A Bankruptcy Risk Factor

Jesse Neumann*

Rutgers University

Draft September 2021

Abstract

Long-short factors sort assets based on specific characteristics which proxies for the level of risk facing a firm. Bypassing these indirect proxies, this paper introduces a factor based on an estimated probability of bankruptcy — a measure of the actual risk a typical investor will lose their investment. Using an underlying model of firm bankruptcy built as a sequence of two random forests, I demonstrate my bankruptcy risk factor has predictive power in equity, bond, options and credit default swap markets earning statistically significant monthly returns of 0.23%, 0.15%, 1.76% and 1.04%, respectively, in all four markets.

1 Introduction

Explanations for why some assets receive higher returns than others is the fundamental question in empirical asset pricing. From an equilibrium theory perspective the answer is clear — higher returns are demanded by investors for holding assets with higher levels of risk. While there are some markets (i.e. bond and option) and long-short characteristic sorted equity factors that conform to the equilibrium theory explanation (see e.g. the Fama-French size factor (1993), the Amihud and Mendelson illiquidity factor (1986), and the analyst uncertainty factor of Ackert and Athanassakos (1996)), there are others that seem to contradict this result. Fama and French (2015) find portfolios of firms with robust profitability have higher average returns than portfolios of firms with low profitability. Tuzel (2010) find portfolios of firms with large real estate holdings have higher average returns than firms with small real estate holdings. Additionally, Beneish, Lee and Nichols (2013) find firms with a lower probability of committing fraud earn higher returns in the cross-section.

Although proposed factors based on various indirect proxies for risk ultimately disagree on whether higher returns are compensation to investors for holding higher levels of risk, perhaps a more appropriate factor would be based on a more direct measure of risk —

*Graduate student, Department of Economics, Rutgers University. Email: jn495@economics.rutgers.edu.
Thanking message to add later.

the probability of firm distress. The asset pricing literature has introduced multiple factors based on direct measures of firm distress. Griffin and Lemmon (2002) and Dichev (1998) form portfolios sorted on Ohlson's O-Score (Ohlson, 1980). Dichev also forms portfolios sorted on Altman's Z-Score (Altman 1968). Vassalou and Xing (2004) form portfolios based on default risk as implied by Merton's (1974) option pricing model. Campbell, Hilscher and Szilagyi (2008) form portfolios based on a logit model estimated failure probability similar to Ohlson. Avramov et al. (2009) form portfolios based on credit downgrade events, and Asness, Frazzini and Pedersen (2019) form portfolios based on a measure of firm quality which includes elements such as profitability, growth and safety. Despite the different definitions of distress, the direction of the sort in these papers is always the same — firms with lower levels of distress (either lower bankruptcy or default probability, higher credit ratings or safer) earn higher expected returns. The exception to these results is Lockwood and Prombutr (2010) who find firms with lower probabilities of default earn lower average returns.

Several explanations have been introduced to motivate this finding. The first is a behavioral finance explanation relating to investors' perception of risk and aversion to downside risk (Harlow, 1991). When the dispersion of returns falls below a benchmark level, investors' perception of risk becomes too large and triggers a shift to less risky assets. The demand for assets with less risk drives returns to these safer assets. A related explanation is the well-known flight to safety[1]. While the flight to safety phenomenon is well-known to occur during crisis periods Asness, Frazzini and Pedersen (2019) present a dynamic model based on residual income which shows how flight to safety also reasonably occurs during normal market conditions. Like the model of Harlow (1991), the increased demand for safe assets by investors drives the higher realized returns.

In this paper I introduce a new factor based on the probability of firm bankruptcy. While previous studies have either used pre-existing bankruptcy risk models, or alternative methods of measuring firm distress, I elect to measure risk directly by estimating a new model of firm bankruptcy using contemporary risk factors. If investor flight to safety is occurring in equity markets, then a factor that is long firms with a low probability of bankruptcy and short firms with a high probability of bankruptcy should earn statistically significant returns. My results indicate that not only does this hold true for equity markets, but my modeled probability of bankruptcy holds predictive power for bond, option and credit default swap (CDS) markets as well. My long-short factor — which I title SSD_e (Safe Subtracting Distressed), SSD_b, SSD_o and SSD_c for the equity, bond, option and CDS markets — produces statistically significant monthly returns of 0.23%, 0.14%, 1.76% and 1.04% in each respective market. Returns to SSD_e are also higher during crisis periods, consistent with the flight to safety explanation of higher returns of less risky firms.

The equity and bond factors also produces statistically significant alpha when regressed on common existing factors. For the equity factor this includes the CAPM and Fama and French five-factor model. The equity factors also contributes to the explanation of the cross-section of returns when included in the factor zoo using the two pass lasso procedure of Feng,

[1] Also termed flight to quality.

Giglio and Xiu (2020). For the bond factor this includes regressions of SSD_b on the excess bond market, six common bond factors (excess bond market, two measures of illiquidity, term structure, default risk and momentum) and the Fama-French equity factors.

Notably, the other factors constructed of measures of firm distress are absent from current discussions of the factor zoo. No measures of distress, bankruptcy risk or default probability are used in papers comparing machine learning methods (see e.g. Gu, Kelly and Xiu, (2020)) and only the quality minus junk factor of Asness, Frazzini and Pedersen (2019) is included in papers introducing methods of inference using the factor zoo (see e.g. Harvey and Liu (2021) and Jensen, Kelly and Pedersen (2021)). However, quality minus junk uses definitions of quality other than distress risk.

A purpose in writing this paper is to bring bankruptcy risk into the discussion of the factor zoo, as this direct measure of the risks facing investors seems a blaring omission. In accomplishing this goal, the paper adds to the individual literatures on equity, bond and option factors, as well as the literature on anomaly factors which have explanatory power in multiple markets (see e.g. Asness et al. (2013) which includes bonds, currencies and commodity futures in addition to stocks; Chordia et al. (2017) which includes stocks and bonds; Frazzini and Pedersen (2014) which includes bonds and futures in addition to stocks; Moskowitz et al. (2012) which includes curencies, commodities and bonds in addition to stocks; and Brooks et al. (2018) which includes options in addition to stocks). To my knowledge this is the first paper to introduce a factor which has explanatory power in both equity and CDS markets.

The literature on modeling firm bankruptcy has proposed hundreds of predictors of bankruptcy. Undoubtedly the most well-known model developed using these predictors is the Altman Z-score (Altman, 1968) which is still used by both academics and practitioners today[2]. Many other models have since been introduced using both accounting based fundamental predictors (see e.g. Ohlson (1980), Altman (1977) and Taffler (1984)) and market based predictors (see e.g. Shumway (2001), Allayannis et al. (2003), and Hillegeist et al. (2004)).

A potential problem with using a previously developed bankruptcy model is that risk factors change over time. Despite the accuracy of the Z-score when it was first developed there is no reason to believe the risk factors facing firms in the 1940s, 50s and 60s (the sample period used to estimate the model) are the same risk factors facing firms today. Likewise, the choice of which potential predictors to choose for inclusion in a bankruptcy risk model can be subjective.

To overcome this problem I propose a two-step procedure to model the probability of bankruptcy. Using bankruptcies from 1990 to 2019, the first step utilizes the feature importance measure of random forests (Breiman, 2001) to choose which variables to include in a bankruptcy model from a large set of potential predictors. In the second step the selected

[2]Edward Altman recently partnered with the Kroll Bond Rating Agency applying his expertise, including use of the Z-score, to the analysis of corporate default risk. Additionally, the original Z-score paper has 20,418 citations on Google Scholar as of September, 2021.

predictors are used as inputs in a second random forest to estimate the probability of firm bankruptcy. A benefit of using this procedure is that the model can be easily updated to both include new predictors and ensure the most relevant predictors are inlcuded in the model at any point in time. The large set of bankruptcy predictors used in this paper contains both variables proposed by the literature and financial ratios which were obtained from the Wharton Research Data Services (WRDS) Financial Ratios Suite. I document the source and construction of the variables in Appendix B.

If my hypothesis of evolving risk factors is true, we should see more accurate bankruptcy prediction and increasing risk adjusted factor returns as we transition from a bankruptcy factor based on Altman's Z-score, to a factor using the same variables but with updated weights, and finally to a factor using all new predictors fit using the two step random forest procedure. We would also expect to see better factor performance later in the sample since the bankruptcy risk model was fit using data from 1990 to 2019 so the model is optimized to determine the risks facing modern firms, not the risks facing firms in the 1960s and 70s. I demonstrate that this does occur providing evidence for the hypothesis of an evolution of risk facing firms.

The remainder of this paper is organized as follows. Section 2 details the creation of the bankruptcy risk measure. Section 3 presents the data sources used to create the bankruptcy predictor variables and the equity, option and CDS factors. Section 4 presents the resulting performance of the equity, option and CDS factors and finally section 5 concludes.

2 Bankruptcy Risk Measure

This section details the two-step procedure which generates the bankruptcy risk measure used to create SSD_i (for $i = e, b, o, c$). There are two purposes in creating a new model of firm bankruptcy. The first is to find a set of predictors that is as parsimonious as previous bankruptcy models in the literature without having to take a stance, ex ante, about which existing predictors are most important. The second is that — insofar as the accuracy of the bankruptcy risk measure matters for the quality of the equity, bond, option and CDS factors — it is important to consider the most relevant risk factors facing firms today. Previous bankruptcy risk models were tuned to the risks facing firms when those papers were published and may not be reflective of the risks facing firms today. One of the benefits of this measure is that — unlike the Z-score and related measures — it is directly interpretable as the probability a firm will declare bankruptcy within 12 months.

Before detailing how the bankruptcy risk measure was developed, it is important to acknowledge that the legal process of filing for bankruptcy is not the only possible measure of firm distress. Failure to pay exchange listing fees, loan default, raising capital for the express purpose of continuing operations (Jones and Hensher, 2004) and reductions in dividends (DeAngelo and DeAngelo, 1990), among others, have also been used as measures of firm distress. However, many of these measures have some degree of subjectivity or could be the result of errors or strategic decisions unrelated to distress. For this reason I choose to measure firm distress as a firm filing for bankruptcy.

Perhaps the biggest difficulty for any model of corporate bankruptcy is that the formal filing of bankruptcy by a publicly traded firm is rare. Even for studies using pre-selected bankruptcy predictors the number of bankrupt firms in the sample can be very small. Altman (1968) uses a sample of only 33 bankrupt firms. Ohlson (1980) uses a sample of 105 bankrupt firms. Lennox (1999) uses a sample of 90 bankrupt firms. This problem is exacerbated here because to be included in the dataset each firm must have non-missing data for all variables necessary to create the full set of predictors instead of the much smaller number of variables needed in previous models. This reflects my competing goals of choosing — in a data driven way — from among the largest possible number of bankruptcy predictors introduced in the literature and including the maximum number of bankruptcies possible in order to obtain the most accurate model.

There are 457 bankruptcies of publicly traded firms listed in the CRSP/Compustat merged database between 1990 and 2019. After accounting for missing data, requiring multiple observations for each firm (necessary because some predictors require differences between adjacent observations or moving averages), and lagging the independent variables I was left with 71 bankrupt firms. Since the year bankruptcy is initiated occurs only once, this means there are 71 bankrupt firm-year observations. However, there are 64,645 firm-year observations for non-bankrupt firms making this a highly imbalanced classification problem. The 0.1% bankruptcy rate in my sample is of the same magnitude as the 0.76% bankruptcy rate found in Zmijewski (1984) and is reasonably less considering the overall decrease in the corporate bankruptcy rate since the 1980s[3]. The similarity between the level of class imbalance provides a level of validity to the composition of my dataset. The econometric and machine learning literatures both acknowledge the problems associated with highly imbalanced classes (see e.g. Fernàndez et al., (2018) and King and Zeng (2001)) although there is no consensus on the best way to resolve the problem.

Consistent with the use of machine learning for modeling bankruptcy, I choose to use machine learning to address the class imbalance as well by using K-nearest neighbors to match a bankrupt firm with a non-bankrupt firm based on size (i.e. market equity). I do this for two reasons. First, it is well known that smaller firms fail at a much higher rate than larger firms (see e.g. Ohlson (1980)). The more interesting question is can we discriminate between similarly sized firms in terms of which will eventually file for bankruptcy. Second, since the ultimate goal is the creation of a bankruptcy risk factor and it is common for equity factors to include a second sort on market equity this second sort will bring in the influence of firm size.

Using the K-nearest neighbors matched sample of 142 firm-years (71 bankrupt and 71 non-bankrupt) the two-step random forest procedure is implemented as follows.

1. Using the full list of bankruptcy predictors for the desired estimation period train a random forest and evaluate each feature's importance using the method of Breiman

[3]Source: Administrative Office of the U.S. Courts - accessed here `https://tradingeconomics.com/united-states/bankruptcies`

(2001).

2. Select the features with importance greater than a given level. Train a second random forest using only those selected features.

A benefit of using this procedure is that it provides a framework by which the bankruptcy risk model can be updated to both include newly proposed predictors and ensure the most relevant predictors are included in the model at any point in time.

For this paper the first forest is grown using 100 trees able to use 16 randomly selected variables at each node to partition the data following the $m_{try} = \sqrt{p}$ rule (Probst et al., 2019). Entropy is used as the criterion function at each split and trees are allowed to grow until there is only observation in each terminal node[4]. Five variables have importance greater than the desired threshold. These five inputs are displayed in table 1 ranked in order of importance.

Table 1: Bankruptcy Predictors Selected by Random Forest

Variable Name	Description	Construction
ROA	Return on assets	$\frac{\text{Income Before Extraordinary Items}}{\text{Total Assets}}$
PI_TA	Scaled pretax income	$\frac{\text{Pretax Income}}{\text{Total Assets}}$
ni	Unscaled net income	Net Income
NI_TA	Scaled net income	$\frac{\text{Net Income}}{\text{Total Assets}}$
CA_LT	Scaled current assets	$\frac{\text{Current Assets}}{\text{Total Liabilities}}$

The second random forest splits the data into a training set of 122 observations (61 bankrupt firms and 61 non-bankrupt firms) a test set of 20 observations (10 bankrupt firms and 10 non-bankrupt firms) to evaluate the model's performance. This forest is grown using 1,000 trees able to choose from only two of the five selected predictors from the first random forest at each node. I then compare the predictive accuracy of the random forest on the test set with the the original Altman Z-Score as well as a weighted least squares re-estimation of the coefficients used in Altman's Z-Score.

If the risk factors facing firms do, in fact, evolve over time as hypothesized we would expect to see two things. First, the coefficients on the inputs used by Altman (1968) should be different than the coefficients on a model estimated with my sample of bankrupt firms using the same inputs. Second, the random forest should select different predictors (which is does) and should be more accurate than both Altman's Z-Score and the re-weighted Z-Score. Table 1 has already shown the random forest selects alternative predictors than those used in Altman (1968). Equations 1 and 2 below compare the coefficients of Altman's Z-Score and the re-estimated Z-Score, whilel Table 2 compares model accuracy.

