Duquesne University
Duquesne Scholarship Collection

Electronic Theses and Dissertations

Spring 1-1-2017

Predicting Bankruptcy and Catastrophic Loss: A Portfolio Approach

Michael McKibben

Follow this and additional works at: https://dsc.duq.edu/etd

Recommended Citation
McKibben, M. (2017). Predicting Bankruptcy and Catastrophic Loss: A Portfolio Approach (Master's thesis, Duquesne University). Retrieved from https://dsc.duq.edu/etd/140

This Worldwide Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact phillipsg@duq.edu.

PREDICTING BANKRUPTCY AND CATASTROPHIC LOSS: A PORTFOLIO APPROACH

A Dissertation

Submitted to the McAnulty College and Graduate School of Liberal Arts

Duquesne University

In partial fulfillment of the requirements for

the degree of Master of Science in Computational Mathematics

By

Michael McKibben

May 2017

Copyright by

Michael McKibben

2017

PREDICTING BANKRUPTCY AND CATASTROPHIC LOSS: A PORTFOLIO APPROACH

By

Michael McKibben

Approved March 22, 2017

Name of Professor Professor of English (Committee Chair)	Name of Professor Professor of Physical Therapy (Committee Member)
Name of Professor Assistant Professor of Music (Committee Member)	Name of Professor Assistant Professor of History (Committee Member)
Name of Dean Dean, School Name Professor of Chemistry and Biochemistry	Name of Department Chair Chair, Department Name Professor of Mathematics

ABSTRACT

PREDICTING BANKRUPTCY AND CATASTROPHIC LOSS: A PORTFOLIO APPROACH

By

Michael McKibben

May 2017

Dissertation supervised by Dr. Abhay Gaur

This paper uses logistic regression to assign risk of catastrophic loss (defined as a loss of 80% or more of market cap value) to companies, and analyzes the subsequent returns of high risk and low risk portfolios. In the final model, the low risk portfolio had a three-year mean return of approximately 47%, with a catastrophic loss rate of 1.1%. The high-risk portfolio had a three-year mean return of approximately .5%, with a catastrophic loss rate of 29%. The paper expands upon a model developed by Dr. Abhay Gaur and Dr. Leo Rebholz in Rebholz's 2002 thesis, Bankruptcy as Cusp Catastrophe. This paper first validates the model, introduces a new variable, which examines financial momentum, and transforms the bankruptcy variable to catastrophic loss. The success of the model was viewed through a comparative approach of high and low risk portfolios.

ACKNOWLEDGEMENT

Special thanks to Dr. Abhay Gaur, my thesis advisor, for his mentorship and encouragement throughout the research process. Dr. Gaur has a gift for believing in his students, and I want to thank him for believing in me. Special thanks also to Dr. John Kern, Dr. Frank D'Amico, and Prof. Sean Tierney for their feedback and role on the thesis committee. I would also like to thank my family and friends for their support through the process and countless readthroughs, especially Sarah Papke, Wei Li, Kaska Adoteye, Maureen Shaughnessy, and Sierra Green.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

Bankruptcy is an ongoing part of the business cycle. Companies may fail outright, and enter Chapter 7 liquidation. Other companies may instead be somewhat profitable but struggle with debts, and renegotiate under Chapter 11. Companies may become vulnerable and merge with competitors or are taken private to refinance themselves. In all cases, a prospective investor would have been better off if they had avoided investing in the company. Various metrics have been used to evaluate the health of the stock and predict bankruptcy, one of the earliest and most well-known being the Altman Z score, published in 1968[1]. This paper will look to validate and expand upon a method used by Leo Rebholz in his thesis at Duquesne University in the early 2000's[2].

The ability to predict bankruptcy is largely dependent on time. If waiting until after bankruptcies had occurred, prediction would have 100% accuracy. If predicting a year in advance, the prediction may still be fairly accurate, especially if isolating a particular industry. The further in advance a prediction is made, the less accurate it tends to be, as there's less information and more random events, such as fraud, recession, new technologies and processes, etc.

The value of the knowledge that a company is going bankrupt also decreases closer to the event. An investor who exits a stock position multiple years before bankruptcy is declared, while the company still appears somewhat healthy, is generally much better off than the investor

---

[1] Altman, 1968
[2] Rebholz, 2002

who waited to exit their position until Chapter 11 is declared, or held their stock through the bankruptcy. With this in mind, the goal of a bankruptcy prediction model is to minimize the trade-off in accuracy of the bankruptcy prediction when predicting further into the future.

This paper extends the cusp catastrophe model introduced by Leo Rebholz and Abhay Gaur in “Bankruptcy as Cusp Catastrophe” in three ways. First, the model is validated by increasing the sample of companies tested from 400 to every publicly available traded company in the US over the past 25 years, from 1990 to present. This represents approximately 130,000 time points. Second, a new variable which examines stock momentum is introduced. Third, a portfolio model is introduced which examines the relative performance of two sets of portfolios, one filled with high risk companies and the other companies with low risk, according to the probabilities of catastrophic loss.

Cusp Catastrophe

The underpinning of a cusp catastrophe model lies in the idea that the combination of moves in cash flow, inventory, liabilities, and assets are apt descriptors of the health of the business. A company undergoing a time of negative profitability can be buffered by a strong balance sheet. However, over time the company will increase liabilities, decrease assets, and run down inventory. There is a moment where the business can be considered to pass a point of no return. They will not have the cash on hand to fund all of their needs, such as capital improvements, payroll, inventory, and other business operations. The length of this period will depend on the original health of the business and the amount of cash burned during the period of unprofitability. Once a company has reached this critical point, it's incredibly difficult for them to recover, short of a debt reorganization or recapitalization. Consider the following view of the cusp from the side:

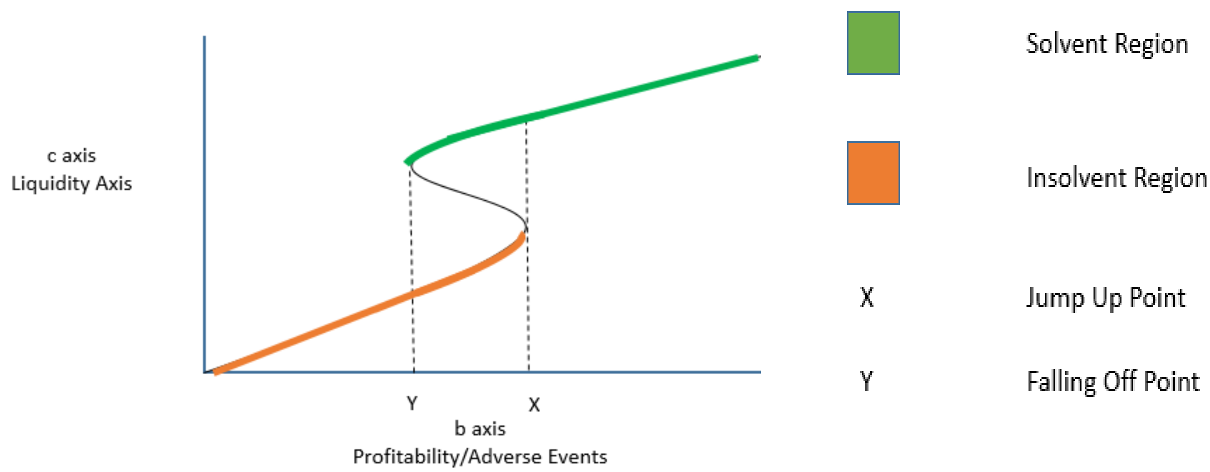


Figure 1

As a company experiences failure, they move from the solvent upper portion to the lower, bankrupt portion. Notice that to move back to the solvent zone, a company must move not only to the right, but must jump up once reaching the point x. The view of the cusp from the top, in figure 2, incorporates factors which drag a company into bankruptcy, and factors which mark the original strength of the company and its balance sheet. As the company is stronger, it must move further to the left to fall of the cusp. The moment a company falls off the cusp is not the point at which Chapter 11 occurs, but is the point at which the eventual occurrence of reorganization becomes inevitable[3].

[3] Rebholz, 2002

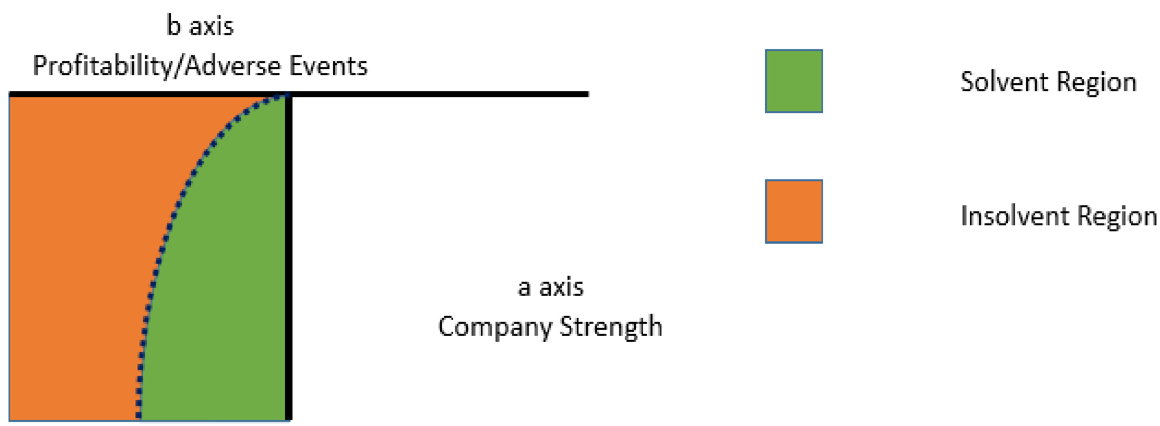


Figure 2

Notice that a company near the start of the curve with little financial strength would need very little to drive it into bankruptcy. This makes sense, as a company with an already maximally leveraged balance sheet is only one bad quarter away from insolvency; this is the equivalent of an employee who is living paycheck to paycheck. A company with a stronger balance sheet must experience either a longer period or much more intense distress to approach the cusp.

Logistic Regression Transformation

While the cusp model is dissimilar from a logistic regression model in that it is not probabilistic, consider the following transformation. Imagine the data from a cusp model is predicted as 0's and 1's, and is then run through a logistic regression model. In this case, we would expect a complete separation. While the logistic model would not be able to be created as maximum likelihood estimators could not be established, the graphical representation would look like the figure below:

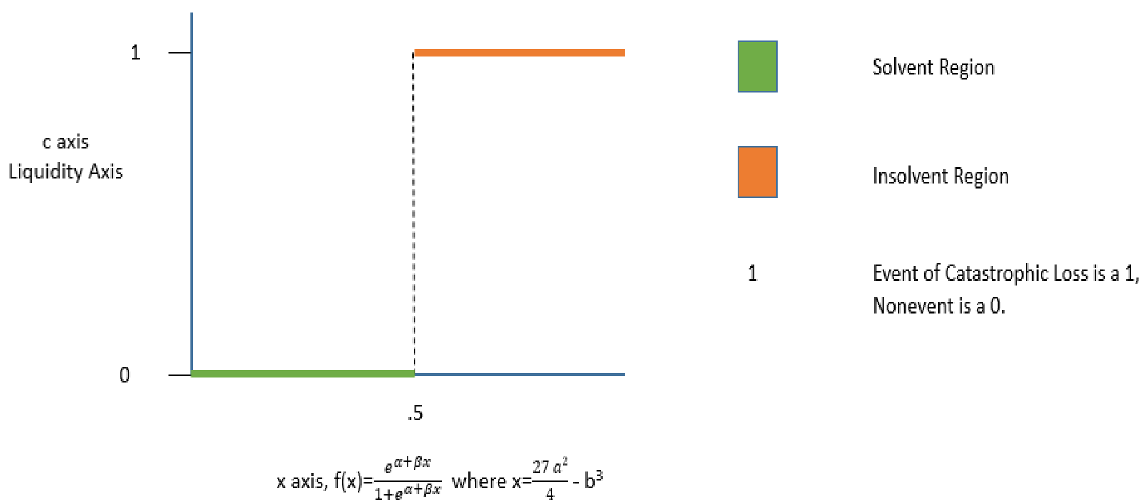


Figure 3

Note that this is a logistic regression model with perfect prediction. However, from the investor perspective, bankruptcy should perhaps not be viewed as a binary event, and two new approaches to the model are worth considering.

