Stock Ranking and Portfolio Selection: Revising and Developing Z-scores

Daniel Scinto, Jo Hardin
Pomona College

April 3, 2009

Abstract

In this study the problems with the current stock ranking system commonly used by portfolio managers, the Z-score, are examined. Metrics are proposed to measure the efficacy of different ranking systems, and then alternative ranking systems are developed. The systems are tested on the metrics and the results presented. Further, the mathematical qualities of the ranking systems and their relationships with each other are explored in various degrees of depth.

To Sage, who was always there with energy and a bark when the stress of Thesis weighed us down.

Contents

1 Introduction

Predicting the stock market has long been a dream of bankrupt men and failed financiers. Most academics consider the market to be a random walk with drift, use a 100 year average of returns for the drift, and advise: "good luck with the random walk." However, this will not stop leagues of portfolio managers and investment professionals from attempting to outperform each other and the indices.

Most recently, quantitative investment management has attempted to create equity portfolios based entirely on mathematical models. One of the most commonly used systems for ranking stocks, the Z-score, is partly responsible for managing billions and likely hundreds of billions of dollars in capital. Hence the motivation for improving such a ranking system to create more accurate results is incredibly high and the the motivation for our study.

2 Background

Over the past few decades, computers have changed the way financial analysts look at the markets. Computers have allowed vast amounts of computations to be done, paving the way for quantitative analysis. The largest area in quantitative finance has been in quantitative equity portfolio management. This is the result of equities being both a very common and a very dynamic investment area.

Although quantitative investment management is generally a new field in finance, there are some accepted practices that have been developed for quantitatively ranking stocks. First, there must be quantitative data that one is analyzing. Fortunately, stocks have a vast amount of readily available quantitative data, such as price, one year return, various items from the financial statements, and binary variables (such as, did the CEO go to a liberal arts school?, or has the company had an initial public offering in the past 5 years?). We will refer to these quantitative pieces of data as *Stock Factors*.

Stock Factor A factor is a quantitative measurement of some aspect of a company's stock. Examples of factors are Price-to-Book ratio, 3 Month Return (or 3 Month Momentum), and Number of Analysts covering the stock.

2.1 Stock Factors used in this study

We will use six different stock factors in our analysis. All six factors were found to be statistically significant in determining returns of stock prices in various studies and are commonly used by portfolio managers[1].

Price-to-Book ratio, or PB-ratio, is equal to:

$$PB = \frac{MarketCapitalization}{BookValue} \tag{2.1}$$

Market Capitalization is the current price of the company, i.e. the price at which a prospective buyer could purchase the company. The Book Value is the amount that the equity is worth according to the balance sheet in the company's financial statements. This is one of, if not the, most commonly used factors in stock analysis. Historically, a low PB corresponds to higher returns, since you are buying equity in the company at a cheaper price relative to its valuation based on Generally Accepted Accounting Principles (GAAP).

Our second fundamental factor, Price-to-Earnings (P/E) ratio to historical earnings growth,or $PEGH$, is the P/E ratio divided by the 5-year earnings growth of the company. Historically, a lower $PEGH$ leads to higher returns.

$$PEGH = \frac{\frac{MarketCapitalization}{NetIncome}}{5YearEarningsGrowthRate} \tag{2.2}$$

Another factor, related significantly to PB, is the $Size$ of the company. It is given simply by $Size = MarketCapitalization$, where Market Capitalization is defined as before.

Our fourth and only technical factor (a technical factor is one which only looks at past price movements of the stock), Three Month Momentum, is the previous 3-month return of the stock. Previous returns is one of the most commonly used factors amongst all technical factors. Three Month Momentum (M3M) is given by:

$$M3M_t = \frac{P_t + DIV}{P_{t-3}} \tag{2.3}$$

Where P_t is the current price of the stock (at time t), P_{t-3} is the price 3 months ago (at time $t-3$), and DIV are the amount of dividends paid out for the stock over the past 3 months.

[1] Chincharini and Kim, 121

We will also include two alternative factors (not commonly used by portfolio managers). The first is $EPSFN$, which is the number of financial analysts covering the stock and predicting what a company's earnings per share will be.

The second alternative factor is $RECC$, or Net Change in Analyst Recommendations. It is given by:

$$RECC = \frac{Upgrades - Downgrades}{EPSFN} \tag{2.4}$$

Where Upgrades is the number of analysts who have upgraded the stock in the past month, and Downgrades is the number of analysts who have downgraded the stock in the past month. $EPSFN$ (defined above) is the total number of analysts covering the stock.

2.2 Introducing the Z-Score

Given all of the quantitative data and different stock factors, an investor will want some way of combining the data systematically to rank various stocks. A traditional approach to combining data is to normalize it first. The current most commonly used method for normalizing the data is the *Original Z-Score.*

Original Z-Score for factor j and stock i

$$Z_{1,i,j} = \frac{X_{i,j} + \overline{X}_j}{S_j} \tag{2.5}$$

$Z_{1,i,j}$ represents the original Z-score for factor j and stock i. $\overline{X}_j$ is the arithmetic mean of the factor over every stock, S_j is the standard deviation of the factor over every stock, and $X_{i,j}$ denotes the actual factor value for stock j. That is, if stock i has a P/B ratio of 7, then $X_{i,PB} = 7$.

In an ideal world,the factors would be distributed normally, and one could say that a Z-score of 2 for stock j implies that it is in the top 2.27% of all stocks. Unfortunately, many factors are non-normal so so we stray away making statements relating the factor Z-score of a stock and its factor percentile.

2.3 The Original Overall Z-Score, Z_1

When building a portfolio of stocks a portfolio manager first computes Z-scores for each factor for every stock. Then, he chooses to combine the Z-scores in some way. The Original Z-Score Method described in equation 3.1 "normalizes" each factor so that different Z-scores can be arithmetically summed in a meaningful way. After ranking the stocks the portfolio manager can buy the best stocks and short the worst ones using various methods. Further, the rankings can be combined with qualitative analysis to decide which stocks to hold.

I should reiterate, a flaw in this idea is that the factors are distributed relatively normally, which they are not; showing the problems with the current Z-score method and providing a solution will be a main tenet of this paper.

Consider m different factors with weights $k_1, ..., k_m$ for these factors such that

$$\sum_{j=1}^{m} |k_j| = 1 \tag{2.6}$$

Note that some of the weights k_j will be negative and some will be positive. For example, for the factor Price-to-Book ratio, a portfolio manager may prefer low Price-to-Book ratio to high ones, and therefore he reverses the P/B Z-score distribution by making $k_{PB} < 0$. Low P/B stocks should have higher Z-scores, and high P/B stocks have lower Z-scores.

An overall Z-score for stock i is given by:

$$Z_{1,i} = \sum_{j=1}^{m} k_j Z_{i,j} \tag{2.7}$$

The Z-score serves first as a ranking of the stocks from best to worst. Further, it ranks the stocks comparably to each other. That is, a stock with with a Z-score of 3 has not only a higher rank than a stock with a Z-score of 1, it is a much, much better stock.

Additionally, a portfolio manager can create a portfolio from the Z-scores by weighting stocks by their Z-scores. However, often portfolio managers use other econometric techniques for weighting stocks and coming up with an actual portfolio.

3 The Problems with Z_1

If the factors that the portfolio manager is looking at are distributed normally, there will be few, if any problems with using Z_1. Unfortunately, many very useful factors are not normal, and some are extremely non-normal. As a further issue, some important factors are binary which creates problems in using Z_1 as a ranking system.

Before delving further and showing histograms of the factors, we should say something about the data. The data for our study is from 1994 to 2006, taken monthly. The stocks we include are from the S&P 1500. It was originally created and used in Pomona College's Spring 2008 Quantitative Investment Management class, taught by Professor Ludwig Chincarini.

3.1 The Percentile Rank Problem

One of the first issues that arises from Z_1 is what will be referred to as the percentile rank problem. The problem is that one would think, in most ranking systems, that a stock that ranks in the first quartile for two factors and the third quartile for one factor would rank higher than a stock that ranks in the first quartile for only one factor and the third quartile for two factors, assuming all factors are to be equally weighted. Because of skewed data and huge outliers, the internal consistency does not always hold. Figures 1,2,3, and 4 show the four histograms of four factors: Price to Book Ratio, 3 Month Momentum, Market Capitalization (Size), and Net Change in Analyst Recommendations.

First, it should be divulged that the x-axes on the Figures for PB, $Size$, and $M3M$ were constrained to what is shown. All three had a good chunk of outliers, although PB had comparatively more outliers and was the only factor with negative outliers. $Size$ and $M3M$ are constrained to being positive by their definition.

As you can see from Figures 1,2,3 and 4, the distributions for each factor are all very different. $M3M$ is reasonably symmetric and bell shaped, whereas $size$ and PB are very skewed. Additionally, size has only positive values, whereas PB has almost all positive values, with a huge drop off at zero and large negative and positive outliers. $RECC$ is unique in its own way, with a vast majority of the data equal to zero and a range of [-1,1].

To illustrate the percentile ranking issue let's compute Z_1 for three factors, equally weighted; in this case pick PB, $M3M$, and $Size$ to be our factors.

