6-22-2023

EDF and short selling as indicator of default

Huiyang Li

EDF and short selling as indicator of default

by

Huiyang Li

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Business Administration

The University of Memphis

May 2023

Acknowledgments

I express my profound gratitude to my dissertation committee, comprising Dr. Thomas McInish, Dr. Ronald Spahr, Dr. Sabatino Silveri, Dr. Jade Planchon, and Dr. Velma Zahirovic-Herbert, for their unwavering support, invaluable insights, and expert guidance throughout my doctoral studies. I am also deeply grateful to my parents, Chang Li and Min Wu, my husband Zixiao Ye, my dear friend Xixi Gu, and my beloved cats Toffee and Clara, for their constant love, companionship, and understanding. Their support made this journey fulfilling and meaningful.

Abstract

Huiyang Li, Ph.D., The University of Memphis. April 2023. EDF and short selling as indicator of default. Major Professor: Ronald Spahr, Ph.D.

This dissertation presents two papers that examine the efficacy of expected default frequency (EDF) when predicting bankruptcy events and whether short-selling utilization could be used as an indicator of default risk change.

The first paper examines the efficacy of Merton's (1974) distance to default model as simplified by Bharath and Shumway (2008) to forecast a firm's Expected Default Frequency (EDF) or probability of defaulting on debt obligations. Merton's model, further developed by the KMV corporation, is based on the Black-Scholes asset pricing model. We apply the simplified Bharath and Shumway model by relating it with Cox's proportional-hazards model (Cox, 1972) to forecast firm-specific expected default frequency (EDF) or probability of default. The accuracy of the Merton/Bharath-Shumway Model is further examined by including its input as one of the variables in a principal components-factor analysis to determine its ability to predict firm bankruptcy and its orthogonality in predicting firm default and potential bankruptcy. We also measure how much of the total variation is explained in a adjusted R squared decomposition analysis.

The second paper investigate the relation between short interests of a firm's common stock and commensurate changes in the firm's expected default probability/risk, using Compustat and NYSE data from 2007 to 2018. Default risk is measured by a simplified Merton distance-to-default (EDF) model, where we find that a firm's short-interests predicts changes in default risk. Further, short-interest levels predict a firm's likelihood of default as measured by it movement into the top default-risk decile. Thus, we recommend that investors, managers, and regulators employ short-interests as an early warning signal for a firm's potential default.

Table of Contents

List of Tables

Chapter 1

The Efficacy of Expected Default Frequency (EDF) Models in Predicting U.S. Public Firm Defaults

1. Introduction and literature review

Firm defaults and potential bankruptcies result from a firm's failure to meet debt service obligations, including required principal payments and interest expenses. Because of serious implications of a firm's default and potential bankruptcy, an accurate default prediction is critical for bond pricing and rating estimation. Generally, if a firm's default/bankruptcy risk increases, its profitability and cash flows may become problematic. Also, higher default probabilities may be exacerbated by increases in firm financial leverage.

Financial ratios provide comprehensive information in measuring a firm's financial condition measurable, thus, beginning with Beaver (1966), a majority of studies focus on predicting firm defaults using financial ratios. Beaver finds that the cash-flow to total assets ratio predicts defaults/bankruptcy significantly better than the liquid assets ratios, and that financial ratios perform better when predicting non-defaults events compared to defaults.

Altman (1968) employs a five-factor multivariate discriminant analysis model in calculating a firm's Altman Z score that estimates a firm's financial strength. Using a similar approach, Ohlson (1980) includes 9 financial ratios in calculating an O score, concluding that the O score has more robust predictive power in predicting defaults and bankruptcies; where, the additional factors significantly improve model predictivity.

A firm's financial ratios represent only one category of factors affecting a firm's probability of default and/or bankruptcy. The economic conditions, the business cycle and the data source may also influence default/bankruptcy risk. Therefore, investors should rely on more than just financial ratios to avoid investing in firms likely to default.

Tinoco and Wilson (2013) create a model using financial ratios, market data, and macroeconomic data to predict defaults, finding that their model's accuracy was superior to a neural network model and Altman Z score, which utilizes only financial data. Additionally, Goudie and Meeks (1991) proposed a macro-micro multivariate model, suggesting that macroeconomic conditions may impact a firm's risk of bankruptcy.

The above models rely on financial ratios and macroeconomic data; however, some studies employ the Black-Scholes (1973) asset pricing model to assess credit risk. Subsequently, Merton (1974) introduces the distance-to-default (DD) model based on the Black-Scholes-Merton asset pricing model. In this model, the expected default frequency (EDF) is calculated as the distance to default and serves as a predictor of default. Merton treats the value of stock equity as a call option on the value of the entire firm, which includes liabilities. If the value of liabilities exceeds total assets, investors will fail to exercise the call option, or liquidate. Otherwise, if the value of assets exceeds liabilities investors would exercise the option. However, market values of both liabilities and total assets is not easily observable.

Bharath and Shumway's (2008) simplified Merton model is the first structural form model that considers bankruptcy as a continuous probability of default and demonstrates its effectiveness in predicting bankruptcies. In contrast, Jarrow and Turnbull (1995) propose a reduced-form model that defines the determinants of bankruptcy as a series of endogenous factors and assumes that investors lack complete knowledge of the firm's financial position. Thus, the assumptions associated with Merton's model (structural form model) are more rigid and sensitive, while those of the reduced-form model are generally more realistic.

In another study, Duffie, Saita, and Wang (2007) estimate term structures of corporate default probabilities over multiple future periods, taking into account firm-specific and macroeconomic variables, concluding that default probabilities are influenced by the state of the economy and the firm's leverage ratio, which can be captured by expected default frequency (EDF). EDF is used by Moody's, one of the largest credit rating agencies, who developed the KMV model to forecast default probabilities for all public firms.

Hillegeist, Keating, Cram, and Lundstedt (2004) compare the accuracy of Altman Z score, Ohlson’s O score, and Black-Scholes-Merton (BSM) model and found that the BSM model outperformed Altman Z score and Ohlson’s O score.

Bharath and Shumway (2004) tested the accuracy of the KMV model for predicting default events and found it to be insufficiently efficient. However, they show that the KMV-Merton model may be improved without solving the simultaneous nonlinear equation required by the KMV model. Therefore, we use the expected default frequency from Merton’s model as a proxy for bankruptcy risk and assess its effectiveness and contribution in predicting actual bankruptcy events.

Other previous studies use multiple models to measure the efficacy of measuring and predicting real world defaults/bankruptcies. As previously mentioned, Altman’s (1968) highly utilized z scores and Goudie and Meeks (1991) use multivariate discriminant analysis to calculate default and bankruptcy predictions, where, multivariate discriminant analysis facilitates the incorporate multiple variables into a single score to predict the probability of default/bankruptcies. However, the accuracy of the prediction depends on the quality and relevance of the predictive variables as well as the data quality. Therefore, variable selection and data sources are crucially important when using multivariate discriminant analysis to predict bankruptcy risk.

Other studies exist that use probit and logit models to forecast business failures and defaults. For example, Kovacova and Kliestik (2007) performed a comparative analysis that applies probit and logit models in predicting bankruptcy for companies in Slovakia. They conclude that the probit model marginally outperforms the logit model in forecasting bankruptcy events.

The proportional hazard model proposed by Cox (1972) is widely used in risk-related studies, including those related to firm default. While it is commonly applied in medical research to explore the relationship between a patient's survival time and predictor variables, it also may be utilized to examine the time to firm default or bankruptcy.

Chava and Jarrow (2004) predict bankruptcies using a hazard model, similar to Cox, discovered that monthly observation intervals provide better predictions of bankruptcy/default compared to annual observation intervals. Furthermore, they emphasized the significance of incorporating industry effects into hazard rate estimation to enhance prediction accuracy.

In this study, we apply both the probit logit model and the Cox proportional hazard model to evaluate the efficacy of using expected default frequency (EDF) in predicting bankruptcy events. Additionally, we decompose the adjusted R-squared to assess the specific contribution of the EDF when combined with other default risk proxies such as the z-score and capital intensity ratio, in predicting firm defaults and bankruptcies.

2. Data and Methodology

2.1 Data

We use firm-level financial data in estimating expected default frequencies, EDFs, to predict firm defaults and assess EDF contributions in predicting defaults.

We collected firm-specific financial data from COMPUSTAT from Oct 1, 1979 through Dec 31, 2021, excluding financial and regulated industries, firms with SIC between 4900 and

4949 and SIC between 6000 and 6999. These firms are excluded because they are subject to regulation and special capital requirements. Also, we exclude firms with no shares outstanding at the end of the fiscal year.

It is logical that a firm's liquidity and profitability ratios tend to deteriorate when a firm approaches bankruptcy. Thus, Legwon and Jager (2020) test for the efficacy of financial ratios, market data, industry and year in predicting bankruptcy. It is logical that firm profitability is an important factor in predicting bankruptcies. Thus, we use NITA (Net Income/Total Assets) and CASHTA (Cash and Cash equivalent/Total Assets) as proxies for a firm's financial condition. Also, we control for firm size by using the ln firm market capitalization at the end of each fiscal year.

S&P500 index historical price data are obtained from Bloomberg, where, historical price data are used to calculate annual market returns and standard deviations. If a firm files for bankruptcy in a given year, we use the one-year annual market return and market standard deviation-volatility for the year prior to the filing date (the day when the firm files bankruptcy) as proxies for market condition. For surviving firms, we use annual market return and market volatility in the fiscal year.

Bankruptcy data, obtained from the UCLA database from 1979 through 2021, include the date and firm specific detailed information when a firm files for bankruptcy. A firm is defined as a "big firm" if the total assets value is greater than $100 million in the last financial report (10-K) before bankrupt.

We match the UCLA bankruptcy data with financial and market data. A firm filing for bankruptcy during a fiscal year is identified by a dummy variable, "default", equal to 1 for the fiscal year or otherwise, equal to 0.

2.2 Methodology

2.2.1 Altman Z Score

The Altman Z-score is calculated in accordance with the five-factor approach outlined by Altman (1968), and is used as a measure of a firm's likelihood of bankruptcy in the near future. A higher Altman Z-score implies a lower probability of bankruptcy. The formula for calculating the Altman Z-score is presented below.

$$Altman\ Z\ score = 1.2 * \frac{Working\ capital}{Total\ Assets} + 1.4 * \frac{Retained\ earnings}{Total\ Assets} + 3.3 * \frac{EBIT}{Total\ Assets} + 0.6 * \frac{Market\ Cap}{Total\ Liabilities} + 1.0 * \frac{Sales}{Total\ Assets}$$

2.2.2 Ohlson O Score

Also, we measure the Ohlson O-Score, Ohlson (1980) that uses 9 factors, more than the 5 factors used by Altman's model. Using more factors may facilitate a more accurate bankruptcy prediction. A larger Ohlson O-score indicates that a firm has a higher chance of filing for bankruptcy in the next 2 years.

$$Ohlson\ O - score = -1.32 - 0.407 log\left(\frac{Total\ Assets}{GNP}\right) + 6.03 * \frac{Total\ Liabilities}{Total\ Assets} - 1.43 * \frac{Working\ Capital}{Total\ Assets} + 0.0757 * \frac{Current\ Liabilities}{Current\ Assets} - 1.72X - 2.37 * \frac{Net\ income}{Total\ Assets} - 1.83 * \frac{Cash\ flow\ from\ OPeration}{Total\ Liabilities} + 0.285Y - 0.521 * \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|}$$

Where X=1 if total liabilities are greater than total assets, 0 otherwise.

