Bankruptcy Prediction

Zeynep Topaloglu
The Graduate Center - CUNY
Department of Economics
ztopaloglu@gc.cuny.edu

Yildiray Yildirim *
Syracuse University
Martin J. Whitman School of Management
yildiray@syr.edu

June 17, 2009

Abstract

The hazard rate models used in the recent bankruptcy literature assume the censoring and the default are two independent events, which means the censored company will eventually default. However we believe there will a portion in the censored group that will be long-term survivors and we propose a mixture model of survivors and risky companies. Moreover this study models the event and the timing of default incident at the same time. For the event of default and the timing of default we utilize a logistic regression. The results have justified the advantage of our model over the standard hazard rate models and proved its predictive power. The companies identified as high default risk by our model proved to deliver extremely low returns in the market.

Key words: credit risk, mixture model

*We would like to thank seminar participants at Goethe-University Frankfurt, Syracuse University and CUNY Queens College.

1 Introduction

Default probability estimation has moved to the center of credit risk literature as credit risk instruments gained a higher share in the investment portfolios. However with the recent break-down in the credit markets which spread through the whole economy, the subject gained higher importance. In this study, we propose an improved statistical model, namely a mixture model, for the default [1] probability which measures the event and timing of the default at the same time and changes the assumption about the censored observations to include the long-term survivors. Besides, we improve the predictor variables to include the government stimulus and macro-economic changes captured by the unemployment rate.

The majority of the discussion regarding default rates evolves around the pioneering works of Altman [?], Ohlson [?], Zmijewski [?] and more recently Shumway [?]. The models proposed by Altman, Ohlson or Zmijewski constitute the static approach to default estimation. Shumway, on the other hand, introducing time variation to the covariates initiated a dynamic interpretation of corporate default events. Authors following Shumway, re-proved the superiority of dynamic modeling over static modeling. Shumway's hazard rate modeling have been improved by Chava and Jarrow [?] to include industry dummies and by Das et al. [?] to include industry frailties. Duffie et al. [?] instead of Shumway's accounting variables, utilized Merton's distance to default measure as Vassalou and Xing [?], Da and Gao [?] and Campbell et al. [?] did. Besides, Campbell et al. used market values of Shumway's accounting variables. Bharath and Shumway [?], on the other hand, found that Merton DD model does not provide a sufficient statistic for default prob-

[1] Two different approaches exist to define default. In the more conservative case, which is also used by Chava and Jarrow [?] and is followed in this study, bankruptcy, being Chapter 7 or 11 filing (refer Swanson [?]) for details), has been set as the default event since a company or its creditors filing with a federal bankruptcy court for protection under either Chapter 7 or Chapter 11 shows officially the company is unable to service its debt or pay its creditors. We are, indeed, by setting the dependent variable as the Chapter 11 (reorganization) and 7 (liquidation) filing, measuring the bankruptcy event. Later in the paper, bankruptcy and default terms will be used interchangeably though. In the other approach, a broader classification has been utilized to include Moody's definition for default referring a missed or delayed disbursement of interest and/or principal, including delayed payments. (Hillegeist et al. [?]) Financially driven de-listings, or D ratings issued by leading credit rating agencies are even considered as failure. (Campbell et al. [?])

ability. Roszbach [?], instead, used a bivariate tobit model with a variable censoring threshold to evaluate the survival of loans. The efforts to find out more efficient indicators, instead of seldom updated accounting variables, or capture contagion effects (Das et al. [?]) through industry indicators or frailties were not sufficient to correct for the fundamental deficiency embedded in hazard rate modeling due to heavy censoring (i.e. majority of the companies in the sample do not observe bankrutcy event). Our proposed model overcomes this problem

All of the prior works of dynamic modeling, either multi-period logistic, weibull, cox proportional hazard or tobit regressions are indeed modeling the timing of default by defining the intensity of default at each point in time. However, in our mixture model, we model event and intensity at the same time. What are the predictor for whether the company will default and when it will default? We answer both questions at the same time, which in turn significantly improves the in-sample prediction power of our model compared to the rest of the model. The results show that mixture model correctly identifies the companies to default in the sample, while hazard rate models fail to do so. The second distinction of the mixture model is the treatment of censored observations. Hazard rate models define a hazard rate for failing observations and a survival rate for the censored ones. Censored observations have the same fate, censoring and default are independent from each other. Therefore all of the censored observations defined to have survived for the sample period, and consequently are assumed to fail altogether in the future. However in the mixture model, we define two cases for censored observations. The censored observations are the combination of long-term survivors and risky companies. We believe it is not fair to consider a healthy survivor in the same group with Lehman Brothers who will fail the next year after the end of observation period. [2]

In the sample data selection, prior works naturally operated with an earlier starting point and on a smaller sample. In the subsequent works the significant distinction is observed at the year 1979. It is preferred to use the data after 1979 in order to avoid a possible shift in the coefficients on account of the new bankruptcy law of 1978. Another common application is to truncate the data at the 1st - 99th or 5th -95th percentiles to prevent outlier effects on the results. In this study

[2]Our sample cover 1980-2007, Lehman Brother enters into our sample period as a healthy company.

both of these precautions have been taken. We have 2,011,977 month per company observations for the period 1980-2007. We have identified 1,640 bankruptcies out of 18,505 publicly traded companies after correcting for the missing and repeating observations. We have not considered the private companies due to the lack of information supplied to the market compared to publicly listed counterparts.

Besides the econometric modeling, we also focus on the fundamentals behind the bankruptcies. Bankruptcies are the 'invisible hand' of financial system to clean out the bad apples from the good ones. The companies that made bad investment decisions and turn out to be insolvent is punished by liquidation. Productivity is the key factor for competitiveness and economic growth, that is why, as Prescott argues, the US is richer than the whole world [**?**]. Unproductive firms need to die for the economy to live happily ever after. The same fear of depressing the efficiency of production is behind the harsh critiques of the current bail-out plans. Thus productivity, profitability and efficiency are the central predictors in our model. Although the main job of bankruptcies in competitive markets is to augment productivity in the end, the productivity and efficiency measures used by Altman [**?**] in his Z-score calculation have been ignored by the recent studies, we re-emphasize their importance. Productivity and efficiency measuring the contribution of the business to the real economy, we have used profitability to figure out the contribution of the company to its owners. With respect to the debt management we have leverage ratio to capture the size of firm's debt burden, solvency ratio to evaluate whether it will be able to pay its debt back and liquidity ratio to deduce the proximity of the payments through available liquid assets. These are the accounting ratios deemed to be critical in default probability estimation together with Shumway's [**?**] market variables; relative size, sigma and excess return.

In addition to the productivity of the business as a means for natural selection to the later generations, we have been interested in two major macroeconomic factors unemployment rate and the stimulus packages. Rising unemployment rate as a proxy for the unfavorable economic conditions, and bankruptcies leading mass layoffs during crisis suggest the positive relationship between the default probability and the unemployment rate. From the official start of the recession in December 2007[3],

[3]NBER Business Cycle Dating Committee concluded that the start of the recession was December 2007 due, in large part, to the decline in jobs that began that month.

3.6 million people on payroll lost their jobs to the end of January 2009[4], about half of it occurring in the months of November, December and January.

Another central issue in our study is the impact of the stimulus packages, measured by the annual government spending, on the bankruptcy filings. Banks to financial institutions, auto industry giants to insurance companies have been waiting for the government to save them from bankruptcy during the recent crisis. The question of whether these stimulus plans will help to rescue the economy from the greatest downturn after Great Depression lays out there, we propose to seek the answer in the historical evidence. And we have found government spending being one of the most influential variable in our model, contrary to the existing fears in the market, history shows governments have been managed to reduce defaults by stimulating the economy, .

Exploiting all of these crucial predictors both on the business and market level, and implementing a mixture model we have been able to identify the default rate as 7.34%, which is very close to 8.14%.

The next step after the estimation of defaults, we discuss the return on companies. Campbell et al. [?]. show the high default risk companies deliver extremely low returns, on the other hand, Gao [?] conclude high-default risk companies provide abnormally high returns, however not as a result of the default risk. Our findings are consistent with Campbell et al. [?].

Lastly, after picking up the bad apples from the barrel, we have turned to the question: 'What should investors do now?' Vassalou and Xing [?], detect abnormal returns associated with distressed stocks. Da and Gao [?], on the other hand, approving high returns, claimed those high returns, indeed, were not a result of increasing default risk. We, however, found distressed companies receive extremely low returns. [5] Through an investment perspective, the investors should better avoid from these companies not to lose money.

The remainder of the paper is organized as follows. Section 2 discusses relevant literature. Section 3 develops the model setting. Section 4 reviews the data used

[4]By January 2009, the unemployment rate was 7.6% according to Census Bureau.
[5]Campbell et al. [?] had the same conclusion of low returns.

to test our model. Section 5 presents the results of our investigation. Section 6 concludes the paper.

2 Literature Review

First, we will explore the static modeling for default. Static model is a terminology first used by Shumway [?] to refer to the single-period classification models for multi-period bankruptcy data. Altman [?] for the calculation of his popular Z-score uses a sample of 66 more or the less proportionate manufacturing firms, 33 of which have failed.[6] On this dataset, using multiple discriminant analysis, 5 of the initial 22 variables, WC/TA, RE/TA, EBIT/TA, MVE/TL and S/TA were selected as more predictive.[7] Z-score provides information about the distress of the company. Specifically, the greater the firm′s distress the lower the discriminant score. Altman [?] finds that Z-score is an accurate forecaster of failure as far as 2 years in advance. Next he estimated ZETA score; he counts 53 bankrupt and 58 healthy firms. Compared to Z, ZETA has more variables in, a higher accuracy as a score, and the ability to forecast up to 5 years prior to the failure.

Shumway [?] criticizing the static models to his time, proposes a dynamic approach, which in time turns out to be the most famous way of bankruptcy modeling, known as hazard rate model.[8] The dependent variable in the model is the time spent by a firm in the healthy group. In other words, for each year the dependent variable shall indicate whether the firm is filed for bankruptcy or not. On the right hand side, the independent variables include firm age, firm size, excess return and standard deviation of stock returns, besides the Altman [?]'s Z score variables and Zmijewski [?]'s variables of NI/TA, TL/TA and CA/CL. His multi-period logit model (or hazard model with logistic intensity of default) is consistent and efficient compared to inconsistent static [9] models, because it corrects for period at risk and allows for time

[6]The sample data lies over the period 1946 to 1965.

[7]For the definition of these variables, please see section: Data-Firm Specific Accounting Variables

[8]He retains 300 bankruptcies in 3,182 non-financial companies for the period 1962-1992. He truncates the accounting data at the 1 and 99 percentiles and defines the default event as filing for any type of bankruptcy.

[9]the single-period models

varying covariates. After him, the hazard rate models became the standard, and the studies to follow here apply hazard rate models mostly with logistic regression, as Shumway [?] does, or one of the exponential, Weibull or Cox regressions for the default timing. He finds that half of the accounting ratios tested before are poor predictors, and the market variables: size, return and standard deviation forecast failure effectively.

Chava and Jarrow [?] compare Altman [?], Zmijewski [?] and Shumway [?] in their respective predictive powers for bankruptcy. While the previous literature uses the broader definition of default, they have employed the simpler definition of Chapter 7 and Chapter 11 filing of the company. The total of 1,197 bankruptcies out of 17,460 companies were extracted for the period 1962-1999, 7 years more than that of Shumway [?]. The forecasting accuracy of the models estimated with 1962-90 data over the years 1991-99 showed the superiority of Shumway [?]'s model to the other two. A private firm model was implemented excluding market variables from the hazard rate estimation, and compared to the public firm model with all the variables; it is their claim that the accounting variables add little predictive power when market variables are in the model. By them, Shumway [?]'s model was extended to contain industry effect dummies using four digit SIC codes, financial companies and monthly data; and all of them were found to be accurate and significant in bankruptcy prediction.

Hillegeist et al. [?] compare the two accounting based measures, Altman [?]'s Z score and Ohlson [?]'s O score, to the market based measure of Black-Scholes-Merton (BSM).[10] BSM-probability defined by him is indeed the Merton distance to default measure used in the upcoming studies. He runs a single model for each score variable, and for BSM-score on the following year's bankruptcies and finds out BSM-score has the highest R^2 among them. He applies discrete hazard rate specification in the models. The results disclose BSM-probability is more powerful than the score variables, that means Merton outperforms the Altman [?] and Ohlson [?] models. After Hillegeist et al. Bharath and Shumway [?] showed the prediction performance of distance to default is relatively robust to its estimation method, and

[10]The default event was defined as the initial filing of bankruptcy, and 756 bankrupt firms out of 14,303 companies listed after the 1978 bankruptcy law for the years 1980 to 2000 were extracted. Consistent with the literature, financial service firms were excluded to leave the total sample to 10,845 firms. Fama-French approach was administered for the industry classification.

they concluded distance to default is not sufficient as a default statistic but as a variable, useful as an explanatory variable to forecast default.

Campbell et al. [?] explores the corporate bankruptcy using a hazard rate model similar to the Shumway [?]'s model expanding the covariates.[11] Their model does outperform Altman [?] and Ohlson [?], and possess greater power than Shumway [?]'s model since it includes additional variation such as new accounting and macro variables. After constructing 10 consecutive portfolios from low to high probability of bankruptcy, they estimated the excess return, three and four factor Fama French model, and detected high risk of bankruptcy delivers low average returns.