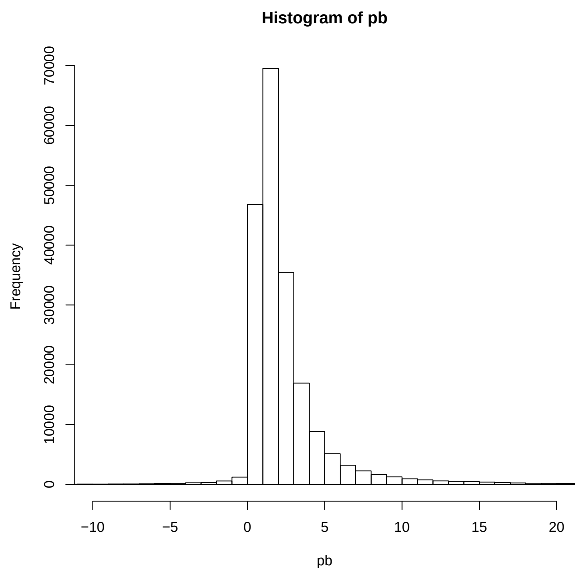

Figure 1: Histogram of PB

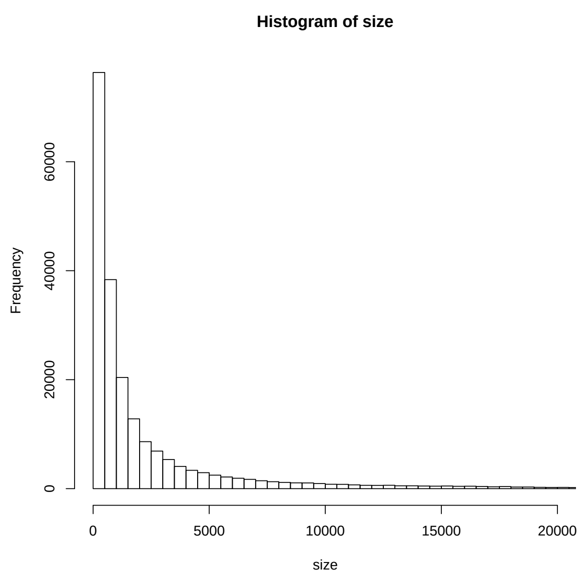

Figure 2: Histogram of size

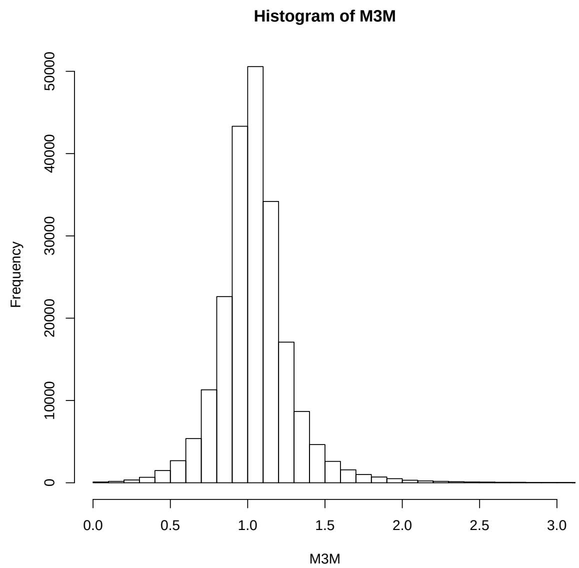

Figure 3: Histogram of M3M

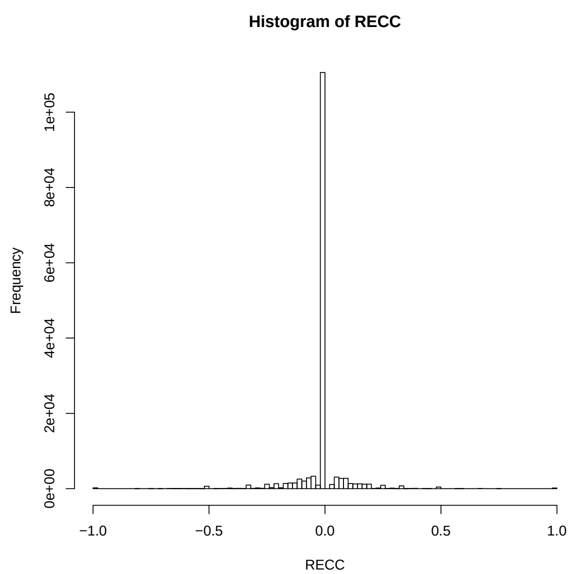

Figure 4: Histogram of RECC

Suppose stock A is in third quartile for PB and $Size$ and the first quartile for $M3M$, and suppose stock B is in the 1st quartile for the first two factors, and the third quartile for the third factor.

Stock A's Z-scores: -.025, -.11, and -.52. Average = -.29

Stock B's Z-scores: -.05, -.29, and .6. Average = .09

The reason for this in the above case is that $M3M$ is a very normal distribution, so it has a standard deviation of .22 and mean of 1.03, resulting in the 75th percentile falling .6 standard deviations away from the mean. PB, on the other hand, has a standard deviation of 88 and mean of 4.7 yet almost all the data lies between 0 and 20; the huge standard deviation is the result of ridiculous outliers. Hence the 75th percentile for PB actually has a negative Z-score that is tiny in absolute terms (its score is -.025 versus .6 for a stock of the same percentile for $M3M$); the same reasoning also applies to $Size$.

When averaged, the third distribution skews the result because it has so much weight: the third quartile for $M3M$ is .6, whereas it is tiny (even negative) for PB and $Size$ leading to a positive Z-score (.09) when we wanted it to be negative. We would have expected A's and B's Z-score to be of different signs: A's positive and B's negative. Most importantly, we would have wanted A to be ranked higher than B.

The above regularities summarize the percentile rank problem of Z_1.

3.2 The Qualitative Ranking Problem

There is an additional problem with Z_1, though this one is both harder to quantify and likely impossible to solve completely. Ideally, we would want a ranking system with some type of meaning, not only to the ranks (i.e. a stock with score 1 is better than a stock with score 0), but also some relative ranking (i.e. how *much* better is a stock of rank 1 than a stock of rank 0.

To illustrate this problem, let's only look at one factor PB, and assume that it is the only factor. As one can see from Figure 1, almost all of the data falls between -5 and 20, yet the standard deviation is 88. The result is that almost all the stocks for the factor Z-score under Z_1 fall between 0 and -.1, telling little about the differences between the relative companies. If the Z-score is not between 0 and -.1, then it is an outlier, and that is about all the information revealed.

In this situation, a portfolio manager would want to differentiate between a stock with PB of 2 and a PB of 8; one stock is valued at 4 times more of its book value, which is a *huge* difference. According to our system, however, it is the difference between a Z-score of -.028 and .086, which qualitatively means nothing (particularly when averaged with our other factors).

Ideally we want our ranking system to have some qualitative meaning in addition to its strict rank. Our current system fails miserably in this regard.

3.3 The Return Problem

An additional problem with Z_1 is how little it explains of actual returns.

To show this problem, I first need to introduce the weights used by Z_1. As I discussed earlier, the weights can be at the portfolio managers' discretion, and it is important that the factor Z-scores are compiled in a way to create a relevant Z-score. For example, if PB has a negative relationship with returns, PB should be weighted negatively. Further, to avoid complicating the issue, we will use equal magnitude weights for all six factors in this segment. To determine whether factors were related positively or negatively with returns we regress each factor against returns and obtain the following results:

Six Factors Against Returns	
	$T - Value$
$M3M$	-9.8
$RECC$	2.7
$EPSFN$	-7.0
$Size$	-2.5
PB	1.1
$PEGH$	-2.4
$Adjusted - R^2$	$.0026$

Fortunately all of our factors except PB are statistically significant, so we have chosen factors that are good to study because of their economic implications. Additionally, based on the sign, we know which factors to weight negatively and which to weight positively. $RECC$ is positive while the rest of the factors are negative. It should be noted that, arguably, we want the sign of the regressions done for each factor individually, not taken all at the same time. Doing a multi-variable regression relates the factors to each other and the response variable (returns), instead of looking at only the effect of the factor (not controlling for anything else).

After completing the individual regressions, none of the signs changed, except the one for PB, which turned negative; this result, multi-collinearity, implies some of the other factors are both correlated with PB and explain returns. This is not surprising; the size of a company, the number of analysts covering the company, and the Price/Book Value are indeed very related. Further, while by itself a *lower* PB means higher returns, if you control for other factors a *higher* PB correlates with higher returns. Given that PB was not statistically significant and that going with the individual model made more intuitive sense, it was decided that we would use the signs from the individual regressions, implying that PB would be weighted negatively.

The most disconcerting part of the results in table 3.3 is the R^2. The factors explain almost nothing of stock returns. It should be noted that the use of a panel regression might yield a much higher R^2 and should be looked at in future work.

Now that we have the correct signs, we can combine the individual factor Z-scores to obtain Z_1, according to the following equation:

$$Z_{1,i} = \frac{-Z.PB_{1,i} - Z.Size_{1,i} - Z.PEGH_{1,i} + Z.RECC_{1,i} - Z.EPSFN_{1,i} - Z.M3M_{1,i}}{6} \tag{3.1}$$

This gives us Z_1 for stock i. Now we can regress the Z_1 for each stock against the return in the *next* month (i.e. we want to know what kind of value for Z_1 will yield what kind of return next month). A linear regression of Z_1 against monthly stock returns from 1994 to 2006 yields the following:

Z_1 **against returns**	
	Z_1
$T - Value$	4.70
$Adjusted - R^2$	$5.61 * 10^{-5}$

Clearly the current ranking system is not explaining anything about stock returns from a significant historical period (monthly returns from 1994-2008). A reason for the high T-Value is the extremely large data set (380,000 data points). Clearly the Z_1 is significant even though it is not explaining much variability in returns. Ideally, we would like our ranking system to more useful in predicting returns. However, this obviously depends to a large extent on the factors.

