University of California
Los Angeles

Number of Stocks in Portfolio and Risk Reduction

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Yuan Zhou

2014

Abstract of the Thesis

Number of Stocks in Portfolio and Risk Reduction

by

Yuan Zhou

Master of Science in Statistics

University of California, Los Angeles, 2014

Professor Yingnian Wu, Chair

A lot of studies have been done on the optimal portfolio size. But not that many of them started by learning the specific relationship between the size of a portfolio and the reduction of the risk. This study looks into this specific relationship and tries to model it. The first part uses data from the up-to-date stock market on a smaller portfolio size. The second part extends the study by including more stocks in the portfolio, up to 1/5 of what is on the market. In the last part, a different metric of return and risk is used to test the conclusion. All the three parts basically agree with the same decreasing asymptotic relationship. The ideal size of portfolio coming from this study would be around 10.

The thesis of Yuan Zhou is approved.

Qing Zhou

Nicolas Christou

Yingnian Wu, Committee Chair

University of California, Los Angeles

2014

To my parents . . .

friends and the UCLA Statistics department

- all these could not have been done

without you . . .

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Since last century, a lot of studies in portfolio theory have been focusing solely on the optimal number of components in a portfolio. What people are trying to learn by studying the portfolio optimization is that how diversification is doing in reducing dispersion. Interestingly, there are not that much studies about the specific relationship between diversification and the reduction of dispersion. As I am researching on this topic, Evans and Archer's empirical studies (1968) [1] caught my attention.

In their studies, regression analysis and some statistical tests are performed to examine the real relationship between "the extent of portfolio diversification and the reduction in the variation (risk) associated with portfolio return"[1]. Some specific questions include the quantitative function about size and risk, the significance of successive increases in portfolio size and the significance on the convergence of the risks. The general idea reveals a whole new aspect on this topic and in my opinion, the study is worth updating. The data used in the study is some 50s-60s security data. Here I try to update the study with modern stock data, extend the study on portfolio size level and also examine the way Evans and Archers calculated the return and standard deviation of a portfolio.

Since we are trying to study the variation rather than the return of a portfolio here, it is important to learn about the nature of variation first. Mokkelbost's

[2] conclusion about the work of Sharpe [3], Lintner [4] and Hastie[5] on portfolio variation could serve as a perfect introduction about the concept of “risk”.

Risk could be measured by the standard deviation of portfolio return and it includes two elements: systematic risk and unsystematic risk. Systematic risk, also known as market risk, is the variation inherent to the entire market rather than a particular stock or industry. It is the “covariation of portfolio rate of return with market rate of return”[2]. Aggregate income, natural events like earthquakes, and world-wide financial crisis could serve as examples of systematic risk. Economists believe that systematic risk could neither be avoided nor mitigated through diversification. In contrast, unsystematic risk, also known as diversifiable risk, is the “difference between total portfolio variation and systematic variation”[2]. Only specific industries or companies are vulnerable to this kind of risk. It could be reduced through diversification.

Therefore, in trying to reduce the general risk of a portfolio, people are actually trying to mitigate the unsystematic risk. If a certain “optimal number” of portfolio size could be achieved, the variation reduced mainly comes from the unsystematic risk. As more and more stocks are added to the portfolio, the overall risks would approach a constant, which should be the estimated systematic risk. This paper examines the rate of the approaching as a function of the portfolio size.

CHAPTER 2

The Problem

In traditional portfolio theory, investors are trying to maximize the return as well as minimize the risk. If they could not reach an optimal situation for both of the factor, which happens in most of the cases, they would optimize the level of one of the factor for a given level of the other. Another factor we should take into account should be the cost of the portfolio. If the cost is a function of the number of shares held, then the traditional portfolio selection theory would satisfy the need. But if the cost associated with the portfolio size, the traditional theory would not be working. Therefore, the marginal analysis of the risk is necessary.

By applying marginal analysis of the variation, it would be easier to choose among portfolios with identical return and risk. What's more, we could even compare portfolios with the same return in a broader aspect. Consider, for example, two portfolios with the same return, but different risks and portfolio sizes. Traditional portfolio theory would lead to the selection of the portfolio with lower risk. But what if the one with lower risk contains more stocks? The additional stocks might cause a higher marginal cost and this is when investors might need marginal analysis of portfolio risks.

The following analysis would focus on the marginal reduction in the portfolio standard deviation as a function of portfolio size. In the portfolio selection process, the assumption are the same as what Evans and Archer[1] have assumed in their study: "(1) The investor is a random buyer of common stocks; and (2) equal dollar amounts are invested in each security/stock in the portfolio".

CHAPTER 3

Data

The data used in the original study was a pool of 470 securities in the Standard and Poor's Index for the year 1958. Semi-annual observations for all the 470 securities were taken from January 1958 to July 1967 and hence there are 19 observations for each security.

Since in this study I am trying to reproduce the result of the paper with some up-to-date data from the stock market, I decide to use the components of Standard and Poor 500 Index (on January 2014) as my pool. Stock prices vary a lot throughout the year and they are comparatively easy to collect, so I am taking the monthly price of all these 500 companies from 2008-12-31 to 2013-12-31 as my observation.

However, initial exploration of the stock price indicates that not all of the 500 companies have completed data for the 5 years. Since the components of the S&P 500 Index keeps changing, some of the companies are fairly new and only have a few stock price records. The companies with completed stock price data for the 5 years are picked out as our pool, where there are 60 data points in total for one stock. There are 473 companies got selected. Appendix A contains a table of all these companies.

CHAPTER 4

Porfolio Diversification with Size up to 40

4.1 Return for Each Stock

The monthly return for period i was calculated in the following formula. "i" here is the indicator for time period, so it takes the integers from 1 to 60. "k" here is an indicator for the company name, so P_i^k is the price of the k^{th} stock at the beginning of period i and P_{i+1}^k is the price of the k^{th} stock at the end of period i. Here R_i^k is the computed return for stock k in period i.

$$R_i^k = \frac{P_{i+1}^k}{P_i^k}, \text{ for } i = 1 \text{ to } 60;\ k = 1 \text{ to } 473$$

4.2 Return and Standard Deviation for each portfolio

For a portfolio with m stocks, the return and standard deviation are calculated in a geometric way. First, the return for each period i is computed from the following:

$$\bar{R}_i = \frac{\sum_{k=1}^{m} R_i^k}{m}, \text{ for } m = 1 \text{ to } 40.$$

Second, the portfolio geometric mean return ($\bar{R}_p$) for the five whole years (2009-2013) is computed:

$$\bar{R}_p = exp(\frac{\sum_{i=1}^{n} \log_e \bar{R}_i}{n}), \text{ for } n = 60.$$

Finally, the portfolio standard deviation (SD_p) is computed:

$$SD_p = \sqrt{\frac{\sum_{i=1}^{n}(\log_e R_p - \log_e \bar{R}_i)^2}{(n-1)}}, \text{ for } n = 60 \ .$$

4.3 Portfolios Selection

In order to investigate the relationship between portfolio size and diversification, I randomly select stocks in the pool to generate portfolios from size 1 to 40.

First, one stock is drawn at random from the 473 stocks and the corresponding portfolio geometric mean return ($\bar{R}_p$) and portfolio standard deviation (SD_p) are recorded. Second, two stocks are chosen from the pool at random. After the return and standard deviation of the portfolio with these two stocks are marked down, the process is repeated for 3, 4, ... , 40 stocks. Hence, the runs generate 40 portfolios with size ranging from 1 to 40

I have it run for 60 times and 60 observations for each portfolio size are generated. In total, there are 2400 (40×60) portfolios. Table 4.1 is the example output from one computer run. It contains the tickers of stocks selected randomly, the corresponding portfolio geometric mean return ($\bar{R}_p$) and portfolio standard deviation (SD_p).

For example, the first portfolio is a portfolio including only the COP, which, according to appendix 8.1, is the ConocoPhillips. The portfolio return is 1.0052 and portfolio Standard Deviation is 0.0796. The second portfolio has size 2 and the components are EXC (Exelon Corporation) and COH (Coach, Inc.). The portfolio return is 1.0042 and standard deviation is 0.0642. We could tell the second portfolio has lower expected return but performs better in terms of risk.

		1	2	3	4	5	...	40
Stocks selected	1	COP	EXC	PCL	AIG	ARG	...	WEC
	2		COH	PFG	D	BLK	...	WEC
	3			IP	LM	STZ	...	FLR
	4				SJM	TAP	...	WMT
	5					MDT	...	OI
	...							
	40						...	GPS
$\bar{R}_p$		1.0052	1.0042	1.0160	1.0422	1.0157	...	1.0168
SD_p		0.0796	0.0642	0.1132	0.1945	0.0529	...	0.0579

Table 4.1: Result from one computer run

Here is a scatter plot about the portfolio risk of all the portfolios generated from the previous process.

