Review of Finance 8: 537–569, 2004.

Bankruptcy Prediction with Industry Effects ⋆

SUDHEER CHAVA[1] and ROBERT A. JARROW[2]
[1] C.T. Bauer College of Business, University of Houston; [2] Johnson Graduate School of Management, Cornell University

Abstract. This paper investigates the forecasting accuracy of bankruptcy hazard rate models for U.S. companies over the time period 1962–1999 using both yearly and monthly observation intervals. The contribution of this paper is multiple-fold. One, using an expanded bankruptcy database we validate the superior forecasting performance of Shumway's (2001) model as opposed to Altman (1968) and Zmijewski (1984). Two, we demonstrate the importance of including industry effects in hazard rate estimation. Industry groupings are shown to significantly affect both the intercept and slope coefficients in the forecasting equations. Three, we extend the hazard rate model to apply to financial firms and monthly observation intervals. Due to data limitations, most of the existing literature employs only yearly observations. We show that bankruptcy prediction is markedly improved using monthly observation intervals. Fourth, consistent with the notion of market efficiency with respect to publicly available information, we demonstrate that accounting variables add little predictive power when market variables are already included in the bankruptcy model.

1. Introduction

With the introduction and expansion of credit derivative markets (see Risk (2000)), bankruptcy hazard rate estimation has taken on a new importance. Default rate intensities are a necessary input to credit derivative pricing models (see Jarrow and Turnbull (1995), Duffie and Singleton (1999)). These inputs can be estimated implicitly using debt prices (see Janosi et al. (2002)) or explicitly using actual bankruptcies, balance sheet and market data (see Altman (1968), Zmijewski (1984), Shumway (2001), Moody's (2000a, b)). A recent paper by Jarrow et al. (2000) links these two estimation procedures.

Most of the bankruptcy prediction models fitted in the academic literature are based on a limited data set containing at most 300 Bankruptcies[1] and yearly observation intervals. This paper uses an expanded database with yearly and monthly observation intervals to reinvestigate bankruptcy prediction models. This expanded bankruptcy database is constructed as part of this research investigation, much of

⋆ We would like to thank Warren Bailey, Jean Helwege, Yongmiao Hong, Yaniv Grinstein, Haitao Li, Amiyatosh Purnanandam, Sophie Shive, Tyler Shumway, Bhaskaran Swaminthan, Sriraman Venkatraman, Martin Wells, seminar participants at the 2001 Financial Management Association annual meetings at Toronto and especially an anonymous referee and the editor Josef Zechner for helpful comments that greatly improved the paper.

[1] See Moody's (2000b), Exhibit 2.1 p. 14.

it by hand from hard copy sources. It consists of most bankruptcies in the U.S. between 1962 and 1999 for publicly listed companies included in both the COMPUSTAT and CRSP databases. The total number of bankruptcies included in this expanded database is 1461.[2]

The first purpose of this paper is to use this expanded database to validate the superiority of Shumway's (2001) dynamic hazard rate model over the models of Altman (1968) and Zmijewski (1984). Using this database, but first restricting it to enable a direct comparison with the previous studies, we re-estimate the models of Altman (1968), Zmijewski (1984) and Shumway (2001) over the 1962–1990 period and forecast bankruptcies over 1991–1999. The bankruptcy database is restricted to include only non-financial firms listed on either the AMEX or NYSE, and yearly observation intervals. Our results confirm the more accurate prediction of Shumway's model.

Second, not much attention has been paid to industry effects in the previous academic literature, most likely due to the limited number of bankruptcies in the databases previously available.[3] Yet, economic intuition suggests that industry effects should be an important component in bankruptcy prediction. This is true for two reasons. First, different industries face different levels of competition and, therefore, the likelihood of bankruptcy can differ for firms in different industries with otherwise identical balance sheets. Second, different industries may have different accounting conventions, again implying that the likelihood of bankruptcy can differ for firms in different industries with otherwise identical balance sheets.

In related research, Lang and Stultz (1992), Shleifer and Vishny (1992), Opler and Titman (1994), Maksimovic and Phillips (1997), and Berkovitch and Israel (1998) support the importance of industry effects on bankruptcy. Berkovitch and Israel (1998) model the decision to declare bankruptcy as a strategic choice variable. In their model, the proportion of firms that enter bankruptcy is higher for firms in mature industries versus firms in growth industries because under-investment problems are less important in mature industries. Maksimovic and Phillips (1997) present evidence that incidence of bankruptcy depends on industry demand conditions. Opler and Titman (1994) find that the adverse consequences of leverage on bankruptcy are more pronounced in concentrated industries. Shleifer and Vishny (1992) argue that when a firm in financial distress needs to sell assets, its industry peers are likely to be experiencing problems themselves, leading to asset sales at prices below value in best use. Finally, Lang and Stultz (1992) investigate the contagion and competitive intra-industry effects of bankruptcy announcements. In a recent study of the determinants of recovery rates of defaulted securities, Acharya et al. (2003) find that in addition to seniority and security of the defaulted securities, industry conditions at the time of default are found to be robust and important determinants of the recovery rates.

[2] The Moody's public model data base (2000b) consists of only 1,406 bankruptcies.

[3] Kavvathas (2000) considers industry effects in his study of rating transitions. Institutional studies do consider industry effects in their bankruptcy models (see Moody's (2000b)).

Given our augmented bankruptcy database, we can investigate the importance of including industry effects in bankruptcy prediction. We approach this investigation in two steps. In the first step we study only non-financial firms traded on either the AMEX or NYSE. Step one is included to enable a direct comparison with the previous literature. Step two augments the investigation to include both financial firms and firms traded on the NASDAQ.

To study the importance of industry effects on bankruptcy prediction, we divide our collection of firms into four industry groupings: (i) finance, insurance and real estate, (ii) transportation, communications and utilities, (iii) manufacturing and mineral, and (iv) miscellaneous industries (the complement of (i)–(iii)). The SIC codes for each of these industry groupings is discussed in the text. For step one, we only consider the non-financial firms listed on either the AMEX or the NYSE with yearly observation intervals. A hazard rate model is fit with intercept and slope coefficient dummy variables for each industry grouping. This estimation is conceptually equivalent to running separate hazard rate models for each industry grouping. The evidence is consistent with the statistical significance of an industry effect (in sample fit). Including industry effects leaves the model's forecasting ability unchanged (out of sample fit).

Next, we augment the estimation to include both the financial firms and firms traded on the NASDAQ. Industry intercept and slope coefficient dummy variables are again included. For financial firms analogous balance sheet ratios are included to those used for the non-financials. Two different models are studied: a private firm model and a public firm model.[4] A private firm model is one that is based on only accounting variables. This reflects an estimation procedure suitable for a private firm, i.e., a firm without publicly traded equity. In contrast, a public firm model is one that uses both market and accounting variables. The market variables relate to the publicly traded equity. The estimation for the augmented database confirms the previous conclusions. For both the private and public firm model, the industry effect is both statistically significant (in sample fit) and it significantly improves the forecasting ability of the model (out of sample fit).

After completing an analysis with yearly observation intervals, we perform a comparative study of the forecasting accuracy of the bankruptcy model based on monthly observation intervals. For a private firm model using only accounting data and a public firm model that also includes market variables, monthly observations significantly improve the forecasting ability of the procedure. This is not a surprising result and it is consistent with the usefulness of timely information contained in quarterly accounting reports and monthly stock price variables.

Next, given the shorter updating frequency of market variables (monthly) versus accounting variables (quarterly), we investigate the efficient markets hypothesis that market variables contain all the information reflected in the quarterly accounting reports with respect to bankruptcy. Although an in-sample test indicates additional explanatory power for the accounting variables, an out-of-sample

[4] This terminology is adopted from Moody's (2000a, b).

prediction test supports this efficient market hypothesis. The most accurate and parsimonious public firm model in terms of forecasting includes no accounting variables.

An outline of this paper is as follows. Section 2 briefly describes our augmented bankruptcy database and provides some summary bankruptcy statistics. Section 3 describes the statistical model. Section 4 reports the re-estimation of the Altman (1968), Zmijewski (1984) and Shumway (2001) models using yearly observation intervals. Section 5 studies the inclusion of industry effects for non-financial firms with equity trading on either the AMEX or NYSE. Financial firms and firms with equity trading on the NASDAQ are included in the estimation in Section 6. Section 7 provides the re-estimation of the models in Section 6 using monthly observation intervals, and Section 8 concludes the paper.

2. The Bankruptcy Data Base

This section describes our augmented bankruptcy database. For pricing corporate debt and credit derivatives, the relevant economic state of the firm is default where default is defined to be the condition that occurs when a firm has a delayed or missing contractual debt payment. Unfortunately, data on defaults is not readily available. For this reason, instead of defaults, this paper studies bankruptcies where a bankruptcy is defined to occur when a firm makes either a Chapter 7 or Chapter 11 filing.[5] Although a Chapter 11 filing may not necessarily mean that the firm will eventually file for Chapter 7 (liquidation), we include Chapter 11 filings because in these filings there is significant uncertainty about both the timing and magnitude of the final payments to the firms' claim holders. This is consistent with the intent to use these estimates for the pricing of corporate debt and credit derivatives.[6]

This database consists of all bankruptcy filings as reported in the Wall Street Journal Index (1962–1980), The SDC Database (Reorganizations module 1980–1999), SEC filings (1978–1999) and the CCH Capital Changes Reporter. As such, the bankruptcy data includes most of the bankruptcy filings between 1962–1999 of publicly traded companies on either the NYSE, AMEX and NASDAQ stock exchanges. This database is the most comprehensive bankruptcy database available. Indeed, it includes all the firms in Shumway's (2001) database[7] and slightly more firms than in Moody's public firm database (2000b). There are a total of 1461 bankruptcies in our initial sample.

Figure 1 contains a partitioning of our bankruptcies in our database by stock exchange listing. As indicated, over half of the companies that filed for bankruptcy

[5] We focus only on the Chapter 7 and Chapter 11 bankruptcy filings as it is difficult to get information on all the private restructurings that occurred between the lenders and borrowers for the sample period. To our knowledge there is no database that has this information.

[6] The subtle difference between defaults and bankruptcies will prove important in subsequent sections of this paper.

[7] We thank Tyler Shumway for making his dataset available to us.