Y=1 if the firm experienced net losses in the past two years, 0 otherwise.

GNP is the gross national product price index level in USD compared with 1968 level (assume GNP of 1968 = 100).

2.2.3 Merton's Distance to Default model

We calculate expected default frequency (EDF). a measure default probabilities using the simplified Merton distance-to-default model proposed by Bharath and Shumway (2008). A larger EDF indicates that a firm is less likely to survive in the near future.

$$DD = \frac{\ln\left(\frac{Firm\ Value_{i,t}}{Debt_{i,t}}\right) + \left(r_{i,t-1} - 0.5\sigma^2_{V_{i,t}}\right) * T_{i,t}}{\sigma_{V_{i,t}} * \sqrt{T_{i,t}}}$$

$$EDF = N\left(-DD_{i,t}\right)$$

$$\sigma_{D_{i,t}} = 0.05 + 0.25 * \sigma_{E_{i,t}}$$

$$\sigma_{V_{i,t}} = \frac{Equity_{i,t}}{Firm\ Value_{i,t}} * \sigma_{E_{i,t}} + \frac{Debt_{i,t}}{Firm\ Value_{i,t}} * \sigma_{D_{i,t}}$$

Also, we use an alternative method, provided by Bharath and Shumway (2008), to calculate debt volatility-standard deviation, $\sigma_{D_{i,t}}$. In Equation (3), 0.05 (5%) represents term-structure volatility. Since firms with higher debt default probabilities demonstrate higher equity volatilities, $0.25*\sigma_{E_{i,t}}$ in Eq. (3) measures default risks associated with equity volatility. $T_{i,t}$ is set to 1 because we use annual financial data. Since the face value of debt is difficult to observe, Merton's distance-to-default model assumes debt is equal to current liabilities plus half of the long-term book-value debt, which is in accordance with Bharath and Shumway (2008).

2.2.4 Cox Proportional Hazard Model

Also, we use the Cox proportional hazard model as another predictor of bankruptcy and measure each variable’s and resulting factor's performance (contribution) in predicting bankruptcy.

The Cox proportional-hazards model (Cox, 1972) is a probit regression model commonly used in medical research to investigate the association between patients' survival time and one or more predictor variables. This also may be apply to time to default/bankruptcy. Firms that do not file for bankruptcy during a given year are define it as a survival firm. In the Cox proportional hazard model, the default rate is defined as the hazard rate.

The Cox model is expressed by the hazard function denoted by h(t).

$$h(t) = h_0(t) + \exp\left(b_1 * x_1 + b_2 * x_2 + \cdots + b_p * x_p\right)$$

where,

- t represents the survival time
- h(t) is the hazard function determined by a set of p covariates ($x_1, x_2, ..., x_p$)
- the coefficients ($b_1, b_2, ... b_p$) measure the variables' impact (i.e., the effect size).

The term h_0 is called the baseline hazard. It corresponds to the value of the hazard if all the x_i (predictors) are equal to zero (the quantity exp (0) equals 1). The ‘t’ in h(t) reminds us that the hazard may vary over time.

The quantities *exp*(b_i)) are called hazard ratios (HR). A value of b_i greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the i^{th} covariate increases, the event hazard increases. Thus, the length of survival (no bankruptcy) decreases.

HR = 1 indicates that a covariate has no effect on predicting bankruptcy.

HR > 1 indicates that a covariate is associated with increased risk (decreased survival). We define a covariate with HR > 1 as a bad predictive factor in the default study.

HR < 1 indicates that a covariate is associated with decreased risk (increased survival). We define a covariate with HR < 1 as a good predictive factor in the default study.

In our case, a hazard ratio (HR) is the probability of bankruptcy relative to firms not declaring bankruptcy over a year. This ratio is a time measure for time-to-bankruptcy.

HRs are similar to relative odds ratios (ORs), where, ORs measure the probabilities of firm bankruptcies and firms not declaring bankruptcy. However, a critical difference exists. Hazard ratios originate from survival analysis that indicate time-to-event. A hazard ratio (HR) measures time-to-event/bankruptcy.

3. Results

3.1 Summary Statistics

Table 1 shows the summary statistics for variables used in this paper. There are 88,386 firm-year observations from 1979 through 2021. Only EBITDASA has a negative skewness, while other variables show positive skewness. The positive skewness of EDF indicates that most of the firms are in the "safe zone".

Table 2 and Figure 1 show the frequency of firm bankruptcies by year and industry. Observe that the number of bankruptcies indicates an opposite trend with GNP, reinforcing the logic that there are fewer bankruptcies when the macro economy performs better, and verse versa. Note that more than 25% of the failed firms were in the more labor and capital-intensive manufacturing industry.

3.2 Factor Analysis

3.2.1 Principal Component Analysis

Factor Analysis and Principal Components Analysis are similar; however, Principal Components Analysis involves extracting linear composites of observed variables. Alternatively, Factor Analysis is based on a formal model predicting observed variables from theoretical latent factors.

Thus, we conduct a principal component analysis to identify a reasonable number of variables to include in predicting bankruptcies. We include thirteen variables, Altman Z score, Ohlson O score, capital intensity ratio, expected default frequency (EDF), cash and cash equivalent to total assets ratio (CASHTA), EBITDA to total assets ratio (EBITDATA), cash flow from operating activity to total assets ratio (CFFOTA), cash and cash equivalent to sales ratio (CASHSA), EBITDA to sales ratio (EBITDASA), cash flow from operating activity to sales ratio(CFFOSA), market return, market volatility, and firm size.

Among the thirteen variables, CASHTA, EBITDATA, and CFFOTA are indicators of a firm's fundamental financial information, profitability and short-term solvency. CASHSA, EBITDASA, and CFFOSA measure the firm's fundamental financial information but use revenue as a benchmark.

Since a firm's leverage (leverage ratio) is previously accounted for in the EDF, we do not include it as a separate variable. Assuming that firms are more likely to go bankrupt during periods of high market volatility or stock price downturns, we use market returns and volatilities to gage the likelihood of firm bankruptcy.

Incorporating both the Altman Z score and Ohlson O score facilitates our comparing the relative contribution of each score in calculating the EDF and prediction of bankruptcy.

Results of the principal component analysis are presented in Table 3, revealing that nine factors have an Eigenvalue greater than 0.5 and collectively account for more than 93.6 percent of the variance. Therefore, we include nine variables in our base model.

3.2.2 General Linear Model

We determined that the general linear model, with variables described in Table 4, may prudently include 9 relevant factors. We run the general linear regression with all 13 variables, and observe the p-value of each variable. We continued to rerun the model until all factor coefficients are statistically significant, excluding variables with a p-values greater than 0.1.[1] Among the thirteen variables, the capital intensity ratio, CASHSA, EBITDASA, and CFFOSA had p-values greater than 0.1 and are deleted from the linear model. After removing these variables and rerunning the general linear model, all estimates were statistically significant at or better than 0.05 significance levels. Therefore, our final model uses the nine remaining variables from Panel B, which we use for the remainder of this paper.

3.2.3 Correlation Matrix and VIF analysis

We used a correlation matrix to indicate the presence of possible multicollinearity. The correlation matrix (Table 5) indicates that most pairwise correlations are between -0.5 and 0.5, which suggests that there is no strong correlation among the variables. The only strong correlations (greater than 0.5 or less than -0.5) are found among capital intensity ratio, CASHSA, EBITDASA, and CFFOSA, which is consistent with the findings of the general linear model.

We also use variance inflation factors (VIFs) in Table 6 to ensure that multicollinearity is not problematic in our model. VIFs for all the twelve variables indicate a similar pattern with correlation matrix, that multicollinearity does not cause problems. Capital intensity ratio, CASHSA, and EBITDASA have the highest VIF values.

Based on these results, we identify nine variables (Altman Z score, Ohlson O score, EDF, CASHTA, EBITDATA, CFFOTA, firm size, market return, and market volatility) to include in our subsequent analysis.

[1] Principal component analysis is a dimensionality-reduction method used to reduce the dimensionality of data sets by transforming a larger set of variables into a smaller one that still contains most of the information in the large set. The principal components or factors are less interpretable and don't have any real meaning since they are constructed as linear combinations of the initial variables. Thus, we initially use all thirteen variables and iteratively reduce the variables until all factors are statistically significant.

3.3 Probit Logit Model

We examine firm default propensities based on firm-specific predictors, market conditions, and distance-to-default model using a probit logit model.

The probit logit model is shown below.

$$\Pr[Y_i = 1|X_{1i}, \dots, X_{Ki}; \beta_0, \dots, \beta_K] = \phi(\beta_0 + \sum_{k=1}^{K} \beta_k X_{ki})$$

In the probit logit model, the dependent variable is a dummy variable of 1 if the firm defaults after the fiscal year-end. Otherwise, if no default occurs, the dependent variable dummy equals 0. The predictor variables are Altman Z score, Ohlson O score, EDF, CASHTA, EBITDATA, COFFOTA, firm size, market return (S&P 500) and market annualized volatility.

Results in Table 7 indicate that for our logistics model, most variable increases increase predictability of firm bankruptcy, except for CFFOTA.

Coefficients for Altman Z score, CASHTA, EBITDATA, market volatility and market return are significantly negative, indicating that firms with increased cash holdings and EBITDA or larger size tend to be less likely to end in bankruptcy. Firms are less likely to go bankrupt when the market is bullish. The coefficient of market volatility (standard deviation) is -153.8 is significant at the 1% level. However, market volatility does not significantly contribute to predicting bankruptcy using the probit logit model.

The coefficient for our main variable, expected distance to failure/bankruptcy, EDF, is 5.5775 and significant at the 1% level. Also, the odds ratio is statistically significant as well. In the logit model, the odds ratio refers to the association between the variable and the outcome. An odds ratio of greater than 1 indicates that the variable is associated with higher odds of the event (bankruptcy) occurring. If the odd ratio equals 1, the variable is unassociated with the event's occurrence, and an odd ratio less than 1 indicates that the variable is associated with a lower event probability.

In our study, both coefficients for EDF and odds ratio support our hypothesis that EDF effectively and significantly predicts the occurrence of bankruptcy events.