[4]Although they are not required to grow until there is only one observation in each terminal node. If multiple observations are close enough to each other the algorithm will stop tree growth.

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \tag{1}$$

$$Z_{WLS} = 0.159X_1 + 0.021X_2 - 0.861X_3 - 0.001X_4 + 0.179X_5 \tag{2}$$

where X_1 is working capital scaled by assets, X_2 is retained earnings scaled by total assets, X_3 is earnings before interest and taxes scaled by total assets, X_4 is market value of equity scaled by total liabilities and X_5 is sales scaled by total assets. While the WLS linear probability model approximates a probability and the Z-Score does not, the relative magnitudes and the signs of the coefficients are telling in how different the risk factors facing firms in the two sample periods are. Altman's Z-Score is interpreted such that a higher Z-Score implies a lower chance of bankruptcy (I use the word chance deliberately because the Z-Score is not interpretable as related to a likelihood function, nor as a probability of bankruptcy). Therefore all of the inputs decrease the chance of bankruptcy. The equivalent interpretation for Z_{WLS} would be negative coefficients for all variables. Clearly this does not occur implying missing risk factors and omitted variable bias.

Table 2: Bankruptcy Model Predictive Accuracy

Panel (A)			
Model	**Accuracy**	**FPR**	**FNR**
Original Altman	55.0%	40.0%	50.0%
	(54.8%, 55.2%)	(39.7%, 40.2%)	(48.8%, 50.2%)
Weighted Least Squares	60.0%	50.0%	30.0%
	(60.0%, 60.0%)	(49.2%, 50.8%)	(29.2%, 30.8%)
Random Forest	80%	20%	20%
	(80.0%, 80.1%)	(18.9%, 21.1%)	(18.9%, 21.1%)
Panel (B)			
Model	**Accuracy**	**FPR**	**FNR**
Original Altman	58.8%	40.2%	42.3%
	(58.6%, 59.0%)	(40.0%, 40.4%)	(42.0%, 42.6%)
Weighted Least Squares	55.6%	24.8%	67.3%
	(55.4%, 55.8%)	(24.3%, 25.3%)	(66.8%, 67.8%)
Random Forest	68.7%	32.5%	29.9%
	(68.6%, 68.8%)	(32.2%, 32.8%)	(29.6%, 30.2%)

Note: This table presents statistics related to the accuracy of the bankruptcy prediction models. Panel (A) uses a test set of 20 observations, Panel (B) uses a test set of 313 observations. Accuracy is the total percentage of test set observations classified correctly. FPR is the false positive rate — the percentage of non-bankrupt firms classified as bankrupt. FNR is the false negative rate — the percentage of bankrupt firms classified as non-bankrupt. The bootstrapped 95% confidence interval is in parentheses.

The results in Panel A of Table 2 — which displays each model's predictive accuracy using the 20 observation test set held out when training the second random forest — are also indicative of evolving risk factors. When applied to a broad range of modern firms the Z-score predicts bankruptcy only marginally better than a naive guess. If only the coefficients are updated with weighted least squares (weighting by market capitalization) using the same sample of firms as the random forest the model's predictive accuracy increases consistent with increased relevance for contemporary risk factors. Further improvements in predictive accuracy occur when new predictors are selected by the random forest, again consistent with the notion of evolving firm risk.

In this relatively small test sample the random forest bankruptcy model correctly predicts 80% of both bankruptcy and non-bankrupt firms. This is an impressive feat considering market capitalization — a well-known predictor of bankruptcy risk — is not available to help discriminate between high and low bankruptcy risk firms. Bootstrapped 95% confidence intervals are in parentheses.

Because of the number of variables necessary to construct the full set of predictors used in this study and the requirement that there be no missing data, there were not many observations available to test these three models out-of-sample. As a further robustness test for the ability of each model to predict corporate bankruptcy, Panel B of Table 2 displays each model's predictive accuracy using an expanded test set of 313 firms — 144 bankrupt firms and 169 non-bankrupt firms. This expanded test set was constructed in the same way as the original 144 observation dataset, but instead of deleting missing observations for all of the variables necessary to construct the full set of predictors used in the first random forest, only missing observations for the variables used to construct the predictors in Table 1 are deleted.

It is to be expected that the predictive accuracy of the random forest decreases somewhat in this extended sample. The firms which comprised the original dataset were the largest of the bankrupt firms since smaller firms are more likely to have missing data leading to their removal. As observations were re-added to form the test set used in Panel B, a higher degree of heterogeneity entered the test set as smaller firms were added back. Despite this added heterogeneity the random forest prediction model performs well, correctly classifying nearly 70% of firms, well above Altman's Z-Score and the WLS models. Since most publicly traded firms are larger than the sample used to fit the random forest, the decreased accuracy with respect to the smallest firms is not a concern when using the level of bankruptcy risk estimated using the random forest to form portfolios.

The power of this measure of bankruptcy probability comes from the fact that despite the relatively small sample size used to train the random forest, it can be used to make predictions for almost all publicly traded companies. This is because the five predictors selected all use common balance sheet items which are regularly reported by firms with listed stocks. Contrast this with a predictor such as the R&D intensity measure of Franzen et al. (2007) which requires a research and development variable available for only 32.6% of firm-year observations in Compustat.

3 Data

This section documents the data used and the locations where the data were obtained. All data were obtained from Wharton Research Data Services (WRDS) and the Bloomberg Terminal.

The full set of bankruptcy predictor variables requires data from Compustat Fundamentals (table COMP.FUNDA) for accounting characteristics, Compustat Names (table COMP.NAMES) for industry codes, Compustat Segments (table COMP.SEG_TYPE) for primary business segment, CRSP (table CRSP.MSF) for price and issuance data, and WRDS financial ratio suite (table WRDSAPPS.FINRATIOFIRM). Compustat, CRSP and the WRDS financial ratio suite were linked using the CRSP-Compustat linking table (CRSP.CCMXPF_LINKTABLE). Of the 241 total predictors, 221 were proposed by the literature while I augment these with 20 ratios from the WRDS financial ratio suite which has observations for all firms in the data set. The formation of all predictors proposed by the literature is detailed in Appendix B.

The bankruptcy risk equity factor is generated using monthly returns data from CRSP from July 1962 to December 2019. Only stock with share codes 10 or 11 and exchange codes 1, 2 or 3 are included in the equity factor formation. That is, only common stock of U.S. based companies which trades on the NYSE, NASDAQ or American stock exchanges are included.

The bankruptcy risk bond factor is generated using data from the Financial Industry Regulatory Authority's (FINRA) Enhanced Trade Reporting and Compliance Engine (TRACE). The enhanced TRACE database contains intraday trade by trade data for corporate bonds in the United States. The enhanced TRACE differs from the standard TRACE only by not truncating the volume of trades at \$1 million for high yield bonds and \$5 million for investment grade bonds. For the period July 2002 to September 2020 dirty bond returns (i.e. including accrued interest and coupon payments) are obtained, already calculated, from WRDS. These returns are generated after applying the filtering procedure of Asquith, Covert and Pathak (2019)[5] and Dick-Nielsen (2009, 2014). Specifically, the following trades are removed: cancelled and updated trades, double-counted dealer trades, trades involving bonds with variable rate coupons, trades of bonds issued by firms not covered by rule 144a, and trades of bonds other than corporate bonds. Returns are winsorized at the 1% level to mitigate the effect of any data errors. These filters remove approximately 32% of trades from TRACE.

Default returns (the bond equivalent of delisting returns in CRSP) are generated following Cici, Gibson and Moussawi (2017). Specifically, investment grade bonds are given a return of -17.67% the month of default and high yield bonds are given a return of -40.17% the month of default. That default returns are not -100% is reflective that defaulted bonds are still tradeable and holders of defaulted bonds often received a strictly positive percent of the principal back. I extend this return series to December 2020 using enhanced TRACE

[5]Revised in 2019, originally submitted in 2013.

trade data and the same filtering procedure used by WRDS.

The bankruptcy risk options factor is generated using data from OptionMetrics from January 1996 to December 2017. This is the entire time series of options available through the WRDS Option Suite. Given the expectation of the underlying equity is increasing returns in riskiness, only call options are used since increasing prices of the underlying asset would imply negative return for put options. Additionally, shorting call options serves a similar purpose as put options. Monthly returns are formed using open interest weighted prices across strike prices from the last trading day of each month.

It is well-known that options data contains observations errors at a rate much larger than equity data (Todorov (2019), Andersen et al. (2021)). Common causes of these errors include the true option value lying within the bid-ask spread (minimum tick sizes for options are usually 5 to 10 times the size of stock minimum tick sizes), liquidity provisions offered by market makers and shifts in option positions which effect only a small range of strike prices. To mitigate the impact of these observation errors I apply a number of filters to the data before calculating monthly returns. The filters include a bid-ask spread larger than the bid-ask midpoint (Gharghori et al., 2017), zero open interest (Goyal and Saretto, 2009), option price less than \$0.10 and winsorizing all variables at the 0.1% level (Muravyev and Pearson, 2020). These filters remove 2.67 million observations, or 3.4% of total observations. To further mitigate any error in reported options pricing I winsorize the generated monthly returns at the 0.1% level as well. This has the added benefit of ensuring my results are not driven by a single outlier return.

Unlike the equity and bond data, delisting or default returns are not a concern for options. In my sample, the options of companies that delist stock for any reason reach a price of zero (or a value that is not different from zero by any economically or statistically relevant amount) well-before the delisting event. This finding is echoed in the literature as no paper I am aware of attempts to add delisting returns to options series.

The bankruptcy risk credit default swap (CDS) factor is generated using data from Bloomberg from October 2001 through December 2020 — the entire period for which CDS prices are available from Bloomberg and accounting data is available from Compustat. CDS contracts exist for a much smaller universe of companies compared to equity and option securities and are generally less liquid. I therefore limit CDS data to firms belonging to the S&P 500 index as of May 2021 to ensure sufficient CDS contracts trade each month to generate portfolios. Following Meine et al. (2016) I use single name CDS contracts with a 5-year term structure because they are the most liquid. Despite being the most frequently traded, only 237 of the 500 firms belonging to the S&P 500 have actively traded credit default swaps.

To protect against possible errors in CDS prices I also impose the following filters before calculating monthly CDS returns which — like option returns — are calculated as the percent change in price on the last trading day of each month. First I drop all prices above 10,000. CDS prices are denominated in basis points of the debt position the CDS is insuring. Therefore a price above 10,000 would imply it costs more to insure the debt than the debt is worth. I then winsorize price at the 0.1% level before generating monthly returns. Lastly I winsorize the return series at the 0.1% level for the same reasons as listed for the options data.

4 Bankruptcy Risk Factor

This section presents the returns generated by the bankruptcy risk factor for the equity, bond, option and CDS markets. For each of the three factors I present evidence of the factors economic benefits both in terms of average monthly returns, cumulative returns and Sharpe ratios.

4.1 Equity Factor

To motivate the formation of the bankruptcy risk factor, I first present decile portfolio sorts on bankruptcy risk as measured by Altman's Z-score, the WLS re-estimation of Altman's Z-score and the random forest probability of bankruptcy. Below the portfolio returns are the associated t-statistic as well as the mean probability of bankruptcy of firms within that portfolio.

Table 3: Equity Decile Mean Returns and Bankruptcy Risk

Sort	Distress (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Safe (10)	(10)-(1)
AltZ 1962	1.13%	1.02%	0.89%	0.93%	0.99%	1.02%	1.01%	0.90%	0.94%	0.90%	-0.23%
t	(5.67)	(5.99)	(5.26)	(5.12)	(5.57)	(5.79)	(5.68)	(5.15)	(5.41)	(4.28)	(1.39)
$\bar{Z}$	2.11	4.12	5.46	6.87	8.56	10.76	13.82	18.64	28.43	169.64	
AltZ 1980	1.12%	1.14%	1.02%	1.01%	1.11%	1.11%	1.10%	0.99%	1.10%	1.03%	-0.09%
t	(4.93)	(5.94)	(5.27)	(4.82)	(5.33)	(5.52)	(5.38)	(4.84)	(5.56)	(4.06)	(0.44)
$\bar{Z}$	2.07	4.20	5.65	7.16	8.23	11.26	14.46	19.49	29.36	185.79	
WLS 1962	0.86%	0.84%	0.96%	0.90%	0.94%	0.97%	0.98%	1.03%	1.07%	1.10%	0.24%
t	(5.25)	(5.09)	(5.58)	(4.90)	(4.81)	(4.89)	(4.75)	(4.90)	(4.99)	(5.62)	(1.41)
$\bar{Z}_{WLS}$	0.58	0.30	0.23	0.19	0.16	0.13	0.10	0.06	0.02	-0.09	
WLS 1980	1.01%	0.91%	1.09%	1.00%	1.10%	1.10%	1.09%	1.23%	1.22%	1.18%	0.17%
t	(5.01)	(4.76)	(5.13)	(4.53)	(4.74)	(4.71)	(4.38)	(4.82)	(4.76)	(5.22)	(1.31)
$\bar{Z}_{WLS}$	0.59	0.30	0.23	0.19	0.15	0.12	0.09	0.06	0.02	-0.09	
RF 1962	0.97%	0.86%	0.86%	0.95%	1.10%	1.03%	0.86%	0.88%	1.00%	1.08%	0.11%
t	(4.05)	(4.19)	(4.37)	(5.50)	(6.33)	(7.71)	(5.17)	(5.53)	(5.92)	(5.28)	(0.78)
$\bar{Z}_{RF}$	0.79	0.58	0.39	0.23	0.13	0.07	0.04	0.03	0.02	0.01	
RF 1980	1.02%	0.88%	0.90%	0.94%	1.12%	1.12%	1.00%	0.96%	1.04%	1.20%	0.18%
t	(3.60)	(3.45)	(3.72)	(4.61)	(5.47)	(5.91)	(4.94)	(4.99)	(5.26)	(4.64)	(1.12)
$\bar{Z}_{RF}$	0.83	0.64	0.43	0.26	0.14	0.08	0.04	0.03	0.02	0.01	

Note: This table presents mean returns for various portfolios formed by sorting firms based on the original Altman Z-score (AltZ), a weighted least squares re-estimation of the weights used to form the Z-score (WLS) and a random forest (RF). Rows 1962 calculate values using the sample from 1962 to 2019. Rows 1980 calculate values using the sample from 1980 to 2019. t-statistics are in parentheses. $\bar{Z}$ represents the mean of the bankruptcy risk measure used in the respective row for each portfolio. Columns Distress (1) to (10)-(1) show mean returns for single sorts on bankruptcy risk.

