In Search of Distress Risk

John Y. Campbell, Jens Hilscher, and Jan Szilagyi[1]

First draft: October 2004
This version: December 2, 2004

[1]Corresponding author: John Y. Campbell, Department of Economics, Littauer Center 213, Harvard University, Cambridge MA 02138, USA, and NBER. Tel 617-496-6448, email john_campbell@harvard.edu. This material is based upon work supported by the National Science Foundation under Grant No. 0214061 to Campbell. We would like to thank Robert Jarrow and Don van Deventer of Kamakura Risk Information Services (KRIS) for providing us with bankruptcy data and Stuart Gilson, John Griffin, Scott Richardson, and seminar participants at Humboldt Universität zu Berlin, HEC Paris, and the University of Texas for helpful discussion.

Abstract

This paper explores the determinants of corporate bankruptcy and the pricing of financially distressed stocks using US data over the period 1963 to 1998. Firms with higher leverage, lower profitability, lower market capitalization, lower past stock returns, more volatile past stock returns, and lower cash holdings are more likely to go into bankruptcy. When predicting bankruptcy at longer horizons, market capitalization, the most persistent predictor, becomes relatively more significant. Our model captures much of the time variation in the US bankruptcy rate. Distressed stocks have delivered anomalously low returns during this period. They have lower returns but much higher standard deviations, market betas, and loadings on value and small-cap risk factors than stocks with low bankruptcy risk. These findings are inconsistent with the conjecture that the value and size effects are compensation for the risk of financial distress.

1 Introduction

The concept of financial distress is often invoked in the asset pricing literature to explain otherwise anomalous patterns in the cross-section of stock returns. The idea is that certain companies have an elevated risk that they will fail to meet their financial obligations, and investors charge a premium for bearing this risk.[2]

While this idea has a certain plausibility, it leaves a number of basic questions unanswered. First, how do we measure the failure to meet financial obligations? Second, how do we measure the probability that a firm will fail to meet its financial obligations? Third, even if we have answered these questions and thereby constructed an empirical measure of financial distress, is it the case that the stock prices of financially distressed companies move together in response to a common risk factor? Finally, what returns have financially distressed stocks provided historically? Is there any evidence that financial distress risk carries a premium?

In this paper we adopt a relatively atheoretical econometric approach to measure financial distress. We say that a firm fails to meet financial obligations if it enters bankruptcy under either Chapter 7 or Chapter 11. That is, we ignore the possibility that a firm may avoid bankruptcy by negotiating a debt restructuring out of court (Gilson, John, and Lang 1990, Gilson 1997). We measure the probability of bankruptcy by estimating a hazard model using a logit specification, following Shumway (2001), Chava and Jarrow (2002), and others.

We extend the previous literature by considering a wide range of explanatory variables, including both accounting and equity-market variables, and by explicitly considering how the optimal specification varies with the horizon of the forecast. Some papers on bankruptcy concentrate on predicting the event that a bankruptcy will occur during the next month. Over such a short horizon, it should not be surprising that the recent return on a firm's equity is a powerful predictor, but this may not be very useful information if it is relevant only in the extremely short run, just as it would not be useful to predict a heart attack by observing a person dropping

[2]Chan and Chen (1991), for example, attribute the size premium to the prevalence of "marginal firms" in small-stock portfolios, and describe marginal firms as follows: "They have lost market value because of poor performance, they are inefficient producers, and they are likely to have high financial leverage and cash flow problems. They are marginal in the sense that their prices tend to be more sensitive to changes in the economy, and they are less likely to survive adverse economic conditions." Fama and French (1996) use the term "relative distress" in a similar fashion.

to the floor clutching his chest. We also explore time-series variation in the number of bankruptcies, and ask how much of this variation is explained by changes over time in the variables that predict bankruptcy at the firm level.

Our empirical work begins with a bankruptcy indicator from Kamakura Risk Information Services (KRIS), used by Chava and Jarrow (2002), which includes all bankruptcy filings in the Wall Street Journal Index, the SDC database, SEC filings and the CCH Capital Changes Reporter. The data cover the months from January 1963 until the end of 1998. We merge this dataset with firm level accounting data from COMPUSTAT as well as monthly and daily equity price data from CRSP. This gives us about 800 bankruptcies, and predictor variables for 1.3 million firm months.

We start by estimating a basic specification used by Shumway (2001) and similar to that of Chava and Jarrow (2002). The model includes both equity market and accounting data. From the equity market, we measure the excess stock return of each company over the past month, the volatility of daily stock returns over the past three months, and the market capitalization of each company. From accounting data, we measure net income as a ratio to assets, and total leverage as a ratio to assets.

From this starting point, we make a number of contributions to the prediction of corporate bankruptcy. First, we explore some sensible modifications to the variables listed above. Specifically, we show that scaling net income and leverage by the market value of assets rather than the book value, adding further lags of stock returns and net income, and including dummies for missing data can improve the explanatory power of the benchmark regression.

Second, we explore some additional variables and find that corporate cash holdings and the market-book ratio also offer a marginal improvement. In a related exercise we construct a measure of distance to default, based on the practitioner model of KMV (Crosbie and Bohn 2001) and ultimately on the structural bankruptcy model of Merton (1974). We find that this measure adds relatively little explanatory power to the reduced-form variables already included in our model.

Third, we examine what happens to our specification as we increase the horizon at which we are trying to predict bankruptcy. Consistent with our expectations, we find that our most persistent forecasting variable, market capitalization, becomes relatively more important as we predict bankruptcy further into the future.

Fourth, we study time-variation in the number of bankruptcies. We compare the realized frequency of bankruptcy to the predicted frequency over time. Although the model underpredicts the frequency of bankruptcy in the 1980s and overpredicts it in the 1990s, the model fits the general time pattern quite well. We show that macroeconomic variables, in particular the default yield spread and the term spread, can be used to capture some of the residual time variation in the bankruptcy rate.

Finally, we explore the risks and average returns on portfolios of firms sorted by our fitted probability of bankruptcy. We find that firms with a high probability of bankruptcy have high market betas and high loadings on the HML and SMB factors proposed by Fama and French (1993, 1996) to capture the value and size effects. However they do not have high average returns, suggesting that the equity market has not properly priced distress risk.

There is a large related literature that studies the prediction of corporate bankruptcy. The literature varies in choice of variables to predict bankruptcy and the methodology used to estimate the likelihood of bankruptcy. Altman (1968), Ohlson (1980), and Zmijewski (1984) use accounting variables to estimate the probability of bankruptcy in a static model. Altman's Z-score and Ohlson's O-score have become popular and widely accepted measures of financial distress. They are used, for example, by Dichev (1998), Griffin and Lemmon (2002), and Ferguson and Shockley (2003) to explore the risks and average returns for distressed firms.

Shumway (2001) estimates a hazard model at annual frequency and adds equity market variables to the set of scaled accounting measures used in the earlier literature. He points out that estimating the probability of bankruptcy in a static setting introduces biases and overestimates the impact of the predictor variables. This is because the static model does not take into account that a firm could have had high levels of unfavorable indicators several periods before going into bankruptcy. Hillegeist, Cram, Keating and Lunstedt (2004) summarize equity market information by calculating the probability of bankruptcy implied by the structural Merton model. Adding this to accounting data increases the accuracy of bankruptcy prediction within the framework of a hazard model. Chava and Jarrow (2002) estimate hazard models at both annual and monthly frequencies and find that the accuracy of bankruptcy prediction is greater at a monthly frequency. They also compare the effects of accounting information across industries.

Duffie and Wang (2003) emphasize that the probability of bankruptcy depends on the horizon one is considering. They estimate mean-reverting time series processes for a macroeconomic state variable—personal income growth—and a firm-specific variable—distance to default. They combine these with a short-horizon bankruptcy model to find the marginal probabilities of default at different horizons. Using data from the US industrial machinery and instruments sector, they calculate term structures of default probabilities. We conduct a similar exercise using a reduced-form econometric approach; we do not model the time-series evolution of the predictor variables but instead directly estimate longer-term default probabilities.

The remainder of the paper is organized as follows. Section 2 describes the construction of the data set, outlier analysis and summary statistics. We compare the distributions of the predictor variables for those observations where the firm went bankrupt and all other observations. This section also considers the pattern of U.S. corporate bankruptcies over time.

Section 3 discusses our basic hazard specification, extensions to it, and the results from estimating the model at one-month and longer horizons. We find that past stock return is less significant and market capitalization is more significant as the horizon increases. This section also considers the stability of the model across industries, firms with high and low leverage, and large and small firms.

Section 4 considers the model's fit to the time-series pattern of bankruptices. When including year dummies to proxy for unobserved factors affecting bankruptcy we reject the null hypothesis that our model completely explains the aggregate history of bankruptcy in the US. Comparing the predicted and realized frequencies of bankruptcy, however, we find that the model has considerable explanatory power for this history.

Section 5 studies the return properties of equity portfolios formed on the fitted value from our bankruptcy prediction model. We ask whether stocks with high bankruptcy probability have unusually high or low returns relative to the predictions of standard cross-sectional asset pricing models such as the CAPM or the three-factor Fama-French model. Section 6 concludes.

2 Data description

In order to estimate a hazard model of bankruptcies we need a bankruptcy indicator and a set of explanatory variables. The bankruptcy indicator we use is taken from Chava and Jarrow (2002); it includes all bankruptcy filings in the Wall Street Journal Index, the SDC database, SEC filings and the CCH Capital Changes Reporter. The indicator is one in a month in which a firm filed for bankruptcy under Chapter 7 or Chapter 11, and zero otherwise; in particular, the indicator is zero if the firm disappears from the dataset for some reason other than bankruptcy such as acquisition or delisting. The data span the months from December 1963 through December 1998.

Table 1 summarizes the properties of the bankruptcy indicator. The first column shows the number of active firms for which we have data in each year; the second column shows the number of bankruptcies; and the third column reports the percentage of active firms that went bankrupt in each year. This series is also illustrated in Figure 1. It is immediately apparent that bankruptcies were extremely rare until the late 1960's. In fact, in the three years 1967–1969 there were no bankruptcies at all in our dataset. The bankruptcy rate increased in the early 1970's, and then rose dramatically during the 1980's to a peak of 1.7% in 1986. It remained high through the economic slowdown of the early 1990's, but fell in the late 1990's to levels only slightly above those that prevailed in the 1970's.

Some of these changes through time are probably the result of changes in the law governing corporate bankruptcy in the 1970's, and related financial innovations such as the development of below-investment-grade public debt (junk bonds) in the 1980's and the advent of prepackaged bankruptcy filings in the early 1990's (Tashjian, Lease, and McConnell 1996). Changes in corporate capital structure (Bernanke and Campbell 1988) and the riskiness of corporate activities (Campbell, Lettau, Malkiel, and Xu 2001) are also likely to have played a role, and one purpose of our investigation is to quantify the time-series effects of these changes.

In order to construct explanatory variables at the individual firm level, we combine quarterly accounting data from COMPUSTAT with monthly and daily equity market data from CRSP. From COMPUSTAT we construct a standard measure of profitability: net income relative to total assets. Previous authors have measured total assets at book value, but we find better explanatory power when we measure the equity component of total assets at market value by adding the book value of liabilities to the market value of equities. We call this series NIMTA (Net Income to Market-valued Total Assets) and the traditional series NITA (Net Income to Total Assets). We also use COMPUSTAT to construct a measure of leverage: total liabilities relative to total assets. We again find that a market-valued version of this series, defined as total liabilities divided by the sum of market equity and book liabilities, performs better than the traditional book-valued series. We call the two series TLMTA and TLTA, respectively. To these standard measures of profitability and leverage, we add a measure of liquidity, the ratio of a company's cash and short-term assets to the market value of its assets (CASHMTA). We also calculate each firm's market-to-book ratio (MB).

In constructing these series we adjust the book value of assets to eliminate outliers, following the procedure suggested by Cohen, Polk, and Vuolteenaho (2003). That is, we add 10% of the difference between market and book equity to the book value of total assets, thereby increasing book values that are extremely small, probably mismeasured, and create outliers when used as the denominators of financial ratios. We also winsorize all variables at the 5th and 95th percentiles of their cross-sectional distributions. That is, we replace any observation below the 5th percentile with the 5th percentile, and any observation above the 95th percentile with the 95th percentile. We are careful to adjust each company's fiscal year to the calendar year and lag the data by two months. This adjustment ensures that the accounting data are available at the beginning of the month over which bankruptcy is measured. The Appendix to this paper describes the construction of these variables in greater detail.

