PREDICTING DEFAULT MORE ACCURATELY: TO PROXY OR NOT TO PROXY FOR DEFAULT?

Koresh Galil and Neta Sher

Discussion Paper No. 15-05

August 2015

Monaster Center for
Economic Research
Ben-Gurion University of the Negev
P.O. Box 653
Beer Sheva, Israel

Fax: 972-8-6472941
Tel: 972-8-6472286

Predicting default more accurately: to proxy or not to proxy for default?

Koresh Galil *, **

Neta Sher *

June 3,2015

Previous studies targeting accuracy improvement of default models mainly focused on the choice of the explanatory variables and the statistical approach. We alter the focus to the choice of the dependent variable. We particularly explore whether the common practice (in literature) of using proxies for default events (bankruptcy or delisting) to increase sample size indeed improves accuracy. We examine four definitions of financial distress and show that each definition carries considerably different characteristics. We discover that rating agencies effort to measure correctly the timing of default is valuable. In predicting default and in explaining CDS spreads, a default model significantly outperforms any other type of financial-distress model, despite being estimated on a substantially smaller sample (72 defaults compared to 409 bankruptcies and 923 delistings).

Keywords: Default; Bankruptcy; Financial Distress; Delisting; Bankruptcy Prediction; Default Prediction.

JEL classifications: G17; G33

The authors would like to thank Zvika Afik, Banita Bissoondoyal-Bheenick, Doron Kliger, Jan Pieter Krahnen and Gal Zahavi and seminar participants in ORIC 2012 meeting in Jerusalem, IFABS 2012 meeting in Valencia and MFC meeting in Prague for fruitful comments.

* Koresh Galil (galilk@bgu.ac.il) and Neta Sher (netashe@bgu.ac.il) are from the Economics department of Ben-Gurion University at the Negev.

** Corresponding author – Ben-Gurion University in the Negev, Department of Economics, P.O. Box 653, Beer-Sheva, Israel. Tel: +972-54-4565570.

Introduction

Economists and accountants have been trying to forecast financial distress for decades. Researchers in the field have been focusing upon two main aspects: the first is the proper choice of explanatory variables. In earlier years, the main question in this context was which financial ratios are the most essential for identifying credit risk potential. The early literature relied upon accounting-based measures as the forecasting variables. The most popular accounting-based predictors are Altman's (1968) Z-score and Ohlson's (1980) O-score. Hillegeist, Keating, Cram and Lundstedt (2004) implemented the Merton (1974) market-based measure of bankruptcy probability. They demonstrated that compared to the two accounting based measures of Altman and Ohlson, this measure provides significantly more information concerning credit risk potential.

The second aspect that researchers focused on was the methodology used to estimate the likelihood of failure. In early literature, the most dominant methodology was discriminant analysis, followed by LOGIT analysis (both methods are also referred to as static models). Beaver (1966), Altman (1968), Ohlson (1980) and Zmijewski (1984) used static models to estimate the failure probability. Shumway (2001) estimated a dynamic LOGIT or hazard model, and pointed out that this technique is preferable to the static models used beforehand.[1]

A third aspect concerning the prediction of financial distress, which had not received enough attention, is the definition of financial distress, which seems to vary significantly among different studies. Altman (1968) and Ohlson (1980) attempted to forecast bankruptcy, which they identified with a firm's filing of a bankruptcy petition. Dichev (1998) outlined a broader definition of distress, by addressing firms that were delisted because of poor performance as his sample for failed firms. Shumway (2001) used bankruptcy and delisting events as an indicator of financial distress. Campbell, Hilscher, and Szilagyi (2008) used a failure indicator that included bankruptcy filings, delisting for financial reasons or receiving a D rating. Bharath and Shumway (2008) defined distress as default; they obtained their default data from the database of firm default maintained by Edward Altman and by using the list

[1] Other papers using hazard models include Chava and Jarrow (2004), and Beaver, McNichols, and Rhie (2005), Duffie, Saita and Wang (2007).

of defaults published by Moody's.

There are several possible reasons for researchers' practice of using proxies for default events. First, a standard dataset of default events among US public companies is non-existent and therefore researchers rely on diverse sources for the construction of their events lists. Second, the number of default events is relatively small. Moreover, once such a list is intersected with other data (e.g. accounting or market data), the final set becomes even smaller. Under these terms it is tempting to use alternative distress definitions (proxies for defaults) in order to expand the set of failure events.

The proximity of the default events to other types of negative events assists in identifying such proxies. A financial default is a state in which a debtor is unable or unwilling to fulfill the terms of a debt contract or a debt instrument. Such an event may come after occurrence of other negative events such as a rating downgrade or a major drop in the value of the equity. A default may also precede other types of financial distress events, such as bankruptcy filing or delisting. Rating agencies exert effort in identifying default events and their exact timing. Moody's definition of default includes three types of credit events: (1) a missed or delayed disbursement of interest and/or principal; (2) bankruptcy filing or legal receivership; (3) a distressed exchange. [2] [3] The time of default is set (by the rating agency) to be the earliest of the above events, because it is then that the major loss is recognized. Missed or delayed payments and distressed exchanges normally precede bankruptcy filings and therefore default events (as defined by rating agencies) normally precede bankruptcy events. [4] Moody's (2000) emphasizes that the alternative definitions of default are not intended to broaden the central idea of non-payment or bankruptcy, but simply to get the timing right. Yet, Moody's definition of default is also slightly broader than the definition of bankruptcy, because it also includes delayed payments. Such events do not necessarily lead to bankruptcies as debtors are paid later. Rating agencies consider them as default events

[2] This definition appears in various default studies by Moody's. See for example, Moody's (2011) page 61. S&P definition of default is similar. See for example Standard and Poor's (2011) p. 65.

[3] A distressed exchange is an event in which the issuer offers bondholders a new security or package of securities that amount to a diminished financial obligation (such as preferred or common stock, or debt with a lower coupon or par -amount) helping the borrower to avoid the other types of default.

[4] Brunner and Krahnen (2008) showed in the context of bank debt that private workout activities usually commence well before formal bankruptcy proceedings are initiated and therefore a bankruptcy filing may be a late indicator of financial distress.

because of the meaningful opportunity costs they load on investors.

We hypothesize that rating agencies' effort for timing accuracy is valuable and hence default-prediction models outperform other financial-distress models in predicting defaults and in explaining default-related financial instruments (e.g Credit Default Swaps). We specify four alternative definitions for distress, three of which are well-known definitions: bankruptcy, default and delisting. Bankruptcy is identified with a firm's deletion from Compustat for bankruptcy or liquidation reasons, or with another indication of bankruptcy in the financial statements (bankruptcy footnote). We follow Bharath and Shumway (2008), using the default definition of distress; we obtain the default data from S&P and Moody's default lists. We also follow Dichev (1998), using exchange delisting for liquidation or poor performance as a proxy for distress. In addition, we examine another proxy for distress: drawdown. A drawdown event occurs when a stock has a significant negative accumulated return from its highest record in the preceding 12 months. We examine this type of event as an example of distress events that precede defaults. [5]

We present a thorough analysis of these four alternative definitions of financial distress and find that, as expected, each type of distress carries different characteristics and outlines a different financial distress environment. For example, we find that the delisting definition is much broader than the bankruptcy and default definitions, as it captures considerably more distress events. We also find that, as expected, the drawdown events precede the delisting events, default and bankruptcy events and default events tend to precede bankruptcy and delisting events. Comparison of the explanatory variables reveals that firms that undergo a default or bankruptcy event tend to be more leveraged than firms that undergo delisting or drawdown events.

We choose the methodology of the LOGIT model to create several different distress prediction models, based on the different types of distress definitions. We perform out-of-sample predictions and

[5] We also examined two other distress definitions: Penny event (the first time stock prices falls below $1 value) and low return (the first time 12 month accumulated return was lower than -80%). These definitions resulted in larger event sets than in other definitions) and prediction models based on these definitions had a poorer performance comparing to the traditional definitions (bankruptcy, default and delisting). We do not present these results and they are available upon request.

evaluate the models' out-of-sample accuracies for predicting all alternative types of distress events. We do so by using two different prediction evaluation methods. We find that alternative definitions applied in different models display different out-of-sample results. For example, in predicting default, the default model shows statistically significant better prediction results than any other prediction model. Therefore, it seems that for predicting default, one should favor the default prediction model over any other model. This result is especially remarkable when taking into account the small sample size of default events (72 defaults compared to 409 bankruptcies and 923 delistings). Consequently, our analysis implies that the advantage of a broader definition of distress, which allows for a larger sample, is not preferable to the definition of default alone. The poor performance of the delisting model is quite surprising, when taking into account the common usage of this definition in credit risk predictions (e.g. Dichev 1998, Campbell et al. 2008).

Our analysis also reveals that for predicting default, one should account for a built-in bias in the default sample. This bias results from the nature of the default events collection; the default sample is collected from the rating agencies' default lists, and therefore does not include defaulted firms that are not rated.

We examine the significance of our results by comparing the out of sample improvement in default-prediction accuracy achieved through two methods. The first is a replacement of the bankruptcy prediction model with the default prediction model as suggested by us. The second is the exchange of the LOGIT model with a hazard model as suggested by Shumway (2001). We do not find any accuracy improvement in replacing the LOGIT model with a hazard model in none of the specifications we examined (exponential hazard model with or without shared-frailty and Cox proportional hazard model). It appears that capturing the adequate timing of a default event is more important than using a hazard model.

In the final section we follow Berndt, Douglas, Duffie, Ferguson and Schranz (2008) and Bahrath and Shumway (2008) by examining whether the financial distress probabilities, which are generated from all financial distress prediction models, are informative explanatory variables for pricing CDS spreads, which is considered to be a market based default measure. We find that all generated probabilities of distress are statistically significant in explaining CDS spreads, if regressed separately. However, regressing CDS spreads against the default probability combined with each of the alternative probabilities, shows that the statistical significance of all models is driven out by the default model. Again, this result is particularly astonishing when taking into account that the default sample is much smaller than the alternative samples.

This study has also applications in the study of stock prices because the differences between the various financial distress events have been ignored in studies searching for financial-distress premium in stock prices. Dichev (1998) found that stocks of firms with higher distress-risk accumulate lower return. Campbell, Hilscher, and Szilagyi (2008), and Avramov, Chordia, Jostova, Philipov (2009) also confirmed this finding. Garlappi, Shu and Yan (2008) found no relation between stock returns and distress-risk, Vassalou and Xing (2004) found this relation to be positive. While all these studies examined the relation between stock returns and distress-risk, they have mostly used alternative measures of distress risk: credit ratings, the probability of delisting, the probability of bankruptcy, the probability of default, the probability of either deleting, bankruptcy or default. It is not clear whether finding absence of a relation between stock returns and one measure of distress-risk (such as default probability as by Garlappi et al., 2008) contrasts finding a negative relation with another measure of distress-risk (e.g. delisting-risk by Dichev, 1998).[6]

The study proceeds as follows: the next section discusses the data. In this section we discuss the alternative definitions of distress that we analyze in the paper and conduct a comparison between the alternative definitions. Section 2 discusses the methodologies used in the study. Section 3 outlines the

[6] Lu and Chollete (2010) demonstrated that the negative relation between distress-risk and stock returns using bankruptcy definition disappears once using a delisting definition.

results. We conclude in section 4.