Looking first at returns, the pattern displayed in the extreme decile long-short portfolio (column (10)-(1)) is as we would expect given my assertion regarding the evolution of the bankruptcy risk measure over time. The portfolios sorted by Altman's Z-score exhibit a much stronger pattern when the sample begins in 1962 compared to when the sample begins in 1980 demonstrating the Z-score was more relevant for predicting returns in the 1960s and 70s than it is today. Similarly, the WLS re-estimation of the weights associated with the five accounting ratios used by Altman (1968) lead to better defined portfolio sorts, but the relationship is still stronger in the 1960s and 70s than it is today. This indicates both the weights and inputs used to construct the Z-score are obsolete in the context of predicting equity returns.

On the other hand, sorts based on the random forest produced probability of bankruptcy lead to better discrimination between high and low bankruptcy risk firms, providing evidence for the benefits of using contemporary risk factors to measure bankruptcy probability.

Despite the better defined division between high and low bankruptcy risk firms produced by the random forest, the extreme decile long-short portfolio is not statistically significant for any of the bankruptcy risk models. This is consistent with Dichev (1998) who shows using both the Altman Z and Ohlson O measure of bankruptcy risk that portfolios sorted on bankruptcy risk do not necessarily form monotonic return patterns. There are two possible explanations for this. The first is that, despite the empirical documentation that firms with lower levels of financial distress tend to have higher average returns (see e.g. Dichev (1998), Vassalou and Xing (2004), and Asness et al. (2019)), the equilibrium theory asset pricing result holds that higher returns are compensation for holding extra levels of risk. Familiar with this theoretical result, some investors may be targeting firms with higher levels of bankruptcy risk driving returns for the riskiest decile.

The second is related to the mean bankruptcy probability in each decile portfolio. Table 3 shows the bankruptcy probability for the five highest deciles are largely similar — even the mean bankruptcy probability for decile 5 is only six percentage points higher than decile 6. It is not surprising then that the returns for these decile portfolios are so similar given the similarity in bankruptcy risk. However, the fact the highest return decile is always among the three highest deciles does provide further evidence in support of investor flight to safety.

The equity bankruptcy risk factor is formed by sorting firms independently by size and my random forest estimated probability of bankruptcy. Sorts are done in June using accounting and market equity information released in December of the previous year. Thus the assumption is made that firm risk information has been fully disseminated and absorbed by investors within six months of its release. Motivated by the observation that bankruptcy risk is very similar among deciles 6 through 10, and bankruptcy risk is very high in decile 1, the bankruptcy risk factor is formed by going long an equal weighted average of the five safest deciles within both the large and small size groups, and going short decile two within the large and small size groups. Shorting decile two instead of decile one avoids the firms with the most uncertain behavior — decile one consistently has the lowest standard errors among the decile portfolios — removing unnecessary risk from the factor. Dichev (1998) also forms portfolios which have unequal numbers of portfolios in the long and short legs.

Dichev (1998) also demonstrates that distress risk and unrelated to the size anomaly. If this is still true, then in my portfolio sorts small firms should still see higher returns than large firms, and safe firms should see higher returns than distressed firms. This means safe small firms should have the highest returns, while large distressed firms should have the smallest. Table 4 confirms these return patterns and presents the average returns and maximum drawdowns of the bankruptcy risk factor termed SSD_e for "safe subtracting distressed"[6].

Table 4: Bankruptcy Risk Equity Factor: 1980 - 2019

			Original Altman			
	High (2)	Med (3-5)	Low (6-10)	SSD_e	t(SSD_e)	Max Drawdown
Small	1.23%	1.20%	1.08%	0.10%	0.73	55.0%
Large	0.98%	0.90%	1.09%			
			WLS Re-estimation			
	High (2)	Med (3-5)	Low (6-10)	SSD_e	t(SSD_e)	Max Drawdown
Small	1.00%	1.03%	1.19%	0.18%	1.73	31.9%
Large	0.89%	0.92%	0.96%			
			Random Forest			
	High (2)	Med (3-5)	Low (6-10)	SSD_e	t(SSD_e)	Max Drawdown
Small	1.02%	1.13%	1.23%	0.23%	2.29	29.4%
Large	0.87%	0.88%	0.96%			

Note: This table presents mean returns for the components of the equity bankruptcy risk factor SSD_e as well as the factor itself. The numbers in parentheses in the columns represent the included deciles from Table 3. Max Drawdown represents the maximum peak to trough decrease from 1980 to 2019 and does not represent any specific period length.

Consistent with Table 3, the sampe pattern emerges as we transition from the factor constructed on Altman's Z-score to the factor constructed using the random forest measure of the probability of bankruptcy. Raw and risk adjusted returns are highest and maximum drawdowns are lowest for the random forest model. This provides yet further evidence supporting my assertion that it is important to not only update the weights associated with a bankruptcy prediction model, but also to select contemporary risk factors as the risks facing firms evolve over time.

Although the average monthly returns of SSD_e are smaller than some indexes such as the S&P 500 index, they are of similar size to other factors in the literature. Alwathainani (2009) documents a monthly return of 0.21% for his factor sorted on earnings consistency which is validated by Chen and Zimmermann (*forthcoming*). Hirschleifer et al. (2013) documents a monthly return of 0.26% for their factor based on patents scaled by R&D expenditures, although Chen and Zimmerman find only 0.21% returns on the same factor. Hou and Robinson (2006) document monthly returns of 0.26% for their factor sorted on industry concentration, although Chen and Zimmermann find only 0.21% returns on the same factor.

[6]A better name may be "safe minus distressed", but the acronym — SMD — is too similar to the Fama-French size factor SMB so I avoid it to avoid confusion.

Additionally, the information contained in SSD_e is independent of existing factors. SSD_e has positive alpha — which can be interpreted as risk-adjusted returns since the influence of other types of risk have been accounted for by the right-hand-side factors — when regressed on the market or the Fama-French five-factors. Although it is commonplace (as it should be) to test new factors against influential asset pricing models such as the Fama-French five factor model, it is still important to test new factors against the CAPM by itself because alphas can move from insignificant to significant as more factors are added (Jensen, Kelly and Pedersen, 2021). A successful new factors should achieve a significant alpha against both.

Table 5: SSD_e Risk-Adjusted Monthly Returns

	CAPM	FF5	Zoo
SSD_e	0.25%	0.23%	5.34
t-statistic	(3.39)***	(2.60)***	(5.93)***

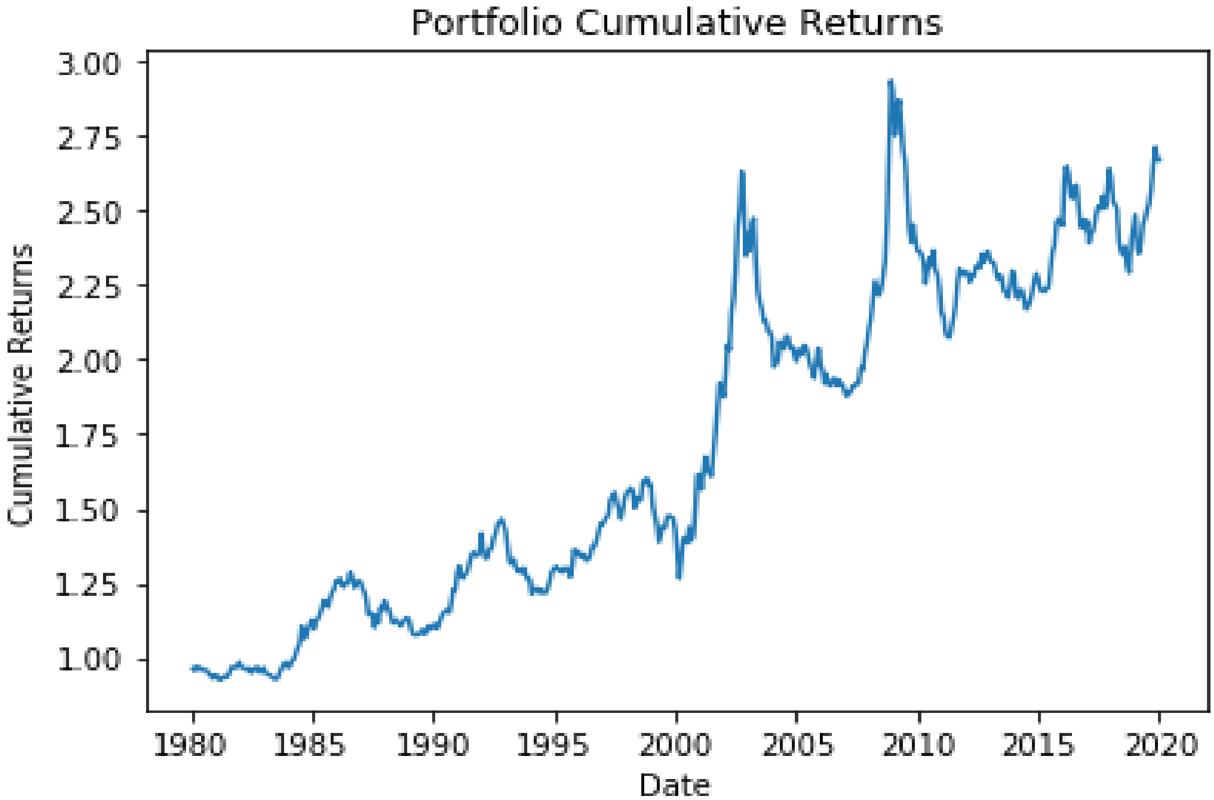
Note: This table presents the statistical significance of SSD_e when regressed on other popular factors. For the columns CAPM (regression on the capital asset pricing model) and FF5 (regression on the five Fama and French factors) the table reports the regression alphas and corresponding t-statistics. The column Zoo reports the coefficient and t-statistic on SSD_e when it is included in cross-sectional regressions along with 15 other factors chosen using the two-step method proposed by Feng, Giglio and Xiu (2020). *** indicates statistical significance at the 1% level.

The coefficient on SSD_e is also statistically significant in cross-sectional regressions when it is included along with 15 other existing factors selected using the two-step procedure developed by Feng, Giglio and Xiu (2020). The procedure of Feng, Giglio and Xiu (2020) was developed as a way to provide evidence of the incremental explanatory ability of a proposed factor for the cross-section of returns compared to the myriad existing factors in the literature — called the "factor zoo". I use the 135 factors in Chen and Zimmermann (*forthcoming*) which have observations from 1964 to 2019 as a proxy for the factor zoo. The 15 factors selected by this procedure of summarized in Appendix C. The SSD_e regressions are summarized in Table 5.

The risk-adjusted returns of SSD_e remain fairly close to their raw value of 0.23% per month when regressed on the CAPM and Fama-French five-factor models. The alpha when regressed on the CAPM is 0.25% per month and 0.23% per month when regressed on the Fama-French five-factor model. The coefficient 5.34 for the Zoo column has a related, but different interpretation. It is the stochastic discount factor loading on SSD_e controlling for the other 15 selected factors when explaining the cross-section of returns. All three coefficients are highly statistically significant.

Figure 1 shows an alternative presentation of the information in Table 4 — the cumulative returns to $1 invested in SSD_e in 1980. Despite the relatively slow growth of SSD_e over time (0.23% per month), there are some desirable features of this series. The maximum drawdown of SSD_e is 29.4%. Conversely, the maximum drawdown of the S&P 500 index, which occured during the Great Recession in 2007 to 2009, was 59%. Likewise, the maximum drawdown of a momentum trading strategy — one of the existing factors with the highest average monthly returns — is 58% and occurs during this same time period. This makes SSD_e in its normal form particularly relevant for leverage investors as these investors could apply more than three-times leverage in SSD_e whereas even two-times leverage would completely kill an investors portfolio if investing in the S&P 500 index or momentum.

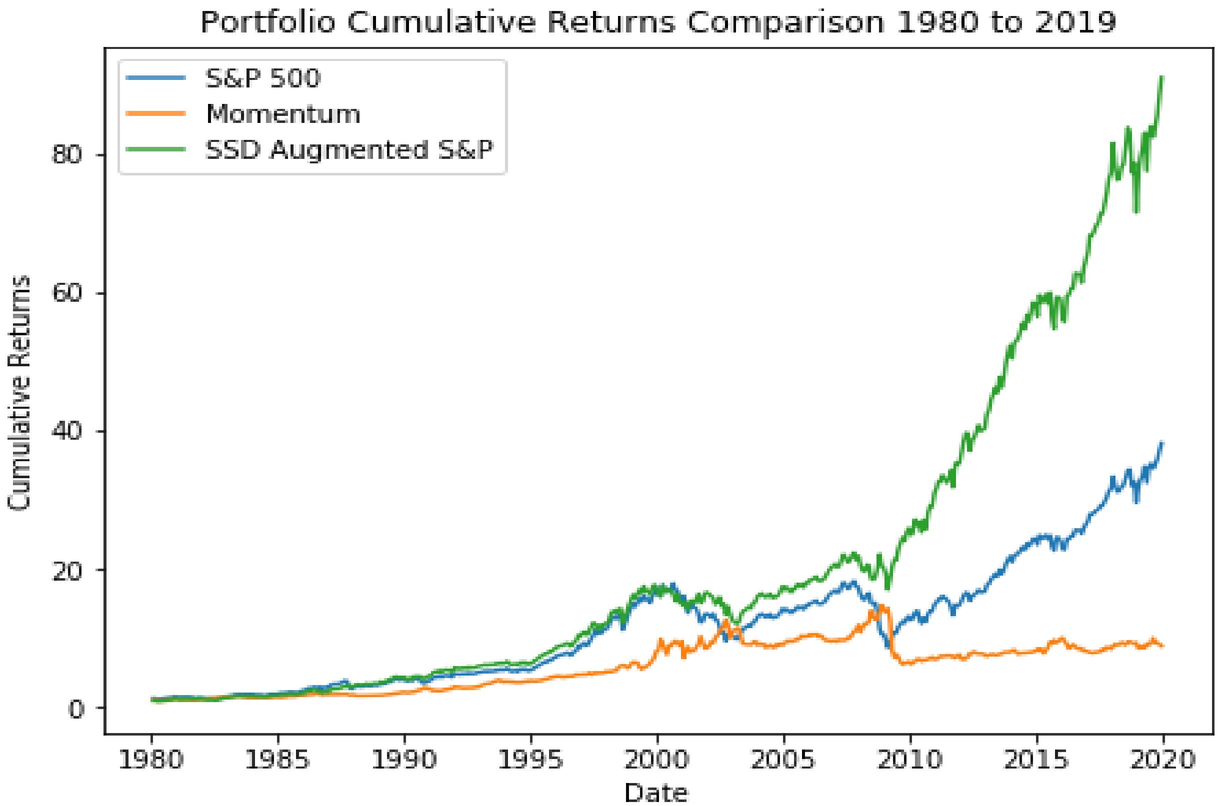
Figure 1: Cumulative Returns to SSD_e Investment Strategy: 1980 - 2019



However, there are also other profitable ways to incorporate SSD_e in an investment strategy. As can be seen in Figure 1 the two periods of sustained high returns overlap with the "dot com" bubble in 2000 to 2002, and the Great Recession in 2007-2009. This is consistent with the idea of investor flight to safety driving up returns for the safest stocks while the market overall declines. These safe stocks are those that make up the long leg of SSD_e. Using this idea of flight to safety once investors observe large negative market returns, a strategy which includes buy and hold on the S&P 500 index until a return less than -5% is observed, then the month after this observation invest in SSD_e until index returns rise above -5% results in a average monthly return of 1.02% — higher than the raw buy and hold returns of 0.80% per month. Figure 2 displays the benefits to this hybrid investment strategy.

