Predicting Corporate Bankruptcy

Heather E. Dempsey[□], Sacred Heart University[†]

July 2017

JEL G33, G34, G17, C10, C22, C32, C53, C59, C87, E32, G39, L25

Keywords: Corporate bankruptcy; firm failure; logit regression model; forecast bankruptcy; default; chapter 7; chapter 11; McNemar test; paired analysis; financial ratios

[□] Tel +1-203-456-5557. E-Mail: dempseyh@mail.sacredheart.edu.

[†] Jack Welch School of Business, 5151 Park Avenue, Fairfield, CT 06825.

Abstract

This paper presents an empirical investigation which endeavors to answer the question of whether corporate bankruptcy[1] can be accurately predicted based on publicly available financial information. In this study, a logistic regression model is specified to forecast the probability of failure three months ahead as a function of publicly available data. The sample includes quarterly financial data from Q12005-Q12015 on [insert number of firms] small, medium and large firms as measured by market capitalization calculated from the Center for Research in Securities and Prices (CRSP)/Compustat Merged Database. Firm-level accounting information from the balance sheet, cash flow, and income statement are taken for the years 2005 to 2015 to predict the probability of failure in 3 months' time. This analysis provides government regulators, business analysts and researchers a framework through which to quantify the likelihood of corporate failure and intervene to obviate that failure. Cash flow and debt ratios, significant at the 1% level, are the most important factors in the likelihood of a firm's continued viability. The profitability ratio, quick ratio and size are all significant at the 5% level, proving reliable indicators of the impact on the likelihood of failure with --% forecast accuracy, overall.

[1] The terms bankruptcy, failure, and default are used interchangeably unless otherwise noted.

1. Introduction

Since 2005, in the United States alone, [insert number of firms] publicly traded companies have filed for bankruptcy[2]. Bankruptcy negatively impacts more than just the firm. It is no wonder the use of prediction methods to forecast failure has been of great interest to government regulatory and private rating agencies, financial market participants, corporate auditors and academics alike. Many agents are adversely affected—suffering substantial financial loss. Consequences following bankruptcy include: 1) employees are out jobs 2) suppliers are out buyers 3) shareholders lose ownership and 4) creditors lose interest payments and principal. The objective of this study is to provide a framework through which to evaluate a firm's likelihood of future failure. This study differs from previous research in the ratios selected and a shorter forecast horizon. I introduce a model using logistic regression which will enable firms to perform a likelihood of failure sensitivity analysis. Model coefficients represent the marginal contribution in determining a firm's probability of failure. Early warning signs can then be identified from those factors with large positive and significant coefficients, and remedial action taken.

The formal hypothesis proposed is that the probability of a firm going bankrupt can be forecast from key financial ratios found on the balance sheet, income, and cash flow statements of the firm. The null hypothesis is simply that corporate bankruptcy cannot be predicted effectively[3]. This study defines bankruptcy as firms that have filed for Chapter 11 or Chapter 7 and have been delisted due to bankruptcy. Throughout this paper, the terms default and failure are used synonymously with bankruptcy. Many firms defaulted due to the unusual circumstances surrounding the recession. Data

---

[2] Source: Bankruptcydata.com, bankruptcy includes chapter 11 and 7 and those delisted in the CRSP/Compustat Merged Database due to bankruptcy.

[3] The McNemar's test is used to access model effectiveness.

for the years 2007-2009 have been controlled for and normalized using dummy variables for each year in the sample data.

This paper is organized as follows: In the next section, a review of the previous literature. The data selection process and variables are described in Section 3. In Section 4, I describe the theoretical and regression model. The results and interpretation are presented in Section 5. In Section 6, the conclusion is drawn from the empirical results. In Section 7, the appendix presents additional tables and figures.

2. Literature Review

Previous literature has grappled with the right method to predict corporate failure for over four decades. To date, no one best measure has been agreed upon. Beginning with William Beaver (1966) who used non-paired and paired sample selection to conduct univariate discriminate analysis and mean difference tests on both failed and non-failed firms. The paired analysis involved sampling failed firms and matching by industry and asset size a non-failed "mate." This study follows Beaver's sample paired selection method. Even before it had been attempted, Beaver alluded to the usefulness of a multiple discriminant approach in future study. He used a univariate discriminant analysis method in conjunction with a mean comparison test and concluded that at *very high* values, a high cash-flow to total-debt ratios can *actually increase* the probability of default, and can be detected up to five years prior to bankruptcy.