We add several market-based variables to these two accounting variables. We calculate the monthly log excess return on each firm's equity relative to the S&P 500 index (EXRET), the standard deviation of each firm's daily stock return over the past three months (SIGMA), and the relative size of each firm measured as the log ratio of its market capitalization to that of the S&P 500 index (RSIZE). In addition, we obtain historical yields on 6-month Treasury bills and 10-year Treasury bonds from the Federal Reserve Board, and historical yields on AAA and BAA rated corporate bonds from Amit Goyal's website.[3]

Finally, we group all firms into four broad sectors using single-digit SIC codes. Industry 1 is miscellaneous (codes 1, 3, 6, 7, 9), industry 2 is manufacturing and minerals (codes 2, 4), industry 3 is transportation, communication and utilities (code 5), and industry 4 is finance, insurance and real estate (code 8).

[3]The Federal Reserve data are at http://www.federalreserve.gov/releases/h15/data.htm and the corporate bond data are at http://www.bus.emory.edu/AGoyal/.

2.1 Summary statistics

Table 2 summarizes the properties of our seven main explanatory variables. The first panel in Table 2 describes non-bankruptcy months and the second panel describes a much smaller sample of bankruptcy months. We exclude months where any of the seven variables have missing values, leaving us with a sample of 1,281,426 observations containing 796 bankruptcy events. We also illustrate the data in a series of graphs. Figures 2 through 8 each have two panels, one showing the distribution of an explanatory variable in non-bankruptcy months, the other showing the distribution in the much smaller sample of bankruptcy months.

In interpreting these distributions, it is important to keep in mind that we weight every firm-month equally. This has two important consequences. First, the distributions are dominated by the behavior of relatively small companies; value-weighted distributions look quite different. Second, the distributions reflect the influence of both cross-sectional and time-series variation. The cross-sectional averages of several variables, in particular NIMTA, TLMTA, and SIGMA, have experienced significant trends since 1963: SIGMA and TLMTA have trended up, while NIMTA has trended down. The downward trend in NIMTA is not just a consequence of the buoyant stock market of the 1990's, because book-based net income, NITA, displays a similar trend. The influence of these trends is magnified by the growth in the number of companies over time, which means that recent years have greater influence on the distribution than earlier years.

These facts help to explain several features of Table 2. The mean level of NIMTA, for example, is only 0.002 or 0.2% per quarter in our dataset, 0.8% at an annual rate. This is three times lower than the median level of NIMTA because the distribution of NIMTA is negatively skewed; it is also lower than the average level of NITA because the market value of equity is on average about twice the book value in our data. Nonetheless these measures of profitability are all strikingly low, reflecting the prevalence of small, unprofitable listed companies in recent years.[4] The distribution of NIMTA shown in Figure 2 has a large spike just above zero, a phenomenon noted by Hayn (1995), suggesting that firms may be managing their earnings to avoid reporting losses.[5]

[4]The value-weighted mean of NIMTA is three times the equal-weighted mean, reflecting the greater profitability of large companies.

[5]There is a debate in the accounting literature about the interpretation of this spike. Burgstahler and Dichev (1997) argue that it reflects earnings management, but Dechow, Richardson, and Tuna

The average value of EXRET is -0.012 or -1.2% per month. This extremely low number reflects both the underperformance of small stocks during the later part of our sample period (the value-weighted mean is almost exactly zero), and the fact that we are reporting a geometric average excess return rather than an arithmetic average. The difference is substantial because individual stock returns are extremely volatile. The average value of the annualized firm-level volatility SIGMA is greater than 50%, again reflecting the strong influence of small firms and recent years in which idiosyncratic volatility has been high (Campbell, Lettau, Malkiel, and Xu 2001).

A comparison of the top and bottom panels of Table 2, and inspection of Figures 2–8, reveal that bankrupt firms have intuitive differences from the rest of the sample. In months immediately preceding a bankruptcy filing, firms typically make losses (the mean loss is 3.5% quarterly or 14% of market value of assets at an annual rate, and the median loss is 4.7% quarterly or almost 19% at an annual rate); the value of their debts is extremely high relative to their assets (average leverage exceeds 75%, and median leverage exceeds 85%); they have experienced extremely negative returns over the past month (the mean is -132% at an annual rate or -11% over a month, while the median is -208% at an annual rate or -17% over a month); and their volatility is extraordinarily high (the mean is 99% and the median is 121%). Bankrupt firms also tend to be relatively small (almost 8 times smaller than other firms on average, and almost 12 times smaller at the median), and they have only about half as much cash and short-term investments, in relation to the market value of assets, as non-bankrupt firms. Finally, the market-book ratio of bankrupt firms has a similar mean but a much higher standard deviation than the market-book ratio of other firms. Figure 8 shows that almost 40% of firms have market-book ratios at the lower winsorization point, while 25% have market-book ratios at the upper winsorization point. It appears that some firms go bankrupt after realized losses have driven down their book values relative to market values, while others go bankrupt after bad news about future prospects has driven down their market values relative to book values.

(2003) point out that discretionary accruals are not associated with the spike in the manner that would be expected if this interpretation is correct.

3 Estimating a hazard model for bankruptcy

The summary statistics in Table 2 show that bankrupt firms have a number of unusual characteristics. However the number of bankruptcy months is tiny compared to the number of other months, so it is not at all clear how useful these variables are in predicting bankruptcy. Also, these characteristics are correlated with one another and we would like to know how to weight them optimally. Following Shumway (2001) and Chava and Jarrow (2002), we now estimate the probability of bankruptcy over the next period using a logit model.

We assume that the marginal probability of bankruptcy over the next period follows a logistic distribution and is given by

$$P_{t-1}\left(Y_{it}=1\right)=\frac{1}{1+\exp\left(-\alpha-\beta x_{i,t-1}\right)} \tag{1}$$

where Y_{it} is a bankruptcy indicator that is equal to one if the firm goes bankrupt in month t, and $x_{i,t-1}$ is a vector of explanatory variables known at the end of the previous month. A higher level of $\alpha+\beta x_{i,t-1}$ implies a higher probability of bankruptcy.

Table 3 reports logit regression results for various alternative specifications. In column 1 we follow Shumway (2001) and Chava and Jarrow (2002), and estimate a model with five standard variables: NITA, TLTA, EXRET, SIGMA, and RSIZE. This model measures assets in the conventional way, using annual book values from COMPUSTAT. It excludes firm age, a variable which Shumway (2001) considered but found to be insignificant in predicting bankruptcy.

All five of the included variables in column 1 enter significantly and with the expected sign. To get some idea of the relative impact of changes in the different variables we compute the effect on the score of a one-standard-deviation change in each predictor variable. This effect is 0.29 for NITA, 0.96 for TLTA, 0.39 for EXRET, 0.36 for RSIZE, and 0.91 for SIGMA. Thus movements in leverage and volatility are more important for explaining bankruptcy than movements in profitability, past stock returns, and firm size.

In column 2 we replace the traditional accounting ratios NITA and TLTA that use the book value of assets, with our ratios NIMTA and TLMTA that use the market value of assets. These measures are more sensitive to new information about firm prospects since equity values are measured using monthly market data rather than quarterly accounting data. Column 2 appears to be an improved specification in several respects. The coefficient on profitability more than doubles and increases its statistical significance, and the pseudo R^2, a measure of fit for the logit model, increases by 0.005. Interestingly, the coefficient on RSIZE becomes small and statistically insignificant, suggesting that its role in column 1 was an artifact of our use of accounting data.

Column 2 has 5% more observations than column 1, because we no longer need to measure book values of assets in the COMPUSTAT dataset. Importantly, the number of bankruptcies increases by 3% from 737 to 762. Since bankruptcies are rare events, it is important to include as many of them as we possibly can. Accordingly in column 3 we expand the dataset further by relaxing the requirement that we be able to measure equity market volatility, SIGMA, over the past three months. In cases where this variable is missing, we create a dummy variable SIGMAMISS and enter it in the column 3 regression. This increases the number of observations by another 1% and the number of bankruptcies by another 4% to 796. As one would expect from these statistics, the dummy variable enters significantly as a bankruptcy predictor. Firms with intermittent equity trading over the past three months are more likely to fail our test for valid construction of equity volatility, and are also more likely to go bankrupt. The magnitude of the SIGMAMISS coefficient implies that missing volatility is equivalent to volatility about 1.5 standard deviations above the cross-sectional mean.

So far we have used only the information in the most recent values of profitability, leverage, excess returns, and volatility. An obvious question is whether lagged values have some additional explanatory power. One might expect that a long history of losses or a sustained decline in stock market value would be a better predictor of bankruptcy than one large quarterly loss or a sudden stock price decline in a single month. Exploratory regressions with lagged values confirm that lags of NIMTA and EXRET enter significantly, while lags of the other variables do not. As a reasonable summary, we impose geometrically declining weights on these lags. We construct

$$NIMTAAVG_{t-1,t-12} = \frac{1-\phi^3}{1-\phi^{12}}\left(NIMTA_{t-1,t-3} + ... + \phi^9 NIMTA_{t-9,t-12}\right), \quad (2)$$

$$EXRETAVG_{t-1,t-12} = \frac{1-\phi}{1-\phi^{12}}(EXRET_{t-1} + ... + \phi^{11} EXRET_{t-12}), \quad (3)$$

where the coefficient $\phi = 2^{-\frac{1}{3}}$, implying that the weight is halved each quarter. The data suggest that this parsimonious specification captures almost all the predictability obtainable from lagged profitability and stock returns. Column 4 of Table 3 reports the regression results using this specification, finding an improvement in the explanatory power of stock returns. This specification causes a further decline in the importance of market capitalization, whose coefficient actually changes sign.

In column 5 we add two other variables that might be relevant for bankruptcy prediction. The ratio of cash and short-term investments to the market value of total assets, CASHMTA, captures the liquidity of the firm. A firm with a high CASHMTA ratio has liquid assets available to make interest payments, and thus may be able to postpone bankruptcy with the possibility of avoiding it altogether if circumstances improve. The market to book ratio, MB, captures the relative value placed on the firm's equity by stockholders and by accountants. Our profitability and leverage ratios use market value; if book value is also relevant, then MB may enter the regression as a correction factor, increasing the probability of bankruptcy when market value is unusually high relative to book value.[6] Column 5 supports both these hypotheses.

Finally, in column 6 we compare our reduced-form model with the structural approach of KMV (Crosbie and Bohn 2001), based on the structural bankruptcy model of Merton (1974). We construct "distance to default", DD, applying a standard formula given in the Appendix. This formula requires an estimate of the volatility of the firm's asset value, and we make the assumption that asset volatility equals equity volatility. For firms with relatively safe debt, this assumption will understate the distance to default. Although our DD measure does predict default with the theoretically expected sign in a univariate logit regression, its explanatory power is much lower than our reduced-form approach (the pseudo R^2 is only about 8%). Column 6 shows that DD adds relatively little explanatory power to the reduced-form variables already included in our model, and in fact enters our multivariate regression with the wrong sign.

[6] Chacko, Hecht, and Hilscher (2004) discuss the measurement of credit risk when the market-to-book ratio is influenced both by cash flow expectations and discount rates.

