Columbia University

STAT W4290

Statistics Method in Finance

Final Project Report

Author:

Ye Zhong (YZ2703)

Jiahong Hu (JH3561)

Rui Sun (RS3345)

Supervisor:

Prof.Hammou Elbarmi

December 14, 2015

1 Summary

In this project, the portfolio consists of the following nine U.S. equity stocks, Alphabet Inc. [GOOGL], International Business Machine [IBM], Amazon.com Inc. [AMZN], Johnson & Johnson [JNJ], Pflizer Inc. [PFE], Merck Co. Inc. [MRK], Cobalt International Energy, Inc. [CIE], CVR Energy, Inc. [CVI], Boardwalk Pipeline Partners, LP [BWP], as shown in **Table 1**[Appendix]. The stocks above were selected based upon the criteria listed below:

1. Risk control by diversification: The underlying companies are in the following three promising industries: Technology, Pharmaceutical and Energy Industry.

2. Medium to large cap category: The conservative investment strategy generates steady medium to long term return with moderate fluctuates.

3. U.S. based companies: The U.S. market has an immediate reflect on the stock price.

4. NYSE or NASDAQ listed companies: The selection enables the team to use S&P 500 index as benchmark and 3-month Treasury bill rate as risk-free rate.

Comprehensive quantitative approach was taken to evaluate the performance of each stock and the portfolio in terms of return and corresponding volatility, with the ultimate goal of identifying optimal investment strategies with specified targets.The five years monthly closing prices starting from June 2010 to June 2015 was analyzed to develop models. We firstly performed the exploratory data analysis by computing the sample statistics and fitting the most suitable distribution for each stock. Secondly, we applied the portfolio theory and asset allocation optimization model and found out the minimum variance portfolio, tangency portfolio, and maximization of sharp ratios portfolio with the assumption on whether the short selling is allowed. Thirdly, we checked the dependency property among nice stocks and performed principal component analysis. Lastly, we estimated value value-at-risk (VaR) and expected shortfall (ES) to evaluate and compare the potential loss of each stock and portfolio by parametric, nonparametric and Copula methods.

2 Descriptive Statistics

In the exploratory data analysis session, we used graphical and numerical descriptive statistics to describe the nine stocks in the portfolio and then studied the marginal distribution of each stocks monthly return.

2.1 Sample Statistics

To begin with our analysis, we calculated the means, variances, skewness coefficients and kurtosis coefficients of nine stocks, as shown in **Table 2**. We found out that CVI generated superior average annualized return, approximately 51.51% every year during the time period between June 2010 and June 2015; however, at the same time, it associated with the second highest average annualized volatility, 47.71%. The stock price of CVI skyrocketed in 2013 because CVI, the petroleum refiner and fertilizer manufacturer gained advantage of cheap crude oil. In addition to CVI, the following three stocks GOOGL, AMZN and PFE also remarkably outperformed the benchmark S&P 500 with 14.62% yearly return on average. As shown in **Figure 1** [Appendix] , the equity curves of the four stocks are above S&P 500, indicating that the return of a same $1 investment in any of those four stocks is higher than in S&P 500 at any time period during the five years.

Moreover, skewness coefficients of the five stocks S&P 500, GOOGL, IBM, AMZN and CVI are approximately equal to 0 and their kurtosis coefficients are around 3, showing that their distributions of monthly return do not have significant departures from normality. However, the stock BMP has a large negative skewness coefficient, which indicates its corresponding distribution has an asymmetric shape with a relative long left tail compared to the right tail. On the contrary, the stock CIE has a large positive skewness coefficient, which means that its distribution is right-skewed. Furthermore, both stocks CIE and BMP have high kurtosis coefficients, stating that both distributions are heavy-tailed compared to normal distribution. The stock BMP should be carefully concerned because of the high possibility of an extreme large negative return. At the same time, if one sells shorts, the high chance of a large positive return, as what happened for CIE, is also worrisome.