Saretto [?] in his paper applies a time-varying duration model as well, to analyze the probability of default.[12] The default event, in his study, is the bond related default and any of the missing payments, covenant violations, restructuring or insolvency is considered default. Besides Altman [?]'s Z-score, Ohlson [?]'s O-score, Zmijewski [?]'s and Shumway [?]'s variables, coverage ratio and Tobin's Q are featured as independent variables. Furthermore, an error classification measure (ECM) was developed to show how costly incorrect specification is. Estimating Altman [?]'s model using discriminant analysis, Ohlson [?]'s, Zmijewski [?]'s and Shumway [?]'s model using a logit analysis, he demonstrates that duration model is more accurate with lowest ECM and outperforms the rest. Additionally, examining the relationship between default probability and Fama French distress factors, he displayed HML is indeed a distress factor in defaults.

Couderc and Renault [?] estimated a duration model with not logistic but exponential intensities for the 1981-2003 period across credit risk classes defined by S&P. Their analysis illustrated financial market variables have weak explanatory power while business cycle and credit market variables are key. In order to cover past shocks, lagged estimation for each covariate up to two years are employed, and to capture the default clustering within each industry, autoregressive conditional duration model is run.

[11] 796 bankruptcy events were identified for the years 1963 to 1998.

[12] He acquires the non financial firms with total assets more than $ 1 million for the years, after the introduction of new bankruptcy law, 1979 to 2000. He had 7,282 healthy and 319 bankrupt firms in his estimation, compared to 756 bankrupt firms out of 14,303 companies used by Hillegeist et al. [?]for the very same period.

Chava et al. [?] addressing the unobservable heterogeneity in the true state of the firm, introduce a latent random variable at the industry level into the model of default probability. 8 different models are derived by adding and subtracting various variables to Shumway [?]'s, Chava and Jarrow [?]'s and Duffie's specifications. It is confirmed that the models for default probability and recovery rate, though inversely related, depend on firm specific and macro economic variables, and the random effects improve the in, but not out of the sample performance.

Roszbach [?] models the decision to extend a loan and the survival time of loan using a bivariate tobit model using data on over 13,000 loan applications processed by a large Swedish bank in the mid-1990s. In his model the loans that are declined are non-observable, and the survival time is right censored only for "the good loans" at the end of sampling period. In other words, all of the loans censored, not defaulted within the sample period are assumed to be good loans, that will never default. Within the observable group of loans, censoring became the indicator of non-failure. This specification is just the opposite of hazard rate assumption, censoring and failure are two independent events.

Here, in this study, a mixture model is employed, and, logistic regression is set for the timing of default event, instead of the hazard rate model. The reason behind this alteration is a fundamental drawback in hazard rate modeling. Primarily it is developed for medical and biological sciences to address the censored data. Consider a study on the effectiveness of a new treatment for cancer, measured as the survival days. At the end of the study, some patients will die, some will survive entire period, however after the study the contact might be lost with some of the survivors. It is not preferred to ignore these censored observations since they essentially denote the success of the treatment. Survival models were developed to model the survival time in these kinds of experiments as standard statistical techniques cannot usually be applied because the underlying distribution is rarely normal and the data are often 'censored'. In order to overcome the censoring problem the way it is defined in biostatistics, censoring is assumed to be independent from the failure event. This is based on the assumption that all of the subjects will die in the long-run. It may be reasonable considering the vast application of the survival models on cancer patients, however in a financial framework, it means a total calamity, all companies filing for bankruptcy in the future. Even in the days of financial crisis right now, we do not believe it is a fair assumption. The financial history shows that there will always be a fraction of companies surviving, and mixture model captures the surviving fraction. Compared to the hazard rate model which assume all censored observations to fail and tobit model which assumes all censored observations will survive, our model proposes a mixture of failures and survivals. Further it corrects the problems aroused by heavy censoring. To expose its superiority, a standard hazard rate shall be run in comparison, using the independent variables shared in the literature.

The only application of mixture model in the financial distress framework is done by Yildirim [?] in his study of commercial real estate loans. Adopting a logistic regression for the intensity, he proposed a three-level mixture model for the 1,194 defaulted loans in the mortgage market for a 5-year period after 2000, the levels to be loan, property type, and region. The very mixture model below is single level.

The next sections explore the statistical models, the data and the results in order.

3 Model

3.1 Mixture Model

In the analysis of default probability, we are faced with a binary choice model where the dependent variable takes values 0 and 1. First issue to consider is that the company either defaults or survives in the period in which the data is taken, and second the default event occurs at a particular point in time. We believe that a set of factors gathered in the vector x explain the event of default, and another set of factors gathered in the vector z_t explain the timing of default. Note that the factors in x do not depend on time while the factors in z_t depend on time.

Then let's first define a binary random variable Y for the default event where $Y = 0$ states that the company is a long-term survisor, while $Y = 1$ states that the company shall eventually default, and the probability of default is given by $p(x) = P(Y = 1; x)$.

The traditional approach is to define a logit regression for the analysis. However as we have pointed out in the previous section more recently with the pioneering work of Shumway [?], existing literature concentrated on discrete time hazard (survival) models (Hillegeist et al. [?], Campbell et al. [?], Chava and Jarrow [?]). Shumway [?] pointed two deficiencies in the single-period models: Sample selection bias due to using only one, non-randomly selected observation per company and failure to benefit from time-varying variables to model bankruptcy. While in effort to correct these limitations, his proposition to use hazard rate/ survival model creates new short-comings.

The underlying assumption in survival models, without the correction for cure probability, is the assumption of eventual default. Classical survival models of cancer research propose that all the subjects in the experiment shall die eventually, and survival rate measures the days of survival before death. In a corporate bankruptcy framework, this proposition would be translated as the eventual bankruptcy of all companies which is not a practical assumption. Besides, it provides only an explanation on the timing of default not the event of the default.

Another dilemma with the survival/hazard rate models stem from the heavy censoring in the financial data. Even without considering the effect of duration, the default is a rare event; the percentage of defaulted companies within a year, or the percentage of defaulted companies over a period of time are always small values. Yet when monthly data is used, the bankruptcy percentage becomes trivial. As pointed out by Campbell et al. [?] the percentage of bankruptcy months becomes literally tiny compared to the other months.

In order to correct for these drawbacks, following Yildirim [?], we shall take advantage of another model popular in the biostatistics known as mixture cure models or long-term survival models. The basic idea in this model is to correct for eventual death of all subjects. The model assumes two groups of subjects, one of which shall never experience the default event (i.e. long-term survisors), while the other group shall eventually default. Moreover we shall be able to model the event and the timing of the default at the same time and correct for heavy censoring.

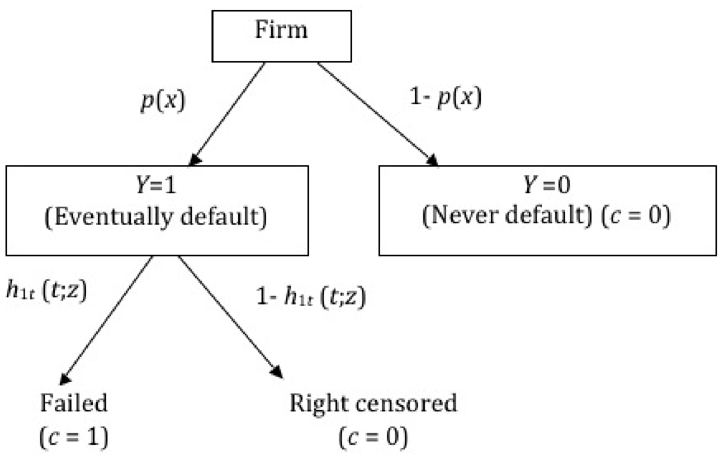
Similar to hazard rate models we shall define the survival rate and hazard rate for our mixture cure model. Beforehand we need to clarify the state of long-term survivors. A fraction of the companies in the data are long term survivors, and they shall never default. The rest shall eventually default. Therefore if the companies in our study do not default there are two possibilities: Either they are long-term survivors, or they shall default in the future but censored in the data. Note that Y is partially observed for right censored cases. Then define a censoring indicator c, where $c = 1$ stands for not censored, $c = 0$ stands for right censored cases and, as a result we shall have 3 states of the art.

We have defined a binary choice variable for the event of default, $Y = 1$ if company default and 0 otherwise, and a censoring indicator $c = 1$ if data is not censored and 0 otherwise. We will have three states based on Y and c as follows:

$c = 1$ & $Y = 1$	not censored,company defaults
$c = 0$ & $Y = 1$	censored,company will default eventually
$c = 0$ & $Y = 0$	censored,long-term survivor

Graphical representation of model is as below:

Figure 1:



The probability of default is p:

$$p = P(Y = 1; x), \tag{1}$$

$$1 - p = P(Y = 0; x), \tag{2}$$

Let τ be a random default time. It is defined only when $Y = 1$ and conditional probability density of it is:

$$f_1(t) : f(t; z|Y = 1) = \lim_{\Delta t \to 0} \frac{P(t \leq \tau < t + \Delta t|Y = 1)}{\Delta t}$$

Note that the company failure time τ is a discrete non-negative random variable, which refers to the month of default in our database and the subscript *1* stands for the condition of default.

Conditional survivor function describes the probability of the company to survive till its default at τ:

$$S_1(t) : S(t; z|Y = 1) = P(t < \tau|Y = 1)$$

Conditional hazard rate describes the probability of the company to default during the next time interval, given that it did not default before:

$$h_1(t) : h(t; z) = P(t = \tau|\tau \geq t, Y = 1) = \lim_{\Delta t \to 0} \frac{P(t \leq \tau < t + \Delta t|\tau \geq t, Y = 1)}{\Delta t} \tag{3}$$

We may relate the conditional probability density to the hazard rate in the following way:

$$f_1(t) = h_1(t) \prod_{j=1}^{t-1} (1 - h_1(j)) \tag{4}$$

Conditional survivor function can be presented in terms of conditional probability density and conditional hazard rate in the following way:

$$S_1(t) = Prob(t < \tau | Y = 1) = 1 - \sum_{j=1}^{\tau-1} f_1(t) \tag{5}$$

$$= \prod_{j=1}^{t-1} (1 - h_1(j)) \tag{6}$$

Then inserting $S_1(t)$ into the equation **??**, we would get;

$$f_1(t) = h_1(t) S_1(t) \tag{7}$$

However in our case, the unconditional survivor function, S_t, has to include the long-term survivors:

$$S_t = S(t; x; z) = (1 - p) + pS_1(t) \tag{8}$$

Therefore in our database, either the companies default, or they follow the above survivor function. For those experiencing the default, the unconditional probability density is given by

$$P(c = 1; x; z) = pf_1(t)$$

and for those being a long-term survivors, unconditional probability density is given by

$$\begin{aligned} P(c = 0; x; z) = S(t; x; z) &= P(Y = 0; x) + P(Y = 1; x) P(t < \tau; z | Y = 1) \\ &= (1 - p) + pS_1(t). \end{aligned}$$

Then the likelihood function of the mixture of long-term survival and eventual default of i^{th} firm at time t can be written as:

$$L_i = [p_i f_{1i}(t)]^c [(1 - p_i) + p_i S_{1i}(t)]^{1-c} \tag{9}$$

where functions p_i,$f_{1i}(t)$ and $S_{1i}(t)$ are p,$f_1(t)$ and $S_1(t)$ for individual firms.

The contribution of each state to the likelihood function is as follows:

$c = 1$ & $Y = 1$	company defaults with density	$p_i f_{1i}(t)$
$c = 0$ & $Y = 1$	company will default but survived in the sample	$p_i S_{1i}(t)$
$c = 0$ & $Y = 0$	long-term survivor	$(1 - p_i)$

If we assume Y is known, the individual likelihood function would be (by substituting equation **??** in):

$$L_i = [(p_i h_{1i}(t) S_{1i}(t))^y]^c [(1 - p_i)^{1-y} (p_i S_{1i}(t))^y]^{1-c} \tag{10}$$

We would get the simpler form of our likelihood equation after rearranging terms:

$$L_i = p_i^y (1 - p_i)^{1-y} h_{1i}(t)^c (1 - h_{1i}(t))^{y-c} S_{1i}(t)^y \tag{11}$$

Note that we had assumed Y as given, while in reality we can only observe the censoring indicator, c. EM-algorithm will serve us to overcome this obstacle.

EM is an iterative maximization algorithm composed of 2 steps. In the first step E-Expectation, the likelihood equation shall be estimated with the best guess of incomplete data and the expectation vector for the incomplete variable is formed. In the second step M-Maximization, the parameters shall be estimated using the expected values for the incomplete variable. These steps shall be repeated by inserting estimated parameters back into the equation and maximizing until convergence achieved.
In our case we shall start by making our best guess to calculate $E(Y_i)$:

$$E(Y) = \begin{cases} 1 & \text{if } c = 1, \\ \frac{p_i S_{1i}(t)}{p_i S_{1i}(t) + (1 - p_i)} & \text{if } c = 0. \end{cases}$$

After substituting expected values into our likelihood function we shall have:

$$L_i = p_i^{E(Y)}(1-p_i)^{1-E(Y)} h_{1i}(t)^c (1-h_{1i}(t))^{(E(Y)-c)} S_{1i}(t)^{E(Y)} \tag{12}$$

Then in the maximization step, the above likelihood function would be maximized. During the search for convergence these two steps would be repeated, however the expectation steps shall not follow the above condition but extract the expected values from the previous maximization step.

Until now, we have not specified a distribution function for neither the event of default nor the timing of default. In a parametric estimation, which we shall follow in our study, a specific distribution would be outlined for both of these events. We shall specify a logistic distribution for the default event and for the time of default. Then the probability that an individual company shall face the event of default depends on the covariate vector x through a logistic distribution:

$$p_i = \frac{e^{\beta' x_i}}{1+e^{\beta' x_i}} \tag{13}$$

And the conditional hazard rate for an individual company depends on the time varying covariates z_t through a logistic regression:

$$h_{1i}(t) = \frac{e^{\beta' z_{it}}}{1+e^{\beta' z_{it}}} \tag{14}$$

With these last specifications our mixture model is done.