It is clear from Figure 2 the benefits of this hybrid strategy come from significantly shortened period of large negative returns. By moving investments into SSD_e after observing large decreases in market returns, investors both avoid continued negative buy and hold returns, and are fully invested in the safe stocks before other investors drive up prices. This strategy results in an annualized Sharpe ratio of 0.90, higher than the annualized Sharpe ratios of the buy and hold return (0.65) and momentum strategy (0.44) during the same period.

Figure 2: Cumulative Returns to SSD_e Augmented Buy and Hold Return: 1980 - 2019



4.2 Bond Factor

As with the equity factor, I motivate the formation of the bankruptcy risk factor by first presenting decile portfolios of bonds sorted on the random forest produced probability of bankruptcy. Since bond trade data is only available through TRACE beginning in 2002 I cannot compare portfolios formed in the 1960s and 1980s like I did for the equity decile portfolios. Additionally, the relative return distribution for the portfolios sorted by Altman'z Z-score, the weighted least squares re-estimation of the Z-score coefficients and the random forest produced bankruptcy probability are very similar to the pattern observed for the equity factor. I therefore only present results for the portfolios sorted on the random forest produced probability of bankruptcy. These results are displayed in Table 6.

Unlike equity returns, bond returns are not driven by changes in the clean price of bond, but rather by the accrued interest and coupon payments. It therefore makes sense that we do not observe the same high returns to bonds with lower bankruptcy risk driven by an investor flight to safety. Bonds issued by companies with higher default risk - which is related to my measure of bankruptcy probability - must generally pay higher interest rates on their bonds leading to higher returns to bonds with high bankruptcy (default) risk. However, Table 6 still highlights the increased uncertainty surrounding returns to the riskiest decile of bonds in the form of the lowest t-statistic for the most distressed decile despite the highest mean returns.

Table 6: Bond Decile Mean Returns and Bankruptcy Risk

Sort	Distress (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Safe (10)	(1)-(10)
RF 2002	0.75%	0.69%	0.73%	0.65%	0.61%	0.58%	0.57%	0.56%	0.58%	0.58%	0.17%
t	(4.35)	(4.70)	(5.62)	(5.19)	(5.48)	(5.08)	(5.06)	(5.11)	(5.42)	(5.33)	(1.50)
$\bar{Z}_{RF}$	0.79	0.56	0.39	0.26	0.17	0.11	0.06	0.04	0.02	0.01	

Note: This table presents mean returns for various portfolios formed by sorting firms based on my random forest produced probability of bankruptcy. The row RF 2002 presents mean returns for each decile. t-statistics are in parentheses. $\bar{Z}$ represents the mean of the bankruptcy risk measure used for each portfolio.

The pattern of returns across the cross-section of bonds is very similar to the cross-section of equities — of course with returns increasing in the opposite direction. Specifically, while returns are not monotonically decreasing across deciles, there is a clear pattern of decreasing returns until the probability of bankruptcy reaches approximately 10%. After that point there is materially no difference in returns among the remainder of the deciles resulting in a difference in extreme decile returns which is not statistically significant. There is one more element of note in Table 6 before I motivate the bankruptcy risk bond factor. The three deciles with the highest bankruptcy risk have lower mean levels of bankruptcy risk for firms which have issued bonds compared to firms with common stock. This makes sense since these firms would have to pay the highest interest rates on bond issues, raising capital via the stock market may be a cheaper alternative.

The bond bankruptcy risk factor is formed in the same way as the equity factor. That is, in June firms are sorted independently by size and my random forest estimated probability of bankruptcy. Motivated by the information in Table 6, the bankruptcy risk factor is formed by going long an equal weighted average of the three riskiest deciles within both the large and small size groups, and going short the five safest deciles within the large and small size groups. The difference between the formation of the bond factor and the equity factor — that the three riskiest deciles are included in the bond factor instead of just decile 2 — reflects the fact that bonds have both built-in compensation for extra risk in higher interest rates and the lower probability of bankruptcy compared to the equity portfolios for the three riskiest deciles.

Table 7: Bankruptcy Risk Bond Factor: 2002 - 2020

	High (1-3)	Med (4-5)	Low (6-10)	SSD_e	t(SSD_e)	Max Drawdown	SR
Small	0.74%	0.65%	0.59%	0.14%	2.77	6.3%	0.69
Large	0.69%	0.56%	0.55%				

Note: This table presents mean returns for the components of the bond bankruptcy risk factor SSD_b as well as the factor itself. The numbers in parentheses in the columns represent the included deciles from Table 6. Max Drawdown represents the maximum peak to trough decrease from 2002 to 2020 and does not represent any specific period length.

Similarly to the equity factor, Crawford et al. (2019) demonstrates that size is a priced risk factor in the bond market where smaller firms — being riskier — require higher interest rates and therefore have higher returns. I should therefore find that smaller firms have higher returns than larger firms for a given level of bankruptcy risk, and within size groups higher bankruptcy risk results in higher returns. Table 7 shows this is indeed the case. The result is the bond factor, termed SSD_b (where the b is for "bond") earns a statistically significant 0.14% return per month.

Table 8: SSD_b Risk-Adjusted Monthly Returns

	Market	Bond6	FF6	FF-Bond12
SSD_b	0.11%	0.08%	0.08%	0.08%
t-statistic	(2.27)**	(2.08)**	(2.08)**	(2.16)**

Note: This table presents the statistical significance of SSD_b when regressed on other popular factors. *Market* indicates a regression on the aggregate bond market return. *Bond6* indicates a regression on six standard bond market factors. *FF6* indicates a regression on the Fama-French six equity factors (including momentum). *FF-Bond12* indicates a regression on all 12 previous used equity and bond factors. ** indicates statistical significance at the 5% level.

Like SSD_e this is not an enormous monthly return. However, it is independent of existing bond and equity factors in the literature. Following Bai, Bali and Wen (2019), I regress SSD_b on both equity and bond factors to obtain the alpha, or risk-adjusted returns. Controlling for the aggregate bond market return in excess of the risk-free rate SSD_b earns a positive alpha of 0.11% per month. This decreases to 0.08% risk-adjusted monthly return when five other standard bond factors are added. These factors are bond momentum (Jostova et al., 2013), two measure of liquidity (Bao, Pan and Wang, 2011; and Bai, Bali and Wen, 2019)[7], and default and term factors (Elton et al., 1995).

SSD_b maintains a risk-adjusted return of 0.08% per month when regressed on the Fama-French six factors and when both equity and bond factors are included in the regressions. That SSD_b maintains a consistent risk-adjusted returns and statistical significance when moving from a univariate regression on the excess bond market to a 12 factor model provides strong evidence for the independence of the information contained in my bond factor.

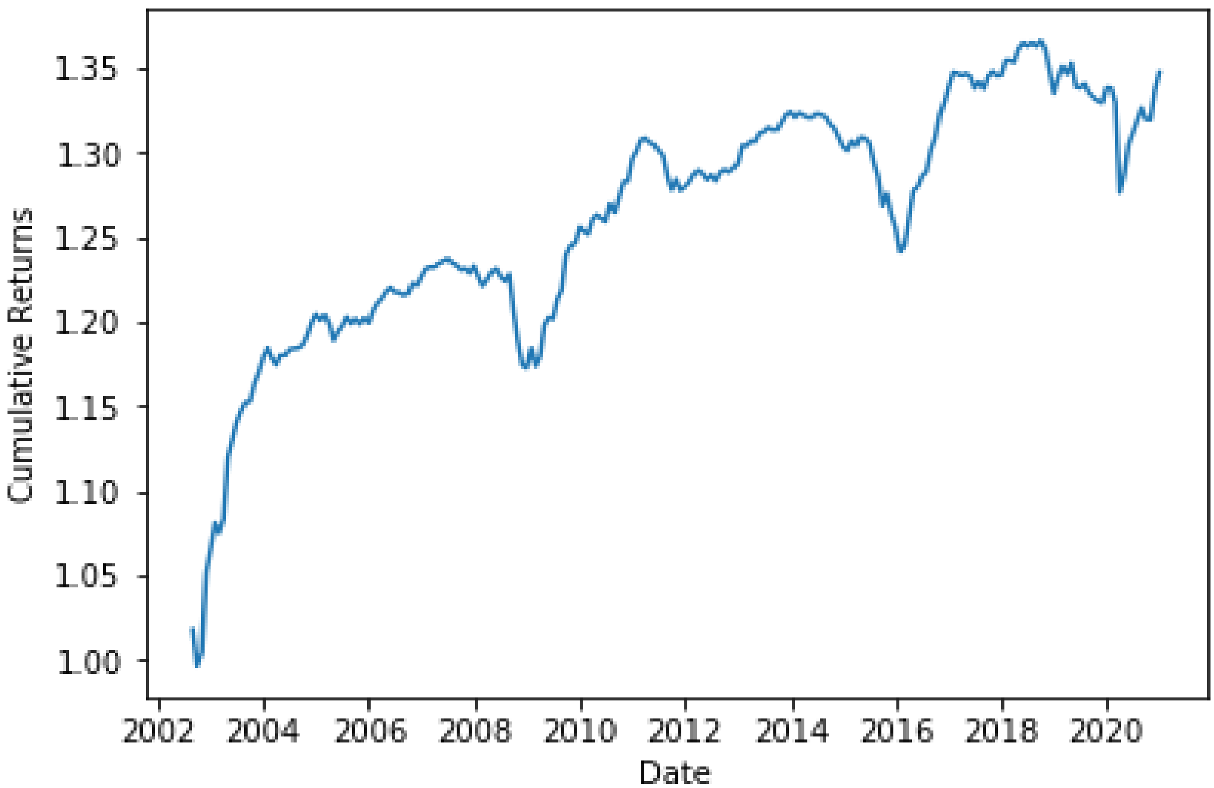
Figure 3 provides an alternative presentation of the information in Table 7. The source of the dips in 2009 and 2020 are obvious — the Great Recession and the onset of the Covid-19 pandemic. The source of the dip in 2016-2017 is less obvious, although all bond factors show this dip — it is not unique to SSD_b.

While the primary benefits of SSD_e are from its role as a leading indicator for upswings in the equity market, the major benefits of SSD_b come from the small maximum drawdowns during periods of systemic distress. These small drawdowns — the largest of which is only 6.3% compared to 11.1% for the bond market and 20.9% for downside risk — and low

[7] I include two liquidity factors because Bai, Bali and Wen (2019) include one liquidity factor in their spanning regressions but they also introduce a new liquidity factor. Since neither liquidity factor subsumes the other in a univariate spanning regression I chose to include both.

levels of overall volatility are such that SSD_b has a Sharpe ratio of 0.69 despite averaging only 0.14% monthly returns. Motivated by the relatively small decreases during periods of systemic distress, a similar strategy to the one proposed in section 4.1 is proposed here.

Figure 3: Cumulative Returns to SSD_b Investment Strategy: 2002 - 2020

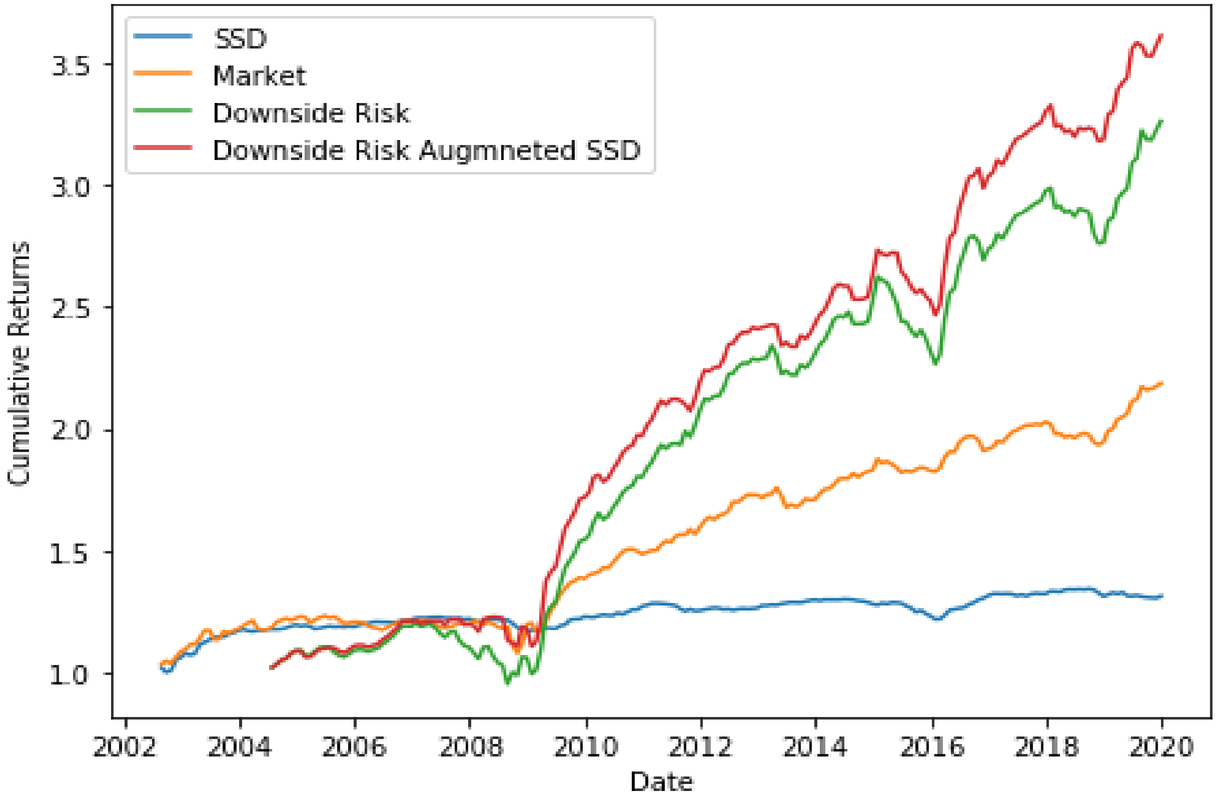


Specifically, beginning with the downside risk factor (DRF) of Bai, Bali and Wen (2019), hold this factor portfolio until observing an end of month decrease of -1%. The subsequent month invest in SSD_b until returns to DRF rise above -1%, then return the investment to DRF the next month. Like the equity hybrid strategy took advantage of SSD_e as a leading indicator, this strategy takes advantage of SSD_b as a safety net during periods of distress as the bottom of the downturn is higher than that of other strategies. The benefits of this strategy can be seen graphically in Figure 4 and in the associated Sharpe ratio which is 1.20 for this bond hybrid strategy, compared to 1.02 for DRF and 0.98 for the aggregate bond market.