The first is how the bankruptcy event is defined. Companies commonly file for Chapter 11 bankruptcy or, with decreased frequency, Chapter 7 bankruptcy. Both events are negative for the stock holder, when compared to investing in a healthy company. However, there are more events that result in negative returns for an investor beyond filing for bankruptcy. Companies which announce they may need to file Chapter 11 but end up selling themselves to another company often experience significant value decline. Dilutive stock issuances to recapitalize are a negative for existing shareholders. A company being taken private during a period of financial struggle generally has a similar result- in each case, the investor of common stock will have lost much, if not all, of their capital. While bankruptcy is somewhat trackable, the other scenarios are related to bankruptcy but much more difficult to track. Instead of bankruptcy, consider a new variable, called catastrophic value loss, defined as when companies which lose at least 80% of their market value over the investment period. This value was chosen as it represents a significant loss the investor, but also results in a large enough sample size to have high risk stocks in every quarter. This will include more companies than bankruptcy, but is also a more applicable variable to the investor, as investors do not care why they lose capital, but that they lose capital.

The second consideration is that bankruptcy is a probabilistic event rather than a certainty. Healthy companies occasionally go bankrupt or lose the bulk of their value in a very short period of time due a variety of factors. Fraud, commodity price shocks, or world events each have their own unique effect on the stock market, business sectors, and individual companies. These events do not happen often, but the probability of a healthy company experiencing catastrophic value loss within a few years is nonzero. Similarly, not all companies that are expected to become insolvent do so. Consider the case of Sirius XM[4], which was a likely bankruptcy candidate. The company received a significant loan to boost liquidity from Liberty Communications, and in short order became a healthy and profitable company. Pier One Imports experienced a similar recovery[5], without any liquidity boost. These are rare events from

[4] Kang, 2009
[5] Schnurman, 2010

companies that are relatively insolvent, but recoveries do happen at a nonzero rate. Depending on the probabilities of healthy default and unhealthy recovery, the logistic regression model of catastrophic value loss could theoretically look like one of the graphs below.

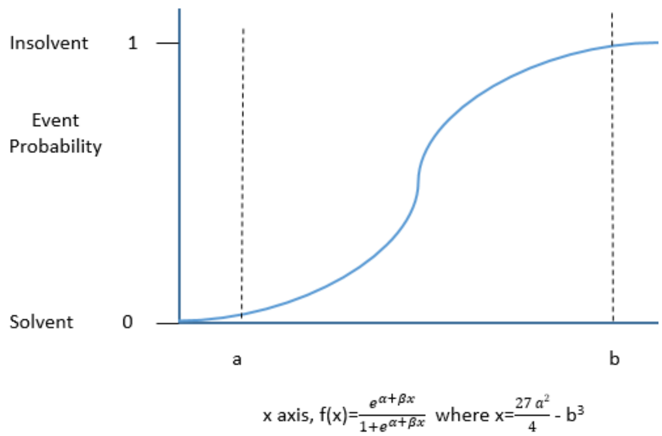


Figure 4a

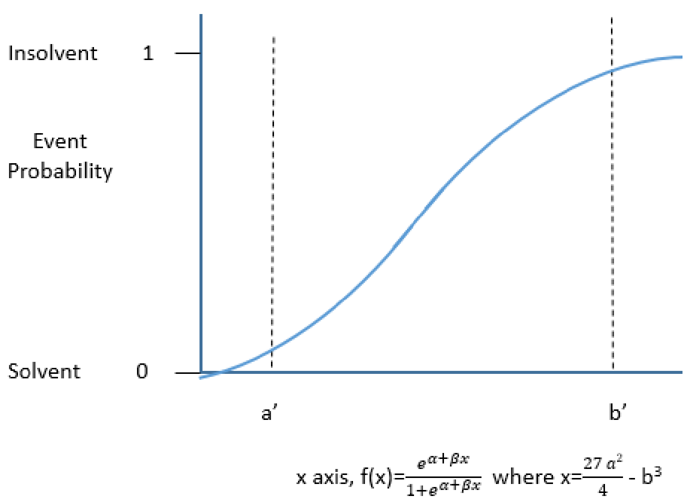


Figure 4b

The diagram in figure 4a indicates relatively low rates of default for healthy companies and high rates of default for insolvent companies. The second model in figure 4b has somewhat more random rates of default and recovery for healthy and insolvent companies. Note that at the end of each range of probabilities, at a and b in figure 4a, and a' and b' in figure 4b, the expected values are still close to zero and n, respectively, where n is the number of companies the high probability range. At the periphery of the model, regardless of the eventual logistic regression model chosen, the transformation to logistic regression will result in similar expected values to the original cusp model transformation, which anticipated perfect prediction. This is pictured in figure 5.

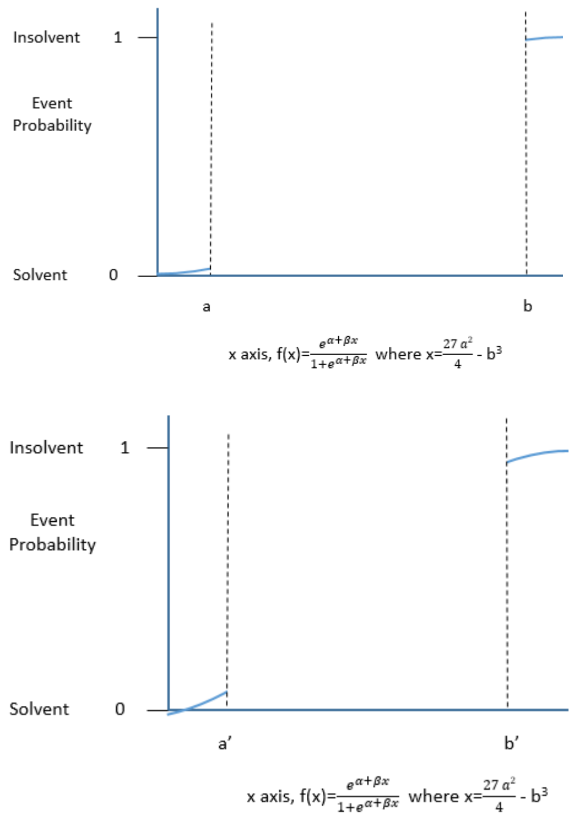


Figure 5

This paper will pursue three avenues of inquiry. The first is a validation of Rebholz's model. The model results of a sensitivity of 82% and specificity of 76% for a bankruptcy prediction three years in advance would be very investable. However, before investing, need to validate this across a larger time period. How do his original coefficients perform when applied to the larger sample of companies? The time period of 1990-present is not as homogeneous as his three-year period in the mid 90's, and includes two significant market crashes. A drop in accuracy should be expected, but it's unclear if the coefficients will need a slight adjustment, or will have a significant drop off in accuracy. It's also unclear how the transformation to a new variable, the 80% drop in market value, will impact the model, as this is likely to be largely influenced by market crashes. Note that in many previous models, bankrupt companies were matched one to one with companies of similar size and industry. This analysis makes no such accommodations, and a substantial drop in accuracy should be expected. However, this makes the results much more generally applicable.

The second is the impact of a new variable. The new variable will be the performance of a company's stock in the quarter after the original prediction is made. If this is statistically significant, the model can be more accurate in its prediction at 2.75 years (which should be expected). If it is not statistically significant, this suggests that an investor could reallocate a portfolio based on stock performance over the previous quarter to maximize profit based on any stock fluctuations. The variable is a similar idea to momentum- a stock which has been trending negatively is unlikely to trend upward without a material change in its prospects.

The third line of inquiry is how the model will perform when the future market conditions are included. Consider the variable quarter, which tracks the time period for a prediction. This variable will essentially measure the economic conditions of the time period by swinging the average default probability higher or lower in a given period. While any bankruptcy model would ideally be forward looking and would not have access to this information, this could be an incredibly useful model for forecasting portfolio performance and creating hedges. The investor would be able to create different portfolios of puts, based on possible future market performance. This should yield a more balanced, if not significantly higher, total return.

With these ideas in mind, the application of this model to a stock portfolio is almost as important as the statistics which describe the model. Beyond statistics such as AUC, we will consider the means and medians of two portfolios of stocks: a high risk group and a low risk group. Few stocks will have a high risk of bankruptcy; any number from a 10% probability of failure to a 50% probability of failure seems reasonable to classify as high risk. The overall failure rate is around 3.5%. A low/typical risk would be a number between 2 and 5%. From a logistic regression model, this idea is akin to taking the probabilities assigned in figure 4, cutting out the middle section, and creating high and low risk portfolios, as in figure 5. These densities will likely not be the same, as most companies are healthy. For this reason, high risk companies are defined as greater than 20% risk of catastrophic lost, and low risk portfolios were considered at both less than 5% and less than 2.5%.

Review of Literature:

Any review of bankruptcy literature will likely start with Altman's "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." This was one of the first pieces of research to gain broad popularity using rigorous statistical techniques to predict bankruptcies. Altman was similarly interested in the broad classification of bankrupt and non-bankrupt companies. Altman drew from a list of 22 possible variables to predict bankruptcy in firms between $1 million and $25 million in market cap. His final model used a linear combination of five variables (Working Capital/Total Assets, Retained Earnings/Total Assets, EBITA/Total Assets, Market Value Equity/Book Value of Debt, and Sales/Total Assets) to sort these companies into solvent and bankrupt groups, with accuracy one year out above 90% and accuracy for predictions two years in advance close to 80%. The Altman Z score is still in existence today, with some modifications, and was one of the first widely cited and used linear models to predict bankruptcy.

"A Review of Bankruptcy Prediction Studies: 1930 to Present" by Bellovary, Giacomino, and Akers provides an excellent synopsis of bankruptcy prediction methods up until the mid 2000's. The publication examines 165 bankruptcy models. Multiple Discriminant Analysis was the most popular method of analysis through the 70's, being overtaken by logistic regression in the 80's. From the 90's through today, neural networks have been increasing in popularity, as machine learning has become more prevalent and computing power has become cheaper.

Many studies with high accuracy have focused on a particular market sector or type of company. For example, El Hennawy and R.C. Morris were able to achieve an accuracy of 100% 5 years of advance in their study in "The Significance of Base Year In Developing Failure Prediction Models." However, their model dealt specifically with English construction, distribution, and manufacturing compabies of a particular size threshold. As models hone in on more specific industries and sectors, they tend to be more accurate. This is due partly because the population is more homogeneous, and partly because the companies are subject to more uniform exigent circumstances. For example, a stock commodity boom and bust cycle may have a large impact globally on all mining and shipping companies, a more muted effect on banks and construction companies, and little effect on consumer spending. In an individual country, that cycle may have a more universal effect (i.e. countries where mining is a larger percentage of

GDP exports/imports). The more homogeneous the population, the higher the expected accuracy of the model. Similarly, it's often easier for a model to distinguish between bankrupt and non-bankrupt firms that are matched side by side for size and industry.

Neural networks, which utilize machine learning, are increasingly growing in popularity, and have achieved strong results. However, Rebholz's model is the primary inspiration for the paper, and this paper will focus on the ideas inherent in the cusp catastrophe model for bankruptcy, and applications to logistic regression. The cusp model is similar to earlier ideas, such as from Dambolena and Shulman, which examines cash flow as a predictor of bankruptcy. Various measures of cash flows and liquidity measures have long been used as a measure of financial health. Fraud is also a somewhat predictable measure[6], based on changes in accounting, restatements, or missing financial disclosures. For this reason, only companies which reported results quarterly were used in the sample.

Methods:

Data was collected from the Compustat database at Duquesne's Palumbo School of Business. The data was originally coded in excel, and each variable had approximately 30,000 rows and 90 columns. The columns each represented a different quarter from 1990 to the present. For example, the first column encapsulates data from quarter 1 of 1990, the second column from quarter 2 of 1990, etc. Data was collected for cash flow, current liabilities, inventories, market value, net income, quick ratio, sales, and total liabilities. Market value was

[6] Dechow, 2011

collected from two files, as Compustat queries market value from active companies and the research set (companies no longer in existence) separately.