3.1 Forecasting bankruptcy at long horizons

At the one month horizon our best specification captures about 30% of the variation in bankruptcy risk. We now ask what happens as we try to predict bankruptcies further into the future. In Table 4 we estimate the conditional probability of bankruptcy in six months, one, two and three years. We again assume a logit specification but allow the coefficients on the variables to vary over time. In particular we assume that the probability of bankruptcy in j months, conditional on survival in the dataset for $j-1$ months, is given by

$$P_{t-1}\left(Y_{i,t-1+j}=1 \mid Y_{i,t-2+j}=0\right)=\frac{1}{1+\exp\left(-\alpha_j-\beta_j x_{i,t-1}\right)}. \tag{4}$$

Note that this assumption does not imply a cumulative probability of bankruptcy that is logit. If the probability of bankruptcy in j months did not change with the horizon j, that is if $\alpha_j=\alpha$ and $\beta_j=\beta$, and if firms exited the dataset only through bankruptcy, then the cumulative probability of bankruptcy over the next j periods would be given by $1-\left(\exp\left(-\alpha-\beta x_i\right)/(1+\exp\left(-\alpha-\beta x_i\right)\right)^j$, which no longer has the logit form. Variation in the parameters with the horizon j, and exit from the dataset through mergers and acquisitions, only make this problem worse. In principle we could compute the cumulative probability of bankruptcy by estimating models for each horizon j and integrating appropriately; or by using our one-period model and making auxiliary assumptions about the time-series evolution of the predictor variables in the manner of Duffie and Wang (2003). We do not pursue these possibilities here, concentrating instead on the conditional probabilities of default at particular dates in the future.

As the horizon increases in Table 4, the coefficients, significance levels, and overall fit of the logit regression decline as one would expect. Even at three years, however, almost all the variables remain statistically significant. The distance to default DD enters the regression with the theoretically expected negative sign at horizons of two and three years, but it is statistically insignificant; this result is particularly disappointing since DD is designed to measure bankruptcy risk at a medium-term horizon of one year.

Two predictor variables are particularly important at long horizons. The coefficient on the market-to-book ratio MB is remarkably stable, and the coefficient on relative size RSIZE becomes increasingly significant with the expected negative sign as the horizon increases. These results imply that the most persistent attributes of a firm, its market capitalization and market-to-book ratio, become increasingly important measures of financial distress at long horizons.

In the top panel of Table 4 the number of observations and number of bankruptcies vary with the horizon, because increasing the horizon forces us to drop observations at both the beginning and end of the dataset. Bankruptcies that occur within the first j months of the sample cannot be related to the condition of the firm j months previously, and the last j months of the sample cannot be used to predict bankruptcies that may occur after the end of the sample. Also, many firms exit the dataset for other reasons between dates $t-1$ and $t-1+j$. In the bottom panel of Table 4 we study a subset of firms for which data are available at all the different horizons. This allows us to compare R^2 statistics directly across horizons. We obtain very similar results to those in the top panel, suggesting that variation in the available data is not responsible for our findings.

3.2 Robustness checks

We have explored industry effects on the bankruptcy models estimated in Tables 3 and 4. The Shumway (2001) and Chava-Jarrow (2002) specification in column 1 of Table 3 appears to behave somewhat differently in the finance, insurance, and real estate (FIRE) sector. That sector has a lower intercept and a more negative coefficient on profitability. However there is no clear evidence of sector effects in the market-based specifications estimated in the other columns of Table 3.

We have also used market capitalization and leverage as interaction variables, to test the hypotheses that other explanatory variables enter differently for small or highly indebted firms than for other firms. We have found no clear evidence that such interactions are important.

4 Matching the time series of bankruptcies

As we noted earlier, there is dramatic variation in the bankruptcy rate over time. Figure 1 shows the bankruptcy rate rising from an average of 0.3% in the 1960's and 1970's to almost 1.7% in 1986. In this section, we ask how well our model fits this pattern. We first calculate the fitted probability of bankruptcy for each company in our dataset using the coefficients from the best fitting regression in Table 3. We then average over all the predicted probabilities for active companies in each month to obtain a prediction of the aggregate bankruptcy rate.

Figure 9 shows annual averages of predicted and realized bankruptcies. Our model captures much of the broad variation in bankruptcy over time, including the strong and long-lasting increase in the 1980's and cyclical spikes in the mid-1970's and early 1990's. However it somewhat overpredicts these spikes and rises in the 1980's much more slowly than actual bankruptcies. In the worst performing year of 1986, the model underestimates the bankruptcy rate by about one half. Also, the model tends to predict more bankruptcies than actually occurred throughout the 1990's.

As an alternative way to understand the time-series variation that is not captured by our basic model, we augment that model by including year dummies that shift the baseline probability of bankruptcy from one year to the next. We restrict our attention to the period since 1975 because of the rarity of bankruptcies in earlier years.

Figures 10 and 11 plot the demeaned year dummies estimated from a model with only these dummies (Figure 10) and a model that also includes our standard set of explanatory variables (Figure 11). We overwhelmingly reject the hypothesis that the year dummies are all equal—that is, that time effects can be omitted from the model—although the Wald statistic for this test does fall from 172 in Figure 10 to 129 in Figure 11, implying that our model captures some of the time-variation in bankruptcies.

To show which years have anomalous variation in bankruptcies, Figures 10 and 11 also plot a two-standard-deviation confidence interval for each demeaned year dummy, constructed under the null hypothesis that all year dummies are equal. The confidence interval is significantly narrower in later years because the number of firm-level observations increases over time. The model with only year dummies has a standard deviation of 0.63 for the dummies and displays unusually high bankruptcy risk throughout the 1980's and early 1990's, whereas the model that uses firm-level market and accounting data has a smaller standard deviation of 0.45 and relies on variation in the year dummies primarily in the early 1980's and the late 1990's. It is possible that positive year dummies in the early 1980's reflect the creation of the junk bond market in that period, and that negative year dummies in the late 1990's reflect the increased tendency of firms to restructure their debt without entering bankruptcy.

4.1 Macroeconomic effects

One interpretation of time effects is that they result from changes in the state of the macroeconomy. Duffie and Wang (2003), for example, give an important role to the growth rate of industrial production in their multi-period bankruptcy model. In order to explore macroeconomic effects, we first added NBER recession dummies to our best fitting model for the period since 1976, both in levels and interacted with the other coefficients to allow a regime change in bankruptcy risk during a recession. We used both contemporaneous and six-month lagged dummies to capture the possibility that bankruptcies respond to macroeconomic conditions with a lag. However none of these variables entered the model significantly.

We obtain more promising results when we allow for an effect of interest rates on bankruptcy risk. We measure the default yield spread between BAA and AAA rated debt, DFY, and the term spread between 10-year Treasury bonds and six-month Treasury bills, TMS. If we add these variables directly to the regression, without interactions, they enter significantly. If we allow them to interact with the other variables in the model, the interactions of DFY with relative size and leverage are significant and drive out the direct effect of DFY, while the interaction of TMS with leverage is significant and again drives out the direct effect of TMS.

Table 5 reports the results. Column 1 shows that an increase in the default yield spread reduces bankruptcy risk for highly leveraged firms and increases it for small firms. This is consistent with the view that a high default yield spread is a signal of a credit crunch, which tends to dry up credit to small firms and increase the average quality of firms that do receive credit. Column 2 shows that an increase in the slope of the yield curve reduces bankruptcy risk for highly leveraged firms. Since the yield spread usually rises when short-term interest rates fall, this may reflect relief to highly indebted firms with floating-rate debt when short rates decline.

The default yield spread is particularly successful at reducing the importance of time dummies in our model. When we include only DFY interacted with relative size and leverage, and not the level of DFY, the Wald test statistic for the significance of the time dummies falls to 66 even though the regression includes no pure time-series variables.

5 Risks and average returns on distressed stocks

We now turn our attention to the asset pricing implications of our bankruptcy model. Recent work on the distress premium has tended to use either traditional risk indices such as the Altman Z-score or Ohlson O-score (Dichev 1998, Griffin and Lemmon 2002, Ferguson and Shockley 2003) or the distance to default measure of KMV (Vassalou and Xing 2004). To the extent that our reduced-form model more accurately measures the risk of bankruptcy at short and long horizons, we can more accurately measure the premium that investors receive for holding distressed stocks.

Before presenting the results, we ask what results we should expect to find. On the one hand, if investors accurately perceive the risk of bankruptcy they may demand a premium for bearing it. The frequency of bankruptcy shows strong variation over time, as illustrated in Figure 1; even if much of this time-variation is explained by time-variation in our firm-level predictive variables, it still generates common movement in stock returns that might command a premium.

Of course, a risk can be pervasive and still be unpriced. If the standard implementation of the CAPM is exactly correct, for example, then each firm's risk is fully captured by its covariation with the market portfolio of equities, and bankruptcy risk is unpriced to the extent that it is uncorrelated with that portfolio. However it seems plausible that bankruptcies may be correlated with declines in unmeasured components of wealth such as human capital (Fama and French 1996) or debt securities (Ferguson and Shockley 2003), in which case bankruptcy risk will carry a positive risk premium.[7] This expectation is consistent with the high bankruptcy risk of small

[7] Fama and French (1996) state the idea particularly clearly: "Why is relative distress a state variable of special hedging concern to investors? One possible explanation is linked to human capital, an important asset for most investors. Consider an investor with specialized human capital tied to a growth firm (or industry or technology). A negative shock to the firm's prospects probably does not reduce the value of the investor's human capital; it may just mean that employment in the

firms that have depressed market values, since small value stocks are well known to deliver high average returns.

An alternative possibility is that investors fail to understand the relation between our predictive variables and bankruptcy risk, and so do not discount the prices of high-risk stocks enough to offset their bankruptcy probability. This investor failure could be consistent with rational learning through the sample period after the increase in bankruptcies during the 1970's, or it could be a deeper failure of the sort postulated by behavioral finance. In either case we will find that bankruptcy risk appears to command a negative risk premium during our sample period. This expectation is consistent with the high bankruptcy risk of volatile stocks, since Ang, Hodrick, Xing, and Zhang (2004) have recently found negative average returns for stocks with high idiosyncratic volatility.

We measure the premium for financial distress by sorting stocks according to their bankruptcy probabilities, estimated using the 12-month-ahead model of Table 4. Each January from 1976 through 1998, we form ten equally weighted portfolios of stocks that fall in different regions of the bankruptcy risk distribution. We hold these portfolios for a year, allowing the weights to drift with returns within the year rather than rebalancing monthly, in order to minimize turnover costs and the effects of bid-ask bounce.[8] Our portfolios contain stocks in percentiles 0–5, 5–10, 10–20, 20–40, 40–60, 60–80, 80–90, 90–95, 95–99, and 99–100 of the bankruptcy risk distribution. This portfolio construction procedure pays greater attention to the tails of the distribution, where the distress premium is likely to be more relevant, and particularly to the most distressed firms.

Because we are studying the returns to distressed stocks, it is important to handle carefully the returns to stocks that are delisted and thus disappear from the CRSP database. In many cases CRSP reports a delisting return for the final month of the firm's life; we have 8,243 such delisting returns in our sample and we use them

firm will grow less rapidly. In contrast, a negative shock to a distressed firm more likely implies a negative shock to the value of human capital since employment in the firm is more likely to contract. Thus, workers with specialized human capital in distressed firms have an incentive to avoid holding their firms' stocks. If variation in distress is correlated across firms, workers in distressed firms have an incentive to avoid the stocks of all distressed firms. The result can be a state-variable risk premium in the expected returns of distressed stocks." (p.77).

[8] In the first version of this paper we calculated returns on portfolios rebalanced monthly, and obtained similar results to those reported here.

where they are available. Otherwise, we use the last available full-month return in CRSP. In some cases this effectively assumes that our portfolios sell distressed stocks at the end of the month before delisting, which imparts an upward bias to the returns on distressed-stock portfolios (Shumway 1997, Shumway and Warther 1999).[9] We assume that the proceeds from sales of delisted stocks are reinvested in each portfolio in proportion to the weights of the remaining stocks in the portfolio. In a few cases, stocks are delisted and then re-enter the database, but we do not include these stocks in the sample after the first delisting. We treat firms that enter bankruptcy as equivalent to delisted firms, even if CRSP continues to report returns for these firms. That is, our portfolios sell stocks of companies that enter bankruptcy and we use the latest available CRSP data to calculate a final return on such stocks.

Table 6 reports the results. Each portfolio corresponds to one column of the table. The top panel reports average monthly returns in excess of the market, with t statistics below in parentheses, and then alphas with respect to the CAPM, the three-factor model of Fama and French (1993), and a four-factor model proposed by Carhart (1997) that also includes a momentum factor. The bottom panel reports estimated factor loadings for excess returns on the three Fama-French factors, again with t statistics, and the standard deviation of each portfolio's excess return. Figures 12 and 13 graphically summarize the behavior of factor loadings and alphas.

