Credit Risk Modeling: Default Probabilities

Jaime Frade

December 28, 2008
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 Comparison and Predications</td>
<td>20</td>
</tr>
<tr>
<td>8.1 Comparison of Results</td>
<td>20</td>
</tr>
<tr>
<td>9 Conclusion</td>
<td>26</td>
</tr>
</tbody>
</table>
List of Figures

6.1 Correlation Matrix .................................................. 12
6.2 SAS Output: Saturated Logistic Model ........................... 13
6.3 SAS Output: Reduced Logistic Model 2 .......................... 14
6.4 SAS Output: Reduced Logistic Model 3 .......................... 16
6.5 Excel Output: Logistic Models ................................. 17

7.1 Confidence Intervals for Accuracy Ratios of Model .......... 19

8.1 Sorted Comparison of Prediction Results: By Reduced Logistic Model 3 ........................................... 22
8.2 Sorted Comparison of Prediction Results: By Reduced CRE Model ............................................... 23
8.3 Sorted Comparison of Prediction Results: By Reduced CDP Model ............................................... 24
8.4 Sorted Comparison of Prediction Results: By Reduced Altman’s Z-Score ........................................ 25
Acknowledgments

I would like to thank Dr. Niu for helping me learn the capabilities to incorporate statistical models and computational techniques in the real world. The foundations and principals in this paper are largely due to his instruction and guidance. I would also like to thank my family and friends for supporting me when my hard drive crashed and I lost all my data and this paper. They helped me to stay focus and build this from scratch.

I would also like to thank my supervisor, Kevin Ceuvorst, and colleague, Ken Hill, at Florida State Board of Administration. They helped me to combine my statistical and computational capabilities with their financial knowledge to produce the models in this paper. They provided guidance and critical ideas that helped me in this paper’s model building process.
Abstract

In this paper, the main objective is to build a quantitative model that estimates the probability that a US issuer will default on public debt within a year, as a function of predictors that include financial ratios and equity market variables. The model is structured to balance timeliness and stability, and also allow company- or industry-level qualitative analysis.

The basic output of the model is a 1 year horizon forward estimate of default probability. The model is uses logistic regression to derive the risk of insolvency within a year as a function of financial and market ratios. The steps taken in this paper are as follows:

- Determine a logistic regression model
- After determining a statistically significant model, will show how to estimate probabilities from current data and interpret a logit model.
- Discuss the significance and relationship of the variables.
- Comparison of the model with other currently accepted models.
- Apply validation techniques.
The source language used to compile the model was SAS and Microsoft Excel. Both programs were used for a confirmation of results. Excel was also incorporated so that future usage by other analysts can be easily updated because of the lack of a SAS license. Macros and programs in Excel were created and used in this paper that go along with the SAS results.

The goal of this paper was to build a statistical model that incorporate financial balance sheet items, which has been done in the past, but also to seek other market equity related variables. The drive in this direction was mainly due to the recent high volatility in late 2008 of the credit arena. The months in late 2008 has experienced many companies defaulting on loans and disappear from the markets, such as Bear Stearns and Lehman Brothers. Analyzing the market as a whole at this current time at the end of 2008, it is projected that many other companies may follow. As of the date of this paper, General Motors and Chrysler Automotive companies are at headline risk.

In this paper, it was found that not all market variables are significant in a logistic regression model. The two main models found that either Market Equity over Total Assets (a measurement of Leverage), Sales over Total Assets, (a measurement of Competitiveness), Earnings before Income Tax Adjusted over Total Assets (a ratio that measures Profitability), and the last traded price of an issuer (the only market indicator) for that date were significant variables to predicting probabilities of default within a 1 year time horizon. In this paper, the predications of the models are compared to other models that are currently accepted in the financial market as credit risk indicators. Overall, there are high correlations between the statistical models and the models of this paper. The logistic model can be used to predict default rates, however, more technical validation techniques need to be incorporated. Also, depending on availability, more issuers that have defaulted may be added to the model.
Introduction

1.1 Credit Risk Methodology

In recent months more emphasis has shifted to the modeling and evaluation of credit risk. There are several forces behind this trend. First, credit markets have grown steadily and the credit derivatives market (CDS) has grown exponentially. Second, credit risk is still a developing field. The understanding and the methodologies in managing and measuring other market risks, including interest rate risk, currency risk etc., have matured and are now assessed based on widely accepted principles.