Several data codes were present in the Compustat data. The codes are as follows, with definitions from the Compustat User Guide[7]:

@AF: Annual Figure (data available at annual, but not quarterly, level)

@CF: Combined Figure (figure is combined in another item)

@IF: Insignificant Figure (the number is immaterial)

@NA: Not Available (company does not disclose information about the item, or was not in existence at that time point)

@NC: Not Calculable (rules for calculation were not met)

@NM: Not Meaningful (item is not meaningful for a company)

@SF: Semi-annual Figure (data available at semiannual, but not quarterly, level).

The most common data code was @NA, which indicated no data was available for the period. This error code was primarily used when a company was not yet in existence for a quarter. Other common error codes were @AF, and @SF (data available only as annual or semi-annual figure, respectively). Only data which was available at the individual quarterly level was used. This decision was made because investors would only invest in companies with information fully available in SEC filings. As noted in Dechow, financial restatements and other questionable financial accounting practices are associated with a higher risk of fraud, and

---

7 Healey, 1999

negative returns; this makes only using companies with full quarterly data a reasonable rule. This choice means the data will be consistent. Companies who do not have quarterly data available may have had accounting changes, non-standardized accounting practices, or technical errors, among other possible issues. A data point was created if data was available for all quarters in both the prediction period and every quarter until the results period. The results period was originally set to 3 years from the prediction period (Case 1 in Figure 6). In some cases, companies had less than 3 full years of data from the prediction point period to the end result- generally when a company underwent a merger, was taken private, went through Chapter 7 liquidation, or was delisted. These companies would ideally be included in the analysis, and still had full data available between the beginning of the research period and the period of removal. These instances were captured by iterating through cases where only the first 2 years and 3 quarters were available (Case 2 in figure 6), then 2 years and two quarters (Case 3, figure 6), etc., until 1 year of data was available. 1 year is generally pushing the boundary of when it is still unknown that a company is likely to merge or go bankrupt, and seemed an appropriate cut off point. Companies with less than 3 full years of data after the prediction point represented less than 5 percent of all data points.

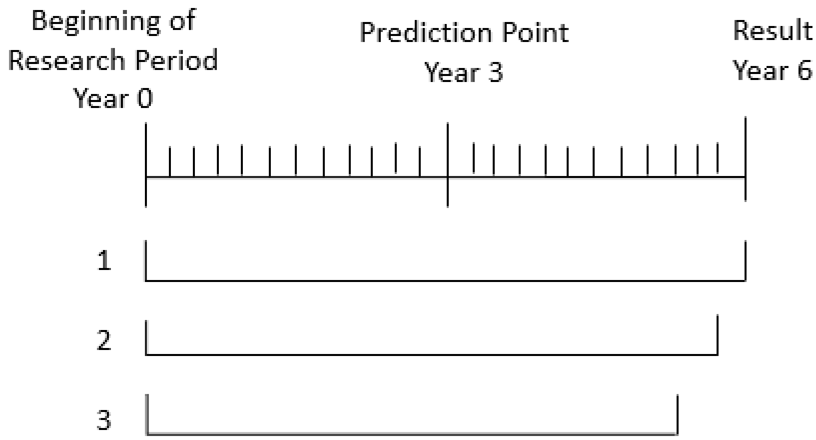


Figure 6

Additional filters were applied for consistency among companies. All of the penny and microcap stocks were removed, using both a stock price and market cap filter. The stock price was set at a minimum of 50 cents for every quarter in the three year prediction period. The market cap was set at 45 million for the first year of the prediction period, and 35 million for the next two years. The market cap filter was set at these prices instead of a constant 50 million, which is used today to delineate microcap stocks, to avoid excluding too many companies from the early 1990's. Even so, the sample has more companies from later in the time period than earlier. This makes sense, as the economy has grown substantially over time, and more businesses are publicly traded today. The stock price was set at 50 cents instead of 1 dollar for a similar reason. No limitations were set during the results period. This filter also removed many companies which underwent Chapter 11 reorganization during the prediction period.

The Compustat database recorded companies which underwent Chapter 7 liquidation, but did not record data on companies which underwent Chapter 11 reorganization. Many of these companies were eliminated using the filters for minimum stock price and market cap, but not all were, especially in cases where the reorganization was very quick and occurred during the results period. The researcher noticed that, in most cases, these companies either had unusually high calculated returns, or large drops in market cap during the prediction period. The researcher checked 1300 entries representing 600 companies by manually searching for journalistic reports detailing bankruptcies, and found 217 such companies[8]. These were manually coded as "excluded" in the dataset where necessary. This is because the prediction is for companies which appear relatively healthy; companies which are currently undergoing chapter 11 reorganization are not healthy, and choosing to invest in a company currently reorganizing requires a different analysis than is done here. Most of the bankrupt companies experienced a market cap decline of over 95% in the prediction period; a cutoff of 90% was used. There are likely still a few companies which underwent Chapter 11 in the data, but this is small number, based on additional random sampling of companies in the dataset.

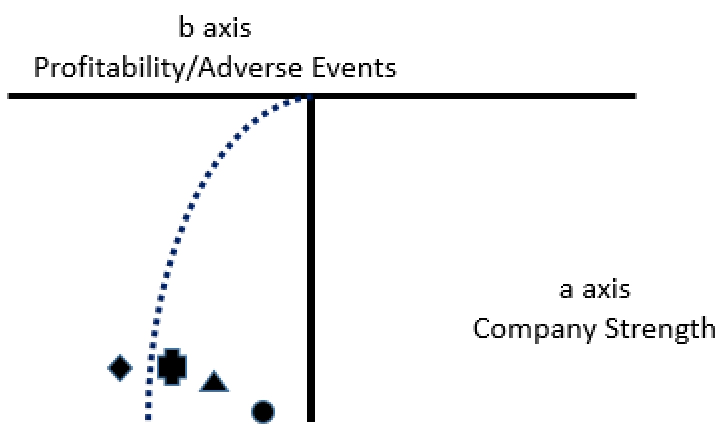


Figure 7

[8] *The New York Times* was used several hundred times to verify chapter 11 filings.

Independence:

A brief note on independence, and the cusp model: consider Figure 7 above, representing four different time points for a given company. It's likely that, in a particular amount of time, a company starting at the point represented by the circle would move no further than the triangle. It's also likely that a company starting at the point represented by the triangle would move no further than the cross in the negative direction or the circle in the positive direction. If viewed from the perspective of a time series model, it would be a problem that the data would very rarely, if ever, move directly from the circle to the diamond or diamond to the circle, and would instead pass through the intermediate points. However, when viewed through the cusp catastrophe model, what matters is the translated location of the company's liquidity and financial data to the plane. Every point to the right of the cusp, represented in figure 7 by the dashed line, will be predicted as solvent, while every point to the left will be predicted as having catastrophic loss. The fact that these points represent the same companies at various points in time and various points of liquidity is not a violation of the fundamental workings of the cusp prediction; the purpose of each point is to generate a binary prediction. The company and time point do not factor into the prediction or location on the plane, therefore they do not factor into the prediction of solvency or catastrophic loss, and thus independence is not violated in the cusp model.

Results and Statistical Methods:

The first model was computed using the exact coefficients from Rebholz's model, where a 1 was coded if the model was in the insolvent region of the cusp, and a 0 if the data was in the solvent region. It suffered from over prediction, with a sensitivity of approximately 25% and specificity around 70%.

		Actual	
		0	1
Predicted	0	112342	5700
	1	30398	1756

Figure 8

Because of the difference in the populations of Leo's original research and 1990-present, due to both increasing the time period and changing from bankruptcy to catastrophic loss, it's possible drop in accuracy is because the coefficients need to be recalculated, rather than any fundamental flaw in the model. A project beyond the scope of this paper is likely finding a way to approximate these values in an ad hoc model. For computational ease, a logistic regression model was used.

When converting the Rebholz model to a logistic regression model, two methods were used. The first stayed closer to the original model. Mathematically, the curve can be rewritten as follows:

From Rebholz model, predict bankruptcy when $b > \sqrt[3]{\frac{27a^2}{4}}$

This is equal to predicting when $0 > \frac{27\ a2}{4}$ - b^3

Where $b = k_0b_0 + k_1b_1 + k_2b_2 + k_3b_3 + k_4b_4 + k_5b_5 + k_6b_6$

a= min(a',0)

$a' = n_0a_0 + n_1a_1 + n_2a_2 + n_3a_3 + n_4a_4$

Where k and n are coefficients for the original liquidity variables a and b.

The a^2 and minimum of a' and zero issues combined are problematic for convergence. Two models were run- one with the cubic polynomial expansion of the b terms and first degree of the a terms, and the second with the first degree of the a and b terms. The b^1 model had an AUC of only .726, compared to an AUC of .73 for the b^3 model. Neither model is a particularly good fit for the data. Two portfolios of high risk and low risk, according to catastrophic loss probabilities of greater than 20% and less than 5%, respectively, were compared. The means were not statistically significantly different, as noted through the 95% confidence intervals (see appendix). When a 90% cutoff was used instead of an 80%, AUC improved to .76 for the b^1 model and .763 for the b^3 model. However, the portfolio comparison was still not significantly different. Note that it should be expected that AUC will increase for the 90% cutoff as opposed to the 80% cutoff, as these should, in aggregate, be a higher risk group of stocks, and thus a bit easier to predict suffering catastrophic loss.

Given how promising the Gaur/Rebholz model was originally, the model is still worth exploring. There are currently a few issues which may be addressed. The primary issue is that, in switching to catastrophic loss, much of the time the companies experience the event in large clusters. Because bankruptcy was reformatted as catastrophic value loss, the rate of when these events happened was very cyclical, and incorporated a high prevalence of events from the dotcom and financial crashes.

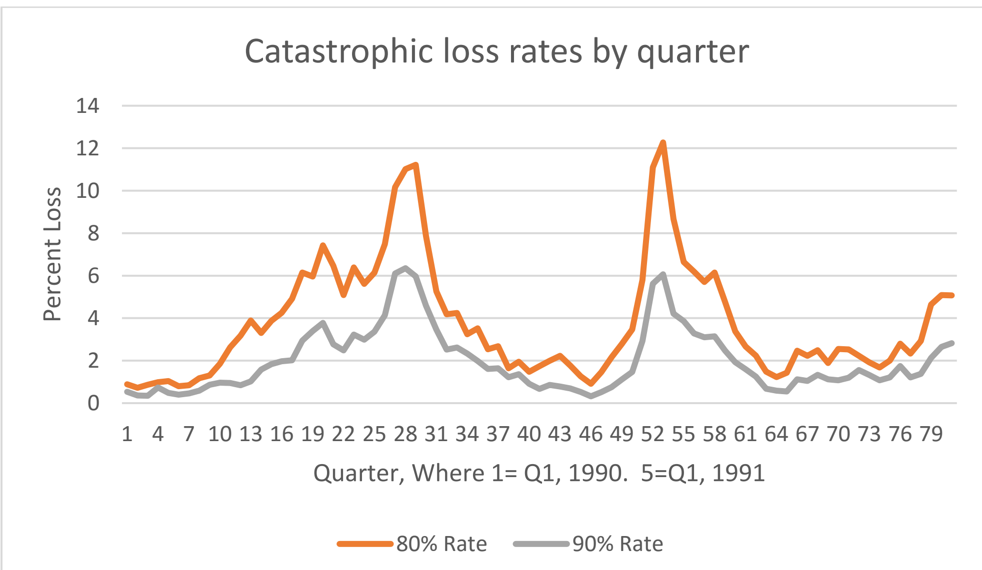


Figure 9

This means that the earlier models not only had to incorporate which companies were likely to collapse, but would also need to incorporate data on overall market performance. This is something which can be argued that no one is capable of doing; consider that no hedge fund has beaten an S & P index fund over a *ten year* period[9]. A reasonable inference is that if a stock

[9] Rekenthaler, 2017

manager knew when the market would crash, they would exit their holdings or short the market, then re-enter once the crash had occurred. Because this hasn't happened, it's reasonable to infer that this skill does not currently exist, for predicting market performance as a whole. By incorporating the variable quarter, we remove this strain on the model. Essentially, instead of asking the model both to predict how many catastrophic value losses will occur in a quarter AND which companies will be in that group, we instead allow the model to know how many value losses will occur, and to only assign the companies risk. The variable quarter is a time variable which incorporates the failure rate for a given quarter.