The average returns in the first row of Table 6 are monotonically declining in bankruptcy risk, with a spread of 1.0% per month between the lowest-risk and highest-risk portfolios. The excess returns for low-risk portfolios are significantly positive, but those for high-risk portfolios are statistically insignificant because of their high volatility. While the low-risk portfolios have monthly standard deviations between 2.0 and 2.5%, the highest-risk portfolio has a standard deviation of 11.2% and the next portfolio has a standard deviation of 7.0%.

There is striking variation in factor loadings across the portfolios in Table 6. The low-risk portfolios have negative market betas for their excess returns (that is, betas less than one for their raw returns), and small or negative loadings on the value factor HML. The high-risk portfolios have positive market betas for their excess returns, and loadings greater than one on HML. Because the portfolios are equally weighted, they all have high loadings on the size factor SMB, but the loadings increase dramatically with bankruptcy risk. This reflects the role of market capitalization in predicting

[9] In the first version of this paper we did not use CRSP delisting returns. The portfolio results were similar to those reported here.

bankruptcies at medium and long horizons.

These factor loadings imply that when we correct for risk using either the CAPM or the Fama-French three-factor model, we worsen the anomalous poor performance of distressed stocks rather than correcting it. The spread in CAPM alphas between the lowest-risk and highest-risk portfolios is 1.2% per month, and the spread in Fama-French alphas is 1.8% per month. The poor performance of distressed stocks, and not just the good performance of relatively safe stocks, is statistically significant when we use the Fama-French model.

One of the variables that predicts bankruptcy in our model is recent past return. This suggests that distressed stocks have negative momentum, which might explain their low average returns. To control for this, Table 6 also reports alphas from the Carhart (1997) four-factor model including a momentum factor. The negative alphas for distressed stocks improve slightly, but remain statistically significant.

Overall, these results are discouraging for the view that distress risk is positively priced in the US stock market. We find that stocks with a high risk of bankruptcy have low average returns, despite their high loadings on small-cap and value risk factors.

6 Conclusion

This paper makes two main contributions to the literature on financial distress. First, we carefully implement a reduced-form econometric model to predict bankruptcy at short and long horizons. Our best model has greater explanatory power than the existing state-of-the-art models estimated by Shumway (2001) and Chava and Jarrow (2002), and includes additional variables with sensible economic motivation. We believe that models of the sort estimated here have meaningful empirical advantages over the bankruptcy risk scores proposed by Altman (1968) and Ohlson (1980). While Altman's Z-score and Ohlson's O-score were seminal early contributions, better measures of bankruptcy risk are available today. We have also presented evidence that bankruptcy risk cannot be adequately summarized by a measure of distance to default inspired by Merton's (1974) pioneering structural model of bankruptcy. While our distance to default measure is not exactly the same as those used by Crosbie and Bohn (2001) and Vassalou and Xing (2004), we believe that this result will be robust to alternative measures of distance to default.

Second, we show that stocks with a high risk of bankruptcy tend to deliver anomalously low average returns. We sort stocks by our 12-month-ahead estimate of bankruptcy risk. Distressed portfolios, containing stocks with high estimated bankruptcy risk, have low average returns but high standard deviations, market betas, and loadings on Fama and French's (1993) small-cap and value risk factors. Thus, from the perspective of any of the leading empirical asset pricing models, these stocks have negative alphas. This result is a significant challenge to the conjecture that the value and size effects are proxies for a distress premium. More generally, it is a challenge to standard models of rational asset pricing in which the structure of the economy is stable and well understood by investors.

This version of our paper is preliminary, and we plan to expand our empirical investigation in several directions. One particularly important question is whether the determinants of bankruptcy risk are stable over time. Our best model has a good in-sample fit, but we would like to show that it predicts bankruptcy when estimated using only data available up to each point of time. Rolling estimates of the model will also allow us to construct portfolios of distressed stocks using contemporaneously available information about the determinants of bankruptcy.

It should be possible to refine our understanding of the bankruptcy risk anomaly by sorting stocks on other characteristics. Our results are consistent with the findings of Dichev (1998), who uses Altman's Z-score and Ohlson's O-score to measure financial distress. Vassalou and Xing (2004) use distance to default as an alternative distress measure; they find some evidence that distressed stocks have higher returns, but this evidence comes entirely from small value stocks. Griffin and Lemmon (2002), using O-score to measure distress, show that distressed growth stocks have particularly low returns. This literature suggests that it will be fruitful to sort stocks on size and book-market ratio as well as bankruptcy risk, in order to identify which types of distressed stocks have positive or negative abnormal returns.

One possible explanation of the bankruptcy risk anomaly is that it results from the preferences of institutional investors, together with a shift of assets from individuals to institutions during our sample period. Kovtunenko and Sosner (2003) have documented that institutions prefer to hold profitable stocks, and that this preference helped institutional performance during the 1980's and 1990's because profitable stocks outperformed the market. It is possible that the strong performance of profitable stocks in this period was endogenous, the result of increasing demand for these stocks by institutions. If institutions more generally prefer stocks with low bankruptcy risk, and tend to sell stocks that enter financial distress, then a similar mechanism could drive our results. This hypothesis can be tested by relating the performance of distressed stocks over time to the changing institutional share of equity ownership.

Appendix

In this appendix we discuss issues related to the construction of our dataset. All variables are constructed using COMPUSTAT and CRSP data. Relative size, excess return, and accounting ratios are defined as follows:

$$\begin{aligned}
RSIZE_{i,t} &= \log\left(\frac{Firm\ Market\ Equity_{i,t}}{Total\ S\&P500\ Market\ Value_t}\right) \\
EXRET_{i,t} &= \log(1+R_{i,t}) - \log(1+R_{S\&P500,t}) \\
NITA_{i,t} &= \frac{Net\ Income_{i,t}}{Total\ Assets(adjusted)_{i,t}} \\
TLTA_{i,t} &= \frac{Total\ Liabilities_{i,t}}{Total\ Assets(adjusted)_{i,t}} \\
NIMTA_{i,t} &= \frac{Net\ Income_{i,t}}{(Firm\ Market\ Equity_{i,t} + Total\ Liabilities_{i,t})} \\
TLMTA_{i,t} &= \frac{Total\ Liabilities_{i,t}}{(Firm\ Market\ Equity_{i,t} + Total\ Liabilities_{i,t})} \\
CASHMTA_{i,t} &= \frac{Cash\ and\ Short\ Term\ Investments_{i,t}}{Total\ Assets(adjusted)_{i,t}}
\end{aligned}$$

The COMPUSTAT data items used are total assets Data6, net income Data172, and total liabilities Data181.

To deal with outliers in the data, we correct both NITA and TLTA using the difference between book equity (BE) and market equity (ME) to adjust the value of total assets:

$$Total\ Assets\ (adjusted)_{i,t} = TA_{i,t} + 0.1 * (BE_{i,t} - ME_{i,t})$$

Book equity is as defined in Davis, Fama and French (2000) and outlined in detail in Cohen, Polk and Vuolteenaho (2003). This transformation helps with the values of total assets that are very small, probably mismeasured and lead to very large values of NITA. After total assets are adjusted, each of the seven explanatory variables is winsorized using a 5/95 percentile interval in order to eliminate outliers.

To measure the volatility of a firm's stock returns, we use a proxy, centered around zero rather than the rolling three-month mean, for daily variation of returns computed as an annualized three-month rolling sample standard deviation:

$$SIGMA_{i,t-1,t-3} = \left(252 * \frac{1}{N-1} \sum_{k \in \{t-1,t-2,t-3\}} r_{i,k}^2 \right)^{\frac{1}{2}}$$

To eliminate cases where few observations are available, SIGMA is coded as missing if there are fewer than five non-zero observations over the three months used in the rolling-window computation. This leads to a loss of about 50,000 observations and 48 bankruptcy events.

To measure distance to default (DD), we use the following formula from Crosbie and Bohn (2001) and Vassalou and Xing (2004):

$$DD = \frac{-\log(BD/MTA) + 0.06 + R_{BILL} - SIGMA^2/2}{SIGMA},$$

where BD is the book value of debt and MTA is total assets at market value. Following the literature, we measure the book value of debt as the book value of short-term debt, plus one-half the book value of long-term debt. This convention 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.

The number 0.06 appears in the formula as an empirical proxy for the equity premium. Vassalou and Xing (2004) instead estimate the average return on each stock, but we believe that it is better to use a common expected return for all stocks than a noisily estimated stock-specific number. Vassalou and Xing also implement an iterative procedure to estimate the volatility of the firm's underlying asset returns. We instead assume that the firm's asset volatility equals its equity volatility, as would be the case if the market value of the debt is as volatile as the equity. This assumption will understate the distance to default of safe firms, but should be increasingly accurate as the distance to default shrinks.

References

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

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2003, The cross-section of volatility and expected returns, unpublished paper, Columbia University.

Asquith, Paul, Robert Gertner, and David Scharfstein, 1994, Anatomy of financial distress: An examination of junk-bond issuers, *Quarterly Journal of Economics* 109, 625–658.

Bernanke, Ben S. and John Y. Campbell, 1988, Is there a corporate debt crisis?, *Brookings Papers on Economic Activity* 1, 83–139.

Burgstahler, D.C. and I.D. Dichev, 1997, Earnings management to avoid earnings decreases and losses, *Journal of Accounting and Economics* 24, 99–126.

Campbell, John Y., Martin Lettau, Burton Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of Finance* 56, 1–43.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chacko, George, Peter Hecht, and Jens Hilscher, 2004, Time varying expected returns, stochastic dividend yields, and default probabilities, unpublished paper, Harvard Business School.

Chan, K.C. and Nai-fu Chen, 1991, Structural and return characteristics of small and large firms, Journal of Finance 46, 1467–1484.

Chava, Sudheer and Robert A. Jarrow, 2002, Bankruptcy prediction with industry effects, unpublished paper, Cornell University.

Cohen, Randolph B., Christopher Polk and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.

Crosbie, Peter J. and Jeffrey R. Bohn, 2001, *Modeling Default Risk*, KMV, LLC, San Francisco, CA.

Davis, James L., Eugene F. Fama and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *Journal of Finance* 55, 389–406.

Dechow, Patricia M., Scott A. Richardson, and Irem Tuna, 2003, Why are earnings kinky? An examination of the earnings management explanation, *Review of Accounting Studies* 8, 355–384.

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

Duffie, Darrell, and Ke Wang, 2003, Multi-period corporate failure prediction with stochastic covariates, unpublished paper, Stanford University.

Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics 33, 3–56.*

Fama, Eugene F. and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Ferguson, Michael F. and Richard L. Shockley, 2003, Equilibrium "anomalies", *Journal of Finance* 58, 2549–2580.

Gilson, Stuart C., Kose John, and Larry Lang, 1990, Troubled debt restructurings: An empirical study of private reorganization of firms in default, *Journal of Financial Economics* 27, 315–353.

Gilson, Stuart C., 1997, Transactions costs and capital structure choice: Evidence from financially distressed firms, *Journal of Finance* 52, 161–196.

Griffin, John M. and Michael L. Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *Journal of Finance* 57, 2317–2336.

Hayn, C., 1995, The information content of losses, *Journal of Accounting and Economics* 20, 125–153.

Hillegeist, Stephen A., Elizabeth Keating, Donald P. Cram and Kyle G. Lunstedt, 2004, Assessing the probability of bankruptcy, *Review of Accounting Studies* 9, 5–34.

Kovtunenko, Boris and Nathan Sosner, 2003, Sources of institutional performance, unpublished paper, Harvard University.

Merton, Robert C., 1974, On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance* 29, 449–470.

Mossman, Charles E., Geoffrey G. Bell, L. Mick Swartz, and Harry Turtle, 1998, An empirical comparison of bankruptcy models, *Financial Review* 33, 35–54.

Ohlson, James A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109–131.

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

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.

Shumway, Tyler, 2001, Forecasting bankruptcy more accurately: a simple hazard model, *Journal of Business* 74, 101–124.