4.3 Options Factor

In the same way I motivated the formation of the equity and bond factors by first looking at decile portfolios sorts, I motivate the options bankruptcy risk factor by examining the average returns and bankruptcy risk of the cross-section of options. It is important to redo this exercise for the cross-section of options because there are fewer firms with actively traded options than there are with actively traded stocks (there are 21,803 firms with returns data

Figure 4: Cumulative Returns to SSD_e Augmented Investment: 2002 - 2020



appearing in CRSP at least one month during the period 1996-2017, but only 2,706 firms with eligible options returns). This could impact the mean bankruptcy risk in each decile of the cross-section, impacting how the options bankruptcy factor is formed.

Since the Ivy OptionMetrics database only has options data beginning in 1996 I cannot compare portfolios formed in the 1960s and 1980s like I did for the equity decile returns. Additionally, the relative return distribution for the portfolios sorted by Altman's Z-score, the weighted least squares re-estimation of the coefficients of Altman's Z, and the random forest bankruptcy probability are very similar to the pattern observed for the equity factor. Therefore for clarity of exposition I only display results for the portfolios sorted on the random forest bankruptcy probability. The equivalent results for portfolios formed on Altman's Z and the WLS models are available upon request. Table 9 presents these results.

Options returns like equity returns are not monotonic. Despite this fact, the pattern in the cross-section of returns is stronger for the options market than for the equity market reflecting the increased sensitivity of options investors to subtle differences in risks facing underlying asset returns. Despite the stronger pattern in mean returns, the pattern of mean bankruptcy risk for each portfolio is largely similar to the equity portfolios. Deciles 6 through 10 have very similar probabilities of bankruptcy and decile 1 has a very high probability of bankruptcy. This pattern suggests a sort consistent with the equity bankruptcy risk factor.

Table 9: Option Decile Mean Returns and Bankruptcy Risk

Sort	Distress (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Safe (10)	(1)-(10)
RF 1996	8.59%	8.88%	8.22%	7.64%	7.44%	6.57%	7.17%	7.15%	7.55%	7.22%	1.37%
t	(5.85)	(5.52)	(5.55)	(6.03)	(5.26)	(5.44)	(5.47)	(5.62)	(5.68)	(4.73)	(1.89)
$\bar{Z}_{RF}$	0.86	0.69	0.48	0.29	0.16	0.08	0.05	0.03	0.02	0.01	

Note: This table presents mean returns for various portfolios formed by sorting firms based on my random forest produced probability of bankruptcy. The row RF 1996 presents mean returns for each decile. t-statistics are in parentheses. $\bar{Z}$ represents the mean of the bankruptcy risk measure used for each portfolio.

The options factor is formed on a univariate sort of the bankruptcy risk measure developed in Section 2. Specifically, each year firms are sorted in June using my bankruptcy risk probability generated with accounting and market equity information released in December of the previous year. The options factor — titled SSD_o where the subscript is for "options" — is formed by going long on call options belonging to the second decile of Table 9 (avoiding the most distressed firms which have an unreasonably high probability of bankruptcy) and short (shorting call options is very similar to, but not exactly the same as, going long on a put option) on call options for the 50% of firms with the lowest probability of bankruptcy. Note that this sort is the opposite direction of the equity factor of the previous section, and the CDS factor of the next section which both go long on the firms with the lowest probability of bankruptcy[8].

Table 10: Bankruptcy Risk Option Factor: 1996 - 2017

	High (9)	Mid (6-8)	Low (1-5)	SSD_o	t(SSD_o)	Max Drawdown	SR
SSD_o	8.88%	7.81%	7.11%	1.76%	2.44	46.8%	0.53

Note: This table presents information related to the option bankruptcy risk factor. SSD_o is the mean monthly returns from the long-short bankruptcy risk factor, t(SSD_o) is the monthly t-statistic associated with the factor. Max drawdown is the maximum peak-to-trough decrease over the full 22 year sample without regard to a specific time period. SR is the annualized Sharpe ratio. The number is parentheses ((6-8) for example) represent the component decile portfolios from Table 9.

The reversed direction of the call option sort can be explained by two risk related concepts — risk aversion and prudence (Kimball, 1990). The causes of demand for ITM call options are obvious. However, Almeida and Freire (2021) describe additional marginal investors who demand OTM call options. As bankruptcy risk increases (going from column (10) to column (1) in Table 9) investors with negative prudence drive up the price of OTM call options as speculators buy them to use a levered bets. This is consistent with the behavior of investors with risk-neutral measures that place more weight on the right tail of the distribution than exists in the physical distribution of underlying assets.

Previous research applying factors discovered in the equity market to options has documented similar patterns where option portfolio sorts are increasing in the opposite direction of equity portfolios sorted on the same characteristic. Brooks et al. (2018) documents that

[8]Despite the reversal in sorting direction I maintain the name "SSD" (safe subtracting distressed) for consistency throughout this paper.

options returns are increasing in size whereas the well-known size equity anomaly is decreasing in market capitalization. Likewise, Cao, et al. (*forthcoming*) documents a negative relationship between cash holdings and options returns whereas Simutin (2010) documents a positive relationship between cash holdings and stock returns. Neither paper provides an explanation for the difference in return patterns between stocks and options, but it is likely also related to the option pricing pattern discussed in Almeida and Friere (2021).

The values in Table 10 compare favorably to other popular investment vehicles. Again using the example of the momentum factor and the buy and hold S&P 500 index returns, the 1.76% mean monthly return over the period 1996 to 2017 is larger than the 0.41% and 0.67% mean monthly returns of momentum and buy and hold strategies, respectively. Given the sensitivity of options to changes in the underlying security price, it is not surprising to find a larger maximum drawdown for SSD_o compared to SSD_e. Since the maximum drawdown of both the momentum factor and the S&P 500 index occurred during the Great Recession, the values of 58% and 59% reported in the previous section still hold here. Therefore, even though the maximum drawdown for SSD_o is larger than the maximum drawdown of SSD_e, it still compares favorably to other strategies.

Figure 5: Cumulative Returns to SSD_o Investment Strategy: 1996 - 2017

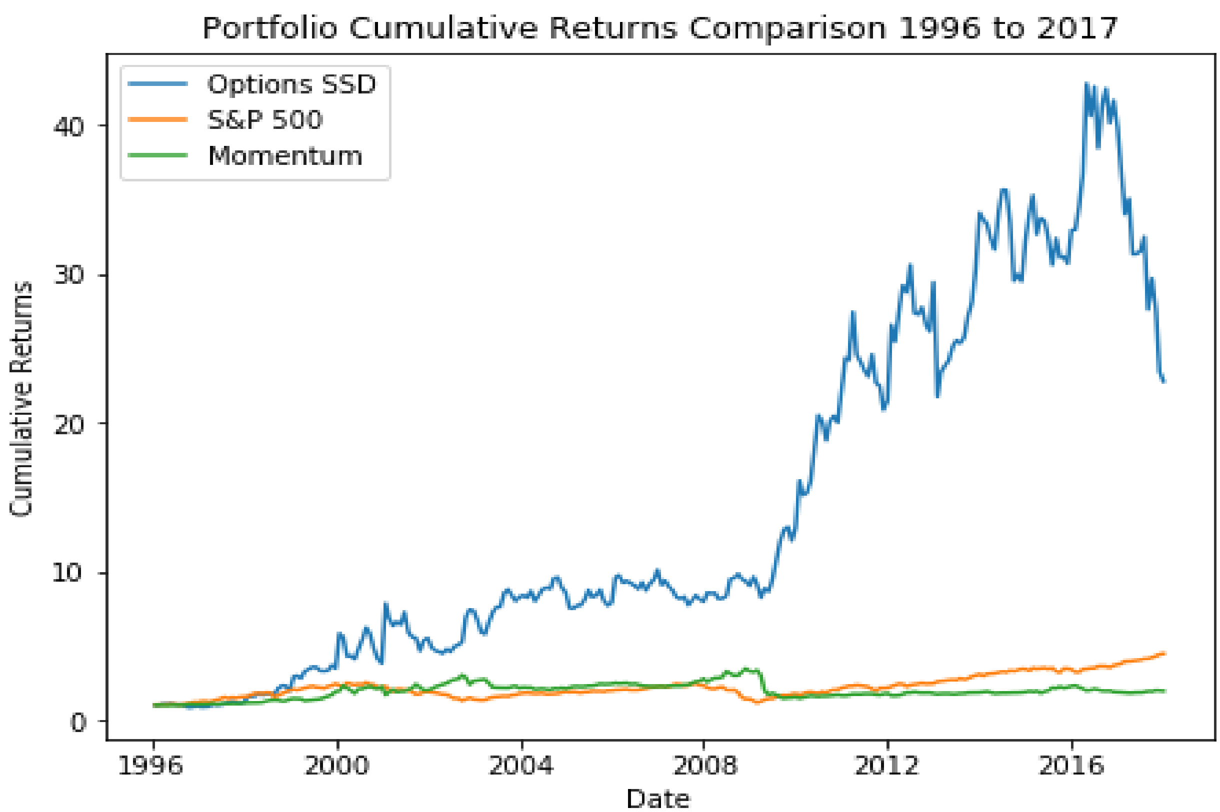


Figure 5 is the call option analogue to Figure 1 and shows the cumulative return to \$1 invested in SSD_o in January 1996 compared to \$1 invested in the S&P 500 index and 1\$ invested in momentum. An investment of \$1 in 1996 would be worth more than \$20 by the end of the sample, with a high of over $40 in early 2016. It is, perhaps, unfortunate the sample ends where it does — at the bottom of a trough that seems comparatively deep. The trough was caused by increasing spreads for options contracts.

Bali et al. (2021) demonstrate that higher bid-ask spreads are associated with lower returns. Spreads are more than twice as large in 2016-2017 than they were during the period of high returns in 2009 and more than six times as large during the initial period of high returns in 1998. Larger spreads make trading options more expensive in general, but the effect is likely especially acute for the investors with negative prudence who use OTM call options as levered bets. As trading these options becomes more expensive, the risk neutral measure used by these investors would require larger and larger right tails to keep demand high. Past a certain weight in the right tail of the risk neutral distribution the option price is too high and demand falls, causing the long-short factor returns to flip.

Even with the decrease from 2016 into 2017, returns to SSD_o dwarfs buy and hold returns over this period. Option spreads do not stay high forever. Although Bali et al. (2021) does not present a time-series of spreads, the ICE BofA US High Yield Index Option-Adjusted Spread obtained from FRED[9] spikes in 2016-2017 — coinciding with the negative returns of SSD_o — but decreases again beginning in 2018. This suggests this period of negative returns is temporary. I can further investigate this once the options return series is expanded.

4.4 Credit Default Swap Factor

Analogously to the equity and option factor, I motivate the formation of the CDS factor by first looking at average returns and bankruptcy risk for decile portfolios sorted on the random forest probability of bankruptcy.

Table 11: CDS Decile Mean Returns and Bankruptcy Risk

Sort	Distress (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Safe (10)	(1)-(10)
RF 2002	0.31%	0.38%	1.11%	2.12%	1.68%	1.52%	1.60%	1.94%	1.52%	1.58%	1.27%
t	(0.29)	(0.44)	(1.16)	(2.05)	(1.97)	(1.87)	(1.75)	(2.22)	(1.57)	(1.60)	(1.46)
$\bar{Z}_{RF}$	0.85	0.70	0.48	0.28	0.18	0.10	0.06	0.03	0.02	0.01	

Note: This table presents mean returns for various portfolios formed by sorting firms based on my random forest produced probability of bankruptcy. The row RF 1996 presents mean returns for each decile. t-statistics are in parentheses. $\bar{Z}$ represents the mean of the bankruptcy risk measure used for each portfolio.

Before describing the decile returns presented in Table 11, it is important to emphasize the differences in what "returns" mean for stocks and options compared to CDSs. When the asset in question is either a stock or option returns represent the percentage difference between the price an investor can purchase the asset for and the price an investor can sell the asset for after a period of one month. However, CDS prices are denominated in basis points. They are the basis point fraction of the debt position being insured which the buyer of CDS protection pays the seller of CDS protection annually[10]. Therefore returns indicate

[9] `https://fred.stlouisfed.org/series/BAMLH0A0HYM2`

[10] Although payments are usually made each quarter.

the percent difference in the basis point rate of purchasing CDS protection in month t and then immediately selling that same level of protection in month $t+1$. In any portfolio that includes a leg where CDSs are "shorted", it follows that shorting a CDS would involve selling CDS protection in month t and then purchasing that same level of protection in month $t+1$.

CDS returns are much more volatile than stock or option returns, as evidenced by the lower t-statistics in Table 11 — only three of the decile portfolios show statistically significant returns. At the same time there is a much clearer pattern of increasing returns to safety than exists in the equity decile sorts even though the relationship is not monotonic for the five safest deciles which have very similar bankruptcy risk and the absolute level of returns is similar. This reflects investor overreaction to changes in perceived risk for firms which were considered safe, while similarly sized changes in perceived risk do not impact firms which investors already deemed risky.

There are unique challenges to the CDS factor that do not exist for the equity and options factor. While there are only approximately 10% as many firms with options contracts compared to firms with exchange listed common stock, the cross-section of CDS contracts is formed using only the 237 firms in the S&P 500 index which have CDS contracts trade at any point between 2002 and 2020. On average there are only 76 CDS contracts trading each month, and some months there are as few as 3 actively traded CDS contracts. This makes portfolio formation using only a single decile impractical, even for firms with high levels of bankruptcy risk.

Given this portfolio size limitation and the fact that firms large enough to be in the S&P 500 index fail at a much lower rate than other firms, regardless of modeled bankruptcy risk, the CDS factor is formed as follows. Each year firms are sorted in June using my bankruptcy risk probability generated using accounting and market equity information released in December of the previous year. The CDS factor — titled SSD_c where the subscript is for "credit default swaps" — is formed by going long on CDSs for the 30% of firms with the lowest probability of bankruptcy an short on CDSs for the 30% of firms with the lowest probability of bankruptcy. The direction of this sort is consistent with the equity factor. Table 12 shows summary statistics related to the CDS bankruptcy risk factor.