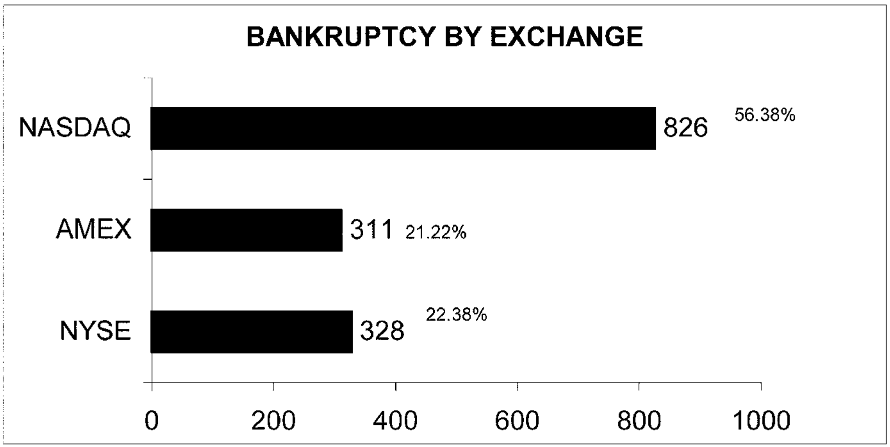


Figure 1. Bankruptcy by exchange. The above figure gives the break up of the bankruptcies in the sample by the exchange in which the firm is listed at the time of bankruptcy filing. The bankruptcy sample includes all bankruptcy filings as reported in the Wall Street Journal Index (1962–1980), The SDC Database (Reorganizations module 1980–1999), SEC filings (1978–1999) and the CCH Capital Changes Reporter. As such, the bankruptcy data includes most of the bankruptcy filings between 1962–1999 of publicly traded companies on either the NYSE, AMEX and NASDAQ stock exchanges. There are a total of 1461 bankruptcies in our initial sample. The number of bankruptcies by the exchange is given along with the percentages (of total number of bankruptcies in the sample).

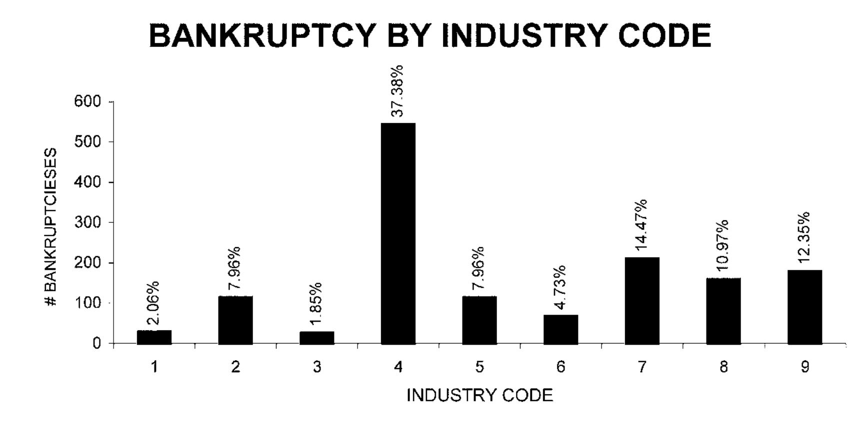
were listed on NASDAQ. The remaining percentage is almost evenly distributed between the NYSE and AMEX exchanges.

Figure 2 contains a partitioning of bankruptcies by industry classification. We considered the four digit SIC codes for each company from CRSP. The industries are grouped into ten major industrial sectors based on the SIC classification. The names of the industrial groupings are also included.

The SIC codes assigned to a company may change through out our sample period. CRSP keeps track of all the historical SIC codes. We considered the relevant SIC code for each company for each year. If a company's SIC code changes then we consider it to be a part of the new industry group in the year of change. Although there are a few well-known problems with the SIC codes as reported by both COMPUSTAT and CRSP (see Kahle and Walkling (1996), Guenther and Rosman (1994)), we use them because these are the most widely available industry classifications covering our sample period. From Figure 2 we see that the majority of bankruptcies occur for group 4 (manufacturing) with groups 7 (retail trade), 8 (finance, insurance and real estate) and 9 (service industries) containing the next highest percentages.

Figure 3 gives the number of bankruptcies by year over our sample period. As seen, the number of bankruptcies has been increasing over time, with the largest number of bankruptcies occurring in the late 1980s and early 1990s. Figure 4 contains the percentage of firms in our database that go bankrupt each year. The

PANEL A: BANKRUPTCY BY INDUSTRY CODE



PANEL B: INDUSTRY CODE CONSTRUCTION

The following table gives the SIC codes corresponding to the industry codes used in the figure in Panel A along with the name of the industry. The last column gives the number of bankruptcies during the sample period 1962-1999 in each of these industries and the percentage of bankruptcies are given in brackets.

IND CODE	SIC CODE	INDUSTRY NAME	# (%) OF BANKRUPTCIES
1	< 1000	Agriculture, Forestry and Fisheries	30 (2.06%)
2	1000 to less than 1500	Mineral Industries	116 (7.96%)
3	1500 to less than 1800	Construction Industries	27 (1.85%)
4	2000 to less than 4000	Manufacturing	545 (37.38%)
5	4000 to less than 5000	Transportation, Communications, and Utilities	116 (7.96%)
6	5000 to less than 5200	Wholesale Trade	69 (4.73%)
7	5200 to less than 6000	Retail Trade	211 (14.47%)
8	6000 to less than 6800	Finance, Insurance, and Real Estate	160 (10.97%)
9	7000 to less than 8900	Service Industries	180 (12.35%)
10	9100 to less than 10000	Public Administration	0 (0%)
		TOTAL # OF BANKRUPTCIES	1461 (100%)

Figure 2. Industry code construction bankruptcy by industry code. The above figure gives the break up of the bankruptcies in the sample by the industry to which the firm belongs. The bankruptcy sample includes all bankruptcy filings as reported in the Wall Street Journal Index (1962–1980), The SDC Database (Reorganizations module 1980–1999), SEC filings (1978–1999) and the CCH Capital Changes Reporter. As such, the bankruptcy data includes most of the bankruptcy filings between 1962–1999 of publicly traded companies on the NYSE, AMEX and NASDAQ stock exchanges. There are a total of 1461 bankruptcies in our initial sample. The percentage of bankruptcies (of the total number of bankruptcies in the sample) in each industry is also given.

percentage bankruptcies mimic the time series behavior of the number of bankruptcies in Figure 4. Conceptually, for each year, this percentage provides the best estimate of the unconditional default probability for any firm with equity trading either on the AMEX, NYSE or NASDAQ exchanges. The mean percentage of firms that go bankrupt over our sample period is 0.65 percent, the median is 0.56 percent.

Figure 5 provides the age of a firm at the time of bankruptcy. The age of the firm is defined as the time that has elapsed since the firm started trading on an organized

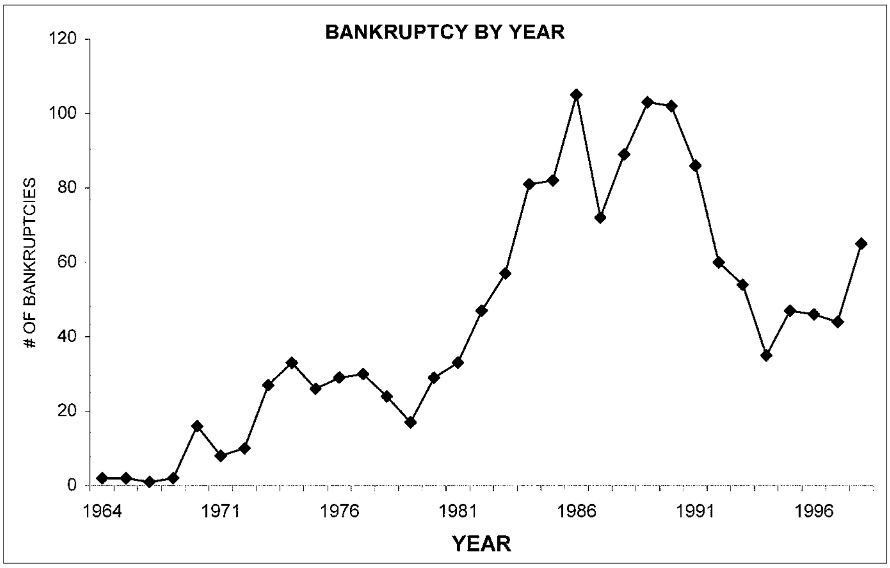
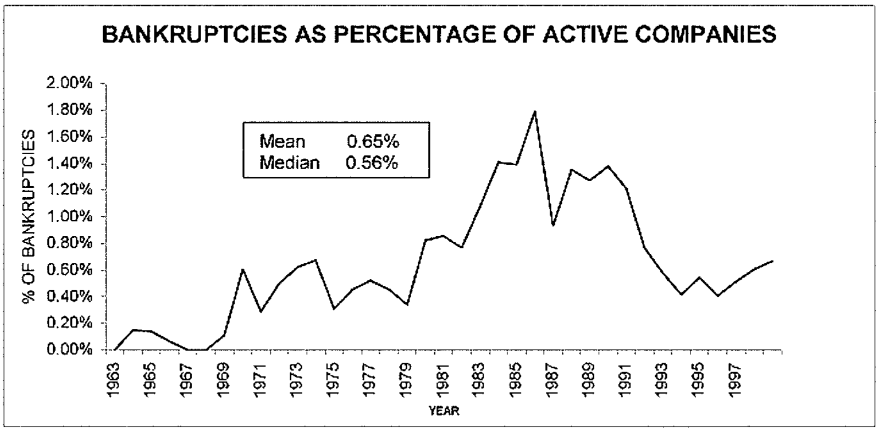


Figure 3. Bankruptcy by year. The above figure shows the number of bankruptcies per year over the sample period of 1962–1999. The bankruptcy sample includes all bankruptcy filings as reported in the Wall Street Journal Index (1962–1980), The SDC Database (Reorganizations module 1980–1999), SEC filings (1978–1999) and the CCH Capital Changes Reporter. As such, the bankruptcy data includes most of the bankruptcy filings between 1962–1999 of publicly traded companies on the NYSE, AMEX and NASDAQ stock exchanges. There are a total of 1461 bankruptcies in our initial sample.

exchange (and not since the firm's incorporation). We considered data only after July 1962 as CRSP contains AMEX data from then onwards. As indicated, bankruptcy occurs most often early in the life of a firm. The average age at the time of bankruptcy is computed to be 9.23 years, the median age is 7 years.