Shumway, Tyler and Vincent A. Warther, 1999, The delisting bias in CRSP's Nasdaq data and its implications for the size effect, *Journal of Finance* 54, 2361–2379.

Tashjian, Elizabeth, Ronald Lease, and John McConnell, 1996, Prepacks: An empirical analysis of prepackaged bankruptcies, *Journal of Financial Economics* 40, 135–162.

Vassalou, Maria and Yuhang Xing, 2004, Default risk in equity returns, *Journal of Finance* 59, 831–868.

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

Table 1: Number of Bankruptcies per year

The table lists the total number of active firms (Column 1) and the total number of bankruptcies (Column 2) for every year of our sample period. The number of active firms is computed by averaging over the numbers of active firms across all months of the year. The last column shows the percentage of bankruptcies.

Year	Active Firms	Bankruptcies	(%)
1963	1251	0	0.00
1964	1297	2	0.15
1965	1372	2	0.15
1966	1446	1	0.07
1967	1542	0	0.00
1968	1641	0	0.00
1969	1817	0	0.00
1970	1951	5	0.26
1971	2044	4	0.20
1972	2282	8	0.35
1973	3531	6	0.17
1974	3546	18	0.51
1975	3544	5	0.14
1976	3570	14	0.39
1977	3574	12	0.34
1978	3910	14	0.36
1979	4041	14	0.35
1980	4146	26	0.63
1981	4500	23	0.51
1982	4687	29	0.62
1983	4923	50	1.02
1984	5354	73	1.36
1985	5360	76	1.42
1986	5531	95	1.72
1987	5954	54	0.91
1988	6026	84	1.39
1989	5942	74	1.25
1990	5906	80	1.35
1991	5918	70	1.18
1992	6213	45	0.72
1993	6732	36	0.53
1994	7408	30	0.40
1995	7637	43	0.56
1996	8011	32	0.40
1997	8302	44	0.53
1998	8175	49	0.60

Table 2: Summary statistics

The tables include the following variables (various adjustments are described in the Data Description section): Net Income over market value of Total Assets (NIMTA), Total Liabilities over market value of Total Assets (TLMTA), log of gross excess return over value weighted S&P 500 return (EXRET) annualized, i.e. log(1+simple excess return), log of firm's market equity over the total valuation of S&P 500 (RSIZ), square root of a sum of squared firm stock returns over a three-month period (annualized) (SIGMA), stock of cash and short term investments over the market value of Total assets (CASHMTA), and market-to-book value of the firm (MB). Market value of total assets was computed by adding market value of firm equity to its total liabilities.
The first group reports summary statistics for the non-bankruptcy firm-month observations and the second panel for the bankruptcy ones. We have a total of 1,281,426 observations, of which 796 are bankruptcy events. In both cases, the panels only contain statistics for values where all variables were non-missing (i.e. those observations that were then actually used in regressions).

Non-Bankruptcy group

	NIMTA	TLMTA	EXRET	RSIZE	SIGMA	CASHMTA	MB
Mean	0.002	0.441	-0.012	-10.272	0.519	0.079	1.991
Median	0.007	0.424	-0.011	-10.400	0.442	0.044	1.542
St. Dev	0.020	0.273	0.108	1.882	0.295	0.087	1.474
Min	-0.055	0.038	-0.226	-13.291	0.155	0.002	0.400
Max	0.029	0.918	0.202	-6.712	1.212	0.318	6.041

Observations: 1280630

Bankruptcy group

	NIMTA	TLMTA	EXRET	RSIZE	SIGMA	CASHMTA	MB
Mean	-0.035	0.762	-0.109	-12.327	0.994	0.043	2.333
Median	-0.047	0.861	-0.173	-12.881	1.212	0.021	1.018
St. Dev	0.025	0.209	0.140	1.264	0.299	0.060	2.333

Observations: 796

Table 3: Logit regressions of default indicator on predictor variables

This table reports results from logit regressions of the default indicator on predictor variables. The value of the predictor variable is known at the beginning of the month over which bankruptcy is measured. Net income and total liabilities are scaled by accounting and market total assets.

	(1)	(2)	(3)	(4)	(5)	(6)
NITA	-16.728					
	(15.47)**					
NIMTA		-33.322	-32.935			
		(19.62)**	(19.91)**			
NIMTAAVG				-42.701	-42.754	-42.768
				(19.47)**	(19.28)**	(19.27)**
TLTA	5.612					
	(25.06)**					
TLMTA		4.866	4.827	4.706	4.479	4.790
		(25.27)**	(25.71)**	(25.06)**	(23.51)**	(23.61)**
EXRET	-3.566	-3.126	-3.111			
	(11.96)**	(10.61)**	(10.72)**			
EXRETAVG				-10.993	-11.674	-11.529
				(13.40)**	(14.16)**	(13.95)**
SIGMA	2.776	2.777	2.749	2.212	2.060	2.497
	(16.16)**	(16.29)**	(16.29)**	(12.74)**	(11.88)**	(12.39)**
RSIZE	-0.123	-0.036	-0.043	0.030	0.038	0.029
	(3.29)**	(0.98)	(1.20)	(0.84)	(1.04)	(0.80)
CASHMTA					-4.954	-4.826
					(7.93)**	(7.74)**
MB					0.124	0.124
					(7.73)**	(7.76)**
DD						0.094
						(4.79)**
SIGMAMISS			1.272	1.095	1.110	1.165
			(6.22)**	(5.31)**	(5.38)**	(5.64)**
Constant	-15.158	-13.586	-13.611	-12.618	-12.234	-13.160
	(36.58)**	(34.88)**	(35.11)**	(32.60)**	(31.02)**	(29.60)**
Observations	1197376	1264108	1281426	1281426	1281426	1281426
No. of bankruptcies	737	762	796	796	796	796
Pseudo R squared	0.268	0.273	0.268	0.282	0.293	0.295

Absolute value of z statistics in parentheses

* significant at 5%; ** significant at 1%

Table 4: Logit regressions on lagged best-model variables

The table below takes our best variables and tests their predictive power as we lag them by 6, 12, 24, and 36 months. Panel A uses all the available observations for each of the 5 regressions. Panel B restricts the available observations to the minimum set so that each regressions has the same number of observations.

Panel A	(1)	(2)	(3)	(4)	(5)
Months lagged by:	0	6	12	24	36
NIMTAAVG	-42.768	-34.854	-25.811	-15.958	-9.931
	(19.27)**	(18.01)**	(12.81)**	(6.87)**	(3.66)**
TLMTA	4.790	2.736	2.080	1.374	1.293
	(23.61)**	(17.85)**	(13.50)**	(8.08)**	(6.83)**
EXRETAVG	-11.529	-10.471	-10.421	-6.505	-5.816
	(13.95)**	(13.14)**	(12.33)**	(6.85)**	(5.45)**
SIGMA	2.497	1.289	1.042	0.768	0.639
	(12.39)**	(8.20)**	(6.56)**	(4.41)**	(3.26)**
RSIZE	0.029	-0.057	-0.057	-0.093	-0.097
	(0.80)	(1.97)*	(2.08)*	(3.29)**	(3.26)**
CASHMTA	-4.826	-6.064	-4.738	-3.019	-2.843
	(7.74)**	(10.36)**	(8.70)**	(5.73)**	(4.93)**
MB	0.124	0.109	0.112	0.096	0.123
	(7.76)**	(6.39)**	(5.90)**	(4.13)**	(4.65)**
DD	0.094	0.020	0.001	-0.019	-0.017
	(4.79)**	(1.31)	(0.10)	(1.34)	(1.13)
SIGMAMISS	1.165	0.071	-0.072	-0.272	-1.541
	(5.64)**	(0.29)	(0.26)	(0.79)	(2.16)*
Constant	-13.160	-10.624	-9.784	-9.266	-9.201
	(29.60)**	(30.31)**	(29.05)**	(26.59)**	(24.82)**
Observations	1281426	1193002	1108671	954603	822626
No. of bankruptcies	796	926	865	729	605
Pseudo R squared	0.295	0.172	0.112	0.052	0.037
Panel B	(1)	(2)	(3)	(4)	(5)
NIMTAAVG	-40.892	-36.679	-25.003	-15.229	-10.082
	(14.04)**	(13.21)**	(8.93)**	(5.17)**	(3.23)**
TLMTA	6.023	4.292	3.455	2.266	1.449
	(17.72)**	(16.28)**	(14.15)**	(10.15)**	(6.68)**
EXRETAVG	-10.922	-9.842	-11.118	-7.666	-6.856
	(10.08)**	(8.66)**	(9.53)**	(6.34)**	(5.60)**
SIGMA	2.893	1.455	1.199	0.922	0.685
	(10.41)**	(6.46)**	(5.51)**	(4.24)**	(3.07)**
RSIZE	0.075	-0.028	-0.044	-0.076	-0.083
	(1.64)	(0.70)	(1.19)	(2.21)*	(2.47)*
CASHMTA	-4.062	-4.756	-4.282	-2.638	-2.561
	(4.85)**	(5.70)**	(5.52)**	(3.95)**	(3.98)**
MB	0.112	0.113	0.091	0.105	0.119
	(5.62)**	(5.04)**	(3.66)**	(3.72)**	(3.94)**
DD	0.100	0.029	0.004	-0.008	-0.014
	(3.62)**	(1.34)	(0.21)	(0.46)	(0.87)
SIGMAMISS	0.838	-0.631	-0.308	-0.419	-1.139
	(2.44)*	(1.22)	(0.67)	(0.82)	(1.59)
Constant	-13.926	-11.845	-10.897	-10.025	-9.378
	(23.80)**	(24.27)**	(23.86)**	(23.41)**	(22.47)**
Observations	773159	773159	773159	773159	773159
No. of bankruptcies	470	470	470	470	470
Pseudo R squared	0.321	0.209	0.147	0.068	0.038

Absolute value of z statistics in parentheses * significant at 5%; ** significant at 1%

Table 5: Logit regressions with macro variables

The regressions below include two macro variables. Regression in column 1 includes default yield spread (DFY), the spread between the yields on BAA and AAA rated debt. Regression in column two includes term spread (TMS), the difference between yields on a 10-year government bond and 6-month T-Bill. Both regressions include interactions of these two variables with relative size (RSIZE) and leverage (TLMTA).

	(1)	(2)
NIMTAAVG	-42.229	-42.407
	(18.95)**	(19.07)**
TLMTA	6.312	5.675
	(14.03)**	(13.53)**
TLMTA*DFY	-175.289	
	(5.09)**	
TLMTA*TMS		-63.873
		(3.36)**
EXRETAVG	-11.454	-11.719
	(13.90)**	(14.12)**
SIGMA	2.184	2.062
	(12.72)**	(11.87)**
RSIZE	0.164	0.048
	(2.13)*	(0.84)
RSIZE*DFY	-13.020	
	(2.25)*	
RSIZE*TMS		-0.252
		(0.10)
CASHMTA	-4.667	-4.893
	(7.46)**	(7.81)**
MB	0.127	0.121
	(7.94)**	(7.58)**
SIGMAMISS	0.971	1.093
	(4.72)**	(5.30)**
DFY	37.025	
	(0.51)	
TMS		53.101
		(1.60)
Constant	-12.831	-13.134
	(13.80)**	(18.18)**
Observations	1243098	1243098
No. of bankruptcies	792	792
Pseudo R squared	0.299	0.294

Absolute value of z statistics in parentheses

* significant at 5%; ** significant at 1%

Table 6: Equity Returns on bankruptcy portfolios

We sorted all stocks based on the predicted probability of bankruptcy 13-months ahead (a 12-month lag to our basic model) and divided them into 10 portfolios based on percentile cutoffs. For example, 0 to 5th percentile (0005) and 99th to 100th percentile (9900). In the table below we show results from regressions on a constant, excess market return (RM), as well as three (RM, HML, SMB) and four (RM, HML, SMB, UMD) FF factor regressions. Panel A shows alphas from these regressions and the corresponding t-stat below. Panel B shows loadings on the three factors, as well as corresponding t-stats below, from the 3-factor regression. The last row in Panel B reports monthly standard deviation of the portfolio returns.