3.2 Hazard Rate Model

For the sake of completeness though, we would compare our model with the popular hazard rate models in the literature. The likelihood function of hazard rate models to replace the equation (10) in our model in case of a model change is :

$$
\begin{aligned}
L_i &= [h_{it}]^y \left[\prod_{j=1}^{t}(1-h_{ij})\right]^{1-y} \\
&= [h_{it}]^y [S_{it}]^{1-y}
\end{aligned}
$$

where h_{it} stands for the unconditional hazard rate of defaulting companies, S_{it} stands for unconditional survival of survivor companies and y stands for the default event.

The distribution function specified for the conditional hazard rate in these models is generally the logistic distribution:

$$
h_{it} = \frac{e^{\beta' z_{it}}}{1 + e^{\beta' z_{it}}}
$$

Notice that, since standard hazard rate models do not model the event but only the timing of default, no distribution would be needed for the default probability.

4 Data

The CRSP dataset originally contains 24,147 publicly traded companies from the period 1980-2007. 1,797 of the companies in our sample have gone bankrupt (7.44%) during the selected period. Monthly observations for each company make a total of 2,447,813 observations in the initial dataset.

We have not considered the time frame before 1980. First reason is to eliminate the bias formed in the data by the Bankruptcy Reform Act of 1978 which took effect on October 1, 1979. Besides, since during the period from World War II through the 1970s, with the exception of railroad failures, bankruptcy was not a major issue in the US; we believe this exclusion shall not affect our research. During the 1970s, there were only two corporate bankruptcies of prominence, Penn Central Transportation Corporation in 1970 and W.T. Grant Company in 1975. However during the 1980s, 1990s and early 2000s record numbers of private and public bankruptcies, of all types, were filed. Table **??** presents the largest ten bankruptcies of the US history.

The first reduction in the data occurred during the merge of CRSP and Compustat databases[13]. Then we have cleaned the missing variables, and we have omitted the companies which were listed in different industries during our selection period. These changes lowered the total number of companies to 18,505 and the number of bankruptcies to 1,640 (8.14%). Besides the missing values, we have corrected for outliers. They were truncated at 99% level. Table **??** and **??** provide the layout of the dataset with respect to the industries and years[14].

Following the convention, we shall use bankruptcy filings for default event. Bankruptcy data is pulled from the US Bankruptcy Courts. Whenever a company files in any federal bankruptcy district, it is recorded as a bankrupt entity. The source of the bankruptcy data is BankruptcyData.com[15] database. Our original database contains a total of 2,659[16] public company bankruptcies for the selected period. The database provides the Chapter 11 (reorganization) and Chapter 7 (liquidation) filings for the companies. In some cases the same company may file for more than one reorganization or both reorganization and liquidation; in those cases we have accepted the first filing as the event of default.

The firm level accounting data has been extracted from quarterly COMPUSTAT files, while the market data has been taken from CRSP database. In order to make sure the estimation covers the data that is available to market participants at the specified time, we have lagged all the corporate data by one quarter.

We have used the following set of covariates in our analysis:

[13] After the merge there were 21,980 companies left.

[14] Figure **??** and **??** at the end of the paper is based on the data in these tables.

[15] BankruptcyData.com is the premier business bankruptcy resource on the market which contains over 400.000 bankruptcies of private and public firms.

[16] Bankruptcy.com database provide the company names and their 4-digit SIC codes. We have merged CRSP/Compustat database and bankruptcy data through company names of the same industry group. Out of 2,659 bankruptcies 1,797 was found in the CRSP/Compustat dataset.

4.1 Firm Specific Covariates

4.1.1 Accounting Variables

All of the accounting variables are calculated with the data from Compustat, except MVETL. The data used to calculate market value of equity is extracted from CRSP. The variables that are employed by Altman [?] in his Z-score are as follows:

WC/TA: Working capital to total assets ratio is a liquidity measure which shows how easily the firm can lay its hands on cash. Both working capital and total asset values are balance sheet values.

RE/TA: Retained earnings to total assets ratio is a cumulative profitability measure which shows the company's ability to accumulate earnings using its total assets. It also measures the leverage of the company. Firms with higher *RE/TA* ratio have financed their assets through profits not debt.

EBIT/TA: Earnings before interest and taxes to total assets is a productivity measure independent of tax and debt factors. *EBIT* is often known as an approximation of cash from operations. The ratio thus, shows the ability of the firm to operate its assets to generate earnings, in other words shows the productivity of assets.

MVE/TL: Market value of equity to book value of total liabilities adds market dimension to the analysis. A higher market value of equity known as market capitalization of a firm compared to its outstanding obligations indicates the market's belief in its solid financial position.

S/TA: Sales to total assets is an efficiency ratio known as the asset turnover ratio. It shows how efficiently the company activates its assets to generate sales.

Z-Score: The Z-Score is calculated based on the original calculation proposed by Altman [?] to measure the financial health of the company in our analysis. As expected healthy companies demonstrate superior z-scores. (Table **??** It is a combination of the five financial ratios stated above, based on the weights of multiple discriminant analysis.

The variables that are introduced by Zmijewski [?] to the bankruptcy literature:

TL/TA: Total liabilities to total assets is the well known leverage ratio named debt ratio which measures how much the company rely on debt to finance its assets.

CA/CL: Current assets to current liabilities is the well known liquidity ratio named current ratio which measures the company's ability to meet its short term debt obligations.

NI/TA: Net income to total assets is the famous profitability ratio known as return on assets. It tells what earnings are generated from invested capital.

The variable that is introduced by Merton [?] to the bankruptcy literature:

DD: Distance to default is a measure of the difference between the asset value of the firm and the face value of its debt scaled by the standard deviation of the firm's asset value. It is used as a practitioner model by Moody's KMV yet it is originated from Merton's [?] structural default model. It enters to the recent literature under the name of "distance to default" as done by Campbell et al. [?], Bharath and Shumway [?], Das et al. [?], Duffie et al. [?], Vassalou and Xing [?], and Da and Gao [?]. Hillegesit et al. [?], on the other hand named a similar measure as "BSM-probability".

Merton [?] defines payoffs to equity as a call option on the firm's total asset value, since equity holders have limited liability for the debt payments in case of a bankruptcy. He defines the face value of the firm's liabilities as the strike price of the call option and the option expires at time T when the debt matures. At time T, two scenarios possible: If the assets are higher than the liabilities of the company, the shareholders will exercise their option and their pay off would be what is left after paying the debtholders. Otherwise, they will let the option expire, in other words, the firm files for bankruptcy and the payoff to the shareholders is zero. Then, in Merton's theory, value of the equity follows the following European call option framework:

$$
\begin{aligned}
E &= A\phi(d_1) - e^{-rT}D\phi(d_2) \\
d_1 &= \frac{ln(A/D) + (r + \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}} \\
d_2 &= d_1 - \sigma_A\sqrt{T}
\end{aligned}
$$

where E is the market value of the equity, A is the total value of assets, $\phi()$ denotes the standard normal distribution function, r is the risk-free interest rate, T is the expiration date, D is the face value of debt, and σ_A is the asset volatility of the firm.

Distance to default is, then, defined as,

$$
DD = \frac{ln(A/D) + (\mu_A - \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}}
$$

where μ_A is the expected value for the return on the firm's assets.

Our construction of the distance to default variable is along the lines used by Hillegeist et al. [?], Vassalou and Xing [?] and Campbell et al. [?] We have defined E as the market value of equity, and A as the market value of assets, which is the total of book value of the liabilities and the market value of the equity. For the strike price, face value of debt, D, we use the KMV definition of short-term liabilities plus one half of long-term liabilities. This convention is also used by many scholars because it is a simple way to take account of the fact that long-term debt may not mature until after the horizon of the distance to default calculation. Risk-free rate is the monthly 1-year treasury bill rate as Hillegeist et al. [?] and as Shumway and Bharath [?] did, because the expiration time, T, is defined as one year. For μ_A, we have implemented the empirical proxy, 0.6, used by Campbell et al. [?] suggesting a common expected return for all stocks is better than a noisily estimated stock-specific number offered by other scholars.

4.1.2 Market Variables

Market variables are calculated from the CRSP data items. The market variables defined by Shumway [?] are as follows:

EXRET: Excess return is measured as the difference between the real market return on the firm and the value-weighted CRSP NYSE/AMEX index return.

SIGMA: Sigma is the standard deviation of each firm's stock returns. We have followed Shumway [**?**] in the calculation. Each stock's monthly returns in year $t-1$ are regressed on the value- weighted NYSE/AMEX index return for the same year. The firm's sigma for year t is the standard deviation of the residual of this regression.

RSIZE: Relative size is calculated as the logarithm of each firm's equity value measured by the market capitalization divided by the total NYSE/AMEX market equity value.

Firm's equity value for a month is its stock price times the number of shares outstanding at the end of the month. The stock price used is the last non-missing closing price of the security for the last trading day of the month. If unavailable, it is replaced with a bid/ask average.

4.2 Industry Specific Covariates

Fixed Effects: We have defined four industry groups based on company SIC codes. Two industries, manufacturing and finance, received special treatment in the earlier studies. Manufacturing sector happened to be the center of the attention for the default analysis as Altman [**?**] used only manufacturing companies in his Z-score calculation. On the other hand financial sector used to be left out. SIC codes and related industries are as follows:

SIC Code	Industry Name
< 1000	Agriculture, Forestry and Fisheries
$1000-1500$	Mineral Industries
$1500-1800$	Construction Industries
$2000-4000$	Manufacturing
$4000-5000$	Transportation, Communications, and Utilities
$5000-5200$	Wholesale Trade
$5200-6000$	Retail Trade
$6000-6800$	Finance, Insurance, and Real Estate
$7000-8900$	Service Industries
$9100-10000$	Public Administration

Based on Chava and Jarrow's [?] industry classification we have converted these SIC codes into four industry groups, manufacturing and finance being two of them.

Industry Code	Industry
1	Agriculture,Forestry and Fisheries,Construction, Wholesale and Retail Trade, Service Industries
2	Manufacturing and Mineral Industries
3	Transportation, Communication and Utilities
4	Finance, Insurance and Real Estate

4.3 Market Specific Covariates

$\triangle$ *UR:* The percentage change in the unemployment rate from the previous year is used to capture the impact of macroeconomic fluctuations. Unemployment rate is the percentage of the labor force that is not employed. The population is defined as 16 years and older civilian non institutional population. The data is extracted from Census Bureau.

$lag\triangle$ *G:* One year lag of the percentage change in the government spending from the last year is included in the model as a proxy for the stimulus packages. Current stimulus packages are specific to the recession, however government spending, being one of the major fiscal policy tools, in general aims to stipulate the economy. This variable displays if ever the government spending can reduce the bankruptcies. The data is from the Bureau of Economic Analysis.

The descriptive statistics for the covariates are presented in Table **??**.

5 Results

5.1 In-Sample Prediction

Initially, named Model-1 and Model-2, two separate specifications for the covariate vectors have been identified for the mixture model. In the first specification, Model 1, the event of default covariates are leverage measured by TL/TA, profitability measured by NI/TA and solvency measured by MVE/TL[17]. The variables in this part, although time-varying originally, are taken fixed at t_0 based on the mixture model specification. How informative today's knowledge with respect to eventual default of the company is explained in this part. Therefore we have considered its leverage, the debt burden of the company, its profitability and its ability to pay relevant. The timing of default covariates are, in addition to those in the first part, liquidity measured by WC/TA, productivity measured by EBIT/TA, efficiency measured by S/TA, excess return, volatility measured by SIGMA, the relative size, unemployment rate, government spending and the industry indicators.

In the other mixture model specification, Model-2, instead of the leverage, profitability and solvency, only Z-Score, for being a combination of financial ratios, is used in the first part and in the second part the leverage and solvency are omitted with the inclusion of Z-score, leaving liquidity, productivity, efficiency, excess return, volatility, relative size, unemployment rate, government spending and industry indicators intact.

The whole set of estimates turn out to be significant at 5 % level for the both models. Table **??** presents the two models.

Before analyzing the results, our expectations for the event and timing of default ought to be discussed. In the event part, leverage is supposed to have an inclining effect over the probability of default, on the other hand the case is vice versa for profitability, solvency and z-score, which measures the financial health.

In the intensity part, the timing of default gets sooner when the company has higher leverage, implying a positive sign on leverage; it gets later when the company is healthier, more productive, more solvent, more efficient, more liquid and profitable, implying a negative sign for the rest of accounting variables. For the market variables, excess return and relative size supposed to be inversely related to the default intensity, notwithstanding default intensity would be higher during more volatile circumstances. As for the macro variable unemployment rate, severe macroeconomic conditions are anticipated to raise the default probability, while

[17] We have checked for the collinearity between TL/TA and MVETL, and we did not find a high correlation. The variance inflation factor (VIF) is 1.89, way below the cut-off value of 10.

government spending would eliminate the defaults. Table **??** summarizes the expectations.

Table **??** presents the coefficients for both mixture models, the absolute value of z statistics are parenthesized. The signs in both models suit the expectations excluding solvency in the first part and the leverage in the second part of Model 1 and relative size in Model 2. With 7.38% estimated default probability, both are good predictors for the real default rate of 8.14% (Table **??**), Model-1 is selected for comparison to the hazard rate model because of its higher log-likelihood[18], lower AIC[19] and higher $pseudo - R^2$ [20] values.