Table 12: Bankruptcy Risk CDS Factor: 2002 - 2020

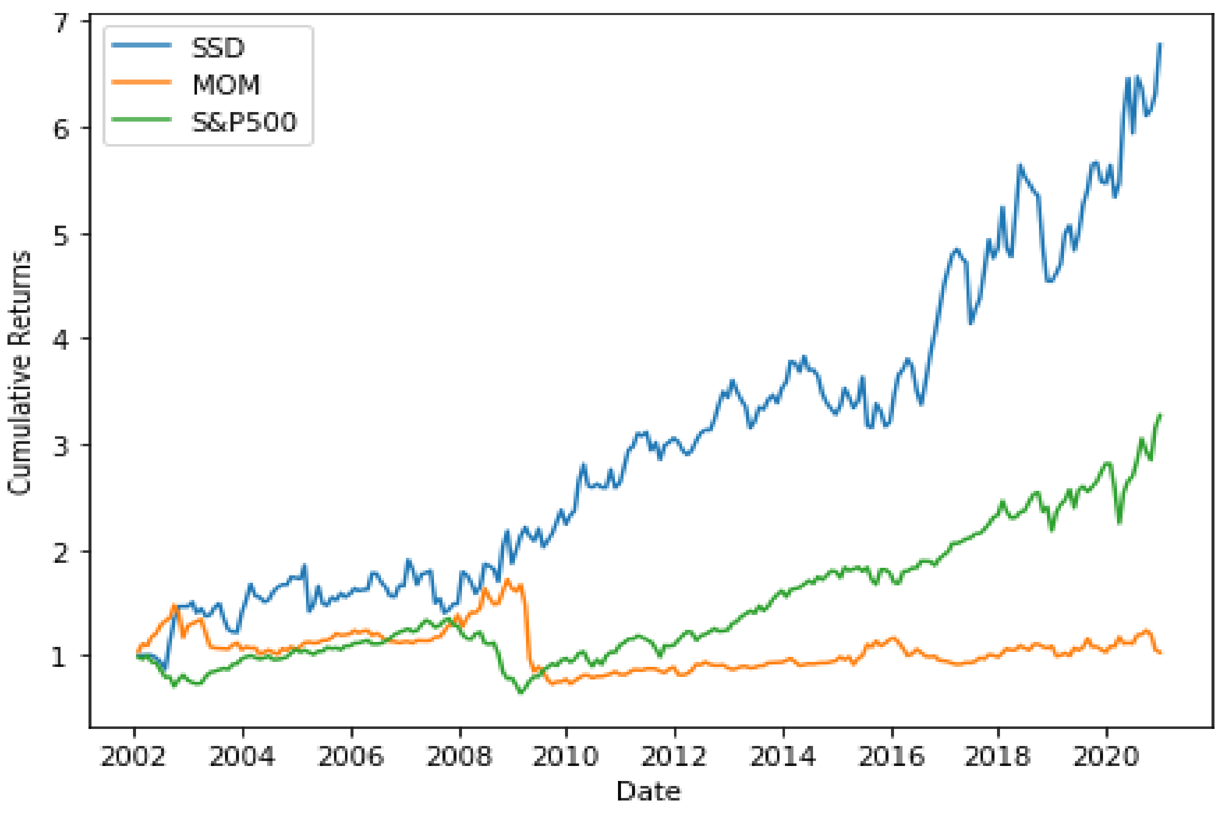
	High (1-3)	Mid (4-7)	Low (8-10)	SSD_o	t(SSD_o)	Max Drawdown	SR
SSD_o	0.83%	1.65%	1.71%	1.04%	2.55	8.9%	0.56

Note: This table presents information related to the option bankruptcy risk factor. SSD_o is the mean monthly returns from the long-short bankruptcy risk factor, t(SSD_o) is the monthly t-statistic associated with the factor. Max drawdown is the maximum peak-to-trough decrease over the full 22 year sample without regard to a specific time period. SR is the annualized Sharpe ratio. The number is parentheses ((6-8) for example) represent the component decile portfolios from Table 9.

The difference between returns on the low bankruptcy risk deciles (decile 8 through 10 in Table 9) and the high bankruptcy risk deciles (decile 1 through 3 in Table 9) is 1.04% per month, on average and is statistically significant with a t-statistic of 2.55. The maximum drawdown is also favorable compared to both a buy and hold return (59%) and the equity and option forms of the bankruptcy risk factor (29.4% and 46.8%, respectively).

Figure 6 shows the cumulative return to \$1 invested in SSD_c in October 2002 to the momentum strategy and a buy and hold strategy. Although CDS data exists for a limited number of firms beginning in 2001, there were not enough firm with debt insured by CDS contracts to form SSD_c until October 2002. There are two notable features of the cumulative return series in Figure 6. First, SSD_c is able to distinguish between firms with low but increasing levels of bankruptcy risk (i.e. those firms with rapidly increasing debt insurance interest rates), and firms with bankruptcy risk so high that CDS contracts do not get materially more expensive during periods of economic downturns. This is particularly evident during the Great Recession where returns to SSD_c continue to increase despite bankruptcy risk increasing across the entire cross-section of CDS contracts. Second, SSD_c has a convex shape and is becoming more profitable as CDS contracts become more liquid. Over the course of only 19 years 1\$ invested in SSD_c increased in value almost 800%.

Figure 6: Cumulative Returns to DMS_c Investment Strategy: 2002 - 2020



4.5 Updating the Factor

Just as it is unreasonable to expect the risk factors included in previous bankruptcy predictor models — including Altman's Z-score — to remain the dominant predictors of risk over time, it is unreasonable to expect my bankruptcy risk model to accurately predict bankruptcy *ad infinitum*. Insofar as safety from bankruptcy risk is representative of the type of asset investors flock to — increasing demand and driving higher returns — this means that it is unreasonable to expect my bankruptcy risk factor to generate returns *ad infinitum*. It will eventually need to be re-estimated. How often, and at what points this re-estimation occurs are both important and difficult questions to answer.

To investigate this problem, I generated long-short factors constructed in the same way as SSD_e using a subset of bankruptcy risk models proposed from 1968 (Altman's Z-score) to 2000 (the F score of Piotroski (2000) and the revised Altman Z (2000)). I chose 2000 as the end date to allow a long enough out-of-sample period to determine how long the bankruptcy risk model successfully predicted returns. I use the average duration from publication date to the date the factor no longer becomes effective as the indicator of how long SSD_i (for $i = e, o, c$) is expected to be an effective investment tool without re-estimation. The date at which the factor is no longer "effective" is either the date of a peak, or the beginning of an extended period of near zero returns.

On average each of the constructed factors took approximately 21 years post publication reaching this effective date. For SSD_i (for $i = e, o, c$) this means the patterns observed in this paper can be expected to last for more than 20 years without the underlying bankruptcy risk model being updated. Of course given the availability of data and cheap processing power, the underlying bankruptcy model can be updated on a yearly basis. As I have demonstrated in this paper it appears that bankruptcy risk can predict returns in multiple markets. Being able to more accurately predict this probability by expanding the number of bankrupt observations in the underlying model should increase the effectiveness of the bankruptcy risk factor.

5 Conclusion

Despite the well-known result from asset pricing equilibrium theory that higher average returns are compensation to investors for holding assets with higher risk, empirically it is seen that assets with lower distress risk have higher average returns in multiple markets. This phenomenon can be explained as a flight to safety by risk averse investors minimizing downside risk by investing in safer assets. Motivated by this empirical fact, this paper proposes a new factor which uses the probability of bankruptcy as a direct proxy for risk. Instead of making the subjective choice of which among the myriad bankruptcy risk models proposed by the literature to use, I use two random forests — one to select five predictors from nearly 250 proposed bankruptcy risk indicators, and one to fit the data — that is, to estimate a new bankruptcy risk model.

The probability of bankruptcy generated by my random forest model has predictive power in four markets — equity markets (SSD_e), bond markets (SSD_b), options markets (SSD_o) and CDS markets (SSD_c). SSD_e earns 0.23% per month and cannot be explained by the CAPM or Fama-French five-factor model. Additionally, DMS_e provides independent information for the cross-section of returns when tested using the two-step method of Feng, Giglio and Xiu (2020). Likewise, SSD_b earns 0.14% per month and cannot be explained by six popular equity factors, six popular bond factors, or the union of those two groups. SSD_o and SSD_c also earn 1.76% and 1.04% per month, respectively.

Just as it is unrealistic to expect the same sources of risk that most impacted firms in the 1960s to be the same risk factors that impact firms today, it is unrealistic to think the bankruptcy risk model introduced in this paper uses risk factors that will best predict bankruptcy *ad infinitum*. Although long-short factors generated by sorting on the output of other bankruptcy risk model successfully predict returns an average of 20 years, it may be beneficial for an investor to update the underlying bankruptcy risk model annually to maximize returns. The two step procedure used in this paper provides a framework by which the underlying bankruptcy risk model can be updated to both include new predictors proposed by the literature and ensure the most relevant predictors are included in the model at any point in time.

Bibliography

Ackert, Lucy; Athanassakos, George. "Prior Uncertainty, Analyst Bias, and Subsequent Abnormal Returns". *The Journal of Financial Research*. Vol 20. Issue 2. pp 263 - 273.

Allayannis, George; Brown, Gregory; Klapper, Leora. "Capital Structure and Financial Risk: Evidence From Foreign Debt Use in East Asia". *The Journal of Finance*. Vol 58. No 6 (2003). pp 2667 - 2710.

Almaida, Caio; Freire, Gustavo. "Pricing of index options in incomplete markets". *Journal of Financial Economics*. In Press. `https://doi.org/10.1016/j.jfineco.2021.05.041` 2021.

Altman, Edward. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *The Journal of Finance*. Vol 23. No 4 (1968). pp 589 - 609.

Altman, Edward; Haldeman, Robert; Narayanan, Paul. "Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations". *Journal of Banking and Finance*. Vol 1. Issue 1 (1977). pp 29 - 54.

Altman, Edward. "Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta® Models". *NYU Stern Working Paper*. 2000.

Alwathainani, Abdulaziz. "Consistency of Firm's Past Financial Performance Measures and Future Returns". *The British Accounting Review*. Vol 41. Issue 3 (2009). pp 184 - 196.

Amihud, Yakov; Mendelson, Haim. "Asset Pricing and the Bid-Ask Spread". *Journal of Financial Economics*. Vol 17. No 2 (1986). pp 223 - 249.

Andersen, Torben; Fusari, Nicola; Todorov, Viktor; Varneskov, Rasmus. "Spatial Dependence in Option Observation Errors". *Econometric Theory*. Vol 37. Issue 2 (2021). pp 205 - 247.

Asness, Clifford; Frazzini, Andrea; Pedersen, Lasse. "Quality Minus Junk". *Review of Accounting Studies*. Vol 24 (2019). pp 34 - 112.

Asness, Clifford; Moskowitz, Tobias; Pedersen, Lasse. "Value and Momentum Everywhere". *The Journal of Finance*. Vol 68. Issue 3 (2013). pp 929 - 985.

Asquith, Paul; Covert, Thom; Pathak, Parag. "The Effects of Mandatory Transparency in Financial Market Design: Evidence from the Corporate Bond Market". *NBER Working Paper No 19417*. 2019.

Avramov, Doron; Chordia, Tarun; Jostova, Gergana; Philipov, Alexander. "Dispersion in Analysts' Earnings Forecasts and Credit Rating". *Journal of Financial Economics*. Vol 91. Issue 1 (2009). pp 83 - 101.

Bai, Jennie; Bali, Turan; Wen, Quan. "Common Risk Factors in the Cross-Section of Corporate Bond Returns". *Journal of Financial Economics*. Vol 131. Issue 3 (2019). pp 619 - 642.

Bali, Turan; Beckmeyer, Heiner; Moerke, Mathis; Weigert, Florian. "Option Return Predictability with Machine Learning and Big Data". SSRN ID 3895984. 2021.

Beneish, Messod; Lee, Charles; Nichols, Craig. "Earnings Manipulation and Expected Return". *Financial Analyst Journal*. Vol 69. Issue 2 (2013). pp 57 - 82.

Breiman, Leo. "Random Forests". *Machine Learning*. Vol 45 (2001). pp 5 - 32.

Breiman, Leo; Friedman, Jerome; Olshen, Richard; Stone, Charles. "Classification and Regression Trees". Wadsworth, New York. 1984.

Brooks, Robert; Chance, Don; Shafaati, Mobina. "The Cross-Section of Individual Equity Option Returns". *LSU Working Paper*. 2018.

Campbell, John; Hilscher, Jens; Szilagyi, Jan. "In Search of Distress Risk". *The Journal of Finance*. Vol 63. Issue 6 (2008). pp 2899 - 2939.

Cao, Jie; Han, Bing; Zhan, Xintong; Tong, Qing. "Option Return Predictability". *Review of Financial Studies*. Forthcoming.

Chen, Andrew; Zimmermann, Tim. "Open Source Cross-Sectional Asset Pricing". *Critical Finance Review. Forthcoming.*

Chen, Nai-Fu; Roll, Richard; Ross, Stephen. “Economic Forces and the Stock Market”. *The Journal of Business*. Vol 59. No 3 (1986). pp 383 - 403.

Chordia, Tarun; Goyal, Amit; Nozawa, Yoshio; Subrahmanyam, Avanidhar; Tong, Qing. “Are Capital Market Anomalies Common to Equity and Corporate Bond Markets?” *Journal of Financial and Quantitative Analysis*. Vol 52. No 4 (2017). pp 1301 - 1342.

Cici, Gjergji; Gibson, Scott; Moussawi, Rabih. “Explaining and Benchmarking Corporate Bond Returns”. SSRN ID 2995626. 2017.

Crawford, Steven; Perotti, Pietro; Price III, Richard; Skousen, Christopher. “Financial Statement Anomalies in the Bond Market”. *Financial Analysts Journal*. Vol 75. Issue 3 (2019). pp 105 - 124.

DeAngelo, Harry; DeAngelo, Linda. “Dividend Policy and Financial Distress: An Empirical Investigation of Troubled NYSE Firms”. *The Journal of Finance*. Vol 45. Issue 5 (1990). pp 1415 - 1431.

Dichev, Ilia. “Is the Risk of Bankruptcy a Systemic Risk?” *The Journal of Finance*. Vol 53. Issue 3 (1998). pp 1131 - 1147.

Dick-Nielsen, Jens. “How to Clean Enhanced TRACE Data”. SSRN ID 2337908. 2014.

Dick-Nielsen, Jens. “Liquidity Biases in TRACE”. *The Journal of Fixed Income*. Vol 19. No 2 (2009). pp 34 - 55.

Fama, Eugene; French, Kenneth. “A Five Factor Asset Pricing Model”. *Journal of Financial Economics*. Vol 116. Issue 1 (2015). pp 1 - 22.