Edward Altman (1968) provided an extension of Beaver's work and used multiple discriminant analysis to improve the predictive power of the model and discriminate between failed and non-failed firms. Altman developed an improvement of the previous Z-score. His design became well-known and is still popular today. He constructed a linear function composed of five financial ratio categories summed together to predict manufacturing firms' failure up to two years in advance if its score fell within a certain range. Similar to, Altman's study, the present research employs financial ratios to determine the probability a firm will file for bankruptcy. James Ohlson (1980) used the logit model for the prediction of corporate bankruptcy. The current research motivated by Olson's work also employs the logit model. Ohlson successfully predicted the probability of a firm's failure with the O-score.

Aydin Ozkan (1996) in his thesis addresses *size* of the firm and relation to *insolvency*. The individual firm characteristics influence on liquidation costs are addressed. In Ozkan's analysis, he puts forth—larger firms are less likely to be liquidated than small firms when they are financially failing. Michele Modina and Filomena Pietrovito (2013) analyze small-to-medium size firms using logit analysis. They conclude the most important factor in firm failure is debt structure. This paper investigates the validity of this claim and finds debt structure does, in fact, have the most impact on the probability of bankruptcy. In Kanstantin Danilov's (2014) thesis, ratios from failed, and not-failed firms were evaluated with similar financial ratios to this current study. Quarterly data is identical to the present study, 5-7 years is a shorter sample period. This study adopts a method of Danilov's, to ensure comparability with each of the firms filing for Chapter 7, or Chapter 11 bankruptcy at different times—the data is then "repositioned"—relative to the bankruptcy event to be equal with that of the firm's counterpart.

An interesting innovation to the body of failed firm research is the use of spline functions taken from financial firms in determining the likelihood of default as in Paolo Giordani, Tor Jacobson, Erik von Schedvin, and Mattias Villani (2014). Authors discovered remarkably, an improved predictability of 70-90% when using spline function (for highly nonlinear relationships) and financial ratios in a logit model, the same regression model employed in the present study.

3. Data

Sample selection begins with all publicly traded firms, excluding the finance and utility industries due to their unconventional capital structure and hence incomparable financial ratios. This sample is limited to data up until 2015 due to availability in the CRSP/Compustat database. This study uses panel data structure to capture cross-section and time series observations allowing for multiple firm-characteristics to be compared over time. The full period of observations is taken from years 2005-2015. The period from the first quarter 2005 to the fourth quarter 2014 is used in-sample to train the model and obtain parameter estimates. The first quarter 2015 is reserved for the out-of-sample forecast and cross-validation. Firms selection is further refined to those with at least 40 consecutive quarterly financial statements. To standardize the sample, I exclude the top 2.5% and the bottom 2.5% of firms in terms of size (market capitalization) Winsorizing the outliers (Ohlson, 1980). Companies with missing observations are removed. The resulting sample is [insert the total number of firms] failed and non-failed firms. Selecting firms that have filed for Chapter 7, Chapter 11, been delisted due to bankruptcy by (CRSP/Compustat Merged Database) or have had the letter "Q" appended to stock ticker symbol by the Securities and Exchange Commission's(SEC)(EDGAR Historical Database), as the failed sample. [insert number of firms] firms are classified as failed are sorted by size and industry classification.

Using the method of paired selection as in Beaver (1966), the non-failed firms are selected one by one based on closest comparable size, the period of operation, and matching industry, to their failed counterpart. The final sample includes [insert number] failed and [insert number] non-failed firms. Failed firms in the dataset are assigned a value of 1 for bankrupt, and non-failed firms are coded 0, for the binary response. This is necessary for a categorical dependent variable. Either bankrupt or nonbankrupt—the only two options, within this study.

Year and industry dummy variables have been created to control for year-specific fixed-effects and comparable ratio variance across each specific industry. Year dummies include the years 2006-2015 and exclude one period—year 2005. Industries identified from the paired sample are crop production, construction of buildings, textile mills, food manufacturing, merchant wholesalers—durable goods, furniture and home furnishing stores, air transportation, telecommunications, real estate and hospitals, merchant wholesaler—durable goods have been left out intentionally. These dummy variables are coded 1 for inclusion and 0 if not applicable. A proxy for growth opportunities has been calculated individually using research and development expenditures (R&D) divided by total assets (Danilov, 2014). The CRSP/Compustat database have a (.) symbol to indicate the absence of R&D. For each of these firms, the value 0 is used to manually overwrite the symbol and record the R&D value as a zero, to be interpreted as the selected firm spent $0.00 on R&D expenditures that period. These firms are then included in the sample.