With the inclusion of quarter, the model is essentially a forward looking forecast tool, in answer to the question of how a portfolio would perform under various expected catastrophic loss rates in the future. The 80% model had an AUC of .789, and the holdout (50% in all cases) had an AUC of .786. The 90% model had an AUC of .816, and a holdout of .814. The stock portfolios were also noticeably different. We'd hope to see decreasing rates of returns, with the highest returns in the low risk group, moderate returns in the middle group, and negligible to negative returns in the high risk group. This would demonstrate a risk that was not being correctly priced by the market. In both the 80% and 90% portfolios, this was achieved. It should be noted that the high risk portfolio in the 90% group is smaller than that of the 80% group, as the variance did not decrease by enough to account for the smaller sample size. This resulted in a larger confidence interval for the mean performance. In the 80% portfolio, using the holdout sample, the low risk portfolio had a return of 46.7%, with a 95% CI of 45.8 to 47.5. The median return was 23.75%, with an IQR of -13.46 to 73.64, with 58655 observations. The catastrophic loss rate was 1.88%. The high risk portfolio had a mean return of 7.35%, with a 95% CI of .41 to 14.29. The median rate of return was -39.97%, with an interquartile range (IQR) of -79.6 to

34.28. The catastrophic failure rate was 24.685%. For charts, see appendix. While the results are stronger, the model is not perfect, as some convergence issues exist.

The variable quarter is a definite improvement. However, it's worth considering another variable which is not as restrictive. Consider a variable named a5. This variable measured the performance of the stock in the quarter after the original three-year data period. (This does shorten the prediction period from 3 years to 2.75 years, and the binary catastrophic loss variable was recoded to reflect this change). The a5 variable can be thought of in the financial sense as momentum- with the changes in the inventory, profitability, etc., is the market optimistic about a stock and driving the price upwards, or continuing to send the price lower? A price up would indicate that there is information beyond what is present in the current metrics and reason to be optimistic. A similar variable, coded as a5v2, looked at the stock performance in the final quarter of the three-year research period. This analyzed the question of if all two-year price change in the stocks are created equal. Theoretically (by the efficient market hypothesis[10]), the price should be all inclusive, regardless of the path it took over the two years, but a stock trending up could indicate a shift in prospects and market sentiment. The a5 variable similarly incorporates not only the prospects of the individual stock, but also the market as a whole. A period of predicted future market turbulence would pressure the price downward and market exuberance would pressure the price upward. It doesn't capture cyclical loss like quarter, but can be conceived of as a guess at both overall market sentiment, and sentiment for a particular stock. Again, a holdout sample of 50% was considered. The b^1 model with the addition of a5 had an AUC of .731, and the holdout had an AUC of .736. While the AUC is not particularly

---

[10] Sewell, 2011

promising, both a5 and a5v2 appear very likely to be statistically significant, and this suggests that they may potentially be variables to include in the model with quarter.

The final model included both the momentum variables (a5) and cyclical market performance adjustment (quarter). Using 80% as the cutoff for catastrophic loss, the original model had an AUC of .803, and an AUC of .792 in the holdout sample. This is a slight improvement from the model with only quarter. This model is also a better identifier of the high risk portfolio. Using the holdout sample for portfolio analysis, the low risk group had a mean return of 41.97%, with a 95% CI of 41.2 to 42.73. The median return was 22.1%, with an IQR of -14.08 to 69.15. The catastrophic loss rate was 1.825%. The high risk portfolio had a mean return of .45%, with a 95% CI of -5.42 to 6.31. The median return was -36.8, and had an IQR of -86.1 to 37.96. The catastrophic failure rate was 29.095%. With this new model, it's worth considering that perhaps the 5% is too generous a term for low risk, when the overall failure rate is around 3%. If we instead use a threshold for the low risk portfolio, identified in the appendix as modified low risk, of below 2.5% probability of catastrophic loss, the mean return is 47.55%, with a 95% CI of 46.7 to 48.41. The IQR is -5.87 to 74.5, and the catastrophic loss rate is 1.148%. The sample size drops from approximately 59,000 to 43,000. A graphical representation of the distribution of returns and rates of catastrophic loss is pictured in figures 10 and 11 below. Figure ten represents the modified low risk portfolio, and figure 11 represents the high risk portfolio. Note that no convergence issues exist with both quarter and a5 in the final model. In the appendix, a modified high risk portfolio with a risk cutoff of 15% was compared. While the sample size increased, the difference in mean returns decreased as well.

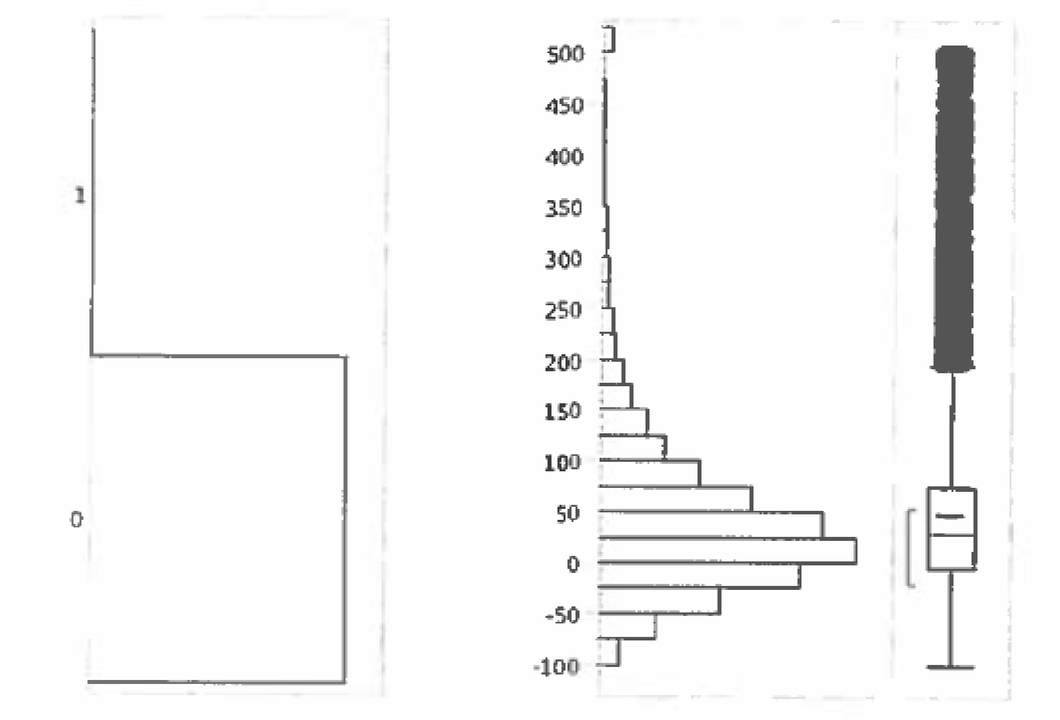


Figure 10

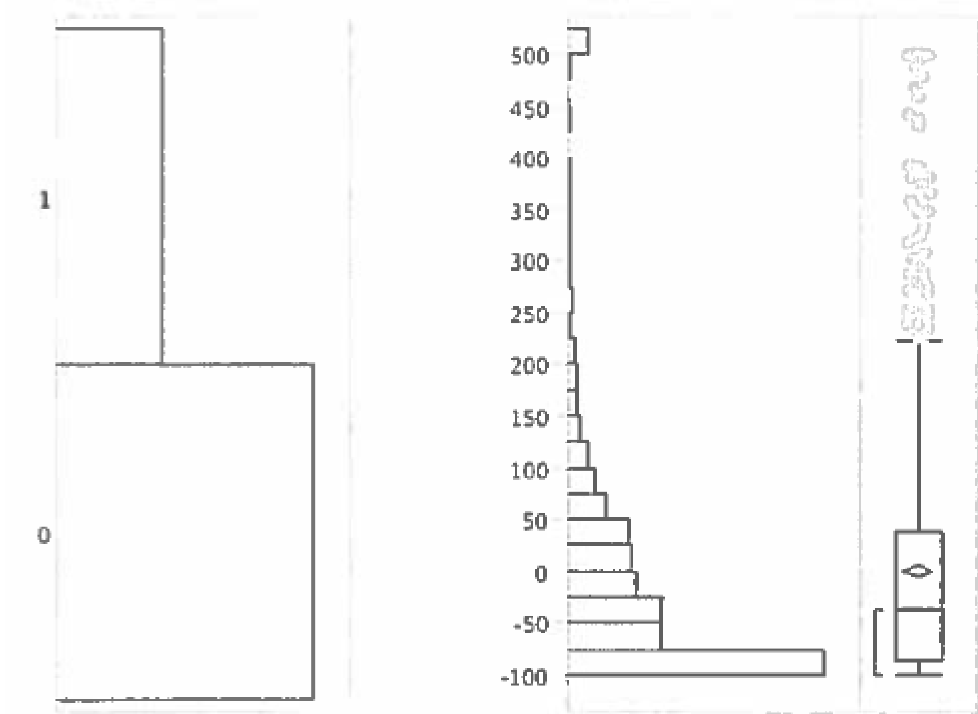


Figure 11

For validity, a chi square test was conducted on the analysis with both the a5 and quarterly variables. For the model as a whole, the chi squared was .867. The chi squared test was also conducted on the low and high risk portfolios themselves, with the high risk portfolio having a chi square of 98.5 and the low risk portfolio having a chi squared of 71.2. With 92 degrees of freedom, neither measurement is significant at even the α=.10 level. However, the 98.5 chi squared value for the high risk portfolio is higher than would be ideal.

Reflection

While catastrophic value loss is a strong idea for capturing the effect of market gains and losses on the individual investor, the market tends to experience the bulk of these events in cyclical patterns. These patterns may be tied to particular industries, or happen to the market at large. While some indicators of the total market risk exist, such as Case Shiller, the exact timing of a large market reversal is difficult to impossible to predict multiple years in advance with current analytics. This is a tradeoff the model has made to be more applicable to the investor.

There are several areas of further exploration. First, while the Rebholz model suffered from computation difficulties in recalculating the parameters of a_0 through a_4 and b_0 through b_6, the coefficients that made the 'a' and 'b' values in the cusp equation, an ad hoc model may be able to converge to a solution. Consider a cycle which parameterized variables, removed any values in which a was positive and replaced them with zero, recalculated, and looped through until a convergence was found. From the portfolio perspective, the chi squared results indicated that number of catastrophic loss expected to happened and which actually happened in each quarter varied due to random chance. This suggests that a strategy of purchasing puts would have a fairly regular success rate. It is possible to calculate the theoretical values for the puts necessary to receive adequate returns, but in most cases the markets have not yet been created, and would likely need to be done privately through an investment bank. It's much more difficult to analyze the expected returns of each quarter, how they compare quarterly, and how they compare for all yearly or three year periods, due to sample size issues. While difference in the means of the portfolios are highly significant, the chi squared test analyzed the expected values of the number of stocks experiencing catastrophic value loss, not the expected portfolio return value. While the median and IQR stock returns are noticeably different, more research is needed to turn shorting the high risk portfolio into a stock trading strategy. From the trading perspective, further research should also be done into the overlap and change in risk for companies under various expectations of market failure, i.e. the odds ratios relative to quarter.

Limitations:

This paper uses logistic regression in a different method than normally used. Given that logistic regression is a robust methodology, it is still valid even though the mean rate of catastrophic loss was around 3.5%. Inflated zeroes would also be a worthwhile investigative tool, as it does well in circumstances with low rates of incidence.

Bibliography

Altman, Edward I. "Exploring the road to bankruptcy." *Journal of Business Strategy* 4, no. 2 (1983): 36-41.

Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The Journal of Finance* 23, no. 4 (1968): 589-609.

Altman, Edward I. "A Further Empirical Investigation of the Bankruptcy Cost Equation." *Journal of Finance* 39 (September 1984): 1067-1089.

Back, Barbro, Teija Laitinen, Jukkapekka Hekanaho, and Kaisa Sere. "The effect of sample size on different failure prediction methods." *Turku Centre for Computer Science Technical Report* 155 (1997): 1-23.

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. "A review of bankruptcy prediction studies: 1930 to present." *Journal of Financial Education* 33 (Winter 2007): 1-42.