3.4 The Outliers Problem

Ideally a good ranking system should be able to handle huge outliers that are the result of certain ratios (price/book, PE/growth rate). Outliers should be considered, but they should not hugely affect the ranking of the rest of the stocks. Clearly, this is an issue for Z_1, which places way to much emphasis on outliers as a result of its use of standard deviation, which tries to haul in outliers by giving them a heavier weight.

3.5 Strange Theoretical Distributions Problem

Ideally a good ranking system should relate the economic theory behind factors to its quantitative methods. The best example of a strange theoretical distribution is the PB factor. In economic theory, a stock with a negative PB is most like a stock with an infinitely positive PB.

$$PB = \frac{Price}{BookValue}$$

The Price of a company is always a large positive value from 10 Million to 1 Trillion dollars. Book Value, on the other hand, can be very small as a company approaches bankruptcy, and even be negative. Hence a Company close to bankruptcy will either have a very, very large positive PB ratio (a result of the very small book value compared to company price), or it will have a negative PB (as a result of the negative book value). These near bankrupt companies are qualitatively very similar yet they have book values ranging from negative infinity to 0 and from 1000 to positive infinity.

Currently, near bankrupt companies can have PB Z-scores including both large positives and large negatives (but not anything in between), even though they are nearly identical in terms of economic theory. Likewise, a stock with a small negative PB will have a nearly identical Z-score to a stock with a small positive PB, when in fact these stocks are at the opposite end of the economic spectrum; the small positive PB stocks are the cheapest to buy, and the small negatives (i.e. small in absolute terms) are the most expensive. In this case, the easiest way to correct this is mapping all the negatives to a huge positive number; this would clump the near bankrupt companies Z-scores together. It is likely for this reason that PB is the only insignificant variable in the regression against returns done earlier.

In sum, a ranking system should be able to handle problems like PB. That is our ranking system should accurately handle negative values that are theoretically the same as infinitely large positive values; if this is not possible, it should at least handle them in such a way that they do not affect the rankings of other stocks unnecessarily. A system that is robust in handling this problem and similar ones is ideal. Z_1 fails in this regard.

4 Development of Metrics

Having identified problems with Z_1, it is a good time to develop metrics to evaluate various ranking systems. This will allows us to identify ways to come up with new Z-score models that address the problems afore mentioned. Each of the metrics described below was developed specifically to address one or more of the problems identified with the Z_1 ranking system.

4.1 The Percentile Rank Metric

The percentile rank metric stems out of the percentile rank problem. This metric seeks to measure "how far away" the ranking system is from a purely "percentile rank" ranking system. Let us define a percentile rank system.

For each factor j, line up all of the stocks from greatest to least, and rank them 1 to n. If the factor is to have a negative weight in the weighting later, then rank them least to greatest from 1 to n. Do this for every factor. For each factor j, convert each stock i into a factor percentile rank $FactPer_{j,i}$ by Equation 4.1.

$$FactPer_{j,i} = \frac{(n - stockRank_{j,i})}{n} \tag{4.1}$$

n is the number of stocks and $stockRank_{j,i}$ is the rank of stock i for factor j by the system described above. In the future, we will describe the process of going from data to a percentile rank as "taking the percentile rank".

Now we combine percentile ranks from the factors for each percentile rank with the following function.

$$PerSum_i = \sum_{j=1}^{m} |k_j| FactPer_{j,i} \tag{4.2}$$

k_j is the weighting for factor j picked by the portfolio manager. Note that the *absolute* weighting is used here because we already accounted for the sign when we did the ranking earlier. $PerSum$ will range from 0 to m, where m is the number of factors. Hence a stock in the 0th percentile for each factor will have a $PerSum$ of 0, and a stock that is in the 99th percentile for each factor will have a $PerSum$ of m*.99.

Next, stocks are ranked from from 1 to n based on $PerSum$ and we perform the following computation (i.e. "take the percentile rank" of $PerSum$).

$$PerRank_i = \frac{(n - PerSum_i)}{n} \tag{4.3}$$

which gives a percentile rank for stock i. Now, each stock has a numerical "percentile" between 0 and 1 that accurately combines its percentiles for each individual factors. One should note that the method described above is the same one used when combining test scores from standardized tests with different subsections (i.e. one could score in the 95th percentile for the math section and the 95th percentile for the verbal section of a test and get a score of the 99th percentile overall).

Let $Z_{\alpha,i}$ denote the overall Z-score for stock i for a certain ranking system α. To compare to the percentile ranks that were developed from the factors themselves it is necessary to create a percentile rank for the ranking system α. Take the rank for $Z_{\alpha,i}$ from 1 to n and convert this to a percentile rank to create a ranking system percentile rank, $P.Z_{\alpha,i}$, for each stock i. Now we calculate how far away the ranking system α is from a straight percentile rank ranking system by the following equation.

$$PM_\alpha = \frac{\sum_{i=1}^{n}(|PerRank_i - P.Z_{\alpha,i}|)}{n} \tag{4.4}$$

PM_α is the average misfire for ranking system α, and is a final numeric metric for comparing different ranking systems. The lower PM_α, the better. In addition to the mean of $|PerRank_i - P.Z_{\alpha,i}|$, one could also look at its median and standard deviation.

4.2 The Outliers Metric

The outliers metric measures the effect of outliers on the rest of the distribution. First, compute any ranking system Z_α. Map the top 2% of each factor to the 98th percentile, and map the bottom 2% of each factor to the 2nd percentile. Now compute Z_α on our revised factors and call this ranking $Z.Out_\alpha$.

To measure how far Z_α is from $Z.Out_\alpha$, perform the following computation:

$$OM_\alpha = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{Z_{\alpha,i} - Z.Out_{\alpha,i}}{\sigma_{Z.Out_\alpha}}\right| \tag{4.5}$$

Here we take the average deviation of $Z.Out_\alpha$ from Z_α. We divide this by $\sigma_{Z.Out_\alpha}$ to scale the difference. The division serves as a scaling factor for the difference between the two ranks (i.e. that way we can compare metrics between different ranking systems). OM_α is the outlier metric for ranking system α, and, as a single number, can be compared between ranking systems. The lower the value of OM_α, the less that outliers are effecting the ranks of the other stocks. Like in the case of the Percentile Metric, looking at the standard deviation and median of $|\frac{Z_{\alpha,i} - Z.Out_{\alpha,i}}{\sigma_{Z.Out_\alpha}}|$ will show us additional information in cases where the means (i.e. $OM\alpha$) are very close together.

4.3 The Strange Theoretical Distributions Metrics

Metrics stemming from the Strange Theoretical Distributions Problem (STDP) need to be made on a case by case basis for each strangely distributed factor. We will lump any metrics developed as a solution to the STDP as the Strange Theoretical Distributions Metrics.

The one we have discovered thus far is the PB problem. As discussed earlier, theoretically the distribution should map negatives to positive infinity (i.e. a PB of -1 is "greater than" a PB of 1000). To handle the specific PB problem map all of the negative PB's to the largest positive PB ratio. Then rank the stocks on the altered factor and the other stock factors using the Z_α system; call this ranking system $Z.PBM_\alpha$. Compute Z_α as usual.

$$PBM_\alpha = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{|Z_{\alpha,i} - Z.PBM_\alpha|}{\sigma_{Z.PBM_\alpha}}\right) \tag{4.6}$$

For reasons similar to those in the above section, we want the lowest PBM_α possible.

4.4 The Returns Metric

While its importance amongst academics is heavily disputed, this research team has decided to include at least one practical metric that does what

Portfolio Managers care about, predicting stock returns. For this metric, we regress our ranking system Z_α against stock returns in the *next* month. That is, we want to predict future returns.

The goal is to achieve the highest R^2 possible. For an example of this test for Z_1 please refer to section 3.3. The R^2's can be compared directly between different ranking systems.

5 Development of Ranking Systems Z_α

Having developed metrics, we turn to developing some alternative ranking systems.

5.1 Z_2, the MAD Median Ranking System

A good first alternative that handles outliers well is the MAD Median Z-score. Instead of the mean in Z_1, it uses the median, and instead of the standard deviation, it uses the the adjusted MAD. The adjusted MAD is the absolute median deviation from the mean, multiplied by 1.4826. The major benefit of the MAD is that it does not "haul in" outliers very much because there is not a squaring effect (that is present in the standard deviation). Hence, in distributions with large outliers, the MAD will not be as large as the standard deviation and hence the Z_2 scores will be larger.

$$Z_{2,i,j} = \frac{X_{i,j} - Median(X)}{AdjustedMAD(X_j)}$$

5.2 Z_3, the Trimmed Mean Ranking System

$$Z_{3,i,j} = \frac{X_{i,j} - trimmed(\overline{X}_j)}{trimmed(S(X_j))}$$

The trimmed mean Z-Score takes the middle 80% of the data and calculates the mean and standard deviation from that. This is to reduce the effect of large outliers, and should perform better on our Percentile Rank and Outliers metrics. However the outliers are still included in the data, just with a revised mean and standard deviation. The logic here is very similar to that of Z_2 since $trimmed(S(X_j)) << S(X_j)$.