Table I, the summary statistics of both classes non-failed and failed firms, reveals larger means and medians for most ratios of the non-defaulted firms only debt ratio was lower than rest of sample. Non-failed firms exhibit lower debt ratio averages, indicating liabilities typically remain low in firms that succeed. The failing firms exhibit high volatility, meeting the author’s previously acknowledged expectations. . Standard deviation and kurtosis are larger for the bankrupt firms as expected they exhibit greater volatility and contain more outliers.

Table I

Summary Statistics

Variables	Mean		Median		Standard deviation		Kurtosis	
	Default	Non-default	Default	Non-default	Default	Non-default	Default	Non-default
Cash flow								
Quick ratio								
Debt ratio								
Profitability								
Size								
Growth								
No. Observations								

Note: Source—CRSP/Compustat Merged Database (2017), SEC Data EDGAR Historical Database

The failed firms have a lower means for growth (measured as research and development expenditures to total assets) cash flow and profitability ratios than those of the non-failed firms. Failed firms also exhibit higher debt ratio than non-failed counterparts (Hol, Westgaard, & Van der Wijst, 2002).

Table II provides descriptive statistics on the entire sample set of failed and non-failed firms both the mean of levels and first differences are calculated with the standard deviation in parenthesis. The mean for each independent variable of entire sample set has dropped below the individual class averages. This can be explained by a large number of outliers in the default sample. The total number of failed firms from each industry is provided in Panel B. The percentage of bankruptcies in each industry out of the total sample of failed firms is reported. Textile mills and crop production fared worse than the rest of the industries with a greater number of bankrupt firms. Conversely, real estate and air transportation failed the least of the sample. Real estate had --% fewer defaults than the remaining industries.

Table II
Descriptive statistics

Panel A			Mean (standard deviation)
Variable		Levels	First differences
Cash flow ratio			
Quick ratio			
Debt ratio			
Profitability			
Size			
Panel B	Number failed firms	Mean	Percentage of sample
Crop production			
Construction of buildings			
Textile mills			
Food manufacturing			
Merchant wholesalers			
Furniture stores			
Air transportation			
Real estate			
Hospitals			

Note: Source CRSP/Compustat Merged Database, SEC Data EDGAR Historical Database (2017).

3.1. Variable Description

Bankruptcy is the dichotomous dependent variable coded one if the company went bankrupt and 0 otherwise. The logit model allows this to be expressed as a probability. Specifically, bankruptcy is defined as companies who have been delisted due to bankruptcy, defaulted, have a letter "Q" appended to stock ticker symbol[4], filed for Chapter 7 or Chapter 11.

The following independent variables were selected from the maximum log likelihood function and stepwise procedure.

- **Cash flow** is calculated as the ratio of operating activities' net cash flow to current liabilities. A cash flow ratio of one or greater is anticipated to have a negative impact on the probability of default. The larger value in the numerator versus the denominator means it can meet its debt obligations. However, a ratio value less than one will have a positive effect on the probability of default because the cash flow ratio measures the ability for the firm to pay its debts for the same period that cash comes in from operations. If the firm's debt obligations exceed the amount of cash or cash equivalents, then the probability of default increases.
- **The quick ratio** is calculated as current assets minus inventories in the numerator, divided by current liabilities. The quick ratio measures the firm's short-term liquidity to its short-term debt obligations. Like the cash flow ratio, a larger value in the numerator indicates the ability to pay its debts and will have a negative effect on the dependent variable, decreasing the probability of default. Conversely, a larger denominator value or the current debt obligations will positively impact the likelihood of default.

[4] Source of historical data: Security Exchange Commission's EDGAR database

- **The debt ratio** is the total liabilities to total assets. Debt ratio describes the percentage of a firm's total assets that were funded with incurring debt. A debt ratio greater than one means the company has more debt than the company is worth and there is an increased probability of default.
- **Growth potential** is quantified by the proxy research and development expenditures divided by total assets. This measure describes a firm's investment in itself, the future potential to increase revenues through new business. Growth potential will have a negative impact on the probability of default.
- **Profitability** is defined as return on equity calculated as the ratio of net income to shareholders' equity. If the firm has more income coming in than revenue it is said to be profitable. Profitability has a negative effect on the probability of failure. Profits allow a company to meet their debt obligations and expand operations (Hol, Westgaard, & Van der Wijst, 2002).
- **Size** is the total number of shares times share price obtained from the monthly CRSP U.S. Stock Database. Stock prices are taken from Friday preceding the announcement of a firm's financial report if it falls on a weekend. Larger firms file for bankruptcy less than smaller firms and smaller firms are less solvent than the large companies (Dang & Li, 2015).
- **Dummy variables** comprise three groups—Industry, year and filing characteristics. Industries that are represented in the failed and non-failed firms are 1) crop production, 2) construction of buildings, 3) textile mills, 4) food manufacturing, 5) merchant wholesalers—durable goods, 6) furniture and home furnishings stores, 7) air transportation, 8) telecommunications, 9) real estate. Hospitals are still observed, but were