Table 2. Mean, Standard Deviation, Skewnwss and Kurtosis of nine stocks

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BMP	S&P500
Annualized Return	20.62%	8.84%	31.27%	14.13%	22.06%	14.75%	17.38%	51.27%	2.46%	14.62%
Annualized Volatility	24.17%	16.14%	26.57%	12.96%	14.86%	15.21%	51.51%	47.61%	31.63%	12.01%
Skewness	0.5086	-0.4112	0.2839	0.3882	-0.0570	0.1802	1.0021	0.0884	-2.5267	-0.0229
Kurtosis	3.0714	3.1103	3.0080	2.8351	2.2717	2.5745	4.6874	3.3780	15.5801	3.3798

2.2 Distribution Modeling Fit

We further studied the return distribution of each stock separately. As shown in **Table 3**, both hypothesis tests Sharprio-Wilk test and Jarque-Beta test conclude at 5% significance level that the stocks BMP and CIE are not normally distributed. The density curves of both stocks

are obviously different from the normal distribution density curve, as shown in **Figure 2** [Appendix]. The normality probability curves of two stocks in **Figure 3** [Appendix] again support the conclusion.

Table 3. Normality, 0utliers, Distribution fitting and Stationality

	Outlier	Normality	Distribution Fit	Stationality
GOOGL	NO	YES	Normal Dist	Stationary
IBM	YES	YES	Normal Dist	Stationary
AMZN	YES	YES	Normal Dist	Stationary
JNJ	YES	YES	Normal Dist	Stationary
PFE	NO	YES	Normal Dist	Stationary
MRK	NO	YES	Normal Dist	Stationary
CIE	YES	NO	Skewed GLE Dist	Stationary
CVI	YES	YES	Normal Dist	Stationary
BMP	YES	NO	Std T-Dist	Stationary
S&P500	YES	YES	Normal Dist	Stationery

Then, we fit each stock by maximum likelihood method with the following five different distributions: skewed t-distribution (SSTD), standard t-distribution (STD), generalized error distribution (GED), skewed GED (SGED) and skewed normal distribution (SNORM). We selected the most suitable model for each stock as the one with lowest Akaike Information Criterion (AIC), as shown in **Table 3**. The most suitable model for the stock BMP and CIE turns out to be standard t-distribution and skewed generalized error model, both of which have heavier tails than normal distribution, which are shown in **Figure 4**[Appendix]. Both fitted distributions perform well on tails with the tradeoff of slight departures from peaks, which ensure more precise estimates of VaR and ES in further analysis.

3 Portfolio Theory

In this section,we examined the portfolio theory we learned from this course and calculated portfolio weight for Minimum Variance Portfolio (MVP) and Tangency Portfolio with or without short-selling.

3.1 MVP and Efficient Frontiers

The weights of nine stocks in the Minimum variance portfolio are showned in **Table 4**. We have seen that under both assumptions (allowing short sale and no allowing short sale), the MVP put more weights on the pharmaceutical industry stocks. The total weights of the three pharmaceutical stocks while allowing short sale and not allowing short sale are about 67.6%

and 65.8%, respectively. It is because that the pharmaceutical stocks have smaller volatility on average compared to the technology and energy industry stocks we have chosen. On the contrary, the three energy stocks as a group have the highest volatility. Hence, MVP under both assumptions put almost no weights on energy stocks.

Table 4. MVP Weights with/without Short Sale

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
Allowing Short Sale	0.065	0.226	0.000	0.303	0.161	0.212	-0.03	0.008	0.056
No Short Sale	0.059	0.218	0.011	0.322	0.128	0.209	0.00	0.000	0.053

Table 5 shows the individual stock's mean return and standard deviation compared with MVP. We have seen that JNJ and MRK have almost the same mean return with MVP. In general, all stocks in pharmaceutical group have mean return and standard deviation close to those of MVP. Technology industry stocks have higher mean returns compared to the MVP's mean return except for IBM. Meanwhile, technology stocks generally have higher standard deviation than MVP's standard deviation. Lastly, the energy sector has showed inconsistent performance. CIE and CVI showed higher mean return than MVP's mean return. The CVI has remarkably high mean return 51.27%, which is almost 4 times of MVP's mean return. However, CVI also has the 2nd largest standard deviation, which is 5 times of MVP's standard deviation. BWP has the worst performance among all stocks. It results in negative return but still has really high volatility.

Table 5. Annulized Retrun and Standard Deviation with Short Sale

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP	MVP
Annulinzed Mean Return	20.62%	8.84%	31.27%	14.13%	22.06%	14.76%	17.38%	51.27%	-2.46%	14%
Annulized Standard Deviation	24.2%	16.1%	26.6%	13.0%	14.9%	15.2%	51.5%	47.6%	31.6%	9.97%

We calclulated 5% VaR with/without short sale. **Table 6** shows all VaRs for individual stocks are much larger than MVP's VaR regardless under short sale or not. The VaR for MVP under short sale is greater than that without short sale.