Fama, Eugene; French, Kenneth. “Common Risk Factors in the Returns on Stocks and Bonds”. *Journal of Financial Economics*. Vol 33. Issue 1 (1993). pp 3 - 56.

Fama, Eugene; MacBeth, James. “Risk, Return, and Equilibrium: Empirical Tests”. *Journal of Politicl Economy*. Vol 81. No 3 (1973). pp 607 - 636.

Feng, Guanhao; Giglio, Stefano; Xiu, Dacheng. “Taming the Factor Zoo: A Test of New Factors”. *The Journal of Finance*. Vol 75. Issue 3 (2020). pp 1327 - 1370.

Fernàndez, Albert; Garcia, Salvador; Galar, Mikel; Prati, Ronaldo; Krawczyk, Bartosz; Herrera, Francisco. “Learning From Imbalanced Data Sets”. First Edition. *Springer*. Berlin, Germany. 2018.

Franzen, Laurel; Rodgers, Kimberly; Simin; Timothy. “Measuring Distress Risk: The Effect of R&D Intensity”. *The Journal of Finance*. Vol 62. Issue 6 (2007). pp 2931 - 2967.

Frazzini, Andrea; Pedersen, Lasse. “Betting Against Beta”. *Journal of Financial Economics*. Vol 111. Issue 1 (2014). pp 1 - 25.

Gharghori, Philip; Maberly, Edwin; Nguyen, Annette. “Informed Trading Around Stock Split Announcements: Evidence From the Options Market”. *Journal of Financial and Quantitative Analysis*. Vol 52. Issue 2 (2017). pp 705 - 735.

Goyal, Amit; Saretto, Alessio. “Cross-Section of Options Returns and Volatility”. *Journal of Financial Economics*. Vol 94. Issue 2 (2009). pp 310 - 326.

Griffin, John; Lemmon, Michael. “Book-to-Market Equity, Distress Risk, and Stock Returns”. *The Journal of Finance*. Vol 57. Issue 5 (2002). pp 2317 - 2336.

Gu, Shihao; Kelly, Bryan; Xiu, Dacheng. “Empirical Asset Pricing Via Machine Learning”. *Review of Financial Studies*. Vol 33. Issue 5 (2020). pp 2223 - 2272.

Harvey, Campbell; Liu, Yan. “Lucky Factors”. *Journal of Financial Economics*. Vol 141. Issue 2 (2021). pp 413 - 435.

Harvey, Campbell; Liu, Yan. “A Census of the Factor Zoo”. SSRN ID 3341728. 2016.

Hillegeist, Stephen; Keating, Elizabeth; Cram, Donald; Lundstedt, Kyle. “Assessing the Probability of Bankruptcy”. *Review of Accounting Studies*. Vol 9 (2004). pp 5 - 34.

Hirschleifer, David; Hsu, Po-Hsuan; Li, Dongmei. “Innovative Efficiency and Stock Returns”. *Journal of Financial Economics*. Vol 107. Issue 3 (2013). pp 632 - 654.

Hou, Kewei; Robinson, David. “Industry Concentration and Average Stock Returns”. *The Journal of Finance*. Vol 61. Issue 4 (2006). pp 1927 - 1956.

Jensen, Theis; Kelly, Bryan; Pedersen, Lasse. “Is There a Replication Crisis in Finance?” *NBER Working Paper No 28432*. 2021.

Jones, Stewart; Hensher, David. “Predicting Firm Financial Distress: A Mixed Logit Model”. *The Accounting Review*. Vol 79. No 4 (2004). pp 1011 - 1038.

King, Gary; Zeng, Langche. “Logistic Regression in Rare Events Data”. *Political Analysis*. Vol 9. Issue 2 (2001). pp 137 - 163.

Lennox, Clive. “Identifying Failing Companies: A Re-evaluation of the Logit, Profit and DA Approaches”. *Journal of Economics and Business*. Vol 51. Issue 4 (1999). pp 347 - 364.

Lewellen, Jonathan; Nagel, Stefen; Shaken, Jay. “A Skeptical Appraisal of Asset Pricing Tests”. *Journal of Financial Economics*. Vol 96. Issue 2 (2010). pp 175 - 194.

Lockwood, Larry; Prombutr, Wikrom. “Sustainable Growth and Stock Returns”. *The Journal of Financial Research*. Vol 33. Issue 4 (2010). pp 519 - 538.

Meine, Christian; Supper, Hendrik; Weiss, Gregory. “Is Tail Risk Priced in Credit Default Swap Premia?” *Review of Finance*. Vol 20. Issue 1 (2016). pp 287 - 336.

Merton, Robert. “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”. *The Journal of Finance*. Vol 29. Issue 2 (1974). pp 449 - 470.

Moskowitz, Tobias; Ooi, Yao; Pedersen, Lasse. “Time Series Momentum”. *Journal of Financial Economics*. Vol 104. Issue 2 (2012). pp 228 - 250.

Muravyev, Dmitriy; Pearson, Neil. “Options Trading Costs Are Lower Than You Think”. *Review of Financial Studies*. Vol 33. Issue 11 (2020). pp 4973 - 5014.

Ohlson, James. “Financial Ratios and the Probabilistic Prediction of Bankruptcy”. *Journal of Accounting Research*. Vol 18. No 1 (1980). pp 109 - 131.

Probst, Phillip; Wright, Marvin; Boulesteix, Anne-Laure. “Hyperparameters and Tuning Strategies for Random Forest”. *Data Mining and Knowledge Discovery*. Vol 9. Issue 3 (2019). pp 1301 - 1316.

Sharpe, William. “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk”. *Journal of Finance*. Vol 19. Issue 3 (1964). pp 425 - 442.

Shumway, Tyler. “Forecasting Bankruptcy More Accurately: A Simple Hazard Model”. *The Journal of Business*. Vol 74. No 1 (2001). pp 101 - 124.

Simutin, Mikhail. “Excess Cash and Stock Returns”. *Financial Management*. Vol 39. Issue 3 (2010). pp 1197 - 1222.

Sloan, Richard. “Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings?” *The Accounting Review*. Vol 71. No 3 (1996). pp 289 - 315.

Sweeney, Richard; Warga, Arthur. “The Pricing of Interest-Rate Risk: Evidence from the Stock Market”. *Journal of Finance*. Vol 41. No 2 (1986). pp 393 - 410.

Todorov, Viktor. “Nonparametric Spot Volatility from Options”. *The Annals of Applied Probability*. Vol 29. No 6 (2019). pp 3590 - 3636.

Tuzel, Selale. “Corporate Real Estate Holdings and the Cross-Section of Stock Returns”. *The Review of Financial Studies*. Vol 23. Issue 6 (2010). pp 2268 - 2302.

Vassalou, Maria; Xing, Yuhang. “Default Risk in Equity Returns”. *The Journal of Fi-*

nance. Vol 59. No 2 (2004). pp 831 - 868.

Zmijewski, Mark. "Methodological Issues Related to the Estimation of Financial Distress Prediction Models". *Journal of Accounting Research*. Vol 22 (1984). pp 59 - 82.

Appendix

Appendix A - Random Forest

The random forest was originally introduced by Breiman (2001). It is an iterative collection of classification and regression tree (CART) algorithms, which were themselves originally developed for machine learning purposes by Brieman et al. (1984). Because random forests are made up of a series of classification and regression trees, they belong to a group of models known as ensemble models. The predictions of individual trees, known as "base learners" are aggregated to make the final prediction. If the random forest is used for classification, predictions are made according to majority rule. If the random forest is used for regression, predictions are averaged over the component trees. Because the prediction of so many base learners are combined, predictions made by random forests are much more stable than predictions of individual trees. In this way the random forest model trades off bias in the prediction for a drastic reduction in the variance.

Each tree in the forest is grown according to the CART algorithm, which can be summarized as follows. Beginning at the root node, the data is split (splits are always binary) on a variable chosen from a bootstrap random sample of the full list of explanatory variables. The resulting two groups are called "branches". The split criteria is selected as to make each resulting group (branch) as homogeneous as possible, where homogeneity is defined with respect to an "impurity function". The impurity function is essentially the loss function to be minimized. The process of making these splits is known as "growing the tree", and splits are not required to be unique. For example, the first split and the third split could be on the same variable if further splitting on an already used variable minimizes the impurity function.

The CART algorithm continues to split the data until a stopping criteria is reached. Examples of stopping criteria include a maximum number of layers (steps), a minimum number of observations in each group after a split and a minimum value of the impurity function. Splits are made in a greedy manner. Greedy algorithms choose paths which are locally optimal, and are therefore not guaranteed to arrive at a globally optimal solution. In practice trees are grown such that they are overfit and then "pruned". This process is undertaken because it is possible for a branch close to the root node to not reduce the impurity function by a meaningful amount, but have a child node that splits the data in a way that significantly improves the tree's predictive accuracy. Pruning the tree in this way ensures these child nodes are reached. The terminal nodes of the tree are known as "leaf" nodes. Predictions are made according to the mean value in each leaf node.

Formally, random forests can be represented mathematically using the following indicator function

$$f(x_i, \phi, N, K, L) = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \phi_k \mathbf{1}_{x_i \in C_k(L)} \tag{3}$$

where x_i is the raw input, ϕ is the mean value of the dependent variable in leaf node k, N is the number of trees in the forest, and K and L describe each tree's number of leaf nodes and depth. The indicator function makes clear that each data point input into the model can only reach one leaf node (leaf node k).

Impurity functions can be any function. However, common impurity functions include the regular L_2 loss function

$$L(\phi, C) = \frac{1}{|C|} \sum_{x_i \in C} (x_i - \phi)^2 \tag{4}$$

where C is the number of observations which fall in leaf node k and the remaining variables and parameters are defined as before.

Appendix B - Bankruptcy Predictor Variable Construction

This appendix lists the full set of bankruptcy predictors used in this study along with the original source of the variable and a brief description of its construction. Studies that emphasize the importance of components of constructed variables (i.e. current assets as a component of the current ratio), those components are also included even if they are not the primary focus of the author's paper.

Table 13: Bankruptcy Predictor Variables

Name	Source	Description
LiqAssets	Altman (1968)	Working capital scaled by total assets
CumProfit	Altman (1968)	Retained earnings scaled by total assets
ScaleEBIT	Altman (1968)	Earnings before interest and taxes scaled by total assets
Solvency	Altman (1968)	Market value of equity scaled by book value of debt
CapTurn	Altman (1968)	Sales scaled by total assets
CurrentRat	Tamari (1966)	Current assets scaled by current liabilities
QuickRatio	Tamari (1966)	Liquid current assets scaled by current liabilities
DebtEquity	Ogachi et al., (2020)	Total liabilities scaled by total shareholder equity
PSolvency	Altman (1983)	Book value of equity scaled by book value of total liabilities
ROA	Piotroski (2000)	Net income before extraordinary items scaled by total assets
OpCash	Piotroski (2000)	Cash flows from operation scaled by total assets
ChgROA	Piotroski (2000)	Year over year change in net income before extraordinary items scaled by total assets
AccrualRat	Piotroski (2000)	Net income in excess of cash flows from operations scaled by total assets
ChgLev	Piotroski (2000)	Year over year change in total debt scaled by total assets
ChgCurRat	Piotroski (2000)	Year over year change in current assets scaled by current liabilities
OfferEq	Piotroski (2000)	Year over year change in shares outstanding
MargRat	Piotroski (2000)	Gross margin scaled by total assets
ChgMargRat	Piotroski (2000)	Year over year change in gross margin scaled by total assets
ChgCapTurn	Piotroski (2000)	Year over year change in sales scaled by total assets
SIZE	Ohlson (1980)	Total assets scaled by the GNP price deflator
Leverage	Ohlson (1980)	Total debt scaled by total assets
TLTA	Ohlson (1980)	Total liabilities scaled by total assets
OENEG	Ohlson (1980)	Indicator variable equal to one if total liabilities is larger than total assets

Continued on next page

Name	Source	Description
NITA	Ohlson (1980)	Net income scaled by total assets
FUTL	Ohlson (1980)	Funds from operations scaled by total liabilities
INTWO	Ohlson (1980)	Year over year percentage change in income
BEME	Fama and French (1993)	Book equity scaled by market equity
Debt	Clayton and Ravid (2002)	Total debt
IntCovRat	Clayton and Ravid (2002)	Earnings before interest and taxes scaled by total interest payments
LogSale	Clayton and Ravid (2002)	Log of total sales
Tax	Allayannis et al., (2003)	Total taxes paid
ME	Allayannis et al., (2003)	Total market equity
TanAssets	Allayannis et al., (2003)	Percentage of total assets made up of tangible assets
CFFO	Gentry et al., (1985)	Cash flows from operations
OpLoss	Hopwood et al., (1994)	Total operating expenses minus gross profits
Industry	Sun (2007)	1-digit industry SIC code
CASALES	Sun (2007)	Current assets scaled by sales
CATA	Sun (2007)	Current assets scaled by total assets
PBTCL	Taffler (1977)	Profits before taxes scaled by current liabilities
CATL	Taffler (1977)	Current assets scaled by total liabilities
CLTA	Taffler (1977)	Current liabilities scaled by total assets
NoCredInt	Taffler (1977)	Ratio of quick assets in excess of total liabilities and sales in excess of profits and depreciation, divided by 365
OCLCL	Marais (1979)	Operating cash flows scaled by current liabilities
OCLCL2	Beaver (1966)	Operating cash flows scaled by sales
CLCR	Beaver (1966)	Operating cash flows in excess of dividends paid scaled by current liabilities
LTDCR	Beaver (1966)	Operating cash flows in excess of dividends paid scaled by long term debt
DebtEq2	Warner (1977)	Total debt scaled by total shareholder equity
STDebtEq	Altman et al., (2015)	Total short-term debt scaled by total shareholder equity
IntCovRat2	Rose-Green and Lovata (2013)	The sum of operating cash flows, interest payments and taxes scaled by interest payments

Continued on next page

Name	Source	Description
DebtEbitda	Shaked and Orelowitz (2017)	Total debt scaled by earnings before interest, taxes, depreciation and amortization
GPTA	Philosophov et al., (2008)	Gross profits scaled by total assets
WorkCap	Philosophov et al., (2008)	Current assets minus current liabilities
LTDTA	Philosophov et al., (2008)	Total long-term debt scaled by total assets
PartCLRat	Philosophov et al., (2008)	Current liabilities minus long-term debt due in one year scaled by total assets
IntRat	Philosophov et al., (2008)	Interest payments scaled by total assets
MLDTA	Philosophov et al., (2008)	Long-term debt due in one year scaled by total assets
MLDTA2	Philosophov et al., (2008)	Long-term debt due in two years scaled by total assets
MLDTA3	Philosophov et al., (2008)	Long-term debt due in three years scaled by total assets
MLDTA4	Philosophov et al., (2008)	Long-term debt due in four years scaled by total assets
MLDTA5	Philosophov et al., (2008)	Long-term debt due in five years scaled by total assets
FinLoss	Pindado et al., (2008)	Earnings before interest, taxes, depreciation and amortization minus financial expenses
ChgME	Pindado et al., (2008)	Year over year change in the market value of equity
Return	Aharony et al., (1980)	One year lagged mean returns
VarRet	Aharony et al., (1980)	Year over year change in the volatility of returns
RelativeME	Shumway (2001)	Market equity scaled by the sum of all stock market equity
TradeCred	Aktas et al., (2012)	Level of trade credit as proxied by accounts payable
TobinQ	Chen et al., (2012)	Tobin's Q
Payout	Chen et al., (2012)	The sum of dividends and repurchases scaled by total assets
CFDebtRat	Beaver (1966)	Cash flows scaled by total debt
Segments	Singhal and Zhu (2013)	The number of business segments