Before getting into the comparison, the coefficients on Model 1 lay out the forces behind the default of a company. Productivity (-6.66) and the profitability (-15.59) with the highest coefficients proved the virtue of macroeconomic theory. It is only by being more productive, the business can survive in the competitive markets, and relieves from distress. Efficiency (-0.70) and liquidity (-0.73) also play a critical role besides the firm's ability to beat the market (excess return:-1.86). However, size has a limited impact compared to the rest (0.06). As we have expected, a higher unemployment rate, signaling the vulnerability of the economy increases the likelihood of the company to fail. (3.13) Anticipated, yet the most surprising result is coming from the government spending (-21.19): Federal and local government incentives and expenditures can change a company's fate, and reduce the default probability significantly. Since the relationship of dependent variable bankruptcy and the independent variables are not a linear one, it is essential to note that the coefficients do not refer to constant increases in the probability of default. The exponential values of the coefficients would refer to the changes in the odds of default as the probabilities defined to follow a logistic distribution. Therefore we avoid to make statements regarding the numerical change in default probability when we mention the coefficients of the variables, rather we compare the coefficients with each other to understand the relative importance of the covariates. Yet in order to

[18]Probabilities are always less than one, so log likelihoods are negative

[19]AIC is computed as $AIC = \frac{(-2lnL+2k)}{N}$ where lnL is the overall likelihood reported by the regression, k is the number of parameters of the model and N is the number of observations.

[20]The $pseudo - R^2$s reported in the table are estimated as the square of the correlation between the observed response and the predicted response based on the calculation suggested by Nicholas J. Cox. [**?**]

understand the impact of each variable on the default probability for our model, we will look at the change in the odds of going bankrupt vs. staying in the market by a unit change in each covariate. As can be seen while some covariates increases the odd of bankruptcy, some will reduce it. Our key variables productivity, profitability and government aids totally rescue the company from bankruptcy.

	Odds of Default/ Non-Default	Odds of Non-Default/ Default		Odds of Default/ Non-Default	Odds of Non-Default/ Default
The event of default			*The time of default*		
Accounting var			*Market var*		
TLTA	3.48		EXRET		5.26
NITA		7.14	SIGMA	5.32	
MVETL	1.04		RSIZE	1.06	
The time of default			*Macro var*		
Accounting var			$\triangle$ UR	22.93	
TLTA	2.76		lag $\triangle$ G	0	∞
NITA	0	∞	*Industry var*		
MVETL		1.39	I1		25.00
WCTA		2.08	I2		25.00
EBITTA	0	∞	I3		33.33
STA		2.00	I4		50.00

Model 1 analyzing the event and timing of default at the same time offers a more inclusive result. We have defined whether the company file for bankruptcy or not, independent from the timing, depends on the debt payability of the company. How levered it is, how profitable to cover its debt and how solvent it is. Model 1 says that 1-unit increase in leverage increases the odds of bankruptcy 3.48 times, while, 1-unit increase in the profitability will increase its likelihood to stay healthy 7.14 times. Any change in solvency turns out to keep the odds of going bankrupt almost constant.

With respect to the timing of the default, the leverage increase the default intensity 2.76 times, while sigma raise it 5.32 times. The increase in unemployment rate, because of the recessive reflections at the macro level will push up the odds of failure by 22.93 times. On the other hand the improvement in solvency, by 1.39 times, in liquidity, by 2.08 times and in excess return, by 5.26 times improves the odds of staying healthy. The key ingredients in this combination are the productivity (EBIT/TA), profitability (NI/TA) and the government stimulus (lag $\triangle$ G). A one-unit increase in any of these will reduce the default probability to zero.

Industry dummies suggest that overall market movement is in the direction of reducing bankruptcy and developing a healthier environment for the companies, however out of all sectors, financial sector, being the backbone for the market, provides more protection against bankruptcies.any company in the financial sector is less likely to default compared to rest of the companies. One reason maybe the government protection for its majority, which can also be observed in today's economic crisis.

The comparison models are named as hazard rate, in the literature it may be seen as hazard, survival, duration or multi period logit models; the intensity of default assumed to have logistic distribution. Table **??** displays five models side by side, first two being the mixture models, the last three being the hazard rate models. The second standard hazard rate model (HR2), is exact the same statistical specification of Shumway [**?**] and the third hazard rate model (HR3) is Chava and Jarrow's [**?**] model which adds dummy variables to Shumway's specification, yet did not change the sign and magnitude of original coefficients significantly. Shumway's model is improved in the literature to include other variables and/or dummy variables. The fourth hazard rate model (HR4) adds z-score. The first hazard rate model (HR1) in Table **??** contains all of the covariates in Model 1 including dummies. Although the statistical specification is exactly the same, with the inclusion of financial companies to the dataset, the coverage of the years after 1992 and the selection of significant covariates, HR1 is already improved with respect to Shumway's model (HR2) and Chava and Jarrow's model (HR3). The $Pseudo - R^2$s delivered by the Shumway's model (HR2), and Chava and Jarrow's model (HR3), and HR4 models are less than HR1's, indicating the lack of information embedded.

The first advantage of mixture model over hazard rate model is its ability to measure the event and timing of the default at the same time. The results for Model 1 stresses the impact of the leverage on the default probability. The increase in the leverage ratio (TL/TA) by 1 unit produces an increase of 1.02[21] in the logarithm of default probability. Solvency has a moderately lower stimulus on the default rate, though positive contrary to negative expectations with 0.08[22] logarithmic increase

[21] From Table **??**. Related odds ratio is 3.48. Likelihood of default goes up by 3.48, by one-unit change in leverage ratio.

[22] From Table **??**. Related odds ratio is 1.04.

in default rate by one-unit increase in solvency ratio (MVE/TL). Yet the effect of profitability (NI/TA) is negative, by increasing the odds of health 7.14 times.

Table **??** presents the differences between Model 1 and HR1. Model 1 proposes a lower weight on leverage (1.02 in Model 1 vs. 2.51 in HR1) and profitability (-15.59 in Model 1 vs. -21.01 in HR1), and a higher weight on liquidity (-0.73 in Model 1 vs. -0.21 in HR1), productivity (-6.66 in Model 1 vs. -2.12 in HR1) and efficiency (-0.70 in Model 1 vs. -0.16 in HR1) while solvency (-0.33 in Model 1 vs. -0.33 in HR1) has the same effect in both. With respect to market variables, in HR1 model excess return (-1.68 in Model 1 vs. -1.95 in HR1) and sigma (1.67 in Model 1 vs. 5.85 in HR1) are more stressed, yet relative size (0.06 in Model 1 vs. -0.17 in HR1) has a contradictory impact, while HR1 model states smaller companies are more likely to default, Model 1 proposes the opposite. Instead of market variables, Model 1 model give more importance to macro variables, the change in unemployment rate (3.13 in Model 1 vs. 1.96 in HR1) and the government spending (-21.19 in Model 1 vs. -10.41 in HR1).

Figure **??** compares the model fit for the two models: Model 1 and HR1. The real annual frequencies of bankruptcies are plotted by the blue line. The red line plots the annual frequencies estimated by the first model, while the green line shows the estimation of HR1. The graph presents Model 1 as a better fit. The average for the real bankruptcies is 0.85% annually, while the same average for Model 1 is 0.95%, it is 1.01% for HR1, displaying the overestimation by the hazard rate model.

Another contribution of our model is the introduction of unemployment rate as macro variable to the default rate modeling. Macro economic variation affects the default intensity, as a proxy to the economic fluctuations an upward trend in unemployment rate considered to be one of the leading indicators of economic slow-downs. Therefore a positive change in unemployment expected to increase the default intensity, and it happened to be so. One percent rise in unemployment rate is found to augment the logarithm of default intensity by 3.13.

In majority of earlier studies a special care is given to the industry clusters either through industry dummies or frailties, industry dummies are used in our model as well. The coefficients of the four industry dummies are significantly negative. The negative coefficients of all dummies imply the market-wide soundness: if everything else set to zero (no variables included in the model), the likelihood of bankruptcy is going down for the entire set of companies. It is noteworthy that the coefficient on the dummy for industry group 4, finance, insurance and real estate sector, is much larger than the rest in absolute value, proposing financial companies are less likely to default.

As clearly seen in the table, in HR models, the introduction of z-score and distance to default to the Shumway's model improves the pseudo-R^2 slightly, yet it is still far from the mixture model specification, the the pseudo-R^2 suggests the superiority of our model over HR.

Following, we have formed 10 portfolios ranking the annualized estimated probabilities for 2003. Having roughly 641 companies in each portfolio, 6,407 was the total number of companies in 2003. 1^{st} portfolio containing the riskiest 641 firms, from portfolio 1 to 10, the default probability was decreasing. Table **??** presents the distribution of actual failures within our in-sample prediction period. It is in-sample, because even though the companies are sorted with respect to their estimated default probabilities of 2003, the whole sample of 1980-2007 is used for estimations. We will use a restricted sample in out-of-sample prediction part. In 2003 there were 84 failures, in 2004 there were 32 failures, in 2005 there were 35 failures, in 2006 there were 22 failures, and in 2007 there were 55 failures where 13 of which were bankruptcies of new companies. We have observed all of the 215 companies to bankrupt in the five years were in our 1^{st} portfolio. Since there is no way to include the companies that are not established yet, it is fair to state our model predicts all bankruptcies over four year period. In order to compare the predictive power of our model to the hazard rate model, from the same sample year with the same companies, we have formed another set of 10 portfolios using the annualized default rates estimated by hazard rate models this time. 1^{st} portfolio would contain the riskiest 10% of the companies again, but the risk is measured by hazard rate models. Table **??** presents the distribution of failures among the sets of portfolios. Panel A is for Model-1 of Table **??**, while Panel B shows the estimation with HR1 of Table **??** and Panel C displays HR2- Shumway's model of Table **??**. 1^{st} portfolio of HR1 can identify 109 failures while 1^{st} portfolio of Shumway's model identifies 199 failures out of 215 failures. Thus our model outperforms that of Shumway [**?**].

Next, in order to clarify possible concerns regarding the impact of financial companies on our model, we have estimated Model-1 for non-financial companies first and then for only financial companies. We have not defined sub-industry groups within financial companies, therefore the mixture model on financial companies does not contain the industry dummies, while the model on non-financial companies has three industry indicators without financials. Table ?? displays the results. It is clearly seen that the inclusion of financial companies do not present any bias to our estimation.

Lastly, we have analyzed the contribution of distance to default to hazard rate and mixture models. Originated from Merton's [?] structural default model, distance to default is widely used in the market. Bharath and Shumway [?] analyze the effect of two different distance to default measures by estimating different hazard rate models by adding extra variables besides distance to default measures and found the distance to default measure to be helpful besides other predictors not alone. Campbell et al. [?] estimated a DD-only model and DD included full model and concluded that distance to default does not improve his model at all. In our estimation, DD is significant yet does not improve the model significantly consistent with the works mentioned.

We have run several models including distance to default as a predictor of failure. Table ?? presents the results from those models. First we ran a mixture model with a constant and distance to default as the only predictor of both event and timing of default. ((1) in Table ??) Then we added industry dummies to the timing part ((2) in Table ??), and then we ran mixture model 1 of Table ?? by adding distance to default on both parts of the model ((3) in Table ??). We followed the same order for hazard rate models. First we ran a multi-period logistic regression on a constant and distance to default only. ((4) in Table ??) Then we added industry dummies ((5) in Table ??) to this model, and then we ran HR1 of Table ?? by adding distance to default ((6) in Table ??). We have added distance to default to Shumway's model [?] ((7) in Table ??) and Chava and Jarrow's model [?]((8) in Table ??). The results shows that distance to default is mostly a significant predictor of failure, however it does not improve the existing models much with respect to model fitting. Especially it barely contributes to our model, so we did not see any virtue to add distance to default to our model specification.

5.2 Out-of-Sample Prediction

In order to measure the out of sample prediction power of our model, we have estimated the conditional probability of default in 6 months, 1, 2 and 3 years using the lagged values of our covariates as Campbell et al. [?] did. In particular, the conditional hazard rate of an individual company defined to depend on the covariates at j months ago:

$$h_{1i}(t) = \frac{e^{\beta' z_{i,t-j}}}{1 + e^{\beta' z_{i,t-j}}} \tag{15}$$

The lagged values are calculated at 6 months, 1 year, 2 years and 3 years; j being 6, 12, 24 and 36 in estimation. The probability of eventual default did not change since it did not depend on time at the first place:

$$p_i = \frac{e^{\beta' x_i}}{1 + e^{\beta' x_i}} \tag{16}$$

For each estimation the same sample period, 1980-2003, has been used. In the 6 months forecast, the failures starting from the second half of 1980 to the end of first half of 2003 are predicted using the 1980-2003 sample. In one year forecast, the failures from 1981 to 2004 are estimated by the predictors of 1980 to 2003. In two year forecast the failures of 1982 to 2005 are estimated by the predictors of 1980 to 2003 and lastly the failures of 1983 to 2006 are estimated by the predictors of 1980 to 2003. Table **??** presents the regression results.

The reason behind the improvement in the loglikelihood as the lag months increase is the reduction in the sample size. Akaike information criteria (AIC) and $Pseudo-R^2$ would be better criteria to compare the models. As the forecast covers longer periods, the model slightly loses its informative power, shown by rising AIC values, and lower $Pseudo-R^2$'s. It is still powerful even after three years. With respect to AIC zero-month model is the best fit for the data. The six month prediction has a higher $Pseudo-R^2$ than our original estimation, indicating the time lag in the information flow from the variables to default event. This is reasonable considering most of our variables are accounting and macro variables, which are likely to affect later. Notice that even in the 3 year forecast most of the covariates are still significant proving the out-of-sample prediction power of our model.