1. Data

We examine all firms in the intersection of the Compustat Industrial File and the CRSP Daily stock return File. The data set consists of financial data from 1990 to 2009.[7] Firms with Standard Industrial Classification (SIC) codes from 6,000 to 6,999 (financial firms) are excluded.

1.1 Financial distress data

We specify four alternative definitions of distress: bankruptcy, default, delisting and drawdown. We document the financial distress events in the following manner: first, we identify the first event for each of the distress definitions, for each company within the data set. Financial distress events may have a long-term effect; therefore, we define a three-year time range as part of the same financial distress event. Accordingly, we remove all data within three years following the first identified event.[8] Only after eliminating the data as reported, we identify the next financial distress event. We repeat this process for all recurring events, for each distress definition separately.[9] In this fashion, we built four data sets, one for each distress definition for the years 1991-2009. As such, our sample also includes the Global Financial Crisis of 2007-2009.

[7] Note that since we predict distress one year ahead, our sample of accounting spans the period of 1990-2008 while the distress event data (including stocks performance) spans the period of 2991-2009 and therefore also includes the Global Financial Crisis.

[8] Firms may default repeatedly. However, many times subsequent default announcements only reflect the same default event. For example, Catalyst Paper Co. defaulted in 2010, 2011 and 2012. Thus, using 2010 observation for predicting default in 2011 is already "contaminated" by the default of the firm in 2010. The subsequent default events only reflect the firm's unsuccessful effort to overcome its financial distress. Default prediction models, in our perspective, focus on credit quality assessment of healthy firms rather than recovery prediction of defaulted firms. Therefore, if a firm defaults in 2000, we estimate its probability to default on 31 December 1999 and then drop this firm from our sample for the years 2000, 2001 and 2002.

[9] For example, if firm A defaulted in 2003, we indicate a default event in the year subsequent to 2002 and remove annual observations for years 2003, 2004 and 2005 from analysis of default events. However, the indication of a default event does not affect the construction of the other distress-events samples. Therefore, if this firm was delisted in year 2004, we indicate a delisting event subsequent to the annual observation of 2003 and remove annual observations for years 2004, 2005 and 2006.

Definitions of financial distress events:[10]

1. **Default** – Firms that appeared in the S&P and Moody's default lists. We define the date of the default event as the earlier of the default dates indicated by the two rating agencies.

2. **Bankruptcy-** We define bankruptcy as one of the following:

   Deletion- The firm was deleted from Compustat, and the reason for the deletion was either bankruptcy or liquidation. We define the date of the bankruptcy event as the first day after the firm's last annual or quarterly report.

   Footnote- The firm's annual or quarterly report included a 'bankruptcy footnote'. We define the date of the bankruptcy as the date of the first report that included the footnote.

3. **Delisting** – The firm was delisted because of poor performance. We analyze firms with CRSP delisting code in the 400 and 500 classes. [11]

4. **Drawdown** - The firm’s stock price fell by 96.6% from its highest record within the previous 12 months.[12]

1.2 Comparison of the different definitions of financial distress

In order to compare the different definitions of financial distress, crosschecking tables are constructed. The crosschecking tables compare the distress events in two aspects: the quantity of the financial distress events and their timing. For this comparison, we analyze only the first distress event, for each of the four financial distress definitions.[13] As an example, if firm A defaulted three times during

[10] We also considered two other types of distress events: Penny event (the first time stock prices fall below $1 value) and low return (the first time 12 month accumulated return was lower than -80%). These definitions resulted in larger event sets than in other definitions, and prediction models based on these definitions had a poorer performance compared to the traditional definitions (bankruptcy, default, drawdown and delisting). We do not present these results and they are available upon request.
[11] A CRSP delisting code in the 400 class means that the firm is being liquidated. The 500 class indicates that the firm is being delisted because of poor performance.
[12] We chose this threshold to assure an amount of drawdown events in years 1990-2009 similar to that of bankruptcy events.
[13] For comparing the different definitions, we include all firm-years. (We do not eliminate outliers and firm-years with missing data).

the sample years and was delisted twice, we only examine the first default event and the first delisting event.

Table 1 presents the comparison. Panel A displays the number of observed events for each of the distress definitions, and the number of observed joint events for every pair of definitions. Panel B displays the joint events, which are divided into six groups, according to their timing of occurrence. Each row compares one pair of definitions.

The comparison reveals that delisting events are more frequent than those of default, bankruptcy and drawdown are. We identify 5,542 delistings, 1,495 bankruptcies, 1503 drawdowns and only 1098 defaults. It might seem strange that we identify more bankruptcies than defaults. This outcome is the first indication that the rating agencies' default lists suffer from selection bias, which results in fewer defaults in our sample. Furthermore, the analysis indicates that there is a connection between the different types of distress events; 74.4% of the defaulted firms and 90.2% of the bankrupt firms undergo a delisting event. However, Only 32.2% of the bankrupt firms experience default. The comparison also shows that drawdown events usually precede the delisting, default and bankruptcy events; 13% of firms that endure both drawdown and bankruptcy events endure a bankruptcy event prior the drawdown, compared to 41% vice versa. Moreover, 8% of firms that experience both drawdown and default events, experience the default event prior the drawdown event, compared to 41% vice versa.

Only 19.7% and 22.6% of firms that experience a drawdown event, experience a default and a bankruptcy event respectively. This result may suggest that the drawdown definition is too broad, as only a small fraction lead to default and bankruptcy. This characteristic might limit its use for credit risk prediction. In this context, the delisting definition demonstrates similar results: Only 14.7% of the delisted firms experience default, and 24.3% of the delisted firms experience bankruptcy.

1.3 Independent variables

We estimate the prediction models with two different sets of independent variables. The forecasting models contain Altman's (1968), and Ohlson's (1980) independent variables, which have been widely used in other studies and in practice. Since our only focus is the dependent variable's effect on financial distress prediction, we choose these common accounting based measures, taking into account that adding market based variables may possibly improve our prediction results.

Altman's variables include the ratios of working capital to total assets (WC/TA), retained earnings to total assets (RE/TA), earnings before interest and taxes to total assets (EBIT/TA), market value equity to total liabilities (MVE/TL), and sales to total assets (S/TA). Ohlson's variables include log of total assets (SIZE), total liabilities to total assets (TL/TA), working capital to total assets (WC/TA), current liabilities to current assets (CL/CA), a dummy variable which gets a value of 1 if total liabilities exceeds total assets, and 0 otherwise (OENEG), net income to total assets (NI/TA), funds provided by operations to total liabilities (FPO/TL), a dummy variable which gets a value of 1 if net income was negative for the last two years and 0 otherwise (INTWO), the ratio $(\mathrm{NI_t} - \mathrm{NI_{t-1}})/(|\mathrm{NI_t}| + |\mathrm{NI_{t-1}}|)$, where $\mathrm{NI_t}$ is the net income for the most recent period (CHIN).

There are a number of extreme values among the observations. In order to ensure that outliers will not heavily influence the results, we eliminate all observations that are higher than the ninety-ninth percentile or lower than the first percentile of each variable. Since a complete set of explanatory variables is not always observable for each firm's annual report, we eliminate all annual reports for which the explanatory data set is not complete.

1.4 Forecasting models

We build four samples of annual accounting data, one for each financial distress definition. We estimate the probability of a firm's financial distress with its annual accounting data. If a firm endures a distress event within 12 months after the annual report of year t, the distress dummy of this firm will be assigned 1 for year t and 0 otherwise. Table 2 shows the distribution of the distress events across the years. The table presents the number of failures in each year, for each of the distress definitions. Table 3 shows the frequency of the distressed firm-years observations in each sample, and the mean values of selected explanatory variables used in the prediction models. We present the mean values by dividing the observations into two groups: observations of healthy firm-years and observations of distressed firm-years.

Most financial ratios come out as expected. The WC/TA and FPO/TL ratios, which are liquidity ratios, tend to be larger for the healthy firm-years. For example, the drawdown sample shows a mean FPO/TL ratio of 0.056 for healthy firm-years, and -0.576 for distressed firm-years. The mean WC/TA ratio in the default sample is 0.258 for healthy observations and only 0.059 for distressed observations.

The TL/TA ratio, which indicates what proportion of the company's assets is being financed through debt, tends to be larger for failing firm-years. Moreover, comparing firms that underwent default events to firms that underwent delisting or drawdown events, underlines an interesting outcome; defaulted firm-years tend to have a larger TL/TA ratio. The mean TL/TA ratio is 0.846 for defaulted firm-years, and only 0.726, 0.680 and 0.702 for distressed firm-years in the bankruptcy, delisting and drawdown samples respectively. This outcome may imply that default events are more correlated with high leverage.

As expected, the SIZE variable tends to be larger for healthy firm-years. This result is robust for all samples, apart from the default sample. The reason for this outcome lies in the nature of the default sample. We collect the default data manually from S&P and Moody's default lists. These lists only include firms that are currently rated or had been rated in the past. Therefore, there are firms in our sample that may have endured a default event that was not documented because they were not ever rated. Consequently, small defaulting firms in our sample, which are less likely to have been rated, may not appear in our default lists. This may explain why the SIZE variable is bigger for defaulters in the default sample. This bias in the default sample may influence the financial distress predictions. To overcome this problem we define an additional financial distress definition; default among rated. For this definition, we only keep firms that have been rated by S&P (as indicated in Compustat) in the sample.

A comparison of the default sample and summary statistics of the default among rated sample shows that addressing the size bias of defaulted firms in our sample results in a larger SIZE variable for healthy firm-years compared with that of the defaulted ones. In the first default sample the mean SIZE variable is 5.21 for healthy observations and 5.973 for distressed observations, in the default among rated sample the mean SIZE variable is 7.37 for healthy observations and 6.62 for distressed observations. This implies that within the rated firms, SIZE is negatively correlated with distress. This result is coherent with the findings of previous studies such as Campbell, Hilscher and Szilagyi (2008), which showed that bankrupt firms tend to be relatively small.

2. Methodology

We divide each of the five samples into two groups: estimation sample and control sample (out-of-sample). The estimation sample consists of all annual financial statements between the years 1990-1998. The control sample consists of all annual financial statements between the years 1999-2008. As stated in the previous section, if a firm undergoes a financial distress event within 12 months after the annual financial statement of year t, the distress dummy of this firm will be assigned 1 for year t and 0 otherwise.

We are only interested in how the choice of the dependent variable affects the financial distress predictions. Therefore, we use a standard static LOGIT method, taking into consideration that using a more advanced method could probably improve the prediction results. We examine discrete-period hazard models in separate section (3.3). We maximize the following likelihood function (L):

(1) $$L = \sum_{i=1}^{N} \sum_{t=1}^{T} \{d_{it} lnG(x_{it}) + (1 - d_{it}) ln[1 - G(x_{it})]\}$$

where d_{it} is 1 if firm i defaults during year $t + 1$ and 0 otherwise and $G(x_{it})$ is the logistic function:

(2) $$G(x_{it}) = \frac{exp(\beta x_{it})}{1+exp(\beta x_{it})}$$

where x_{it} is the explanatory variable.