excluded from the dummy variable construction. Public utilities and finance firms have been left out intentionally. The exclusion is due to utility and finance companies' capital structure being significantly different from other industries, rendering their financial ratios incomparable (Chen, Hu, & Pan, 2008). Each industry (Table A.2) has unique norms reflected in their financial ratios which necessitate dummy variables to control and compare data within the industry to which it is assigned. Coded, 1 when an industry-specific data item is being analyzed, and a 0 otherwise. Two more dummy variables specify whether a firm has failed to file on time, or has restated its financials. Filing late is identified by an "E" appended to its stock ticker symbol for companies listed on the Nasdaq Stock Market or the letters "LF" for New York Stock Exchange-listed companies. Restated quarterly data is taken from CRSP/Compustat Merged Database (2017). Yearly dummy variables from 2005 to 2015 have been included to control for time specific fixed-effects. These variables are coded 1 for in-year, and 0 for all other years. Table III describes the expected coefficient signs of non-categorical variables in the model.

Table III
Expected coefficient signs of non-dichotomous variables

Variable	***Expected sign***	***Explanation***
Cash flow ratio	(+/-)	Cash flow of one or greater will enable the firm to meet its obligation, reducing default risk, less than one will increase the chances of default
Quick ratio	(+/-)	A value of one or greater will reduce default probability, otherwise, increases probability of default
Debt ratio	(+/-)	Lower ratio will have a negative effect; a higher ratio will increase probability of default
Growth	(-)	Growth will have a negative impact on the probability of default. This measure is increasing firm value.
Profitability	(-)	Positive return on equity (ROE) decreases the probability of default; negative ROE value has positive impact on the odds of default
Size (market cap)	(-)	Larger firms have more resources, access to credit and default less on loans; smaller firms are less stable, positively increases probability of default

A summary of all variables included can be found in the appendix tables A.1 and A.2.

4. Methodology

There are two prominent models used to classify failure and non-failure, these are the logit model and multiple discriminant analysis. In this research, the logit model is employed as previous research has shown higher classification accuracy than discriminant analysis (Mihalovic, 2016). Quarterly data, spanning the period January 1, 2005, to March 31, 2015, has been retrieved from CRSP/Compustat Merged Database (2017). The observations are first differenced to correct for the three dynamics which obfuscate time-series results: 1) cyclicality, 2) seasonality and 3) trend (Diebold, 2007). Cyclicality represents any dynamic not evident in the trend or seasonal patterns. Business cycles, in general, present a less rigid oscillation than other cycles. To forecast, I want to know at least that the mean and covariance structure (covariances between current and historical values) are constant over time (Diebold, 2007). Covariance Stationarity is verified through the presence of the following three conditions: The mean is 0, the variance is constant, and serial correlation has been removed (Diebold, 2007). y is serially independent if it also serially uncorrelated. y is said to be identically and independently distributed as

$$y_t = \sim iid(\,0, \sigma^2)\,.$$

This study is testing whether a firm will fail or not which requires bankruptcy to be a dichotomous variable. Bankruptcy is coded 1 if the firm defaults and 0 if the firm is viable. With a logistic function, a binary variable can be expressed in terms of probability. Since probability is bound to values between 0 and 1, a linear regression is not appropriate as it can take on negative values and those greater than one. The nonlinear association between the independent variables and binary dependent variable violates one of the basic assumptions of linear regression; linearity. To correct for this restriction, a logarithmic transformation is used. First quarter 2005 to fourth quarter

2014 comprise the in-sample data from which the model parameters are estimated. The logit of the probability of bankruptcy is the natural logarithm of the odds ratio of bankruptcy.

b_{it} = probability of bankruptcy

$$logit(b_{it}) = \ln\left(\frac{b_{it}}{1-b_{it}}\right) = \beta_0 + \ \beta \Delta \boldsymbol{X}_{it-1} \tag{1}$$

where $\boldsymbol{X}$ is a matrix of independent variables Matrix $\boldsymbol{X}$ in equation (1) denotes all independent variables (both the continuous and binary dummy variables) lagged one period, to control for time fixed effects (Stat501, 2017).

$$y_{it} = a + \beta_1 x_{it-1} + \beta_2 x_{it-1} + \cdots + \beta_n x_{it-1} + \varepsilon_{it} \tag{2}$$
$$\text{for } i = 1,2, \dots , n$$

$$\begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & \cdot \\ & \cdot \\ & \cdot \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \tag{3}$$

$$z = \beta_0 + \ \beta \Delta \boldsymbol{X}_{it-1} \tag{4}$$

$$b_{it} = f(z) = \frac{1}{1+e^{-z}} \tag{5}$$

where $f(z)$ is the estimated probability of bankruptcy with values between 0 and 1.