For the first sections of this paper, to be consistent with the previous literature, all estimation is done on a yearly basis. For the explanatory variables: (i) all accounting data is taken from COMPUSTAT (annual – active and research files), and (ii) all market data is taken from CRSP. Both the accounting and market data are lagged (in terms of their usage) so that they are observable by the market at the beginning of each year. The market prices for each year are considered as of the December of the previous year. The accounting data are lagged so as to ensure that the accounting report's release occurs at least three months prior to the end of the fiscal year. Otherwise, we use the previous year's accounting data. Again, this is an attempt to ensure that the accounting data is available to the market at the time of estimation. In the case where accounting or market data are missing, we substitute the previous available observations. We do not consider any interim financial statements that may be available after the release of the annual financial statements.

PANEL A: BANKRUPTCIES AS PERCENTAGE OF ACTIVE FIRMS DURING 1962-1999



PANEL B: BANKRUPTCIES AS PERCENTAGE OF ACTIVE FIRMS

The following table gives the number of bankruptcies, total number of active firms and the percentage of bankruptcies each year during the sample period of 1962-1999.

YEAR	# of bankrupt firms	# of active firms	% of bankrupt to active firms	YEAR	# of bankrupt firms	# of active firms	% of bankrupt to active firms
1963	0	1294	0.00%	1981	35	4090	0.86%
1964	2	1362	0.15%	1982	36	4699	0.77%
1965	2	1444	0.14%	1983	50	4645	1.08%
1966	1	1536	0.07%	1984	72	5112	1.41%
1967	0	1612	0.00%	1985	73	5240	1.39%
1968	0	1721	0.00%	1986	94	5250	1.79%
1969	2	1842	0.11%	1987	52	5584	0.93%
1970	12	1988	0.60%	1988	80	5911	1.35%
1971	6	2089	0.29%	1989	74	5814	1.27%
1972	11	2208	0.50%	1990	79	5726	1.38%
1973	27	4355	0.62%	1991	69	5689	1.21%
1974	29	4310	0.67%	1992	45	5874	0.77%
1975	13	4215	0.31%	1993	36	6246	0.58%
1976	19	4197	0.45%	1994	29	6971	0.42%
1977	22	4229	0.52%	1995	40	7387	0.54%
1978	19	4182	0.45%	1996	31	7620	0.41%
1979	14	4110	0.34%	1997	41	8023	0.51%
1980	33	4017	0.82%	1998	49	8079	0.61%

Figure 4. Bankruptcies as % of firms listed on NYSE–AMEX–NASDAQ. The above figure shows the bankruptcies as percentage of active firms every year over the sample period of 1962–1999. The bankruptcy sample includes all bankruptcy filings as reported in the Wall Street Journal Index (1962–1980), The SDC Database (Reorganizations module 1980–1999), SEC filings (1978–1999) and the CCH Capital Changes Reporter. As such, the bankruptcy data includes most of the bankruptcy filings between 1962–1999 of publicly traded companies on the NYSE, AMEX and NASDAQ stock exchanges. There are a total of 1461 bankruptcies in our initial sample.

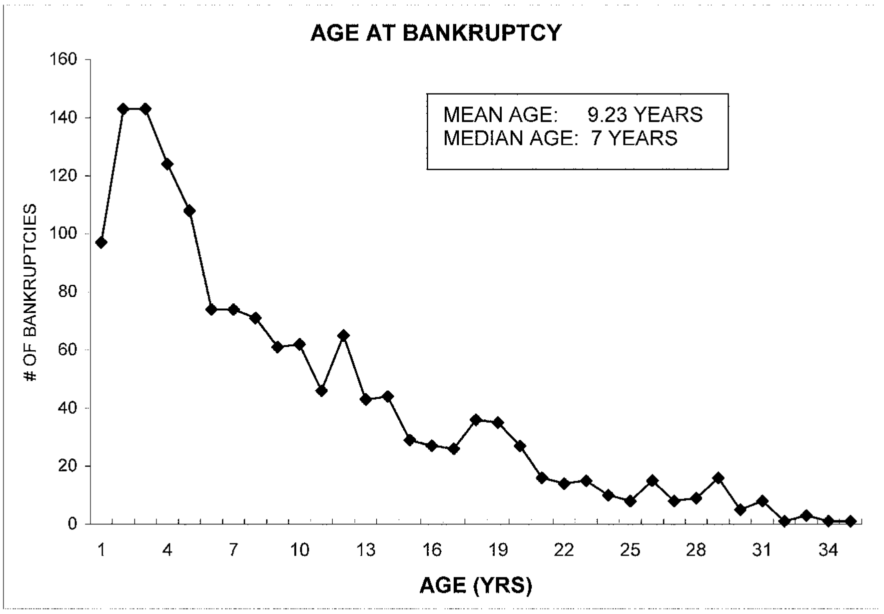


Figure 5. Age at bankruptcy. The above figure shows the age of the firm at the time of bankruptcy over the sample period of 1962–1999. The bankruptcy sample includes all bankruptcy filings as reported in the Wall Street Journal Index (1962–1980), The SDC Database (Reorganizations module 1980–1999), SEC filings (1978–1999) and the CCH Capital Changes Reporter. As such, the bankruptcy data includes most of the bankruptcy filings between 1962–1999 of publicly traded companies on the NYSE, AMEX and NASDAQ stock exchanges. There are a total of 1461 bankruptcies in our initial sample. The age of the firm is defined as the number of years since the firm was listed on AMEX, NYSE or NASDAQ. The mean and median age at the time of bankruptcy are also given (9.23 years and 7 years respectively).

When using monthly observation intervals, a similar procedure is employed, but modified to the shorter observation period. For example, both the accounting and market data are lagged by a month so that they are available to the market at the time of the estimation. If the accounting data is not yet available, we use the previous quarter's accounting data. In the case where accounting or market data are missing, we substitute the previous available observation.

To be consistent with the previous literature (see Shumway (2001)), the accounting data is truncated at the 1 and 99 percentiles for all the variables. This truncation is intended to remove outliers (typos, recording errors) from the accounting database. All the tests in the paper were run with untruncated data with no significant impact on the results.

3. The Statistical Model

This section describes the statistical model used for the subsequent hazard rate estimation. Suppose that we observe a total of n firms ($i = 1, \ldots, n$) with bankruptcy times for each firm denoted by τ_B^i. Following the standard survival analysis literature (see Kiefer (1988), Klein and Moeschberger (1997)), we define the hazard rate or intensity rate for the bankruptcy time τ_B^i, a random variable, as:

$$\lambda_B^i(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq \tau_B^i < t + \Delta t \mid \tau_B' \geq t)}{\Delta t}. \tag{1}$$

Define the point process for this bankruptcy time as $N_t^i = \{1 \text{ if } \tau_B^i \leq t;\ 0 \text{ otherwise}\}$.

In aggregate, we observe these n firms discretely (at times $t = 1, 2, 3 \ldots, T$), starting at the beginning ($t = 1$) until the end ($t = T$) of our sample period. However, the observations of any particular firm i continue from some starting time t_i (the start of its equity trading when it is not bankrupt) until some time $T_i < T$ when the firm experiences bankruptcy (τ_B^i) or is censored (T_i). Censoring means that the firm is observed at time T_i but not at time $T_i + 1$. Time T_i is usually the last date in our sample period, but it could be otherwise. For example, the firm could experience a merger and vanish from the data set. We assume that censoring is non-informative.

We define the random time $Y_i = \min(\tau_B^i, T_i)$ corresponding to the last date that we observe firm i over our observation period. Also, let X_{it} denote the time varying covariates (market variables, accounting variables etc.) of firm i at time t.

Following Allison (1982), we define the discrete time conditioned hazard rate[8] process as:

$$P_t^i = \Pr[\tau_B^i = t \mid \tau_B^i \geq t, X_{it_i}, \ldots, X_{iY_i}] \quad \text{for} \quad t_i + 1 \leq t \leq T_i. \tag{2}$$

We assume given the information sets generated by the covariates X_{it} for all firms $i = 1, \ldots, n$ and times $t = 1, \ldots, T$ that the default times τ_B^i are independent. This conditional independence is a standard assumption in hazard rate modeling.

The data observed for firm i is $\{N_{Y_i}^i\}$. This data is either a 1 if $Y_i = \tau_B^i$ and bankruptcy occurs at time τ_B^i, or the data is a zero if $Y_i = T_i$ and the firm is censored at time T_i. The likelihood function L of firm i's data conditioned on the covariates is:

$$\begin{aligned} L(N_{Y_i}^i \mid X_{it_i}, \ldots, X_{iY_i}) &= \begin{cases} \Pr(\tau_B^i = Y_i) & \text{if} \quad N_{Y_i}^i = 1 \\ \Pr(\tau_B^i > Y_i) & \text{if} \quad N_{Y_i}^i = 0 \end{cases} \\ &= \Pr(\tau_B^i = Y_i)^{N_{Y_i}^i} \Pr(\tau_B^i > Y_i)^{1 - N_{Y_i}^i}, \end{aligned} \tag{3}$$

[8] Note that $P_t^i \approx \lambda_B^i(t)\Delta t$ where Δt is the observation interval and that the conditioning is using the entire history of the covariates $\{X_{it_i}, \ldots, X_{iY_i}\}$.

where for simplicity of notation $\Pr(\cdot) = \Pr(\cdot \mid X_{it_i}, \ldots, X_{iY_i})$.