5.3 Z_4, the Logarithmic Ranking System

$$Z_{4,i,j} = \frac{ln(adjX_{i,j}) - \overline{ln(adjX_j)}}{S(ln(adjX_j))}$$

For Z_4 the bottom 2% of the data is taken out if Z_4 contains negatives. If the 2nd percentile is negative then all the rest of the data is made positive by adding the value of the 2nd percentile to the data. We take $ln(adjX)$, where $adjX$ is the data with the previous adjustment, and compute $initial.Z_4$ as shown above. Finally, the bottom 2% is mapped to $min(initial.Z_4)$. The result of this final mapping is Z_4. The goal of Z_4 is to again to haul in outliers with the use of a logarithm, while hopefully not decreasing performance on the Returns and Strange Distributions Metrics. In order to use a logarithm, all the data values need to be positive, which explains the mapping to turn all of the data values positive before the logarithm is used.

5.4 Z_5, the Percentile Ranking System

This ranking system, which takes the bull by the horns with respect to the percentile rank metric, *is* the percentile rank metric. That is, the percentile rank of each factor is found, the factors are summed by the given weighting system, and then the percentile rank of that distribution is taken. Hence this ranking system will have a PM_α equal to zero. For a computation of the Z_5, please see section 4.1; in that section $PerRank_i$ is Z_5 for stock i.

6 Testing Results: The Ranking Systems on the Metrics

After developing metrics and ranking systems, the natural course of action is to test the ranking systems on the metrics.

6.1 The Percentile Rank Test

The first metric we tested was the percentile rank metric. The following results were found.

Percentile Test			
Z_α	PM_α	$\sigma Misfire$	$MedianMisfire$
Z_1	.15	.13	.12
Z_2	.15	.15	.10
Z_3	.10	.10	.07
Z_4	.06	.09	.04
Z_5	.00	.00	.00

Based on PM_α, Z_1 ties for last in this metric. This is not surprising since all of the other Z-scores we created had this metric in mind. Z_5 is zero, given that it *is* the percentile rank metric, this should not be surprising. Most fascinating is that our most creative metric, Z_4, performs quite well, with an average misfire of .07.

One unfortunate result of the above is that Z_2, the MAD-Median Ranking, was no better than Z_1; given that the MAD-Median is supposed to handle outliers better (i.e. it uses a trimmed standard deviation, which should result in a lower standard deviation that better differentiates between the bulk of the data close to the mean) we would have expected it to outperform Z_1 in percentile rank.

There is nothing too surprising with the median and standard deviation of the misfires. For all the ranking systems (save Z_5), the data was skewed upward from 0. Given that the data lies so close to 0, yet 0 is a hard limit (i.e. there cannot be a negative misfire), the skew upward was almost expected. Considering how small the mean and median misfire were for Z_4, its standard deviation of .09 is quite large.

Given that every ranking system created performed better or as well as Z_1, the above results are promising.

6.2 The Outliers Test

The outliers tests takes the bottom 2 and top 2 percentiles of each factor and maps them to the 2nd and 98th percentile, respectively. Then it computes the ranking on these revised factors, and sees "how far away" the revised rankings are from the original ones. The results, the mean, median, and standard deviation of the "misfire", are shown in Figure 6.2.

Outliers Test			
Z_α	OM_α	$\sigma Misfire$	Median Misfire
Z_1	.45	.96	.31
Z_2	.30	9.7	.04
Z_3	.45	21	.00
Z_4	.15	.60	.05
Z_5	.001	.01	.00

These results are very interesting. First we note that Z_1 comes in dead last; it does nothing to control for outliers so this is expected. Also, first place goes to the percentile rank metric, once again. Given that Z_5 accords no credence to outliers and grades by percentile, this was also relatively expected. Z_2 and Z_3 perform admirably, however both have huge standard deviations for the misfire ($|\frac{Z_{\alpha,i} - Z.Out_{\alpha,i}}{\sigma_{Z.Out_\alpha}}|$) that would be worth exploring further. Note how low their medians are compared to their standard deviation; the skew and outliers are pulling the mean significantly.

Z_4 performs pretty well all around, and its misfire suffers from a skew as well, but not as pronounced. Given that Z_4 is logarithmic, performing relatively well on this metric was expected.

6.3 The PB Test

The PB test, stemming from the Strange Theoretical Distributions Problem, maps the negatives of the PB factor to the maximum PB, recalculates the rankings on the altered factors, and tests "how far" they are from the original rankings. The results from our data set are shown in Table 6.3.

PB **Test**			
Z_α	OM_α	$\sigma Misfire$	Median Misfire
Z_1	.12	.71	.05
Z_2	.16	5.4	.01
Z_3	.13	1.0	.00
Z_4	.22	.97	.06
Z_5	.05	.19	.03

Z_5, the factor least affected by extreme values, once again outperforms. As expected, Z_2 and Z_3 come in far ahead of Z_1 if one uses the median as the yardstick, which we will do. There are some extreme values in Z_2 (and to a lesser extent Z_3) that throw off the mean.

Most surprising is Z_4. Normally a top performer, it comes in as the worst ranking on this metric (though its median is almost equivalent to Z_1). PB is a factor that has negative values; as such, normally Z_4 takes all of the negatives out below the second percentile, adds the second percentile to the data, calculates Z_4 on the adjusted data according to Equation 5.3, and then maps back in all of the negative points to the new minimum value.

However, Z_4 does not do this for the adjusted PB factor because all the negatives of PB are mapped to the maximum. Therefore, Z_4 skips the mapping portion (since there are no negatives) and calculates the logarithm instead. Hence, due to how Z_4 was defined, it is calculated in two very different ways under this metric. The difference in calculation, the mapping and adding the second percentile in normally and not doing it for $Z.PBM_4$, explains why it performs so badly on this metric.

6.4 The Returns Test

The most practical of our tests looks at how much of stock returns is predicted by our ranking system. Hence each of the rankings was regressed against the next month's stock returns.

Returns Test		
Z_α	$T-Value$	R^2
Z_1	4.70	$5.6 * 10^{-5}$
Z_2	1.04	$2.0 * 10^{-7}$
Z_3	0.78	$-1.0 * 10^{-6}$
Z_4	17.0	$7.4 * 10^{-4}$
Z_5	12.6	$4.2 * 10^{-4}$

Fascinating results! The most striking thing from the data are the adjusted R^2s - they are meaninglessly small and one is even negative(Z_3). So almost none of stock returns are being explained by our model. Note that this was done in a regression without panels for time, which skews the R^2's downward, given that the correlation between stock returns in any given month is incredibly high.

Getting beyond that, some of our ranking systems are definitely statistically significant because of their high T-statistics and p-values (although they are not explaining much). Z_4 is the clear winner, with Z_5 not far behind; Z_1 is still significant, and Z_2 and Z_3 fail miserably at predicting returns.

7 Ranking System Distributions

Having developed ranking systems and metrics and having tested the systems on the metrics, a natural follow up questions is: How does each ranking system relate to each other? Why did some systems perform differently? A good place to start the analysis is to look at the histograms of some of our ranking Z-scores.

7.1 The Distribution of Z_1

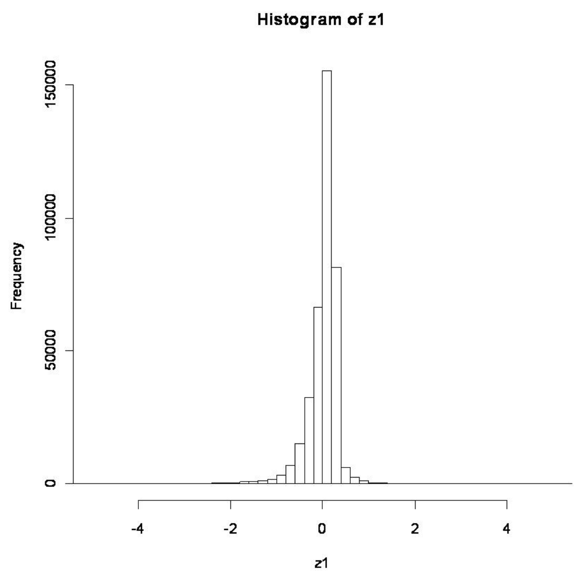

Figure 5: Histogram of Z_1

As evidenced by Figure 5 which has a constrained x-axis, Z_1 has a slight skew to the positive side. Almost all the data is concentrated between -1 and 1. This is why figure 6 is useful; it shows how spread out the outliers are. The reason that the median, the box, and the whiskers aren't visible is that they are all focused on the line in the middle; i.e. they are all approximately 0 compared to the outliers. The most extreme outliers (3 positives and about 20 negatives) stray out all the way to positive and negative 100. In addition, there are a lot of outliers that are less extreme that lie either between -5 and

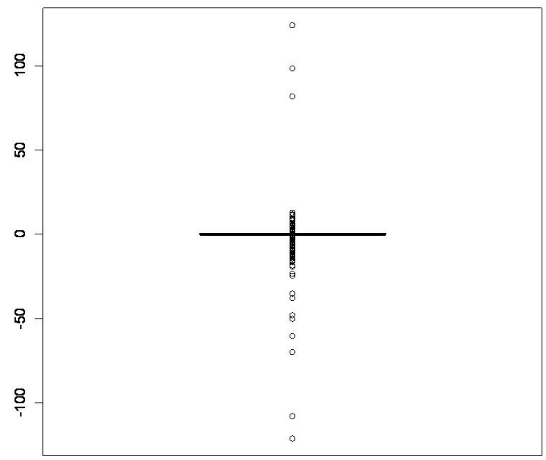

Figure 6: Box Plot of Z_1

-20 or 5 and 20. These outliers explain why the standard deviation for Z_1 is so large.