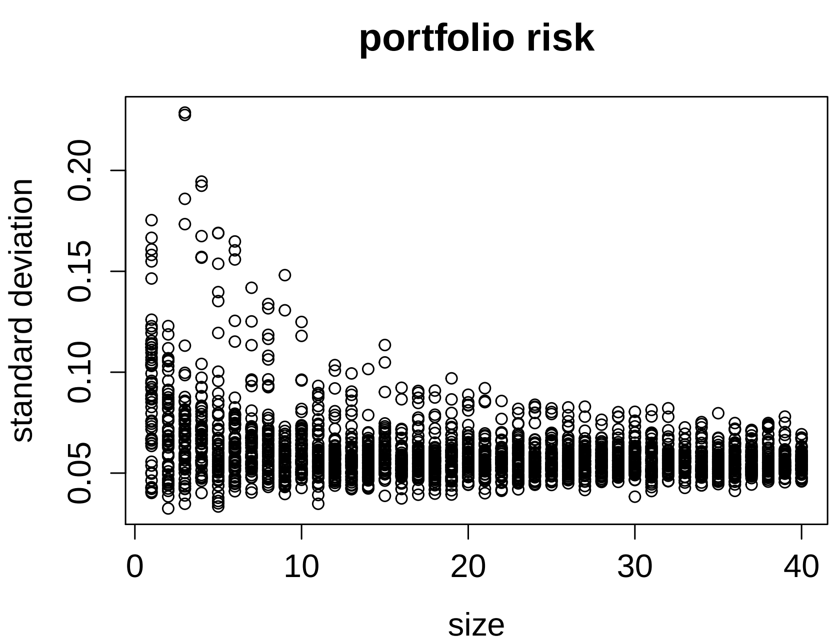

Figure 4.1: Scatter plot of portfolio risk

Figure 4.1 shows that as the portfolio size is getting larger, the variation of the standard deviations is getting smaller. And the mean of them seems to be asymptotic.

4.4 Regression

In order to examine the relationship between size and risk, I constructed a regression model

$$y = b(1/x) + a,$$

where y is the mean portfolio standard deviation (SD_p) and x is the size. I fit this model with the 40 means of portfolio standard deviation with different sizes generated from section 4.3 and it turns out to be a good fit. The R^2, also known as coefficient of determination, is 0.8549 and the fitted model is the following:

$$\hat{y} = 0.0407(1/\hat{x}) + 0.0562.$$

Both coefficients are extremely significant, indicated by a p-value less than $2e^{-16}$ for both of them. The coefficient of determination is not as good as that in the article, which is 0.9863, but 0.8549 is a fair enough R^2 for linear model. This is also supported by the fitted plot of the model.

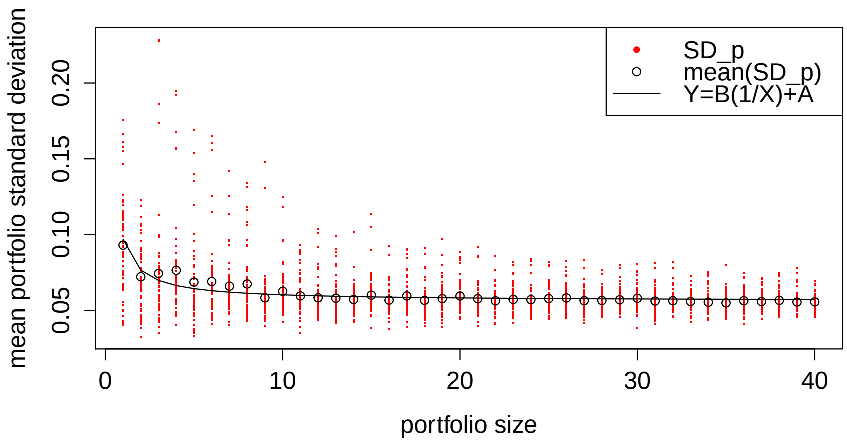

Figure 4.2: Scatter and fitted line for the model

In Figure 4.2, we could see the line is a pretty good fit and it is also an asymptote. As the size of the portfolio is growing, the mean portfolio standard deviation is approximating to a certain number, which is believed to be the systematic variation in the market. If we include all the 473 stocks in our portfolio and calculated the portfolio standard deviation (SD_p), we could get the estimated systematic variation, which is 0.0530. The result is very similar to what was got from the 1950s security data.

4.5 t-test

The unsystematic variation is believed to be attributable to individual stocks. In order to analyze the unsystematic variation, some statistical tests are necessary.

The t-test is applied on successive mean portfolio standard deviation. By now we have already got the 60 different portfolios for each size in section 4.3, which means, we have 60 data points for the portfolio risks in each of the size. The t-test is performed on these successive 40 groups to see whether there is a significant difference between the mean of the standard deviation. The main purpose of the t-test is to detect how many more stocks we need in the portfolio to drop the standard deviation significantly at 5% level. Here is the null and alternative hypothesis. Note that μ_1 is the mean of the standard deviations from the smaller size portfolios and μ_2 is the mean from the larger ones.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

First, I take the 60 data points coming from portfolios with size 1, which are SD_{1_1}, SD_{1_2}, SD_{1_3}, ... $SD_{1_{60}}$, and perform a one-side t-test with the 60 data points coming from SD_2, which are SD_{2_1}, SD_{2_2}, SD_{2_3}, ... $SD_{2_{60}}$. The test is used to see whether there is a significant difference in the mean of these two data sets. Since the portfolios of larger size are supposed to have lower risk, a significant result of the t-test means that the portfolio with larger size has a smaller mean of the standard deviation. The p-value resulting is 4.6671e-05, which means that adding one stock on the portfolio with size 1 would significantly decrease the standard deviation in the portfolio. Then I repeat the same procedure starting with portfolio of size 2. The p-value of the t-test is pretty big until 5 stocks are added in the portfolio. The p-value of the t-test after the addition of 5 stocks is 0.0471, indicating that the ideal significant decrease in the standard deviation of portfolio with size 2 is achieved and the examination of portfolio with size 2 is finished. Third, do the t-test again starting with size 3, 4, ... 39 until the number of additional stocks making a significant drop in the the standard deviation is found. In the t-test, the significant reduction outside the range of 40 stocks is not considered. And here is a figure showing the result of the t-test.

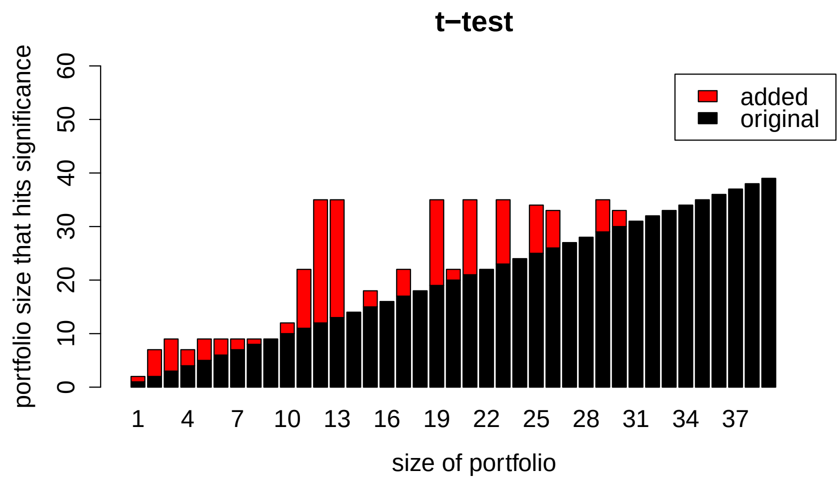

Figure 4.3: Bar plot for the result from the t-test

Figure 4.3 is a bar plot indicating how many additional stocks are needed to generate a significant difference in the mean of standard deviations. The black part of the bars represents the original size, while the red bars represents the additional size. For those sizes only with black bars, the number of additional stocks would make the total size exceed 40, and those cases are not considered here. Table 4.2 is a summary of the additional size for both the t-test and F-test. As we could tell from figure 4.2, it is more difficult to decrease the portfolio standard deviations significantly after size 9. The mean of the portfolio standard deviation with size 9 is 0.0583 and the systematic variation is believed to be 0.0530 according to section 4.4, which means the unsystematic variations is largely eliminated by the time the 9^{th} stock is added to the portfolio. The result is also supported by Figure 4.1, the scatter plot and fitted line of the model.

4.6 F-test

For the F-test, the idea is similar. The goal of this test is to examine the convergence of each observations on the mean value in standard deviation. For each size of portfolio, the F-test is performed on successive portfolio standard deviation to figure out how many more stocks are needed reduce the standard deviation of the observed $\{SD_p\}_{60}$ significantly at 5% level.

Here is the null and alternative hypothesis. Note that σ_1 is the standard deviation of the standard deviations from the smaller size portfolios and σ_2 is the standard deviation from the larger ones.

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 > \sigma_2^2$

First, I take the 60 data points coming from portfolios with size 1, which are SD_{1_1}, SD_{1_2}, SD_{1_3}, ... $SD_{1_{60}}$, and perform a one-side F-test with the 60 data points coming from SD_2, which are SD_{2_1}, SD_{2_2}, SD_{2_3}, ... $SD_{2_{60}}$. The test is used to see whether there is a significant difference in the standard deviation of these two data sets. Since the portfolios of larger size are supposed to have more convergent standard deviation, a significant result of the F-test means that the portfolio with larger size has more convergent risks. The p-value resulting is 0.0011, which means that adding one stock on the portfolio with size 1 would significantly decrease the standard deviation in the portfolio risk. Then I repeat the same procedure starting with portfolio of size 2. The p-value of the F-test is pretty big until 7 stocks are added in the portfolio. The p-value of the t-test after the addition of 7 stocks is 0.0275, indicating that the ideal significant decrease in the standard deviation of portfolio risk with size 2 is achieved and the examination of portfolio with size 2 is finished. Third, do the F-test again starting with size 3, 4, ... 39 until the number of additional stocks making a significant drop in the standard deviation of portfolio risk is found. Again, the significant reduction outside the range of 40 stocks is not recorded here.

The result roughly agrees with what I get from the t-test. Here is a figure showing the output from the F-test.