Beneish, Messod D. "The detection of earnings manipulation." *Financial Analysts Journal* (1999): 24-36.

Carty, Lea V. "Corporate credit-risk dynamics." *Financial Analysts Journal* (2000): 67-81.

Dambolena, I. G., & Shulman, J. M. "A primary rule for detecting bankruptcy: Watch the cash." *Financial Analysts Journal* 44, no. 5 (September/October 1988): 74-78.

Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan. "Predicting material accounting misstatements." *Contemporary Accounting Research* 28, no. 1 (2011): 17-82.

Francis, Jack Clark, Harold M. Hastings, and Frank J. Fabozzi. "Bankruptcy as a mathematical catastrophe." *Research in Finance* 4 (1983): 63-89.

Fulmer, John G., James Moon, Thomas Gavin, Michael Erwin. "A Bankruptcy Classification Model for Small Firms." *Journal of Commercial Bank Lending* (July 1984): 25-37.

Gilson, Stuart C., Edith S. Hotchkiss, and Richard S. Ruback. "Valuation of Bankrupt Firms." *Association for Investment Management and Research* (2000): 43-74.

Harrell, Frank. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

Healey, Ryan, *Research Insight (Compustat) User's Guide*. New Brunswick, NJ: Rutgers University Library, 2009. healey.library.wikispaces.com/file/view/ Compustat+Rutgers+Guide.pdf

Hennawy, R. H. A., & Morris, R. C. "The significance of base year in developing failure prediction models". *Journal of Business Finance & Accounting* 10, no. 2 (1983): 209-223.

Kahya, Emel and Theodossiuo, Panayiotis. "Predicting Corporate Financial Distress: A Time-Series CUSUM Methodology." *Review of Quantitative Finance and Accounting* 13 (1999): 323-345.

Kang, Cecilia. "Liberty Extends $530 Million Loan to Bail Out Sirius XM." *The Washington Post,* February 18, 2009.

Kao, Duen-Li. "Estimating and pricing credit risk: An overview." *Financial Analysts Journal* (2000): 50-66.

Rebholz, Leo. "Bankruptcy as Cusp Catastrophe." Master's Thesis, Duquesne University, 2002

Rekenthaler, John. "Buffett Brings Out the Victory Cigar." Rekenthaler Report, February 28, 2017

Scapens, Robert W., Robert J. Ryan, and Leslie Fletcher. "Explaining corporate failure: a catastrophe theory approach." *Journal of Business Finance & Accounting* 8, no. 1 (1981): 1-26.

Schnurman, Mitchell. "How Pier 1 Imports Made a Turnaround." *D Magazine*, September 2010.

Settles, Burr. *Active learning literature survey.* Madison, WI: University of Wisconsin, Madison, 2010. Accessed January 10, 2017. http://burrsettles.com/pub/settles.activelearning.pdf

Sewell, Martin. "History of the efficient market hypothesis." *RN* 11, no. 04 (2011): 04.

Shumway, Tyler. "Forecasting Bankruptcy More Accurately: A Simple Hazard Model." *Journal of Business* 74, no. 1 (January 2001): 101-124.

Thom, Rene. *Structural Stability and Morphogenesis*. Reading, MA: WA Benjamin, 1975.

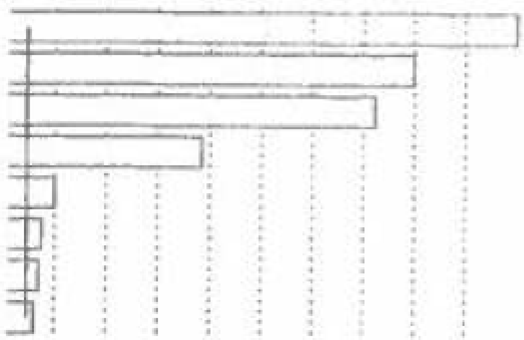
Appendix: Model for b[1], 80% threshold

JMP_finaldata_withexclusions - Fit Nominal Logistic

Nominal Logistic Fit for predictorA

Effect Summary

Source	LogWorth		PValue
b0	556.775		0.00000
b4	40.108		0.00000
b3	36.387		0.00000
b6	19.226		0.00000
a0	4.971		0.00001
a2	3.645		0.00023
a1	3.306		0.00049
capped_a4	2.996		0.00101



Converged in Gradient, 6 iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1902.567	8	3805.133	[illegible]
Full	22266.940			
Reduced	24169.507			

RSquare (U)	0.0787
AICc	44551.9
BIC	44641.1
Observations (or Sum Wgts)	148440

Measure	Training	Definition
Entropy RSquare	0.0787	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0911	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.1500	∑ -Log(ρ[j])/n
RMSE	0.1891	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0713	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0384	∑ (ρ[j]≠ρMax)/n
N	148440	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	148371	22266.940	44533.88
Saturated	148379	0.000	**Prob>ChiSq**
Fitted	8	22266.940	1.0000

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	4.06508969	0.023416	30138	[illegible]
b0	-0.5901994	0.0113434	2707.1	[illegible]
b3	-0.0035869	0.0002723	173.49	[illegible]
b4	-0.0039298	0.0002821	194.08	[illegible]
b6	0.00044392	4.8033e-5	85.41	[illegible]
a0	-0.0008337	0.0001844	20.45	[illegible]
a1	-0.000718	0.0002015	12.70	[illegible]
a2	-0.0016338	0.0004415	13.70	[illegible]
capped_a4	-0.0004426	0.0001338	10.94	[illegible]

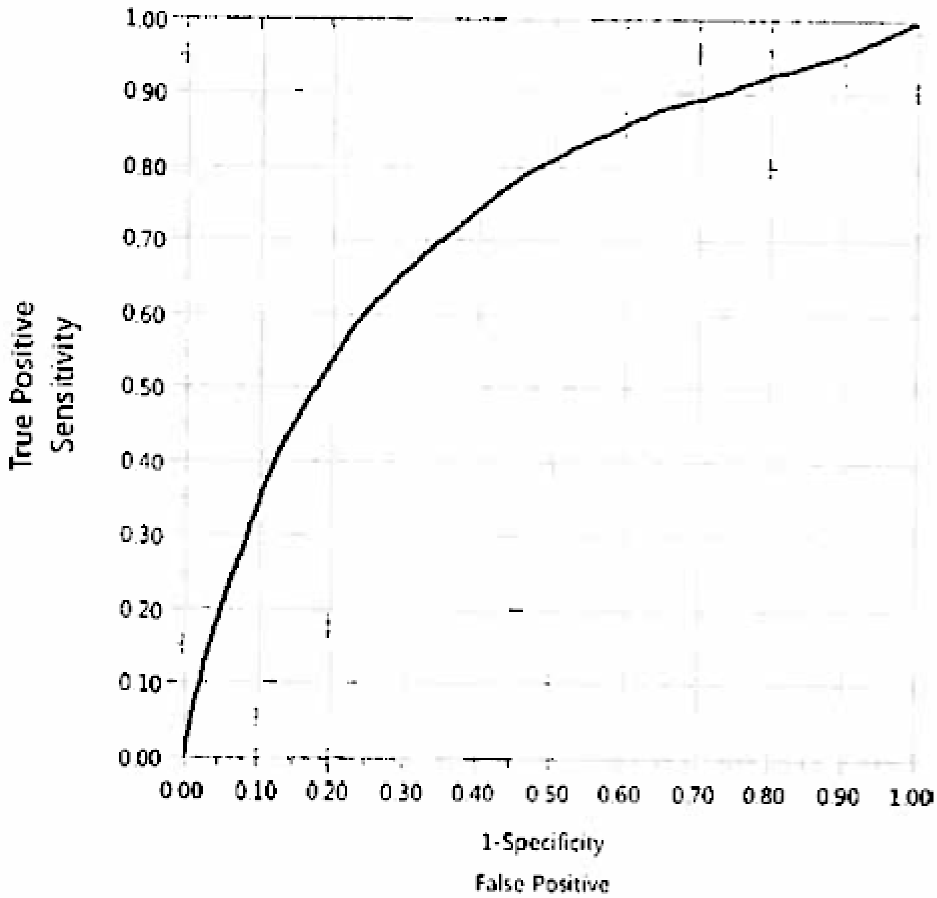
For log odds of 0/1

Nominal Logistic Fit for predictorA

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
b0	1	1	2555.74517	[illegible]
b3	1	1	162.01571	[illegible]
b4	1	1	179.054434	[illegible]
b6	1	1	83.6389304	[illegible]
a0	1	1	19.3855209	[illegible]
a1	1	1	12.1354043	[illegible]
a2	1	1	13.595853	[illegible]
capped_a4	1	1	10.8124601	[illegible]

Receiver Operating Characteristic



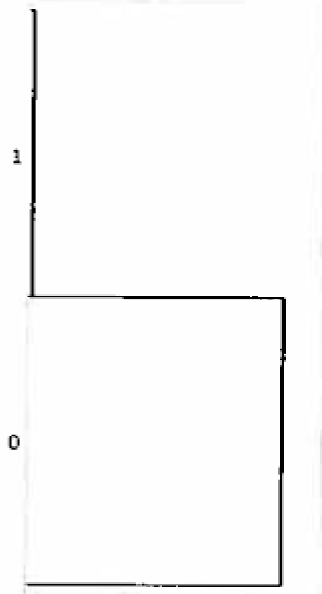
Using predictorA='1' to be the positive level

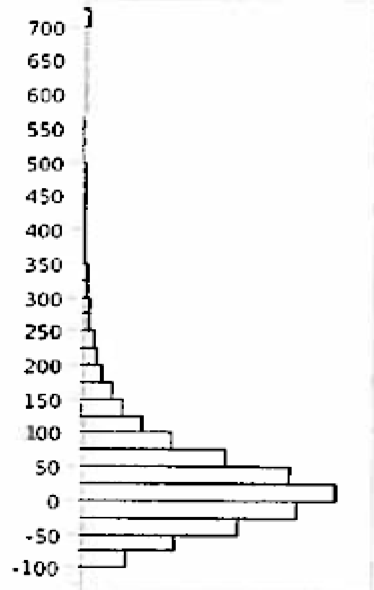
AUC

0.72624

b[1] portfolio comparisons

Low Risk portfolio:





Frequencies

Level	Count	Prob
0	115935	0 97672
1	2763	0.02328
Total	118698	1.00000

N Missing 86

2 Levels

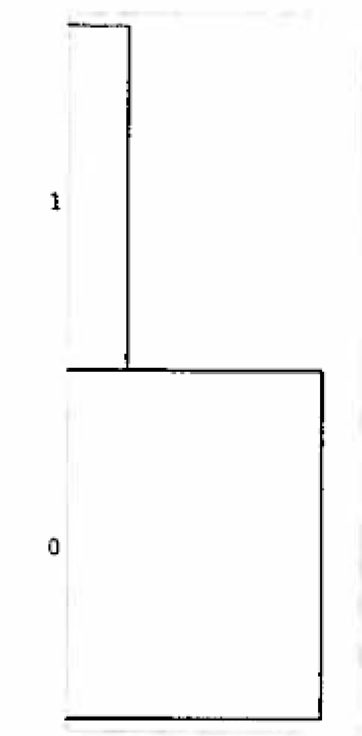
Quantiles

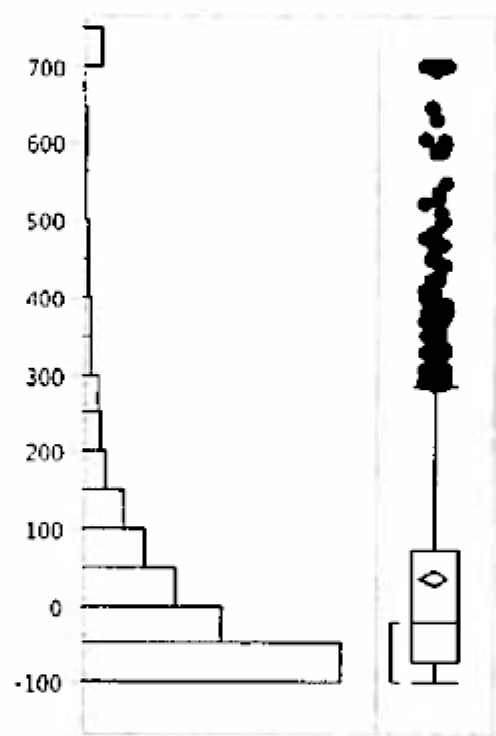
100.0%	maximum	700
99 5%		612 26389552301
97 5%		299.6459865275
90 0%		140 33844192
75 0%	quartile	67 5612557275
50 0%	median	20 521103135
25.0%	quartile	-16 81325767
10.0%		-49 065703719
2 5%		-79.01498298825
0 5%		-95 64500710485
0 0%	minimum	-99.99446971