7.2 The Distribution of Z_2 and Z_3

As shown in Figures 5, 7, and 8, the graph of every ranking system is skewed to the negative side. Like the histogram of Z_1, the axes of the histograms for Z_2 and Z_3 are constrained. A view of the box and whisker plots for Z_2 and Z_3 would result in a similar view as the box plot for Z_1; a ton of outliers lie outside the box and whiskers. Z_3, the trimmed mean, has a particularly heavy skew to the left.

7.3 The Distribution of Z_4

The use of a logarithm has resulted in a much more bell shaped distribution for Z_4, visible in Figure 9. Also interesting is the recognizable bump in the histogram of Z_4 (Figure 9) on the positive side.

We shall address the bell shaped nature of the distribution of Z_4 first, and

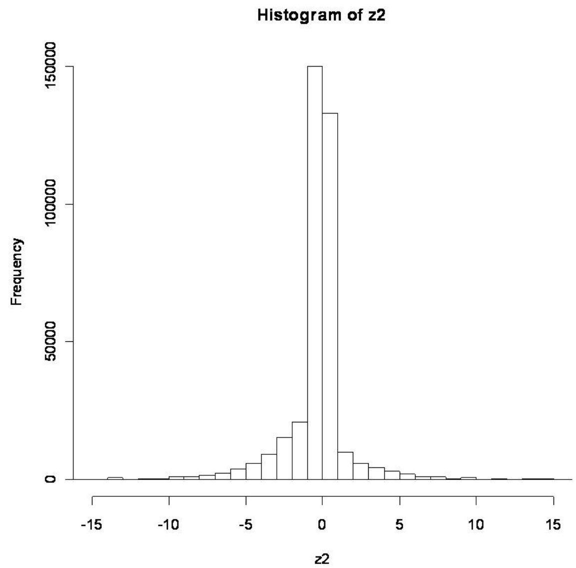

Figure 7: Histogram of Z_2

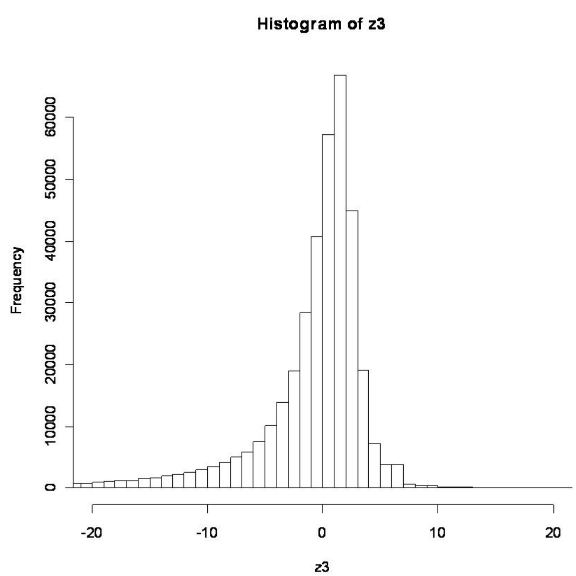

Figure 8: Histogram of Z_3

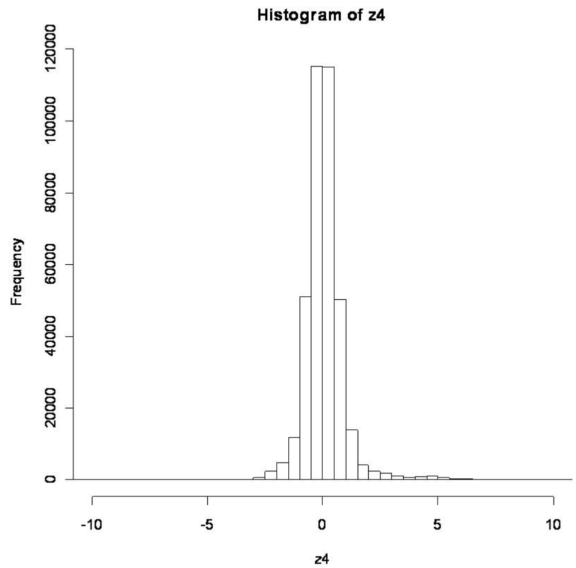

Figure 9: Histogram of Z_4

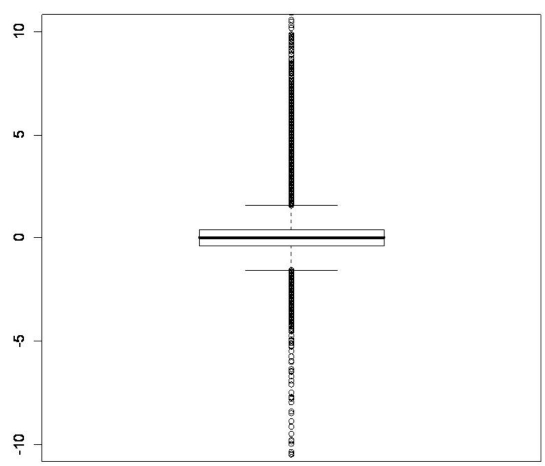

Figure 10: Box Plot of Z_4

then its bump on the positive side. Note that the logarithm brings outliers in and clumps the data together. Although to create Z_4 we added logarithms together (the Z-scores for each factor are added together to create an overall Z_4), this results in the multiplication of the factors inside a logarithm because:

$$ln(j_1) + ... + ln(j_m) = ln(j_1 * ... * j_m)$$

So net, we are taking a logarithm, which typically results in a more bell shaped distribution. As additional evidence to this clumping phenomenon, Figure 10 shows the box and whisker plot for Z_4. It is mildly useful to read, as we can make out the box, the whiskers,and the median. The distribution is centered around the median at 0 and the first and third quartiles are located at -.4 and .4; the whiskers lie at about -2 and 2. The min and max are much tighter than the other distributions, located at -18 and 22, respectively.

7.3.1 The Positive Bump of Z_4

Note that if a factor has negative values, then the bottom 2% of the values are taken out, and all of the data shifted upward by the 2nd percentile to make the data positive. Then the factor Z-score, $Z_{4,Factor}$ is taken, and the previous negative values mapped back to the minimum of the factor Z-score. In our case, only half of our factors, PB, *pegh*, and $RECC$ have negative values and are subject to this. Hence we would expect a bump on the negative side on the histogram of $Z_{4,PB}$,$Z_{4,RECC}$, and $Z_{4,pegh}$. The negative bump is visible in Figure 11, located at -12.

The weights for $Z_{4,pegh}$ and $Z_{4,PB}$ are negative, so this bump get transferred to the positive side in the combined weighting. Further, a large negative *pegh* and large negative PB are highly correlated (a company with a negative book value also probably has negative earnings). As we can see from Figure 11 that shows the distribution of $Z_{4,pegh}$ (the histogram of $Z_{4,PB}$ is similar), these stocks will have values for both $Z_{4,pegh}$ and $Z_{4,PB}$ of about -12. When combined with the other factors Z-scores that are close 0, shown in Equation 7.1, we would expect these stocks to have a Z_4 of approximately 4.

$$\begin{aligned} Z_4 &= \frac{1}{6}(-Z_{4,pegh} - Z_{4,PB} + Z_{4,RECC} - Z_{4,Size} - Z_{4,EPSFN} - Z_{4,M3M}) \quad (7.1) \\ &\approxeq \frac{1}{6}(-(-12) - (-12) - 0 - 0 - 0 - 0) = 4 \end{aligned}$$

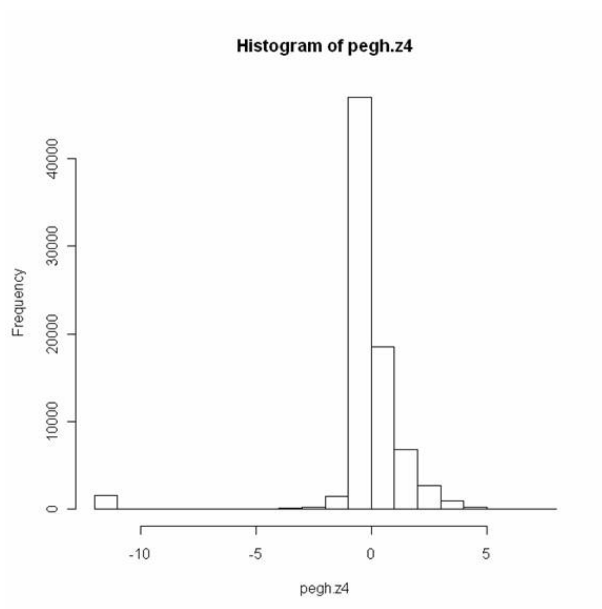

Figure 11: Histogram of $Z_{4,pegh}$

If we look back at Figure 9, we see that the small positive bump of Z_4 is centered around none other than 4, just as we suspected. Voila!

A natural question arises, why is there no bump for $Z_{4,RECC}$ on the negative side of the histogram of Z_4. The weight for $Z_{4,RECC}$ is positive, so its negative bump should be visible in Z_4. The bump would be centered around (-12/6) = -2; Z_4 already has a lot of points around -2, and hence it is not noticeable.

7.4 The Distribution of Z_5

As expected, Z_5 is an even distribution between 0 and 1. The points are evenly spaced by construction as each Z-score is a percentile. In many senses this is the "opposite" of Z_1 because it does not give special value to extreme data values.