For each of the four distress types we first estimate the Altman and Ohlson models' coefficients using the estimation sample and then we test the models' power using the control sample. That is, we construct two prediction models for each of the four definitions; one by using Ohlson's variables, and the other by using Altman's variables. Consistent with much of the prior literature, we examine each updated measure's ability to explain the five distress outcomes over the following year.

2.1 Prediction ability

In the case of distress models, validation involves examining a model along two aspects: Model

Calibration and Model Power. Calibration addresses the accuracy of a model's predicted probability, whereas a model's power is its ability to discriminant between distressed and non-distressed observations. To examine the models' prediction ability, we present findings from two validation methods. Both evaluation methods are power tests, and as such only require the ranking of the firms' distress probabilities, and not the estimation of the actual probabilities of distress.

Deciles method:

Following Shumway (2001), we sort all observations in the control sample into deciles, based on their failure probabilities. We then examine whether the observations of distressed firm-years show up in the riskier groups. If the model predicts financial distress properly, we would see failing firm-years extensively in the first few deciles. The deciles method is useful for providing an intuitive foundation but is limited in the information it provides. Furthermore, we are not aware of a statistical inference or other tests that allow proper quantitative evaluation or ranking of this method's results.

ROC curve method:

Another method of forecast ability evaluation is the Receiver Operating Characteristic (ROC) curve. As stated above, the prediction models produce predicted probabilities of distress for each firm-year. A critical probability value (cutoff point) is defined as the value that outlines all observations with higher probability of failure as "risky" (classified as distressed) and all the observations with lower probability of failure as "safe" (classified as non-distressed). For every cutoff point, the type I and type II error rates can be measured, as presented in figure 1. A type I error is said to occur when the observation's probability of distress is greater than the cutoff point, but the observation (a specific firm in a specific year) does not experience a financial distress event the following year. In a similar fashion, a type II error will occur if the observation's probability of distress is less than the cutoff point, and the observation does in fact endure a financial distress event the following year.

The ROC curve generalizes the contingency table representation of the model performance through all potential cutoff points. The ROC curve provides information on the performance of the model at all possible cutoff points, measuring the tradeoff between the type I and type II error rates for the entire range of cutoff points. The x-axis presents the false positive rate (type I error) and the y-axis presents the true positive rate (1- type II error). A point is plotted on the graph for each of the cutoff points. These plotted points form the ROC curve. Figure 2 presents a schematic of the ROC curve.

A well-known index associated with an ROC curve is the Area Under Curve (Swets and Pickett, 1982). The Area Under Curve (AUC) is an index for measuring the performance of the model. The greater the AUC, the better the model classifies the failed and non-failed observations. The AUC range is between 0.5 (random model) and 1 (perfect model). We use the De-Long test (DeLong, DeLong, and Clarke-Pearson, 1988) for AUC statistical comparison.

2.2 CDS regressions

After evaluating the prediction ability of the forecast models, we go on to examine the generated default probabilities ability to explain Credit Default Swaps (CDS) spreads. Following Berndt et al. (2008), and Bharath and Shumway (2008) we regress the log of the CDS spread against the log of default probabilities produced from the different prediction models. We run the following regressions:

$$\log CDS_{it} = \alpha + \beta_j \cdot \log FDF_{it}^{j} + \varepsilon_{it}, \tag{3}$$

$$\log CDS_{it} = \alpha + \beta_j \cdot \log FDF_{it}^{j} + \beta_k \cdot \log FDF_{it}^{k} + \varepsilon_{it}, \tag{4}$$

where CDS_{it} is the CDS spread of firm i in time t and FDF_{it}^{j} is the financial distress probability of firm i in time t based on model j (drawdown, delisting, bankruptcy or default). In equation (4) we compete between two models j and k where $j \neq k$. In this way, we examine whether the produced probabilities of default are consistent with market's estimates of default risk (CDS spreads) and whether one model outperforms the other in explaining CDS spreads. It should be noted that the purpose of this analysis is not to estimate the determinants of the CDS spreads, rather to identify whether one explanatory variable (financial distress probability given by one model) is a sufficient statistics for the other explanatory variables (financial distress probabilities according to the other models).

3. Results and discussion

3.1 Out of sample results

We construct two prediction models for each of the five alternative definitions; one by using Ohlson's variables, and the other by using Altman's variables. We generate the prediction models by using the estimation samples, which include all observations between the years 1990-1998. Table 4 reports on the number of observations and distress events in each of the five estimation samples. [14] We present the updated coefficients for each model in Table 5. We estimate one set of coefficients for each model. Altman's variables are statistically significant in most prediction models. While four of the five coefficients (WC/TA, RE/TA, EBIT/TA, MVE/TA) have the same signs as their counterparts in Altman's original model, the S/TA variable has a different sign than its original counterpart. [15] For the Ohlson models, we find that while seven of the nine variables are statistically significant in most prediction models, the CL/CA and OENEG variables are mostly insignificant. Most of the coefficient signs come out as expected. We find that the SIZE coefficient is negative for all models apart from the default model. This outcome results from the selection bias in the default sample, after accounting for the bias by conditioning the observations on S&P ratings, the SIZE coefficient is indeed negative.

The following sections contain the out-of-sample results of the prediction models. We present findings from two validation methods: The decile method and the ROC curve method.

3.1.1 Out of sample forecasts- Decile method

Following Shumway (2001) we sort all firms-years into deciles based on their failure probabilities. Then we tabulate the number of financial distress events that actually take place in each of the decile groups. We examine the prediction ability of the five alternative definitions of distress by using all five prediction models; we do so for each set of explanatory variables.

Table 6 reports on the success of all forecasting models using Altman's independent variables; each panel displays the results of the out-of-sample accuracy in predicting a certain type of event. Figure 3 shows the schematics of the out-of-sample results, each panel displays the schematic of the out-of-sample accuracy in predicting a certain type of distress event. In each of the figures, the x-axis represents

[14] We include the default among rated model.

[15] Consistent with Hillegeist, Keating, Cram and Lundstedt (2004), we find that a few coefficients have substantially changed from their original values.

the percentiles of the failure probabilities and the y-axis represents the accumulated percentage of the occurring events. In order to assess the successfulness of the prediction models, we conduct a comparison between the plotted curves; if two curves never cross, we determine that one dominates the other in all cases. If two curves do cross, we favor the model that offers better results in identifying failure among the worst quality observations, by examining the first two deciles.

This analysis, while only providing a basic understanding, shows several interesting outcomes. It appears that for predicting a specific type of distress event, it is best to use the prediction model of the same type of event, or from what seems to be a similar type of event. To be precise, default prediction shows by far the best results when using the default model or the default among rated model (panel A). The default model classifies 67% of all defaults in the highest default probability decile. Both the bankruptcy and delisting models cannot match this accuracy, classifying only 43% of all defaults in the highest decile. This result is particularly remarkable when taking into consideration the small sample size of defaulted firms. The drawdown model shows even worse results, classifying only 26% of all defaults in the highest default probability decile. Looking at Figure 3 (panel A); it is apparent that the default model outperforms all alternative models in predicting default.

Out-of-sample prediction results of bankruptcies (panel B) illustrate the best results for the delisting, default among rated and bankruptcy models, while showing poor results for the drawdown model. The bankruptcy, default, default among rated and delisting models appear fairly accurate, assigning approximately 50% of bankrupt observations to the highest bankruptcy probability decile. Both bankruptcy and delisting models classify 78% of bankrupt observations in the two highest deciles. These results are fairly close to the bankruptcy prediction results presented by Shumway (2001). When using Altman's explanatory variables, Shumay's hazard model classified 82% of all bankrupt firms in the two highest deciles. The drawdown model is of inferior quality, classifying only 27% of bankrupt observations in the highest decile.

In the case of predicting delisting events (panel C), it seems that the differences between the alternative prediction models (except the drawdown model) are relatively small. The delisting and bankruptcy models present the best results, by successfully classifying 55% and 53% of the delisted firms in the highest delisting probability deciles. The default model classifies approximately 50% of the delisting events in the highest decile. The drawdown model displays the worst results, classifying only 27% of the events in the highest decile.

The drawdown model shows moderate results in predicting the out-of-sample drawdown events (panel D), successfully assigning 55% of the events in the highest decile, and classifying 73% of the events above the median probability. The default and bankruptcy models display inferior results, classifying only 32% and 33% of the drawdown events to the highest deciles. Figure 3 (panel D) reveals that the drawdown model dominates all alternative models in predicting drawdown events. The analysis shows that it is somewhat difficult to predict this type of event, as even the drawdown model classifies only 42% of the events in the highest probability decile.

The default-among-rated model shows poor results in predicting defaults in the unconditional sample. It only correctly predicts 28% and 62% of defaults in the two highest deciles respectively, compared to 72% and 86% by the default model. However, in predicting default-among-rated the accuracy profiles swap and the default-among-rated model outperforms the unconditional default model (Panel E). These differences demonstrate that the selection bias in the default sample significantly affects the prediction accuracy.

We continue this analysis with a second set of explanatory variables in order to examine the robustness of the results. Table 7 reports on the success of all forecasting models using Ohlson's independent variables. Figure 4 demonstrates schematics of the out-of-sample results. The O-score prediction models show quite similar results to the Z-score models. In the default prediction, the default model again shows the best result, classifying 72% of defaults in the highest decile. The delisting model shows very poor results, classifying merely 29% of defaults to the highest decile. Bankruptcy out-of-sample prediction results show decent results for the bankruptcy and default models, both classifying approximately 53% of bankrupt observations to the highest decile. The delisting model shows very poor results, classifying only 36% to the highest decile, very similar to the drawdown model, which classifies 35% of bankrupt observations to the highest decile. The poor performance of the delisting model is quite surprising, when taking into account the common usage of this definition for the purpose of credit risk prediction. For example, Dichev (1998) selected the delisting definition as his failure indicator. By defining the failure indicator as delisting, he was able to collect 1,121 delisting events. In comparison, Shumway (2001) hand collected only 300 bankruptcies. However, our analysis may imply that the advantages of a broader definition of distress, which allows for a larger sample, may not always be preferable to the definition of bankruptcy alone.

It appears that different types of financial distress events have different characteristics and therefore one type of event cannot necessarily be used to predict a different type of event. Furthermore, the analysis shows that for predicting credit risk potential, the usage of broader distress definitions may not always be preferable.

As mentioned previously, the deciles technic of measuring the prediction ability can offer an intuitive understanding, but is limited in the information it provides. In order to compare the different models and examine the statistical significance of the differences between them, we use the ROC curve method.

3.1.2 ROC analysis

Figures 5-9 display the ROC curves of the forecast models. We present Only the ROC curves that we formed by the O-score prediction models. [16] The different AUC are compared to a gold-standard, using the De-Long et al. (1998) test. The gold standard is defined as the AUC that is created by using the same definition of financial distress to forecast a certain distress definition. For example, for examining the forecast ability of bankruptcy events, we compare the AUC of all models to the AUC of the bankruptcy model. Table 8 summarizes the results.