Another way frequently used in the industry to measure the predictive power of a default model is to graph a 'power curve' or a CAP (Cumulative Accuracy Profile) curve comparing the model's prediction of default rate from 0% to 100% on the x-axis to the cumulative observations of default event from 0% to 1% on the y-axis. To obtain the power curve, all firms are first ordered by their respective scores from riskiest to safest. For a given fraction x of the total number of firms, the power curve is constructed by calculating the percentage $d(x)$ of the defaulters whose default rates are equal to or lower than the maximum score of fraction x. This is done for x ranging from 0% to 100%. The curve is expected to be concave with the observed defaults concentrated among those firms with the highest predicted default rate. A model with no predictive power would produce a 45° straight line from 0% to 100% default rate; in that case, actual defaults would be evenly distributed over predicted default frequencies. A model with full predictive power would produce a straight line up on the y-axis bending 90° at 100% and stays at 100% to the end. Panel A in Figure **??** graphs one-year, and two-year power curves estimated in 2003 for our model compared to Shumway's hazard rate model. Panel B in the same figure graphs one-year, and two-year power curves estimated in 1998 for our model compared to Shumway's hazard rate model. On the horizontal axis the predicted default rate of the companies is given sorted from the riskiest to the safest, while on the vertical axis the realized defaults of the same companies corresponding the order on the x-axis is give. So the first graph in Panel A of the Figure **??** can be read as following: The riskiest 10% estimated by both models depicted by Model-1 and Shumway curves cover 70% of the firms that actually defaulted in the subsequent one-year period. Likewise the other power curve in Panel A shows the subsequent two-year period. Panel B gives the same curves estimated in 1998. With respect to power curves, our model is close to Shumway's model. Shumway's model performs slightly better in the one-year curves, while our model performs better than Shumway in the two-year prediction.

In Figure **??**, we have presented the estimated default probabilities for Lehman Brothers and Washington Mutual. Recently in 2008, these two companies have filed for the two largest bankruptcies of all times.[23] Our estimation extends through the end of 2007, therefore Lehman Brothers and Washington Mutual are in the dataset

[23]Please see Table **??**

as healthy companies for our estimation, however the graphs display the upward movement in the default probability for both until 2007. They are also listed in the 1^{st} portfolio of Model 1, constructed in 2003, as the riskiest 5% of all companies in the market.

We have the bankruptcy data up until 2008, however the CRSP database did not provide the market variables for the last year at the time of our research. In the dataset from Bankruptcy.com, there are 90 companies [24] that filed for bankruptcy in 2008. 65 of them could be identified among the companies listed in the dataset of 2007. Because of the lack of market information, we have not employed full range of bankruptcies in Model 1 estimation. In order to utilize 2008 data of bankruptcies, we have forecasted the annual default rates for two years into the future, 2008 and 2009, using Model 1, and compared the 2008 prediction with the actual default percentage of the year. Figure **??** plots the curves for the Model 1 forecast and the real data. Model 1 forecasts 1.22% of the companies would default in 2008, the realized rate is 1.09% [25] . For 2009, it estimates 1.35% of the companies will default, it means approximately 80 companies[26] will file for bankruptcy.

5.3 Risk and Return Analysis

Estimated the default rates and analyzed the predictive power of our model, we shall now focus on the returns offered by various risk groups. Our bankruptcy model predicts the upcoming failures entirely up to three years; then the next question is, what would be the investment implications of our predictions? What would be the return of an investor investing based on our model? In order to analyze the return structure, we will reconstruct portfolios. Using the dataset for the period 1980-2003, we have estimated Model 1 model and the associated annual default probability for each company. For the last year in the dataset, 2003, we have sorted the annualized default probabilities of the companies from highest to the lowest. For the bankruptcy prediction we had ten portfolios of equal number of companies in

[24] In the original dataset there were 140 companies that filed for bankruptcy, however 90 of them can be identified in the CRSP/Compustat merged database.

[25] 65 is 1.09% of 5,940 companies. 5,940 is the total number of companies in our dataset for 2007 listed in Table **??**

[26] 1.35% of 5,940 companies. 5,940 is the total number of companies in our dataset for 2007 listed in Table **??**

each; now we shall form ten portfolios again, however the number of companies will not be even, more stress will be given to the tails instead of the center. Portfolios are still ordered according to their default probability, 1^{st} portfolio containing the riskiest companies, 10^{th} portfolio containing the safest. The average relative size[27], z-score values and excess returns[28] are calculated below for the ten portfolios.

Portfolios of 2003						
Portfolio	# of Companies	Percentage	Relative Size	Z-Score	Distance to Default	Excess Return in 2003
1	320	4.99 %	-9.87	1.20	0.36 %	-1.98 %
2	320	4.99 %	-12.29	1.29	0.26 %	-0.69 %
3	641	10.00 %	-11.72	1.29	0.28 %	0.66 %
4	641	10.00 %	-10.95	1.28	0.32 %	0.53 %
5	1,282	20.01 %	-10.88	1.77	0.34 %	1.28 %
6	1,282	20.01 %	-10.98	2.48	0.39 %	1.13 %
7	641	10.00 %	-11.13	2.93	0.37 %	1.49 %
8	641	10.00 %	-11.28	3.80	0.34 %	1.48 %
9	320	4.99 %	-11.22	4.74	0.35 %	2.44 %
10	319	4.98 %	-10.87	6.16	0.39 %	1.86 %
Total	6,407	100 %	-11.12	2.69	0.34 %	0.82 %

The first portfolio, being the highest risk portfolio of all, contains relatively smaller companies. Altman's Z-score increases as the risk goes down, displaying the increasing financial health in the portfolios, however for couple portfolios it presents a confusion, suggesting the fourth portfolio is riskier than the second and the third. Since our model proved to predict upcoming defaults entirely, this confusion signals the deficiency of Z-score in bankruptcy prediction. Distance to default on the other hand is far from being intuitive in any ways. Our new portfolio specification gets thinner at the tails and fatter at the center. It is now easier to track the returns on the tails.

The excess returns follows an upward trend with respect to the portfolios since the estimated default rate is reduced. Figure **??** displays average excess returns for each portfolio in 2003. The excess return makes a jump on the left tail, however on the right tail, there is a slight difference in excess returns. The first portfolio with the riskiest 5% of the companies, causing a loss to its holders, makes 1.98% below

[27]Relative size is defined as the logarithm of each firm's equity value measured by the market capitalization divided by the total NYSE/AMEX market equity value.

[28]Excess return is defined as the difference between the real market return on the firm and the value-weighted CRSP NYSE/AMEX index return.

the market return on average. The return on the next portfolio improves by roughly 1.3%, yet on average it is still costly compared to the market. From the second to the third the change is around 1.2%. On the other hand, the gain on the safest portfolios are close to each other, and the return on the 9^{th}, for instance, provides a higher return than the 10^{th}. Hence, as an investment tool, our model is more intuitive on the lower side. It is possible to avoid a loss of around 2% over the market return by refraining risky portfolios, although the model offers around 0.50% extra profit moving to the safer companies. Average excess return for the entire portfolio space being almost equal to zero, .01%, proves it indeed covers the market. Figure **??** presents the average excess returns for the same portfolios the next two years, 2004 and 2005. Our first portfolio produces the lowest returns in the the three-year period, yet the excess returns smoothen both on the left and the right tails. It is possible to observe a general decline in the overall return on our portfolios, which cover the entire company space of 2003. The slowing economy would be the reason for this general reduction in returns, considering almost 1.5% of these companies went bankrupt before the year ends[29]. For the investment purposes however, the smoothening in the returns signals the reduction in the informative power in the long-run. Our model would be most intuitive for investments up to one year.

6 Conclusions

In this study we have explored the bankruptcy prediction with a new approach. Our first contribution has been the mixture model. Mixture model is advantageous in two major ways. We had the opportunity to model the event and timing of the defaults at the same time, and most importantly we managed to overcome the eventual default assumption of the standard models.

Second, besides the statistical specification, our model contributed to the literature by the extensive dataset it covers, all publicly traded companies in the US from 1980 up until 2007.

Third, we have used critical indicators to predict the default event. Productivity and profitability proved to be central in default probability estimation. Our

[29]See Table **??** for the distribution of bankruptcies within years.

model incorporated macro-economic variation through unemployment rate and we have introduced the government spending as a proxy to the stimulus plans. Government spending significantly reduces the default probability suggesting the success of government stipulation against bankruptcies.

Our fourth contribution is the higher predictive power provided by our model. A well-specified mixture model performs better in forecasting the default probability compared to the popular standard hazard rate model both in-sample and out-of-sample. With the improvements of mixture model, our default prediction delivered, less than 1% lower than the actual default rate of 8.14% as in-sample, and kept its predictive power up to three years.

Our fifth contribution is with respect to the risk and return analysis. Our highest default rate portfolio received the lowest return in the market, which proves the claims about the extremely low returns associated with bankruptcy risk.

References

[1] E. I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, September 1968.

[2] E. I. Altman. Predicting financial distress of companies: Revisiting the z-score and zeta® models. working paper, downloadable at pages.stern.nyu.edu/ ealtman/Zscores.pdf, July 2000.

[3] S. T. Bharath and T. Shumway. Forecasting default with the merton distance to default model. *The Review of Financial Studies*, 21(3):1339–1369, 2008.

[4] J. Y. Campbell, J. Hilscher, and J. Szilagyi. In search of distress risk. *Journal of Finance*, 63(6):2899–2939, December 2008.

[5] S. Chava and R. Jarrow. Bankruptcy prediction with industry effects. *Review of Finance*, 8:537–569, 2004.

[6] S. Chava, C. Stefanescu, and S. M. Turnbull. Modeling expected loss with unobservable heterogeneity. downloadable at http://wehner.tamu.edu/finc.www/schava/ChavaStefanescuTurnbull.pdf, February 2006.

[7] F. Corbière and P. Joly. A sas macro for parametric and semiparametric mixture cure models. *Comput. Methods Prog. Biomed.*, 85(2):173–180, 2007.

[8] F. Couderc and O. Renault. Times-to-default:life cycle, global and industry cycle impact. FAME Research Paper Series rp142, International Center for Financial Asset Management and Engineering, 2005.

[9] D. R. Cox and N. Wermuth. A comment on the coefficient of determination for binary responses. *American Statistician*, 46: 1—4, 1992.

[10] S. R. Das, D. Duffie, N. Kapadia, and L. Saita. Common failings: How corporate defaults are correlated. *Journal of Finance*, 62(1):93–117, 02 2007.

[11] Z. Da and P. Gao. Clientele change, liquidity shock, and the return on financially distressed stocks. forthcoming, Journal of Financial and Quantitative Analysis.

[12] D. Duffie, L. Saita, and K. Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83, 2007.

[13] E. F. Fama and K. R. French. Dissecting anomalies. *Journal of Finance*, 63(4):1653–1678, 08 2008.

[14] S. Hillegeist, E. Keating, D. Cram, and K. Lunstedt. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9:5–34, 2004.

[15] R. C. Merton. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29(2):109–131, 05 1974.

[16] J. Ohlson. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18:109–131, 1980.

[17] S. Parente and E. Prescott. Barriers to Riches. *MIT Press*, 2002.

[18] K. Roszbach. Bank Lending Policy, Credit Scoring, and the Survival of Loans. *The Review of Economics and Statistics*, 86(4):946–958, 2004.

[19] A. Saretto. Predicting and pricing the probability of default. AFA 2006 Boston Meetings Paper; EFA 2004 Maastricht Meetings Paper No. 4510. Available at SSRN: http://ssrn.com/abstract=556226, 2004.

[20] T. Shumway. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74:101–124, 2001.

[21] F. Steele. A discrete-time multilevel mixture model for event history with long-term survivors, with an application to an analysis of contraceptive sterilization in bangladesh. *Lifetime Data Analysis*, 9:155–174, 2003.

[22] J. Swanson, P. Marshall, H. Lokey and L. Norley. A Practitioner's Guide to Corporate Restructuring. *City & Financial Publishing*, 1st edition, 2008.

[23] M. Vassalou and Y. Xing. Default risk in equity returns. *Journal of Finance*, 59:831–868, 2004.

[24] Y. Yildirim. Estimating default probabilities of cmbs with clustering and heavy censoring. *The Journal of Real Estate Finance and Economics*, 37(2), 2008.

[25] M. E. Zmijewski. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22:59–82, 1984.

Table 1: **The Largest Ten Bankruptcies in the US History**

	Company	Bankruptcy Date	Total Assets (bn)
1	Lehman Brothers Holdings, Inc.	9/15/2008	$639
2	Washington Mutual	9/26/2008	$307
3	Worldcom, Inc.	7/21/2002	$103.91
4	Enron Corp.	12/2/2001	$63.39
5	Conseco, Inc.	12/18/2002	$61.39
6	Texaco, Inc.	4/12/1987	$35.89
7	Financial Corp. of America	9/9/1988	$33.86
8	Refco Inc.	10/17/2005	$33.33
9	Global Crossing Ltd.	1/28/2002	$30.18
10	Pacific Gas and Electric Co.	4/6/2001	$29.77
11	UAL Corp.	12/9/2002	$25.20
12	Delta Air Lines, Inc.	9/14/2005	$21.80

This table lists the largest ten bankruptcies of the US history. As can be seen the list contains 1987 bankruptcy earliest. Total assets are the pre-bankruptcy values in billions of dollars. We do not have the data for 2008 in our sample period, therefore the top ten bankruptcy list for our sample would include UAL Corp. and Delta Air Lines.

Table 2: **Distribution of Companies within Industries**

Industry	Number of Bankrupt Companies	Number of Healthy Companies	Total Number of Companies	Total Number of Observations (Company per Month)
1	705(10.11%)	6,270(89.89%)	6,975	438,076
2	621(7.71%)	7,433(92.29%)	8,054	746,877
3	190(12.30%)	1,355(87.70%)	1,545	134,879
4	124(3.47%)	3,447(96.53%)	3,571	43,005
Total	1,640(8.14%)	18,505(91.86%)	20,145	1,362,837

This table lists the total number of active firms and bankruptcies and the total number of monthly observations for 4 industry groups of our sample period. The percentages in parenthesis are as of the total number within an industry group. The number of active firms may change for reasons other than the bankruptcies. Industries are defined based on SIC codes. The groups are 1-Agriculture, Forestry and Fisheries, Construction, Wholesale and Retail Trade, Service Industries, 2-Manufacturing and Mineral, 3-Transportation, Communication and Utilities, 4-Finance, Insurance and Real Estate.