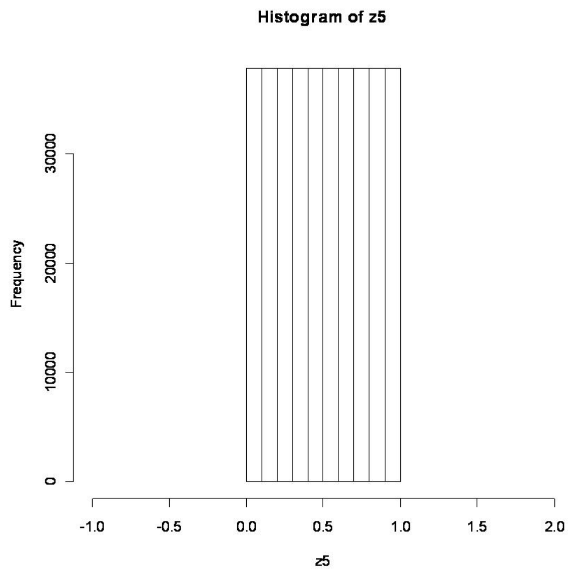

Figure 12: Histogram of Z_5

8 Ranking System Distributions Compared To Each Other

Three outliers were removed that had extraordinarily high z-scores for Z_1 to Z_3 in this analysis; three outliers in a data set of 385000 points should not affect results much.

Figure 13 of the various Z-scores against each other yields some very interesting results. The first is the three distinctive double-linear relationships amongst Z_1,Z_2, and Z_3. That is, all the data points appear to lie on one of two linear relationships between Z_1 and Z_2 or in the mass of points in the middle . The same is true of Z_1 vs. Z_2, Z_2 vs. Z_3, and Z_1 vs. Z_3. Z_4 and Z_5 differ the most from the first three metrics mathematically, and hence the more bizarre nature of their graphs seems to be a natural result.

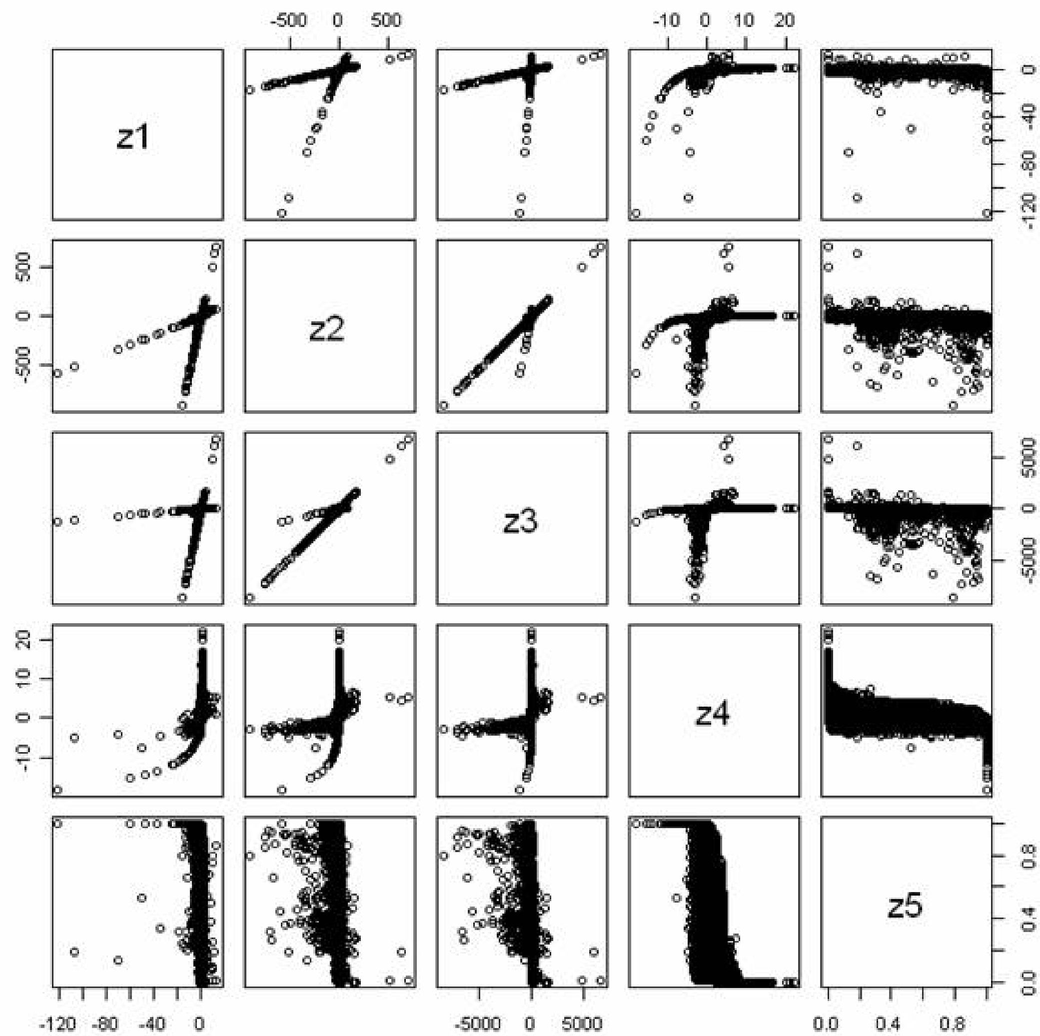

Figure 13: Distributions Plotted Against Each Other

8.1 Z_1 vs. Z_2

To get a better idea of the double linear relationship of the first three ranking system, we will look at an example. Figure 14 zooms in on the specific case of the relationship of Z_1 and Z_2. It is a clear that there is a double linear relationship; however at their intersection, where almost all of the 380,000 data points lie, the linearity breaks down.

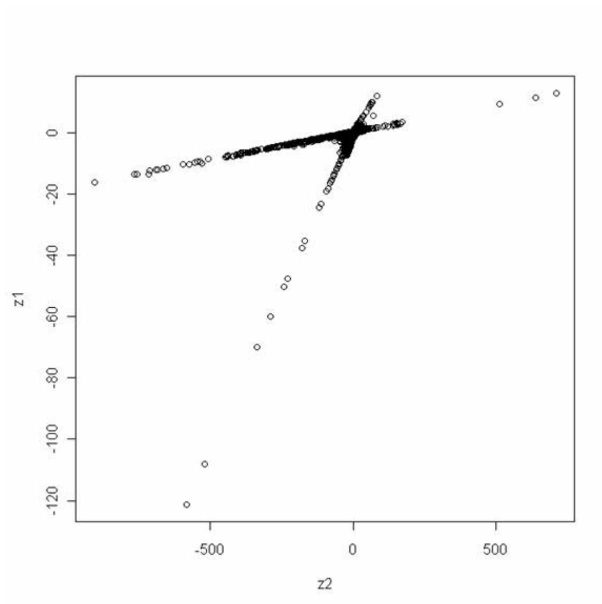

Figure 14: z1 vs. z2

To understand the double linearity, it is helpful to look at the factor components of Z_1 and Z_2, $PB.Z1$ and $PB.Z2$, in Figure 15. $PB.Z1$ and $PB.Z2$ represent the normalization of the PB factor for $Z1$ and $Z2$ before being linearly combined (by addition) with the other factors.

A perfectly linear relationship! Of course there must be a mathematical reason for such results; a simple proof of this linear relationship is shown below.

$$Z_{1,i,j} = \frac{X_i - \overline{X}}{S(X)} \forall i$$

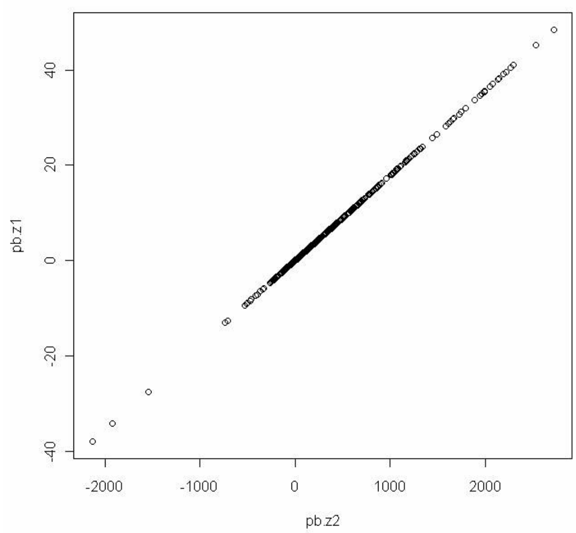

Figure 15: pb.z1 vs. pb.z2

$$Z_{2,i,j} = \frac{X_i - Median(X)}{Adj.MAD(X)}$$

$$X_i = \overline{X} + S(X) * Z_{1,j} = Adj.Mad(X) * Z_{2,j} + Median(X)$$

$$Z_{1,j} = (Median(X) - \overline{X}) + \frac{Adj.Mad(X)}{S(X)} * Z_{2,j}$$

Hence Z_1 is linearly related to Z_2 by the above equation — one would expect a graph of the two of them to be a straight line, which was revealed in Figure 15. The proof of this linear relationship, when using only one factor, helps us explain the relationships of different ranking systems to each other in Figure 13.

Firstly, it will not necessarily be true that because the individual factor Z-scores are linearly related that the sum of factor Z-scores will be linearly related. Indeed, assuming that the factors are not highly correlated, adding even two factor Z-scores together should result in a mass of points.

After pulling out some points from the ends of the two lines, it was found that the more horizontal of the two lines consists of points with a very large PB factor Z-scores (either negative or positive), and 0, N/A, or near 0 values for all of the other factor Z-scores. The large negative PB stocks are on the far right side of the line and the large positives are on the far left side of the line. The complete linearity between $PB.Z_1$ and $PB.Z_2$ we found in Figure 8.1 essentially "hijacks" the relationship of Z_1 and Z_2 for those points; the values for PB, and hence $PB.Z_1$ and $PB.Z_2$, are so large that the input from the other 5 factors is nil.