Using properties of conditional probabilities we can write

$$\Pr(\tau_B^i = Y_i) = P_{Y_i}^i \prod_{t=t_i+1}^{Y_i-1} [1 - P_t^i] \quad \text{and} \quad \Pr(\tau_B^i > Y_i) = \prod_{t=t_i+1}^{Y_i} [1 - P_t^i].$$

Substituting into the likelihood function and taking the logarithm yields the log-likelihood function

$$\begin{aligned} \log L(N_{Y_i}^i \mid X_{it_i}, \ldots, X_{iY_i}) &= N_{Y_i}^i \log\left(P_{Y_i}^i \prod_{t=t_i+1}^{Y_i-1} [1 - P_t^i]\right) \\ &\quad + (1 - N_{Y_i}^i) \log\left(\prod_{t=t_i+1}^{Y_i} [1 - P_t^i]\right) \\ &= N_{Y_i}^i \log\left(\frac{P_{Y_i}^i}{(1 - P_{Y_i}^i)}\right) + \sum_{t=t_i+1}^{Y_i} \log(1 - P_t^i). \qquad (4) \end{aligned}$$

Using the independence across firms, the log-likelihood function for the entire data is

$$\begin{aligned} &\log L(N_{Y_1}^1, \ldots, N_{Y_n}^n \mid X_{1t_1}, \ldots, X_{1Y_1}; \ldots; X_{nt_n}, \ldots, X_{nY_n}) \\ &= \sum_{i=1}^{n} N_{Y_i}^i \log\left(\frac{P_{Y_t}^i}{(1 - P_{Y_i}^t)}\right) + \sum_{i=1}^{n} \sum_{t-t_i}^{Y_i} \log(1 - P_t^i). \end{aligned}$$

For a point process we have that $N_{Y_i}^i = \sum_{t=t_i+1}^{Y_i} [N_t^i - N_{t-1}^i]$. Substitution yields the final result.

$$\begin{aligned} &\log L(N_{Y_1}^1, \ldots, N_{Y_n}^n \mid X_{1t_1}, \ldots, X_{1Y_1}; \ldots; X_{nt_n}, \ldots, X_{nY_n}) \\ &= \sum_{i=1}^{n} \sum_{t=t_i+1}^{Y_i} [N_t^i - N_{t-1}^i] \log\left(\frac{P_t^i}{(1 - P_t^i)}\right) + \sum_{i=1}^{n} \sum_{t=t_i+1}^{Y_i} \log(1 - P_t^i). \qquad (5) \end{aligned}$$

But this is identical (with respect to maximization over the data) to the log-likelihood function for a regression analysis of dichotomous dependent variables (see Cox (1970)). This implies that the computer programs used for the analysis of dichotomous data can also be used to estimate discrete-time hazard rate models.[9]

[9] Brown (1975) noted that the discrete-time hazard models could be estimated using programs for the analysis of dichotomous data. We use the statistical package SAS to do the maximum likelihood estimation.

We use a logistic model with parameters (α_t, β') and time varying covariates X_{it} for our discrete time hazard rate:

$$P_t^i = 1/[1 + \exp(-\alpha_t - \beta' X_{it})]. \tag{6}$$

Formally, we see that the likelihood function for the discrete-time survival model as derived above, under non-informative censoring, coincides with the binomial likelihood function that would be obtained by treating the annual bankruptcy indicator variable as independent binomials. This observation is important in the construction of statistics for the estimated parameters. It is important to note that we do not assume that the multiple observations for each firm are independent.[10]

4. Bankruptcy Models for Non-financials with No Industry Effects

This section investigates the re-estimation of the existing bankruptcy prediction models using our augmented bankruptcy database. This database is reduced somewhat to enable a comparison to the existing literature. First, we restrict our analysis to only NYSE and AMEX listed companies and yearly observation intervals. Of the 1461 Chapter 11 bankruptcy filings in our database, only 585 are listed on either the NYSE or AMEX stock exchange. Also, in keeping with the previous literature, we initially consider only non-financial firms. Finally, to utilize the data available in both COMPUSTAT and CRSP, our bankruptcies are further reduced to 464 firms. For comparison purposes, Shumway (2001) had 300 bankruptcy filings in his database. Our sample includes all of his bankruptcies.

For bankruptcy estimation, in every year of the observation period, both firms filing for bankruptcy and those not filing for bankruptcy are included. A total of 5282 firms listed on both the NYSE and AMEX are included in the hazard rate estimation. If a company had multiple bankruptcy filings, we considered only the year of the first bankruptcy filing.

The dependent variable equals one if a bankruptcy filing is made in a particular year, zero otherwise. A firm contributes an observation in the estimation equation for every year after it starts trading until the end of the observation period or until it files for bankruptcy. A firm that merges or is taken over by another firm contributes observations until the restructuring. This firm always has the dependent variable equal to zero. A firm that is de-listed but did not file for bankruptcy would contribute observations until the delisting year. This firm's dependent variable would always be zero. A firm that was de-listed and that ultimately filed for bankruptcy would contribute observations until the delisting year with the de-listing year counted as a bankruptcy filing.

We estimate the hazard rate model with Altman's (1968), Zmijewski's (1984) and Shumway's (2001) variables. We restrict ourselves to these variables since they

[10] See Allison (1982, 1989).

are well accepted explanatory variables often used in bankruptcy prediction models. Altman's variables are the ratios of working capital to total assets (WC/TA), retained earnings to total assets (RE/TA), earnings before interest and taxes to total assets (EBIT/TA), market equity to total liabilities (ME/TL), and sales to total assets (SL/TA). Zmijewski's variables are the ratio of net income to total assets (NI/TA), the ratio of total liabilities to total assets (TL/TA), and the ratio of current assets to current liabilities (CA/CL). We use the best performing model in Shumway (2001) with variables (NI/TA), (TL/TA), relative size (RSIZ) defined as the logarithm of each firm's equity value divided by the total NYSE/AMEX market equity value, excess return (EXRET) defined as the monthly return on the firm minus the value-weighted CRSP NYSE/AMEX index return cumulated to obtain the yearly return, and the stock's volatility (SIGMA). The stock's volatility for the present year (SIGMA) is computed as the sample standard deviation using the last sixty observable daily market prices.

Summary statistics for these independent variables are contained in Table I. As indicated, there is significant cross sectional variation in these variables. For example, net income to total assets (NI/TA) varies from a minimum of -1.0454 to a maximum of 0.2753.

Table II contains the hazard rate estimation for the Altman (1968), Zmijewski (1984), Shumway (2001), private firm and public firm models using data from 1962–1999. Given are the estimated coefficients and a chi-squared statistic (χ^2) for the significance of each coefficient and its associated P-value. These chi-squared statistics provide an in-sample goodness of fit measure.

To gauge the out-of-sample model performance we also perform a bankruptcy prediction test. The bankruptcy prediction test is the same test as reported by Shumway (2001). First, we re-estimated these models using only data from 1962–1990 for use in predicting bankruptcies for the period 1991–1999. Year by year from 1991–1999 we rank the firms into deciles based on their computed bankruptcy probabilities. The firms most likely to default in the subsequent year are placed into the first decile, the next most likely to default in the second decile, and so forth. Second, year by year from 1991–1999, we count the number of firms in each decile that actually file for bankruptcy. The number of firms in each decile that file for bankruptcy across the 1991–1999 observation period are aggregated and for each decile the percentage of bankrupt companies occurring in that decile are reported in Panel B of Table II.

For the in sample test, the hazard rate model coefficients reported in Table II are similar to those reported in Shumway (2001) in terms of sign, magnitude and statistical significance. Differences between Shumway's and our estimates are due to our augmented database. Concentrating on Shumway's model in Table II, we see that all of the market and accounting variables are significant except NI/TA. Furthermore, the signs of these coefficients are consistent with simple economic intuition. Higher NI/TA, lower TL/TA, increasing excess return (EXRET), increas-

Table I. Summary statistics of independent variables

The following table gives summary statistics of the independent variables and where applicable, the relevant COMPUSTAT item numbers (in brackets) used to calculate the variable. The sample period is 1962–1999. The summary statistics for annual observations are given in Panel A. There are a total of 1,197 bankruptcies with relevant information in COMPUSTAT and CRSP. Firms listed on all the three main exchanges – NYSE, AMEX and NASDAQ are considered. There are a total of 17,460 firms and 167,617 observations in the sample. Each firm year is considered a separate observation.

Panel A: Summary statistics for annual observations

Variable	Definition	Median	Mean	Min	Max	Std
WCTA	Working capital/Total assets (179/6)	0.2707	0.2692	−0.4545	0.8445	0.2433
RETA	Retained earnings/Total assets (36/6)	0.1506	0.0147	−4.1502	0.7641	0.7013
EBTA	Earnings before interest and taxes/Total assets (178/6)	0.0791	0.0517	−0.9020	0.3898	0.1797
METL	Market equity/Total liabilities (ME/181)	1.3413	5.5948	0.0191	106.556	14.4470
SLTA	Sales/Total assets (12/6)	1.0991	1.1954	0.0325	4.9073	0.9060
CACL	Current assets/Current liabilities (4/5)	2.0284	2.7217	0.2106	19.7696	2.7365
NITA	Net income/Total assets (172/6)	0.0398	0.0030	−1.0454	0.2753	0.1816
TLTA	Total liabilities/Total assets (181/6)	0.5327	0.5347	0.0398	1.2585	0.2479
EXRET	Excess annual return over the value-weighted NYSE, AMEX and NASDAQ return (CRSP)	−0.0553	−0.0021	−0.9694	2.2639	0.5198
RSIZ	Log (firm's market capitalization/Total NYSE, AMEX, NASDAQ market cap) (CRSP)	−10.882	−10.745	−14.994	−5.7767	1.9958
SIGMA	Std. dev. of 60 prior day's stock returns (CRSP)	0.5247	0.6538	0.0938	2.7123	0.4728

Panel B: Summary statistics for monthly observations

The summary statistics for monthly observations are given in Panel B. The sample period is 1962–1999. There are a total of 1,120 bankruptcies with relevant information in COMPUSTAT and CRSP. Firms listed on all the three main exchanges – NYSE, AMEX and NASDAQ are considered. There are a total of 16,816 firms and 1,900,834 observations in the sample. Each firm month is considered a separate observation.

Variable	Definition	Median	Mean	Min	Max	Std
NITA	Net income/Total assets (69/44)	0.0089	−0.0028	−0.3440	0.0949	0.0590
TLTA	Total Liabilities / Total Assets (54/44)	0.5395	0.5361	0.0330	1.1830	0.2499
EXRET	Excess annual return over the value-weighted NYSE, AMEX, NASDAQ return (CRSP)	−0.0080	−0.0016	−0.3430	0.4916	0.1294
RSIZ	Log (firm's market capitalization/Total NYSE, AMEX, NASDAQ market cap) (CRSP)	−10.780	−10.645	−14.831	−5.735	1.9639
SIGMA	Std. dev. of the past month's stock returns (CRSP)	0.4602	0.5869	0.0000	2.6337	0.4572

Table II. Altman (1968), Zmijewski (1984), Shumway (2001), private firm and market variables models with yearly observations
The hazard model is estimated for 1962–1999 data with yearly observations. The data set is limited to NYSE and AMEX non-financial companies to facilitate comparison with previous literature. Parameter estimates are given first followed by chi-square values in brackets. The chi-square of the likelihood ratio test for the model fit is reported in the model fit column. There are 5282 firms in the sample.