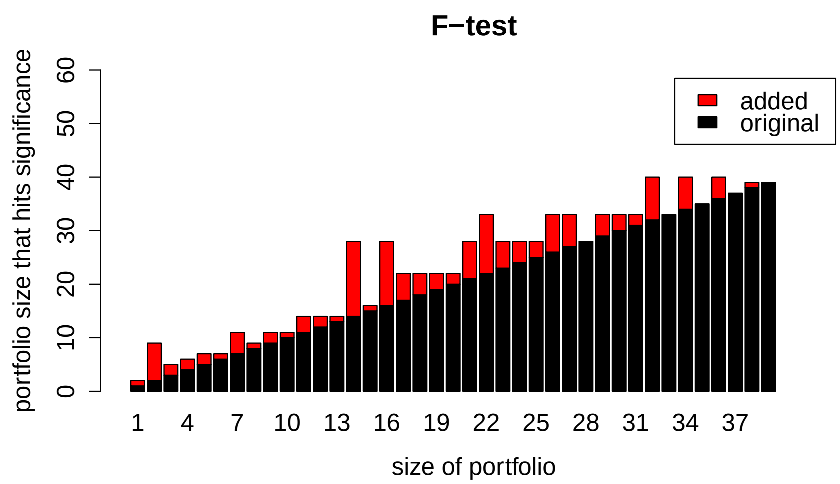

Figure 4.4: Bar plot for the result from the F-test

Figure 4.3 gives a similar conclusion to the output of the t-test. We could see that it is getting less easier to make a significant difference in the standard deviation for the portfolios with size larger than 10. However, within the range of 40 stocks, the addition of stocks tends to reduce the standard deviation of portfolio risks in more size cases than the result from the t-test. There are 22 cases in the t-test where adding stocks within the range of size 40 would significantly decrease the portfolio standard deviations. However, the number of available cases in the F-test is 34. The convergence of the individual observations on the mean values seems to be less of an issue than significance of successive increase in portfolio size.

Here is a completed table of the outputs from the two tests.

size	significant size(t-test)	significant size(F-test)
1	2.00	2.00
2	7.00	9.00
3	9.00	5.00
4	7.00	6.00
5	9.00	7.00
6	9.00	7.00
7	9.00	11.00
8	9.00	9.00
9	NA	11.00
10	12.00	11.00
11	22.00	14.00
12	35.00	14.00
13	35.00	14.00
14	NA	28.00
15	18.00	16.00
16	NA	28.00
17	22.00	22.00
18	NA	22.00
19	35.00	22.00
20	22.00	22.00
21	35.00	28.00
22	NA	33.00
23	35.00	28.00
24	NA	28.00
25	34.00	28.00
26	34.00	33.00

27	NA	33.00
28	NA	NA
29	35.00	33.00
30	33.00	33.00
31	NA	33.00
32	NA	40.00
33	NA	NA
34	NA	40.00
35	NA	NA
36	NA	40.00
37	NA	NA
38	NA	39.00
39	NA	NA

Table 4.2: Additional stocks needed for the t-test and F-test

Table 4.2 gives a clear description about the output of both tests. For the portfolios with size less than 10 stocks, results from both tests are similar. This may due to the decrease of unsystematic variation. For size less than 10, the risk reduced by the increase in portfolio size mainly comes from unsystematic variation. Therefore, after a certain point where the unsystematic variation is largely cut down, it is hard to decrease the mean or standard deviation of portfolio sizes by just adding stocks.

4.7 Comparison

Basically, what I get from the analysis with modern stock data agrees with the result of the analysis with 50s-60s security data. The two fitted models are the following:

for the model with 50s-60s security data: $\hat{y} = 0.08625(1/\hat{x}) + 0.1191$

for the model with modern stock data: $\hat{y} = 0.0407(1/\hat{x}) + 0.0562$

Here is a comparison of the two lines.

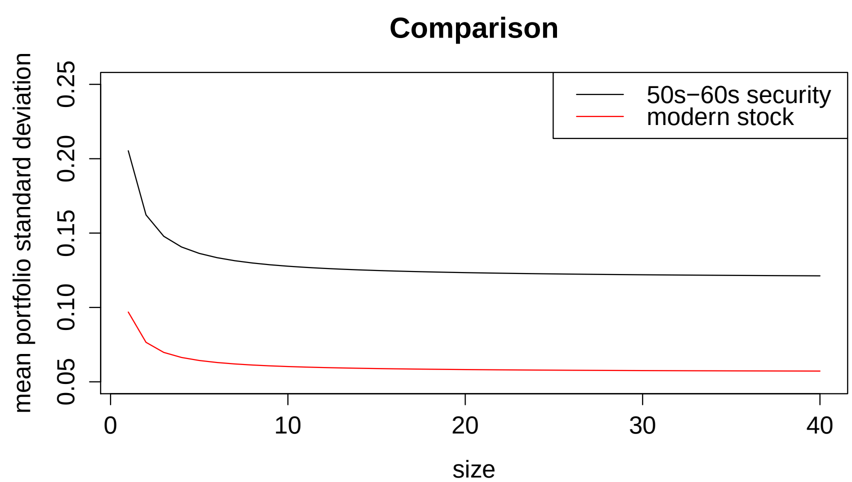

Figure 4.5: Comparison of modern and the 50s-60s models

Figure 4.5 shows that the models generated from the two sets of data are pretty similar in shape and almost parallel. The gap might come from the natural difference between stocks and securities. In the calculation of return in securities, the dividends are taken into consideration. Also, the goodness of fit for both models is over 0.85, which is a high enough coefficient of determination. We could tell from both curves, the systematic variation of portfolios are largely eliminated after the 8^{th} or 9^{th} security/stock is added to the portfolio.

CHAPTER 5

Portfolio Diversification with Different Sizes

5.1 Introduction

Intense discussion on the optimal number of portfolio size has been raised since the 20^{th} century. Evans and Archers [1] proved that the optimal number would be no more than 10. Gup [7] believed that “the diversifiable risk is reduced as the number of stocks increases from one to about eight or nine”. Reilly [8] took the number to 12-18. Statman[6] challenged the previous statements by raising that the optimal number should be at least 30-40.

Since there is hardly any agreement on the number of portfolio size that would derive the most of the benefits of diversification, an further examination on portfolio size of the study in Chapter 4 is necessary. Here I would continue the study by extending portfolio size up to 100 and compare the results with what we have got from the previous section.

Note that in this section the random selection process for portfolios is still the same as section 4.3. It is just this time I am doing it from size 41 to 100. Here is a scatter plot about all the portfolios resulting (including what we have got from Chapter 4).

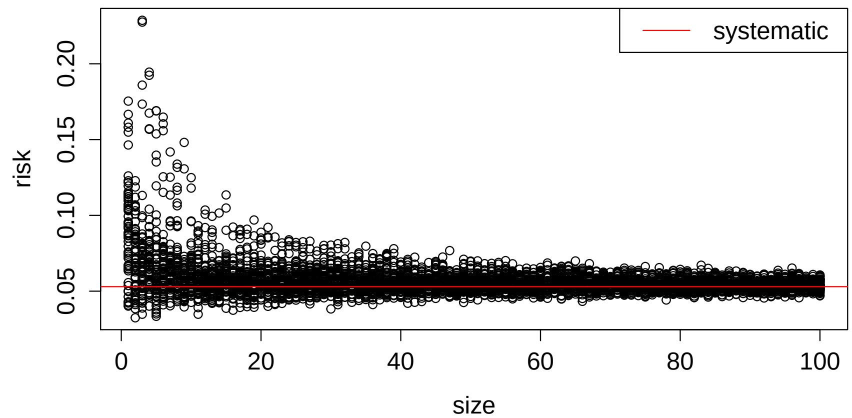

Figure 5.1: Comparison of modern and the 50s-60s models

Figure 5.1 shows all the standard deviations from portfolio size 1 to size 100. In this chapter, the focus is the standard deviation for portfolio from size 41 to size 100. We could see that the range of standard deviation for portfolio with size 41 to 100 is mainly around [0.047, 0.064]. Note that in section 4.4 the estimated systematic variation is 0.0530 (red line). As is mentioned in section 4.5, much of the unsystematic variation is eliminated by the time the 9^{th} stock is added to the portfolio, which is to say, the estimated standard deviation for the portfolio with size 41-100 would be close to the systematic variation 0.0530. What is seen on the scatter plot agrees with this conclusion. As for the convergence of the risk of portfolio with larger size, portfolio with larger size do have smaller standard deviation on those risks.

5.2 Regression

Figure 5.2 shows the trend on the mean of portfolio standard deviations as portfolio size is going up. The curvature resembles the model fitted in chapter 4 a lot. Therefore, if a regression model would be fitted to all these data points, a similar one to the $y = b(1/x) + a$ model would be taken into serious consideration.

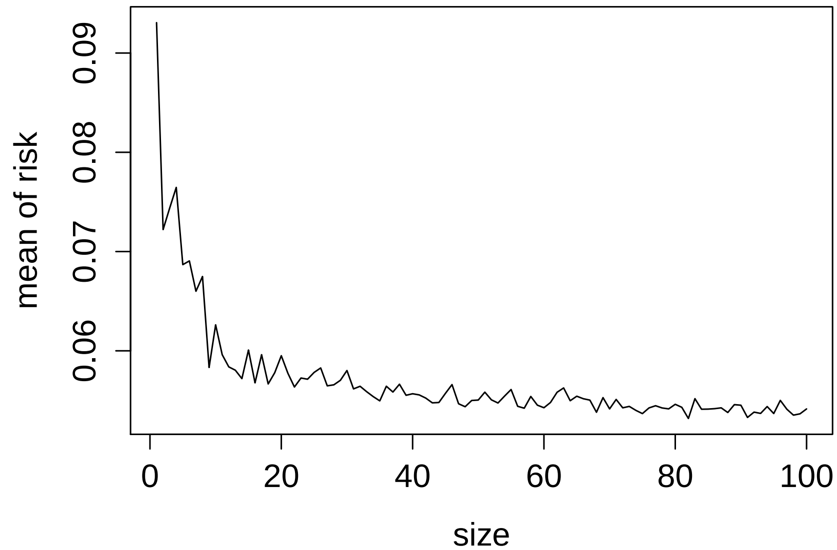

Figure 5.2: the Mean of Portfolio Risk

This time the model fitting procedure is applied to all the mean standard deviation data generated so far, from size 1 to size 100. As is mentioned in previous section, for each size, there are 60 data points and in order to make the model a better fit, the mean of all the standard deviations coming from the same portfolio size is taken out as the response in our model. The only factor is the portfolio size. And here is what is got out of the regression.

$$\hat{y} = 0.0444(1/\hat{x}) + 0.0547$$

The regression seems to be a good fit with the $R^2 = 0.8557$. And the R^2 of the regression model with size up to 40 is 0.8549. Again, Both coefficients are extremely significant, indicated by a p-value less than $2e - 16$. Basically, what is got out of the regression analysis with all the 100 data points is very close to the output in the previous regression analysis with the first 40 data points.