Credit markets are implying significant levels of default over the next twelve months, but rating agencies, such as Moody’s or Standard and Poors, downgrades and realized defaults have yet to follow suit. This is apparent in the latest sub prime bubble bursts and the more recent defaults such as Lehman Brothers that have surprised the economy. Therefore, in order to stay afloat in the financial, one needs to become more active in credit risk management to avoid substantial losses in any position.

Any approach to credit risk management has a balance many often competing requirements. Objectivity, accuracy, and stability are all important, but not at the expense of timeliness. Last, but not clear, coverage is important for consistent decision-making across large portfolios. Given these requirements, the approach to credit risk is to steer a course between market signals and fundamental creditworthiness, but mixing financial statement and equity market data together within a credit scoring framework. By using financial statements, which will focus on the internal factors that drive company credit risk, and the inclusion of equity market signals adds timeliness, recognizing (as Merton did in 1974) that equity prices and equity volatility will be useful factors for the measurement of riskiness of an issuer’s debt. In this paper, have collected a database of risk factors and risk measures of individual companies to identify a either company level, or an industry level, probabilities of default using statistical models and analysis.
Chapter 2

Preliminaries

2.1 Financial Definitions

Throughout this paper, some financial terms are used. Below is provided a list of terms with definitions, also, what each measures in a financial aspect. When a corporation offers a stock or bond for sale, or a government offers a bond, the security is known as an issue, and the company or government is the issuer.

- **Bonds**: A debt instrument issued for a period with the purpose of raising capital by borrowing for the issuer.

- **Default**: The inability of a borrower to pay the interest or principal on a debt when it is due.

- **Liquidity**: The asset’s ability to be easily converted through an act of buying or selling without causing a significant movement in the price and with minimum loss of value.
Chapter 3

Methods of Modeling Credit Risk

3.1 Existing Methods of Quantitatively Scoring

Many of the common types of qualitative methods for evaluating risk target to explain the following general categories of a company. Overall, predictors should target explaining/correlating with this following framework. The most commonly used typology of risk factors for issuers is the one regulators use the CAMEL framework. The focus in this paper, is to determine predictors that targeted the following general areas which affect default concerns.

- **Capital Adequacy:** Amount of core capital to a issuer’s risk-weighted assets. This area is designed to relate the amount of core capital (less intangibles and other items) to a bank’s risk-weighted assets. Capital adequacy can be used to define insolvency itself, as well as to predict future insolvency. The key to avoiding confusion is that capital ratios used as a predictor (along with all other predictors) must be measured before the institution’s insolvency. That is, the question is whether poorly capitalized institutions later tend to strengthen their capital positions or to become yet more poorly capitalized to the point that they are so undercapitalized as to be defined as insolvent.

- **Asset quality:** Chosen on the basis that poor asset quality should increase the risk of failure. A common one is the ratio of nonperforming assets to total assets.

- **Management quality:** A qualitative measurement on the past actions and strength of management. While an important part of qualitative analysis of a issuer’s risk profile, is not measurable quantitatively.

- **Earnings:** Most common index of earnings is the ratio of net income to total assets. Research has shown that growth in either of these two quantities likewise should predict a reduced risk of insolvency. [5]

- **Asset Liquidity:** Issuers with assets that are more liquid should have lower failure risk, all else equal.
Each of the above areas may not pertain to all issuers, however, in the ultimate determination in deciding the stability of a company or entity the areas are basic tools to help focus attention in the right area. It is up the analyst in determining the methods/models of how to accurately and precisely target the general areas. The assessment made can either be qualitative, using the areas above, or quantitative. A quantitative credit risk scoring model can complement all the sources of qualitative judgment. Furthermore, a quantitative model offers breadth of coverage, including firms that are too small or that issue securities too infrequently to substantiate ongoing qualitative coverage. Furthermore, a quantitative model can be updated far more frequently than any other type of evaluation, and timely evaluations are more likely to be accurate. Finally, a quantitative model is valuable in the context of an integrated analytic approach because it provides a systematic algorithm for scoring risk that can be used as a check and starting point for in-depth qualitative analysis.

3.2 Two Classes of Quantitative Models

Existing quantitative models for scoring issuers’ credit risk fall into two broad classes: statistical and structural.

3.2.1 Statistical Models

Statistical Models: Use historical data on characteristics of issuer (for example, measures of earnings or liquidity) to determine the set of characteristics that best predict the occurrence of the selected outcome. The precise form of the relationship between the inputs and the outcome is specified by the particulars of the statistical model used. These models are historically specific: the model parameters depend on the data used to create the model.