Table 3: **Distribution of Companies within Years**

Year	Number of Bankrupt Companies	Number of Healthy Companies	Total Number of Companies	Total Number of Observations (Company per Month)
1980	2,605(99.85%)	4(0.15%)	2,609	28,208
1981	3,915(99.85%)	6(0.15%)	3,921	35,547
1982	4,850(99.90%)	5(0.10%)	4,855	50,067
1983	5,348(99.83%)	9(0.17%)	5,357	58,701
1984	5,604(99.63%)	21(0.37%)	5,625	63,085
1985	5,725(99.62%)	22(0.38%)	5,747	63,885
1986	6,073(99.43%)	35(0.57%)	6,108	65,949
1987	6,264(99.40%)	38(0.60%)	6,302	70,018
1988	6,171(99.18%)	51(0.82%)	6,222	69,383
1989	5,951(98.97%)	62(1.03%)	6,013	67,295
1990	5,806(98.88%)	66(1.12%)	5,872	66,498
1991	5,942(98.85%)	69(1.15%)	6,011	66,672
1992	6,178(99.23%)	48(0.77%)	6,226	69,140
1993	7,291(99.40%)	44(0.60%)	7,335	75,149
1994	7,897(99.50%)	40(0.50%)	7,937	88,381
1995	8,193(99.38%)	51(0.62%)	8,244	90,789
1996	8,655(99.40%)	52(0.60%)	8,707	95,635
1997	8,903(99.19%)	73(0.81%)	8,976	99,137
1998	8,649(98.69%)	115(1.31%)	8,764	97,376
1999	8,405(98.48%)	130(1.52%)	8,535	92,148
2000	8,051(98.02%)	163(1.98%)	8,214	90,698
2001	7,347(97.56%)	184(2.44%)	7,531	84,158
2002	6,729(98.19%)	124(1.81%)	6,853	77,831
2003	6,323(98.69%)	84(1.31%)	6,407	72,606
2004	6,183(99.49%)	32(0.51%)	6,215	70,375
2005	6,116(99.43%)	35(0.57%)	6,151	69,414
2006	5,996(99.63%)	22(0.37%)	6,018	68,014
2007	5,885(99.07%)	55(0.93%)	5,940	65,818
Total	18,505 * (91.86%)	1,640(8.14%)	20,145*	2,011,977

This table lists the total number of active firms, bankruptcies and the total number of observations for every year of our sample period. The percentages in parenthesis are as of the total company number within that year. The number of active firms may change for reasons other than the bankruptcies.

*18,505 is the total number of healthy companies and 20,145 is the total number of all companies in the entire dataset, not the column totals.

Table 4: **Descriptive Statistics for Covariates**

Variable Names	Bankrupt Companies Mean	(Std. Dev.)	Healthy Companies Mean	(Std. Dev.)	All Companies Min	Max
$WCTA$	0.366	(0.367)	0.374	(0.368)	-0.512	1.005
$RETA$	0.013	(0.413)	0.021	(0.414)	-0.779	0.886
$EBITTA$	0.025	(0.047)	0.026	(0.047)	-0.065	0.108
STA	0.310	(0.241)	0.310	(0.242)	-0.303	0.795
$MVETL$	2.303	(3.180)	3.17	(3.61)	-5.579	10.54871
$Z-Score$	1.937	(2)	2.77	(2.27)	-3.530	9.925717
$CLCA$	0.738	(0.500)	0.739	(0.503)	-0.427	1.537
$NITA$	0.003	(0.027)	0.003	(0.027)	-0.049	0.059
$TLTA$	0.532	(0.279)	0.525	(0.277)	-0.272	1.324
$EXRET$	-0.007	(0.126)	-0.006	(0.124)	-0.289	0.271
$SIGMA$	0.131	(0.079)	0.128	(0.078)	0	0.321
$RSIZE$	-10.914	(2.349)	-10.886	(2.361)	-16.787	0.000
$\triangle UR$	-0.007	(0.120)	-0.006	(0.120)	-0.219	0.276
$lag \triangle G$	0.060	(0.026)	0.060	(0.025)	0	0.157
DD	0.3494815	(0.200)	0.441	(0.227)	0.019	1
$I1$	0.280	(0.449)	0.273	(0.446)	0	1
$I2$	0.463	(0.499)	0.464	(0.499)	0	1
$I3$	0.083	(0.276)	0.083	(0.275)	0	1
$I4$	0.173	(0.379)	0.181	(0.385)	0	1
# of obs.	148,538		1,863,439		2,011,977	

The following covariates are presented in this table: Working capital to total assets, WC/TA; Retained earnings to total assets, RE/TA; Earnings before interest and tax to total assets, $EBIT/TA$; Sales to total assets, S/TA; Market value of equity to total liabilities, MVE/TL; $Z-score$ (calculated as Z = 1.2WC/TA + 1.4RE/TA + 3.3EBIT/TA + .6MVE/TL + .999S/TA); Current liabilities to current assets, CL/CA; Net income to total assets, NI/TA; Total liabilities to total assets, TL/TA; Excess return, $EXRET$; $SIGMA$; Relative Size, $RSIZE$; The change in unemployment rate, ΔUR; One year lag of the change in government spending, $lag\Delta G$; Distance to Default, DD and Industry dummies, $I1-I4$. A total of 2,011,977 monthly observations are available, of which 148,538 are for bankrupt companies. Mean and the standard deviation of each predictor is calculated for bankrupt and non-bankrupt company observations. Minimum and maximum values are delivered for the entire dataset.

Table 5: **Expected Signs for The Covariates**

The Event of Default		
$TLTA$	Leverage	Positive
$NITA$	Profitability	Negative
$MVETL$	Solvency	Negative
$Z-Score$	Financial Health	Negative
The Time of Default		
$TLTA$	Leverage	Positive
$MVETL$	Solvency	Negative
$NITA$	Profitability	Negative
$Z-Score$	Financial Health	Negative
$EBITTA$	Productivity	Negative
STA	Efficiency	Negative
$WCTA$	Liquidity	Negative
$EXRET$	Excess Return	Negative
$SIGMA$	Volatility	Positive
$RSIZE$	Relative Size	Negative
$\triangle UR$	Unemployment Rate	Positive
$lag \triangle G$	Government Spending -1 Year Ago	Negative
DD	Distance to Default	Negative

This table presents the expected relationship between the bankruptcy indicator and the predictor variables. The expectations are derived from literature and general economic intuition.

Table 6: **Mixture Models vs Hazard Rate Models**

Variable	Mixture Models				Std. Haz. Rate Models							
	Model 1		Model 2		HR1		HR2 - Shumway [?]		HR3 - Chava and Jarrow [?]		HR4	
Names	Est.	(Z)	Est.	(Z)	Est.	(Z)	Est.	(Z)	Est.	(Z)	Est.	(Z)
The Event of Default												
$TLTA$	1.25	(11.85)										
$NITA$	−1.95	(2.30)										
$MVETL$	0.04	(4.71)										
$Z-Score$			0.00	(1.55)								
The Time of Default												
$TLTA$	1.02	(10.78)			2.51	(25.07)	3.16	(40.92)	3.17	(40.57)	2.36	(24.63)
$NITA$	−15.59	(11.69)	−17.95	(13.87)	−21.01	(13.77)	−19.38	(16.90)	−18.39	(15.97)	−17.18	(14.53)
$MVETL$	−0.33	(22.39)			−0.33	(14.15)						
$Z-Score$			0.00	(29.88)							−0.36	(14.18)
$WCTA$	−0.73	(10.88)	−1.02	(15.70)	−0.21	(2.61)						
$EBITTA$	−6.66	(8.81)	−6.39	(8.58)	−2.12	(2.65)						
STA	−0.70	(5.86)	−0.59	(5.10)	−0.16	(1.34)						
$EXRET$	−1.68	(11.74)	−1.65	(11.60)	−1.95	(12.29)	−2.19	(13.79)	−2.16	(13.66)	−2.00	(12.66)
$SIGMA$	1.67	(5.65)	1.94	(6.61)	5.85	(18.45)	6.15	(20.30)	5.87	(19.29)	5.79	(18.89)
$RSIZE$	0.06	(5.71)	0.07	(7.86)	−0.17	(10.29)	−0.28	(17.18)	−0.28	(17.20)	−0.19	(11.33)
$\triangle UR$	3.13	(14.54)	2.97	(14.02)	1.96	(9.17)						
$lag \triangle G$	−21.19	(15.34)	−21.01	(15.83)	−10.41	(7.37)						
DD												
$I1$	−3.12	(35.69)	−2.26	(51.13)	−11.06	(42.76)			−13.97	(66.65)		
$I2$	−3.35	(38.80)	−2.49	(56.22)	−11.22	(43.73)			−14.17	(68.97)		
$I3$	−3.37	(34.90)	−2.47	(39.79)	−11.09	(43.91)			−13.77	(65.73)		
$I4$	−3.84	(34.65)	−2.82	(38.84)	−11.56	(40.89)			−14.48	(67.33)		
$Constant$							−14.16	(69.82)			−11.94	−50.01)
# of obs.	2,011,977		2,011,977		2,011,977		2,011,977		2,011,977		2,011,977	
Log L	-13,242.75		-13,365.52		-10,122.45		-10,376.18		-10,350.62		-10,246.49	
AIC	0.01318		0.01330		0.01007		0.01032		0.01029		0.01019	
$Pseudo-R^2$	0.06544		0.06477		0.01553		0.01180		0.01181		0.01282	
Default Probability	7.38%		7.38%									

This table reports results from mixture models (Model 1-Model 2) and the standard hazard rate models (HR1-HR2-HR3-HR4-HR5) of bankruptcy indicator on predictor variables. The default probability and default intensity are defined to follow a logistic distribution in Model 1 and Model 2, and HRs are multi period logistic regressions. They only model the default intensity. HR1 is the hazard rate model specification with our covariates in Model 1. HR2 is the hazard model applied by Shumway [?] and HR3 is the Chava and Jarrow's [?] model. Chava and Jarrow adds industry dummies to Shumway's model, which indeed did not change the sign and magnitude of the original variables significantly. HR4 is the Shumway's model plus z-score. The absolute value of z-statistics are reported in parenthesis. A higher log-likelihood, a lower AIC or a higher R^2 values of a model indicate a better fit for the data.

Table 7: **Distribution of Bankruptcies for 2004, 2005, 2006 and 2007**

Year	2003	2004	2005	2006	2007	Total
Total number of companies	6,407	6,215	6,151	6,018	5,940	
Total number of bankruptcies	84	32	35	22	55	228
Bankruptcies of new companies	0	0	0	0	13	
Bankruptcies of old companies	84	32	35	22	42	215
Panel A - Model 1						
Bankruptcies from 1^{st} portfolio	84	32	35	22	42	215
Bankruptcies from 2^{nd} portfolio	0	0	0	0	0	0
Bankruptcies from 3^{rd} portfolio	0	0	0	0	0	0
Bankruptcies from 4^{th} portfolio	0	0	0	0	0	0
Bankruptcies from 5^{th} portfolio	0	0	0	0	0	0
Bankruptcies from 6^{th} portfolio	0	0	0	0	0	0
Bankruptcies from 7^{th} portfolio	0	0	0	0	0	0
Bankruptcies from 8^{th} portfolio	0	0	0	0	0	0
Bankruptcies from 9^{th} portfolio	0	0	0	0	0	0
Bankruptcies from 10^{th} portfolio	0	0	0	0	0	0
Panel B - HR1						
Bankruptcies from 1^{st} portfolio	71	20	11	3	4	109
Bankruptcies from 2^{nd} portfolio	7	8	7	5	8	35
Bankruptcies from 3^{rd} portfolio	1	1	4	1	4	11
Bankruptcies from 4^{th} portfolio	2	1	2	5	9	19
Bankruptcies from 5^{th} portfolio	0	0	0	1	1	2
Bankruptcies from 6^{th} portfolio	2	0	4	1	5	12
Bankruptcies from 7^{th} portfolio	0	2	3	4	6	15
Bankruptcies from 8^{th} portfolio	1	0	1	1	1	4
Bankruptcies from 9^{th} portfolio	0	0	2	1	2	5
Bankruptcies from 10^{th} portfolio	0	0	1	0	2	3
Panel C - HR2						
Bankruptcies from 1^{st} portfolio	69	22	33	35	40	199
Bankruptcies from 2^{nd} portfolio	7	4	12	18	24	65
Bankruptcies from 3^{rd} portfolio	4	2	2	6	10	24
Bankruptcies from 4^{th} portfolio	0	2	6	10	16	34
Bankruptcies from 5^{th} portfolio	2	1	4	5	12	24
Bankruptcies from 6^{th} portfolio	1	0	1	3	7	12
Bankruptcies from 7^{th} portfolio	1	0	2	4	9	16
Bankruptcies from 8^{th} portfolio	0	1	3	4	6	14
Bankruptcies from 9^{th} portfolio	0	0	3	3	3	9
Bankruptcies from 10^{th} portfolio	0	0	1	1	4	6

This table displays that all of the failures through 2004-2007 have been predicted by our model. Ten portfolios are formed in 2003 based on annualized default probabilities estimated by Model 1 in Table **??**. There are equal number of companies in each portfolio. Panel A presents upcoming failures within these portfolios. Panel B compares the ten portfolios formed again in 2003 but based on annualized default probabilities of HR1 estimation in Table **??**. Panel C compares the ten portfolios formed again in 2003 but based on annualized default probabilities of HR2 estimation in Table **??**. All of the companies filed for bankruptcy within the following 3 years were listed in the 1^{st} portfolio of Model 1 estimation. New companies are defined as the companies established later during 2004 to 2007, which by default can not enter our portfolio selection process in 2003. Old companies are the ones established before 2004.