Equation (4) $z = \beta_0 + \ \beta \Delta \boldsymbol{X}_{it-1}$ is exponentiated always to be positive. The denominator is specified slightly larger than the numerator to ensure the value of the function does not exceed 1. Taking the antilog of the logit and solving for b_{it} yields the logistic function.

$$\hat{b}_{it} = f(z) = \frac{1}{1+e^{-(\alpha+\,\beta X_{it-1}\,)}} \tag{6}$$

Multicollinearity is an issue when regressing financial variables that contain the same factors within the ratio. The stepwise procedure is used to eliminate independent variables that are highly correlated with one another as well as correct for serial correlation (Diebold, 2007). Serial correlation is a result altering issue when the error terms are correlated with past values of itself. These issues, multicollinearity, and serial correlation are resolved with the forward stepwise procedure. This procedure begins with the simplest form of the model only regressing the intercept and one variable. In the next iteration adds an independent variable and so on for n independent variables in the model, then evaluates which predictors are the most significant. From stepwise logistic regression, the remaining independent variables cash flow ratio, quick ratio, debt ratio, profitability, and size are kept in the final model.

The acceptable level of correlation (Danilov, 2014) is verified from a Pearson correlation matrix that the remaining ratios were below 0.3. The model is chosen to maximize the log likelihood ratio and the covariance proportion in relation to the sum of bias and variance proportion (Orlowski, 2017). The forecast estimates three months ahead the probability a firm will go bankrupt. The forecast function meets the selection criteria when the root mean square error[5] is minimized (Orlowski, 2017).

[5] (RMSE) = $\sqrt{\frac{\sum_{t=1}^{T} e^2}{T}}$

5. Results

The Durbin-Watson statistic (DW) has been computed (Diebold, 2007) as

$$\mathrm{DW} = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \; . \qquad (7)$$

DW statistic is close to 2, indicating the absence of serial correlation. The low Akaike Information Criterion (AIC)[6] lends support that the model selection is ideal and has a cross-validated forecast error variance which is stable (Diebold, 2007). Stepwise regression allows for (Diebold, 2007) the continuous variables 1) cash flow ratio, 2) quick ratio, 3) debt ratio, 4) profitability, and 5) size, to be identified as significant. These five factors are kept in the final model (Modina & Pietrovito, 2013). These variables are significant at the 1% and 5% level and exhibit low correlation with one another.

The main results recorded in Table IV support the rejection of the null hypothesis, namely that the prediction of bankruptcy is not possible from financial ratios found on a firm's balance sheet, cash flow, and income statements. Concerning classification accuracy, the predicted sample did well with --% accurate and --% misclassified. Type I errors, --% were classified as failed and survived. Type II errors—firms which were predicted not-to-fail and did, in fact, fail—comprise --% of the entire sample. A confusion matrix summarizing these results is found appended after references (Table A.7).

[6] AIC $= e^{\left(\frac{2k}{T}\right)} * \frac{\sum_{t=1}^{T} e_t^2}{N}$

The McNemar test (Daniel, 1978) is applied to the number of failed and nonfailed firms for years 2005 and 2014 (Table A.8) and is calculated from the following equation.

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} \tag{8}$$

McNemar's test is a measure of the differences in paired proportions for a dichotomous response variable and two classification variables dependent on one another (MedCalc Software bvba, 2017), at two points in time. The frequency of each of the four mutually exclusive groups is recorded in a 2 x 2 results table. The adjusted R^2, calculated as

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-k}\sum_{t=1}^{T} e_t^2}{\frac{1}{T-1}\sum_{t=1}^{T}(y_t-\bar{y})^2} \tag{9}$$

indicates --% of the variation in whether a firm fails or not *is* explained by the logistic regression model (Diebold, 2007). The Chi-Squared test is used in place of the F-test due to nonlinearity a small value of [insert number] indicates the observed data fit the estimated data successfully. The coefficients on each predictor variable are significantly different from zero (Chi-square tests, 2017), p-values for each variable is below 0.05, and I reject the null hypothesis.