3.2.2 Structural Models

Structural Models: models are form and parameters are, in principle, specified by a theory of economic structure. In this model, the common framework is to us a contingent claims models, which is based on the concept that the equity of a firm as a call option on the assets of a firm. In this approach, it uses the Black-Scholes (1973) framework and applies it to a debt pricing model by Merton Approach (1974). In this model, the strike price of the option represented by the firm’s equity is a function of the firm’s debt level, and if the value of the firm’s assets falls below that point the equity holders decline to exercise the call, turning over the assets of the firm to the debt holders. This is the point of default. The probability of that event is determined by the difference between the firm’s asset value and default point, along with the volatility of the firm’s asset value. The asset value is not observed, these quantities are generally estimated as a function of the firm’s equity value and volatility. [5]

There are challenges of applying contingent claims models to banks specifically, the models more generally share certain limitations. Through various studies, the default probabilities implied by Merton-type contingent claims models are inconsistent with historically observed default rates [3] (Falkenstein,
Although there are models that do attempt to compensate for this, the underlying fact is that exists market inefficiencies for many of the firms that make market valuation different. The main concern today, because of the deteriorating conditions is that market prices may no longer possess the desirable properties they exhibit under perfect and complete markets. (Beaver, Datar, and Wilson 1992: 261). If market prices were fully efficient and complete summaries of all relevant information, no additional factors could improve the prediction of firm riskiness. [2] But as Krainer and Lopez (2003) show in the case of banks, financial ratios are strong predictors of subsequent regulatory ratings for bank holding companies, and remain so even with the inclusion of equity variables as potential predictors. [4]

The Merton model is not statistically estimated. It is a theoretically specified mathematical statement of relationships among the market value of a firm’s equity, the value of its liabilities, and the volatility of its equity value. Statistical models specify only their component variables and the family of forms to which the relationships among them must belong, with the data being used to estimate the exact nature of those relationships. Statistical default models estimate those relationships using variants of regression-based techniques.

Overall, there are several challenges make contingent claims models difficult to successfully apply to certain industries, such as banks: a lack of any single dominant model for banks, data requirements that severely limit the scope of application for some models, and dependence of some models’ results on distributional assumptions. Furthermore, that market imperfections exist implies that nonmarket measures may add significant power in predicting default beyond that possible with even an ideal contingent claims model.

The importance note is that contingent claims model uses, but not limited to, equity markets information as a source of evaluations for the current conditions of any firm. In this paper, I will address the need to encompass not only equity data, but to incorporate financial ratios to build a statistical model that best predicts an issuer’s insolvency. The goal is to determine a logistic regression model that includes financial ratios in addition to equity-driven measures. While this model strategy attempts to explicitly predict default probabilities, this or any model, is not guaranteed from modifications due to the environment. It is a useful tool in examining credit risk, and provides another indicator to a broad spectrum of qualitative and quantitative tools that identify risk factors that are emerging and/or difficult to quantify.

Altmans Z-score

The earliest, Altmans Z-score (1968), used discriminant analysis. [5] Discriminant analysis shares with linear regression several assumptions (and adds assumptions of its own): predictors with multivariate normal distributions, normally distributed errors, linear relationships between independent and dependent variables, and a dependent variable with roughly similar size groups. Violating those assumptions, as default data and many other kinds of data often do, can produce biased results.

In the decades since Altman published his Z-score model, logistic regression has replaced discriminant analysis as the preferred tool for dichotomous outcomes, largely because logistic regression has assumptions that are less restrictive, and so less frequently violated. This model allows for predictors that
need not be normally distributed and that are nonlinear in their effects on the probability of an event, and for dependent variables with different sized groups.
Chapter 4

Methodology

4.1 A Statistical Model: Logistic Regression

I apply a statistical model to historical data on issuers' characteristics. The particular form of statistical model is a discrete-time event history model. This model is designed to predict the risk of an event occurring, as a function of specified variables measured before the event occurs.

The linear regression (a discrete time model) can be used to predict the risk of an event within a certain time period. This is equivalent and estimated by applying a logistic regression to issuer-year of data.

The logistic regression takes the following form

\[
\log \left( \frac{p}{1-p} \right) = \sum_{k=1}^{K} \beta_k x_k \tag{4.1}
\]

where \( p \) is the probability of the event occurring, and \( K \) independent variables, \( x \), are each weighted by a coefficient, \( \beta \).

From 4.1, this different from the conventional linear regression model

\[
y = \sum_{k=1}^{K} \beta_k x_k \tag{4.2}
\]

From 4.2 differs from 4.1, because does not predict the value of the dependent variable, \( y \), but rather the natural logarithm of the ratio of the probability of the event occurring to the probability of the event not occurring.