Panel A

Portfolios	**0005**	**0510**	**1020**	**2040**	**4060**	**6080**	**8090**	**9095**	**9599**	**9900**
Mean return	0.006	0.004	0.003	0.002	0.001	0.001	0.000	-0.002	-0.002	-0.004
	(3.74)**	(2.64)**	(2.19)*	(1.48)	(0.40)	(0.55)	(0.05)	(0.54)	(0.57)	(0.66)
CAPM alpha	0.006	0.003	0.002	0.001	0.000	0.001	-0.001	-0.003	-0.003	-0.006
	(3.68)**	(2.25)*	(1.47)	(0.98)	(0.29)	(0.52)	(0.23)	(0.82)	(0.78)	(0.88)
3-factor alpha	0.005	0.004	0.002	0.001	-0.001	-0.002	-0.004	-0.007	-0.008	-0.013
	(6.28)**	(4.56)**	(4.11)**	(1.37)	(2.15)*	(1.68)	(2.74)**	(2.87)**	(2.52)*	(2.04)*
4-factor alpha	0.004	0.003	0.001	0.000	-0.001	-0.001	-0.004	-0.006	-0.008	-0.010
	(5.22)**	(3.37)**	(2.59)*	(0.61)	(2.12)*	(0.98)	(2.41)*	(2.46)*	(2.25)*	(1.52)

Panel B

Portfolios	**0005**	**0510**	**1020**	**2040**	**4060**	**6080**	**8090**	**9095**	**9599**	**9900**
RM	-0.129	-0.095	-0.035	-0.032	-0.046	-0.034	0.034	0.052	0.059	0.204
	(6.27)**	(4.89)**	(2.71)**	(2.60)**	(3.31)**	(1.47)	(0.93)	(0.90)	(0.71)	(1.30)
HML	0.014	-0.136	-0.139	0.027	0.222	0.384	0.527	0.592	0.766	1.087
	(0.43)	(4.25)**	(6.52)**	(1.36)	(9.78)**	(10.02)**	(8.80)**	(6.30)**	(5.69)**	(4.24)**
SMB	0.853	0.786	0.705	0.741	0.821	0.957	1.299	1.509	1.767	1.944
	(27.43)**	(26.60)**	(35.76)**	(40.34)**	(39.06)**	(26.95)**	(23.42)**	(17.29)**	(14.16)**	(8.17)**
St. Deviation	0.025	0.024	0.021	0.020	0.023	0.029	0.041	0.054	0.070	0.112

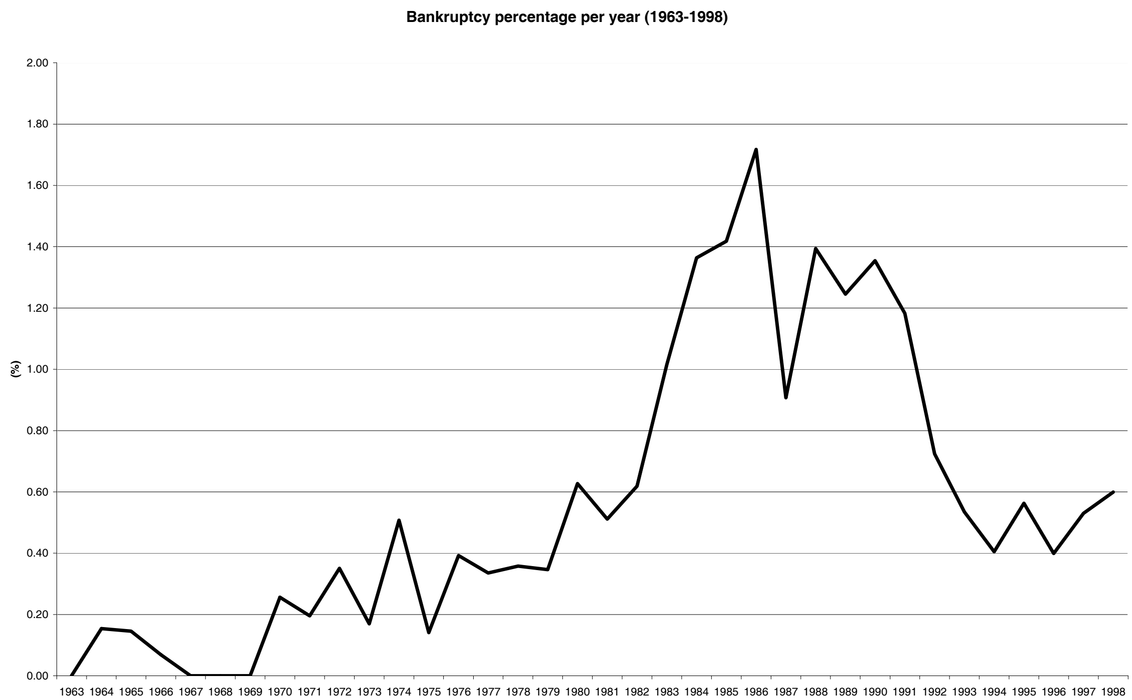

Figure 1 The graph shows the percentage of active firms (where the number of firms in a year is taken as the average over the year) that were bankrupt in any given year over the period 1963-1998.

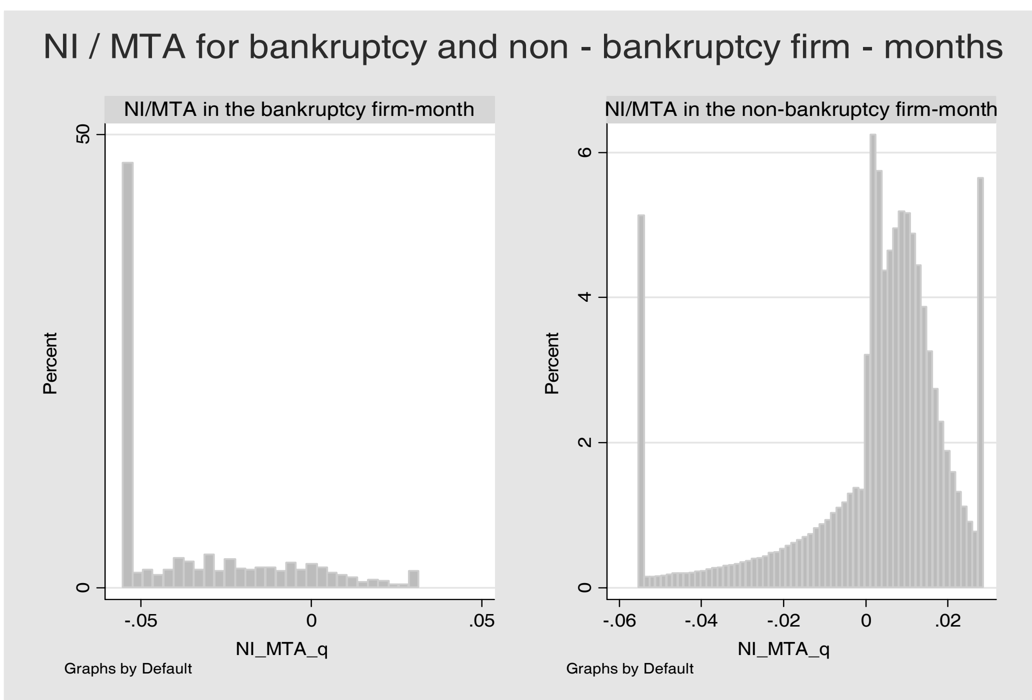

Figure 2 The graph above shows the distribution of the company's Net Income over market-adjusted Total Assets (the sum of market value of equity and book value of Total Liabilities) for the firm-month when the firm was in bankruptcy (left) and all others (right). The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

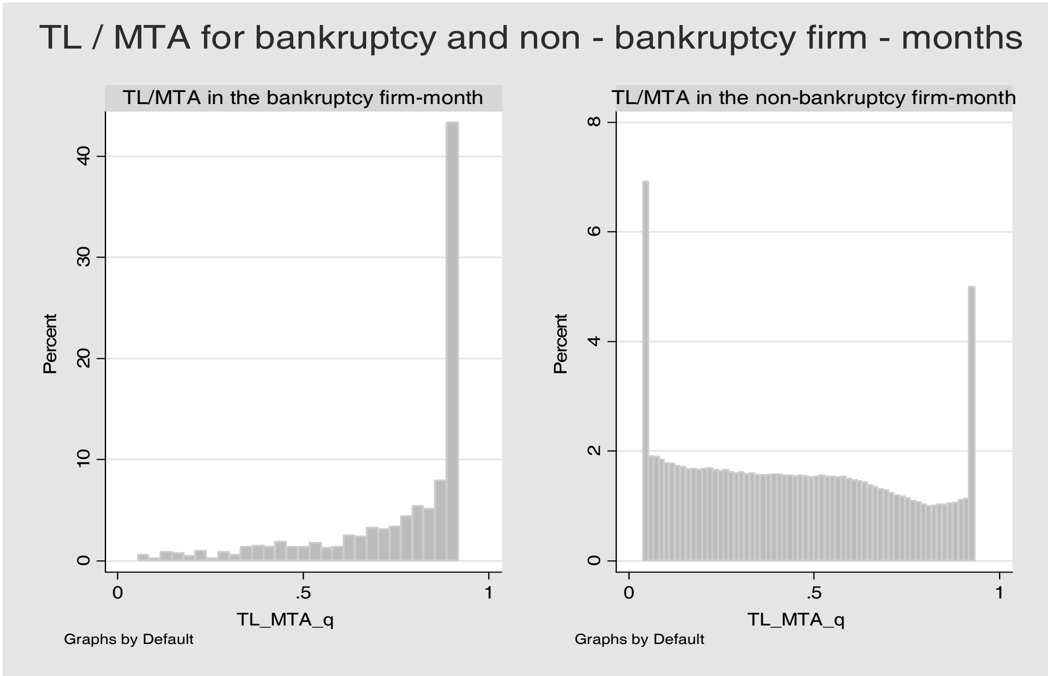

Figure 3 The graph above shows the distribution of the company's Total Liabilities over market-adjusted Total Assets (the sum of market value of equity and book value of Total Liabilities) for the firm-month when the firm was in bankruptcy (left) and all others (right). The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

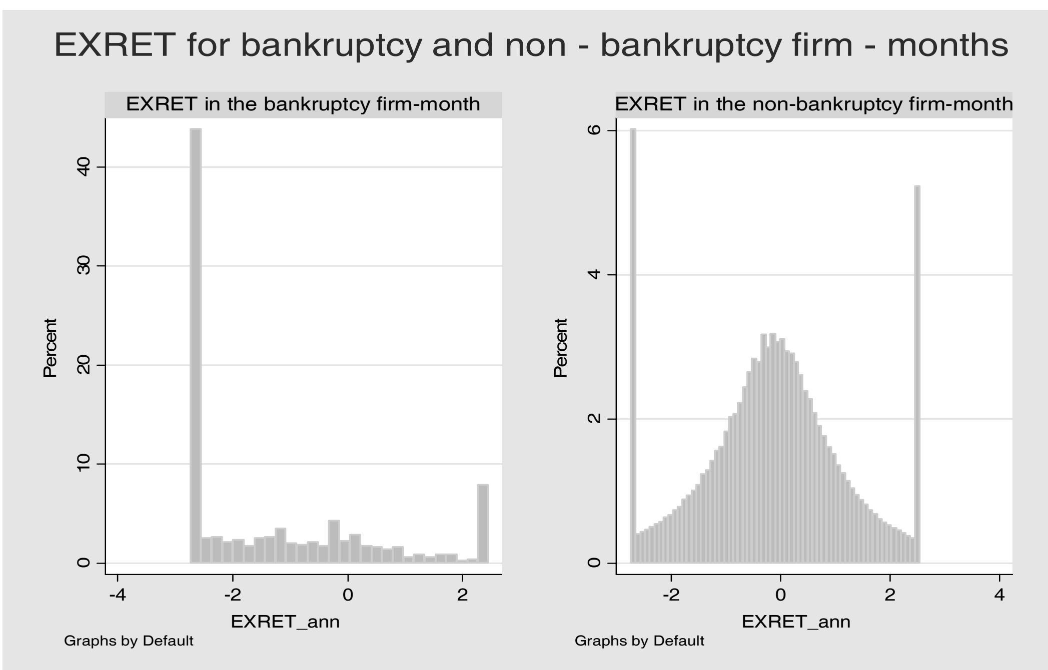

Figure 4 The graph above shows the distribution of the log gross excess monthly return on company stock over the value-weighted S&P500 Index (EXRET) for the firm-month when the firm was in bankruptcy (left) and all others (right). It was constructed as the log of a gross return based on the simple difference between the return on the company stock and return on the index in a given month. The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