3.4 Cox Proportional Hazard Model

The Cox proportional hazard model indicates similar results as compared to the above probit logit model in measuring the contribution and efficacy of EDF. The results are shown in Table 8.

The coefficient for CFFO is negative but is not statistically significant. The coefficient for Ohlson O score is positive but only significant at the 0.1 level. On the other hand, coefficients for Altman Z score, CASHTA, EBITDATA, market volatility, firm size, and market return are all statistically significantly negative, with hazard ratios less than 1. This suggests that the Altman Z score, CASHTA, EBITDATA, market volatility, firm size, and market return are reliable predictors of firm bankruptcy. Firms with higher cash reserves, better profitability, larger size, and with bullish market conditions are less vulnerable to default.

The coefficient of 5.6576 for EDF is statistically significant at the 1% level, and a hazard ratio of 286.451 strongly confirms our initial hypothesis that EDF is an effective predictive factor, our strongest predictor of bankruptcy. As the EDF value increases, the probability of the firm filing for bankruptcy at the end of the fiscal year also increases.

3.5 Decomposing the adjusted R^2

We apply the partial adjusted R square method to indicate each variable's contribution to the model's coefficient of variation (R^2) or each factor's ability to explain total OLS variation. Our results in Table 9 indicate that 65.32% of the adjusted R2 is explained by the expected default frequency (EDF). Ohlson O score accounts for 27.95% of the adjusted R2, while Altman Z score accounts for 4.23%. Based on these findings, we can infer that EDF is the most significant contributor to predicting actual bankruptcy events, alongside other default risk proxies and market condition proxies.

4. Conclusions

The main purpose of this study is to develop a model to predict firm default and/or bankruptcy and examine the efficacy of Merton's (1974) distance to default model as simplified by Bharath and Shumway (2008) to forecast a firm's Expected Default Frequency (EDF) or probability of defaulting on debt obligations with likely bankruptcy. Merton's model, further developed by the KMV corporation, is based on the Black-Scholes asset pricing model. We apply the simplified Bharath and Shumway model by relating it with Cox's proportional-hazards model (Cox, 1972) to forecast firm-specific expected default frequency (EDF) or probability of default. Using principal components-factor analysis, we find that EDF output from the Merton/Bharath-Shumway Model is our most important orthogonal variable in predicting firm default and potential bankruptcy. We also measure how much of the total variation is explained in an adjusted R squared decomposition analysis.

Multiple models were employed in this paper to analyze the predictability of the expected default frequency (EDF) or risk of default that may lead to a firm's bankruptcy. The results from both the probit logit model and Cox Proportional Hazard model support the assumption that EDF is a valid and effective variable for predicting the occurrence of bankruptcy events. The squared partial correlation analysis further supports this conclusion.

References

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589-609.

Black, F., & Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy, 81(3), 637-654.

Bharath, S. T., & Shumway, T., 2004. Forecasting default with the KMV-Merton model. In AFA 2006 Boston Meetings Paper.

Chava, S., & Jarrow, R. A., 2004. Bankruptcy prediction with industry effects. Review of Finance, 8(4), 537-569.

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

Goudie, A. W., & Meeks, G., 1991. The exchange rate and company failure in a macro-micro model of the UK company sector. The Economic Journal, 101(406), 444-457.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G., 2004. Assessing the probability of bankruptcy. Review of Accounting Studies, 9, 5-34.

Jarrow, R. A., & Turnbull, S. M., 1995. Pricing derivatives on financial securities subject to credit risk. The Journal of Finance, 50(1), 53-85

Ledwon, A. V., & Jäger, C. C., 2020. Cox proportional hazards regression analysis to assess default risk of german-listed companies with industry grouping. ACRN Journal of Finance and Risk Perspectives, 9.

Lennox, C., 1999. Identifying failing companies: a re-evaluation of the logit, probit and DA approaches.

Journal of Economics and Business, 51(4), 347-364.

Ohlson, J. A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 109-131.

Reid, N., & Cox, D. R., 2018. Analysis of survival data. Chapman and Hall/CRC.

Tinoco, M. H., & Wilson, N., 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. International Review of Financial Analysis, 30, 394-419.

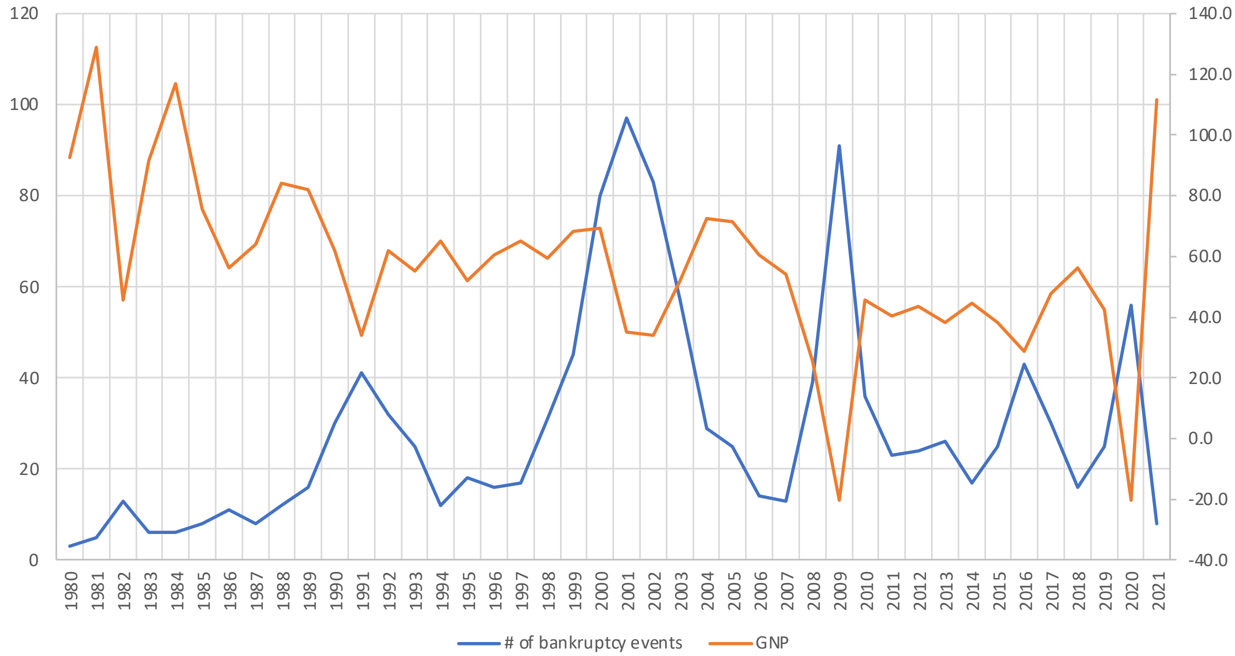

Figure 1. Failed firms sorted by year

We sort the failed firms by year and compare the pattern with the historical GNP from 1979 through 2021.

Table 1
Summary Statistics

Variable	N	Mean	Median	Std	Min	Max	Skewness
Altman Z score	86,049	4.3772	3.1611	7.0205	-74.4081	341.4542	12.9156
Ohlson O score	71,548	-0.0169	-0.0200	2.4853	-102.0440	186.1744	4.4648
Capital Intensity Ratio	87,987	3.9652	1.0694	96.7315	0.0481	17765.1897	122.4789
EDF	87,987	0.0317	0.0000	0.0893	0.0000	1.0000	3.9519
CASHTA	87,987	0.1546	0.0822	0.1852	0.0000	0.9969	1.8976
EBITDATA	87,987	0.1151	0.1235	0.1346	-2.7788	10.6985	3.1776
CFFOTA	77,144	0.0805	0.0864	0.1158	-1.4149	8.4674	3.6820
CASHSA	87,987	1.6108	0.0877	40.6795	0.0000	5712.3267	70.5881
EBITDASA	87,987	-0.3614	0.1259	11.9037	-987.5000	4.1491	-44.9059
CFFOSA	77,144	-0.2666	0.0919	12.4545	-846.7624	1660.1129	27.5422
Log Firm size	87,981	6.6455	6.4899	1.8676	-6.9078	14.4916	0.3554

Table 2
Failed firm sorted by sector

SIC Group	Number of Cases
A: Agriculture Production Crops	2
B: Mining	131
C: Construction	23
D: manufacturing	363
E: Tranportation, Communication, Electric, Gas	199
F: Wholesale Trade	40
G: Retail Trade	147
H: Finance, Insurance, and Real Estate	163
I: Services	149
Grand Total	1218

Table 3
Principal Component Analysis

Panel A: Principal Conponent Analysis

Factor	Eigenvalue	Difference	Proportion	Cummulative
Factor 1	2.9912	0.5067	0.2301	0.2301
Factor 2	2.4845	0.9969	0.1911	0.4212
Factor 3	1.4876	0.165	0.1144	0.5356
Factor 4	1.3226	0.3712	0.1017	0.6374
Factor 5	0.9514	0.1945	0.0732	0.7106
Factor 6	0.7569	0.0211	0.0582	0.7688
Factor 7	0.7358	0.1149	0.0566	0.8254
Factor 8	0.6209	0.05	0.0478	0.8731
Factor 9	0.5709	0.1708	0.0439	0.9171
Factor 10	0.4001	0.0392	0.0308	0.9478
Factor 11	0.3609	0.147	0.0278	0.9756
Factor 12	0.2139	0.2139	0.0165	0.9921
Factor 13	0.1033	0	0.0079	1.0000

Panel B: Variance Explained by Each Factor

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
2.9912	2.4845	1.4876	1.3226	0.9514
Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
0.7569	0.7358	0.6209	0.5709	0.4001
Factor 11	Factor 12			
0.3609	0.2139			

Table 4
General Linear Analysis

Panel A: GLM (13 variables)						
Variables	**Estimates**	**p value**	**Type I Error**	**p value**	**Type III Error**	**p value**
Altman Z Score	0.00011	0.0021	0.4907	<.0001	0.03998	0.0021
Ohlson O Score	0.00115	<.0001	3.2114	<.0001	0.3355	<.0001
Capital Intensity Ratio	0.000001	0.6960	0.0024	0.4488	0.0006	0.6960
EDF	0.122	<.0001	7.3992	<.0001	6.6270	<.0001
CASHTA	-0.0049	0.0011	0.0152	0.0577	0.0446	0.0011
EBITDATA	-0.1233	<.0001	0.0691	<.0001	0.0723	<.0001
CFFOTA	0.0067	0.0572	0.0114	0.0668	0.0153	0.0572
CASHSA	-0.00001	0.2061	0.0078	0.1747	0.0067	0.2061
EBITDASA	-0.00003	0.4535	0.00005	0.9100	0.0024	0.4535
CFFOSA	0.00004	0.0780	0.0142	0.1000	0.0131	0.078
Market Volatility	-0.0269	<.0001	0.1328	<.0001	0.1511	<.0001
Firm Size	-0.0003	0.0218	0.0231	0.0194	0.0222	0.0218
Market Return	-0.0042	0.0257	0.0210	0.0257	0.0210	0.0257