Table 6. 5% VaR with/without Short Sale (in$)

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP	MVP
VaR with Short Sale	9469	6239	9602	3598	5466	5318	20331	17210	10513	4011
Var without Short Sale	9488	6251	9620	3632	5471	5269	20393	17595	10567	3606

3.2 Sharpe Ratio and Tangency Portfolio

In this section we compute Sharpe ratio for each stock and tangency portfolio under two conditions, and we also showed the efficient frontier. **Table.7** shows the mean return, standard deviation, and variace for both condtions as below:

Table.7 Tangent Portfolio with/without Short Sale

	Mean Return	Standard Deviation	Variance
Tangent Portfolio Allow Short Sale	2.43%	4.15%	17.2%
Tangent Portfolio Not Allow Short Sale	2.04%	3.65%	13.3%

The tangent portfolio without short sale has smaller standard deviation but also has smaller mean return compared to the tangent portfolio with short sale. This means short sale might have higher return with the trade off increasing the overall portfolio risk.

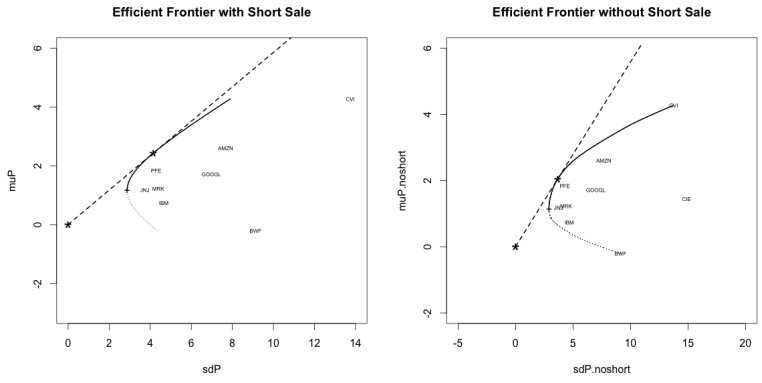
Next we computed the Sharpe ratio for tangent portfolio under both condtions:

Table.8 Tangent Portfolio Sharpe Ratio Comapre to Each Stock

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP	Tangent Short	Tangent no Short
Sharpe Ratio	0.246	0.157	0.339	0.313	0.427	0.279	0.0971	0.31	-0.023	0.585	0.558

We have seen that the stock PFE has the highest sharpe ratio. However, the sharpe ratio of the tangent portfolio outperforms any individual stocks. Compared to tangent portfolio under short sale, the sharpe ratio under short sale is higher but not by much.

Further, we would like to show the efficient frontiers for both cases. Compared with the short-selling case, the efficient frontier without short sale is relatively rightward shifted and narrowed since we put nonnegative restriction on portfolio weights. As a result, we could not a achieve the desired portfolio returns with the same risk under short-selling. Also the MVP return in no short is lower than the MVP's return in allowing short sale case.



4 Asset Allocation

To achieve the expected monthly return of 0.5%, we need to allocate our asset as (without short-selling). According to **Table 7**, we have the results : Monthly risk = 0.0432 , 5%VaR = 3928 and 5% ES = 8529.

Table.7 Expected Monthly Return 0.5%

GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
0.00%	48.20%	0.00%	7.70%	0.00%	10.10%	0.00%	0.00%	34.00%

The average monthly return for each stock is as showed in **Table 8**. We have discovered IBM takes a significant weight (48.2%) because its average monthly return is close to 0.5%. Stocks with returns notably greater than 0.5% are excluded from portfolio, since there is no short sale.

Table.8 Aeverage Monthly Return

GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
1.719%	0.737%	2.606%	1.178%	1.838%	1.230%	1.449%	4.272%	-0.205%

If we included risk-free Treasury Bill in the portfolio, we should allocate money as in **Table 9** we have results: monthly risk = 0.00887, 5%VaR = 1112 and 5% ES = 1345.