This logit function preferable to the linear regression function because it limits \( p \), the probability of the event, to be between 0 and 1 (or zero and 100%), where applying the conventional linear regression to dichotomous outcomes would instead allow nonsensical results like probabilities greater than 1 or less than zero.

A transformation of 4.1, obtain the logistic model which gives the probability of the event as

\[
p = \frac{e^{\sum_{k=1}^{K} \beta_k x_k}}{1 + e^{\sum_{k=1}^{K} \beta_k x_k}} \tag{4.3}
\]
This difference in structure between this model and more familiar linear regression models gives rise to differences in how the results of the model can be interpreted. Changing one independent variable by a fixed amount changes the level of the dependent variable by an amount that is identical, no matter what the levels of the other independent variables.

If interpreting a linear regression to predict the risk of a bank failing, whether its current risk was 10%, decreasing its ratio of net income to assets by a fixed quantity would increase its risk of failure by the same amountlet’s say 1%, respectively 11%.

In contrast, in a logistic regression, a change in one factor changes the risk by an amount that is proportional to the level of the other factors. If the other factors produce a failure probability of 2%, decreasing the ratio of net income to assets might double the risk to 4%.

### 4.2 Choosing a Significant Model

An estimating algorithm is used to find the coefficients, \( \beta \)'s that best satisfy the relationship expressed in the regression equation for the estimation data sample. The technique used to find those coefficients for logistic regression, was using maximum likelihood estimation. Basically, method tries coefficients until it finds the set that maximizes the value of a mathematical function that gives the joint probability of observing the given data.

That function, \( L \), the likelihood function, forms the basis of a statistical test of how well the model fits the observed data:

\[
L^2 = 2(\log L_T - \log L_B)
\]  \hspace{1cm} (4.4)

where \( L_T \) is the likelihood function of the first model with smaller variables and the \( L_B \) is the likelihood function of a baseline model.

\( L^2 \) is a statistic that will be compared with the standard \( \chi^2 \) table to determine whether the tested model fits significantly better than a baseline model. This procedure is completed to establish whether the added variables significantly improve the fit to the data, or whether conversely the smaller subset is equally sufficient. We use this procedure to omit from the model variables that do not significantly improve our ability to predict insolvency.
Chapter 5

Data Description

In this paper, I obtained quarterly financial and market information from different issuers (ranging from publicly-traded US bank holding companies to industrial industries, etc) during 1996-2008. Issuers consisted of myriad of companies that have either defaulted or are currently stable. The data for the predictors was obtained from the financial statements from each issuer. There were 186 individual issuers that were collected. Each issuer reported 4 data points of each variable for one year.

5.1 Defaulted Bonds

To provide a encompass a wider selection of issuers, defaulted bonds were chosen industry and time independent. Issuers that defaulted most (but not all) of these subsequently facing regulatory intervention as well. In total there were 37 issuers that have defaulted in the time span of this model’s data collection.

- **Airlines**: Delta Air Lines Inc, Northwest Airlines Corporation, FLYi, Inc,
- **Financial Insitutions**: Lehman Brothers, Bear Stearns, WAMU
- **Delphi Corporation, Levitz Home Furnishings, Inc, Winn-Dixie Stores Inc, Calpine Generating Co LLC**

5.2 Data

As mentioned in an early section, predictors should target explaining/correlating with the following CAMEL framework. This is the most commonly used typology of risk factors for banks is the one regulators use to summarize the results of their on-site examinations. These framework is setup to qualitatively or quantitatively represent the amount of cash the company has on hand, incoming cash flow of a company, the strength of the corporate structure, and how easily the company can liquidate the assets. The financial variables listed below were chosen to target the stability of the company. A subset of the variables listed below are used in common practice in calculation of the Altman Z-score (Altman 1968) [1]. Altman, using Multiple Discriminant Analysis, combined a
set of financial ratios to come up with the Z-Score. This score uses statistical techniques to predict a company’s probability of failure.

The predictors will use information from the financial and balance sheet of an issuer. Listed below are the definitions and/or explanation of the parameters. Also to incorporate market behavior, there are equity-driven measures added to the possible set of parameters.

- **Working Capital**: Operating liquidity available to a business, calculated as current assets minus current liabilities.

- **Retained Earnings**: The percentage of net earnings not paid out as dividends, but retained by the company to be reinvested in its core business or to pay debt.