Panel B: GLM (9 variables)						
Variables	**Estimates**	**p value**	**Type I Error**	**p value**	**Type III Error**	**p value**
Altman Z Score	0.0001	0.0031	0.5004	<.0001	0.0374	0.0031
Ohlson O Score	0.0012	<.0001	3.3007	<.0001	0.3229	<.0001
EDF	0.1237	<.0001	7.6375	<.0001	6.8788	<.0001
CASHTA	-0.0051	0.0008	0.0178	0.0412	0.0481	0.0008
EBITDATA	-0.0086	<.0001	0.0677	<.0001	0.0649	<.0001
CFFOTA	0.00004	0.0248	0.0221	0.0228	0.0215	0.0248
Market Volatility	-0.2695	<.0001	0.1353	<.0001	0.1521	<.0001
Firm Size	-0.0003	0.0204	0.0238	0.0182	0.0229	0.0204
Market Return	-0.004	0.0346	0.019	0.0346	0.019	0.0246

Table 5
Correlation Matrix

Variables	Altman Z score	Ohlson O score	Capital Intensity Ratio	EDF	CASHTA	EBITDATA	CFFOTA	CASHSA	EBITDASA	CFFOSA	Market volatility	Firm size	Market return
Altman Z score	1	-0.4087	0.0308	-0.1616	0.3185	0.1531	0.1456	0.0741	-0.0552	-0.0319	-0.0156	0.1422	0.0285
Ohlson O score		1	0.0187	0.3576	-0.1989	-0.4223	-0.4791	0.0233	-0.0597	-0.1071	0.0402	-0.3316	0.0064
Capital Intensity Ratio			1	0.0121	0.0557	-0.0628	-0.0516	0.6480	-0.5671	-0.3077	-0.0028	-0.0006	-0.0022
EDF				1	-0.0745	-0.2088	-0.1793	0.0158	-0.0123	-0.0117	0.1691	-0.3344	-0.1179
CASHTA					1	-0.2920	-0.1794	0.1230	-0.1393	-0.0896	0.0042	0.0487	-0.0283
EBITDATA						1	0.4694	-0.1002	0.1638	0.1181	-0.0332	0.2107	0.0238
CFFOTA							1	-0.0804	0.1423	0.1502	-0.0080	0.2482	-0.0051
CASHSA								1	-0.8662	-0.3780	-0.0011	-0.0010	-0.0007
EBITDASA									1	0.5159	0.0034	0.0102	0.0021
CFFOSA										1	0.0000	0.0106	0.0033
Market volatility											1	-0.073	-0.3209
Firm size												1	0.0446
Market return													1

Table 6
Variance inflation Factor (VIF) Analysis

Variables	Variance Inflation
Altman Z Score	1.3209
Ohlson O Score	1.8501
Capital Intensity Ratio	2.2680
EDF	1.2813
CASHTA	1.4020
EBITDATA	2.8016
CFFOTA	2.6653
CASHSA	5.5921
EBITDASA	4.8914
CFFOSA	2.3819
Market Volatility	1.1143
Firm Size	1.2238
Market Return	1.0889

Table 7
Probit Logit Results

Logit Regression				
Variables	Maximum likelihood Estimates	Standard error	p vlaue	Odds Ratio
Altman Z score	-0.0708	0.0150	<.0001	0.932
Ohlson O score	0.0517	0.0225	0.0219	1.053
EDF	5.5775	0.3213	<.0001	264.401
CASHTA	-3.0679	0.5584	<.0001	0.047
EBITDATA	-1.0957	0.2745	<.0001	0.334
CFFOTA	-0.0002	0.0067	-0.9770	1.000
Market volatility	-3.3823	1.1517	0.0033	0.034
Firm size	-0.2726	0.0357	<.0001	0.761
Market return	-0.5586	0.4129	0.1761	0.572
	Chi-Square	DF	p value	
Likelihood Ratio	1038.6541	9	<.0001	
Score	2645.9487	9	<.0001	
Wald	1072.3159	9	<.0001	

Table 8
Cox Proportional Hazard Model

Cox Proportional Hazard Model				
Variables	Estimates	Standard error	p vlaue	Hazard Ratio
Altman Z score	-0.0703	0.0104	<.0001	0.932
Ohlson O score	0.0078	0.0046	0.0924	1.008
EDF	5.6576	0.2820	<.0001	286.451
CASHTA	-2.0403	0.5766	0.0004	0.130
EBITDATA	-0.7079	0.1467	<.0001	0.493
CFFOTA	-0.0012	0.0042	0.7822	0.999
Market volatility	-2.8957	1.0619	0.0064	0.055
Firm size	-0.3195	0.0236	<.0001	0.727
Market return	-1.0064	0.4091	0.0139	0.366
	Chi-Square	DF	p value	
Likelihood Ratio	1123.8570	9	<.0001	
Score	3352.7297	9	<.0001	
Wald	1596.3328	9	<.0001	

Table 9

Decomposition of Adjusted R squared

Variables	% of adjust R2 explained
Altman Z score	4.23%
Ohlson O score	27.95%
EDF	65.32%
CASHTA	0.16%
EBITDATA	0.59%
CFFOTA	0.19%
Market volatility	1.19%
Firm size	0.21%
Market return	0.17%

Chapter 2

Short selling as an indicator of default risk

1. Introduction

Default risk may be defined as financial deteriorate where a firm's cash flows fail to cover its operating expenses and its debt service cost, including interest payments and repayment of principal at maturity. Detection and prediction of financial deterioration is a major component of a firm’s default risk. We find that short-interest levels may serve as an early warning signal for a firm's deteriorating financial health and indicator of default risk.

Merton (1974) develops a distance to default (DD) model that quantifies a firm’s expected default frequency (EDF). Subsequently, Bharath and Shumway (2008) simplify Merton's model, concluding that his DD model is a useful measure for forecasting default. Thus, we apply Bharath and Shumway's (2008) approach to measure expected default frequency EDF.

We extend relevant literature by examining whether short-selling levels are predictors of EDF changes. If short-selling levels are effective in predicting or forecasting EDF changes, monitoring shorting levels may proxy as another early warning signal for investors, managers, and regulators regarding a firm’s deteriorating financial condition and increased risk of default or bankruptcy.

In theory, increases in a firm’s default risk motivates investors to demand higher expected returns for a firm's shares and increasing cost of its debt, hence reducing both the value of its stock and debt. Related to this relationship between default risk and value, Clark and Weinstein (1983) conduct an event study using a sample of US bankruptcy cases between 1938 and 1979, finding that investors experience significant stock price decreases during the month a bankruptcy occurs, and further major losses occur within a three trading-day-window surrounding the bankruptcy filing date.[1] Clark and Weinstein's conclusion indicates that bankruptcy filings convey important unanticipated information to the market. Given significant price changes resulting from bankruptcy filing dates, the ability for investors to predict increases in a firm's default risk (probabilities to file bankruptcy) by monitoring shorting levels, may offer profitable opportunities.

Prior literature documents that short sellers generally are informed, sophisticated investors. Aitken, Frino, McCorry, and Swan (1998), Boehmer, Jones, and Zhang (2008), Diether, Lee, and Werner (2009), and Cohen, Diether, and Malloy (2007) prove that short sales are predictors of negative returns.

Further studies, Boehmer and Wu (2013), demonstrate that short sales improve price discovery, and Engelberg, Reed, and Ringgenberg (2012) suggest that short sellers are skilled information processors surpassing just "informed" traders.

Another study by Christophe, Ferri, and Hsieh (2010) shows that firms with high pre-announcement abnormal shorting levels have lower stock returns over the six months subsequent to a debt downgrade as compared to those with low pre-announcement abnormal short selling levels. Further, Avromov, Chordia, Jostova, and Philipov (2007) find that short sellers exploit profit opportunities provided by downgrade announcements, and Dechow, Hutton, Meulbroek, and Sloan (2001) indicate that short sellers generate positive returns by focusing on deteriorating fundamentals.

Henry, Kisgen, and Wu (2015) indicate that short interests tend to be 40% higher in the month prior to credit downgrades than for one-year prior. Overall, we find a consensus that short

[1] Clark and Weinstein (1983) indicate that a bankruptcy filing increases the likelihood that a firm's shares will become worthless. Iskandar-Dattab and S.Datta, (1995) report that bankruptcy announcements lead to significant negative returns .

sellers may have the ability to forecast defaults based on publicly available information better than the marginal investor and profit thereby.

As indicated above, we apply the modified distance to default model, developed by Merton (1974) and simplified by Bharath and Shumway (2008), to measure default risk.

Merton models a firm's equity as a European call option on the firm's underlying-inherent value with a strike price equal to the firm's debt's face value. Since both firm value and implied volatility are difficult to observe, Merton simplifies the model by using observable variables such as the market value of equity and stock price volatility. The Merton distance-to-default model assumes that the normal cumulative distribution of distance-to-default measures a firm’s expected default frequency (EDF).

Many other studies use the Merton distance-to-default model.[2] Bharath and Shumway (2008) test the Merton distance to default model's accuracy and find that Merton’s model is useful for forecasting defaults. Following Bharath and Shumway (2008), Brogaard, Li, and Xia (2017) analyze stock liquidity's impact on default. Guo and Wu (2019) also explore the relationship between short interest and default risk and conclude that short interest levels tend to be the highest for financially distressed firms.

We extend the literature by focusing on changes of EDF and examine short interest levels ability to predict changes in EDF. We measure the ability to predict firms moving from a low default risk group to a high default risk group, firms remaining in the same default risk level or moving from a high default risk group to a low default risk group. We find that changes in short-sale interests precedes and may predict changes in EDF.

---

[2] Vassalou and Xing (2004) use the model to compute default measures for individual firms and assess the effect of default risk on equity returns. These authors argue that the use of equity data to measure default probability is an important innovation of the Merton (1974) model.

We find that short-interest levels demonstrate significant forecasting accuracy vis-à-vis securities' movements into and out of EDF deciles.

Also, we investigate intra-year change of EDF and find supportive evidence that short interest change between the first 6 months and second 6 months are significantly related to EDF changes.