The ROC curves results support the conclusions from the previous section, and give the conclusions statistical validation. The main advantage of the ROC curve validation method is that it enables statistical inference with the non-parametric test suggested by De-Long et al. (1998).

In the default prediction, once again the default model displays the best results. The ROC curves (Figure 8) show that the default model dominates all other models in predicting defaults. The default

[16] The Z-score models achieved fairly similar results which are omitted because of space considerations.

model demonstrates an AUC of 0.8905. The De-Long et al. (1988) test reveals that one can indeed reject the hypothesis that the AUC of the bankruptcy, delisting and drawdown models equal the AUC of the default prediction model. The bankruptcy model shows decent result, with an AUC of 0.8397. The drawdown model shows quite poor results with an AUC of 0.7345. It should be noted that the superiority of the default model in predicting defaults may not be attributed to fewer 'false-positive' cases because of the smaller number of default events. The ROC curve and the AUC consider both type-I errors and type-II errors and therefore reducing one of type of error on the account of another type of error would not support such a significant improvement in prediction power.

The out-of-sample results of bankruptcy prediction show the best results for the bankruptcy model, which demonstrates an AUC of 0.8115. The default model also displays fairly good results, displaying an AUC of 0.8073. Moreover, the Delong test reveals that one cannot reject the hypothesis that the AUC of the bankruptcy model equals the AUC of the default model. Looking at the ROC curves (Figure 9), it is quite obvious that the bankruptcy and default models outperform the alternative models. Similar to the case of default prediction, the drawdown model shows poor results with an AUC of 0.7593 and the delisting model shows a similar AUC of 0.7592. According to the De-Long test, AUC of both models are significantly different from the bankruptcy model.

The AUC results reinforce the findings from the previous section. The results illustrate significant differences between the models' prediction powers. For example, analysis of the default prediction shows that the smallest gap between the two models is 0.0558, whereas the largest gap between the two models is 0.1560. For comparison, Afik, Arad and Galil (2012) found that the largest gap between several alternative specifications of Merton (1974) model was 0.014.

3.2 Default vs. default among rated

As discussed in the section 2, there is an inherent bias in the default sample that results from the nature of the default events data collection. We manually collected the default events from S&P and Moody's default lists, which are naturally composed of rated firms alone. Therefore, there are firms in both the default sample and the default out-of-sample that may have endured a default event that was not documented, because the firms were not ever rated. Subsequently, the default event definition is problematic and actually refers to a joint event of defaulters and rated firms.

Considering this, one may claim that the alternative models show lower performance in predicting defaults, not because they fail to predict defaults but because they fail in predicting the existence of ratings. We only include in our default-among-rated sample firm-years that are currently rated or have been rated by S&P in the past. In this manner, we do not let defaulters appear as non-defaulters, merely because they are not rated.

Table 9 summarizes the AUC results of the predictions of the alternative distress definitions. Panel A compares the AUC generated by using the default prediction model to the AUC generated by using the default among rated prediction model. For the delisting and drawdown predictions, the default-among-rated model shows better results than the default model. For drawdown prediction, the first default model displays AUC of 0.7652 whereas the default-among-rated model shows AUC of 0.7999. The delisting prediction shows an AUC of 0.7305 when applying the default model and an AUC of 0.8366 when applying the default-among-rated model. In predicting bankruptcy, the default model and the default among rated model show very similar results, AUC of 0.8073 and 0.8035 respectively. [17] These results illustrate that for predicting financial distress, one should take into account the selection bias in the rating agencies' defaults lists.

We also alternate the out-of-sample data, by eliminating all the non-rated firms from our control sample. In this fashion, we generate a new out-of-sample of the default-among-rated definition. Panel B in Table 9 summarizes the AUC results for the default among rated prediction, in comparison to the AUC results for the prediction of the default events as defined initially. Although one cannot compare prediction models by comparing their AUC on different out-of-samples, it seems that conditioning the observations on S&P ratings improves the default prediction ability of the prediction models, as the differences between the models are significantly smaller. The fact that in the initial default definition, non-rated defaulted observations cannot appear as defaulted observations causes the prediction of the

[17] The difference between the models is not statistically significant.

default events to be a prediction of a joint event of default and existence of rating. According to our findings, by conditioning the default out-of-sample and correcting the selection bias, all prediction models show better results in predicting defaults. The results suggest that the poor performance of these models in predicting "defaults" are partly caused by their failure in predicting the existence of rating

However, the main conclusion remains; there are differences in the prediction ability of the different models, and for predicting defaults, the default model is superior to all alternative models and this superiority is statistically significant.

3.3 The significance of accuracy improvement

To examine the significance of our results, we compare the accuracy improvement achieved through the choice of the dependent variable to the accuracy improvement achieved through the choice of the statistical approach. To assure a fair competition, we use the most accurate LOGIT model in prediction of default (beside the default model) – the bankruptcy model- as a benchmark. We now compare the accuracy improvement (in prediction of default) when we switch from a bankruptcy model to a default-among-rated model, to the accuracy improvement achieved when switching from a LOGIT model to a hazard model. Table 10 shows these results.

As shown above, the out-of-sample AUC for the default-among-rated model (M2) is greater than that of the bankruptcy model (M1) – 0.8892 vs. 0.8753. The difference is also statistically significant. Now we examine the out-of-sample accuracy of four hazard models (M3-M6) in prediction of default. We estimate these models using the same estimation sample and explanatory variables as in M1. Models M3-M5 are exponential hazard models, where M3 ignores frailty, M4 contains a Gamma-distributed frailty component and M5 contains an inverse-Gaussian frailty component.[18] The M6 is a Cox-proportional hazard model commonly used in the literature. None of the hazard models exhibit an improvement compared to the LOGIT model. In fact, the Cox-proportional hazard model (M4) appears to be significantly inferior to the LOGIT model (M1) in prediction of default. As illustrated by Shumway (2001) the LOGIT model is inferior to a hazard model because it ignores the dependency of

[18] Frailty is the parallel concept of random effects in hazard models.

the probability of default in time t on the survival time until t. These results may indicate that this problem is not severe, and perhaps there is not much age effect in the probability of default among public firms. These results are also consistent with the empirical analysis of Shumway (2001) that only showed an outperformance of the hazard model over Altman (1968) Multivariate Discriminant Analysis but not over Zmijewski (1984) LOGIT model. Shumway (2001) estimation of the hazard model also discovered no statistically significant age effect.

To conclude, changing the dependent variable as we suggest, results in a significant accuracy improvement, while changing the estimation method from a LOGIT model to a hazard model does not.

3.4 CDS spread regressions

The previous results demonstrate that for predicting defaults, it is best to use the default financial distress prediction model. The next set of results examines whether the financial distress probabilities, which are generated from all financial distress prediction models, are informative explanatory variables for pricing CDS spreads, which is considered to be a market based default measure. For this analysis, we use the new definition of default; default-among-rated, that was shown to have better prediction results for all alternative prediction models. We derive all other probabilities for distress from the same prediction models that were specified in previous sections, from the estimation sample (1990-1998). We follow Berndt et al. (2008), and Bahrath & Shumway (2008)) and regress the log of the CDS spread against the log of the default probabilities generated from the various prediction models, time dummies, and the fixed effect approach.[19] The purpose of this analysis is to detect whether one explanatory variable (distress probability given one model) is a sufficient statistics for the other explanatory variables (distress probabilities given by the other models).

We obtain the CDS data from Markit for the period of January 2002 to December 2009. We predict a firm's probability of default in the next year (Ohlson's model), following the process described in previous sections. We use all four distress definitions, thus generating four default probabilities for every firm-year in our control sample. We then pair every probability with the firm's compatible CDS

[19] We performed the Hausman test to determine whether a random or fixed effects model is more appropriate. The test indicated that a fixed effects model is a superior option in this case.

spread. That is, we regress firms' log of CDS spreads ($\log \mathrm{CDS}_{\mathrm{it}}$) on the log of financial distress probabilities ($\log FDP_{it}^{j}$), where FDP_{it}^{j} is the financial distress probability of firm i on time t based on model j. The probabilities of distress are estimated for the end of financial year and the CDS spreads are the observed 5-years CDS spread on the same day+6 month. Using this procedure, we are able to collect 3,619 paired CDS-FDP observations, for each distress definition. The sample includes CDS spreads of 587 firms. Panel A of table 11 shows the summary statistics of the log CDS spreads and the log FDP. Panel B provides a Spearman correlation matrix.

The matrix reveals a relatively low correlation between the drawdown probabilities and the default probabilities of 0.560. As anticipated, the bankruptcy model's probabilities are highly correlated with the default model's probabilities (0.963). There is a very high correlation between the default model and the delisting model (0.870).

Table 12 displays the regression coefficients results, regressing the log CDS against the distress probabilities that are derived from the different prediction models. In Panel A we regress the log of CDS spreads against each of the different default probabilities separately (models 1-4). The coefficients on the default probabilities are all positive and significant at the 1% level. The log FDP of the default model shows the highest R^2, with a value of 0.248. The log FDP of the delisting model shows a very similar R^2 of 0.242. The log FDP of the drawdown model shows the lowest R^2, with a value of 0.104. The R^2 values are lower than Bharath & Shumay's estimates of 0.26 and 0.38; this may be explained by our usage of accounting data alone. Hillegeist et al. (2004) demonstrated that market based measures provide significantly more information about the probability of bankruptcy than do either of the popular accounting-based measures of Altman and Ohlson. The lower R^2 may also be explained by the special features of our CDS sample period, which also includes the crisis period of 2007-2009.

In panel B we regress the log of CDS against the Default model's probability combined to each of the alternative models' probabilities. Interestingly, adding each of the probabilities to the default model's probability in the same regression (models 5-7) shows that the statistical significances of the bankruptcy and delisting models are driven out by the default model.[20] The coefficient of the drawdown probability is statistically significant but with the 'wrong' sign (negative), indicating that this measures something else than credit risk. These results show that the default model outperforms all other models in explaining CDS spreads. Given that the CDS spreads are actually market based default measure, this outcome supports the conclusions from previous sections. This result is again especially astounding given the default's model considerably smaller sample size.

4. Summary and conclusions

In this paper, we examine several different proxies for firm distress. We outline three well-known definitions of distress: default, bankruptcy and delisting, as well as a new proxy for distress: drawdown events. We find that the delisting definition is much broader than the default and bankruptcy definitions, as it seems to capture considerably more distress events. We also find that, as expected, the drawdown events commonly precede the delisting, default and bankruptcy events.

We apply the methodology of the LOGIT model to create several different distress prediction models, which are based on the different types of distress definitions. We evaluate the models' out-of-sample accuracies for predicting all alternative types of distress events. Our analysis shows that there are significant differences between the models' prediction abilities. This outcome implies that using an unsuitable proxy for distress might limit the prediction ability. Moreover, it seems that for predicting a certain type of event, one should favor the model which uses the same type of event or from what seems to be a similar type of event. The accuracy improvement achieved through the fine-tuning of the dependent variable is significant. For comparison, we do not find any accuracy improvement by switching from a LOGIT model to hazard models as suggested by Shumway (2001).

We also examine the ability of the different financial distress probabilities (FDP) to explain CDS spreads. Again, we demonstrate that default is well explained by default model probabilities, while other types of distress model probabilities do not achieve the same level of explanation.