Panel A: Estimation using 1962–1999 data (NYSE-AMEX non-financial companies)										
	Altman (1968)		Zmijewski (1984)		Shumway (2001)		Private firm model		Market variables	
INTERCEPT	−4.5088***	(549.37)	−7.6379***	(709.49)	−8.8085***	(416.58)	−7.5532***	(2120.66)	−7.5370***	(349.62)
WCTA	−1.2417***	(21.46)								
RETA	0.0207	(0.04)								
EBTA	−4.6448***	(163.95)								
METL	−0.6269***	(78.32)								
SLTA	0.2591***	(28.84)								
LNAGE	0.0858	(1.70)	0.2347***	(16.87)						
NITA			−1.7777***	(69.06)	−0.0364	(0.02)	−1.6901***	(67.71)		
TLTA			3.4902***	(182.82)	2.7926***	(162.02)	3.7713***	(294.83)		
CACL			−0.1500***	(6.02)						
EXRET					−1.6566***	(115.85)			−1.6015***	(115.32)
RSIZ					−0.2275***	(39.43)			−0.2453***	(49.26)
SIGMA					0.9572***	(93.79)			1.2488***	(193.47)
MODEL FIT	607.08***		593.13***		1259.10***		572.64***		1100.68***	
# bankruptcies (#firm-year obs)	394	(67276)	404	(71922)	404	(72184)	409	(72682)	430	(75476)

Table II. Continued

Panel B: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated and then the companies are grouped into deciles based on the default probabilities. The number of bankruptcies in each decile for each year are aggregated over 1991–1999 and reported in the above panel.

Percentage of the 125 bankruptcies between 1991–1999 predicted by the model in each decile

Decile	Altman (1968)	Zmijewski (1984)	Shumway (2001)	Private firm model	Market variables
1	63.2	43.20	74.40	44.00	72.00
2	14.4	14.40	12.00	16.00	13.60
3	4.8	4.00	5.60	8.80	5.60
4	4	8.80	3.20	5.60	3.20
5	4.8	8.00	2.40	8.00	3.20
6	2.4	5.60	0.80	4.80	
7	4	4.80	1.60	1.60	0.80
8	0.8	5.60		4.80	
9	0.8	4.80		5.60	
10	0.8	0.80		0.80	1.60

Panel C: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data – area under the ROC curve

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated for each firm and then an ROC curve is constructed from the predicted probabilities and the status of the firm in that year. The area under the ROC curve is computed as a measure of the forecasting accuracy of the model. The above procedure is repeated on 1000 bootstrapped samples for each model and the descriptive statistics are computed for the area under the ROC curve.

Model	Mean	Median	Min	Max	Std. dev
Altman (1968)	0.8662	0.8669	0.8141	0.9152	0.0156
Zmijewski (1984)	0.7392	0.7395	0.6704	0.8078	0.0226
Shumway (2001)	0.9113	0.9120	0.8682	0.9421	0.0109
Private firm model	0.7513	0.7513	0.6779	0.8204	0.0216
Market variables	0.9022	0.9027	0.8478	0.9335	0.0125

ing size (RSIZ), and decreasing volatility (SIGMA) indicates a smaller chance of bankruptcy.

Based on forecasting ability, Shumway's model significantly outperforms both Altman's and Zmijewski's model. For Shumway's model 74.4 percent of the bankruptcies are correctly identified (in the first decile), for Altman's model the corresponding number is 63.2 percent and for Zmijewski's model it is 43.2 percent. For the top two deciles (in aggregate) the correct predictions are Shumway (86.4 percent), Altman (77.6 percent) and Zmijewski (57.6 percent). In terms of incorrect predictions, aggregating across the bottom 5 deciles, Shumway misclassifies only 2.4 percent of the firms, Altman 8.8 percent, and Zmijewski 21.6 percent. This forecasting prediction re-confirms the validity of Shumway's hazard rate model.

For subsequent comparison, we estimated two additional models. The first is called the private firm model, adopting terminology from Moody's (2000a, b). A private firm model is one that is useful for firms without publicly traded equity. Our private firm model is Shumway's model without the market variables and only using the accounting variables. The results are also contained in Table II.

In this private firm model, both NI/TA and TL/TA are significant with the appropriate signs. The forecasting accuracy is not as good as with the previous model that we subsequently call the "public firm" model. The top decile includes only 44 percent of the companies that filed for bankruptcy. The top two deciles contain only 60 percent. This model compares favorably, however, with both the Altman and Zmijewski models.

Second, we estimated Shumway's model with just the market variables providing a modified public firm model. The results are contained in Table II. Here, all the market variables are significant and with the appropriate signs. Furthermore, the market variables appear to predict bankruptcy better than do the accounting variables, correctly identifying 72 percent of the bankruptcies in the first decile and 85.6 percent of the bankruptcies in the top two deciles.

In fact, the performance of the market variable model is comparable to the more complex public firm model including both market and accounting variables. This is consistent with an efficient market hypothesis stating that market prices reflect all publicly available information regarding bankruptcy, including that contained in the accounting variables. This hypothesis is more easily validated using monthly observation periods[11] in a subsequent section of this paper. For yearly observations, we utilize the accounting plus market variables model in the subsequent investigation. This is called the public firm model.

An alternate and widely used measure of out-of-sample accuracy is the area under the power[12] or ROC curve (receiver operating characteristic) (see Sobehart et al. (2000, 2001)). For comparison across models, the area under the ROC curve

[11] For yearly data, updating of both accounting and market variables occurs at the same frequency. For monthly data, market variables are updated more quickly than are the quarterly accounting variables.

[12] We thank the referee for suggesting the ROC curve analysis.

is usually measured relative to the area of the unit square. A value of 0.5 indicates a random model with no predictive ability, and a value of 1.0 indicates perfect discrimination. To compute the ROC measure, we use the 1962–1990 period to estimate the parameters of each model and then we use these parameter estimates to forecast the bankruptcy probabilities of the sample firms over the out-of-sample period 1991–1999. To obtain the sampling distribution for statistical significance, we generate 1000 bootstrap samples using the bankruptcy probabilities over 1991–1999 in order to reduce dependence on any particular sample (See Efron and Tibshirani (1993)). For each of these 1000 samples we calculate the area under the ROC curve. The descriptive statistics for the area under the ROC curve based on this sampling distribution are reported in Panel C of Table II.

The area under the ROC curve for Shumway's model is 0.9113 and this is statistically significantly[13] (at 1% confidence level) different from the area under the ROC curve for the other models. The modified public firm model using only the market variables also performs well with an area under the ROC curve equaling 0.9022. Altman's (1968) model, with an ROC area of 0.8662, performs well when compared to the other accounting based models such as Zmiejewsky (1984) (area under the ROC curve = 0.7392) and the private firm model (area under the ROC curve = 0.7513). These results are consistent with the results previously reported for the out-of-sample decile analysis. Combined, the ROC analysis confirms the conclusion that the Shumway and the modified public firm model are the best performing of the five models analyzed in this section.

5. Bankruptcy Models for Non-financials on AMEX and NYSE with Industry Effects

This section investigates the importance of including industry effects in the hazard rate model estimation. For easy comparison with the preceding section, this estimation is done for yearly observation intervals using only non-financial firms with equity trading on either the AMEX or NYSE. The estimation with financial firms and firms trading on NASDAQ is contained in the next section.

As we previously discussed, Figure 2 provides bankruptcies by industry grouping. We separated ten industries by the 4-digit SIC code. For estimation purposes, this partitioning is too fine. Instead, we combine these classifications into four sub-groups.[14] The four industry groupings are: (i) finance, insurance and real estate, (ii) transportation, communications and utilities, (iii) manufacturing and mineral, and (iv) miscellaneous industries (the complement to (i)–(iii)). These clas-

[13] To check the statistical significance for the area under ROC curves, we run a t-test. We also perform non-parametric tests that confirm the results.

[14] The construction of the industry variable is as follows: Miscellaneous industries (sic code is in (1–1000, 1500–1800, 5000–6000, 7000–8900)), Manufacturing and mineral industries ((sic code is in (1000–1500, 2000–4000)), Transportation, communications and utilities ((sic code is in (4000–5000)), Finance, insurance and real estate (sic code is in (6000–6800)).

sifications were selected based on different regulatory environments and different asset/product structures. Of course, for this section's analysis we exclude firms in the finance, insurance and real estate grouping.

To investigate the impact of industry effects on bankruptcy prediction, we estimate both the public firm and private firm models of the preceding section with intercept and slope dummy variables for the industry groupings. This estimation is conceptually equivalent to running a separate hazard rate model for each industry classification. The results are contained in Table III.

Table III contains the tests for the private firm model, both in-sample estimation and out-of-sample prediction. As seen, all industry dummy variables are significant except for NITA*IND2 and TLTA*IND3. Nonetheless, the entire set of industry variables is statistically significant using a chi-squared test. This is strong evidence consistent with an industry effect.

Examining the signs of the coefficients, the intercepts indicate that the unconditional default probabilities differ across industry – from largest to smallest they are: the miscellaneous grouping (industry 1), manufacturing and minerals (industry 2), and transportation, communications and utilities (industry 3). A negative coefficient on a dummy variable indicates that the relevant industry's default probability is, on average, less than that of industry 1. For example, this is true for industry 2 which has a negative coefficient of (−0.8968). This evidence is consistent with the bankruptcy experience by industry illustrated in Figure 2. The slope dummy variables indicate different sensitivities to the net income variable NITA and total liabilities variable TLTA. Both industries 2 and 3 exhibit less sensitivity to NITA and more sensitivity to TLTA than does industry 1.

With respect to out-of-sample prediction, the industry variables provide a modest improvement. The relevant comparison is the private firm model with industry effects in Table III versus the private firm model in Table II. The model with the industry effect in Table III classifies two less firms correctly in the first decile. However, the correct classifications for the private firm model with industry effects for the top two deciles in Table III are 60.8 percent, slightly greater than that for the private firm model in Table II (60 percent). An improvement also occurs in reducing the incorrect classifications. For the bottom five deciles, the private firm model with industry effects in Table III has 15.2 percent compared with 17.6 percent for the private firm model in Table II.