2. Data, methodology, and summary statistics

2.1 Data

We collect firm-specific financial data from Compustat and monthly stock returns from CRSP for the NYSE, AMEX, and NASDAQ stocks for 2007 through 2018. Since financial or regulated industries are subject to regulation and special capital requirements, we exclude firms with SIC codes between 6000 and 6999.[3] Also, we drop firms with missing equity market capitalization. We compute the total capitalization or value of each firm *i* for fiscal year *t* as the sum of total debt and equity market-value ($Equity_{i,t}$), measured as the product of shares outstanding and share price at the end of the fiscal year. Total debt of firm *i* in fiscal year *t* ($Debt_{i,t}$) is the sum of current liabilities and one-half of long-term debt at the fiscal year-end. $\sigma_{E_{i,t}}$, the stock volatility of firm *i* in fiscal year *t*, is calculated using monthly equity returns for fiscal year *t*. Also, we calculate the log price relative for each month. The return for year *t*, $r_{i,t}$, is the sum of the twelve-monthly returns.

Our shorting variables are obtained from Market Securities Finance daily lending data for 2007 through 2019 from which we compute monthly means from the daily data. Participants in the securities lending market necessary for short selling, include prime brokers, custodians, asset

[3] In finance research, it is standard practice to exclude financial or regulated industries as they are quite different from industries in terms of operations and regulations. Thus, we exclude these finance firms, which have SIC 4 digit code from 6000 to 6999.

managers, and hedge funds, report these lending data. *Available Value* (*AV*) is the inventory available to lend (based on share value) and, hence, to short.

Our proxy for short interest, *Borrowed Value* (*BV*), is the total debt on loan, net of double counting, (based on share value). *Utilization* is the *Borrowed Value* divided by the *Available Value* (another proxy for short interest). *Borrowing Fee* is the value-weighted average borrowing fee for all transactions. We measure all Market variables as of the end of the first month of year *t* (except for Table 9).

2.2 Methodology

We measure default probabilities using a simplified Merton distance-to-default model modified by Bharath and Shumway (2008). Expected default frequency (*EDF*) is the probability that a firm's total assets will be insufficient to cover its total debt. *EDF* is measured as the cumulative standard normal distribution of distance-to-default (*DD*).

DD is

$$\frac{\ln\left(\frac{Firm\ Value_{i,t}}{Debt_{i,t}}\right)+\left(r_{i,t-1}-0.5\sigma^2_{V_{i,t}}\right)*T_{i,t}}{\sigma_{V_{i,t}}*\sqrt{T_{i,t}}} \tag{1}$$

$$EDF = N(-DD_{i,t}) \tag{2}$$

$$\sigma_{D_{i,t}} = 0.05 + 0.25 * \sigma_{E_{i,t}} \tag{3}$$

$$\sigma_{V_{i,t}} = \frac{Equity_{i,t}}{Firm\ Value_{i,t}} * \sigma_{E_{i,t}} + \frac{Debt_{i,t}}{Firm\ Value_{i,t}} * \sigma_{D_{i,t}} \tag{4}$$

We use an alternative method, provided by Bharath and Shumway (2008), to calculate debt volatility, $\sigma_{D_{i,t}}$. In Equation (3), 0.05 (5%) represents term-structure volatility. Since firms with higher debt default probabilities demonstrate higher equity volatilities, 0.25* $\sigma_{E_{i,t}}$ in Eq. (3) measures default risks associated with equity volatility. $T_{i,t}$ is set to 1 because we use annual financial data.[4]

Notice that for fiscal year *t*, the *EDF* values are for fiscal year-end, and *Utilization* and *Borrowing Fee* values are for the first month of the year (except for Table 9). Hence, *Utilization* and *Borrowing Fee* for fiscal year *t* are lagged relative to *EDF*.

All statistical tests are significant at the 1% level unless otherwise stated.

2.3 Summary statistics

Figure 1 shows the distribution by decile of expected default frequency (*EDF*) for the 37,938 firm-year observations from 2007 through 2018. Over 80% of the sample has an *EDF* of almost 0.00. The mean *EDF* for the ninth and tenth deciles are 0.02 and 0.84, respectively. Figure 2 illustrates the mean and standard deviation of *EDF* by year. Due to the GFC, it is not surprising that *EDF* has its highest means of 0.12 in 2008 and 0.24 in 2009. This figure also shows a positive relation between default risk and standard deviation. The years with the highest mean *EDF* also have the highest standard deviations of *EDF*. Figure 3 shows the histogram of *Utilization* by decile. The distribution is concave up, with about 60% of observations having less than 10% *Utilization*.

Table 1 provides summary statistics for the entire sample. The mean and standard deviation (in parentheses) of *EDF*, *Utilization*, and *Borrowing Fee* are 0.09 (0.26), 18.59% (24.68%), and 2.84 (9.46), respectively. The statistics for *Equity*, *Debt*, *Return*, *Volatility* also look reasonable.

3. Research question

We explore the relation between default risk and short selling. Specifically, we investigate whether short sellers can predict changes in *EDF*. Many previous studies find that short sellers are

---

[4] Li and Spahr (2023) have a working paper that examines The Efficacy of Expected Default Frequency (EDF) Models in Predicting U.S. Public Firm Defaults that confirms the accuracy and efficiency of using EDF as a measure of firm defaults/bankruptcies.

exceptionally well informed. Boehmer, Jones, and Zhang (2008) report that institutional short sellers make decisions based on fundamental financial information, making the stock price more efficient. *EDF* is calculated from known fundamental financial information and predicts the probability that the firm's total assets will not be sufficient to cover its total debt. Of course, short-sellers also predict firms' likelihood of default. Therefore, it is not surprising that short selling is higher for firms with higher *EDF*s, shown in the Appendix.

However, we test the more interesting question of whether short interest predicts changes in *EDF*. Engelberg, Reed, and Ringgenberg (2012) report that short sellers' information advantage partly arises because they are better than other market participants at processing public information. We hypothesize that shorting is a leading indicator of deteriorating default probability.

4. Results

4.1 Shorting and default probabilities

4.1.1 Portfolio sort analysis

For each year, we form portfolios by sorting stocks into deciles by their level of *EDF*. Table 2 presents statistics for *EDF* for deciles 10, 9, and 1-8 combined. Mean *EDF* is 0.8384 for decile 10, 0.0168 for decile 9, and 4.57E-10 for deciles 1-8. We present an additional analysis of the mean *Utilization* for each *EDF*-portfolio decile for all years combined and by year in the Internet Appendix. Mean *Utilization* increases monotonically with *EDF*. For all years, the mean *Utilization* for the highest-default-risk stocks is 11.12% higher than the mean *Utilization* for the lowest-default-risk stocks. This difference is statistically significant.

4.2 Shorting and increasing default risk

4.2.1 Portfolio sort analysis

We consider our research question: Shorting is a leading indicator of deteriorating default probabilities. To investigate, we compare *Utilization* for companies' experiencing an increase versus a decrease in default risk. In Table 3, we find that for fiscal year *t Utilization* anticipates changes in *EDF*.[5] For 2008-2018, mean *Utilization* is 19.76% for companies experiencing an increase in default risk and 17.07% for companies undergoing a decrease in default risk. These values are statistically significantly different. We find similar results when testing each year separately for 2008, 2011, 2012, 2013, 2014, 2017, 2018. We believe that the market dislocation following the financial crisis is responsible for the lack of statistical significance in 2009 and 2010.

Given that short sellers are particularly interested in the companies expected to have the largest *EDF* increase, we sort companies into deciles by the change in *EDF* (see Table 4). We expect that shorting will be higher for companies in the highest decile (the largest increase in default risk). For all years, the mean *Utilization* for decile 10 (the highest-increase-in-default-risk stocks) is 7.23% higher than the mean *Utilization* for deciles 1-8 and 2.87% higher than decile 9. These values are statistically significantly different at the 1% level. We find consistent results when testing each year separately from 2011 to 2018. For 2008-2010, we do not find significance, likely due to the market dislocation following the financial crisis.

[5] Recall that $Utilization_t$ is measured the first month of the fiscal year and *EDF* at the end of the fiscal year of year *t*. Therefore, for companies with a December 31 fiscal year-end, $Utilization_t$ is measured in January and *EDF* on December 31 (based on annual results) of year *t*.

4.2.2 Cross-sectional regression analysis

To extend our portfolio sort analysis, we estimate the following regression:

$$\Delta EDF_{i,t} = b_0 + b_1\, Utilization_{i,t} + b_2\, Ln(Equity)_{i,t} + b_3\, Ln(Debt)_{i,t} + b_4\, Stock\ volatility_{i,t} + \text{Year fixed effects} + \text{Firm fixed effects} + \varepsilon_{i,t} \quad (1)$$

where, $\Delta EDF_{i,t} = EDF_{i,t} - EDF_{i,t-1}$. $Utilization_{i,t}$ is the monthly average short interest in the first month after the fiscal year ends. Ln(Equity)$_{i,t}$ is the market capitalization of the firm *i* at the end of fiscal year *t*. Ln(Debt)$_{i,t}$ is the log value of debt of the firm *i* at the end of fiscal year *t*. Stock volaitlity$_{i,t}$ refers to the annualized standard deviation of stock return within the fiscal year.

Table 5, Columns 1-3, shows results for all deciles. We find a positive and significant relation between shorting as measured by *Utilization* and *ΔEDF* for all three regressions for all deciles (Columns 1-3) and deciles 9 and 10, combined (Columns 4-6). For all firms, almost all the coefficients are statistically significant. However, in Columns 4-6, for firms experiencing the most default risk increase, the coefficient of short utilization remains significant but the stock volatility is no longer a determinant.

The regression analysis results and the portfolio analysis support the view that shorts predict changes in *EDF*.

4.3 Shorting and movement into and out of the highest EDF group

To further our research question analysis, we ask the following additional question: Does the lag of *Utilization* anticipate movement into and out of the highest *EDF* group?

In Table 6, we define four mutually exclusive and exhaustive categories of companies: (1) *HNever* are not in decile 10 for EDF_t or EDF_{t-1}; (2) *HLeaver* are in decile 10 for EDF_{t-1} but not for EDF_t; (3) *HMover* are in decile 10 for EDF_t but not for EDF_{t-1}; and (4) *HStayer* are in decile 10 for EDF_t and EDF_{t-1}. We compare the *Utilization* for each of these groups and predict that it will be highest for *HMover* and *HStayer*.

We find that *HStayer* has the highest *Utilization* at 26.58%, which is statistically higher than *HLeaver* (21.30%). Further, *HMover* has the second-highest *Utilization* at 24.07%, which is statistically higher than *HLeaver*. Shorts predict which companies will stay in the highest *EDF* decile and which companies will join the highest *EDF* decile.

To further investigate shorts' ability to predict movement into the highest *EDF* decile, we conduct a two-way portfolio sort. In Table 7, stocks are sorted by lagged *EDF* decile and then by the dummy variable *HMover* (=1, if in decile 10 for *EDF* but not for lagged *EDF*; =0, otherwise). We expect that shorts' will concentrate on firms expected to have increasing default risk (i.e., *HMover* = 1). We find that for each group of lagged *EDF* decile (except for decile 2), *HMover* = 1 is higher than *HMover* = 0. For lagged *EDF* decile 1, *Utilization* of 28.75% for *HMover* = 1 as compared to 13.76% for *HMover* = 0. In the lagged EDF decile 9, we find *Utilization* of 24.95% for *HMove*r = 1 as compared to 20.82% for *HMover* = 0.