[20] It should be noted that the purpose of this analysis is to examine whether any of the probability estimates in our sample of probability estimates is a sufficient statistics. By definition, this analysis is not affected by omission of other variables that may explain CDS spreads.

The outcomes of this study indicate that different definitions of distress should not necessarily be viewed as different signals for the same occurrence, but rather be regarded as different types of distress events, which may carry different features and characteristics. We demonstrate that for predicting defaults, one should use the default prediction model, even if it is based on a much smaller sample. Rating agencies' effort to catch the timing of defaults accurately is valuable. A default model should also account for a selection bias that exists in default lists provided by rating agencies.

References

Afik, Z., Arad, O., and K. Galil, 2012, "Using Merton model: an empirical assessment of alternatives", working paper, Ben-Gurion University.

Altman, E. I. 1968. "Financial ratios, discriminate analysis, and the prediction of corporate bankruptcy". Journal of Finance 23, 589–609.

Beaver, W. H., 1966, "Financial ratios as predictors of failure", Journal of Accounting Research 4, 71–111.

Beaver, W. H., McNichols, M. F., and Jung-Wu Rhie, 2005," Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy". Review of Accounting Studies 10, 93–122.

Berndt, A., Douglas, R., Duffie, D., Ferguson, M., and D. Schranz, 2008, "Measuring default-risk premia from default swap rates and EDFs, Working paper, Stanford University.

Bharath, Sreedhar, and Tyler Shumway, 2008, "Forecasting default with the Merton distance to default model", Review of Financial Studies 21, 1339–1369.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi. 2008. "In Search of Distress Risk". Journal of Finance 63, no. 6: 2899-2939.

Chava, S., and R. Jarrow, 2004. "Bankruptcy Prediction with Industry Effects". Review of Finance 8, 537–69.

DeLong, E.R., DeLong, D.M., and D.L., Clarke-Pearson, 1988. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach", Biometrics 44, 837-845.

Dichev, Ilia, 1998. "Is the risk of bankruptcy a systematic risk?" Journal of Finance 53, 1141–1148.

Duffie, D., Saita, L., and K., Wang, K. 2007. "Multi-period corporate default prediction with stochastic covariates." Journal of Financial Economics 83(3), 635-665.

Hillegeist, S.A., Keating, E.K., Cram, D.P., Lundstedt, K.G., 2004. "Assessing the Probability of Bankruptcy". Review of Accounting Studies 9, 5–34.

Macey, J., O'Hara, M. and D., Pompilio, 2008. "Down and out in the stock market: The law and finance of the delisting process", Journal of Law and Economics, forthcoming.

Merton, R. C. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. Journal of Finance 29:449–70.

Ohlson, J.S, 1980, "Financial ratios and probabilistic prediction of bankruptcy". Journal of Accounting Research 19, 109-31.

Moody's, 2000, "Moody's Approach to Evaluating Distressed Exchanges", available on Moody's website.

Moody's, 2011, " Corporate Default and Recovery Rates, 1920-2010", available on Moody's website.

Shumway, T., 2001. "Forecasting bankruptcy more accurately: a simple hazard Model", Journal of Business 74, 101—124.

Standard and Poor's, 2010, "2010 Annual Global Corporate Default Study and Transition Matrices, " available at Standard and Poor's website.

Standard and Poor's, 2011, "2011 Annual Global Corporate Default Study And Rating Transitions, " available at Standard and Poor's website.

Swets, J.A, Pickett, R.M, 1982, "Evaluation of Diagnostic systems: Method from signal Detection Theory". New York: Academic Press.

Zmijewski, M.E, 1984, "Methodological related to the estimation of financial distress prediction models". Journal of Accounting Research 22, 59-82.

Tables

Table 1: Crosschecking alternative definitions of financial distress

The following table reports on the comparison between every pair of distress definitions. The table compares the definitions along two dimensions: the number of events and their timing. For the purpose of this analysis, only the first event of each firm is kept in the sample. That is, we do not account for recurring events. Panel A displays the number of observed events for each of the distress definitions, and the number of observed joint events for every pair of definitions. For example, the comparison between delisting and default shows 1,098 firms which underwent a default event, 5,542 firms which underwent a delisting event, and 812 firms which underwent both types of events. Panel B displays the joint events, which are divided into 6 different groups, according to their timing of occurrence. For example, out of the 812 firms, which underwent both a default event and a delisting event, 45 firms experienced the delisting event more than 3 years prior to the default event.

Panel A: Frequency of distress events

	Default	**Bankruptcy**	**Delisting**	**Drawdown**
Default	1098 **(100.0)**	481 (43.8)	817 (74.4)	296 (27.0)
Bankruptcy	481 (32.2)	1495 **(100.0)**	1,349 (90.2)	339 (22.7)
Delisting	817 (14.7)	1,349 (24.3)	5542 **(100.0)**	1090 (19.7)
Drawdown	296 (19.7)	339 (22.6)	1090 (72.5)	1503 **(100.0)**

Panel B: Timing of distress events

		Definition 1 prior to definition 2					Definition 2 prior to definition 1				
	Years between events	**3+**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**3+**	**Sum**
Def 1: default	Events	18	8	19	119	235	46	5	5	26	481
Def 2: bankruptcy	%	4%	2%	4%	25%	49%	10%	1%	1%	5%	100%
Def 1: default	Events	32	10	21	82	488	100	23	16	45	817
Def 2: delisting	%	4%	1%	3%	10%	60%	12%	3%	2%	6%	100%
Def 1: default	Events	7	4	1	15	147	84	17	6	15	296
Def 2: drawdown	%	2%	1%	0%	5%	50%	28%	6%	2%	5%	100%
Def 1: bankruptcy	Events	33	5	19	275	650	208	67	30	62	1,349
Def 2: delisting	%	2%	0%	1%	20%	48%	15%	5%	2%	5%	100%
Def 1: bankruptcy	Events	9	1	6	27	153	98	15	15	15	339
Def 2: drawdown	%	3%	0%	2%	8%	45%	29%	4%	4%	4%	100%
Def 1: delisting	Events	11	7	2	12	536	324	77	24	97	1090
Def 2: drawdown	%	1%	1%	0%	1%	49%	30%	7%	2%	9%	100%

Table 2: Distribution of distress events over time

This table reports the number of failures for every year of the sample period.

Year	Defaults	Bankruptcies	Delistings	Drawdown
1991	14	56	71	42
1992	9	41	87	15
1993	9	33	84	9
1994	7	35	65	10
1995	7	39	71	6
1996	7	42	81	12
1997	14	47	96	36
1998	21	64	170	61
1999	29	52	198	37
2000	31	42	161	107
2001	60	34	221	203
2002	32	27	134	96
2003	13	17	121	22
2004	10	12	61	4
2005	11	13	70	3
2006	4	12	61	1
2007	4	4	50	5
2008	14	12	84	112
2009	13	6	75	89
Total	**309**	**588**	**1,961**	**870**

Table 3: Mean values of explanatory variables

The following table reports the frequency of distressed firm-years and the mean values for selected variables used in the prediction models. The table includes both the estimation and the control samples (1990-2009) after dropping outliers and missing data. WC/TA is the working capital divided by total assets; EBIT/TA is earnings before interest and taxes divided by total assets; MVE/TA is the market value equity divided by total liabilities; TL/TA is total liabilities divided by total assets; SIZE is the log of total assets; FPO/TL is funds provided by operations divided by total liabilities. The mean values are divided into two groups: observations of healthy firm-years (distress dummy equals 0) and observations of distressed firm-years (distress dummy equals 1).

	Distress	Freq.	%	WC/TA	EBIT/TA	MVE/TA	TL/TA	SIZE	FPO/TL
Default sample	0	73,173	99.58	0.258	0.011	1.381	0.491	5.210	0.054
	1	309	0.42	0.059	-0.090	0.389	0.846	5.973	-0.094
Bankruptcy sample	0	72,861	99.20	0.258	0.012	1.382	0.490	5.223	0.056
	1	588	0.80	0.100	-0.181	0.775	0.726	4.236	-0.238
Delisting sample	0	71,147	97.32	0.261	0.017	1.392	0.487	5.263	0.064
	1	1,961	2.68	0.105	-0.235	0.835	0.680	3.420	-0.308
Drawdown sample	0	75,322	98.85	0.203	-0.007	1.446	0.553	5.335	0.056
	1	870	1.15	0.057	-0.347	2.061	0.702	4. 705	-0.576
Default among rated sample	0	19,447	98.92	0.150	0.077	1.007	0.625	7.370	0.159
	1	212	1.08	0.066	-0.042	0.264	0.880	6.602	-0.028

Table 4: Frequency of distressed firm-years in the estimation samples

This table displays the frequency of distressed firm-years in each of the estimation samples (1990-1998).

Sample	Distress dummy	Freq.	Percent
Default	0	36,786	99.68
	1	117	0.32
Bankruptcy	0	36,432	98.89
	1	409	1.10
Delisted	0	35,912	97.49
	1	923	2.51
Drawdown	0	38,107	99.41
	1	228	0.59
Default amongst rated	0	8,441	99.15
	1	72	0.85

Table 5: Updated coefficients for Altman (1968) and Ohlson (1980) models

The updated coefficients are estimated in a LOGIT regression that includes all available firm-years in each of the estimation samples (1990-1998). Altman's variables include the ratios of working capital to total assets (WC/TA), retained earnings to total assets (RE/TA), earnings before interest and taxes to total assets (EBIT/TA), market value equity to total liabilities (MVE/TL), and sales to total assets (S/TA). Ohlson's variables include log of total assets (SIZE), total liabilities to total assets (TL/TA), working capital to total assets (WC/TA), current liabilities to current assets (CL/CA), a dummy variable which gets a value of 1 if total liabilities exceeds total assets, and 0 otherwise (OENEG), net income to total assets (NI/TA), funds provided by operations to total liabilities (FPO/TL), a dummy variable which gets a value of 1 if net income was negative for the last two years and 0 otherwise (INTWO), the ratio $(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$, where NI_t is the net income for the most recent period (CHIN).