Table III also contains the tests for the public firm model, both in-sample estimation and out-of-sample prediction. Similar to the private firm model, these tests confirm the importance of an industry effect. Indeed, for the in-sample estimation, all industry dummy variables are significant except for IND3, NITA*IND2 and TLTA*IND3. As before, the entire set of industry variables is statistically significant using a chi-squared test. The signs of the industry variables have the identical interpretation as in the case of the private firm model. With respect to out-of-sample prediction, the industry variables leave the forecasting ability of the model unchanged. The relevant comparison is the public firm model in Panel B of

Table III. Private firm and public firm models with industry effects using yearly observations

Panel A: Estimation using 1962–1999 data (NYSE-AMEX non-financial companies)

The hazard model is estimated for 1962–1999 data with yearly observation. The data set is limited to NYSE and AMEX companies to facilitate comparison with previous literature. Parameter estimates are given first followed by chi-square values in brackets. The chi-square of the likelihood ratio test for the model fit is reported in the model fit column. There are 5282 firms in the sample.

	Private firm model with industry effects		Public firm model with industry effects	
INTERCEPT	−6.8590***	(560.28)	−8.2221***	(273.53)
NITA	−1.4343***	(16.59)	0.2318	(0.30)
TLTA	3.2148***	(70.71)	2.2177***	(34.47)
EXRET			−1.6587***	(117.13)
RSIZ			−0.2237***	(37.10)
SIGMA			0.9524***	(92.84)
IND2	−0.8968***	(6.40)	−0.8231***	(5.59)
IND3	−1.4319*	(3.00)	−0.1790	(0.05)
NITA*IND2	−0.1305	(0.08)	−0.1602	(0.10)
TLTA*IND2	0.8579*	(3.30)	0.9717**	(4.33)
NITA*IND3	−2.1303***	(5.73)	−2.7675***	(9.05)
TLTA*IND3	0.3470	(0.10)	−0.4049	(0.16)
Model Fit	613.13***		1275.27***	
# bankruptcies	409		404	
# firm-year obs.	72682		72682	

Panel B: Are industry dummies significant?

To check the whether industry dummies are significant a likelihood ratio test is performed considering the model with industry dummies and interaction variables as the unconstrained model and the model without industry dummies as a constrained model.

	Private firm model with industry effects	Public firm model with industry effects
UNCONSTRAINED −2LOG(LF)	4444.51	3724.91
CONSTRAINED −2LOG(LF)	4485	3741
CHI-SQUARE	40.49	16.17
P-VALUE	0.0000	0.013

Table III. Continued

Panel C: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated and then the companies are grouped into deciles based on the default probabilities. The number of bankruptcies in each decile for each year are aggregated over 1991–1999 and reported in the above panel.

Percentage of the 125 bankruptcies between 1991–1999 predicted by the model in each decile

Decile	Private firm model with industry effects	Public firm model with industry effects
1	42.40	72.80
2	18.40	13.60
3	8.00	5.60
4	11.20	3.20
5	4.80	0.80
6	4.00	1.60
7	2.40	1.60
8	4.80	0.80
9	1.60	
10	2.40	

Panel D: Descriptive statistics for the area under the ROC curve

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated for each firm and then an ROC curve is constructed from the predicted probabilities and the status of the firm in that year. The area under the ROC curve is computed as a measure of the forecasting accuracy of the model. The above procedure is repeated on 1000 bootstrapped samples for each model and the descriptive statistics are computed for the area under the ROC curve.

Model	Mean	Median	Min	Max	Std. dev
Private firm model with industry effects	0.7646	0.7651	0.6961	0.8280	0.0208
Public firm model with industry effects	0.9101	0.9108	0.8636	0.9426	0.0116

Table II with the public firm model with industry effects in Panel C of Table III. As seen therein, the numbers in each decile are almost identical, except perhaps for a minor switching of three firms between deciles one and two.

Panel D of Table III gives the descriptive statistics for the area under the ROC curve for both the private and public firm models with industry effects based on the sampling distribution using 1000 bootstrapped samples as previously described. As expected, the public firm model with industry effects (area under the ROC curve

= 0.9101) performs better than the private firm model with industry effects (area under the ROC curve = 0.7646). The addition of industry effects has a marginal, although statistically significant impact[15] (at 1% confidence level), on the accuracy of both the public and private firm models. Comparing Panel C of Table II with Panel D of Table III, we can see that the addition of industry effects increases the area under the ROC curve for the private firm model from 0.7513 to 0.7646 and for the public firm model from 0.9022 to 0.9101. Combined with the remaining results in Table III, it appears that industry effects are statistically significant in-sample, but do not dramatically increase out-of-sample accuracy.

6. Bankruptcy Models for all Firms on the AMEX, NYSE and NASDAQ with Industry Effects

This section fits a hazard rate model using yearly observation intervals with industry effects to both financial and non-financial firms listed on the AMEX, NYSE and NASDAQ. As such, it utilizes our entire database consisting of 1197 bankruptcies and 17,460 different firms. The estimation is performed in two stages. Stage one fits the private and public firm models employed for the non-financials listed on the AMEX/NYSE, but this time to the expanded database. This estimation allows a direct comparison with the private and public firm models in Table II. These estimates, both in-sample and out-of-sample prediction, are contained in Table IV. Stage two fits the private and firm models with industry effects. This enables a direct comparison to the private and public firm models with industry effects in Table III. These estimates, both in-sample and out-of-sample prediction, are contained in Table V.

Let us first consider the estimates for the public firm and private firm models in Table IV. For the in-sample estimation, the coefficients are similar in magnitude and sign to the estimates for the public firm and private firm models obtained using the restricted database in Table II. The most striking difference occurs in the out-of-sample prediction. For the private firm model, the prediction accuracy declines significantly in the expanded database. Indeed, the first decile explains only 31.76 percent of the bankruptcies, and the first two deciles explain only 37.94 percent of the bankruptcies. This prediction is much less accurate than in the restricted database (first decile is 44 percent, first two deciles are 60 percent). For the public firm model, the prediction accuracy also declines. The first decile explains 60.29 percent of the bankruptcies for the public firm model in Table III. The first two deciles explain 78.82 percent of the bankruptcies. This is less than the prediction accuracy for the public firm model contained in Table II (first decile is 74.40 percent, first two deciles are 86.4 percent). The results from the area under the ROC curve reported in Panel C confirm the results from the decile analysis.

[15] To check the statistical significance for the area under ROC curves, we run a t-test. We also perform non-parametric tests that confirm the results.

Table IV. Private firm and public firm models with financials included using yearly observations

Panel A: Estimation using 1962–1999 data for NYSE-AMEX-NASDAQ companies

The hazard model is estimated for 1962–1999 data with yearly observation intervals. The data set includes NASDAQ companies in addition to NYSE and AMEX companies. Financial companies are also included. Parameter estimates are given first followed by chi-square values in brackets. The chi-square of the likelihood ratio test for the model fit is reported in the model fit column. There are 17460 firms in the sample.

	Private firm model		Public firm model	
INTERCEPT	−6.4215***	(5603.52)	−9.4398***	(1362.01)
NITA	−1.6408***	(297.03)	−0.1680	(2.16)
TLTA	2.2816***	(375.38)	2.5447***	(457.67)
EXRET			−1.6406***	(282.45)
RSIZ			−0.1667***	(55.51)
SIGMA			0.6803***	(155.90)
MODEL FIT	899.53***		2282.22***	
# bankruptcies	1090		1066	
# firm-year obs.	146102		146102	

Panel B: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated and then the companies are grouped into deciles based on the default probabilities. The number of bankruptcies in each decile for each year are aggregated over 1991-99 and reported in the above panel.

Percentage of the 340 bankruptcies between 1991–1999 predicted by the model in each decile

Decile	Private firm model	Public firm model
1	31.76	60.29
2	6.18	18.53
3	15.88	6.76
4	14.71	4.41
5	6.76	3.24
6	9.41	2.35
7	6.18	1.76
8	5.00	0.88
9	2.06	0.59
10	2.06	1.18

Panel C: Descriptive statistics for the area under the ROC curve

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991-1999. Each year the probabilities are calculated for each firm and then an ROC curve is constructed from the predicted probabilities and the status of the firm in that year. The area under the ROC curve is computed as a measure of the forecasting accuracy of the model. The above procedure is repeated on 1000 bootstrapped samples for each model and the descriptive statistics are computed for the area under the ROC curve.

Model	Mean	Median	Min	Max	Std. Dev
Private firm model	0.7228	0.7226	0.6865	0.7533	0.0119
Public firm model	0.8675	0.8676	0.8407	0.8931	0.0081

Table V. Private firm and public firm models with industry effects using yearly observations with financials included

Panel A: Estimation using 1962–1999 data for NYSE-AMEX-NASDAQ companies

The hazard model is estimated for 1962–1999 data with yearly observation intervals using industry intercepts and slopes. The data set includes NASDAQ companies in addition to NYSE and AMEX companies. Financial companies are also included. Parameter estimates are given first followed by chi-square values in brackets. The chi-square of the likelihood ratio test for the model fit is reported in the model fit column. There are 17460 firms in the sample.

	Private firm model with industry effects		Public firm model with industry effects	
INTERCEPT	−5.9090***	(1713.57)	−8.9134***	(972.96)
NITA	−1.0466***	(38.19)	0.3163*	(2.67)
TLTA	2.2036***	(118.46)	2.3573***	(136)
IND2	−0.9619***	(26.54)	−0.6968***	(13.62)
IND3	−0.7524*	(3.37)	−0.1127	(0.08)
IND4	−0.8315**	(3.63)	−0.6524	(2.19)
NITA*IND2	−0.2354	(1.13)	−0.4373*	(3.32)
TLTA*IND2	0.8275***	(9.59)	0.6785***	(6.51)
NITA*IND3	−1.4547***	(11.87)	−1.4423***	(10.36)
TLTA*IND3	0.1174	(0.04)	−0.1211	(0.05)
NITA*IND4	−2.2822***	(27.31)	−1.6189***	(10.28)
TLTA*IND4	−0.5104	(0.94)	−0.1991	(0.14)
EXRET			−1.6135***	(279.17)
RSIZ			−0.1571***	(48.01)
SIGMA			0.6568***	(143.93)
MODEL FIT	1074.02***		2354.26***	
# bankruptcies	1066		1066	
# firm-year obs.	146102		146102	

Panel B: Are industry dummies significant?