Table 8 uses a logit model to test short interests' ability to predict movement into the highest *EDF* group. In column 1, we find that *Utilization* predicts movement into the highest *EDF* group (defined by the dummy variable *HMover*). This relation remains statistically significant when we control for lagged EDF decile (*Decile10*), *Equity*, *Debt*, and *Volatility* (columns 2-4). These results support the view that *Utilization* anticipates movement into and out of the highest *EDF* group.

4.4 Intra-fiscal-year change in Utilization

In Table 9, we test whether changes in movements into the highest short interest decile anticipate changes in default risk. Let $\Delta EDF = EDF_{i,t} - EDF_{i,t-1}$ where t is a fiscal year for firm i. *Dummy_Util* = 1 if the firm's *Utilization* at fiscal year-end moves from decile 1-9 in fiscal year *t-*

1 into decile 10 in fiscal year *t* and 0 otherwise. *ΔUtilization* is the change in *Utilization* between the first and second six months of fiscal year *t*. We estimate *EDF* using financial information released a few weeks after the fiscal year-end. Short interest is released with significantly higher frequency. Hence, *ΔUtilization* is more current than *EDF*. Our results show a significant and positive relation between intra-fiscal-year changes in short interest and subsequent changes in *EDF*.

5. Conclusion

We investigate whether shorts are able to predict changes in expected default risk. Our proxy for short interest, *Utilizaton*, is the ratio of shares shorted to shares available to short. We measure expected default risk (EDF) using the distance to default model originated by Merton (1974) and simplified by Bharath and Shumway (2008). We have 37,938 firm-year observations from 2007 through 2018. There were about 1,600 instances in which firms moved from decile 1-9 of EDF in one year to decile 10 in the next. *Utilization*, which is lagged, was 36.6% higher for these firms than for the 28,200 firms that did not move into decile 10 of EDF. For the twelve instances in which firms moved from decile 1 of EDF in year t-1 to decile 10 in year t, the *Utilization* in year t-1 was more than twice as high in year t-1 as for other firms. Therefore, we conclude that Utilization anticipates movements into and out of the highest *EDF* decile.

We also investigate the relation between change of *Utilization* and change of *EDF* and find a significantly positive relation, indicating that changes in *Utilization* are a predictor of movement into the highest *EDF* decile. Exploring further, we compare the *Utilization* change between the first and second six months of a fiscal year and find that the intra-fiscal-year change in Utilization is a leading predictor of firms' subsequent change in *EDF*.

Reference

Aitken, M.J., Frino, A., McCorry, M.S., Swan, P.L., 1998. Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange. Journal of Finance 53, 2205-2223.

Bharath, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. Review of Financial Studies 21, 1339-1369.

Boehmer, E., Jones, C.M., Zhang, X., 2008. Which shorts are informed? Journal of Finance 63, 491-527.

Boehmer, E., Wu, J., 2013. Short selling and the price discovery process. Review of Financial Studies 26, 87-322.

Brogaard, J., Li, D., Xia, Y., 2017. Stock liquidity and default risk. Journal of Financial Economics 124, 486-502.

Christophe, S.E., Ferri, M.G. Hsieh, J., 2010. Informed trading before analyst downgrades: Evidence from short sellers. Journal of Financial Economics 95, 85-106.

Clark, T. A., & Weinstein, M. I., 1983. The behavior of the common stock of bankrupt firms. *The Journal of Finance*, *38*(2), 489-504.

Cohen, L., Diether, K. B., Malloy, C. J., 2007. Supply and demand shifts in the shorting market. Journal of Finance 62, 2061-2096.

Dechow, P.M., Hutton, A.P., Meulbroek, L., Sloan, R.G., 2001. Short-sellers, fundamental analysis, and stock returns. Journal of Financial Economics 61, 77-106.

Diether, K.B., Lee, K.H., Werner, I.M., 2009. Short-sale strategies and return predictability. Review of Financial Studies 22, 575-607.

Engelberg, J.E., Reed, A.V., Ringgenberg, M.C., 2012. How are shorts informed?: Short sellers, news, and information processing. Journal of Financial Economics 105, 260-278.

Guo, X., & Wu, C. (2019). Short interest, stock returns and credit ratings. *Journal of Banking & Finance*, *108*, 105617.

Henry, T. R., Kisgen, D. J., & Wu, J. J., 2015. Equity short selling and bond rating downgrades. *Journal of Financial Intermediation*, *24*(1), 89-111.

Iskandar-Dattab, M. E., and Datta, S., 1995, The information content of bankruptcy filing on security holders of the bankrupt firm: An empirical investigation, Journal of Banking and Finance 19, 903-919.

Li, H. and Spahr, R. W., 2023, The Efficacy of Expected Default Frequency (EDF) Models in Predicting U.S. Public Firm Defaults. working paper University of Memphis.

Mensing, R., 2015, Informational efficiency and market reaction to bankruptcy announcement. Masters Thesis, University of Amsterdam.

Merton, R.C., 1974. On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29, 449-470.

Vassalou, M., Xing, Y., 2004. Default risk in equity returns. Journal of Finance 59, 831-868.

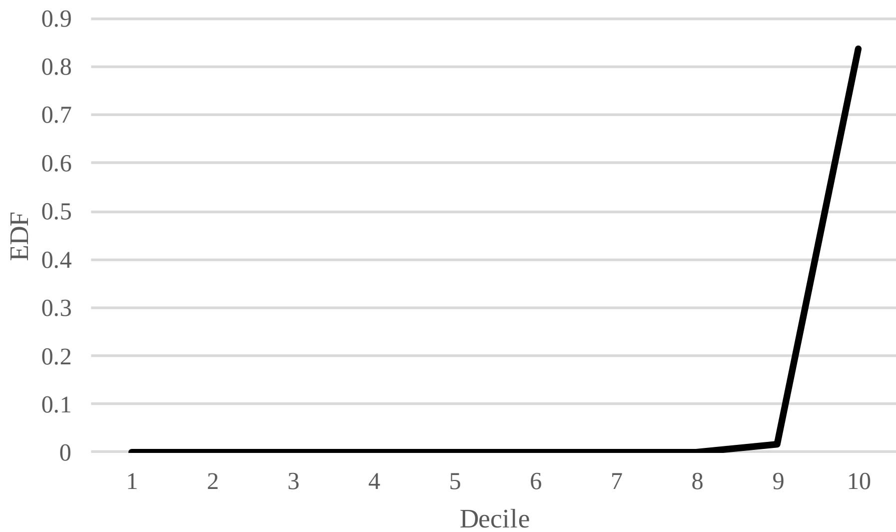

Figure 1. *EDF*, by decile. We present the mean for firm-year observations for *EDF* from 2007 to 2018 (n = 37,938). We measure *EDF* at each firm's fiscal year-end.

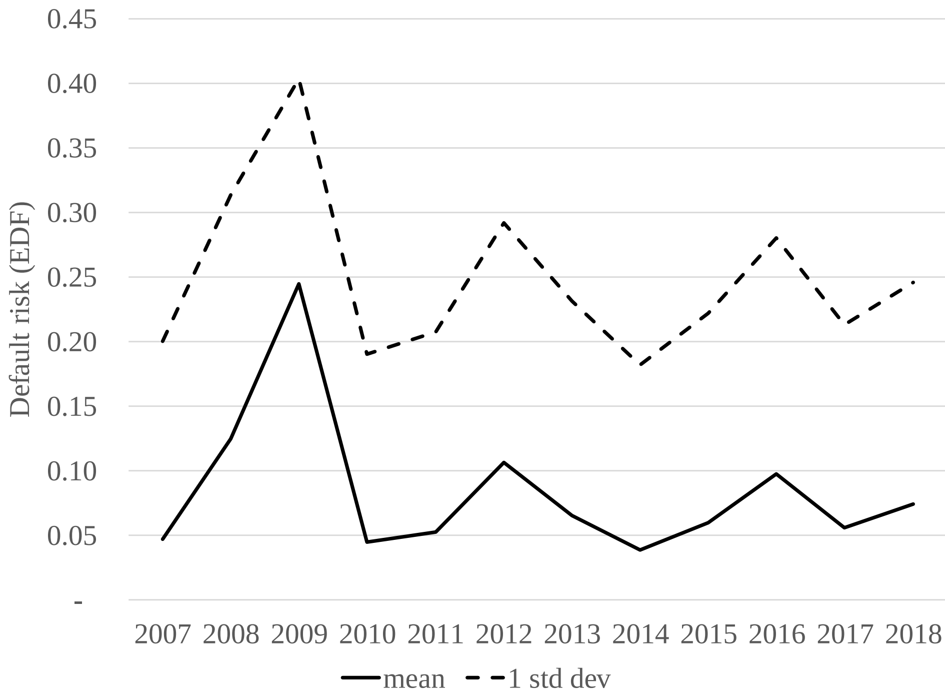

Figure 2. *EDF*, by year. For each year from 2007 through 2018, we present the mean and standard deviation of default risk (*EDF*).

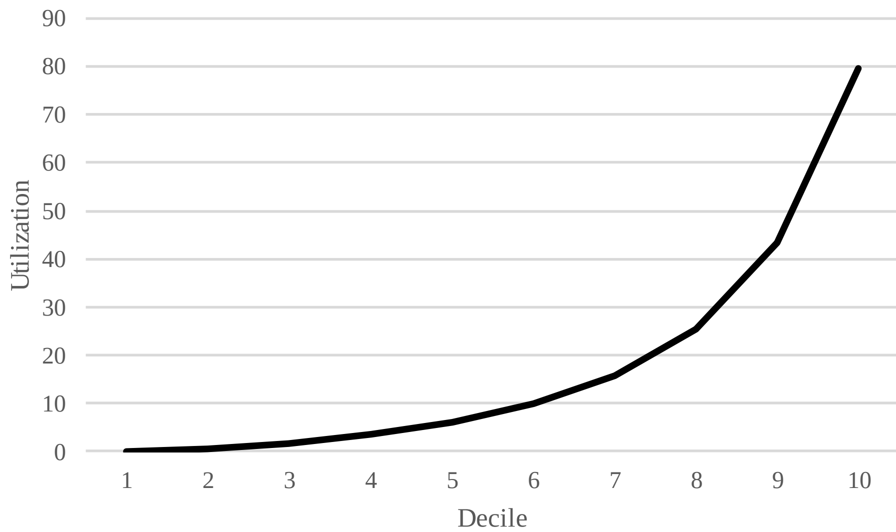

Figure 3. *Utilization,* by decile. We present the mean for firm-year observations for *Utilization*—our proxy for short interest—from 2007 to 2018 (n = 37,938). *Utilization* is the dollar amount borRowed divided by the dollar amount available to borRow. For each firm, *Utilization* is the mean of daily Utilization for the month following the fiscal year-end.