Altman updated model	WC/TA	RE/TA	EBIT/TA	MVE/TA	S/TA	Constant					Obs.	Distress events	LR χ^2
Default	-2.26***	-0.01	-2.42***	-2.05***	0.02	-4.03***					36,903	117	197.86***
Bankruptcy	-2.14***	-0.12***	-2.93***	-0.74***	0.54***	-4.30***					36,841	409	648.21***
Delisting	-2.44***	-0.31***	-3.39***	-0.69***	0.35***	-3.36***					36,835	923	1903.14***
Drawdown	-0.536***	0.268***	-1.71***	-0.11***	-0.10	-4.77***					38,335	228	81.23***
Default among rated	-1.40*	-0.55**	-4.10***	-6.01***	0.01	-1.89***					8,513	72	202.91***
Ohlson updated model	**SIZE**	**TL/TA**	**WC/TA**	**CL/CA**	**OENEG**	**NI/TA**	**FPO/TL**	**INTWO**	**CHIN**	**Constant**	**Obs.**	**Distress events**	
Default	0.22***	2.30***	-1.55*	-0.09	-0.24	-0.12***	-0.38*	1.64***	-0.34*	-9.00***	36,903	117	248.93***
Bankruptcy	-0.20***	2.24***	-1.29***	0.00	-0.47**	-0.37*	-0.19*	0.84***	-0.75***	-5.19***	36,841	409	675.83***
Delisting	-0.58***	1.88***	-1.59***	0.01	-0.11	-0.57***	-0.09	0.87***	-0.59***	-2.81***	36,835	923	2329.33***
Drawdown	-0.14***	-0.56**	-0.59***	-0.001	1.23***	-0.005	-0.15***	1.52***	-1.39***	-5.27***	38,335	228	393.05***
Default among rated	-0.36***	2.57***	-2.18*	0.07	-1.00*	-0.11	1.76***	-1.20***	-0.66***	-4.62***	8,513	72	174.43***

*** (**) [*] significant at the 1% (5%) [10%] level (two-sided test).

Table 6: Out-of-sample prediction results (Z score)

This table reports on the success of the different prediction models, with the purpose of predicting the alternative definitions of distress. Each panel reports on the success of predicting a certain type of event, using all prediction models that are generated from Altman's explanatory variables. All the models use the explanatory variables identified by Altman, and are estimated with annual data between the years 1990-1998. The out-of-sample data contains all annual data between the years 1999-2008.

Panel A: Default prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	67%	43%	43%	26%	67%
2	84%	77%	77%	44%	84%
3	92%	90%	89%	53%	89%
4	96%	95%	94%	60%	93%
5	97%	97%	97%	68%	97%
10	100%	100%	100%	100%	100%

Panel B: Bankruptcy prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	53%	49%	50%	27%	58%
2	69%	78%	78%	43%	70%
3	74%	83%	85%	50%	76%
4	78%	87%	88%	55%	80%
5	82%	88%	92%	61%	84%
10	100%	100%	100%	100%	100%

Panel C: Delisting prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	50%	53%	55%	27%	53%
2	64%	71%	73%	40%	67%
3	73%	79%	80%	47%	75%
4	79%	83%	84%	53%	80%
5	84%	88%	89%	61%	85%
10	100%	100%	100%	100%	100%

Panel D: Drawdown prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	32%	33%	36%	41%	32%
2	42%	46%	48%	55%	41%
3	48%	52%	54%	64%	48%
4	54%	56%	58%	70%	54%
5	57%	60%	60%	73%	57%
10	100%	100%	100%	100%	100%

Panel E: Default among rated prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	65%	53%	65%	33%	72%
2	81%	79%	78%	44%	86%
3	90%	87%	89%	49%	91%
4	93%	94%	94%	56%	93%
5	96%	98%	98%	64%	96%
10	100%	100%	100%	100%	100%

Table 7: Out-of-sample prediction results (O-score)

This table reports on the success of the different prediction models, with the purpose of predicting the alternative definitions of distress. All the models use the explanatory variables identified by Ohlson, and are estimated with annual data between the years 1990-1998. The out-of-sample data contains all annual data between the years 1999-2008.

Panel A: Default prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	72%	53%	29%	22%	28%
2	86%	78%	49%	51%	62%
3	89%	83%	66%	72%	82%
4	92%	88%	78%	84%	88%
5	93%	91%	86%	88%	92%
10	100%	100%	100%	100%	100%

Panel B: Bankruptcy prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	54%	52%	36%	35%	39%
2	66%	69%	56%	57%	65%
3	77%	80%	65%	70%	80%
4	84%	84%	78%	85%	87%
5	89%	88%	89%	89%	89%
10	100%	100%	100%	100%	100%

Panel C: Delisting prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	40%	55%	55%	38%	50%
2	54%	73%	74%	63%	72%
3	64%	81%	83%	77%	83%
4	71%	87%	89%	85%	88%
5	78%	91%	94%	90%	92%
10	100%	100%	100%	100%	100%

Panel D: Drawdown prediction – 3,213 events

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	38%	38%	31%	43%	37%
2	58%	64%	53%	64%	63%
3	70%	78%	69%	80%	77%
4	80%	85%	79%	88%	87%
5	86%	89%	86%	90%	91%
10	100%	100%	100%	100%	100%

Panel E: Default among rated prediction

Decile	Default model	Bankruptcy model	Delisting model	Drawdown model	Default among rated model
1	57%	63%	59%	44%	65%
2	84%	81%	81%	77%	86%
3	90%	87%	87%	88%	89%
4	91%	91%	91%	89%	91%
5	94%	93%	95%	89%	94%
10	100%	100%	100%	100%	100%

Table 8: AUC comparison (O-score)

This table summarizes the AUC (Area Under Curve) for all financial distress predictions, using all four prediction models. P values from the Delong et al. (1988) test, for the difference from the gold standard are in parentheses. In each prediction, the gold standard is the AUC that is created by using the same definition of financial distress to forecast the certain distress definition. For example, for examining the forecast ability of bankruptcy events, we compare the AUC of all models to the AUC of the bankruptcy model.

	AUC			
	Default prediction	Bankruptcy prediction	Delisting prediction	Drawdown prediction
Default model	**0.8905** **g. standard**	0.8073 (0.701)	0.7305 (0.000)	0.7652 (0.000)
Bankruptcy model	0.8397 (0.000)	**0.8115** **g. standard**	0.8416 (0.000)	0.7943 (0.274)
Delisting model	0.7398 (0.000)	0.7592 (0.000)	**0.8573** **g. standard**	0.7544 (0.000)
Drawdown model	0.7345 (0.000)	0.7593 (0.000)	0.7904 (0.000)	**0.8016** **g. standard**

Table 9: Default vs. and Default among rated (O-score)

This table reports on the differences in results when using Default sample vs. using Default amongst rated sample. Panel A. displays the AUC of the Default model in predicting all alternative distress events, and the AUC of the Default among rated model in predicting all alternative distress events. P values from the Delong et al. (1988) test, for the difference between the two prediction models are in parentheses. Panel B. reports the AUC of predicting default events and default among rated events, using all prediction models. P values from the Delong et al. (1988) test, for the difference from the gold standard are in parentheses. In each of the predictions, the gold standard is the AUC, which is created by using the same definition of financial distress to forecast the certain distress definition. For example, for examining the forecast ability of default events, we compare the AUC of all models to the AUC of the default model.

Panel A: Comparison between the Default model and Default among rated model

	AUC	
	Default model	**Default among rated model**
Bankruptcy out-of-sample prediction	0.8073	0.8035
Delisting out-of-sample prediction	0.7305	0.8366
Drawdown out-of-sample prediction	0.7652	0.7999

Panel B: Prediction of Default vs. Prediction of Default among rated

	Default prediction	Default amongst rated prediction
Default model	**0.8905** **g. standard**	0.8722 (0.0496)
Bankruptcy model	0.8397 (0.000)	0.8753 (0.000)
Delisted model	0.7398 (0.000)	0.8657 (0.000)
Drawdown model	0.7345 (0.000)	0.8168 (0.000)
Default among rated model	0.7973 (0.000)	**0.8898** **g. standard**

Table 10: Accuracy improvement – hazard models vs. default definition

This table reports on the differences in prediction of default among rated firms when using bankruptcy prediction models estimated through hazard model or LOGIT model vs. default among rated model estimated through a LOGIT model. The table displays the AUC of the various models. P-values from the Delong et al. (1988) test, for the difference between the two prediction models are in parentheses. The gold standard is the AUC, which is created by using bankruptcy prediction model, estimated through a LOGIT regression. The models are the exponential model (without frailty or with Gamma/Inverse-Gaussian frailty) and the Cox proportional hazard model. All prediction models use Ohlson (1980) explanatory variables.

Model	Default amongst rated prediction
Benchmark model:	
M1 - Bankruptcy LOGIT model	**0.8753** **g. standard**
Altering the dependent variable	
M2 - Default among rated LOGIT model	0.8898 (0.000)
Altering the statistical method (hazard models)	
M3 - Bankruptcy exponential-hazard model	0.8719 (0.263)
M4 - Bankruptcy exponential-hazard model with Gamma-distributed frailty	0.8735 (0.538)
M5 - Bankruptcy exponential-hazard model with Inverse-Gaussian-distributed frailty	0.8724 (0.332)
M6 - Bankruptcy Cox-proportional-hazard model	0.8679 (0.000)

Table 11- CDS spreads

The following panel reports summary statistics for the variables that are used in the regression. Log CDS is the log of the CDS spread (basis points). Log FDP Drawdown is the log of the financial distress probabilities (decimal fraction) that are generated from the drawdown prediction model. Log FDP Delisting is the log of the financial distress probabilities that are generated from the delisting prediction model. Log FDP Bankruptcy is the log of the financial distress probabilities that are generated from the bankruptcy prediction model. Log FDP default is the log of the financial distress probabilities that are generated from the default among rated model.

Panel A: Summary statistics

Variable	Mean	Median	Std. Dev	Min	Max
Log_CDS	4.395	4.194	1.103	1.140	9.949
Log FDP_Drawdown	-6.925	-7.050	0.829	-8.795	-3.202
Log FDP_Delisting	-7.005	-7.054	0.878	-9.542	-0.005
Log FDP_Bankruptcy	-5.711	-5.775	0.724	-8.249	-0.004
Log FDP_Default	-6.711	-6.746	1.011	-11.229	-0.000

Panel B: Spearman's Correlation matrix

	Log CDS	Log FDP Drawdown	Log FDP Delisting	Log FDP Bankruptcy	Log FDP Default
Log_CDS	1.000				
Log FDP_Drawdown	0.234	1.000			
Log FDP_Delisting	0.410	0.629	1.000		
Log FDP_Bankruptcy	0.372	0.676	0.853	1.000	
Log FDP_Default	0.421	0.560	0.870	0.963	1.000

Table 12: CDS regressions

This table reports on the results of regressing the log of the CDS spreads (basis points) against the log of distress probabilities (FDP in decimal fractions) and time dummies, using fixed effects. The CDS data are obtained from Bloomberg for the period January 2002 to December 2009. The total number of firm-years observations is 3,619; the total number of firms is 644. P values are shown in parentheses (*** significant at the 1% level). In panel A we regress the Log of the CDS spreads against each of the different default probabilities separately (Models 1-5). In panel B we regress the log of CDS against the Default model's probability combined to each of the alternative models' probabilities.