Table 8: **Mixture Model on Non-Financial and Financial Companies**

Variable Names	All Companies Est.	(Z)	Non-Financial Companies Est.	(Z)	Financial Companies Est.	(Z)
The Event of Default						
$TLTA$	1.25	(11.85)	1.21	(10.66)	2.10	(6.06)
$NITA$	−1.95	(2.30)	−2.37	(2.69)	−5.66	(1.32)
$MVETL$	0.04	(4.71)	0.02	(1.88)	0.14	(4.65)
The Time of Default						
$TLTA$	1.02	(10.78)	0.67	(6.60)	1.29	(2.81)
$NITA$	−15.59	(11.69)	−12.98	(9.26)	−31.35	(5.96)
$MVETL$	−0.33	(22.39)	−0.36	(21.40)	−0.13	(3.23)
$WCTA$	−0.73	(10.88)	−1.28	(16.20)	−0.12	(0.68)
$EBITTA$	−6.66	(8.81)	−8.76	(10.64)	3.12	(1.02)
STA	−0.70	(5.86)	−0.67	(5.26)	0.39	(0.76)
$EXRET$	−1.68	(11.74)	−1.63	(10.82)	−1.28	(2.56)
$SIGMA$	1.67	(5.65)	1.52	(4.93)	3.74	(3.46)
$RSIZE$	0.06	(5.71)	0.05	(4.64)	−0.03	(0.91)
$\triangle UR$	3.13	(14.54)	2.84	(12.62)	4.18	(4.82)
$lag \triangle G$	−21.19	(15.34)	−20.05	(13.97)	−25.01	(4.12)
$I1$	−3.12	(35.69)	−2.96	(31.59)		
$I2$	−3.35	(38.80)	−3.16	(34.35)		
$I3$	−3.37	(34.90)	−3.17	(31.13)		
$I4$	−3.84	(34.65)				
$Constant$					−5.40	(18.05)
# of obs.	2,011,977		1,662,963		349,014	
Log L	-13,242.75		-11,882.44		-1,075.85	
AIC	0.01318		0.01431		0.00625	
$Pseudo-R^2$	0.06544		0.06948		0.04388	
Default Probability	7.38%		8.11%		3.11%	

This table reports results from mixture Model 1 of bankruptcy indicator on predictor variables for all companies in the first column, non-financial companies in the second column and for the financial companies in the third column. The default probability and default intensity are defined to follow a logistic distribution in Model 1. The absolute value of z-statistics are reported in parenthesis. A higher log-likelihood, a lower AIC or a higher R^2 values of a model indicate a better fit for the data.

Table 9: **Distance to Default Models**

	(1) Cons. + DD	(2) Ind. Dummies + DD	(3) Model-1 + DD		
Panel A - Mixture Models					
The Event of Default					
DD	0.29	0.40	0.19		
	(4.09)	(5.57)	(1.85)		
The Time of Default					
DD	-6.04	-5.99	-0.37		
	(61.39)	(60.76)	(2.67)		
$LogL$	$-14,470.42$	$-14,418.38$	$-13,239.96$		
AIC	0.01439	0.01434	0.01318		
$Pseudo-R^2$	0.04953	0.05225	0.06558		
Panel B - Hazard Rate Models					
	(4) Cons. + DD	(5) Ind. Dummies + DD	(6) HR1 + DD	(7) Shumway + DD	8 C&J + DD
DD	-5.50	-5.35	1.16	0.30	0.40
	(30.80)	(29.83)	(6.47)	(1.64)	(2.12)
$LogL$	$-12,646.96$	$-12,595.20$	$-10,104.30$	$-10,374.89$	$-10,348.49$
AIC	0.01257	0.01253	0.01006	0.01032	0.01030
$Pseudo-R^2$	0.00184	0.00190	0.01590	0.01186	0.01189

This table reports the coefficient on distance to default, log-likelihood, Akaike Information Criteria and $Pseudo-R^2$for various models incorporating distance to default. Panel A presents the mixture models, and Panel B presents the hazard rate models. (1) runs a mixture model with a constant and distance to default as a predictor for both event and timing of default. (2) runs a mixture model with a distance to default as a predictor for both event and timing of default and industry dummies on the timing part. (3) runs Model 1 in Table **??** including distance to default on both event and timing parts. (4) runs a hazard rate model with a constant and distance to default as predictor variables. (5) runs a hazard rate model with distance to default and industry dummies. (6) runs HR1 in Table **??** adding distance to default as predictor. (7) runs Shumway's model [**?**] adding distance to default. (8) runs Chava and Jarrow's model [**?**] adding distance to default. The absolute value of z-statistics are reported in parenthesis. A higher log-likelihood, a lower AIC or a higher R^2 values of a model indicate a better fit for the data.

Table 10: **Mixture Model on Lagged Values**

Lag (Months)	0	6	12	24	36
The Event of Default					
$TLTA$	1.22	0.88	0.78	0.67	0.75
	(11.18)	(7.8)	(6.95)	(5.66)	(5.91)
$NITA$	−1.93	−2.92	−3.24	−3.43	−3.18
	(2.18)	(3.24)	(3.58)	(3.61)	(3.1)
$MVETL$	0	0	0	0	0
	(4.32)	(1.33)	(0.25)	(1.03)	(0.41)
The Time of Default					
$TLTA$	0.98	0.59	0.36	0.31	0.4
	(9.93)	(5.33)	(3.16)	(2.44)	(2.94)
$NITA$	−13.49	−9.29	−9.51	−5.41	−3.45
	(9.57)	(6.57)	(6.21)	(3.58)	(2.11)
$MVETL$	−0.28	−0.22	−0.14	−0.09	−.05
	(21.5)	(15.75)	(13.1)	(8.48)	(4.33)
$WCTA$	−0.73	−0.62	−0.2	0.13	0.18
	(10.43)	(7.76)	(2.37)	(1.4)	(1.79)
$EBITTA$	−6.75	−4.83	−3.34	−1.63	−1.62
	(8.44)	(5.48)	(3.31)	(1.54)	(1.42)
STA	−0.78	−0.66	−0.75	−0.69	−0.57
	(6.19)	(4.88)	(5.56)	(4.79)	(3.71)
$EXRET$	−1.64	−0.86	−1.03	−0.36	−0.34
	(10.96)	(5.42)	(6.21)	(1.96)	(1.64)
$SIGMA$	1.91	5.59	4.97	4.33	2.81
	(6.13)	(14.89)	(13.2)	(10.85)	(6.42)
$RSIZE$	0.05	0.09	0.09	0.08	0.07
	(4.48)	(7.94)	(7.97)	(6.8)	(5.43)
$\triangle UR$	3.12	3.93	3.46	3.05	1.71
	(13.37)	(16.2)	(14.7)	(11.52)	(5.64)
$lag \triangle G$	−22.28	−22.53	−19.6	−20.9	−20.3
	(15.77)	(16.32)	(14.7)	(15.82)	(16.01)
$I1$	−3.12	−2.78	−2.6	−2.42	−2.48
	(34.25)	(29.83)	(28.1)	(25.25)	(24.3)
$I2$	−3.35	−3.03	−2.87	−2.69	−2.75
	(37.3)	(32.76)	(31.4)	(28.47)	(27.58)
$I3$	−3.36	−2.99	−2.8	−2.59	−2.63
	(33.27)	(29.14)	(27.3)	(24.05)	(22.94)
$I4$	−3.85	−3.48	−3.38	−3.33	−3.47
	(32.99)	(29.32)	(28.7)	(26.93)	(25.97)
# of obs.	1,738,356	1,591,887	1,585,305	1,442,150	1,314,028
# of bank.	1,132	1,145	1,164	1,194	1,248
Log L	−12,027.51	−11,859.69	−12,322.87	−11,532.09	−10,302.76
AIC	0.01385	0.01492	0.01556	0.01601	0.01570
$Pseudo-R^2$	0.06586	0.06096	0.05871	0.05791	0.05610

This table takes our best-model variables (Model 1) and reports their predictive power for lags of 6, 12, 24, and 36 months. The dependent variable is failure and the sample period is 1980 to 2003. The estimates for '0 months' is Model 1 model of Table **??** for 1980-2003. The estimates for '6 months' is the prediction of the first 6 months of 2004 by the sample period. 12, 24 and 36 months are the predictions of periods following 2003 as well. The absolute value of z-statistics is reported in parentheses. A higher log-likelihood, a lower AIC or a higher R^2 values of a model indicate a better fit for the data.

Figure 2: Bankruptcies by Industry Group
The figure depicts the values in Table ??. It shows the number of healthy and bankrupt companies within each industry group for the period 1980-2007. The percentage of bankruptcies are defined as of each industry. The highest percentage is observed in the third industry group.

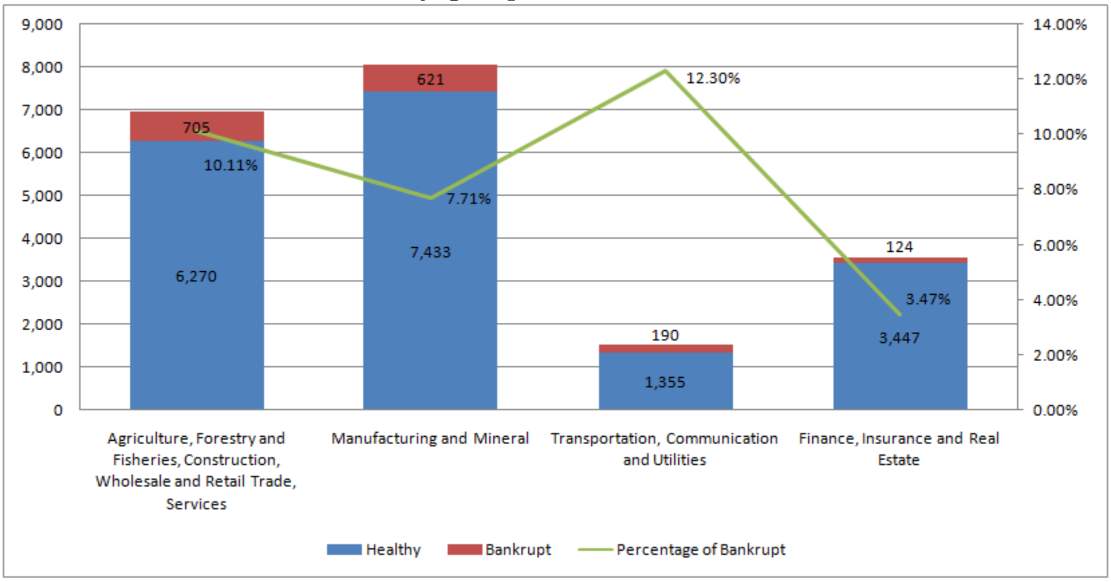


Figure 3: Bankruptcies by Year
The figure depicts the values in Table ??. It shows the number of bankrupt companies within each year for the period 1980-2007. The percentage of bankruptcies are defined as of the total number of companies each year. The highest level and percentage is observed in 2001.

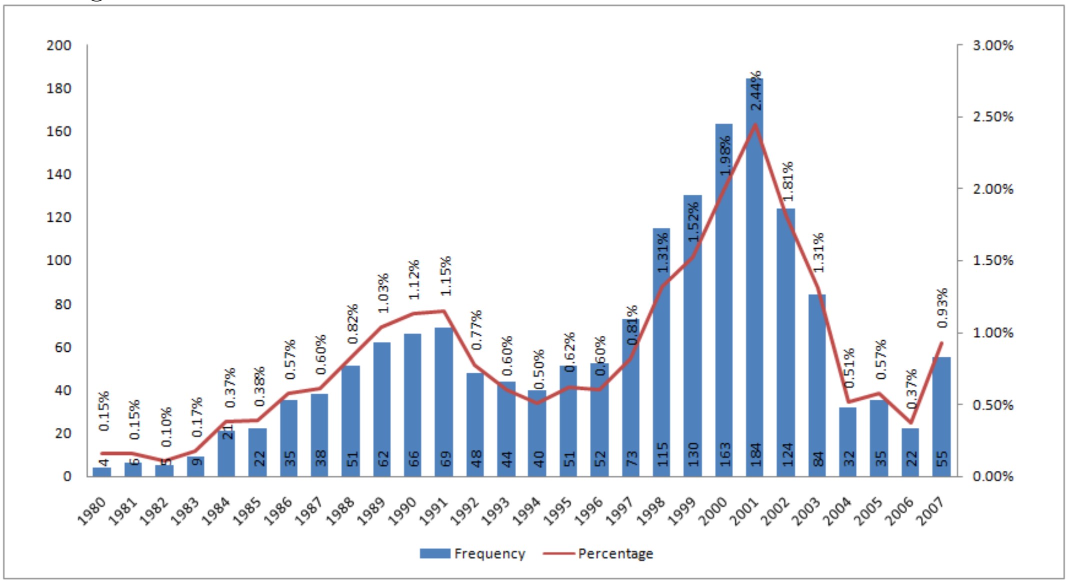
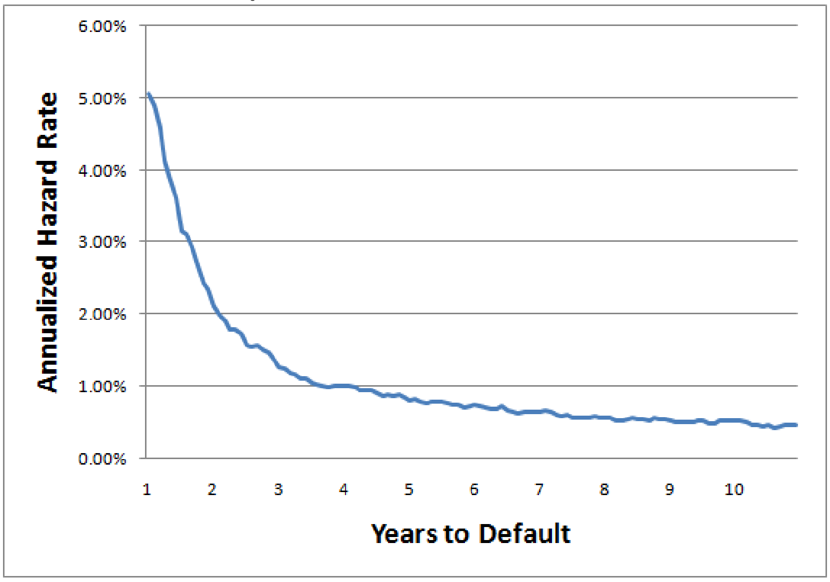


Figure 4: Estimated Hazard Rates
Hazard rates are estimated by Mixture Model-1 in Table **??**.