- **Earnings Before Debit and Interest (EBIT)**: A company’s earning power, represents the earnings which the company has achieved. Measure of a issuer’s profitability.

- **Market Value of Equity**: The price at which investors buy or sell an issue.

- **Equity Market Volatility**: Movements in the market in which it is traded.

- **Sales**: Total dollar amount collected for goods and services provided.

- **Total Liabilities**: The liabilities found by adding current liabilities to long-term debts.

- **Total Assets**: All the property owned by a corporation.

- **Net Income**: The company’s total earnings, minus expenses and taxes.

Equity market indicators may serve as vital flags of risk that may not be reflected in the company’s financial statements. Equity markets identify those believed to be taking on greater risk, by a higher total equity volatility. Companies with greater equity volatility should be more likely to become insolvent.

For the model in this paper, ratios of the above financial parameters were used so that the predictors can not only encapsulate the standard CAMEL framework, but also incorporate market behaviors. For example, a company’s debt ratio is a leverage ratio calculated by dividing total liabilities by total assets. This ratio measures the extent to which total assets have been financed with debt.

### 5.3 Predictors for Logistic Regression Model

The following are the ratios and equity market information that were calculated from the collected parameters for each issuer. From financial analysis, when analyzing certain ratios each targets certain areas of credit risk fundamentals.

- **Working Capital/Total Assets**: Measures the short-term liquidity of the issuer.
• **Retained Earnings/Total Assets**: Measures Historic Profitability
• **EBIT/Total Assets**: Measures Current Profitability
• **Market Equity/Total Liabilities**: Measures Leverage
• **Sales/Total Assets**: Approximated the competitive situation of the company
• **Net Income/Sales**: An equivalent measurement of the Profit Margin
• **Equity Market Shares Volatility**
• **Price of Last Trade on Market**
Chapter 6

Results

6.1 Saturated Logistic Regression Model 1

Preliminary analysis began by creating a “saturated” model including all predictors (ratio) described in the previous section. There were concerns of that several ratios are correlated. The correlation matrix of the variables is provided below. A decision was made to include only parameters with insignificant correlation in the model. It should be noted, however, that in the aggregate the results would have been essentially identical with either variable, given that the two ratios are empirically virtually indistinguishable.

A correlation matrix is provided. In preliminary analysis, the correlations between EBITA/TA and EBIT/TA was high. Running the model with each variable, it was determined to use only EBITA/TA. Also, NI/TA, also showed higher correlations across several other parameters. The variable was dropped from the model possibilities, and furthermore, the variable was found to be insignificant, which will be illustrated in the outputs.

![Figure 6.1: Correlation Matrix](image-url)
6.1.1 Saturated Logistic Regression Model 1: Output and Diagnostics

Using SAS and Excel, I obtained the following outputs. SAS is listed below and Excel is provided later. Both programs confirmed the results. In the saturated model it was found that **EQY SH OUT** (Equity Share Values), **WC/TA** (Working Capital/Total Assets), **RE/TA** (Retained Earnings/Total Assets), and **EBIT/TA** to be insignificant to the model. These parameters were taken out and the likelihood ratio will be calculated with the second model. Although the p-values were insignificant on **EBITA/TA** and **PX LAST** (Last trading price), each was left in the analysis to determine a reduced model.

**SAS OUTPUT: Log Model 1**

![SAS Output](image)

6.1.2 Saturated Logistic Regression Model 1 Interpretation

- Then if **ME/TL** changes increases by 1, the odds that the default takes the value 1 increase $e^{-0.2396} = 76.69\%$
6.2 Logistic Regression Reduced Model 2

After rerunning the model with parameters removed in the saturated model it was found that all remaining parameters to be significant to the model. However, analyzing the p-value from the likelihood ratio test, it was determined that we can not reject the null hypothesis and there is no different between either models. The likelihood ratio test for the hypothesis for LR- Test for \( b(EQY\ SH\ OUT) = b(WC/TA) = b(RE/TA) = b(EBIT/TA) = 0 \) in model 1 is based on the log likelihoods of the two models. It has a \( \chi^2 \) distribution with 5 degrees of freedom because imposed 4 restrictions.

Because of the variables and what each represents, the tendency may toward model 2. Further validation techniques will be discussed in a later chapter to compare all three models with accuracy ratios and a bootstrapping method will be used to create confidence intervals from the predictions. From the Excel output, the LR=4.81 and the SAS output LR=4.82.

From the Excel output, the p-value can be interpreted as if were to add the eliminated variables to model 2, there is a probability of 30.76% that we do not add explanatory power.