Table.9 Assets Allocation with Risk Free

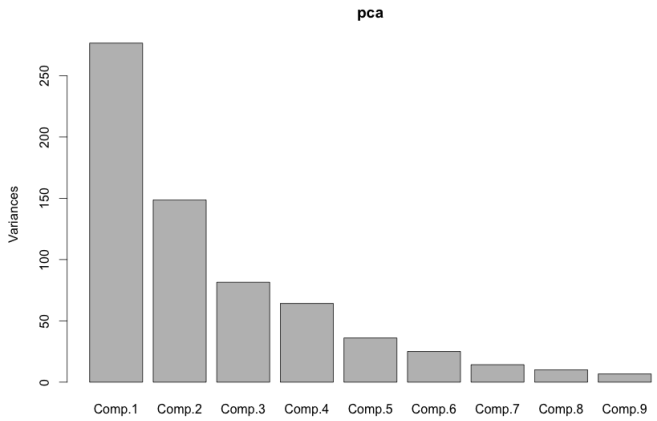
GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP	RF
1.436%	0.000%	3.353%	2.624%	10.498%	3.718%	0.000%	2.673%	0.000%	75.7%

With Treasury bill in the portfolio, the weight in each asset changed from prior result. More importantly, the monthly risk decreased significantly, as well as the 5% Value at Risk and Expected Shortfall. The reason of the lower monthly risk is that a big portion of the capital was invested in risk free asset; therefore the total risk of the portfolio was reduced.

5 Principal Component Analysis

In this session, we checked the dependency properties of each stock with time and the correlation among the nine stocks. As shown in **Table 3**, all the time series of the nine stocks are stationary according to augment Dickey-Fuller test at 5% significance level. JNJ and PFE, both from

pharmaceutical industry, are the most correlated because they have the highest sample correlation coefficient among nine stocks, which is approximately 0.5080, as in **Table 4**[Appendix]. Moreover, as shown in **Figure 5** [Appendix], both stocks move along with S&P 500 in certain degrees. Also the scatter plot for the two stocks supports the moderate positive correlation with a slight upward trend, as shown in **Figure 6**[Appendix]. BWP and MRK are the least correlated with lowest sample correlation coefficient -0.0039. Most of the estimated correlations are relatively small and there is no obvious trend between most pairs. Therefore, we think that diversification will reduce risk for the portfolio with nine assets.



According to the PCA plot we discovered that the first 5 principle components can represent almost 92% of the total variance and those five could be used to represent the portfolio.

6 Risk Management

If we assumed that returns are normally distributed, estimated 5% VaR and ES for nine stocks are shown as below, and we discovered that CIE has the highest VaR and highest ES while JNJ has lowest VaR and lowest ES.

Table.9 Normal VaR and ES (in$)

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
VaR	9760	6929	10011	4974	5217	5991	23011	18335	15228
ES	12676	8877	13216	6537	7010	7825	29224	24079	19044

If we use nonparametric method, estimated 5% VaR and ES for nine stocks list below. Also, CVI has the highest VaR and ES while JNJ still has lowest VaR and lowest ES.

Table.10 Nonparametric VaR and ES (in $))

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
VaR	9834	6457	9939	4236	5549	5792	21510	24534	11546
ES	10137	9328	9991	5151	6317	6818	23551	25592	25098

By using bootstrap method, if we assume each stock follows normal distribution, we could calculate 95% confidence interval for VaR and ES:

Table.11 Bootstrap Under Normal VaR and ES (in $)

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
VaR Upper	11950	9002	12649	6157	6758	7485	28310	24008	24032
VaR Lower	7569	4857	7373	3790	3677	4496	17711	12662	6424
ES Upper	15243	11264	16293	7917	8756	9542	35828	30677	29614
ES Lower	10109	6489	10138	5156	5263	6109	22621	17480	8474

If we used nonparametric method, by applying bootstrap method, the 95% confidence interval for VaR and ES is as follows:

Table.12 Nonparametric with Bootstrap VaR and ES (in$)

	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BWP
VaR Upper	11991	10319	11337	5967	7446	7696	27767	34324	31837
VaR Lower	7677	2596	8541	2505	3652	3888	15252	14745	8745
ES Upper	11432	13490	10789	7005	7670	8450	28256	31937	48651
ES Lower	8842	5165	9194	3297	4964	5186	18846	19247	1545

By comparing the results using the two different methods, we noticed, in general, the Value at Risk and Expected Shortfall calculated by nonparametric method are smaller than those calculated by parametric method. Therefore, using nonparametric method may underestimate the risk. In addition, it is in our concern that we have a relatively small sample size (60 monthly returns), the result could be less accurate using nonparametric method because of the data size.