To check the whether industry dummies are significant a likelihood ratio test is performed considering the model with industry dummies and interaction variables as the unconstrained model and the model without industry dummies as a constrained model.

	Private firm model with industry effects	Public firm model with industry effects
−2LOG(LF) UNCONSTRAINED	11815.6	10275.76
−2LOG(LF) CONSTRAINED	11990.09	10347.8
χ^2	174.48	72.04
P-VALUE	0.0000	0.0000

Table V. Continued

Panel C: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated and then the companies are grouped into deciles based on the default probabilities. The number of bankruptcies in each decile for each year are aggregated over 1991–1999 and reported in the above panel.

Percentage of 340 bankruptcies between 1991–1999 predicted by the model in each decile

Decile	Private firm model with industry effects	Public firm model with industry effects
1	36.76	59.12
2	16.47	20.00
3	14.41	8.82
4	9.71	2.35
5	3.82	3.53
6	5.88	2.06
7	3.82	0.88
8	4.71	2.06
9	2.35	0.29
10	2.06	0.88

Panel D: Descriptive statistics for the area under the ROC curve

The model estimated with data from 1962–1990 with yearly observation intervals is used to forecast bankruptcy probabilities for 1991–1999. Each year the probabilities are calculated for each firm and then an ROC curve is constructed from the predicted probabilities and the status of the firm in that year. The area under the ROC curve is computed as a measure of the forecasting accuracy of the model. The above procedure is repeated on 1000 bootstrapped samples for each model and the descriptive statistics are computed for the area under the ROC curve.

Model	Mean	Median	Min	Max	Std. dev
Private firm model with industry effects	0.7590	0.7594	0.7179	0.7931	0.0119
Public firm model with industry effects	0.8724	0.8726	0.8452	0.8990	0.0079

Next, let us now consider the estimation including industry effects. For the private firm model in Table V, the in-sample estimation, the industry variables are similar in magnitude and sign to the estimates contained in Table III. All the coefficients but three (NITA*IND2, TLTA*IND4 and TLTA*IND4) are statistically significant. Using a chi-squared test, the industry variables as a whole are statistically significant. This evidence supports the validity of an industry effect in bankruptcy prediction for the expanded database. Next, for the out-of-sample prediction, the industry variables significantly improve performance. Indeed, the first decile explains 36.76 percent of the bankruptcies, and the first two deciles explain 53.23 percent of the bankruptcies. This is a significant improvement over the prediction accuracy for the private firm model in Table IV (first decile is 31.76 percent, first two deciles are 37.94 percent).

For the public firm model in Table V, the industry variables are also similar in magnitude and sign to the estimates for the public firm model contained in Table III. All the coefficients except four are statistically significant (IND3, IND4, TLTA*IND3 and TLTA*IND4). Using a chi-squared test, the industry variables as a whole are statistically significant. This evidence again supports the validity of an industry effect in bankruptcy prediction using the expanded database. Next, for the out-of-sample prediction, the industry variables only marginally improve performance. As seen in Panel C of Table V, the first decile explains 59.12 percent of the bankruptcies, and the first two deciles explain 79.12 percent of the bankruptcies. This is comparable with the prediction accuracy in Panel B of Table IV (first decile is 60.29 percent, first two deciles are 78.82 percent). For prediction using the public firm model, the industry effect is seen to be less important due to the inclusion of the market variables. The area under ROC curve reported in Panel C confirms the results from the out-of-sample decile analysis. The addition of industry effects increases the area under the ROC curve for the private firm model from 0.7228 to 0.7590 (with the difference being statistically significant at 1% confidence level[16]). As before, the inclusion of industry variables only marginally increases the accuracy of the public firm model (from 0.8675 to 0.8724).

Another interesting comparison is between the models without (Table III Panel C) and with financials (Table V Panel C). The predictive ability of the model including financials declines from 42.40 percent for the private firm model with industry effects to 36.76 percent, and from 72.80 percent for the public firm model with industry effects to 59.12 percent. The ROC curve analysis confirms this decline in forecasting accuracy when including financial firms. This reduction in the predictive ability of the model including financial firms is an indication that bankruptcy prediction for financial firms is a more difficult exercise than it is for non-financial firms.

7. Bankruptcy Prediction with Monthly Observation Intervals

This section investigates the private and public firm models of the previous section using monthly observation intervals. The procedure is analogous to that used with yearly observations. For bankruptcy estimation, in every month of the observation period, both firms filing for bankruptcy and those not filing for bankruptcy are included. A total of 17,460 firms are included in the hazard rate estimation. If a

[16] To check the statistical significance for the area under ROC curves, we run a t-test. We also perform non-parametric tests that confirm the results.

company had multiple bankruptcy filings, we considered only the month of the first bankruptcy filing. The dependent variable in the estimation is one if a bankruptcy filing is made in a particular month, zero otherwise. A firm contributes an observation in the estimation equation for every month after it starts trading until the end of the observation period or until it files for bankruptcy. A firm that merges or is taken over by another firm contributes observations until the restructuring. This firm always has the dependent variable equal to zero. A firm that is de-listed but did not file for bankruptcy would contribute observations until the delisting month. This firm's dependent variable would always be zero. A firm that was de-listed and that ultimately filed for bankruptcy would contribute observations until the delisting month with the de-listing month counted as a bankruptcy filing.

We estimate the hazard rate model with variables: net income to total assets (NI/TA), total liabilities to total assets (TL/TA), relative size (RSIZ) defined as the logarithm of each firm's equity value divided by the total NYSE/AMEX market equity value at the end of the month, excess return (EXRET) defined as the monthly return on the firm minus the monthly value-weighted CRSP NYSE/AMEX index return, and the stock's volatility (SIGMA). The stock's volatility for the present month (SIGMA) is computed as the sample standard deviation using the all of the previous month's observable daily market prices. Industry intercept and slope dummy variables are included.

Table VI contains the hazard rate estimation with monthly observation intervals for the private and public firm models using data from 1962–1999. Given are the estimated coefficients and a chi-squared statistic (χ^2) for the significance of each coefficient and its associated P-value. A bankruptcy prediction test is also included. For this prediction test, we re-estimated these models using only data from 1962–1990 for use in predicting bankruptcies for the period 1991–1999. Month by month from 1991–1999 we rank the firms into deciles based on their computed bankruptcy probabilities. The firms most likely to go bankrupt in the subsequent month are placed into the first decile, the next most likely to default in the second decile, and so forth. Second, month by month from 1991–1999, we count the number of firms in each decile that actually file for bankruptcy. The aggregate number of firms in each decile that file for bankruptcy across the 1991–1999 observation period are reported as well as its percentage of the total number of bankruptcies that occurred.

First, consider the private firm model in Table VI. The NI/TA variable (for industry 1 – no dummy variable) is significant and negative, with a P-value of less than 0.0001. Although the NI/TA*IND2 variable is insignificantly different from zero, all the remaining NI/TA*IND variables are negative and significant (for industries 3 and 4). The sum of NI/TA and NI/TA*IND gives the appropriate magnitude for the NI/TA effect. It is negative for industries 2–4 indicating that for these industries as NI/TA increases the likelihood of bankruptcy declines. The total liabilities to total assets (TL/TA) variable for industry 1 is significant with the appropriate sign, i.e., as TL/TA decreases the likelihood of bankruptcy declines.

Table VI. Private firm, public firm and public firm model-efficient markets version with monthly observations

Panel A: Estimation using 1962–1999 data for NYSE-AMEX-NASDAQ companies including financials

Parameter estimates are given first followed by chi-square values in brackets. The chi-square of the likelihood ratio test for the model fit is reported in the model fit column. There are 17460 firms in the sample. The hazard model is estimated for 1962–1999 data with monthly observations using the quarterly accounting variables and industry dummies. The data set is consists of NYSE, AMEX and NASDAQ companies including financials.

Variable	Private firm model		Public firm model		Public firm model – efficient markets version	
INTERCEPT	−10.0856***	(2903.82)	−14.8859***	(1825.30)	−13.9331***	(2332.66)
NITA	−4.9183***	(137.34)	−1.9236***	(19.41)		
TLTA	4.5326***	(377.86)	4.0338***	(326.11)		
IND2	−0.8143***	(10.65)	−0.4597*	(3.53)		
IND3	−0.9550*	(3.31)	−0.0178	(0.01)		
IND4	−1.6629**	(5.16)	−1.2260*	(3.49)		
NITA*IND2	0.5775	(1.03)	0.3414	(0.34)		
NITA*IND3	−3.3136***	(10.02)	−2.5921***	(6.08)		
NITA*IND4	−6.5245***	(39.78)	−3.4877***	(10.46)		
TLTA*IND2	0.6232**	(4.07)	0.3547	(1.44)		
TLTA*IND3	0.1187	(0.03)	−0.3423	(0.35)		
TLTA*IND4	−0.1138	(0.02)	0.2175	(0.09)		
EXRET			−2.6620***	(198.25)	−2.9564***	(264.13)
RSIZ			−0.3475***	(180.03)	−0.4411***	(317.13)
SIGMA			0.8312***	(298.19)	1.2070***	(731.72)
MODEL FIT	2654.62***		4006.73***		3044.97***	

Panel B: Are industry slopes and intercept dummies significant for the public firm model?

To check the whether industry slopes and intercept dummies are significant a likelihood ratio test is performed considering the model with industry dummies and interaction variables as the unconstrained model and the model without industry dummies and interaction variables as a constrained model.

	−2LOG(LF)	χ^2	P-value
Unconstrained	12181.77	74.394	0.0000
Constrained	12256.16		

Table VI. Continued

Panel C: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data

The model estimated with data from 1962–1990 with monthly observations is used to forecast bankruptcy probabilities for 1991–1999. Each month the probabilities are calculated and then the companies are grouped into deciles based on the default probabilities. The number of bankruptcies in each decile for each month are aggregated over 1991–1999 and reported in the above panel.