Summary Statistics

Mean	40.003931
Std Dev	98.7416
Std Err Mean	0.2866015
Upper 95% Mean	40 565665
Lower 95% Mean	39.442196
N	118698

High Risk Portfolio





Frequencies

Level	Count	Prob
0	1002	0.80611
1	241	0.19389
Total	1243	1.00000

N Missing 12

2 Levels

Quantiles

100.0%	maximum	700
99.5%		700
97.5%		689.59646915999
90.0%		237.0052541
75.0%	quartile	71.31157969
50.0%	median	-21.22323953
25.0%	quartile	-71.31046086
10.0%		-94.372473042
2.5%		-99.649799746
0.5%		-99.9510048868
0.0%	minimum	-99.97349004

Summary Statistics

Mean	36.428316
Std Dev	168.24118
Std Err Mean	4.7719594
Upper 95% Mean	45.790308
Lower 95% Mean	27.066324
N	1243

Model for b^3

Effect Summary

Source	LogWorth	PValue	
b0	100.958	0.00000	
b3	21.975	0.00000	
b4	16.068	0.00000	
b12	6.239	0.00000	
b4*b4*b4	5.563	0.00000	
b12*b12*b12	5.422	0.00000	
capped_a4	5.053	0.00001	
a0	5.024	0.00001	
b6	4.084	0.00008	
b4*b4	3.925	0.00012	^
a1	2.920	0.00120	
b3*b3*b3	2.893	0.00128	
b6*b6	2.426	0.00375	
b6*b6*b6	2.049	0.00894	
b5*b5	1.854	0.01400	
b12*b12	1.784	0.01646	^
a2	1.761	0.01733	
b0*b0	1.364	0.04328	
b3*b3	0.945	0.11358	^
b0*b0*b0	0.677	0.21033	
b5*b5*b5	0.229	0.59048	
a3	0.045	0.90223	
b5	0.024	0.94543	^

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1971.442	23	3942.884	[illegible]
Full	22198.065			
Reduced	24169.507			

RSquare (U)	0.0816
AICc	44444.1
BIC	44681.9
Observations (or Sum Wgts)	148440

Measure	Training	Definition
Entropy RSquare	0.0816	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0943	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.1495	∑ -Log(ρ[j])/n
RMSE	0.1889	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0713	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0384	∑ (ρ[j]≠ρMax)/n
N	148440	n

Lack Of Fit

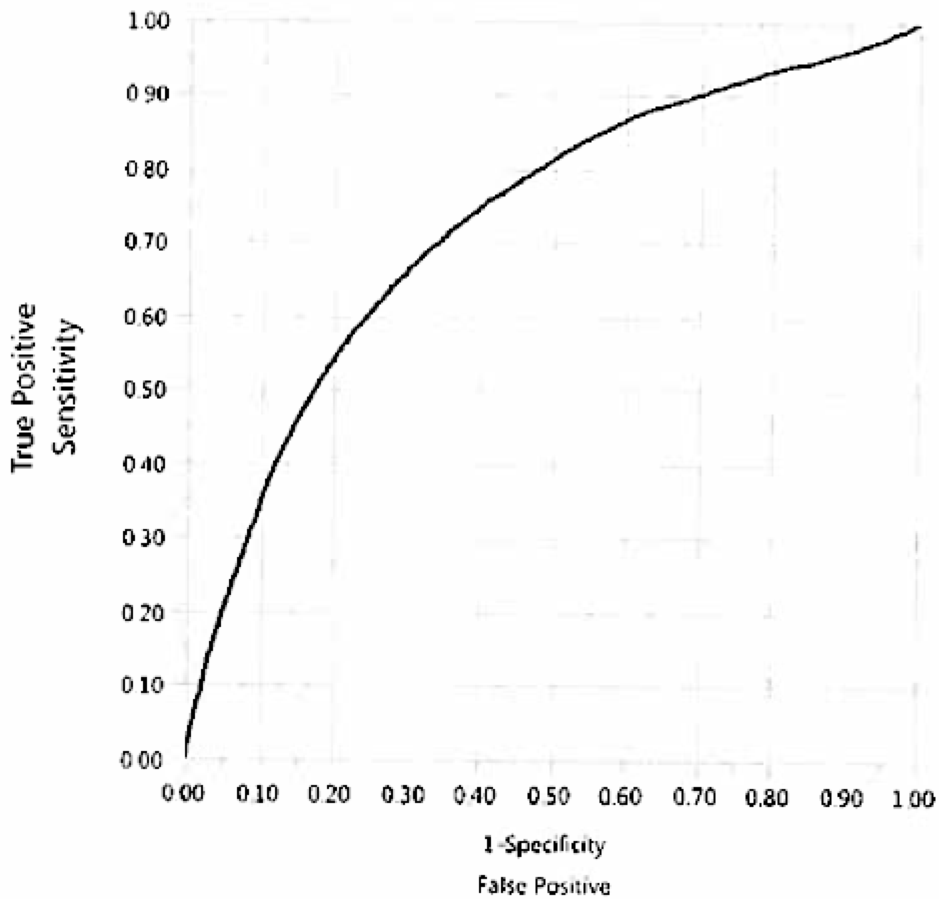
Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	148356	22198.065	44396.13
Saturated	148379	0.000	**Prob>ChiSq**
Fitted	23	22198.065	1.0000

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	4.12072969	0.0418415	9699.2	<.0001*
b0	-0.6411301	0.0299468	458.34	<.0001*
(b0-0.6876)*(b0-0.6876)	0.12271681	0.0607215	4.08	0.0433*
(b0-0.6876)*(b0-0.6876)*(b0-0.6876)	-0.0303104	0.024197	1.57	0.2103
b12	0.00074473	0.000149	24.99	<.0001*
(b12+103.938)*(b12+103.938)	1.91803e-7	7.9964e-8	5.75	0.0165*
(b12+103.938)*(b12+103.938)*(b12+103.938)	-7.33e-10	1.586e-10	21.37	<.0001*
b3	-0.0052002	0.0005303	96.16	<.0001*
(b3-18.4214)*(b3-18.4214)	-1.5374e-5	9.7163e-6	2.50	0.1136
(b3-18.4214)*(b3-18.4214)*(b3-18.4214)	1.70126e-7	5.2829e-8	10.37	0.0013*
b4	-0.0045697	0.000549	69.28	<.0001*
(b4-16.7297)*(b4-16.7297)	-3.7459e-5	9.7337e-6	14.81	0.0001*
(b4-16.7297)*(b4-16.7297)*(b4-16.7297)	2.48872e-7	5.3066e-8	21.99	<.0001*
b5	8.6246e-6	0.000126	0.00	0.9454
(b5-16.2918)*(b5-16.2918)	-1.8626e-7	7.5798e-8	6.04	0.0140*
(b5-16.2918)*(b5-16.2918)*(b5-16.2918)	-7.612e-11	1.415e-10	0.29	0.5905
b6	0.0005126	0.0001302	15.50	<.0001*
(b6-13.2156)*(b6-13.2156)	-1.9195e-7	6.6221e-8	8.40	0.0037*
(b6-13.2156)*(b6-13.2156)*(b6-13.2156)	-3.659e-10	1.399e-10	6.84	0.0089*
a0	-0.0008347	0.0001885	19.62	<.0001*
a1	-0.0006652	0.0002054	10.49	0.0012*
a2	-0.0010887	0.0004575	5.66	0.0173*
a3	-5.829e-5	0.0004745	0.02	0.9022
capped_a4	-0.0006127	0.0001379	19.74	<.0001*

For log odds of 0/1

Receiver Operating Characteristic



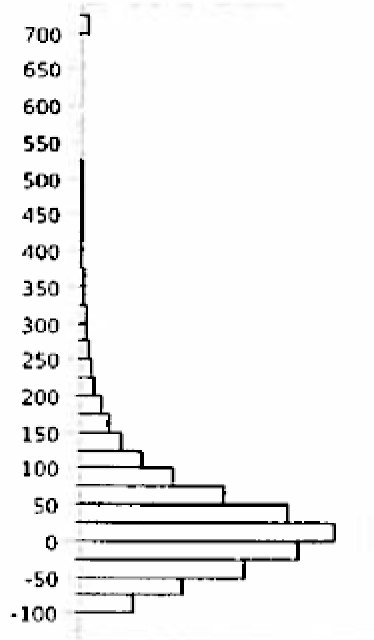
Using predictorA='1' to be the positive level

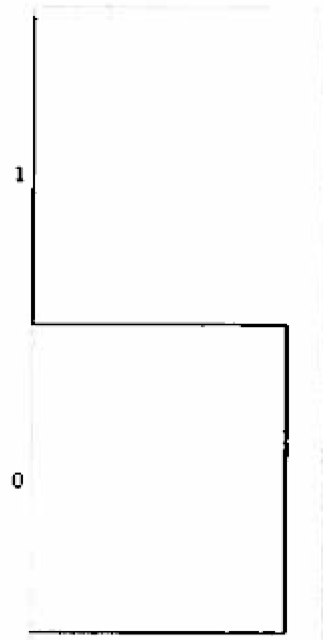
AUC

0.73028

b^3 portfolio comparisons

Low Risk portfolio:





100.0%	maximum	700
99.5%		683.54613618
97.5%		320.918433085
90.0%		146.59293994
75.0%	quartile	68.923922685
50.0%	median	19.95853965
25.0%	quartile	-18.724479745
10.0%		-52.122711244
2.5%		-82.448768084
0.5%		-97.248774632
0.0%	minimum	-99.99446971

Summary Statistics

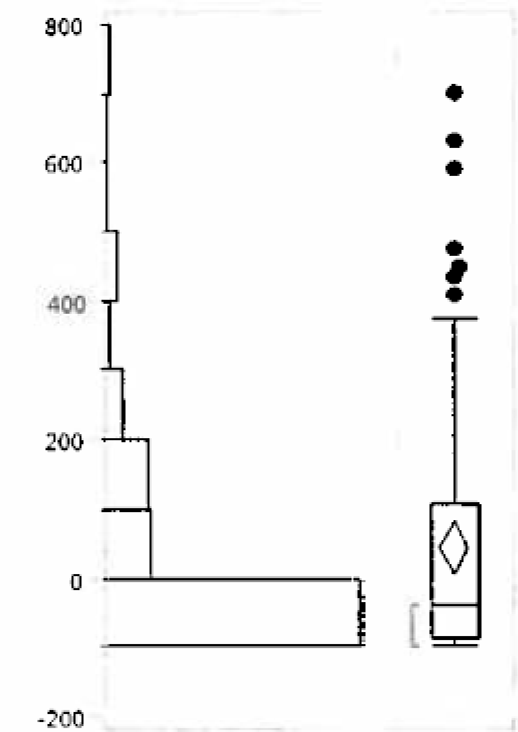
Mean	41.253742
Std Dev	104.67452
Std Err Mean	0.2872456
Upper 95% Mean	41.816739
Lower 95% Mean	40.690746
N	132793

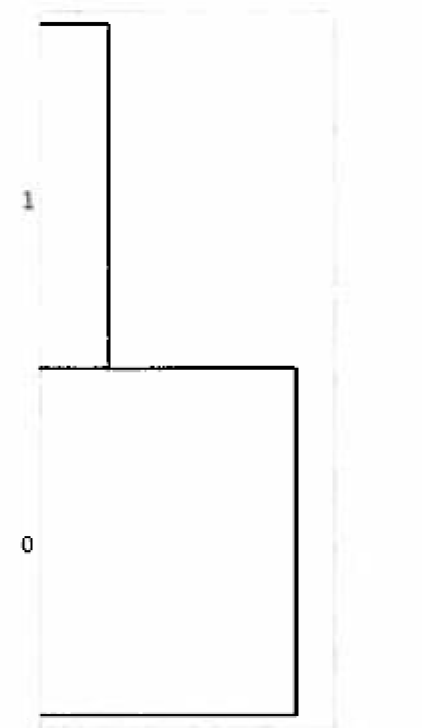
Level	Count	Prob
0	130989	0.98641
1	1804	0.01359
Total	132793	1.00000