7 Copula

In this session, we applied the copulas method to model the joint distribution of the nine stocks based on the most suitable marginal distribution of each stock mentioned in session 2.2. We fit four different copulas, normal copula, t copula, clayton copula and gumbel copula. Then, we selected the best copula model with the lowest AIC score, which is gumbel copula, as shown in the **Table 13**. The fitted distribution of portfolio return under gumbel copula model performs well on its tail, as shown in **Figure 7**[Appendix].

Table 13. AIC for Coupulas

	Normal Distribution	T Distribution	Clayton Distribuion	Gumbel Distribution
AIC	-29.8338	-29.9389	-41.5466	-50.0983

In the end, we calculated the 95% confidence interval VaR for the tangent portfolio (short selling allowed) under the joint distribution estimated by gumbel copula. The result is as shown in **Table 14: (95% CI for VaR for** $\$100,000$ **tangent portfolio when short selling allowed under gumbel copulas model)**. We found out that the 95% confidence interval is relatively tight, which means that the estimation under copula model is precise.

Table 14. 95% VaR CI Tangent Portfolio No Short

Statistics of Interest	Mean of VaR	VaR Lower Bound	VaR Upper Bound
	5995	5793	6195

8 Appendix

Table.1 Description of the portfolio in the project with nine stocks

No.	Company Name	Tickle	Stock Exchange	Asset Class	Industry	Share Outstanding	Closing Price 12/1/2015	Market Cap
1	Alphabet Inc	GOOGL	NASDAQ	Equity	Tech	291.33B	$783.79	$534.95B
2	Amazon	AMZN	NASDAQ	Equity	Tech	468.76B	$679.09	$316.89B
3	Interational Business Machines Corporation	IBM	NYSE	Equity	Tech	970.11B	141.28	135.52B
4	Johnson & Johnson	JNJ	NYSE	Equity	Pharmaceuticals	2.77B	$102.36	$282.37B
5	Pfizer Inc.	PFE	NYSE	Equity	Pharmaceuticals		6.17B $33.62	$203.11B
6	Merck Co., Inc.	MRK	NYSE	Equity	Pharmaceuticals	2.79B	$54.57	$150.35B
7	CRV Energy Inc.	CRI	NYSE	Equity	Energy	86.83M	$47.37	$4.00B
8	Cobalt International Energy	CIE	NYSE	Equity	Energy	41.45M	$7.66	$3.15B
9	Boardwalk Pipeline Partners, LP	BWP	NYSE	Equity	Energy	250.3M	$12.62	$2.94B