As an additional piece of evidence of this explanation, one should notice that as one goes along the line towards the mass of data points in the center on graph, the points jitter outwards from a straight line. The jittering, or noise, is the result of the other factor Z-scores pulling the data points off of the perfect line when they are non-zero but still small.

Having explained the horizontal line, we move to the more vertical line of the two. Upon closer inspection, these are actually two rays (half-lines) emanating from the center mass of points driven by different causes; the one going downwards contains points with incredibly large positive $M3M$ values, and the ray going up contains points with very large negative *pegh* values. The jittering and analysis for these linear relationships are the same as with PB. A natural question emerges: why does PB have a full line of points, whereas $M3M$ and *pegh* are only rays (half-lines)? $M3M$ cannot have negative values since it is the three month momentum; hence it has no large negative outliers that would "complete" the line. *pegh*, while it does have large positive values, they are not *as* large as the negatives by a factor of two. Given that the *pegh* ray is small to begin with on the negative side, its positive outliers are not large enough to see a distinguishable line emerging from the mass of points in the center.

Hence, outliers in the PB, *pegh*, and $M3M$ factors completely explain the interesting relationship between Z_1 and Z_2.

8.2 Z_1 and Z_2 versus Z_3

Building on Section 8.1, we now additionally consider Z_3. Z_1 vs. Z_3 has the exact same analysis as Z_1 vs. Z_2. The proof of the one component linear relationship is identical so won't be repeated, but the resulting equation is:

$$Z_{1,j} = (Trimmed(\overline{X_j}) - \overline{X_j}) + \frac{trimmed(S(X_j))}{S(X_j)} * Z_{3,j} \tag{8.1}$$

The PB line remains almost entirely unchanged. The rays for $M3M$ and *pegh* are slightly shifted.

The similarity between Z_1 vs. Z_3 and Z_1 vs. Z_2 is almost expected because Z_2 and Z_3 both attempt to trim down the normalization by giving less credence to outliers; Z_2 subtracts the MAD and divides by the median whereas Z_3 cuts out the ends (the bottom and top 10%) and still uses the mean and standard deviation.

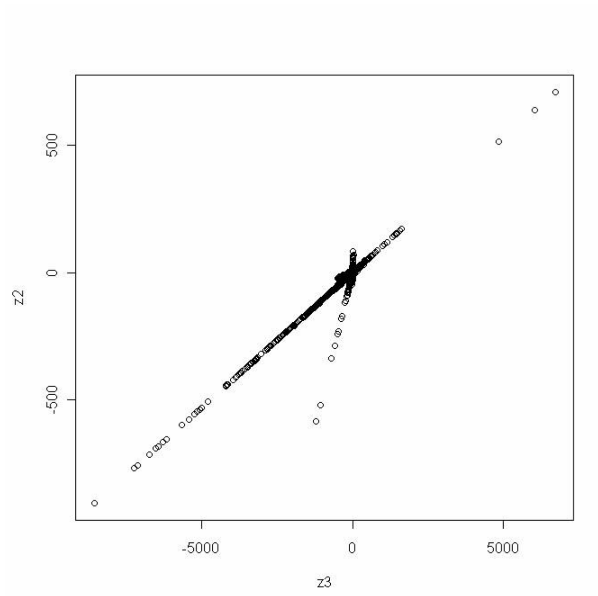

Figure 16: z2 vs. z3

The distribution for Z_2 vs. Z_3, visible in Figure 16, looks less random than the distribution for Z_1 because of the similarity between Z_2 and Z_3. However, the same linearity proof for one factor holds, resulting in Equation 8.2.

$$Z_{2,j} = (Trimmed(\overline{X_j}) - Median(X_j)) + \frac{trimmed(S(X_j))}{MAD(X_j)} * Z_{3,j} \qquad (8.2)$$

Again we see the same line from PB, and the rays (albeit less visible than previously) for $M3M$ and $pegh$.

8.3 Z_4 versus Z_1, Z_2, and Z_3

Having thoroughly explained the relationships evident in the figures of Z_1, Z_2, and Z_3, we now turn to the relationship of these three ranking systems with Z_4.

The main theoretical difference between Z_4 and the first three ranking systems is its use of a logarithm and its exclusion and then reinsertion of the first two percentiles for factors that contain negative points. All three graphs of Z_1,Z_2, and Z_3 vs. Z_4 look relatively similar, visible in Figure 13.

There are a total of 4 groups of data in the figure, and then there is the mass of data in the middle. Before breaking down these groups, let us analyze the factor distributions.

8.3.1 The All Positive Factors of Z_4: $M3M$, $Size$, and $EPSFN$

Let us deal with factors that are all positive first, like $M3M$. These factors don't get adjusted before the logarithm because they are entirely positive. That is $adjX_i = X_i$ because there is no mapping adjustment.

$$Z_{1,i,j} = \frac{X_i - \overline{X_j}}{S(X_j)}$$

$$Z_{4,i,j} = \frac{ln(adjX_i) - \overline{adjX_j}}{S(ln(adjX_j))}$$

Since $adjX_i = X_i$, we can use the same replacement in earlier proofs (i.e. set both equations equal to X_i and solve for Z_1). After algebra, this yields the following:

$$Z_{1,j} = \frac{1}{S(X_j)}(\overline{X_j}S(X_j)^{Z_{4,j}} - \overline{X_j}) \tag{8.3}$$

Unlike the relationships of our first three ranking systems with each other, this relationship is exponential. There is an exponential relationship for Z_2 and Z_3 with Z_4 as well(the algebra and equations for Z_2 and Z_3 vs. Z_4 are almost identical to the above and will not be reproduced here). This exponential relationship is shown in Figure 17.

8.3.2 The Other Factors of Z_4: PB, $RECC$, and $pegh$

These factors undergo a mapping whereby the bottom 2% is taken out, the 2nd percentile is added to the data, the natural logarithm of the adjusted

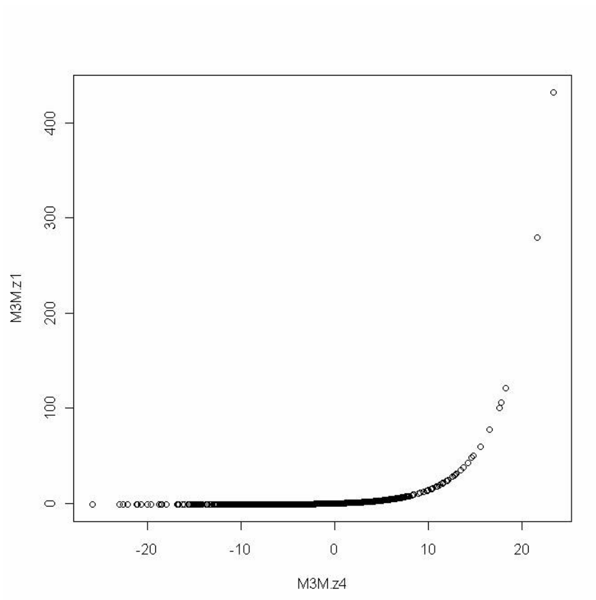

Figure 17: M3M.Z1 vs. M3M.Z4

data taken, and then it is normalized by subtracting the mean of the adjusted data and dividing by the standard deviation; finally the bottom 2% is mapped back in to the minimum factor Z-score. The result is that our data, our X_i, are slightly different than they would otherwise be. Call the adjustment to X_i a constant SP, where SP_j stands for value of the second percentile for the factor j (which is the value of the adjustment). Then $X_i + SP_j = adjX_i$; Z_1 and Z_4 are still given by:

$$Z_{1,i,j} = \frac{X_{i,j} - \overline{X_j}}{S(X_j)}$$

$$Z_{4,i,j} = \frac{ln(adjX_{i,j}) - \overline{adjX_j}}{S(ln(adjX_j))}$$

Then instead of $adjXi$ we substitute in $X_{i,j} + SP_j$ and solve for Z_1. The result is displayed in Equation 8.4.

$$Z_{1,j} = \frac{1}{S(X_j)}(\overline{X_j}S(X_j)^{Z_{4,j}} - \overline{X_j} - SP_j) \tag{8.4}$$

For the top 98% of the data, we should then still see roughly an exponential curve, like the positive factors. However, for the bottom 2 percent of the factor data, we will see a vertical line because all of the Z_4 data is equal to the minimum of its top 98% while the data for Z_1,Z_2, and Z_3 continues onward. Figure 18 shows the phenomenon in action. As expected there is an exponential curve with a vertical line at the end. Excellent!

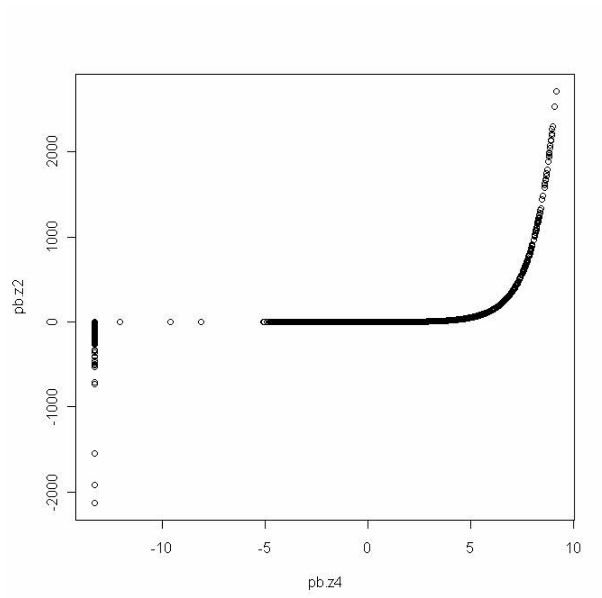

Figure 18: PB.Z2 vs. PB.Z4

8.3.3 Looking Again at Z_4 vs. the first 3 ranking systems

Now we look back at Figure 13. Note that they are relatively similar. We will look only at Z_2 versus Z_4 and generalize; the arguments explaining the graphs of Z_1 vs. Z_4 and Z_3 vs. Z_4 are the same.