Panel A – Single variable regressions

Independent variable				
Variable	Model 1	Model 2	Model 3	Model 4
Constant	6.349***	7.620***	7.051***	7.143***
	(0.000)	(0.000)	(0.000)	(0.000)
Log FDP Drawdown	0.131***			
	(0.000)			
Log FDP Delisting		0.308***		
		(0.000)		
Log FDP Bankruptcy			0.2820***	
			(0.000)	
Log FDP Default				0.253***
				(0.000)
Observations	3,296	3,296	3,296	3,296
N. of firms	587	587	587	587
R-squared	0.105	0.242	0.204	0.248
Time effect	yes	yes	yes	yes
Fixed effect	yes	yes	yes	yes

Panel B – Multiple variables regressions

Independent variable			
Variable	Model 5	Model 6	Model 7
Const.	7.141***	7.217***	7.129***
	(0.000)	(0.000)	(0.000)
Log FDP Drawdown	-0.058**		
	(0.021)		
Log FDP Delisting		0.035	
		(0.685)	
Log FDP Bankruptcy			-0.044
			(0.554)
Log FDP Default	0.314***	0.227***	0.289***
	(0.000)	(0.001)	(0.000)
Observations	3296	3296	3296
N. of firms	587	587	587
R-squared	0.257	0.251	0.252
Time effect	yes	yes	yes
Fixed effect	yes	yes	yes

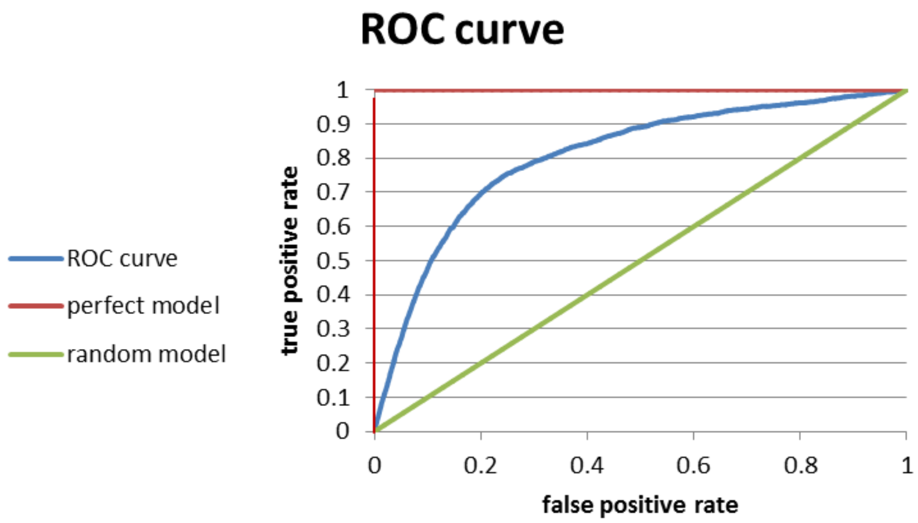
FIGURES

FIGURE 1: Simple contingency table

		Actual value	
		Failed	**Non failed**
Prediction outcome	**Failed**	**-**	**Type I error**
	Non failed	**Type II error**	**-**

This figure shows the simplest case of prediction; the model creates only two rankings (failed and non-failed). These are shown along with the actual prediction outcomes. The cells specify the number of type I errors, the number of type II errors and the number of successful predictions.

FIGURE 2: Illustration of ROC curves

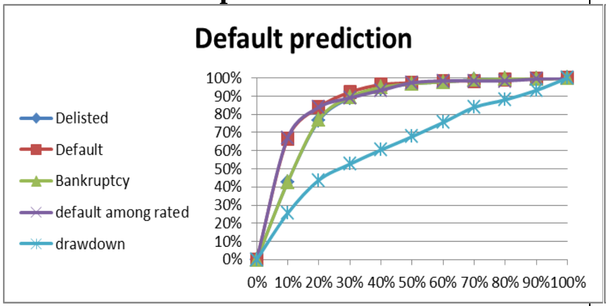


This figure illustrates a ROC curve compared to the perfect model and the random model.

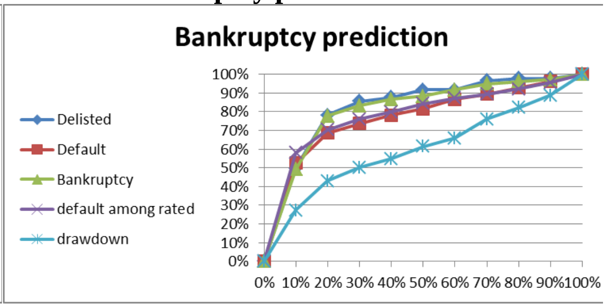
FIGURE 3: Deciles method out-of-sample accuracy (Z score)

The following panels illustrate the out-of sample prediction results of predicting each type of financial distress. Each panel shows the prediction accuracy of a certain distress definition, using all models that are generated with Altman's variables. The X-axis represents the percentiles of the failure probabilities. The Y-axis represents the accumulated percentage of the occurring distress events.

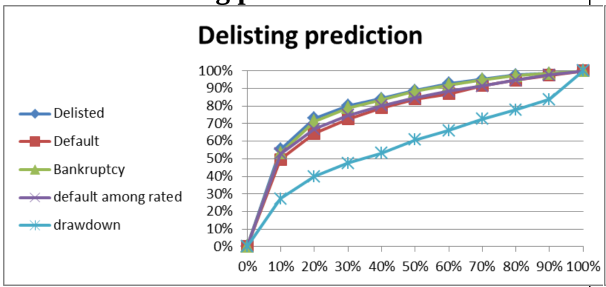
Panel A: Default prediction



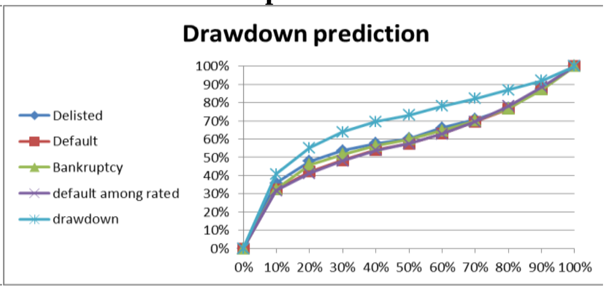
Panel B: Bankruptcy prediction



Panel C: Delisting prediction



Panel D: Drawdown prediction



Panel E: Default among rated

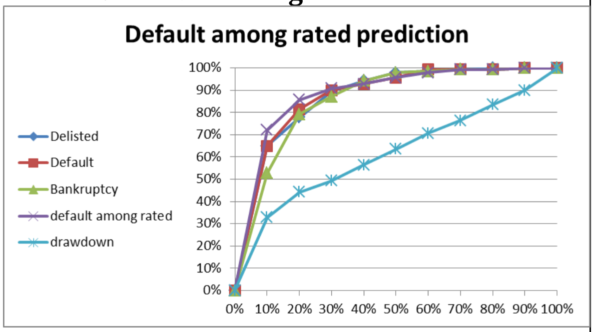
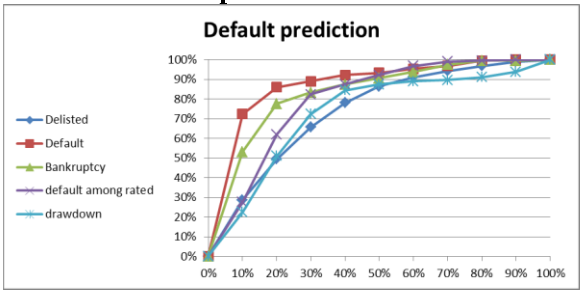


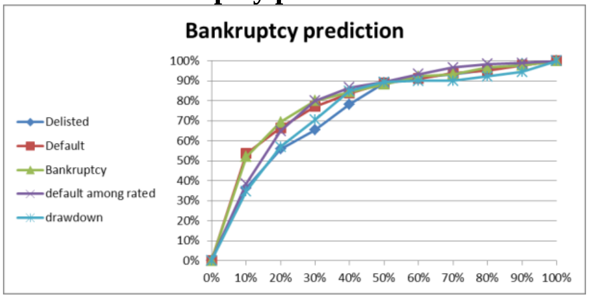
FIGURE 4: Deciles method out-of-sample accuracy (O score)

The following panels illustrate the out-of sample prediction results of predicting each type of financial distress. Each panel shows the prediction accuracy of a certain distress definition, using all models that are generated with Olson's variables. The X-axis represents the percentiles of the failure probabilities. The Y-axis represents the accumulated percentage of the occurring distress events.