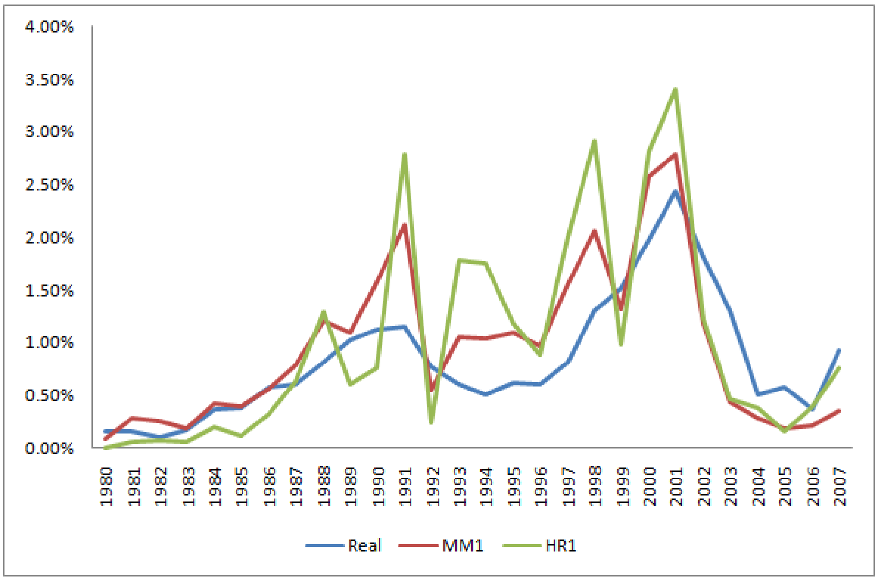
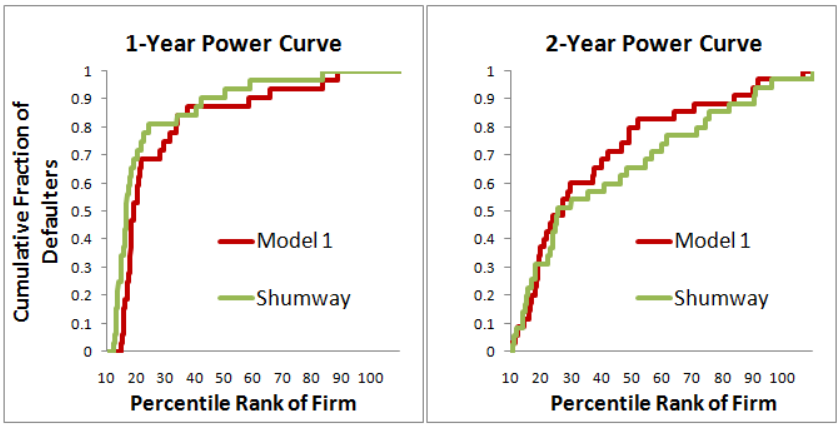


Figure 5: Actual vs. Predicted Default Rates
The figure shows the actual vs. estimated default rates for the period 1980-2007. The default is defined as the Chapter 7 or 11 filing of the company. The predicted rates are calculated using the fitted values of models Model 1 and HR1 in Table **??**.

Panel A. 2003



Panel B. 1998

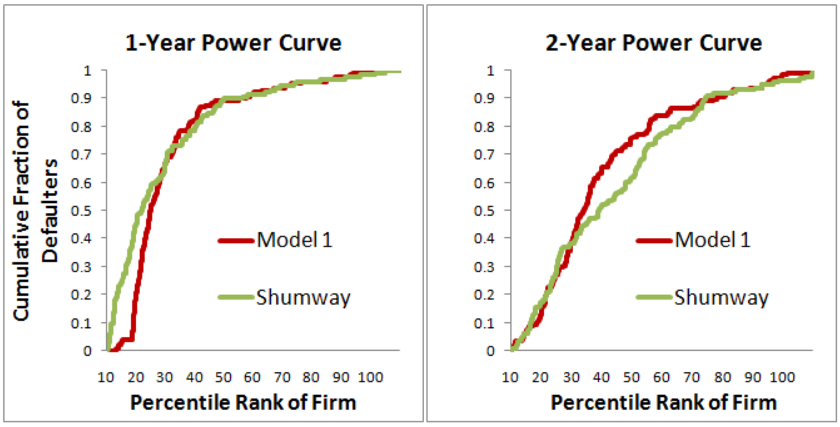


Figure 6: Power Curves
Average out-of-sample power curves for one-year, two-year, three-year and four-year default predictions, after 2003. In the one year power curve, for each fraction x on the horizontal axis, the curve shows the fraction of the firms that defaulted within one year. In two-year curve, within two year, in three-year curve within three year, and in four-year curve, within four years. The fractions on the x-axis are the portfolios. The portfolios in Model 1 curve formed according to estimated default probabilities at the year 2003 by Model 1 in Table **??** and the portfolios in HR1 curve formed according to estimated default probabilities by HR1 of the same table. 10 portfolios cover the entire company space of 2003. There are equal number of firms in each portfolio. Portfolio-1 refers to the riskiest 10% in both estimations. Riskiest 10% predicts all of upcoming defaults in three years in Model 1. Model 1 is more accurate compared to HR1 in one, two and three-year predictions.

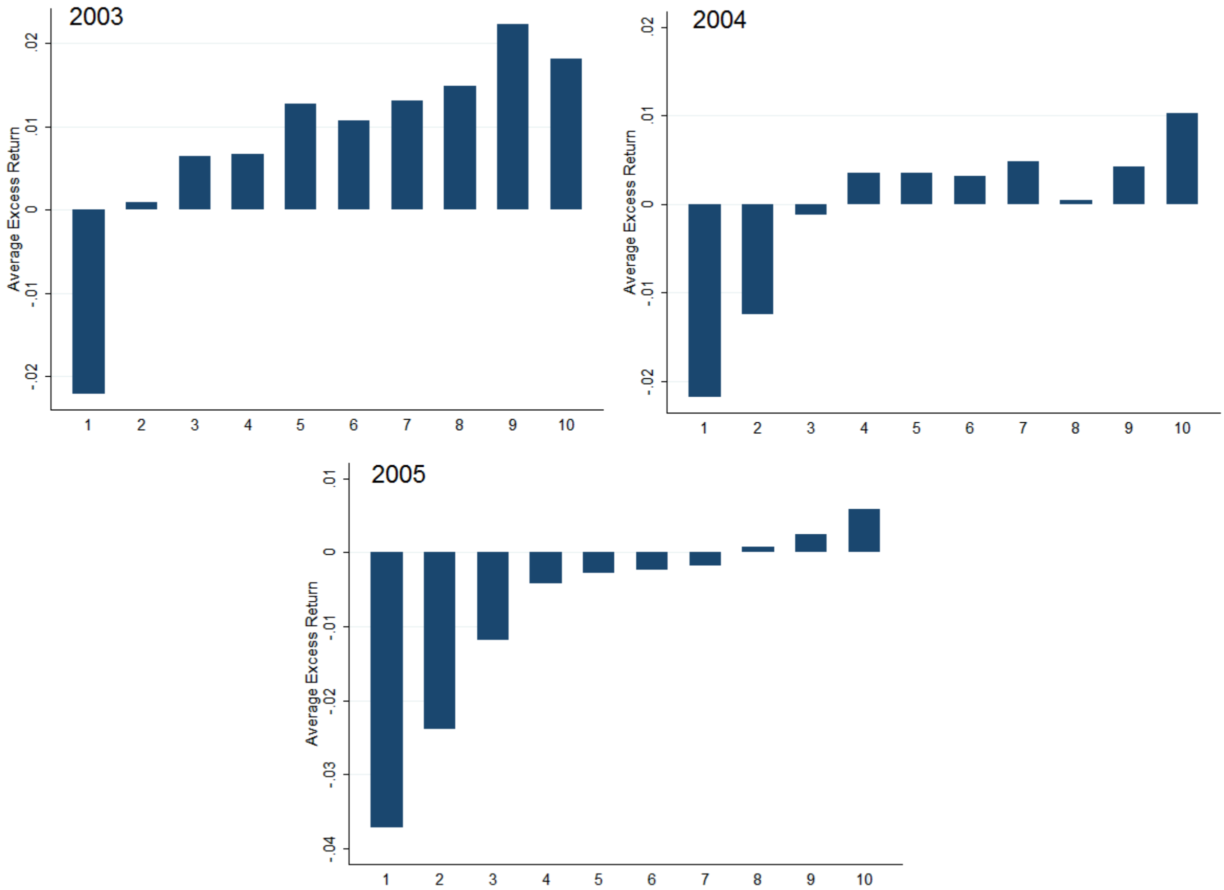
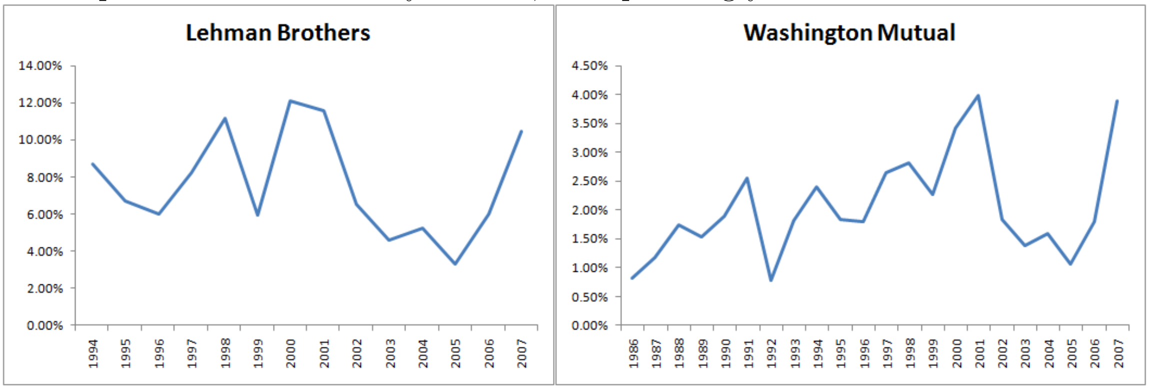


Figure 7: Average Excess Returns in 2003, 2004 and 2005
The graphs plot the average excess return in 2003, 2004 and 2005 for each portfolio formed in 2003. On the horizontal axis, the first one being the riskiest, portfolios from one to ten present lower default risk. The portfolios are not equal size. From the first one to the last one, they contain the 5%, 5%, 10%, 10%, 20%, 20%, 10%, 10%, 5%, and 5% of the companies respectively. Excess return is the real return on a firm's equity over the value-weighted average return on S&P500.

Figure 8: Sample Companies

The graphs plot the annualized default probabilities estimated by Model-1. On the vertical axis, the estimated default probabilities are given in percentages. On the horizontal axis, the years are given. For Lehman Brothers, the data starts from 1994, while for Washington Mutual it starts from 1986. Both of these companies exist in the the riskiest first portfolio, constructed in 2003. They filed for the largest bankruptcies of the US history in 2008, the upcoming year for the dataset.



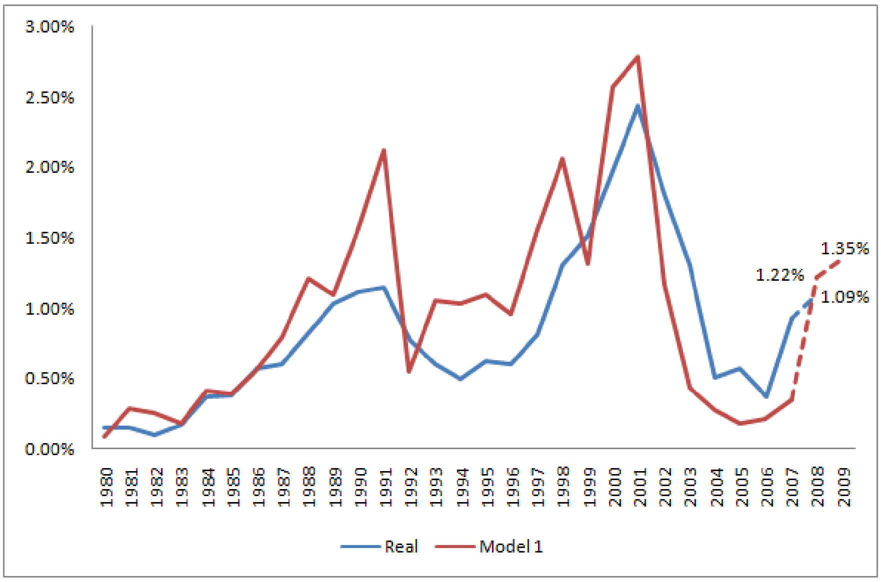


Figure 9: Two-year Forecast by Model 1
The graph displays the two-year forecast of the annual default probability by Model 1. The forecasted part of the Model 1 is presented by the dashed portion. The real default percentages are also displayed. We have the bankruptcy data up until 2008, however it was not employed in the estimation due to the lack of accounting data. The dashed part in the real data refers to the part that is not utilized in the estimation.