N Missing 158

2 Levels

High Risk Portfolio





100.0%	maximum	700
99.5%		700
97.5%		663.7294727875
90.0%		357.39183404
75.0%	quartile	106.1419524
50.0%	median	-38.745562215
25.0%	quartile	-86.63152292
10.0%		-98.890257717
2.5%		-99.89523174475
0.5%		-99.95638413
0.0%	minimum	-99.95638413

Summary Statistics

Mean	43.997185
Std Dev	188.97365
Std Err Mean	18.897365
Upper 95% Mean	81.493658
Lower 95% Mean	6.5007126
N	100

Level	Count	Prob
0	79	0.79000
1	21	0.21000
Total	100	1.00000

N Missing 1

2 Levels

Final Model

Effect Summary

Source	LogWorth	PValue
quarter	205.814	0.00000
b0	125.648	0.00000
b12	45.588	0.00000
b3	14.832	0.00000
b4	8.355	0.00000
a1	8.004	0.00000
a5	4.171	0.00007
a5v2	3.375	0.00042
a4	2.790	0.00162
b6	2.079	0.00833
a0	1.150	0.07074
a2	0.327	0.47138
b5	0.212	0.61345
a3	0.142	0.72041

onverged in Gradient, 8 iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1721.966	94	3443.933	<.0001*
Full	9447.086			
Reduced	11169.053			

RSquare (U)	0.1542
AICc	19084.4
BIC	19956.7
Observations (or Sum Wgts)	72035

Measure	Training	Definition
Entropy RSquare	0.1542	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.1751	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.1311	∑ -Log(ρ[j])/n
RMSE	0.1792	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0639	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0362	∑ (ρ[j]≠ρMax)/n
N	72035	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	71927	9447.0864	18894.17
Saturated	72021	0.0000	Prob>ChiSq
Fitted	94	9447.0864	1.0000

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Lower 95%	Upper 95%
Intercept	4.3630214	0.0401765	11793	<.0001*	4.28517105	4.44272981
b0	-0.4912297	0.0205424	571.83	<.0001*	-0.5314401	-0.4509075
b12	0.00127076	8.8936e-5	204.16	<.0001*	0.00109576	0.00144443
b3	-0.0034436	0.0004316	63.67	<.0001*	-0.0042839	-0.0025919
b4	-0.0027051	0.000461	34.43	<.0001*	-0.0036016	-0.0017942
b5	7.14159e-5	0.0001414	0.26	0.6135	-0.0002065	0.00034769
b6	0.00037413	0.0001418	6.96	0.0083*	9.47914e-5	0.00065068
a0	-0.0006696	0.0003705	3.27	0.0707	-0.0013885	6.44154e-5
a1	-0.0021261	0.0003709	32.86	<.0001*	-0.002847	-0.0013928

Parameter Estimates

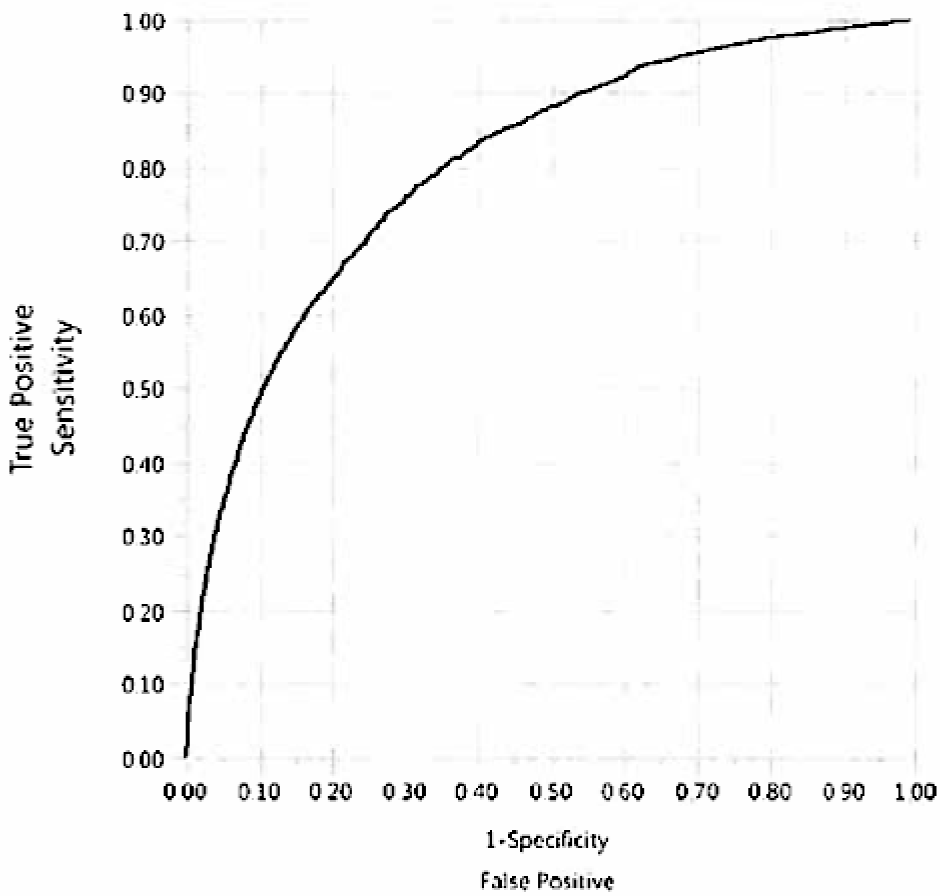
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Lower 95%	Upper 95%
a2	0.00040956	0.0005686	0.52	0.4714	-0.0006956	0.00153359
a3	0.00021198	0.0005923	0.13	0.7204	-0.000939	0.00138282
a4	-0.0002258	7.1657e-5	9.93	0.0016*	-0.0003675	-8.5548e-5
a5	0.00281187	0.0007056	15.88	<.0001*	0.00143954	0.00420552
a5v2	0.00257395	0.00073	12.43	0.0004*	0.0011543	0.00401602
quarter[0]	0.88303831	0.4493847	3.86	0.0494*	0.10855948	1.90735798
quarter[1]	1.32678938	0.5736267	5.35	0.0207*	0.37644175	2.70569956
quarter[2]	0.74479817	0.4085527	3.32	0.0683	0.03402757	1.66405838
quarter[3]	0.77558358	0.4094585	3.59	0.0582	0.06271532	1.69625329
quarter[4]	1.29634216	0.498507	6.76	0.0093*	0.45170338	2.45820555
quarter[5]	2.00428659	0.7017188	8.16	0.0043*	0.88112124	3.78032592
quarter[6]	1.01694027	0.4094578	6.17	0.0130*	0.30408362	1.93761079
quarter[7]	0.80368995	0.3800337	4.47	0.0344*	0.1360197	1.64832424
quarter[8]	0.52480959	0.3355696	2.45	0.1178	-0.0722247	1.25870547
quarter[9]	0.38817473	0.3353822	1.34	0.2471	-0.208411	1.12178059
quarter[10]	0.69675638	0.3554054	3.84	0.0499*	0.0682104	1.4798886
quarter[11]	-0.4316996	0.2200728	3.85	0.0498*	-0.8384896	0.02824417
quarter[12]	0.11455175	0.2813771	0.17	0.6839	-0.3948005	0.71710836

quarter[13]	0.03970559	0.2544147	0.02	0.8760	-0.4246413	0.57919674
quarter[14]	-0.5200459	0.2048109	6.45	0.0111*	-0.9007646	-0.0948105
quarter[15]	-0.45044	0.1931083	5.44	0.0197*	-0.8112423	-0.0518055
quarter[16]	-0.6511496	0.1849739	12.39	0.0004	-0.997766	-0.2705382
quarter[17]	-0.7633182	0.1839129	17.23	[illegible]	-1.107742	-0.3846434
quarter[18]	-0.9176827	0.1625759	31.86	<.0001*	-1.2247022	-0.5860154
quarter[19]	-1.1902077	0.1496726	63.24	<.0001*	-1.474043	-0.8862744
quarter[20]	-1.0968426	0.1510345	52.74	<.0001*	-1.3830796	-0.789947
quarter[21]	-0.7300074	0.1628371	20.10	<.0001*	-1.037445	-0.3977466
quarter[22]	-0.4432346	0.1911249	5.38	0.0204*	-0.800097	-0.0483649
quarter[23]	-0.8199144	0.1614274	25.80	<.0001*	-1.1247273	-0.490604
quarter[24]	-1.1162598	0.1462948	58.22	[illegible]	-1.3939007	-0.8194506
quarter[25]	-0.8582064	0.1612162	28.34	<.0001*	-1.1624301	-0.5290499
quarter[26]	-1.231461	0.1340462	84.40	<.0001*	-1.4870432	-0.9608783
quarter[27]	-1.6127733	0.1205374	179.02	<.0001*	-1.8438413	-1.3708574
quarter[28]	-1.5129703	0.118978	161.71	[illegible]	-1.7409999	-1.27414
quarter[29]	-1.2822098	0.1329982	92.95	<.0001*	-1.5359301	-1.0139106
quarter[30]	-0.8716151	0.1568272	30.89	[illegible]	-1.168535	-0.5526182
quarter[31]	-0.570803	0.1657187	11.86	[illegible]	-0.8837006	-0.232757
quarter[32]	-0.4947103	0.185145	7.14	0.0075*	-0.8416941	-0.1138
quarter[33]	0.06849947	0.2185228	0.10	0.7539	-0.3366131	0.52368069
quarter[34]	0.15560603	0.236249	0.43	0.5101	-0.2790234	0.65210643
quarter[35]	0.16887814	0.2428139	0.48	0.4867	-0.2771011	0.68018045
quarter[36]	0.73364594	0.2831162	6.71	[illegible]	0.22046422	1.33914478
quarter[37]	0.64908679	0.2639092	6.05	0.0139*	0.16807417	1.20977005
quarter[38]	1.12593556	0.3189396	12.46	0.0004*	0.55547064	1.81896427
quarter[39]	0.83572341	0.2920594	8.19	0.0042*	0.3089706	1.46390992
quarter[40]	0.75381085	0.2811113	7.19	0.0073*	0.24498633	1.35587646
quarter[41]	0.56921872	0.2721	4.38	0.0364*	0.07484158	1.14945885
quarter[42]	0.4181045	0.2480779	2.84	0.0919	-0.0360969	0.94233143
quarter[43]	0.6069975	0.2543027	5.70	0.0170*	0.1428601	1.14627117
quarter[44]	0.91050898	0.2915377	9.75	0.0018*	0.38494597	1.53783822
quarter[45]	1.20350864	0.3349559	12.91	0.0003*	0.60787508	1.93640889
quarter[46]	1.28502671	0.3543334	13.15	0.0003*	0.65896742	2.06645533
quarter[47]	0.74995046	0.2808898	7.13	0.0076*	0.24166008	1.35166905