Continued on next page

Name	Source	Description
LogAsset	Singhal and Zhu (2013)	Log of total assets
IATS	Singhal and Zhu (2013)	Intangible assets scaled by sales
NetIncome	Singhal and Zhu (2013)	Net income
LabProd	Ho et al., (2013)	Total employment scaled by total assets
WCMan	Kieschnick et al., (2013)	The sum of accounts receivable, accounts payable and inventory
ChgSale	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Year over year percentage change in sales
FinProf	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Gross profits scaled by shareholder equity
WorkCap2	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The sum of shareholder equity, provision for risk and expenses and long-term debt minus fixed assets and other non-current assets
WorkCapReq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The sum of subscribed shares not paid in, accrued expenses and bank loans minus the sum of short-term financial investments, cash, accrued income and current liabilities
Treasury	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Stock holdings plus cash minus loans
Equilibrium	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The sum of shareholder equity, non-current liabilities and long-term debt scaled by total fixed assets
WCR2	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "WorkCapReq" scaled by sales
Treasury2	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "Treasury" scaled by sales
DebtSale	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Total debt scaled by sales
Debtness	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Liabilities scaled by the sum of liabilities and shareholder equity

Continued on next page

Name	Source	Description
EqCapRat	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Shareholder equity scaled by non-current liabilities
PayCap	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The sum of long-term debt and current liabilities scaled by the sum of sales, depreciation and operating and investing provisions
ShareFund	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Profits scaled by shareholder equity
RetCapEmp	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Profits before interest payments scaled by the sum of shareholder equity and non-current liabilities
Margin	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Profits scaled by asset turnover
NetTurn	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Operating revenue scaled by the sum of shareholder equity and non-current liabilities
IntCover	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Operating profits scaled by interest payments
StockTurn	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Operating revenue scaled by asset turnover
ShareLiq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Shareholder equity scaled by non-current liabilities
Gearing	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The sum of non-current liabilities and loans scaled by shareholder equity
ChgSaleSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Squared year over year percentage change in sales
CapTurnSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Squared capital turnover ratio
GPTASq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of profits scaled by total assets

Continued on next page

Name	Source	Description
FinProfSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of profits scaled by shareholder equity
WorkCap2Sq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "WorkCap2" squared
WCRSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "WorkCapReq" squared
TreasurySq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "Treasury" squared
EquilSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "Equilibrium" squared
WCSaleSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of working capital scaled by sales
WCR2Sq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "WCR2" squared
Treas2Sq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "Treasury2" squared
DebtSaleSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of debt scaled by sales
DebtnessSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "Debtness" squared
EqCapRatSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "EqCapRat" squared
PayCapSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio "PayCap" squared
QRSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The squared quick ratio

Continued on next page

Name	Source	Description
CRSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The squared current ratio
ShareFundSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “ShareFund” squared
RCESq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “RetCapEmp” squared
GPTASq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of profits scaled by total assets
MarginSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of profits scaled by asset turnover
NetTurnSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “NetTurn” squared
IntCoverSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “IntCover” squared
StTurnSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “StockTurn” squared
ShareLiqSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “ShareLiq” squared
GearSq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	Constructed ratio “Gearing” squared
WCTASq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of working capital scaled by total assets
EBITTASq	Acosta-Gonzalez and Fernandez-Rodriquez (2014)	The square of earnings before interest and taxes scaled by total assets
NATCTC	Altman et al., (1977)	Net available for total capital scaled by total capital
SaleTC	Altman et al., (1977)	Sales scaled by total capital

Continued on next page

Name	Source	Description
EBITSale	Altman et al., (1977)	Earnings before interest and taxes scaled by sales
NATCSale	Altman et al., (1977)	Net available for total capital scaled by sales
LogTanAss	Altman et al., (1977)	Log of tangible assets
LogIntCov	Altman et al., (1977)	The log of earnings before interest and taxes scaled by interest payments
LogWCLTD	Altman et al., (1977)	Log of working capital scaled by long-term debt
WCLTD	Altman et al., (1977)	Working capital scaled by long-term debt
FCCR	Altman et al., (1977)	The sum of earnings before interest and taxes and fixed charges scaled by the sum of fixed charges before taxes and interest payments
EBITDebt	Altman et al., (1977)	Earnings before interest and taxes scaled by debt
CFFC	Altman et al., (1977)	Cash flows scaled by fixed charges
WCCE	Altman et al., (1977)	Working capital scaled by cash expenses
BETC	Altman et al., (1977)	Book equity scaled by total capital
METC	Altman et al., (1977)	Market value of equity scaled by total capital
EBITDrop	Altman et al., (1977)	Year over year change in earnings before interest and taxes
MDrop	Altman et al., (1977)	Year over year change in profits
SaleFA	Altman et al., (1977)	Sales scaled by fixed assets
DivPayRat	Wilcox (1971)	Total dividends paid scaled by net income
LiqValueIF	Wilcox (1976)	Net income minus total dividends paid
LiqValue	Wilcox (1976)	Sum of real estate, equipment and inventory
ChgAsset	Wilcox (1976)	Year over year change in total assets
QATA	Beaver (1968)	Quick assets scaled by total assets
CashTA	Beaver (1968)	Total cash holdings scaled by total assets
QACL	Beaver (1968)	Quick assets scaled by current liabilities
QASale	Beaver (1968)	Quick assets scaled by sales
WCSale	Beaver (1968)	Working capital scaled by sales
CashSale	Beaver (1968)	Total cash holdings scaled by sales

Continued on next page

Name	Source	Description
EATTA	Beaver (1966)	Earnings after taxes scaled by total assets
QuickInv	Beaver (1966)	Quick assets scaled by inventory
OpMargin	Edminster (1972)	Operating earnings scaled by revenue
InvWC	Edminster (1972)	Inventory scaled by working capital
CADebt	Edminster (1972)	Current assets scaled by total debt
FAEquity	Edminster (1972)	Fixed assets scaled by total shareholder equity
CLEquity	Edminster (1972)	Current liabilities scaled by total shareholder equity
OwnAssets	Edminster (1972)	The sum of shareholder equity and long-term debt scaled by fixed assets
InvSale	Edminster (1972)	Inventory scaled by total sales
FASale	Edminster (1972)	Fixed assets scaled by total sales
TASale	Edminster (1972)	Total assets scaled by sales
SESale	Edminster (1972)	Total shareholder equity scaled by sales
EBITSE	Edminster (1972)	Earnings before interest and taxes scaled by total shareholder equity
EBITDD	Edminster (1972)	The sum of earnings before interest and taxes and depreciation scaled by total debt
CATOR	Altman (1973)	Current assets scaled by total operating revenue
IBITTA	Altman (1973)	Income before interest and taxes scaled by total assets
RevProp	Altman (1973)	Operating revenue scaled by the value of total property
OpEff	Altman (1973)	Operating expenses scaled by operating revenue
GrowRate3	Altman (1973)	Three year percentage growth in operating revenue
LiqAss	Sinkey (1975)	The sum of cash and treasury securities scaled by total assets
LoanTA	Sinkey (1975)	Total loans scaled by total assets
OEOI	Sinkey (1975)	Operating expenses scaled by operating income
LoanRev	Sinkey (1975)	Total loans scaled by total revenue
TresRev	Sinkey (1975)	Total holdings of U.S. treasury securities scaled by revenue
IntRev	Sinkey (1975)	Total interest paid scaled by revenue
NIBTTC	Korobow and Stuhr (1975)	Net income before taxes scaled by total capital
DivTC	Korobow and Stuhr (1975)	Dividends paid scaled by total capital
BorrowTC	Korobow and Stuhr (1975)	Total borrowing scaled by total capital
TCTA	Korobow and Stuhr (1975)	Total capital scaled by total assets

Continued on next page

Name	Source	Description
OccExpTA	Korobow and Stuhr (1975)	Net occupancy expenses scaled by total assets
TresTA	Korobow and Stuhr (1975)	Total holdings of U.S. treasury securities scaled by total assets
LoanCA	Korobow and Stuhr (1975)	Loans scaled by current assets
LLTC	Martin (1977)	The sum of loans and leases scaled by total capital
CATC	Martin (1977)	Current assets scaled by total capital
NITACash	Martin (1977)	Net income scaled by the difference between total assets and cash holdings
LoanTAC	Martin (1977)	Total loans scaled by the difference between total assets and cash holdings
NLATACash	Martin (1977)	Net liquid assets scaled by the difference between total assets and cash holdings
Loans	Martin (1977)	Total loans held
NIMEA	Martin (1977)	Net interest margin scaled by earning assets
ProdEff	Martin (1977)	Non-interest expenses scaled by the difference between operating revenue and interest expenses
DivType	Martin (1977)	Common stock dividends scaled by the difference between net income and preferred stock dividends
NIATTA	Altman and Lorris (1976)	Net income after taxes scaled by total assets
LLSE	Altman and Lorris (1976)	The sum of total liabilities and loans scaled by total shareholder equity
EndBegCap	Altman and Lorris (1976)	The difference between this year's capital and capital additions scaled by the prior year's capital
CapLag	Altman and Lorris (1976)	One period lagged capital
NIATCap	Altman and Lorris (1976)	Net income after taxes scaled by the prior year's capital
NetGross	Altman (1977)	Net operating income scaled by gross operating income
NetTotal	Altman (1977)	Net income scaled by total income
OPTA	Altman (1977)	Total operating expenses scaled by total assets
NWTA	Altman (1977)	Net worth scaled by total assets
PPETA	Altman (1977)	The total value of plants property and equipment scaled by total assets
NOINI	Altman (1977)	Non-operating income scaled by net income
NINW	Altman (1977)	Net income scaled by net worth
OfficeExp	Altman (1977)	Office building expenses scaled by operating income

Continued on next page

Name	Source	Description
ESTA	Altman (1977)	Earned surplus scaled by total assets
PaidEquity	Santomero and Vinso (1977)	Total shareholder equity
TotCap	Santomero and Vinso (1977)	Total capital
CapAssRat	Santomero and Vinso (1977)	Capital scaled by total assets
NISale	Lee et al., (2012)	Net income scaled by total sales
EBITDATA	Lee et al., (2012)	Earnings before interest and taxes, depreciation and amortization scaled by total assets
SETA	Lee et al., (2012)	Total shareholder equity scaled by total assets
SEIATA	Lee et al., (2012)	The difference between shareholder equity and intangible assets scaled by total assets
Lev2	Lee et al., (2012)	Total assets minus intangible assets, cash holdings and the value of land and buildings
CLCTA	Lee et al., (2012)	The difference between current liabilities and cash holdings scaled by total assets
ARSale	Lee et al., (2012)	Accounts receivable scaled by total sales
APSale	Lee et al., (2012)	Accounts payable scaled by total sales
InvGrow	Lee et al., (2012)	Year over year percentage change in inventory
LTGrow	Lee et al., (2012)	Year over year percentage change in total liabilities
CashGrow	Lee et al., (2012)	Year over year percentage change in cash holdings
aftret_eq	WRDS Ratio	After-tax return on average common equity
aftret_equity	WRDS Ratio	After-tax return on total shareholder equity
capital_ratio	WRDS Ratio	Fraction of capital made up by debt
cash_lt	WRDS Ratio	Cash balance scaled by total liabilities
cash_ratio	WRDS Ratio	Cash and cash equivalents scaled by current liabilities
curr_debt	WRDS Ratio	Current liabilities scaled by total liabilities
de_ratio	WRDS Ratio	Total debt scaled by total equity
debt_assets	WRDS Ratio	Debt to assets ratio - different specification to constructed ratio "Leverage" above
debt_at	WRDS Ratio	Alternative specification of the debt to assets ratio
debt_capital	WRDS Ratio	Total debt scaled by total capital
debt_ebitda	WRDS Ratio	Total debt scaled by earnings before interest, taxes, depreciation and amortization
evm	WRDS Ratio	Enterprise value scaled by earnings before interest, taxes, depreciation and amortization
gprof	WRDS Ratio	Alternative specification of profits scaled by assets
invt_act	WRDS Ratio	Inventory scaled by current assets
lt_debt	WRDS Ratio	Long-term debt scaled by total liabilities

Continued on next page

Name	Source	Description
lt_ppent	WRDS Ratio	Total liabilities scaled by total tangible assets
profit_lct	WRDS Ratio	Profits before depreciation scaled by current liabilities
quick_ratio	WRDS Ratio	Acid test ratio
rd_sale	WRDS Ratio	Research and development expenditures scaled by total sales
rect_act	WRDS Ratio	Total receivables scaled by current assets

Appendix C - Factor Zoo Selected Variables

Table 14: Feng, Giglio & Xiu Factor Zoo Selected Variables

Name	Source	Description
AdExp	Chan, Lakonishok and Sougiannis (2001)	Advertising expenditures
BMdec	Fama and French (1992)	Book-to-market using December market equity
ChInvIA	Abarbanell and Bushee (1998)	Change in capital expenditures above or below an industry benchmark
ChNWC	Soliman (2008)	Change in working capital
ChTax	Thomas and Zhang (2011)	Quarterly changes in tax expenses
CompEquIss	Daniel and Titman (2006)	Share issuance for both cash and services (i.e. including employee stock plans)
DivSeason	Hartzmark and Salomon (2013)	Month in which a dividend is expected to be issued
EarningsSurprise	Foster, Olsen and Shevlin (1984)	Sign and magnitude of earnings forecast errors
EarnSupBig	Hou (2007)	Sign and magnitude of earnings forecast errors for large firms
EBM	Penman, Richardson and Tuna (2007)	Net operating assets scaled by price
EP	Basu (1977)	Price earnings ratio
NetDebtPrice	Penman, Richardson and Tuna (2007)	Market value of debt scaled by market value of equity
FirmAge	Barry and Brown (1984)	Months since listed on an exchange
RD	Chan, Lakonishok and Sougiannis (2001)	R&D expenditures scaled by market capitalization
VolMkt	Haugen and Baker (1996)	Trading volume scaled by market equity