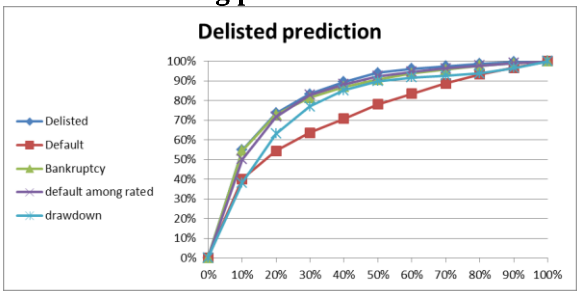
Panel A: Default prediction



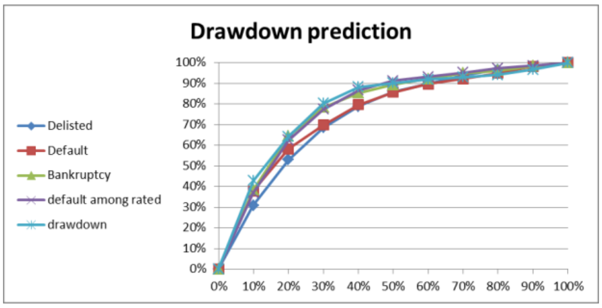
Panel B: Bankruptcy prediction



Panel C: Delisting prediction



Panel D: Drawdown



Panel E: Default among rated

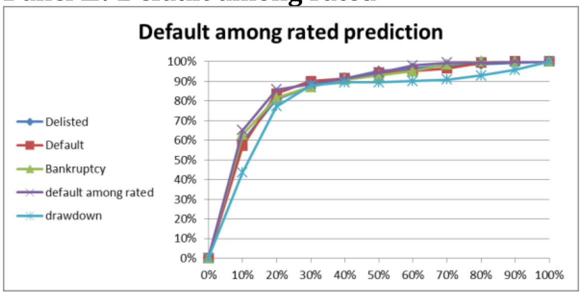
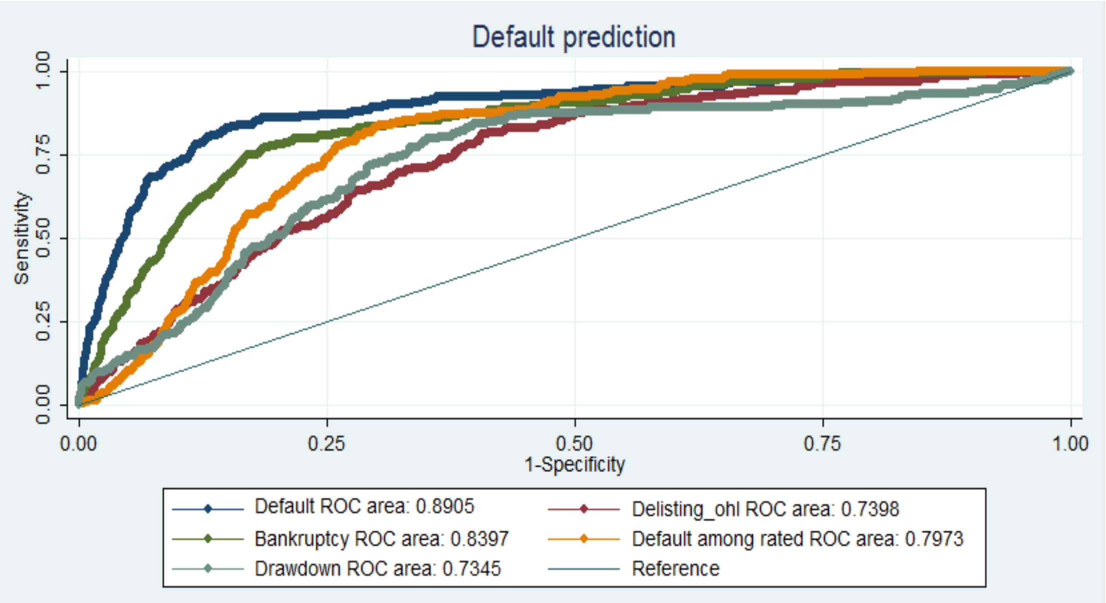
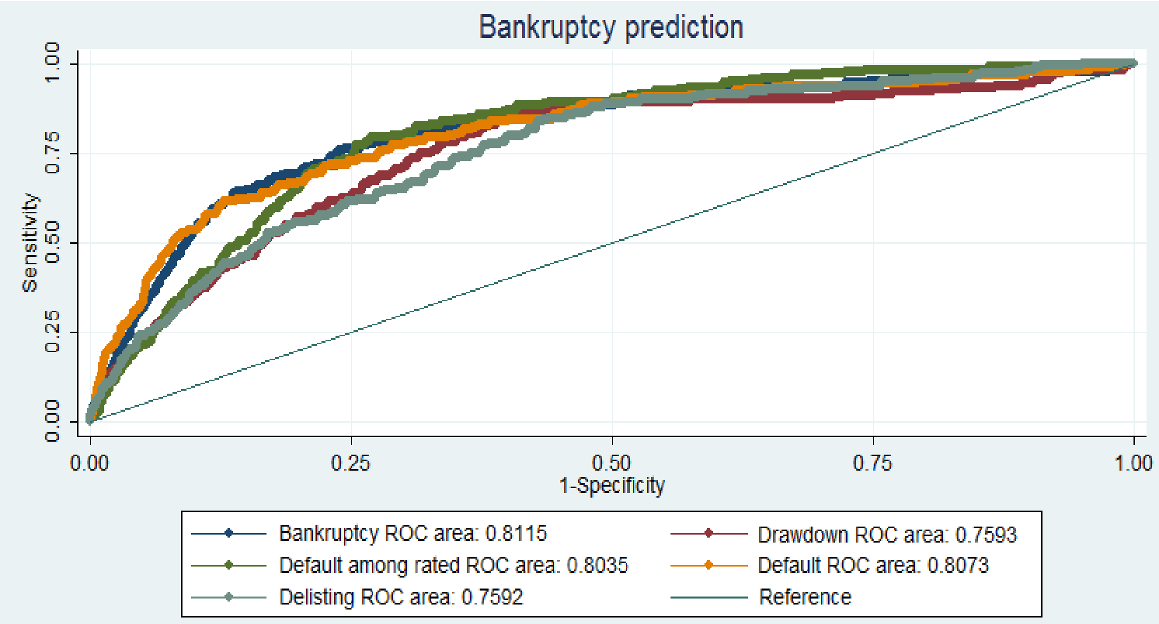


FIGURE 5: ROC curves of the out-of-sample default prediction (O-score)



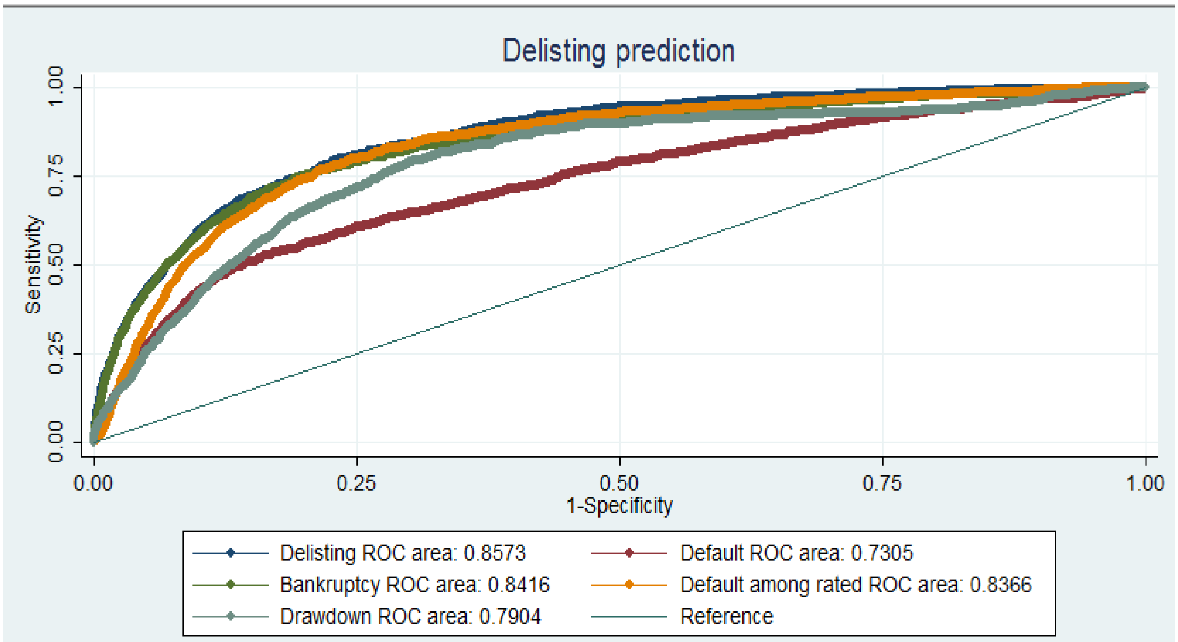
Roc curves computed from all five prediction models for predicting default events. All ROC curves are compared to the default model's ROC curve.

FIGURE 6: ROC curves of the out-of-sample bankruptcy prediction (O-score)



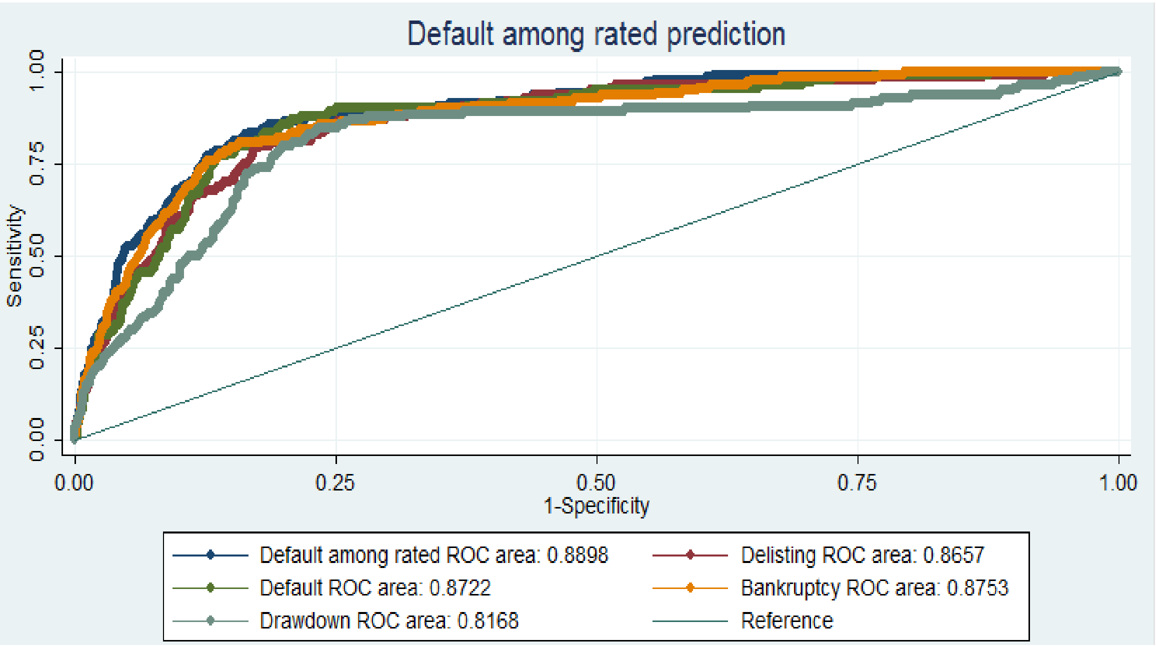
Roc curves computed from all five prediction models for predicting low returns events. All ROC curves are compared to the low return model's ROC curve.

FIGURE 7: ROC curves of the out-of-sample delisting prediction (O-score)



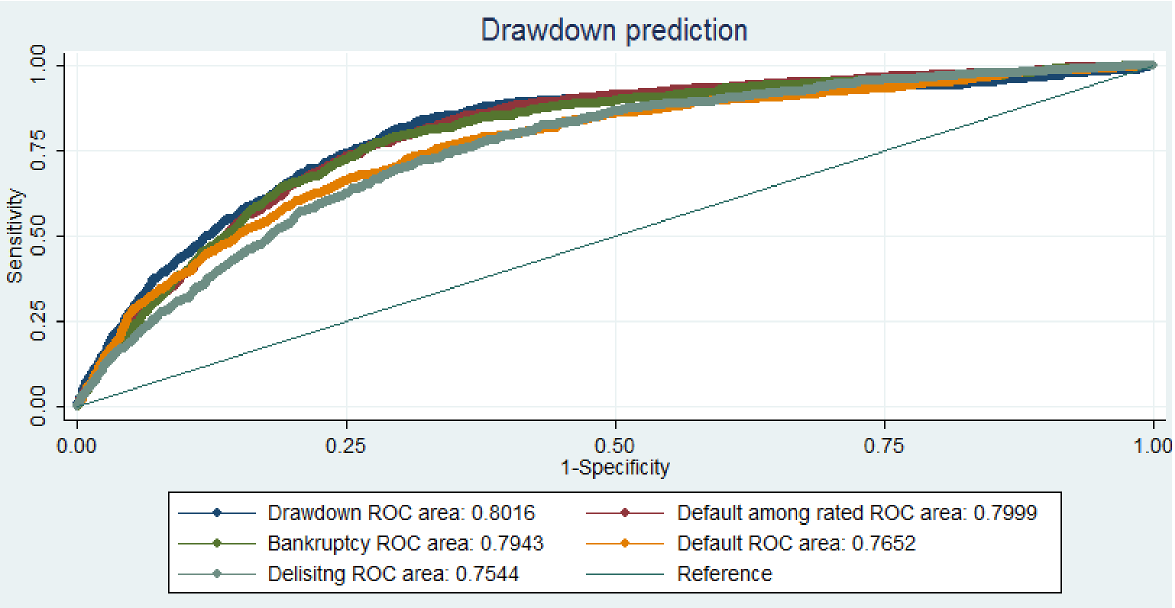
Roc curves computed from all five prediction models for predicting delistings. All ROC curves are compared to the delisting model's ROC curve.

FIGURE 8: ROC curves of the out-of-sample default among rated prediction (O-score)



Roc curves computed from all five prediction models for predicting defaults. All ROC curves are compared to the default model's ROC curve.

FIGURE 9: ROC curves of the out-of-sample drawdown prediction (O-score)



Roc curves computed from all five prediction models for predicting drawdown. All ROC curves are compared to the drawdown model's ROC curve.