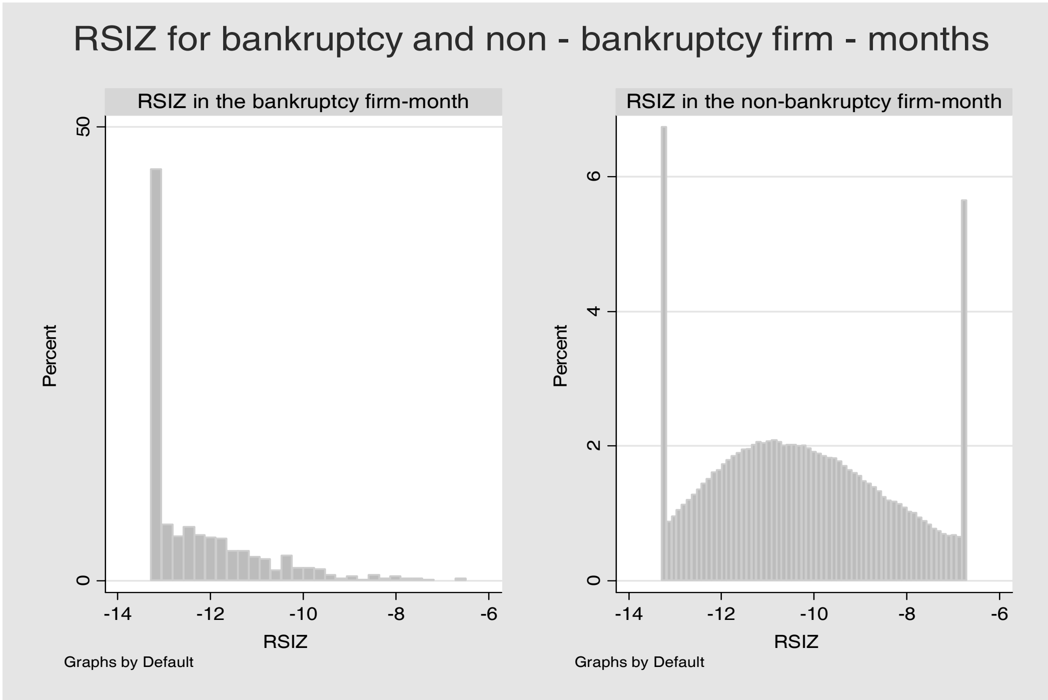

Figure 5 The graph above shows the distribution of the log of the ratio of company's market equity and S&P 500 index total market valuation (RSIZ) for the firm-month when the firm was in bankruptcy (left) and all others (right). The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

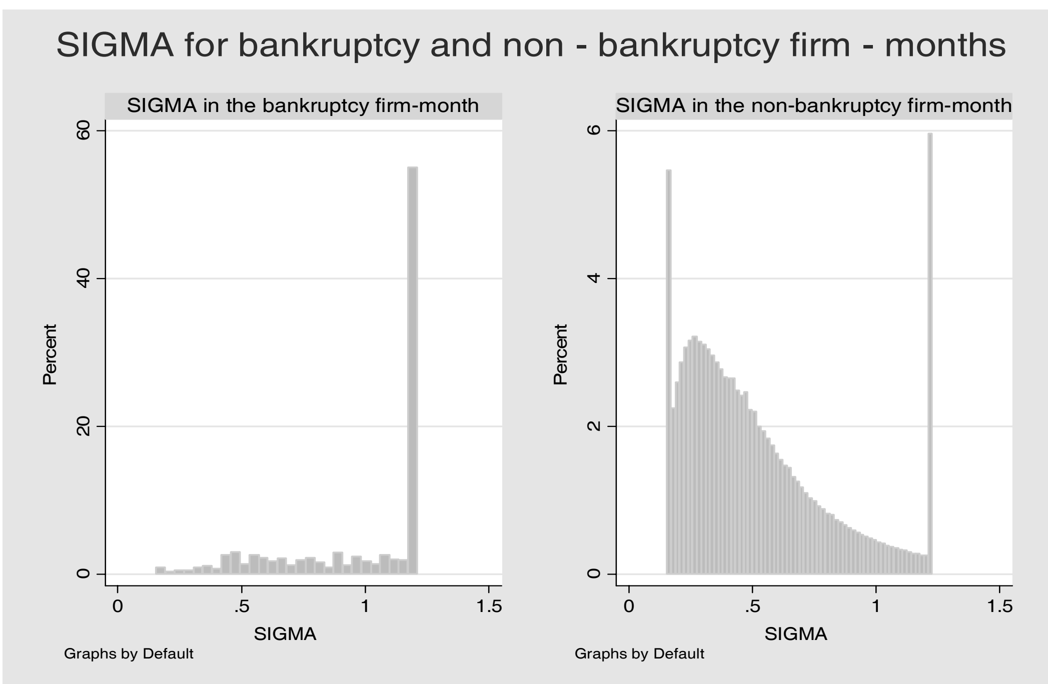

Figure 6 The graph above shows the distribution of the estimate of daily return variations (annualized) -- constructed by summing over squared daily returns in any given month and diving by the number of observations less 1 -- for the firm-month when the firm was in bankruptcy (left) and all others (right). The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

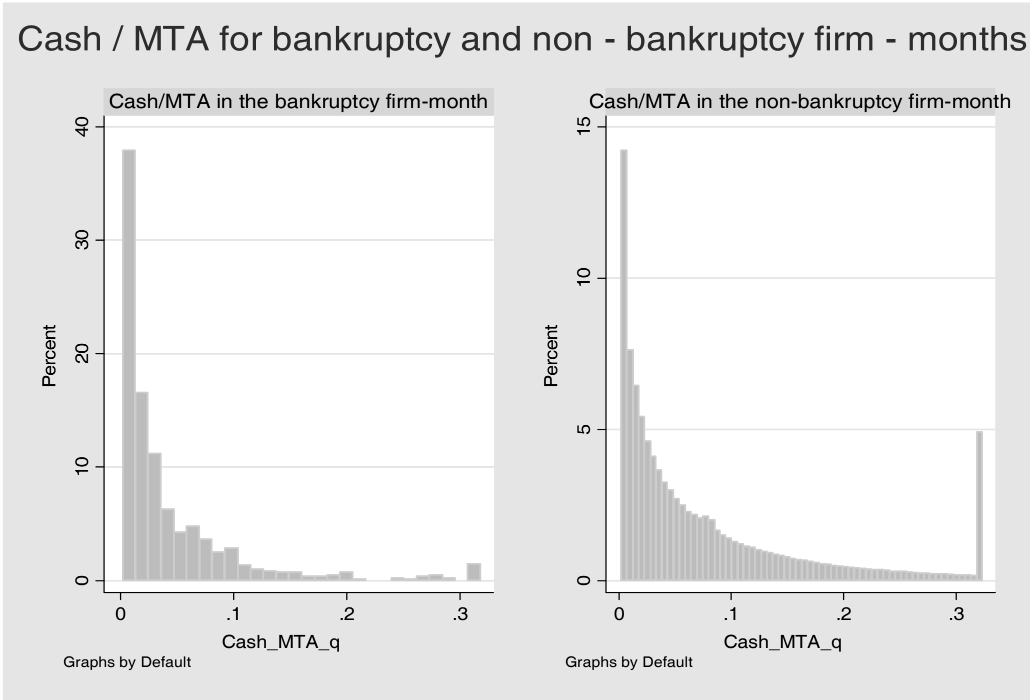

Figure 7 The graph above shows the distribution of the company's Cash and Cash equivalent securities over Total Assets (adjusted as described in the Data section) for the firm-month when the firm was in bankruptcy (left) and all others (right). The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

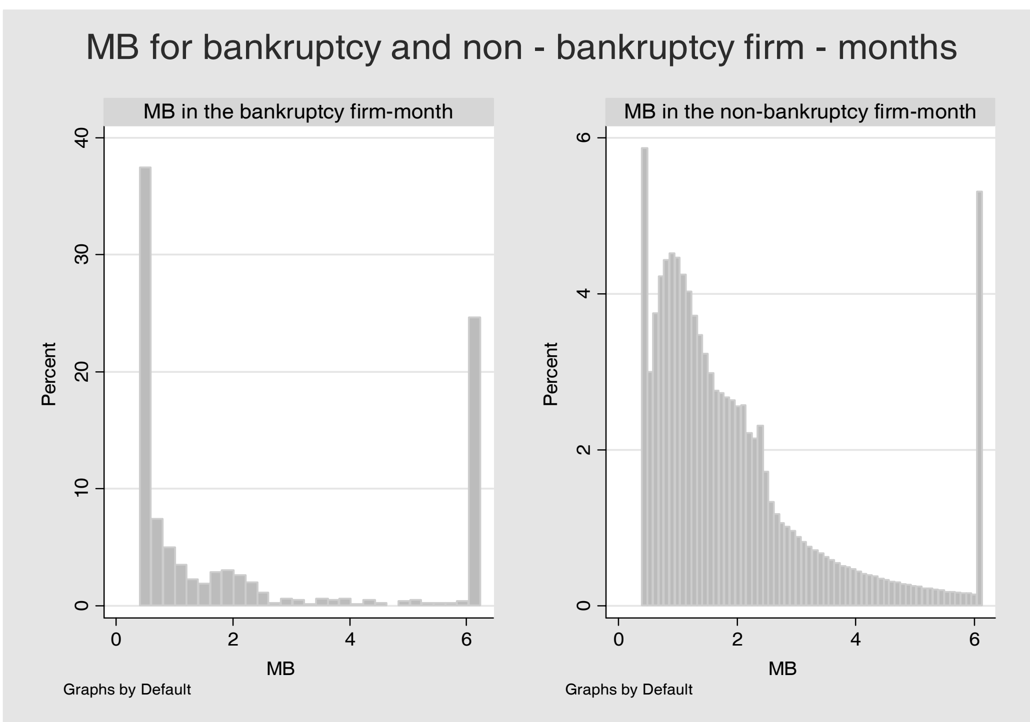

Figure 8 The graph above shows the distribution of the company's Market-to-Book value (adjusted as described in the Data section) for the firm-month when the firm was in bankruptcy (left) and all others (right). The mass points in either graph represent the effect of winsorization (i.e. the resulting mass points). The variable has also been lagged to ensure that the information was the latest publicly available in the month with which it is associated**.**

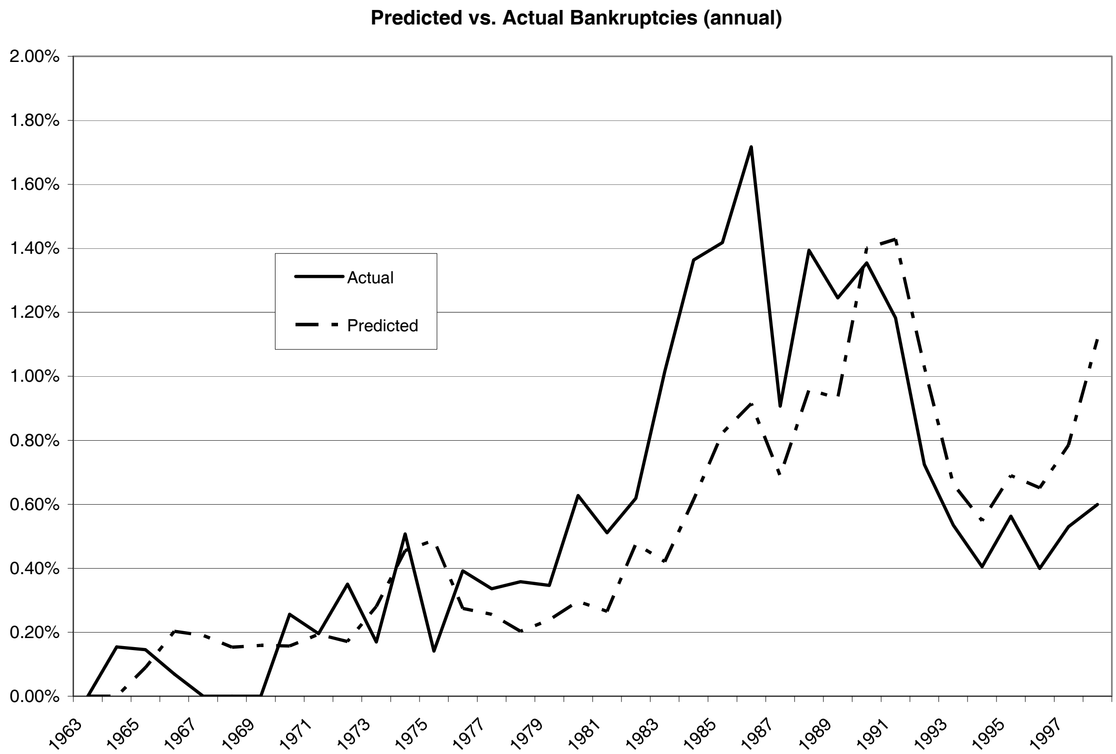

Figure 9 The graph above represents the percentage of aggregate bankruptcies as predicted by our model and the actual bankruptcies that occurred in that year. To get the predicted percentage of bankruptcies, we average over the predicted probability (based on our model) of all active firms in a given month.