The standard errors of the regression (SER) were computed using the following equation.

$$\text{SER} = \sqrt{s^2} = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{T-k}} \tag{10}$$

The SER data units are the same as the error terms($_n$), making it simpler to interpret than standard errors (SE) which would require conversion (Diebold, 2007). The reliability of the forecast is indicated by a minimized Theil inequality coefficient (Orlowski, 2017). The coefficient is [insert number close to zero] indicating the forecast is reliable. See Appendix Figure B.2 for empirical results.

The debt ratio has the largest impact on the probability of default with a coefficient of [insert number] and t-statistic of [insert number] significant at the 1% level. This odds ratio, which is the estimated linear function exponentiated reveals for every one percent increase in the debt ratio the firm is [insert number] times more likely to fail. Cash flow ratio with coefficient negative [insert number] and t-statistic [insert number] is also significant at the 1% level. Cash flow is a vital component of a company's viability enabling the firm to meet its contractual obligations and operate with few financial constraints (Succurro & Mannarino). Higher cash flow ratios negatively affect the probability of bankruptcy. A firm that lowers its liabilities or increases its cash flow from operations, will decrease their likelihood of bankruptcy. The quick ratio with negative [insert number] coefficient decreases the probability of bankruptcy. Firms with highly liquid assets have a larger negative impact on the odds of default. However, firms with a low quick ratio value will have a smaller negative effect on the likelihood of default. The negative coefficient for quick ratio verifies the earlier expectation. Profitability (ROE) with negative [insert number] coefficient decreases the probability of bankruptcy. Firms with positive returns significantly decrease the likelihood of default. The size of the firm (market capitalization) decreases their probability of bankruptcy by [insert number] for every one unit increase in size holding other variables constant. Larger firms have more capital and access to the markets. Large firms have the advantage with specialized departments (Ozkan, 1996), such as accounting and marketing. Large investments and risks are not as feasible for smaller firms. These reasons are consistent with the earlier expectation increasing the size of the firm decreases the likelihood of firm failure. Growth has a negative [insert number] coefficient signifying firms with large research and development expenditures ($^{\text{R\&D}}/_{TA}$) relative to total assets, greatly reduces the probability of bankruptcy. Growth is an investment in the future revenues of the firm. The more invested in research and development the greater opportunity the firm has to increase its assets.

Table IV
Main Results

The dependent variable: b_{it} = probability of bankruptcy				
	Coefficient	Standard Error	*p*-value	Odds ratio
Cash flow ratio				
Quick ratio				
Debt ratio				
Profitability				
Size				
Growth				
Chi-Sq				
Log-L				
$\bar{R}^2$				
Theil Coef.				
RMSE				
Number of observations				

Note: t-statistics are in parenthesis *, **, *** indicates significance at the 10%, 5% and 1% respectively. The in-sample data set includes observations from 01/03/2005 to 12/31/2014. Out-of-sample data includes observations from 01/03/2015 to 03/31/2015. Dummy variables for the industry, filed on time, restatement and year were included in the regression. Chi-Square test is used for the non-linear function.

6. Conclusion

This paper examines several accounting ratios and their impact in predicting corporate failure. The McNemar test statistic indicates that corporate failure can be effectively predicted. Specifically, the results identify debt structure and the cash flow ratio to be the most important factors in determining the probability of bankruptcy. Armed with these results, corporate finance departments can evaluate and design financial interventions that will reduce a firm's overall likelihood of failure. Further research could expand on the work presented here by 1) exploring the use of Tobit analysis to forecast the likelihood of a company going bankrupt and how long it will take for the event to occur, 2) using classification trees, artificial neural networks and other methods to evaluate the existence or non-existence of possible non-linear relationships.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 589-609.

Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research, 4*, 71-111.

Chen, R.-R., Hu, S.-y., & Pan, G.-G. (2008). *On the Determinants of Bankruptcy Prediction.* Ping-Tung: National Ping-Tung Universities of Sciences and Technologies.

Chi-square tests. (2017, July 7). Retrieved from Inspire.stat: http://inspire.stat.ucla.edu/unit_13/

Dang, C., & Li, F. (2015, April 8). *Measuring Firm Size in Empirical Corporate Finance.* Retrieved from extranet.sieo.org: https://extranet.sioe.org/uploads/isnie2015/li_dang.pdf

Daniel, W. W. (1978). *Applied nonparametric statistics.* Boston: Houghton Mifflin.

Danilov, K. A. (2014). *Corporate Bankruptcy: Assessment, Analysis and Prediction of Financial Distress, Insolvency, and Failure.* Boston: MIT Sloan School of Management.