Percentage of 349 bankruptcies between 1991–1999 predicted by the model in each decile

Decile	Private firm model	Public firm model	Public firm model – efficient markets version
1	65.33	81.38	75.93
2	14.61	10.6	13.47
3	6.88	4.01	4.87
4	3.15	2.01	1.43
5	4.3	0.86	2.29
6	1.72		0.29
7	2.29	0.57	0.86
8	1.15	0.57	0.29
9	0.57		0.57

Panel D: Forecasting accuracy over 1991–1999 using model estimated with 1962–1990 data – area under the ROC curve

The model estimated with data from 1962–1990 with monthly observations is used to forecast bankruptcy probabilities for 1991–1999. Each month the probabilities are calculated for each firm and then an ROC curve is constructed from the predicted probabilities and the status of the firm in that month. The area under the ROC curve is computed as a measure of the forecasting accuracy of the model. The above procedure is repeated on 1000 bootstrapped samples for each model and the descriptive statistics are computed for the area under the ROC curve.

	Mean	Median	Min	Max	Std. dev
Private firm model with industry effects	0.8812	0.8815	0.8592	0.9002	0.0081
Public firm model with industry effects	0.9449	0.9453	0.9303	0.9567	0.0055
Public firm model: Efficient markets version	0.9151	0.9161	0.8980	0.9304	0.0064

All the remaining industry variables TL/TA*IND are insignificantly different from zero except TLTA*IND2. The sum of TL/TA and TL/TA*IND for industries 2–4 give the appropriate magnitude for the TL/TA effect. They are all positive: +5.1558 (IND2), +4.6503 (IND3), and +4.4188 (IND4).

Based on forecasting ability, the private firm model with monthly observation intervals is significantly more accurate than the yearly observation interval model.

Indeed, the first decile in Table VI contains 65.33 percent of the bankrupt firms versus 36.76 percent for the yearly observation interval in Table V. The accuracy of the first two deciles for the monthly observation interval is 79.94 percent versus 53.23 percent for the yearly observation interval. This improved performance is due to the accuracy and timeliness of the information contained in the quarterly accounting reports.

Second, two public firm models are estimated. The first public firm model includes market and accounting variables. Its estimation is contained in Table VI. The second public firm model – the efficient markets version – just uses the market variables excess return (EXRET), relative size (RSIZ) and volatility (SIGMA). Its estimation is also contained in Table VI.

First consider the public firm model in Table VI. All variables (both market and accounting) are statistically significant, except for one industry intercept (IND3) and four industry slope dummy variables (NITA*IND2, TLTA*IND2, TLTA*IND3 and TLTA*IND4). The market variables – excess return EXRET, relative size RSIZ and volatility SIGMA – all have the correct signs. Increasing excess return (EXRET), increasing size (RSIZ), and decreasing volatility (SIGMA) indicates a smaller chance of bankruptcy. For the accounting variables, a chi-square test shows that all the industry variables (intercept and slope) are significant as a group.

Individually, for industry 1, NI/TA has the proper sign and it is significant. As NI/TA increases, the likelihood of bankruptcy decreases. To get the NI/TA effect for industries 2–4, we need to consider the sum of NI/TA and NI/TA*IND. For industry 2–4 the signs of these sums are in the proper direction. Indeed, the sums are -1.5822 (IND2), -4.5157 (IND3), and -5.4113 (IND4). Next consider the TL/TA variable. For industry 1, TL/TA is positive (the correct sign) and significant. To get the TL/TA effect for industries 2–4, we need to consider the sum of TL/TA and TL/TA*IND. For industry 2–4 the signs of these sums are in the proper direction: $+4.3885$ (IND2), $+3.6915$ (IND3) and $+4.2513$ (IND4).

Based on forecasting ability, the public firm model in Table VI significantly outperforms the private firm model. For the first decile, the public firm model correctly classifies 81.38 percent of the bankrupt firms versus 65.33 percent for the private firm model. With respect to the monthly observation interval, the public firm model also significantly outperforms the yearly observation interval model in Table V. For example, in the yearly observation interval, only 59.12 percent of the firms in the first decile are correctly classified. This is not a surprising result and it is consistent with the usefulness of the timely information contained in stock price variables.

We next study the notion of market efficiency with respect to publicly available accounting information. If the market variables included in this estimation incorporate all the bankruptcy information contained in the quarterly accounting reports, then one would not expect the accounting variables to be statistically significant in Table VI. As mentioned earlier, all the accounting variables except IND3, NITA*IND2, TLTA*IND2, TLTA*IND3 and TLTA*IND4 are statistically significant. However, this is an in-sample test of the efficient markets hypothesis. As common with in-sample tests, this statistical significance could be due to over-fitting the noise in the data.

To investigate this possibility, in Table VI we re-estimated the public firm model without the accounting variables. As seen therein all market variables are significant with signs and magnitudes similar to those for the public firm model contained in Table VI. An out-of-sample forecasting comparison provides a better test of the efficient markets hypothesis. In this comparison, the forecasting performance of the simpler model is nearly identical to the more complex model. For example, the first decile contains 75.93 percent for the simpler model (Table VI, Panel C) versus 81.38 percent for the more complex model. Continuing, the sum of the number of bankruptcies in the first two deciles is almost identical across the two models at 89.40 percent versus 91.98 percent. The remaining deciles are nearly identical as well. The results from the ROC curve analysis reconfirm the decile analysis. The public firm model with industry effects is the best performing model (area under ROC curve = 0.9449), followed by public firm model – efficient markets version (area under the ROC curve = 0.9151), and the private firm model with industry effects (area under the ROC curve = 0.8812). Comparing these results with those in Panel D of Table V shows that monthly forecasting increases the accuracy of all models in a statistically significant way. This out-of-sample comparison is strong evidence consistent with the efficient markets hypothesis. If one selects the best model based both on the accuracy of its prediction and parsimony of the number of independent variables, then the efficient markets version of the public firm model is the preferred bankruptcy model.

8. Conclusion

This paper investigates the forecasting accuracy of bankruptcy hazard rate models for U.S. companies over the time period 1962–1999 using yearly and monthly observation intervals. The contribution of this paper is multiplefold. One, we validate the superior forecasting performance of the bankruptcy hazard rate model of Shumway (2001) as opposed to the models of Altman (1968) and Zmijewski (1984). Two, we demonstrate the importance of including industry effects in hazard rate models. Industry groupings are shown to significantly affect both the intercept and slope coefficients in the forecasting equations. Three, we extend the standard hazard rate model to apply to financial firms. Analogous forecasting variables to those used with non-financials, both balance sheet ratios and market variables, can still be used to forecast bankruptcy for financial firms, although the forecasting accuracy is slightly reduced. Four, we investigate the performance of this bankruptcy hazard rate model using monthly data. The existing academic literature usually employs only yearly observations. We show that bankruptcy prediction is improved using this shorter observation interval. Five, we demonstrate that accounting variables add little predictive power when market variables are already included in the bankruptcy model. This supports the notion of market efficiency with respect to publicly available accounting information.

References

Acharya, V., Bharath, S., and Srinivasan, A. (2003) Understanding the recovery rates of defaulted securities, working paper, London Business School.

Allison, P. (1982) Discrete-time methods for the analysis of event histories, *Sociological Methodology* **1**, 61–99.

Allison, P. (1999) *Survival Analysis Using the SAS System: A Practical Guide*, SAS Institute.

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

Berkovitch, Elazar and Israel, R. (1998) The bankruptcy decision and debt contract renegotiations, *European Finance Review* **2**, 1–27.

Brown, C. (1975) On the use of indicator variables for studying the timed dependence of parameters in a response-time model, *Biometrics* **31**, 863–872.

Cox, D. (1970) *The Analysis of Binary Data*, Methuen.

Duffie, D., Jarrow, R., Purnanandam, A., and Yang, W. (2003) Market pricing of deposit insurance, *Journal of Financial Services Research* **24**, 93–119.

Duffie, D. and Singleton, K. (1999) Modeling term structures of defaultable bonds, *Review of Financial Studies* **12**(4), 197–226.

Duffee, G. (1999) Estimating the price of default risk, *The Review of Financial Studies* **12**(1), 197–226.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman and Hall/CRC Press.

Guenther, D and Rosman, A. (1994) Differences between compustat and CRSP SIC codes and related effects on research, *Journal of Accounting and Economics* **18**, 115–128.

Janosi, T., Jarrow, R., and Yildirim, Y. (2002) Estimating expected losses and liquidity discounts implicit in debt prices, *Journal of Risk* **5**(1), 1–38.

Jarrow, R. (2001) Default parameter estimation using market prices, *Financial Analysts Journal* **57**, 75–92.

Jarrow, R. and Turnbull, S. (1995) Pricing derivatives on financial securities subject to credit risk, *Journal of Finance* **50**(1), 53–85.

Jarrow, R., Lando, D., and Yu, F. (2000) Default risk and diversification: Theory and applications, forthcoming in *Mathematical Finance*.

Kahle, K. and Walkling, R. (1996) The impact of industry classifications on financial research, working paper, Ohio State University.

Kavvathas, D. (2000) Estimating credit rating transition probabilities for corporate bonds, working paper, University of Chicago

Keifer, N. (1988) Economic duration data and hazard functions, *Journal of Economic Literature* **26**, 646–679.

Klein, J. and Moeschberger, M. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*, Springer-Verlag.

Lang, L. and Stultz, R. (1992) Contagion and competitive intra-industry effects of bankruptcy announcements: An empirical analysis, *Journal of Financial Economics* **32**, 45–60.

Madan, D. and Unal, H. (1998) Pricing the risks of default, *Review of Derivatives Research* **2**, 121–160.

Maksimovic, V. and Phillips, G. (1997) Asset efficiency and reallocation decisions of bankrupt firms, working paper, University of Maryland.

Moody's Investors Service (2000a), *Moody's Public Firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk*, March, New York.

Moody's Investors Service (2000b) *RiskCalc Private Model: Moody's Default Model for Private Firms*, May, New York.

Opler, Tim and Titman, Sheridan (1994) Financial distress and corporate performance, *Journal of Finance* **49**, 1015–1040.

Risk Magazine (2000) *Credit Risk: A Risk Special Report*, March.

Shleifer, Andrei and Vishny, Robert (1992) Liquidation values and debt capacity: A market equilibrium approach, *Journal of Finance* **47**, 1343–1366.

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

Warga, A. (1999) *Fixed Income Data Base*, University of Houston, College of Business Administration (www.uh.edu/∼awarga/lb.html).

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