Here is some visualized comparison of the two model.

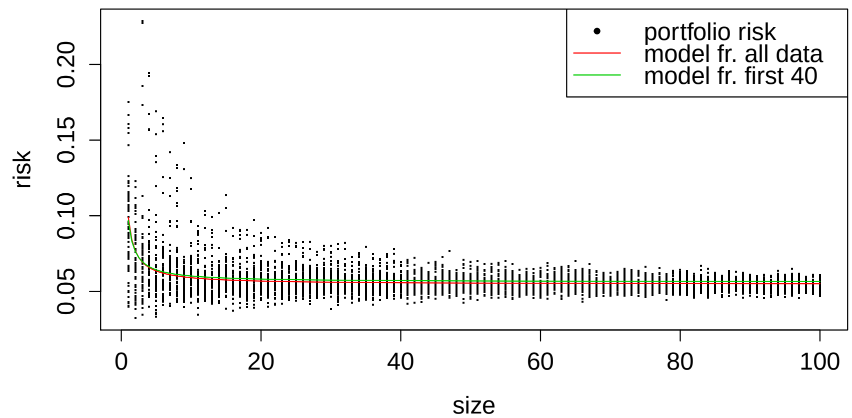

Figure 5.3: Models Comparison

Figure 5.3 shows how close the two models are. They are both good fit for the mean portfolio standard deviations. The predicted risk from the first model for sizes larger than 40 are pretty decent, demonstrated by the green line for size larger than 40. Both lines are asymptotic and approximating to a constant as the portfolio size is growing. It seems that the theory that portfolio risk could not be reduced more after a certain size is supported here. Also, we could tell the predicted standard deviation for portfolio with an even larger size than 100 would still be really close the constant here. What's more, the sharp drop for both lines for size ranging from 1 to 10 is similar. It verifies the conclusion from the previous section that after the 9^{th} stock most of the unsystematic variation is eliminated and what is left is mainly the systematic variation.

5.3 t-test and F-test

Just like what has been done in Chapter 4, t-tests and F-tests are performed to study how the unsystematic variation was reduced. How the tests are performed and the null hypotheses are basically the same. The main difference is that the range of portfolio sizes is extended to [1, 100] instead of [1, 40].

The t-tests are performed on successive mean portfolio standard deviations, trying to learn how many more stocks are needed to drop the risks significantly.

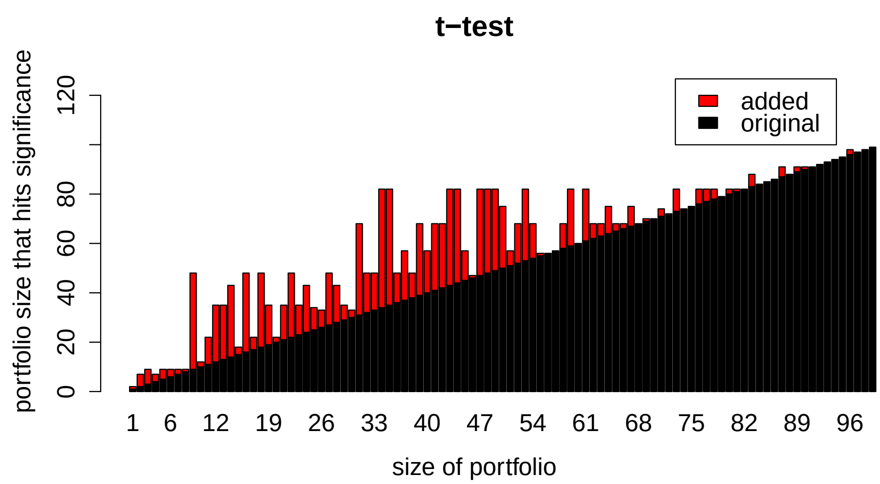

Figure 5.4: Bar plot for the result from the t-test

Just as Figure 4.3, Figure 5.4 is a bar plot for the result of the updated t-tests. The "added" red bar indicates how many more stocks are needed to produce a significant drop in the mean of standard deviation. This time, the upper limit for the size is 100. However, the conclusion about the optimal portfolio size remains unchanged. As we could see from Figure 5.4, the red bar tends to be longer after size 9, which is to say, it is more difficult to decrease the portfolio standard deviations significantly after the 9^{th} stock is added to the portfolio. Noted that the mean of portfolio standard deviation of the portfolios with size 9 is 0.0583 and the systematic variation is estimated to be 0.0583. The unsystematic variation is largely reduced by then.

The F-test are performed on successive standard deviation about the mean portfolio standard deviation, indicating the significance of the convergence of the standard deviations when portfolio sizes are increasing.

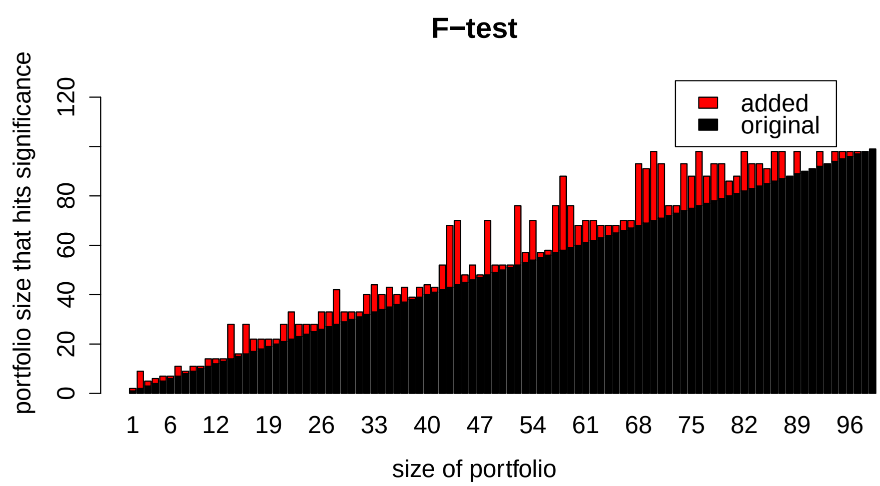

Figure 5.5: Bar plot for the result from the F-test

Just as Figure 4.4, Figure 5.5 is a bar plot for the results coming from the extended F-tests. It is not as obvious as the figure from the t-test, but we do see some similar conclusions here. The red bar tends to be longer after the 9^{th} stock is added. That is to say, achieving the convergence of the standard deviations by enlarging the portfolio size is getting harder after size 9.

There is a completed table about the outputs for both tests in the appendix (table 8.2). The significant sizes for both the t-test and F-test are similar and comparatively smaller for portfolio size smaller than 9, which might tell us that within the major range for the unsystematic variation, it is easier to reduce the risks as well as to achieve significant convergence in risks by increasing portfolio size. After size 9, which means after the major part of unsystematic variation is reduced, the results coming from both tests are not on the same page that much.

5.4 Conclusion

Basically, what is got from the extended study reinforces the conclusions before. The decreasing asymptotic relationship stays even for larger portfolio size. It is easier to reduce risk by increasing portfolio size for smaller portfolio size. And after the unsystematic risk is largely reduced, the remaining risk mainly comes from the market and hence is more difficult to reduce. Also, the optimal portfolio size is still no more than 10, since increasing portfolio size after that does not make a significant difference in the reduction of the risk.

CHAPTER 6

Examining the Calculation of Return and Standard Deviation

6.1 Introduction

It might be noted that the calculations for both returns and standard deviations are somewhat different from the traditional formula. It was the calculation from Evans and Archer's empirical studies (1968) [1]. They used the metric for various reason. Some involves the nature of securities, which were the data used in that study. And the metric has its own good when it comes to the capture of change and the measurement of risk.

However, here in the chapter, the traditional way to calculation returns and risks is used to see whether the metric makes a huge difference in the study, which is the arithmetic mean and the arithmetic standard deviation. If the change of metric would jeopardize the conclusion, the conclusion is not stable enough.

Noted that they are calculated from the following formulas (same notation as in Chapter 4):

$$\bar{R}_p = \frac{\sum_{i=1}^{n} \log_e \bar{R}_i}{n}, \text{ for } n = 60.$$

$$SD_p = \sqrt{\frac{\sum_{i=1}^{n} (R_p - \bar{R}_i)^2}{(n-1)}}, \text{ for } n = 60 \text{ .}$$

6.2 Portfolio Selection

In Chapter 4, 60 portfolios are randomly drawn for each size from 1 to 40, which made a totally of 2400 data points in our dataset. The goal of the study in this chapter is to compare the metrics, and hence the same set of portfolio selections as that in chapter 4 is used here. That is to say, the stocks selected in each portfolio are exactly the same in these two chapters and it is only how the returns and risks are calculated that differs. Here is some examples of the comparison.