Diebold, F. X. (2007). *Elements of Forecasting.* M ason: Thomson South Western.

Hensher, D. A., Jones, S., & Greene, W. H. (2007). *An Error Component Logit Analysis of Corporate Bankruptcy and Insolvency Risk in Australia.* Sydney: The economic society of Australia.

Hol, S., Westgaard, S., & Van der Wijst, N. (2002). *Capital Structure and the Prediction of Bankruptcy.* Trondheim: Norwejian University of Science and Technology.

Lim, T. C., & Lim, X. Y. (2012). Bankruptcy Prediction: Theoretical Framework Proposal. *International Journal of Management Sciences and Business Research*, 69-74.

Logit Models for Binary Data. (2017, June 22). Retrieved from Princeton: http://data.princeton.edu/wws509/notes/c3.pdf

McNemar's Test. (2017, July 6). Retrieved from Statistics Solutions : http://www.statisticssolutions.com/non-parametric-analysis-mcnemars-test/

MedCalc Software bvba. (2017, July 7). Retrieved from McNemar Test: https://www.medcalc.org/manual/mcnemartest2.php

Mihalovic, M. (2016). Performance Comparison of Multiple Discriminant Analysis and Logit Models in Bankruptcy Prediction. *Economics and Sociology*, 101-118.

Modina, M., & Pietrovito, F. (2013). *Rating, Capital Structure and Bankruptcy Prediction.* Campobasso: Universita degli Studi del Molise.

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 109-131.

Orlowski, L. (2017). *Lectures 1,2 and 3.* Stamford: Sacred Heart University.

Ozkan, A. (1996, March). *Costs of Financial Distress and Capital Structures of Firms.* Retrieved from http://etheses.whiterose.ac.uk/2502/1/DX197774.pdf

Stat501. (2017, July 7). Retrieved from Online courses science: https://onlinecourses.science.psu.edu/stat501

Stock Markets & Sector. (2017, June 22). Retrieved from Bloomberg: https://www.bloomberg.com/research/sectorandindustry/overview/sectorlanding.asp

Succurro, M., & Mannarino, L. (n.d.). *The Impact of Financial Structure on Firm's Probability of Bankruptcy: a Comparison Across Western Europe Convergence Regions.* Calabria: Department of Economics, Statistics and Financial Sciences.

7. Appendix

Table A.1
Variable Description

Dependent variable	***Symbol***	***Description***
Bankrupt	b	Probability a firm will default ranges from 0 to 1. Bankrupt firms are coded 1; viable firms are coded 0.
Independent variables	***Symbol***	***Description***
Cash flow ratio	c	Cash flow from operations/ current liabilities
Quick ratio	q	(Current assets – inventories) / current liabilities
Debt ratio	r	Total liabilities/ total assets
Profitability	p	Return on equity = Net income/ shareholders' equity
Size (market cap)	s	Total number of shares × share price
Growth	g	Research & development expenditures/ total assets
Dummy variables		***Description***
Failed to file statements on time	d_1	Dummy variable coded 1 for firms that failed to file financial statements on time; 0 otherwise
Restatement	d_2	Dummy variable coded 1 for firms that restated financial statements; 0 otherwise
Dummy variables		***Control for year specific economic climate***
Year 2005	d_3	Dummy coded 1 when year is 2005; 0 if not in year
Year 2006	d_4	Dummy coded 1 when year is 2006; 0 if not in year
Year 2007	d_5	Dummy coded 1 when year is 2007; 0 if not in year
Year 2008	d_6	Dummy coded 1 when year is 2008; 0 if not in year
Year 2009	d_7	Dummy coded 1 when year is 2009; 0 if not in year
Year 2010	d_8	Dummy coded 1 when year is 2010; 0 if not in year
Year 2011	d_9	Dummy coded 1 when year is 2011; 0 if not in year
Year 2012	d_{10}	Dummy coded 1 when year is 2012; 0 if not in year
Year 2013	d_{11}	Dummy coded 1 when year is 2013; 0 if not in year
Year 2014	d_{12}	Dummy coded 1 when year is 2014; 0 if not in year
Year 2015	d_{13}	Dummy coded 1 when year is 2015; 0 if not in year

Note: All accounting data are taken from CRSP/Compustat Merged Database and Securities and Exchange Commission historical EDGAR database. Years 2005-2015, quarterly periodicity.