6.2.1 Logistic Regression Model 2: Output and Diagnostics

![SAS OUTPUT: Log Model 2](image)

Figure 6.3: SAS Output: Reduced Logistic Model 2

6.2.2 Logistic Regression Model 2: Interpretation

- Then if ME/TL changes increases by 1, the odds that the default takes the value 1 increase \( e^{-0.200} = 0.8187 \).
• Then if $\frac{SA}{TA}$ changes increases by 1, the odds that the default takes the value 1 increase $e^{0.0638}$
6.3 Logistic Regression Reduced Model 3

To analyze the market variables in comparison to a model that is solely on financial balance sheet information, model 3 is compared to model 2. The incorporation of market was included in the model building process to accommodate for large market movements, such as the the financial distress in recent times of 2008. The variable, PX LAST, is the last trade price of the issuer on the market.

Further validation techniques will be discussed in a later chapter to compare all three models with accuracy ratios from the predictions. From the Excel output, the LR=10.04 and the SAS output LR=7.17 with a degree of freedom of 1. In this model, with the p-value < 1%, model 3 can be interpreted as a better model.

From the Excel output, the p-value can be interpreted as if were to add the eliminated variable to model 3 PX LAST, there is a probability of 0.15% that we do not add explanatory power. This is a smaller contribution, so model 3 will be considered, however, further validation will be completed as well as predicting and comparing results to other statistical and structural model used in the financial market.

6.3.1 Logistic Regression Model 3: Output and Diagnostics

**SAS OUTPUT: Log Model 3**

![Figure 6.4: SAS Output: Reduced Logistic Model 3](image)

6.3.2 Logistic Regression Model 3: Interpretation

- Then if ME/TL changes increases by 1, the odds that the default takes the value 1 increase $e^{-0.1887} = 81.87\%$
• Then if SA/TA changes increases by 1, the odds that the default takes the value 1 increase $e^{0.0641}$

• Then if EBITA/TA changes increases by 1, the odds that the default takes the value 1 increase $e^{50.3147}$

6.3.3 Logistic Regression Model: Excel Output

<table>
<thead>
<tr>
<th>Model 1</th>
<th>CONST</th>
<th>ME/TL</th>
<th>S_TA</th>
<th>EBITA/TA</th>
<th>PX_LAST</th>
<th>EOY_SH</th>
<th>OUT</th>
<th>W/C/TA</th>
<th>R_ETA</th>
<th>EBITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>-3.736</td>
<td>0.246</td>
<td>2.9753</td>
<td>867</td>
<td>-0.035</td>
<td>-0.082</td>
<td>-1.637</td>
<td>-0.687</td>
<td>20.570</td>
<td></td>
</tr>
<tr>
<td>SE(b)</td>
<td>0.781</td>
<td>0.102</td>
<td>1.6678</td>
<td>0.999</td>
<td>0.036</td>
<td>0.303</td>
<td>1.2967</td>
<td>1.430</td>
<td>59.935</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0.32</td>
<td>2.41</td>
<td>1.448</td>
<td>1.27</td>
<td>-1.83</td>
<td>-0.90</td>
<td>0.3595</td>
<td>-0.480</td>
<td>0.643</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.016</td>
<td>0.062</td>
<td>0.205</td>
<td>0.067</td>
<td>0.970</td>
<td>0.697</td>
<td>0.631</td>
<td>0.520</td>
<td></td>
</tr>
<tr>
<td>Pseudo R^2/ # iter</td>
<td>0.325</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR test / p-value</td>
<td>76.2</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inf / Inf0</td>
<td>-71.6</td>
<td>-115.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2</th>
<th>CONST</th>
<th>ME/TL</th>
<th>S_TA</th>
<th>EBITA/TA</th>
<th>PX_LAST</th>
<th>EOY_SH</th>
<th>OUT</th>
<th>W/C/TA</th>
<th>R_ETA</th>
<th>EBITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>-3.6728</td>
<td>0.1999</td>
<td>29061</td>
<td>762</td>
<td>-0.0037</td>
<td>-0.0037</td>
<td>-0.0037</td>
<td>-0.0037</td>
<td>-0.0037</td>
<td></td>
</tr>
<tr>
<td>SE(b)</td>
<td>0.49814</td>
<td>0.056164</td>
<td>13817</td>
<td>1563</td>
<td>1654862</td>
<td>0.02372</td>
<td>0.066081</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>-4.4362</td>
<td>3.509902</td>
<td>2.10392</td>
<td>2.10392</td>
<td>2.10392</td>
<td>2.10392</td>
<td>2.10392</td>
<td>2.10392</td>
<td>2.10392</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.00072</td>
<td>0.054544</td>
<td>0.00072</td>
<td>0.054544</td>
<td>0.00072</td>
<td>0.054544</td>
<td>0.00072</td>
<td>0.054544</td>
<td>0.00072</td>
<td></td>
</tr>
<tr>
<td>Pseudo R^2/ # iter</td>
<td>0.300034</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR test / p-value</td>
<td>71.35699</td>
<td>1.176-14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inf / Inf0</td>
<td>-80.1632</td>
<td>-115.845</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Diagnostics**