		1	2	3	4	5	...	40
Stocks selected	1	COP	EXC	PCL	AIG	ARG	...	WEC
	2		COH	PFG	D	BLK	...	WEC
	3			IP	LM	STZ	...	FLR
	4				SJM	TAP	...	WMT
	5					MDT	...	OI
	...							
	40						...	GPS
$\bar{R}_{p_{geometric}}$		1.0052	1.0042	1.0160	1.0422	1.0157	...	1.0168
$SD_{p_{geometric}}$		0.0796	0.0642	0.1132	0.1945	0.0529	...	0.0579
$\bar{R}_{p_{arithmetic}}$		0.0155	0.0089	0.0251	0.0272	0.0186	...	0.0201
$SD_{p_{arithmetic}}$		0.0680	0.0652	0.1200	0.1172	0.0532	...	0.0573

Table 6.1: Result from one computer run

Table 6.1 records the returns and risks from the two metrics for one computer run. For the return, it seems that portfolio with a higher return under the geometric metric still has a higher return under the arithmetic metric. However, the same rule does not apply to portfolio standard deviation. Some of the portfolios have comparatively higher risk under one metric but have lower risk in the other. The comparison of the 3^{rd} and 4^{th} portfolio in the table could serve as an example.

Under the geometric calculation, the portfolio of size 3 here (PCL, PFG, IP) has a smaller standard deviation than the portfolio of size 4 here (AIG, D, LM, SJM). But when the standard deviation is calculated in the arithmetic way, the portfolio of "PCL, PFG and IP" does have a larger standard deviation. Compared to the 4^{th} portfolio, the 3^{rd} portfolio performs worse under the geometric metric in terms of risk but better under the arithmetic metric. This might be due to the natural difference in the two calculations. The following scatter plot might gives a clearer understanding of the arithmetic metric.

Here is a scatter plot about the portfolio risk of all the portfolios.

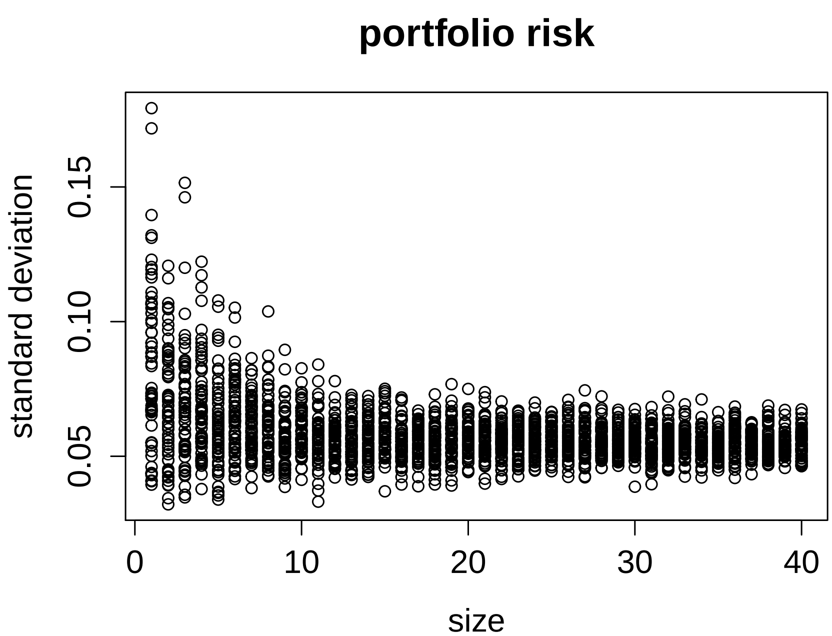

Figure 6.1: Scatter plot of portfolio risk

Figure 4.1 and Figure 6.1 share some really similar trends for the standard deviations. The variation of the standard deviations is getting smaller as portfolio size increases. And after size 10, the average level of the variation seems to remain the same. Both Table 6.1 and Figure 6.1 resemble the cases in Chapter 4 a lot, which might be indicating that the following results would also be close. If it is true, then the conclusion from chapter 4 would be strongly reinforced.

6.3 Regression

The same regression model is fitted:

$$y = b(1/x) + a,$$

where y is the mean portfolio standard deviation (SD_p) and x is the size. And here is the result:

$$\hat{y} = 0.0328(1/\hat{x}) + 0.0546,$$

This is a even better fit than the previous models. The coefficient of determination is 0.9076, and both the model and coefficients are extremely significant. All these numbers are implying that this is a good model fitting. And this is also supported by the plot here.

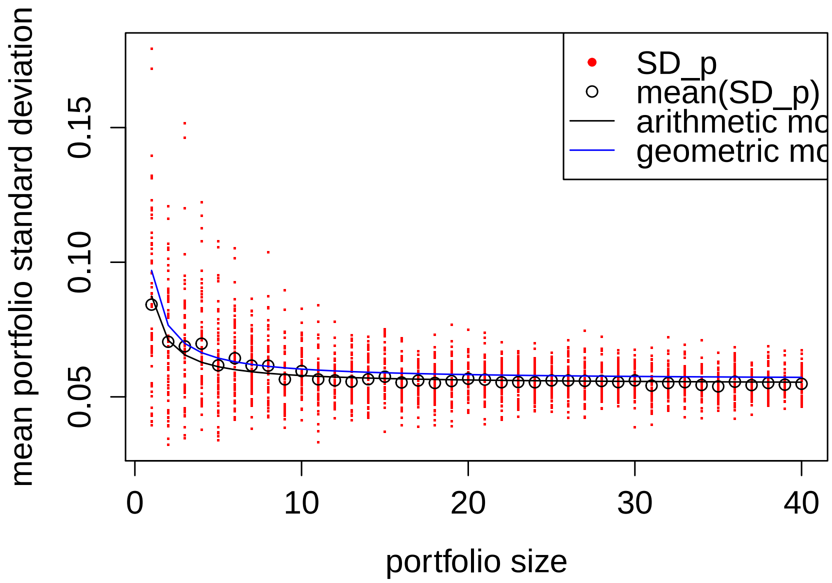

Figure 6.2: Scatter and fitted line for the model

In Figure 6.2, we could see the line is a pretty good fit and it is also an asymptote, just as the way it is in Figure 4.2. As the size is increasing, the mean portfolio standard deviation is approximating to what is believed to be the systematic variation. Under this the arithmetic metric for returns and risks here, the systematic variation is calculated by including all the 473 stocks in the portfolio and it is 0.0536. Noted that the systematic variation from the other metric is 0.0530. They are pretty close and the difference might come from the nature of both metrics.

Another interesting fact is that lines coming from both models are almost parallel. Also, with a higher estimated systematic variation, the black line is surprisingly lower all the time. the range of the mean of portfolio standard deviations for size 31 to 40 in the arithmetic model is around [0.054, 0.056]. It seems that the risk calculated from the traditional way is closer to what is believed to be the market risk. But again, this does not imply any superiority in metric choice and might just due to the nature of the metric.

6.4 t-test and F-test

Once again, t-test and F-test are performed to study how the unsystematic variation was reduced. The purpose of the test and the null hypothesis stay the same as previous chapters. It is just this time, the tests are performed on data calculated from the arithmetic mean and standard deviation.

The t-tests are performed on successive mean portfolio standard deviations to study the significance of successive increase in portfolio sizes.

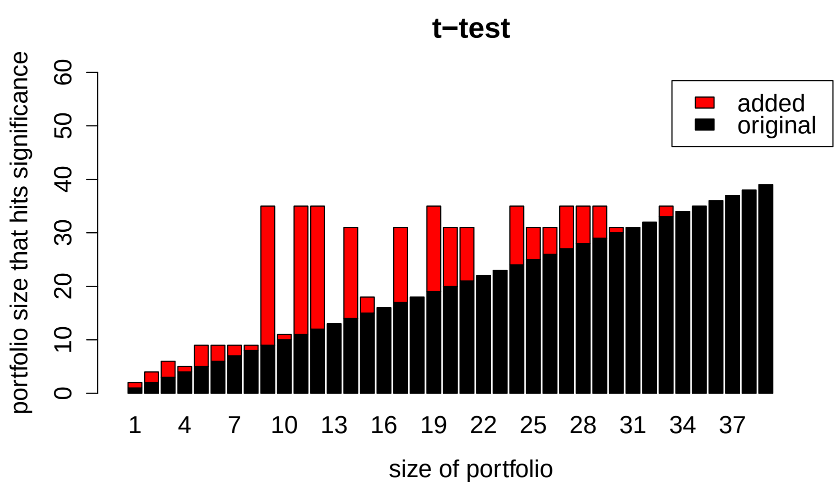

Figure 6.3: Bar plot for the result from the t-test

Just as Figure 4.3 and 5.4, Figure 6.3 is a bar plot for the result of the updated t-tests. The "added" red bar indicates how many more stocks are needed to produce a significant drop in the mean of standard deviation.As is shown, the red bar tends to be longer after size 8, which is to say, it is more difficult to decrease the portfolio standard deviations significantly after the 8^{th} stock is added to the portfolio. Noted that the mean of portfolio standard deviation of the portfolios with size 8 is 0.0616. The systematic variation is estimated to be 0.0536 and hence the unsystematic variation is largely reduced by then.

The F-test are performed on successive standard deviation about the mean portfolio standard deviation to study how the convergence of the standard deviations works when portfolio sizes are increasing.