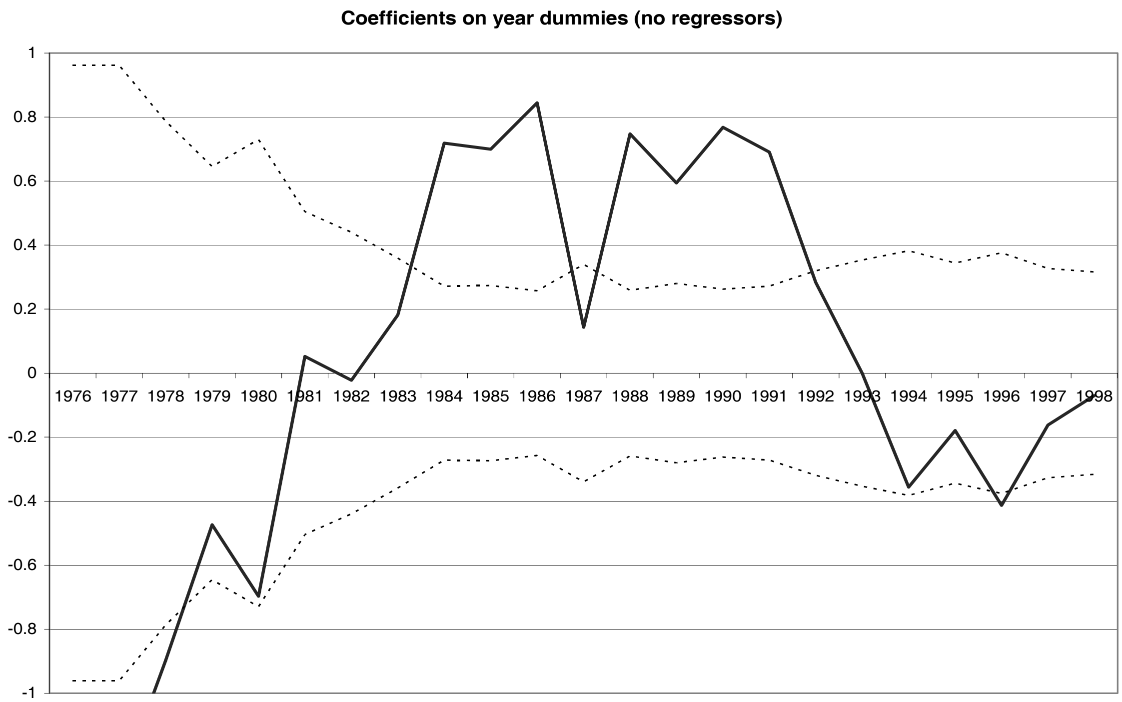

Figure 10 The graph above shows the deviation of the year dummy coefficients from their sample mean. The coefficient estimates are from the regression that included only dummies. The number of observations is equal to that in our best model regression. The dotted line shows a 2 standard deviation confidence interval around zero for each coefficient value.

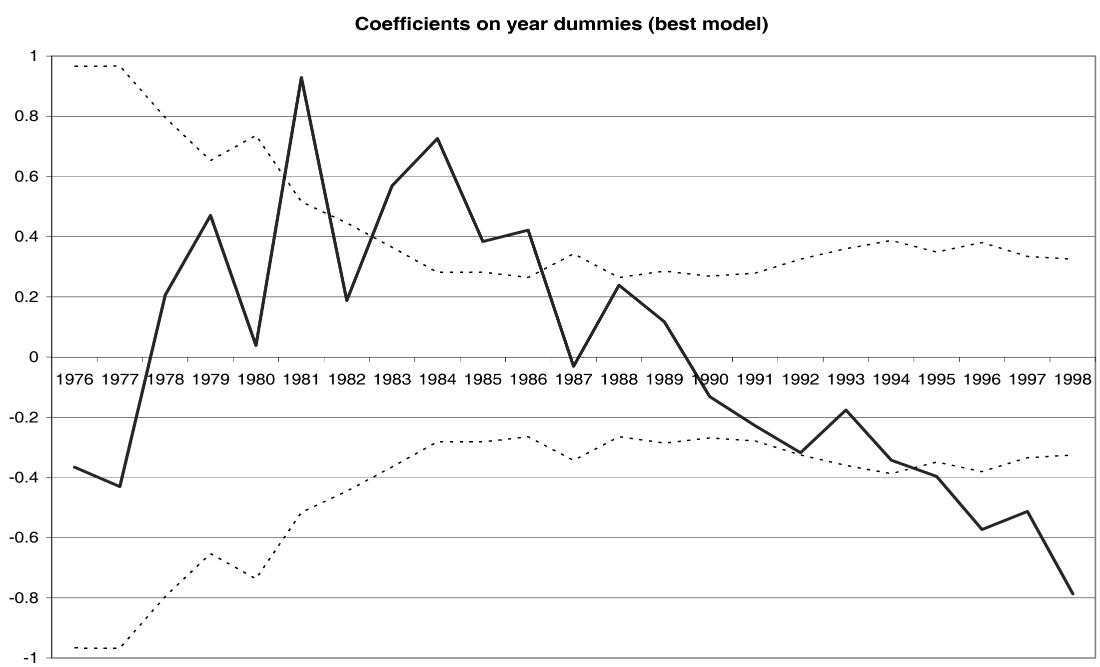

Figure 11 The graph above shows the deviation of the year dummy coefficients from their sample mean. The coefficient estimates are from our best fitting regression. The dotted line shows a 2 standard deviation confidence interval around zero for each coefficient value.

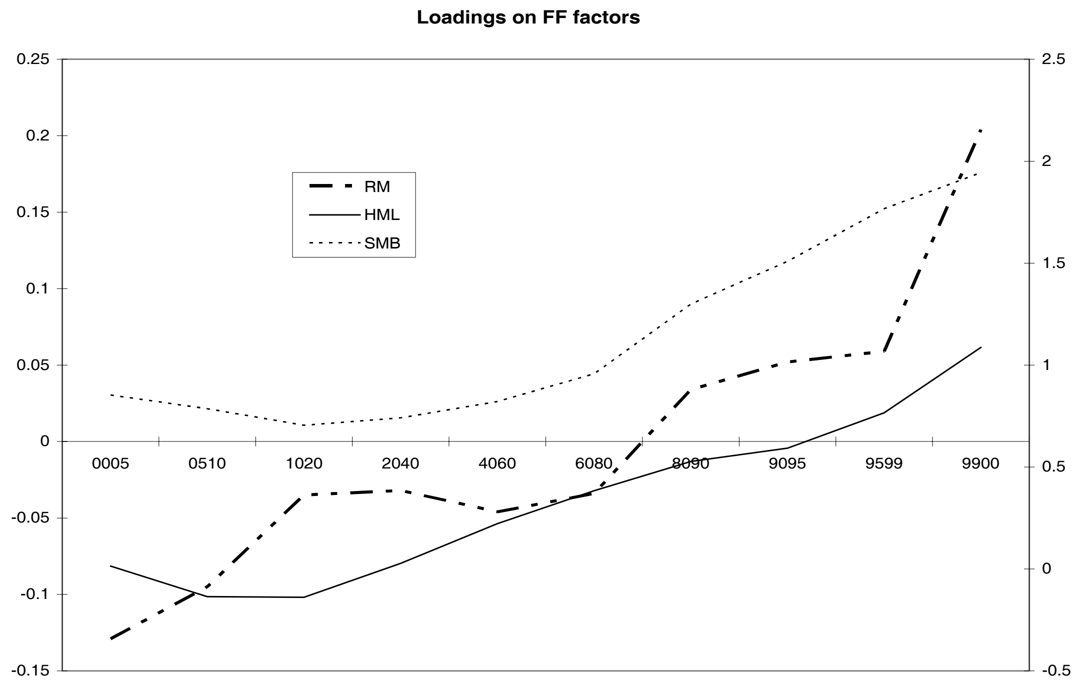

Figure 12 The graph above shows the loadings (betas) on Fama-French factors from the regressions of bankruptcy portfolios on the three factors (RM, HML, SMB).

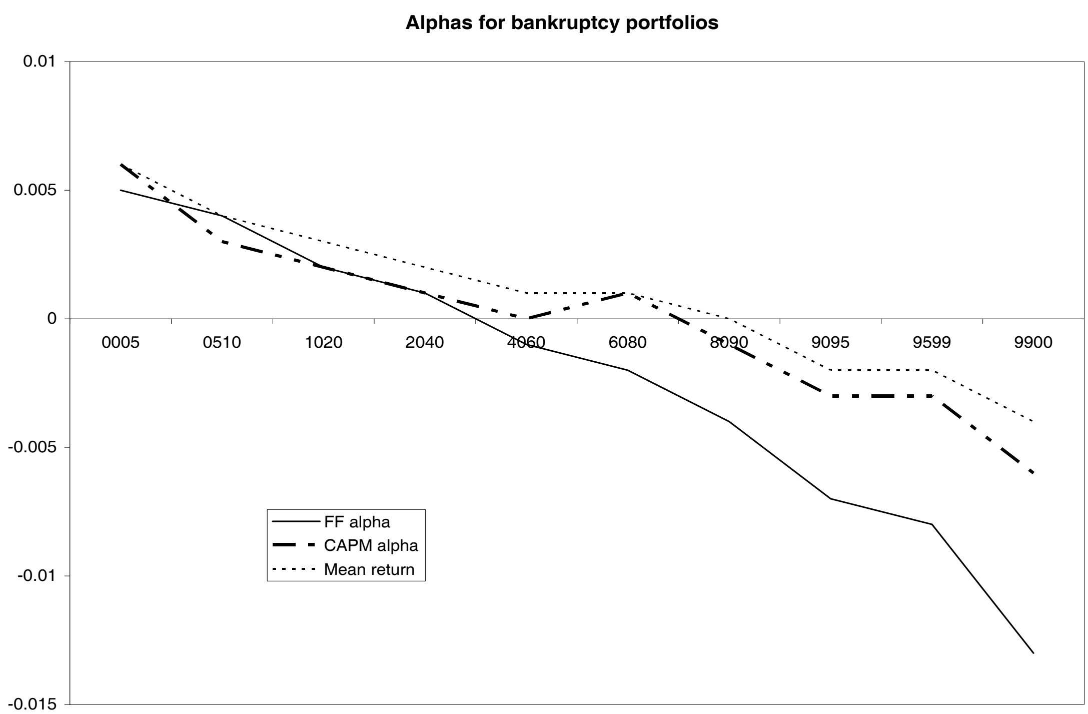

Figure 13 The graph shows alphas from various regressions: regression on a constant (Mean return), three-factor regression (FF alpha) and the CAPM regression (CAPM alpha).