LR Test for b(EOY_SH_OUT)<b(W/C/TA)<b(RETA)<b(EBITA)=0 in model 1

| LR    | 4.81 |
| p-value | 0.3076 |

<table>
<thead>
<tr>
<th>Model 3</th>
<th>CONST</th>
<th>ME/TL</th>
<th>S_TA</th>
<th>EBITA/TA</th>
<th>PX_LAST</th>
<th>EOY_SH</th>
<th>OUT</th>
<th>W/C/TA</th>
<th>R_ETA</th>
<th>EBITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>-4.522</td>
<td>0.271</td>
<td>29074</td>
<td>632</td>
<td>-79.663</td>
<td>12591</td>
<td>227</td>
<td>13.883</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE(b)</td>
<td>0.416</td>
<td>0.047</td>
<td>12591</td>
<td>227</td>
<td>13.883</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>-10.48</td>
<td>5.97</td>
<td>2.31</td>
<td>-9.74</td>
<td>2.31</td>
<td>-9.74</td>
<td>2.31</td>
<td>-9.74</td>
<td>2.31</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.021</td>
<td>0.000</td>
<td>0.021</td>
<td>0.000</td>
<td>0.021</td>
<td>0.000</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>Pseudo R^2/ # iter</td>
<td>0.255</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR test / p-value</td>
<td>61.4</td>
<td>0.699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inf / Inf0</td>
<td>-85.2</td>
<td>-115.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Diagnostics**

LR Test for b(P_LAST)=9 in model 2

| LR    | 0.04 |
| p-value | 0.815 |

Figure 6.5: Excel Output: Logistic Models
Chapter 7
Validation Techniques

Once a model is set up, there is a need to access its quality. In order to capture how well do the estimated probabilities of defaults match the true defaults, there is a need to derive a method beyond goodness of fit techniques. The common technique of assessment is to analyze the accuracy ratio of the estimated probabilities. The ratio is a calculation of the cumulative accuracy ratio (CAP) curve. The CAP curve measures discriminatory power of the model's estimates. The curve is a way to visually see if the defaults are occurring with those issuers that have lower ratings. The area between the CAP curve and above the diagonal is the accuracy ratio. The diagonal line represents the probability of guessing. There is a need to maximize this area. The area above the curve is used because the latter gives the expected CAP curve of an uninformative rating system which does not discriminate at all between low and high default risks. [6] No CAP calculation of the cap curve was reported in this paper, because of the lack of information of the actual ratings of all the issuers in the sample space, such as AAA, Aa3,... However, for each model, the accuracy ratio was calculated.

7.1 Bootstrapping Confidence Intervals for Accuracy Ratios

The accuracy ratios are estimates of the models ability to discriminate between high and low credit risks against the actually defaulted issuers. In this paper, to obtain a confident estimate, used bootstrapping simulations to derive a confidence interval. After reading related papers on default models, much of the validation was presented using CAP and accuracy ratios. Furthermore, confidence intervals were presented, therefore, this paper will follow this procedure. Analyzing the outputs, there is small difference in the accuracy ratio, however, overall model 2 and 3 do substantiate results from previous sections.
7.1.1 Bootstrapping Confidence Intervals: Output

Figure 7.1: Confidence Intervals for Accuracy Ratios of Model.

| Default probability Model 1 | 85.29% | 68.84% | 95.28% |
| Default probability Model 2 | 84.13% | 68.48% | 96.43% |
| Default probability Model 3 | 85.22% | 68.42% | 94.24% |

7.1.2 Accuracy Ratios Confidence Intervals: Interpretation

- **Model 1**: The accuracy ratio is 85.29%, the bootstrapped 95% confidence interval is [68.84%, 95.28%]
- **Model 2**: The accuracy ratio is 84.13%, the bootstrapped 95% confidence interval is [68.48%, 96.43%]
- **Model 3**: The accuracy ratio is 85.22%, the bootstrapped 95% confidence interval is [68.42%, 94.24%]
Chapter 8

Comparison and Predications

8.1 Comparison of Results

In this paper, there are three other default models that are used to compare the results of the logistics regression outputs to currently accepted models that are structural and statistically based. The models and a brief description is provided below.