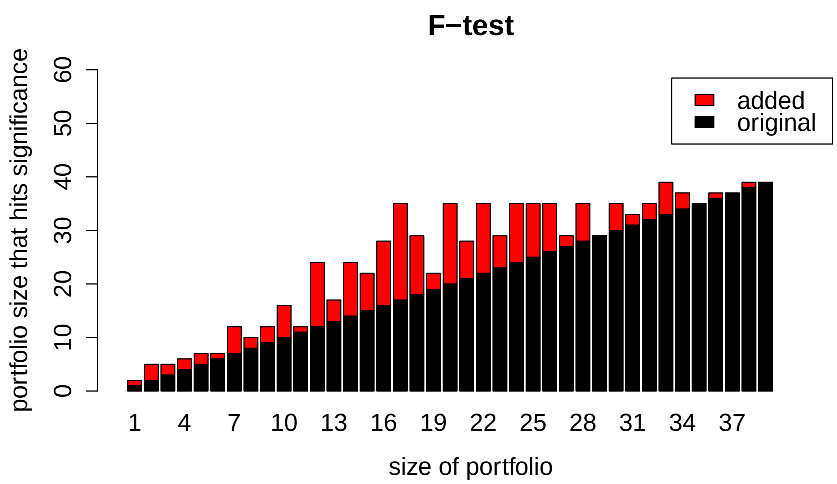

Figure 6.4: Bar plot for the result from the F-test

Just as Figure 4.4 and 5.5, Figure 6.4 is a bar plot for the result of the updated F-tests. The "added" red bar indicates how many more stocks are needed to produce a significant drop in the standard deviation of the risks. This provides similar result to the t-test. Things are getting hard after size 10.

Here is a table showing the exact result from both tests.

size	t-test(geo*)	t-test(arith*)	F-test(geo*)	F-test(arith*)
1	2.00	2.00	2.00	2.00
2	7.00	4.00	9.00	5.00

3	9.00	6.00	5.00	5.00
4	7.00	5.00	6.00	6.00
5	9.00	9.00	7.00	7.00
6	9.00	9.00	7.00	7.00
7	9.00	9.00	11.00	12.00
8	9.00	9.00	9.00	10.00
9	NA	35.00	11.00	12.00
10	12.00	11.00	11.00	16.00
11	22.00	35.00	14.00	12.00
12	35.00	35.00	14.00	24.00
13	35.00	NA	14.00	17.00
14	NA	31.00	28.00	24.00
15	18.00	18.00	16.00	22.00
16	NA	NA	28.00	28.00
17	22.00	31.00	22.00	35.00
18	NA	NA	22.00	29.00
19	35.00	35.00	22.00	22.00
20	22.00	31.00	22.00	35.00
21	35.00	31.00	28.00	28.00
22	NA	NA	33.00	35.00
23	35.00	NA	28.00	29.00
24	NA	35.00	28.00	35.00
25	34.00	31.00	28.00	35.00
26	33.00	31.00	33.00	35.00
27	NA	35.00	33.00	29.00
28	NA	35.00	NA	35.00
29	35.00	35.00	33.00	NA
30	33.00	31.00	33.00	35.00

31	NA	NA	33.00	33.00
32	NA	NA	40.00	35.00
33	NA	35.00	NA	39.00
34	NA	NA	40.00	37.00
35	NA	NA	NA	NA
36	NA	NA	40.00	37.00
37	NA	NA	NA	NA
38	NA	NA	39.00	39.00
39	NA	NA	NA	NA

Table 6.2: Additional stocks needed for the t-test and F-test

∗ "arith" represents the results coming from the arithmetic metric. "geo" represents the results coming from the geometric metric.

Table 6.2 gives a clear description about the output as well as the comparison of both tests. The addition number coming from both metric are similar, which reinforces the previous conclusion that after the 10^{th} stock, it is really hard to reduce the risk and make the risk more convergent by adding stocks.

6.5 Conclusion

The result generating from a different metric is surprisingly similar to what was captured by the geometric metric before. Therefore, the metric does not make a huge difference in the result.

CHAPTER 7

Conclusion

The result of the above study basically agrees with the conclusion from Evans and Archer[1]. There is a strong relationship between portfolio size and risk, and this relationship could be captured by decreasing asymptotic function. The asymptotic approximation is the systematic variation, which means by increasing portfolio size, the unsystematic variation could be largely reduced. The "ideal" number of components in a portfolio would be no more than 10 under this study, since increasing portfolio size larger than 10 could hardly decrease the risks. And if the cost of increasing portfolio size is taken into concern, it would be even more unnecessary to increase the size after 10.

All of these conclusions survive the challenges of both security and stock data, different data time period in the financial industry, large portfolio size, and different metric in calculation risk and return, which means the conclusions are stable.

CHAPTER 8

Appendix

8.1 Ticker Symbol Reference

Num in S&P500	Ticker symbol	Company
1	MMM	3M Company
2	ABT	Abbott Laboratories
4	ACE	ACE Limited
5	ACN	Accenture plc
6	ACT	Actavis plc
7	ADBE	Adobe Systems Inc
9	AES	AES Corp
10	AET	Aetna Inc
11	AFL	AFLAC Inc
12	A	Agilent Technologies Inc
13	GAS	AGL Resources Inc.
14	APD	Air Products & Chemicals Inc
15	ARG	Airgas Inc
16	AKAM	Akamai Technologies Inc
17	AA	Alcoa Inc
18	ALXN	Alexion Pharmaceuticals
19	ATI	Allegheny Technologies Inc
21	AGN	Allergan Inc

22	ADS	Alliance Data Systems
23	ALL	Allstate Corp
24	ALTR	Altera Corp
25	MO	Altria Group Inc
26	AMZN	Amazon.com Inc
27	AEE	Ameren Corp
28	AEP	American Electric Power
29	AXP	American Express Co
30	AIG	American International Group, Inc.
31	AMT	American Tower Corp A
32	AMP	Ameriprise Financial
33	ABC	AmerisourceBergen Corp
34	AME	Ametek
35	AMGN	Amgen Inc
36	APH	Amphenol Corp A
37	APC	Anadarko Petroleum Corp
38	ADI	Analog Devices, Inc.
39	AON	Aon plc
40	APA	Apache Corporation
41	AIV	Apartment Investment & Mgmt
42	AAPL	Apple Inc.
43	AMAT	Applied Materials Inc
44	ADM	Archer-Daniels-Midland Co
45	AIZ	Assurant Inc
46	T	AT&T Inc
47	ADSK	Autodesk Inc
48	ADP	Automatic Data Processing
49	AN	AutoNation Inc

50	AZO	AutoZone Inc
51	AVB	AvalonBay Communities, Inc.
52	AVY	Avery Dennison Corp
53	AVP	Avon Products
54	BHI	Baker Hughes Inc
55	BLL	Ball Corp
56	BAC	Bank of America Corp
57	BK	The Bank of New York Mellon Corp.
58	BCR	Bard (C.R.) Inc.
59	BAX	Baxter International Inc.
60	BBT	BB&T Corporation
61	BEAM	Beam Inc.
62	BDX	Becton Dickinson
63	BBBY	Bed Bath & Beyond
64	BMS	Bemis Company
66	BBY	Best Buy Co. Inc.
67	BIIB	BIOGEN IDEC Inc.
68	BLK	BlackRock
69	HRB	Block H&R
70	BA	Boeing Company
72	BXP	Boston Properties
73	BSX	Boston Scientific
74	BMY	Bristol-Myers Squibb
75	BRCM	Broadcom Corporation
76	BF.B	Brown-Forman Corporation
77	CHRW	C. H. Robinson Worldwide
78	CA	CA, Inc.
79	CVC	Cablevision Systems Corp.

80	COG	Cabot Oil & Gas
81	CAM	Cameron International Corp.
82	CPB	Campbell Soup
83	COF	Capital One Financial
84	CAH	Cardinal Health Inc.
86	KMX	Carmax Inc
87	CCL	Carnival Corp.
88	CAT	Caterpillar Inc.
89	CBG	CBRE Group
90	CBS	CBS Corp.
91	CELG	Celgene Corp.
92	CNP	CenterPoint Energy
93	CTL	CenturyLink Inc
94	CERN	Cerner
95	CF	CF Industries Holdings Inc
96	SCHW	Charles Schwab
97	CHK	Chesapeake Energy
98	CVX	Chevron Corp.
99	CMG	Chipotle Mexican Grill
100	CB	Chubb Corp.
101	CI	CIGNA Corp.
102	CINF	Cincinnati Financial
103	CTAS	Cintas Corporation
104	CSCO	Cisco Systems
105	C	Citigroup Inc.
106	CTXS	Citrix Systems
107	CLF	Cliffs Natural Resources
108	CLX	The Clorox Company

109	CME	CME Group Inc.
110	CMS	CMS Energy
111	COH	Coach Inc.
112	KO	The Coca Cola Company
113	CCE	Coca-Cola Enterprises
114	CTSH	Cognizant Technology Solutions
115	CL	Colgate-Palmolive
116	CMCSA	Comcast Corp.
117	CMA	Comerica Inc.
118	CSC	Computer Sciences Corp.
119	CAG	ConAgra Foods Inc.
120	COP	ConocoPhillips
121	CNX	CONSOL Energy Inc.
122	ED	Consolidated Edison
123	STZ	Constellation Brands
124	GLW	Corning Inc.
125	COST	Costco Co.
126	COV	Covidien plc
127	CCI	Crown Castle International Corp.
128	CSX	CSX Corp.
129	CMI	Cummins Inc.
130	CVS	CVS Caremark Corp.
131	DHI	D. R. Horton
132	DHR	Danaher Corp.
133	DRI	Darden Restaurants
134	DVA	DaVita Inc.
135	DE	Deere & Co.
137	DAL	Delta Air Lines

138	DNR	Denbury Resources Inc.
139	XRAY	Dentsply International
140	DVN	Devon Energy Corp.
141	DO	Diamond Offshore Drilling
142	DTV	DirecTV
143	DFS	Discover Financial Services
144	DISCA	Discovery Communications
146	DLTR	Dollar Tree
147	D	Dominion Resources
148	DOV	Dover Corp.
149	DOW	Dow Chemical
150	DPS	Dr Pepper Snapple Group
151	DTE	DTE Energy Co.
152	DD	Du Pont (E.I.)
153	DUK	Duke Energy
154	DNB	Dun & Bradstreet
155	ETFC	E-Trade
156	EMN	Eastman Chemical
157	ETN	Eaton Corp.
158	EBAY	eBay Inc.
159	ECL	Ecolab Inc.
160	EIX	Edison Int'l
161	EW	Edwards Lifesciences
162	EA	Electronic Arts
163	EMC	EMC Corp.
164	EMR	Emerson Electric
165	ESV	Ensco plc
166	ETR	Entergy Corp.