quarter[48]	0.39506911	0.2291038	2.97	0.0846	-0.0266754	0.87608471
quarter[49]	-0.0559723	0.2027479	0.08	0.7825	-0.4324889	0.36542194
quarter[50]	-0.3741172	0.1806899	4.29	0.0384*	-0.7124159	-0.0020288
quarter[51]	-0.9342601	0.1436481	42.30	<.0001	-1.2067085	-0.6426131
quarter[52]	-1.565078	0.1134357	190.36	<.0001	-1.7829131	-1.337849
quarter[53]	-1.5322002	0.1081691	200.64	<.0001	-1.7402611	-1.3159015
quarter[54]	-1.3232495	0.1182726	125.17	<.0001	-1.5497863	-1.085674
quarter[55]	-0.800513	0.1384204	33.45	<.0001	-1.0638526	-0.5204445
quarter[56]	-0.9895079	0.1258375	61.83	<.0001	-1.2298225	-0.73595
quarter[57]	-0.4820933	0.1493308	10.42	0.0012	-0.7647718	-0.1783078
quarter[58]	-0.5066531	0.1442719	12.33	0.0004	-0.7802999	-0.2137814
quarter[59]	-0.3889481	0.1547652	6.32	0.0120*	-0.6814762	-0.0735749
quarter[60]	-0.1606187	0.1646838	0.95	0.3294	-0.470495	0.17663958
quarter[61]	0.28954102	0.1965895	2.17	0.1408	-0.076512	0.69690208
quarter[62]	1.01818578	0.2722246	13.99	0.0002*	0.52357174	1.59865835
quarter[63]	0.67911828	0.2352181	8.34	0.0039*	0.24677118	1.17385838
quarter[64]	0.73389342	0.2621326	7.84	0.0051*	0.2568272	1.29156497
quarter[65]	0.91279376	0.2920536	9.77	0.0018*	0.38605208	1.54096938
quarter[66]	0.42427315	0.2291413	3.43	0.0641	0.0024892	0.90538379
quarter[67]	0.02515233	0.1856296	0.02	0.8922	-0.3211329	0.40896081
quarter[68]	0.49546081	0.2232433	4.93	0.0265*	0.08404957	0.9634993
quarter[69]	0.33040287	0.2135492	2.39	0.1218	-0.0642802	0.77662285
quarter[70]	-0.0477062	0.1895979	0.06	0.8013	-0.4012052	0.34453465
quarter[71]	0.28760699	0.2183768	1.73	0.1878	-0.1154848	0.74459967
quarter[72]	0.46653926	0.2344532	3.96	0.0466*	0.03587882	0.95996903
quarter[73]	1.14296291	0.3188159	12.85	0.0003*	0.57288115	1.83584348
quarter[74]	0.22625341	0.2191465	1.07	0.3019	-0.1785226	0.68457841
quarter[75]	0.24461909	0.2230907	1.20	0.2729	-0.1665106	0.71235874
quarter[76]	0.05692823	0.1985026	0.08	0.7743	-0.3119879	0.46912207
quarter[77]	-0.1834544	0.1719784	1.14	0.2861	-0.5060419	0.16995361
quarter[78]	0.03731614	0.1919196	0.04	0.8458	-0.3200959	0.43489805
quarter[79]	-0.3694205	0.1648093	5.02	0.0250*	-0.6794864	-0.0318691
quarter[80]	-0.6950387	0.1460743	22.64	<.0001*	-0.9715059	-0.3978152

For log odds of 0/1

Receiver Operating Characteristic



Using predictorA='1' to be the positive level

AUC

0.80297

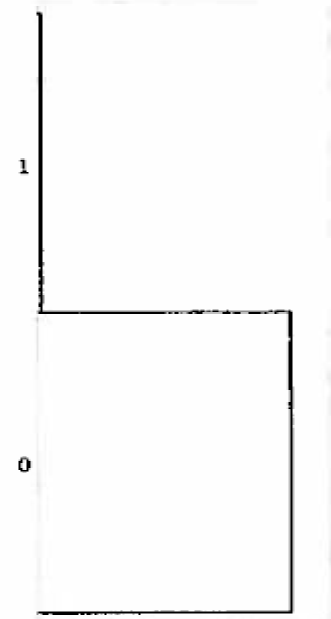
Confusion Matrix

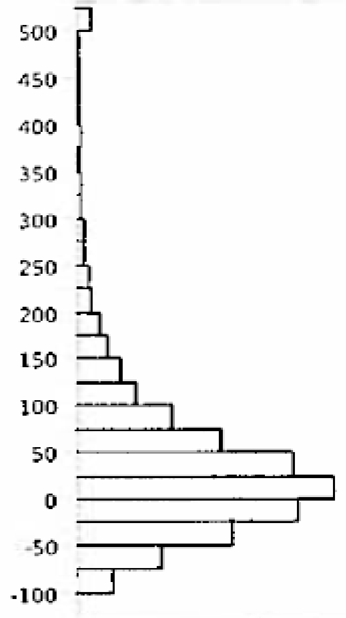
Training

Actual	Predicted	
predictorA	**0**	**1**
0	69382	59
1	2549	45

Final Model

Low Risk Portfolio





Frequencies

Level	Count	Prob
0	57926	0.98175
1	1077	0.01825
Total	59003	1.00000

N Missing 58933

2 Levels

Quantiles

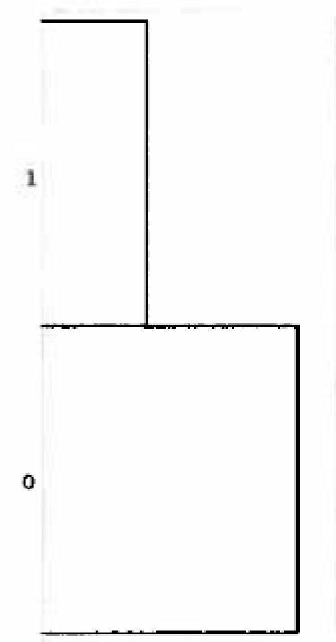
100.0%	maximum	500
99.5%		500
97.5%		315.14815899
90.0%		144.26185534
75.0%	quartile	69.15370797
50.0%	median	22.09922423
25.0%	quartile	-14.07600874
10.0%		-45.072307446
2.5%		-75.112215794
0.5%		-93.7836937798
0.0%	minimum	-99.97426825

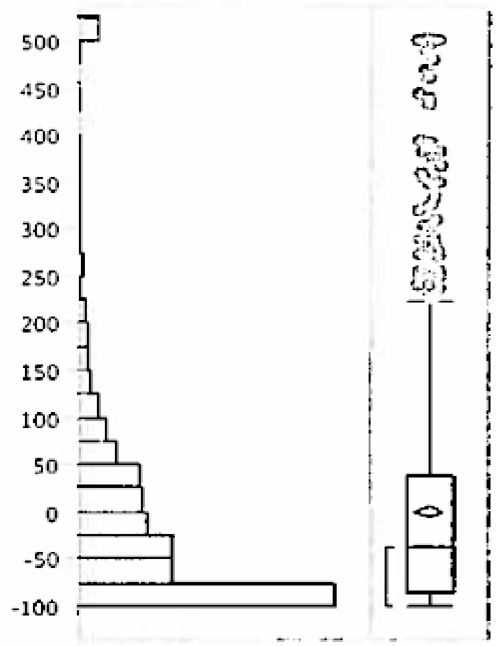
Summary Statistics

Mean	41.971272
Std Dev	94.346216
Std Err Mean	0.3884074
Upper 95% Mean	42.732552
Lower 95% Mean	41.209992
N	59003

Note that n missing refers to training sample. Analysis conducted on holdout sample

Final Model, High Risk portfolio





Frequencies

Level	Count	Prob
0	1199	0 70905
1	492	0 29095
Total	1691	1 00000

N Missing 1695

2 Levels

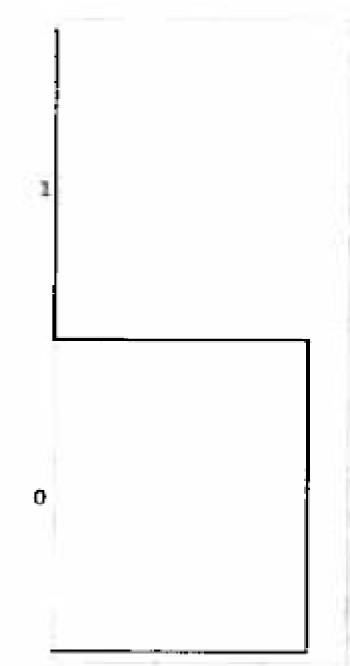
Quantiles

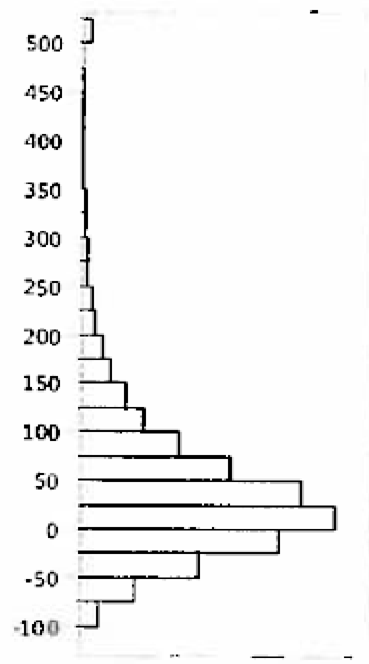
100 0%	maximum	500
99 5%		500
97 5%		499 44281168
90 0%		130 52846518
75 0%	quartile	37 95538664
50 0%	median	-36 82988039
25 0%	quartile	-86 12904873
10 0%		-97 789410648
2 5%		-99 788970613
0 5%		-99 9458102688
0.0%	minimum	-99 98486648

Summary Statistics

Mean	0 4484221
Std Dev	122 96627
Std Err Mean	2 9902962
Upper 95% Mean	6 3134955
Lower 95% Mean	-5 416651
N	1691

Final Model, Modified Low Risk Portfolio (p<.025 used as cutoff)





Frequencies

Level	Count	Prob
0	42788	0.98852
1	497	0 01148
Total	43285	1.00000

N Missing 43118

2 Levels

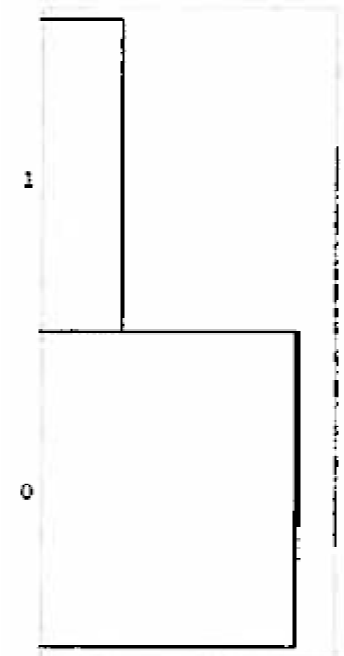
Quantiles

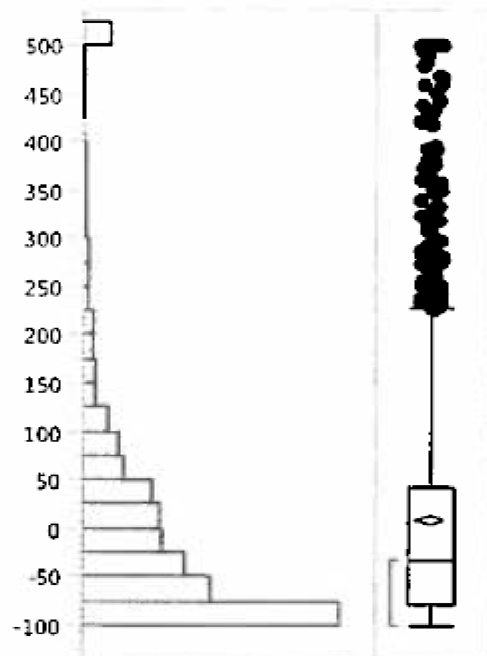
100.0%	maximum	500
99.5%		500
97 5%		306 003968285
90.0%		146.82257136
75.0%	quartile	74 504814775
50.0%	median	27 93124278
25.0%	quartile	-5 866294449
10 0%		-35.823832836
2 5%		-67 0823545475
0 5%		-90 3530243242
0 0%	minimum	-99 97426825

Summary Statistics

Mean	47 552372
Std Dev	90 90491
Std Err Mean	0 4369368
Upper 95% Mean	48 408776
Lower 95% Mean	46 695967
N	43285

Final Model, Modified high risk portfolio (p>.15)





Frequencies

Level	Count	Prob
0	2081	0 75508
1	675	0 24492
Total	2756	1 00000

N Missing 2734

2 Levels

Quantiles

100 0%	maximum	500
99 5%		500
97 5%		500
90 0%		153 33467911
75 0%	quartile	43 271125605
50 0%	median	-31 22339969
25.0%	quartile	-79 3915577225
10 0%		-96.144849002
2 5%		-99 7167780705
0 5%		-99.93515813825
0 0%	minimum	-99 99780593

Summary Statistics

Mean	8 96648
Std Dev	130.4553
Std Err Mean	2 4849755
Upper 95% Mean	13 839083
Lower 95% Mean	4 0938768
N	2756