***	-0.0001	
	0.000		0.0000	
EarnFrc $_t$	-0.00001		0.0000	
	0.000		0.0000	
1/sigma(EarnFrc)	-4.577	***	-1.664	***
	0.351		0.089	
N	143,416		143,416	
LRT	311.57	***	1495.25	***

Panel B: Goodness-of-fit Deciles

Decile	Model 1	Model 2
1	47.95	56.85
2	10.96	11.3
3	9.93	8.9
4	7.53	5.14
5	8.22	3.77
6	6.16	4.79
7	3.42	3.77
8	2.4	1.37
9	1.71	2.4
10	1.71	1.71

This table shows the importance of the functional form of PD. Panel A reports results from the multiperiod logistic regression of the bankruptcy indicators. Parameter estimates for all variables in each model are reported along with their standard errors below. The logistic model is estimated for the sample period of 1968-2002 with 143,416 observations and 1,198 bankruptcies. The chi-square of the likelihood ratio test for each model is reported in the row labeled LRT. *** denotes significance at the 1% level. Panel B compares the out-of-sample accuracy. Parameter estimates are used to predict bankruptcies for the validation period of 2003 to 2013. This validation sample includes 45,835 firm years and 292 bankruptcies. For every year we rank firms into deciles based on their fitted bankruptcy probability values, where the firms with the highest values fall into the first decile. We report the percentage of bankrupt firms that are classified into each probability deciles.