Table A.2

Industry Dummy Variable Description

Industry		***Control for the effect of industry sector***
Crop production	d_{14}	Dummy variable coded 1 for firms in the crop production industry; 0 otherwise
Construction of buildings	d_{15}	Dummy variable coded 1 for firms in the construction of buildings industry; 0 otherwise
Textile mills	d_{16}	Dummy variable coded 1 for firms in the textile mills industry; 0 otherwise
Food manufacturing	d_{17}	Dummy variable coded 1 for firms in the food manufacturing industry; 0 otherwise
Merchant wholesalers	d_{18}	Dummy variable coded 1 for firms in the merchant wholesalers, durable goods industry; 0 otherwise
Furniture stores	d_{19}	Dummy variable coded 1 for firms in the furniture and home furnishings stores industry; 0 otherwise
Air transportation	d_{20}	Dummy variable coded 1 for firms in the air transportation industry; 0 otherwise
Real estate	d_{21}	Dummy variable coded 1 for firms in the real estate industry; 0 otherwise
Hospitals	d_{22}	Dummy variable coded 1 for firms in the hospital industry; 0 otherwise

Note: All industry-specific accounting data taken from CRSP/Compustat Merged Database. Years 2005-2015.

Table A.3
Analysis of Moments

Regressors	Mean	Median	Skewness	Kurtosis
Cash flow				
Quick ratio				
Debt ratio				
Profitability				
Size				
Growth				

Table A.4
Total Number of Sampled Firms in Each Industry

Industry	Number of companies
Construction of buildings	
Textile mills	
Food manufacturing	
Merchant wholesalers	
Furniture stores	
Air transportation	
Real estate	
Hospitals	

Note: Table format inspired by Ozkan (1996). Source CRSP/Compustat Merged Database.

Table A.5
The Number of Failed Firms in Each Year

Year	Number of companies
2005	
2006	
2007	
2008	
2009	
2010	
2011	
2012	
2013	
2014	
2015	

Note: Source CRSP/Compustat Merged Database. Table inspired by Ozkan (1996).

Table A.5 describes the frequency (number of firms) of each industry's bankrupt population annually, for the entire sample period.

Table A.6
Annual Means of Variables

Variable	2005	2006	2007	2008	2009	2010	2011	2013	2014	2015
Cash flow ratio										
Quick ratio										
Debt ratio										
Profitability										
Size										

Note: The means of 2015 are calculated from the first quarter only. Compustat Merged Database.
Table format influenced by Ozkan (1996).

Table A.6 presents the mean independent variable value, annually, reporting on in-sample dataset characteristics.

Table A.7
Confusion Matrix

		Actual outcome	
		Non-fail	Fail
Predicted outcome	Fail	Type I error	Correct prediction
	Non-fail	Correct prediction	Type II error

Table A.7 describes the contingency table for calculating the *classification accuracy*. Predicted outcomes reported on the left and actual outcomes are on the right. There are only four possible outcomes. They are (**a**) correct fate of firm predicted—firms who were forecast survived and were actually a survivor. Erroneous failed predictions (**b**)—firms who would have survived if they had been forecast a survivor. Erroneous failures in (**c**)—firms who were predicted to survive and later failed. Correct failures (**d**)—firms who were predicted to fail, who actually do end up failing. (McNemar's Test, 2017)

Table A.8
Contingency Table

	Year 2014			
Year 2005		Test 2 Non-Bankrupt	Test 2 Bankrupt	Row total
	Test 1 Non-Bankrupt	**a**	**b**	**a** + **d**
	Test 1 Bankrupt	**c**	**d**	**c** + **d**
	Column total	**a** + **c**	**b** + **d**	**n**

In Table A.8 the proportion of correct *classification accuracy* is computed for each year by summing the number of correct predictions (groups **a** and **d**) and dividing this total by $\boldsymbol{n} - -$ the total number of predictions (MedCalc Software bvba, 2017). The *classification accuracy* is evaluated in both the years [2005, and 2014]. Classification accuracy is *the* basis for comparing the two years' accuracy rates—how well the predictions did, compared with the actual outcomes.

7.1 Figures

Figure B.1 Forecast Summary

Figure B.1 is an evaluation summary of the forecasted model.

Figure B.2 Histogram

Frequency

Logit

The histogram in Figure B.2 is taken with the Chi-square test for nonlinear data to determine frequency and distribution. This test analyses data as if it had been normally distributed. The sample appears normally distributed, confirmed by a small p-value of [insert number]. The mean probability is slightly positive and the small Jarque-Bera statistic close to 0 indicates the data is normally distributed.

Figure B.3 Forecast

Figure B.3 displays the regression of y on x and z, and the fitted and actual values are presented.

.

Figure B.4 Residuals Plot

Figure B.4 the residuals plot shows no obvious pattern in the errors which is ideal (Diebold, 2007)