Table 1
Summary statistics
We present fiscal-year-end summary statistics for the 37,938 firm-year observations for 2007–2018. We follow Bharath and Shumway (2008) to calculate yearly default risk (*EDF*). For the shorting variables, we compute monthly means from daily data. We combine the yearly default risk at firms' fiscal year-end with the next months' shorting variables. As an example, for 2018, for a firm with a 12/31 (5/31) fiscal year-end, we combine the 2018 *EDF* calculation with January 2019 (June 2018) shorting variables. *Utilization*—our proxy for short interest—is borRowed value divided by available value. *Borrowing Fee* (in basis points) is the value-weighted average borRowing fee for all transactions. *Equity* is the market value of equity (in millions of dollars) calculated as the product of shares outstanding and share price. *Debt* is the sum of current liabilities and one-half of long-term debt. *Return* is the sum of twelve-monthly log price relatives. *Stock volatility* is the standard deviation of monthly equity returns over the previous year.

	Mean	Std.	Minimum	Maximum
EDF	0.09	0.26	0.00	1.00
Utilization	18.59%	24.68%	0.00%	100.00%
Borrowing Fee	2.84	9.46	0.17	229.10
Equity	$5,025	$23,152	$0.00	$867,507
Debt	$1,514	$5,941	$0.00	$163,734
Return	-0.03	0.70	-11.99	9.43
Stock volatility	0.14	0.12	0.00	5.81

Table 2
Statistics for *EDF*, sorted by deciles
We present statistics for *EDF* for the indicated deciles.

	Deciles		
	10th	**9th**	**1st—8th**
Mean	0.8384	0.0168	4.57E-10
Std Deviation	0.2522	0.0352	4.10E-17
Skewness	-1.4083	2.5179	11.9655
Kurtosis	0.5278	5.6172	60.3173
100% Max	1.0000	0.1664	<0.0001
99%	1.0000	0.1537	<0.0001
95%	1.0000	0.1103	<0.0001
90%	1.0000	0.0654	<0.0001
75% Q3	1.0000	0.0120	<0.0001
50% Median	0.9960	0.0004	<0.0001
25% Q1	0.7485	<0.0001	<0.0001
10%	0.3647	<0.0001	<0.0001
5%	0.2558	<0.0001	<0.0001
1%	0.1811	<0.0001	<0.0001
0% Min	0.1668	<0.0001	<0.0001
N	3,793	3,794	30,351

Table 3

Does *Utilization* anticipate changes in *EDF*?

Let ΔEDF = the difference between *EDF* at the end of fiscal years *t* and *t-1*. $\Delta EDF > 0$ for 15,183 companies, < 0 for 14,491 companies, and equal to 0 for 8,264 companies. Let $\Delta EDF+$ be a dummy variable that equals 1 if $\Delta EDF >= 0$ and 0 otherwise. We calculate the mean of *Utilization* for fiscal year *t* (measured the first month of the fiscal year) for companies that experience an increase in *EDF* and a decrease in *EDF*. Row 3 reports the difference between the means of *Utilization* for $\Delta EDF+ = 1$ and = 0. We test the null hypothesis that the means are equal against the alternate hypothesis that the mean of *Utilization* for $\Delta EDF+ = 1$ is higher and presents the resulting *t*-statistics in parentheses. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

	Utilization											
ΔEDF+	All years	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
0	17.07	21.99	19.93	14.92	13.77	14.66	15.08	16.08	15.85	22.86	18.85	20.97
1	19.76	24.95	18.71	15.90	15.88	17.71	20.71	19.91	16.04	20.04	22.26	23.01
Row 0 minus Row 1	2.69†	2.96†	-1.23	0.99	2.11†	3.05†	5.62†	3.83†	0.19	-2.83†	3.41†	2.04*
t-statistic	(9.72)	(2.82)	(-1.12)	(1.07)	(2.88)	(3.70)	(5.61)	(3.72)	(0.20)	(-2.66)	(3.23)	(1.84)

Table 4

Distribution of *Utilization* by *ΔEDF*

We calculate portfolio deciles of *ΔEDF,* which is the difference between *EDF* at the end of fiscal years *t* and *t-1*. We consolidate the means for deciles 1-8 and report the mean of *Utilization* in Row 1 for all years (Column 2) and each year. Rows 2 and 3 report the results for deciles 9 and 10, respectively. Row 4 (Row 6) reports the difference between the means of decile 9 (10) and deciles 1–8. We test the null hypothesis that the mean of *Utilization* for decile 1–8 and decile 9 (10) are equal against the alternate hypothesis that the mean of *Utilization* for decile 9 (10) is higher and present the resulting *t*-statistics. There are 37,938 firm-year observations from 2008–2018, and all variables are for firm *i* in period t. Note that *Utilization* is lagged relative to *EDF*. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

Deciles of	Utilization											
ΔEDF	All years	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1-8	16.82	23.53	19.04	15.07	13.62	15.75	14.95	14.70	13.56	18.58	17.23	18.06
9	21.18	24.09	16.84	15.64	15.22	19.90	22.47	21.31	21.04	25.13	23.39	27.04
10 (highest)	24.05	25.67	19.58	15.96	16.86	20.74	24.09	27.35	22.83	30.88	29.13	30.81
(1-8) minus 9	-4.36†	-0.56	2.2	-0.57	-1.6	-4.15	-7.52	-6.61†	-7.48	-6.55*	-6.16†	-8.98
t-statistic	-4.12	-0.72	-1.46	-0.19	-0.89	-0.41	-0.64	-2.35	-0.82	-2.18	-2.34	-1.38
(1-8) minus 10	-7.23†	-2.14	-0.54	-0.89	-3.24†	-4.99†	-9.14†	-12.65†	-9.27†	-12.3†	-11.90†	-12.75†
t-statistic	-16.21	-1.41	-0.39	-0.69	-2.7	-3.87	-6.19	-8.27	-7.03	-7.35	-7.26	-7.58

Table 5
Regressions of *ΔEDF* on *Utilization*
We present the results of our estimation of the OLS regression of *ΔEDF* on *Utilization* (measured the first month of fiscal year *t*) and the remaining variables listed in the first column. *ΔEDF* is the difference between *EDF* at the end of fiscal years *t* and *t-1*. We report all firms' results in Columns 1–3 and deciles 9 and 10 (combined) in Columns 4–6. *t*-statistics are in parentheses. There are 37,938 firm-year observations from 2007–2018ables are for firm *i* in period *t*. Note that *Utilization* is lagged relative to *EDF*. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

	Dependent variable: ΔEDF_t					
	All firms			Deciles 9 and 10		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$Utilization_{i,t}$	0.001†	0.0003†	0.0003*	0.001†	0.0013†	0.0010†
	(7.23)	(3.89)	(2.21)	(8.21)	(7.77)	(2.86)
$Ln(Equity)_{i,t-1}$	-0.016†	-0.011†	-0.010†	-0.148†	-0.106†	-0.144†
	-(11.62)	-(8.38)	-(3.08)	-(37.73)	-(27.87)	-(15.67)
$Ln(Debt)_{i,t-1}$	0.009†	0.007†	0.019†	0.114†	0.081†	0.114†
	(6.62)	(5.32)	(4.55)	(32.04)	(23.74)	(8.52)
Stock $volatility_{i,t-1}$	0.003†	0.002†	0.003†	0.000	-0.002†	0.003
	(13.83)	(6.87)	(8.90)	-(0.34)	-(3.04)	(1.80)
Intercept	0.027†	0.035†	-0.012	0.497†	0.380†	0.793†
	(5.02)	(4.63)	-(0.04)	(37.66)	(20.80)	(3.69)
Year fixed effects	No	Yes	Yes	No	Yes	Yes
Firm fixed effects	No	No	Yes	No	No	Yes
Adjusted R^2	0.01	0.07	0.19	0.20	0.35	0.68

Table 6

Does *Utilization* anticipate movement in and out of the highest *EDF* group?

We report mean *Utilization* based on firms moving in and out of the *EDF* decile 10. For firm *i* in year *t*, let $EDF10_{i,t}$ = *HNever* if neither $EDF_{i,t-1}$ nor $EDF_{i,t}$ are in the 10th decile; = *HLeaver* if $EDF_{i,t-1}$ is in the 10th decile and $EDF_{i,t}$ is not; = *HMover* if $EDF_{i,t}$ is in the 10th decile $EDF_{i,t-1}$ is not; and = *HStayer* if both $EDF_{i,t-1}$ and $EDF_{i,t}$ are in the 10th decile. In Row 5, we present the difference between Row 4 and Row 2. In Row 6, we present the *t*-statistic for the test of the null hypothesis that the means of these two Rows are equal against the alternative hypothesis *HStayers* have higher *Utilization* than HLeavers. In Row 7 – 10, we present similar tests for *HStayer* minus *HMover* and *HMover* minus *HLeaver*. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

$EDF10_{i,t}$	Utilization	N =
1) HNever	17.06	25,790
2) HLeaver	21.30	1,346
3) HMover	24.07	1,613
4) HStayer	26.58	1,064
5) HStayer minus HLeaver	5.28†	
6) t-statistic	(4.33)	
7) HStayer minus HMover	2.51*	
8) t-statistic	(2.16)	
9) HMover minus HLeaver	2.77†	
10) t-statistic	(2.65)	

Table 7

Does *Utilization* anticipate moving into the highest *EDF* group?