167	EOG	EOG Resources
168	EQT	EQT Corporation
169	EFX	Equifax Inc.
170	EQR	Equity Residential
171	EL	Estee Lauder Cos.
172	EXC	Exelon Corp.
173	EXPE	Expedia Inc.
174	EXPD	Expeditors Int'l
175	ESRX	Express Scripts
176	XOM	Exxon Mobil Corp.
177	FFIV	F5 Networks
179	FDO	Family Dollar Stores
180	FAST	Fastenal Co
181	FDX	FedEx Corporation
182	FIS	Fidelity National Information Services
183	FITB	Fifth Third Bancorp
184	FSLR	First Solar Inc
185	FE	FirstEnergy Corp
186	FISV	Fiserv Inc
187	FLIR	FLIR Systems
188	FLS	Flowserve Corporation
189	FLR	Fluor Corp.
190	FMC	FMC Corporation
191	FTI	FMC Technologies Inc.
192	F	Ford Motor
193	FRX	Forest Laboratories
194	FOSL	Fossil, Inc.
195	BEN	Franklin Resources

196	FCX	Freeport-McMoran Cp & Gld
197	FTR	Frontier Communications
198	GME	GameStop Corp.
199	GCI	Gannett Co.
200	GPS	Gap (The)
201	GRMN	Garmin Ltd.
202	GD	General Dynamics
203	GE	General Electric
205	GIS	General Mills
207	GPC	Genuine Parts
208	GNW	Genworth Financial Inc.
209	GILD	Gilead Sciences
210	GS	Goldman Sachs Group
211	GT	Goodyear Tire & Rubber
212	GOOG	Google Inc.
213	GWW	Grainger (W.W.) Inc.
214	HAL	Halliburton Co.
215	HOG	Harley-Davidson
216	HAR	Harman Int'l Industries
217	HRS	Harris Corporation
218	HIG	Hartford Financial Svc.Gp.
219	HAS	Hasbro Inc.
220	HCP	HCP Inc.
221	HCN	Health Care REIT, Inc.
222	HP	Helmerich & Payne
223	HES	Hess Corporation
224	HPQ	Hewlett-Packard
225	HD	Home Depot

226	HON	Honeywell Int'l Inc.
227	HRL	Hormel Foods Corp.
228	HSP	Hospira Inc.
229	HST	Host Hotels & Resorts
230	HCBK	Hudson City Bancorp
231	HUM	Humana Inc.
232	HBAN	Huntington Bancshares
233	ITW	Illinois Tool Works
234	IR	Ingersoll-Rand PLC
235	TEG	Integrys Energy Group Inc.
236	INTC	Intel Corp.
237	ICE	IntercontinentalExchange Inc.
238	IBM	International Bus. Machines
239	IGT	International Game Technology
240	IP	International Paper
241	IPG	Interpublic Group
242	IFF	Intl Flavors & Fragrances
243	INTU	Intuit Inc.
244	ISRG	Intuitive Surgical Inc.
245	IVZ	Invesco Ltd.
246	IRM	Iron Mountain Incorporated
247	JBL	Jabil Circuit
248	JEC	Jacobs Engineering Group
249	JNJ	Johnson & Johnson
250	JCI	Johnson Controls
251	JOY	Joy Global Inc.
252	JPM	JPMorgan Chase & Co.
253	JNPR	Juniper Networks

254	KSU	Kansas City Southern
255	K	Kellogg Co.
256	KEY	KeyCorp
257	KMB	Kimberly-Clark
260	KLAC	KLA-Tencor Corp.
261	KSS	Kohl's Corp.
263	KR	Kroger Co.
264	LB	L Brands Inc.
265	LLL	L-3 Communications Holdings
266	LH	Laboratory Corp. of America Holding
267	LRCX	Lam Research
268	LM	Legg Mason
269	LEG	Leggett & Platt
270	LEN	Lennar Corp.
271	LUK	Leucadia National Corp.
272	LIFE	Life Technologies
273	LLY	Lilly (Eli) & Co.
274	LNC	Lincoln National
275	LLTC	Linear Technology Corp.
276	LMT	Lockheed Martin Corp.
277	L	Loews Corp.
278	LO	Lorillard Inc.
279	LOW	Lowe's Cos.
280	LSI	LSI Corporation
282	MTB	M&T Bank Corp.
283	MAC	Macerich
284	M	Macy's Inc.
285	MRO	Marathon Oil Corp.

287	MAR	Marriott Int'l.
288	MMC	Marsh & McLennan
289	MAS	Masco Corp.
290	MA	Mastercard Inc.
291	MAT	Mattel Inc.
292	MKC	McCormick & Co.
293	MCD	McDonald's Corp.
294	MHFI	McGraw Hill Financial
295	MCK	McKesson Corp.
297	MWV	MeadWestvaco Corporation
298	MDT	Medtronic Inc.
299	MRK	Merck & Co.
300	MET	MetLife Inc.
301	MCHP	Microchip Technology
302	MU	Micron Technology
303	MSFT	Microsoft Corp.
304	MHK	Mohawk Industries
305	TAP	Molson Coors Brewing Company
306	MDLZ	Mondelez International
307	MON	Monsanto Co.
308	MNST	Monster Beverage
309	MCO	Moody's Corp
310	MS	Morgan Stanley
311	MOS	The Mosaic Company
312	MSI	Motorola Solutions Inc.
313	MUR	Murphy Oil
314	MYL	Mylan Inc.
315	NBR	Nabors Industries Ltd.

316	NDAQ	NASDAQ OMX Group
317	NOV	National Oilwell Varco Inc.
318	NTAP	NetApp
319	NFLX	NetFlix Inc.
320	NWL	Newell Rubbermaid Co.
321	NFX	Newfield Exploration Co
322	NEM	Newmont Mining Corp. (Hldg. Co.)
324	NEE	NextEra Energy Resources
326	NKE	NIKE Inc.
327	NI	NiSource Inc.
328	NE	Noble Corp
329	NBL	Noble Energy Inc
330	JWN	Nordstrom
331	NSC	Norfolk Southern Corp.
332	NTRS	Northern Trust Corp.
333	NOC	Northrop Grumman Corp.
334	NU	Northeast Utilities
335	NRG	NRG Energy
336	NUE	Nucor Corp.
337	NVDA	Nvidia Corporation
339	ORLY	O'Reilly Automotive
340	OXY	Occidental Petroleum
341	OMC	Omnicom Group
342	OKE	ONEOK
343	ORCL	Oracle Corp.
344	OI	Owens-Illinois Inc
345	PCG	P G & E Corp.
346	PCAR	PACCAR Inc.

347	PLL	Pall Corp.
348	PH	Parker-Hannifin
349	PDCO	Patterson Companies
350	PAYX	Paychex Inc.
351	BTU	Peabody Energy
352	PNR	Pentair Ltd.
353	PBCT	People's United Bank
354	POM	Pepco Holdings Inc.
355	PEP	PepsiCo Inc.
356	PKI	PerkinElmer
357	PRGO	Perrigo
358	PETM	PetSmart, Inc.
359	PFE	Pfizer Inc.
360	PM	Philip Morris International
362	PNW	Pinnacle West Capital
363	PXD	Pioneer Natural Resources
364	PBI	Pitney-Bowes
365	PCL	Plum Creek Timber Co.
366	PNC	PNC Financial Services
367	RL	Polo Ralph Lauren Corp.
368	PPG	PPG Industries
369	PPL	PPL Corp.
370	PX	Praxair Inc.
371	PCP	Precision Castparts
372	PCLN	Priceline.com Inc
373	PFG	Principal Financial Group
374	PG	Procter & Gamble
375	PGR	Progressive Corp.

376	PLD	Prologis
377	PRU	Prudential Financial
379	PSA	Public Storage
380	PHM	Pulte Homes Inc.
381	PVH	PVH Corp.
383	PWR	Quanta Services Inc.
384	QCOM	QUALCOMM Inc.
385	DGX	Quest Diagnostics
386	RRC	Range Resources Corp.
387	RTN	Raytheon Co.
388	RHT	Red Hat Inc.
389	REGN	Regeneron
390	RF	Regions Financial Corp.
391	RSG	Republic Services Inc
392	RAI	Reynolds American Inc.
393	RHI	Robert Half International
394	ROK	Rockwell Automation Inc.
395	COL	Rockwell Collins
396	ROP	Roper Industries
397	ROST	Ross Stores
398	RDC	Rowan Cos.
399	R	Ryder System
400	SWY	Safeway Inc.
401	CRM	Salesforce.com
402	SNDK	SanDisk Corporation
403	SCG	SCANA Corp
404	SLB	Schlumberger Ltd.
405	SNI	Scripps Networks Interactive Inc.