- CreditSights: BondScore CRE Default Model: Statistical logistic Model that incorporate economic factors (GDP)
- Barclay’s (formerly Lehman): BASE CDP Default Model: Structural based that incorporates on and off financial balance sheets items
- Altman Z-Score Model: Structural model that uses financial balance sheet items.

Of the three models, only CRE and CDP provide actual probabilities of default. The Altman Z-score provides a score that can be interpretation as a certain standard deviation away from an issuer’s likelihood of default. The layout is provided below. Also, not all models incorporate the same issuers that was used in this paper’s model building process. Because of proprietary usage, only the results of each of the three models can be provided. If the model does not have a score or probability, a #N/A is labeled. The results in the figure below are sorted by the yellow column. The rankings are also provided based on this paper’s reduced logistic model 3. It can be compared that many of the issuers that a likely to default are pairing up with the current models. There was a relatively high positively correlation CRE Model and model 1 (0.861727971), CRE Model and model 2 (0.905134001),CRE Model and model 3 (0.585281804). This may be because of the fact that all models are statistically based. There was a relatively high positively correlation CDP Model and model 1 (0.69824494), CDP Model and model 2 (0.813915174), CDP Model and model 3 (0.43992355). Each current model were more correlated with logistic model 2. Because of lack of data, the Altman Z-score was unable to be calculated. This Z-score is still
able derived, however, more modern techniques since its original release in 1968 have replaced this calculation.
Figure 8.1: Sorted Comparison of Prediction Results: By Reduced Logistic Model 3
Figure 8.2: Sorted Comparison of Prediction Results: By Reduced CRE Model
### Figure 8.3: Sorted Comparison of Prediction Results: By Reduced CDP Model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>American Home Lines</td>
<td>0.05%</td>
<td>0.03%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Boeing Aircraft</td>
<td>0.51%</td>
<td>0.48%</td>
<td>0.46%</td>
<td>0.44%</td>
<td>0.42%</td>
<td>0.40%</td>
<td>0.38%</td>
</tr>
<tr>
<td>Coca-Cola Enterprises</td>
<td>0.15%</td>
<td>0.10%</td>
<td>0.05%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Electric Capital Corp</td>
<td>0.05%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>FMC Corp</td>
<td>0.05%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>General Electric Co</td>
<td>0.15%</td>
<td>0.10%</td>
<td>0.05%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>General Motors Corp</td>
<td>0.05%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>IBM Corp</td>
<td>0.15%</td>
<td>0.10%</td>
<td>0.05%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

### Synopsis of Conclusions from Other Models

**CRE Model:**
- Credit Sights Model
  - Issuers with probabilities greater than or equal to 1% were classified as highly likely to default or be downgraded.

**Adjusted Corporate Default Probability:**
- Issuers with probabilities greater than or equal to 1% were classified as highly likely to default or be downgraded.

**The Interpretation of Altman Z-Score:**
- **Z-Score Above 1.5:** The company is in a healthy state. (See chart for details.)
- **Z-Score between 1.5 and 2.0:** (See chart for details.)
- **Z-Score Below 1.5:** Protocol of financial distress. (See chart for details.)
Figure 8.4: Sorted Comparison of Prediction Results: By Reduced Altman’s Z-Score
Chapter 9

Conclusion

In this paper, it was found that not all market variables are significant in a logistic regression model. The models found that the financial ratio variables that targeted a company’s current leverage position, the company’s current competitive advantages, the profitability margins were important factors in likelihood that the company would go into default on outstanding debt. Also, a market indicator that illustrates recent movements in the issuer’s value, price of last trade, was found to be significant in the model. The objective of this process was to incorporate more timely affected data to an issuer because of the recent down trends in credit quality across the spectrum of companies in the current financial markets.

The models provided in this paper help to point a credit analyst in the right direction, however, more validation techniques need to be provided on the past predictions of the model. Accuracy ratios are provided in this paper, however the lack of the capabilities to actually calculate CAP curves because of the lack of data in historical ratings may be researched in later updates of the model. For the time being, the model does have high correlations with other substantial metrics in the financial markets, but still lacks the testing and calibration techniques of a true quantitative analyst.
Bibliography