Ignoring *EDF* cases in decile 10 in both fiscal years *t* and *t-1*, $HMover_{i,t}$ = 1 if *EDF* moves from a decile ≠ 10 in year *t-1* to decile 10 in year *t* and 0 otherwise. Of course, *EDF* can be in deciles 1-9 in year *t-1*. For each decile of *EDF* in year *t-1*, we present the mean of *Utilization* for fiscal year *t* for cases when *EDF* does not move into decile 10 in Row 1 and for cases when it does move into decile 10 in Row 3. We test the null hypothesis that the mean *Utilization* for Row 1 and Row 3 is equal against the alternate hypothesis that the mean for Row 3 is higher. *t*-statistics are in parentheses. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

	Decile of EDF									
HMover	All	1	2	3	4	5	6	7	8	9
	Utilization (N in italics)									
0	17.62	13.76	15.51	15.88	16.17	16.84	18.16	19.19	19.57	20.82
N =	*28,200*	*3,610*	*2,713*	*3,151*	*3,067*	*3,017*	*2,871*	*2,796*	*2,509*	*2,093*
1	24.07	28.75	9.50	25.59	24.20	26.36	24.69	20.80	24.38	24.93
N =	*1,613*	*12*	*20*	*22*	*56*	*81*	*152*	*237*	*388*	*655*
Diff 1-0	6.44†	14.99†	-6.01	9.71*	8.02†	9.52†	6.52†	1.61	4.81†	4.12†
t-statistic	(10.82)	(2.78)	(-0.01)	(2.22)	(2.87)	(3.84)	(3.35)	(0.98)	(3.45)	(3.37)

Table 8

Logit model for prediction of movement to the highest *EDF* group

We estimate a Logit model with $HMover_{i,t}$ as the LHS variable and $Utilization_{i,t-1}$, $Decile10_{i,t-1}$, $Ln(Equity)_{i,t-1}$, $Ln(Debt)_{i,t-1}$, and $Stock\ volatility_{i,t-1}$ as RHS variables.We drop firms with *EDF* in decile 10 in year *t-1* (N=2,410). $HMover_{i,t}$ = 1 (n = 1,613) if *EDF* is in decile 10 in year *t* but not in year *t-1* and 0 (N=25,790) otherwise. The probability of movement into *EDF* decile 10 is $F_i = (1 + \exp(-D_i))^{-1}$, where $D_i = X_i\beta$ is a linear index of financial variables relevant to the movement into *EDF* decile 10. Positive coefficients for β indicate a positive relation between the variables and the probability of moving into *EDF* decile 10, and vice versa. We report Wald chi-square statistics for the hypothesis test that an individual predictor's regression coefficient is zero in parentheses. *L-Ratio* is the log-likelihood ratio statistic for testing the null hypothesis that all coefficients are statistically insignificant. AIC is the Akaike Information Criterion. Pseudo-R^2 is from the log-likelihood function. Note that *Utilization* is lagged relative to *HMover*. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

	$HMover_{i,t}$			
Variables	(1)	(2)	(3)	(4)
Intercept	-3.00†	-6.00†	0.36†	-0.40*
	(7,793)	(3,163)	(16.37)	(3.86)
$Utilization_{i,t}$	0.01†	0.01†	0.02†	0.02†
	(142.04)	(46.30)	(262.13)	(241.86)
$Decile10_{i,t-1}$		0.58†		0.09†
		(1,334)		(17.10)
$Ln(Equity)_{i,t-1}$			-1.79†	-1.68†
			(2,287)	(1,383)
$Ln(Debt)_{i,t-1}$			1.43†	1.33†
			(1,717)	(1,044)
Stock $volatility_{i,t-1}$			-0.04†	-0.02†
			(33.15)	(10.39)
L-Ratio	127.5	2,211	4,231	4,248
AIC	12,144	10,062	8,017	8,013
Psuedo-R^2	0.01	0.18	0.35	0.35

Table 9

Does *ΔUtilization* anticipate *ΔEDF*?

We present results of the regression of *ΔEDF* against *Dummy_Util* (Column 2) and *ΔUtilization* (Column 3). *Dummy_Util* = 1 if the firm's monthly average *Utilization* moves from decile 1-9 in fiscal year *t-1* into decile 10 in year *t* and 0 otherwise. *ΔUtilization* is the change in *Utilization* between the first and second six months of fiscal year *t*. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

Variables	ΔEDF	
Intercept	0.01†	0.00262
	(4.50)	(1.44)
Dummy_Util	0.05†	
	(6.51)	
ΔUtilization		0.00035†
		(3.71)

Note: The definition of *ΔUtilization* in this table differs from that in earlier tables.

Internet Appendix

Relation of Short selling to *EDF*

Table IA1 reports the mean $Utilization_{t+1}$ for each *EDF*-portfolio decile for all years combined (Column 2) and by year. $Utilization_{t+1}$ is measured the month after the fiscal year-end, whereas we measre EDF_t at the fiscal year-end.[6] Mean $Utilization_{t+1}$ increases monotonically with *EDF* in Column 2. The penultimate Row reports the difference in the mean $Utilization_{t+1}$ between the highest and lowest deciles of *EDF* stocks. For all years, the mean $Utilization_{t+1}$for the highest-default-risk stocks is 11.12% higher than the mean $Utilization_{t+1}$ for the lowest-default-risk stocks. This difference is statistically significant. The differences in values for the penultimate Row are positive and statistically significant for 2012 to 2018. However, there is no clear pattern from 2007 to 2011, which is likely caused by market dislocation and heightened borrowing costs during and surrounding the financial crisis.

To extend our portfolio sort analysis, we estimate the following regression:

$$Utilization_{i,t+1} = b_o + b_1\, EDF_{i,t} + b_2\, Borrowing\ Fee_{i,t+1} + b_3\, Ln(Equity)_{i,t} + b_4\, Ln(Debt)_{i,t} \quad (1)$$
$$+ b_5\, Stock\ volatility_{i,t} + \text{Year fixed effects} + \text{Firm fixed effects} + \varepsilon_{i,t}$$

Table IA2, Columns 1–9, presents the estimation of Equation 1with and without firm and year fixed effects. We find a positive and significant relation between shorting as measured by $Utilization_{t+1}$ and EDF_t for eight of the nine regressions.

Table IA2, Columns 4–9, includes various control variables and consistently finds a positive relation between shorting and default risk. Columns 4 to 6 reports the results of regressions with *Borrowing Fee* as a control variable. Columns 7 to 9 reports the results with the following control variables: *Borrowing Fee*, *Ln(Equity)*, *Ln(Debt)*, and *Volatility*. In all but the last regression, the point estimates for *EDF* are statistically significant.

[6] We measure *Utilization* the month following the fiscal year-end as there is a delay in the reporting of financial results of up to 60 days following year-end. Therefore, the financial data necessary to compute EDF_t is available around the time of the $Utilization_{t+1}$.

In Table IA3, we incorporate the interactive term *EDF X EDF10* in our regression analysis to analyze the stocks with the highest default risk more closely. *EDF10* equals 1 if the observation is in the top decile of *EDF* for its given year and 0 otherwise. By including this term, we allow the slope to differ for the companies with the highest default risk. We find that *EDF X EDF10* is positive and statistically significant in all of our regressions. When including the year and firm fixed effects, *EDF* becomes insignificant.

The regression analysis results and the portfolio analysis support the idea that short interest is higher for stocks with the highest default risk.

Table IA1

Distribution of *Utilization* by *EDF* portfolios

We form portfolios by sorting EDF_t into deciles for each year 2007–2018. We report the mean of $Utilization_{t+1}$ for each year for each decile. The penultimate Row indicates the difference between the means of the highest EDFs and lowest EDFs. We test the null hypothesis of equality of Row 1 and Row 10, and present *t*-statistics are in parentheses in the last Row. † indicates significance at the 0.01 level.

Deciles of	$Utilization_{t+1}$												
EDF_t	All years	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1 (lowest)	13.53	20.06	16.54	15.12	15.01	16.40	14.06	10.31	11.77	12.80	12.71	11.11	10.21
2	15.27	22.54	18.95	15.69	13.04	16.74	12.81	11.46	13.85	14.66	13.66	11.82	13.56
3	15.72	23.49	20.63	16.84	14.09	15.70	14.82	13.62	11.25	14.27	13.94	14.81	14.33
4	16.49	23.69	19.77	13.58	14.64	17.26	17.60	14.76	13.95	16.33	14.68	15.17	15.88
5	17.40	25.82	18.15	16.14	14.54	13.03	17.25	15.78	13.17	18.15	19.53	19.42	17.42
6	18.69	25.49	19.74	14.96	12.15	16.91	16.43	18.34	18.45	22.14	18.92	20.50	19.90
7	20.09	24.16	19.29	14.85	13.00	16.14	18.28	19.27	18.19	27.53	24.32	21.92	24.10
8	21.10	24.74	16.70	13.72	13.58	16.23	19.74	22.32	21.02	25.69	25.84	27.68	26.26
9	23.33	21.57	16.75	14.52	14.30	18.71	20.63	25.63	19.14	32.65	30.07	31.78	34.66
10 (highest)	24.65	17.35	15.51	14.25	16.30	15.05	25.04	29.01	20.14	43.23	28.24	40.93	32.93
Row 10 minus 1	11.12†	-2.71	-1.04	-0.87	1.30	-1.35	10.98†	18.70†	8.37†	30.42†	15.52†	29.82†	22.72†
t-statistic	(19.61)	(-1.41)	(-0.64)	(-0.56)	(0.71)	(-0.80)	(5.41)	(9.65)	(19.61)	(13.10)	(6.90)	(13.13)	(12.88)

Table IA2

Regressions of *Utilization* on *EDF*

We present the ordinary least square (OLS) regressions of $Utilization_{t+1}$ on EDF_t and additional control variables. We control for both firm and year fixed effects as indicated. *t*-statistics are in parentheses. There are 37,938 firm-year observations from 2007–2018 so that all variables are for firm *i* in period *t*. $*$ and † indicate significance at the 0.05 and 0.01 levels, respectively.

	Dependent variable: $Utilization_{t+1}$		
Variable	(1)	(2)	(3)
EDF_t	7.091†	8.267†	0.878
	(12.89)	(14.73)	(1.79)
Borrowing Fee_{t+1}	0.949†	0.946†	0.669†
	(72.97)	(72.12)	(44.54)
$Ln(Equity)_t$	0.861†	0.966†	-0.741†
	(8.00)	(8.90)	(-4.47)
$Ln(Debt)_t$	-2.098†	-2.175†	1.426†
	(-21.7)	(-22.42)	(7.04)
Stock $volatility_t$	-0.002*	-0.002†	-0.068†
	(-2.48)	(-2.59)	(-4.26)
Intercept	21.334†	20.620†	27.557*
	(52.04)	(35.86)	(2.38)
Year fixed effects	No	Yes	Yes
Firm fixed effects	No	No	Yes
Adjusted R^2	0.18	0.19	0.64

Table IA3

Regressions of *Utilization* on *EDF* including an interaction term

We present the ordinary least square (OLS) regressions of $Utilization_{t+1}$ on EDF_t. We interact *EDF* with *EDF10*, which equals 1 if the observation is in the top decile of *EDF* for a given year and 0 otherwise. We control for both firm and year fixed effects as indicated. *t*-statistics are reported in parentheses. There are 37,938 firm-year observations from 2007–2018 so that all variables are for firm *i* in period *t*. ∗ and † indicate significance at the 0.05 and 0.01 levels, respectively.

Variable	(1)	(2)	(3)
EDF_t	-2.0284	0.3280	-0.5329
	(-1.85)	(0.28)	-(0.56)
EDF_t X $EDF10_t$	8.5709†	6.42†	3.14†
	(7.13)	(5.11)	(3.05)
Borrowing Fee_{t+1}	1.0184†	1.02†	0.67†
	(79.51)	(79.47)	(44.80)
Intercept	16.27†	15.11†	27.61*
	(125.36)	(36.66)	(2.40)
Year fixed effects	No	Yes	Yes
Firm fixed effects	No	No	Yes
Adjusted R^2	0.16	0.17	0.64