406	STX	Seagate Technology
407	SEE	Sealed Air Corp.(New)
408	SRE	Sempra Energy
409	SHW	Sherwin-Williams
410	SIAL	Sigma-Aldrich
411	SPG	Simon Property Group Inc
412	SLM	SLM Corporation
413	SJM	Smucker (J.M.)
414	SNA	Snap-On Inc.
415	SO	Southern Co.
416	LUV	Southwest Airlines
417	SWN	Southwestern Energy
418	SE	Spectra Energy Corp.
419	STJ	St Jude Medical
420	SWK	Stanley Black & Decker
421	SPLS	Staples Inc.
422	SBUX	Starbucks Corp.
423	HOT	Starwood Hotels & Resorts
424	STT	State Street Corp.
425	SRCL	Stericycle Inc
426	SYK	Stryker Corp.
427	STI	SunTrust Banks
428	SYMC	Symantec Corp.
429	SYY	Sysco Corp.
430	TROW	T. Rowe Price Group
431	TGT	Target Corp.
432	TEL	TE Connectivity Ltd.
433	TE	TECO Energy

434	THC	Tenet Healthcare Corp.
435	TDC	Teradata Corp.
436	TSO	Tesoro Petroleum Co.
437	TXN	Texas Instruments
438	TXT	Textron Inc.
439	HSY	The Hershey Company
440	TRV	The Travelers Companies Inc.
441	TMO	Thermo Fisher Scientific
442	TIF	Tiffany & Co.
443	TWX	Time Warner Inc.
444	TWC	Time Warner Cable Inc.
445	TJX	TJX Companies Inc.
446	TMK	Torchmark Corp.
447	TSS	Total System Services
448	RIG	Transocean
450	FOXA	Twenty-First Century Fox
451	TSN	Tyson Foods
452	TYC	Tyco International
453	USB	U.S. Bancorp
454	UNP	Union Pacific
455	UNH	United Health Group Inc.
456	UPS	United Parcel Service
457	X	United States Steel Corp.
458	UTX	United Technologies
459	UNM	Unum Group
460	URBN	Urban Outfitters
461	VFC	V.F. Corp.
462	VLO	Valero Energy

463	VAR	Varian Medical Systems
464	VTR	Ventas Inc
465	VRSN	Verisign Inc.
466	VZ	Verizon Communications
467	VRTX	Vertex Pharmaceuticals Inc
468	VIAB	Viacom Inc.
469	V	Visa Inc.
470	VNO	Vornado Realty Trust
471	VMC	Vulcan Materials
472	WMT	Wal-Mart Stores
473	WAG	Walgreen Co.
474	DIS	The Walt Disney Company
475	GHC	Graham Holdings Company
476	WM	Waste Management Inc.
477	WAT	Waters Corporation
478	WLP	WellPoint Inc.
479	WFC	Wells Fargo
480	WDC	Western Digital
481	WU	Western Union Co
482	WY	Weyerhaeuser Corp.
483	WHR	Whirlpool Corp.
484	WFM	Whole Foods Market
485	WMB	Williams Cos.
486	WIN	Windstream Communications
487	WEC	Wisconsin Energy Corporation
489	WYN	Wyndham Worldwide
490	WYNN	Wynn Resorts Ltd
491	XEL	Xcel Energy Inc

492	XRX	Xerox Corp.
493	XLNX	Xilinx Inc
494	XL	XL Capital
496	YHOO	Yahoo Inc.
497	YUM	Yum! Brands Inc
498	ZMH	Zimmer Holdings
499	ZION	Zions Bancorp

Table 8.1: Companies used to make portfolios

8.2 Completed output from the extended t-test and f-test

size	significant size(t-test)	significant size(F-test)
1	2.00	2.00
2	7.00	9.00
3	9.00	5.00
4	7.00	6.00
5	9.00	7.00
6	9.00	7.00
7	9.00	11.00
8	9.00	9.00
9	48.00	11.00
10	12.00	11.00
11	22.00	14.00
12	35.00	14.00
13	35.00	14.00
14	43.00	28.00
15	18.00	16.00
16	48.00	28.00
17	22.00	22.00
18	48.00	22.00
19	35.00	22.00
20	22.00	22.00
21	35.00	28.00
22	48.00	33.00
23	35.00	28.00
24	43.00	28.00
25	34.00	28.00

26	33.00	33.00
27	48.00	33.00
28	43.00	42.00
29	35.00	33.00
30	33.00	33.00
31	68.00	33.00
32	48.00	40.00
33	48.00	44.00
34	82.00	40.00
35	82.00	43.00
36	48.00	40.00
37	57.00	43.00
38	48.00	39.00
39	68.00	43.00
40	57.00	44.00
41	68.00	43.00
42	68.00	52.00
43	82.00	68.00
44	82.00	70.00
45	57.00	48.00
46	47.00	52.00
47	82.00	48.00
48	82.00	70.00
49	82.00	52.00
50	75.00	52.00
51	57.00	52.00
52	68.00	76.00
53	82.00	57.00

54	68.00	70.00
55	56.00	57.00
56	NA	58.00
57	NA	76.00
58	68.00	88.00
59	82.00	76.00
60	NA	68.00
61	82.00	70.00
62	68.00	70.00
63	68.00	68.00
64	75.00	68.00
65	68.00	68.00
66	68.00	70.00
67	75.00	70.00
68	NA	93.00
69	70.00	91.00
70	NA	98.00
71	74.00	93.00
72	NA	76.00
73	82.00	76.00
74	NA	93.00
75	NA	88.00
76	82.00	98.00
77	82.00	88.00
78	82.00	93.00
79	NA	93.00
80	82.00	86.00
81	82.00	88.00

82	NA	98.00
83	88.00	93.00
84	NA	93.00
85	NA	91.00
86	NA	98.00
87	91.00	98.00
88	NA	NA
89	91.00	98.00
90	91.00	NA
91	NA	NA
92	NA	98.00
93	NA	NA
94	NA	98.00
95	NA	98.00
96	98.00	98.00
97	NA	98.00
98	NA	NA
99	NA	NA

Table 8.2: Additional stocks needed for the t-test and F-test

References

[1] John L. Evans and Stephen H. Archer , *"Diversification and the Reduction of Dispersion: An Empirical Analysis"* The Journal of Finance, Vol. 23 No. 5 (December, 1968), pp. 761-767

[2] Per B. Mokkelbost, *"Unsystematic Risk Over Time"* Journal of Financial & Quantitative Analysis, Vol. 6 Issue 2, (March, 1971), pp. 785-796

[3] William F. Sharpe, *"A Simplified Model for Portfolio Analysis"* Management Science, Vol. 9 Issue 2, (January, 1963), pp. 277-293

[4] John Lintner, *"Security Prices, Risk, And Maximal Gains From Diversification"* Journal of Finance, Vol. 20 (December, 1965), pp. 587-615

[5] K. Larry Hastie, *"The Determination of Optimal Investment Policy"* Management Science, Vol. 13 (August, 1967), pp. B-757-S-774

[6] Meir Statman, *"How Many Stocks Make a Diversified Portfolio?"* Journal of Financial & Quantitative Analysis, Vol. 22 No.3, (September, 1987), pp. 353-363

[7] Benton E. Gup, *"The Basics of Investing. 2nd ed."* New York: John Wiley & Sons (1983), pp. 363-364

[8] Frank K. Reilly, *"Investment Analysis and Portfolio Management. 2nd ed. "* San Francisco: Dryden Press (1985), pp. 101

[9] Harry M. Markowitz, *"Portfolio Selection"* The Journal of Finance, Vol. 7, No. 1 (Mar., 1952), pp. 77-91

[10] Julian J. Faraway, *"Linear Models with R"* Taylor & Francis, 2009

[11] Robert C. Merton, *"An Analytic Derivation of the Efficient Portfolio Frontier"* The Journal of Financial and Quantitative Analysis, Vol. 7, No. 4 (Sep., 1972), pp. 1851-1872

[12] Oldrich A. Vasicek, *"A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas"* The Journal of Finance, Vol. 28, No. 5 (Dec., 1973), pp. 1233-1239

[13] Marshall E. Blume, *"On the Assessment of Risk"* Source: The Journal of Finance, Vol. 26, No. 1 (Mar., 1971), pp. 1-10

[14] Marshall E. Blume, *"Betas and Their Regression Tendencies"* The Journal of Finance, Vol. 30, No. 3 (Jun., 1975), pp. 785-795

[15] R. J. Kuo and C. W. Hong, *"Integration of Genetic Algorithm and Particle Swarm Optimization for Investment Portfolio Optimizatio"* Applied Mathematics & Information Sciences, Vol. 7, No. 6 (Nov., 2013), pp. 2397-2408

[16] Edwin J. Elton, Martin J. Gruber and Manfred W. Padberg, *"Simple Criteria for Optimal Portfolio Selection with Upper Bounds"* Operations Research, Vol. 25, No. 6 (Nov. - Dec., 1977), pp. 952-967

[17] Edwin J. Elton, Martin J. Gruber and Manfred W. Padberg, *"Simple Criteria for Optimal Portfolio Selection: Tracing out the Efficient Frontier"* The Journal of Finance, Vol. 33, No. 1 (Mar., 1978), pp. 296-302

[18] Edwin J. Elton, Martin J. Gruber and Manfred W. Padberg, *"Simple Criteria for Optimal Portfolio Selection: the Multi Group Case* The Journal of Finance and Quantitative Analysis, Vol. 12, No. 3 (Sept., 1977), pp. 329-345

[19] Paul Teetor, *"R Cookbook"* O'Reilly Media; 1 edition (March 22, 2011)