Looking now at Figure 19, the horizontal line extending to the right from the center in all 3 graphs (Z_1, Z_2, Z_3 vs. Z_4) are stocks that have very small positive decimal values for $M3M$, but are close to the factor mean or N/A for the other 5 factors. Since the ln is taken in computing Z_4 and the ln of a positive value less than 1 is a large negative number, the small positive decimal values of $M3M$ become very large negative numbers in $M3M.Z_4$.

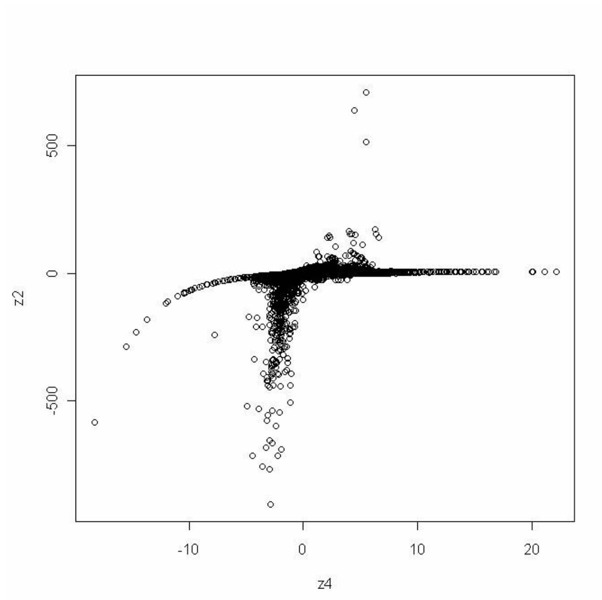

Figure 19: Z2 vs. Z4

These points also have NA values for the other factors. The exponential curve extending out of the mass of points horizontally to the left and then downwards are stocks that have very large values for $M3M$, and are NA for the other factors.

Similar to the PB line in the case of the first three ranking systems which jutted out of both ends of the mass of points in the middle, the $M3M$ exponential curve juts out horizontally of both ends of the middle mass in Figure 19. To the casual observer it looks like the graph of y = ln(x) with a bunch of noise in the middle of the graph. Because of the outliers in $M3M$, $M3M.Z_4$ ends up being very large, dominating the other factors (especially if the other factor have values close to the factor mean or NA). The reason it is "upside down", i.e. it looks logarithmic instead of exponential, is because of both the axis choice and that $M3M.Z_4$ has a negative coefficient when it is linearly combined with the other factor Z-scores. Also similar the jittering off of the PB line in the first three ranking systems, we see the example of jittering off of the $M3M$ exponential curve for Z_4.

Looking again at Figure 19, the group of points going directly downward toward the left-center of the graph, and the group of points going upward in the center of the graph, have very large positive and very large negative PB's, respectively. This is the faux exponential curve of PB that we discussed in the last section (refer to Figure 18)! The part going downwards is composed of the points with the highest PB's and is exponential in nature (the jittering makes this unintelligible however); the part going upwards is composed of the stocks with the lowest PB's, these got lopped off and mapped to the 2nd percentile.

Hence $M3M$ and PB explain all the deviations from the center mass of points in the graphs of Z_4 versus Z_1, Z_2, and Z_3.

8.4 Z_5 versus Z_1, Z_2, Z_3, and Z_4

Z_5, based on the percentile rank metric, has the most unique relationships with the other ranking systems.

The first thing to note is that it does a very good job of bringing outliers into its mass, which was done by design. All of the outliers from the first three ranking systems are brought in; most of them are not even near the 0 or 1 on Z_5.

Most noteworthy about this is the relationship between Z_4 and Z_5. A regression between these two variables would probably give a high level of significance, but an inverse relationship! That is, the highest Z_4 values are in the 0th percentile for Z_5, and visa versa. Given that both of these systems perform very well on our metrics, this is fascinating and should be looked at in future work.

9 Conclusion

The reason for this study, finding a better stock ranking system, has proven to be quite a bit more complicated than we had originally thought. That being said, we should present our results of testing our systems on our metrics.

9.1 Testing Results

Table 9.1 shows how each Z_α performed on the different metrics, by rank.

Testing Results				
Z_α	Returns Test	Percentile Test	Outliers Test	PB Test
Z_1	3	5	5	3
Z_2	4	4	4	2
Z_3	4	3	2	1
Z_4	1	2	2	4
Z_5	2	1	1	1

Even a cursory glance at Table 9.1 shows that Z_5 and Z_4 are at the head of the pack, Z_3 is towards the middle, and Z_2 and Z_1 come in last. However, more needs be said, because some of these test are very related and do not show things that are that different from each other.

The most important metric, from a practical standpoint (and statistics is an *applied* mathematics) is the returns test. Z_4 comes in a solid first in this test, with a t-statistic of 17, followed by Z_5, with a t-statistic of 12. Z_2 and Z_3 values are insignificant, and considering that they cannot explain 14 years of monthly stock returns, all but entirely rules them out as good systems from the start. Z_1 is definitely significant (t-statistic of 5), so it stays a contender.

The percentile rank metric, designed to test the problems of overreliance on outliers and overweight of certain factors that results from this, is our second most important metric. Our first and second place, Z_4 and Z_5 are the same, although in reverse order; it should be mentioned that the percentile ranking system, which is the same as the metric, was clearly going to get first on this. Surprisingly Z_4 again finishes at the top of the pack, at 2nd, far ahead of Z_1,Z_2, and Z_3.

In the outliers metric, we have the same first and second place, Z_5 followed by Z_4, as the percentile rank metric, however Z_3 performs as well as Z_4 in this case. Unsurprisingly, given the relative similarity of the Percentile Test and the Outliers Test, Z_1 again comes in last.

I struggle to consider the PB test by itself, as it was developed and intended to be one of many strange theoretical distribution metrics. In its creation it was expected that multiple other tests would be considered with it, and hence we could see how robust a ranking system was to strange theoretical distributions. Here we only have a sample size of 1 test; given this, it is worthwhile to note that Z_5 again came in first and Z_1 underreported.

To summarize, Z_5 comes as a clear winner amongst the ranking systems considered, with Z_4 as a clear second. Both metrics not only outperform

Z_1 handily in their significance predicting future stock returns, they also outperform heavily on our theoretical metrics. Although it could be argued that Z_4 might be better than Z_5 because of its slightly higher t-statistic in the regression, ultimately Z_5's simplicity and its first place finish in every other metric make it preferable. Further Z_5 is already used in many other applications; the combining of subsections in exam scores is the most obvious of these applications.

In the future Portfolio Managers would be wise to switch to a percentile ranking system (Z_5) to rank stocks (or even logarithmic Z_4), instead of the more complicated, less robust, and less effective Z_1.

Of course, a more complicated system is sometimes preferable when explaining investment decisions to clients...

9.2 Interesting Tangents

So many interesting things were found about the ranking systems and their relationships to each other, it is hard to conclude any single thing from them.

The first would be that the intricacies of Z_4 were examined; its strange nature and relationship with the other factors proved fascinating. While Z_1, Z_2, and Z_3 were easier to explain, it was useful to be able to explain the relationships between them and their distributions. Our analyses should prove useful in trying to come up with metrics and other ranking systems in the future.

9.3 Future Studies

It is impossible to have a flawless study, and flaws and incomplete areas are abound here. First, another ranking system should be tested, one that takes Z_1 as normally each factor, and then maps each factor Z-score above 5 or below -5 to 5 and -5. This may effectively cure the outliers problem more effectively than Z_5; however, given the number of systems tested and thought expended on the systems, we find a more effective Z-score than Z_5 to be highly suspect.

More strange theoretical distributions problems should be found, and metrics created to test these, to complement the PB problem. The PB test by itself was much less meaningful. Ultimately we want a ranking system that is robust to many different types of STDP's, not just the PB problem.

Additionally, the regressions done in this paper should be done as panel regressions as opposed to simple regressions, to account for the time variable. If inserted, the time variable (i.e. the month) would explain a lot about stock returns. It may also yield different statistical results of which factors and ranking systems are significant.

The relationship of Z_5 and the other systems would be useful to look at in the future, if only from a mathematical perspective. Further, the test statistics for the Outliers Test and PB Test should be reexamined, to ensure that the test statistic we used matches up with our qualitative concerns.

References

[1] Chincarini, Ludwig, and Kim, Daehwan. *Quantitative Equity Portfolio Management.* New York, NY: McGraw Hill, 2006.

[2] DeGroot, Morris and Mark Schervish. *Probability and Statistics. 3rd ed.* New York, NY: Addison Wesley, 2002.

[3] Hardin,Jo. *An R Tutorial.* Claremont,CA: 2008.

[4] R Core Team. *R Language Definition.* Version 2.5.1, 2007.

[5] Venables, W.N. and D.M. Smith. *An Introduction to R.* Version 2.5.1, 2006.