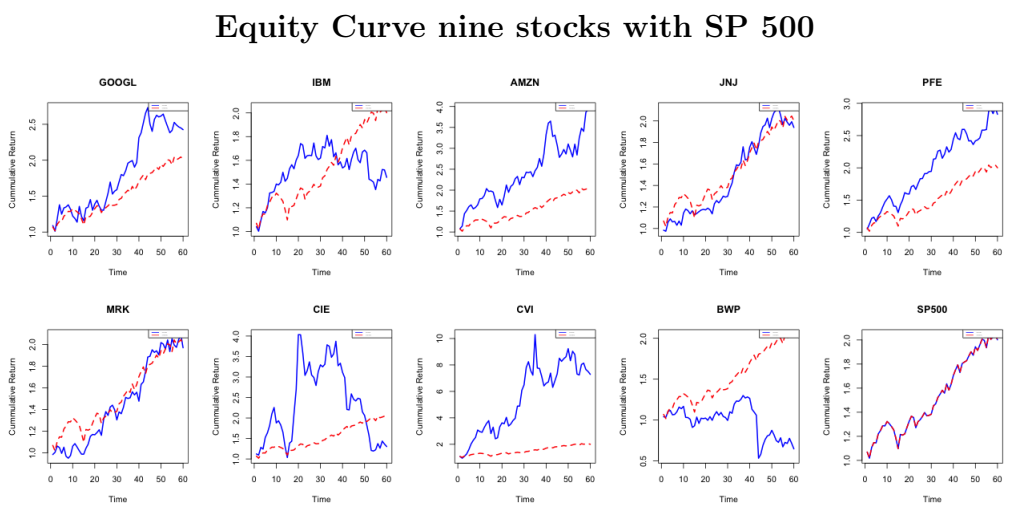


Figure 1: Equity Curve nine stocks with SP 500

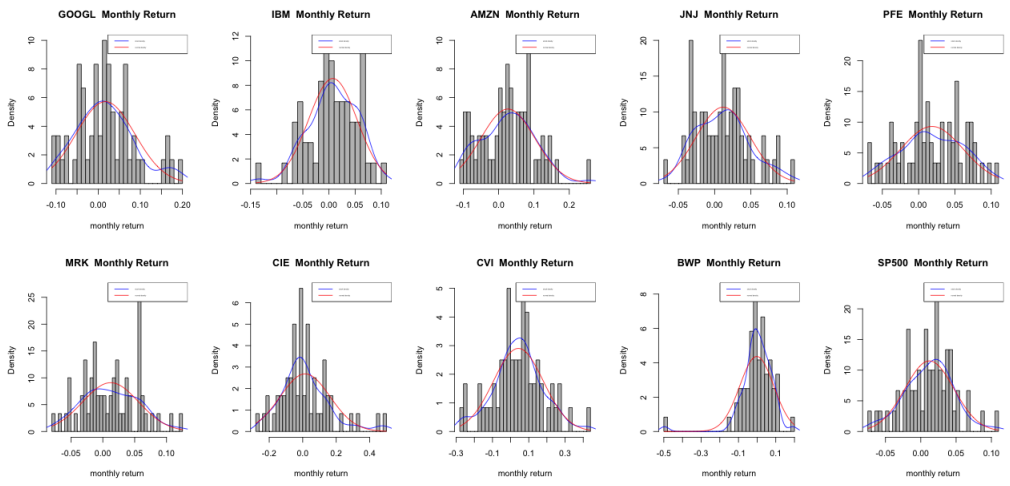


Figure 2: Histogram and Density Curve of Nine stocks

Normal Probability Plots of Nine stocks and S&P 500

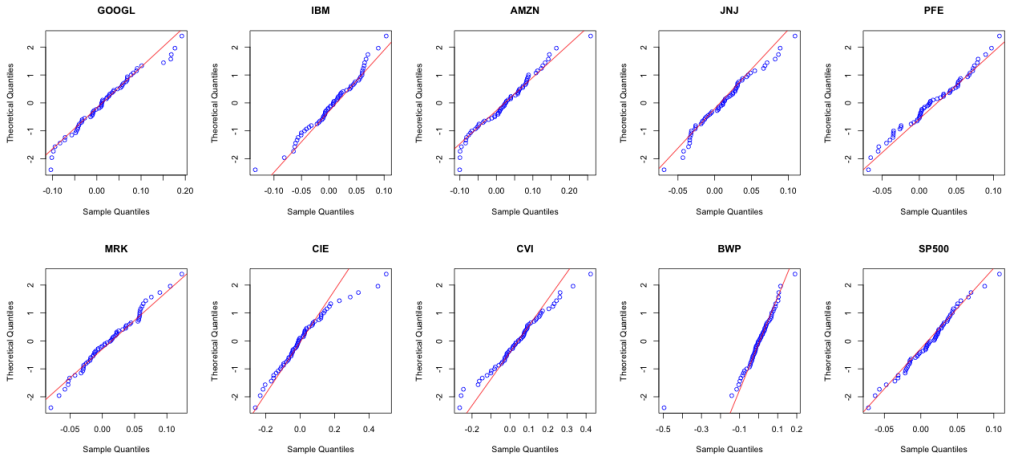


Figure 3: Normal Probability Plots of Nine stocks and S&P 500

Table 4.Correlation Matrix of Nine Stocks and S&P 500

Correlation Matrix	GOOGL	IBM	AMZN	JNJ	PFE	MRK	CIE	CVI	BMP	S&P 500
GOOGL	1.0000	0.3286	0.2649	0.1329	0.2809	0.0769	0.2417	0.1894	0.1136	0.5200
IBM	0.3286	1.0000	0.3133	0.2005	0.2602	0.1622	0.2730	0.3806	0.1998	0.5200
AMZN	0.2649	0.3133	1.0000	0.2364	0.2191	0.1808	0.0113	0.2793	0.0321	0.3820
JNJ	0.1329	0.2005	0.2364	1.0000	0.5080	0.4667	0.1170	0.0590	0.1363	0.5051
PFE	0.2809	0.2602	0.2191	0.5080	1.0000	0.4214	0.3645	0.0697	0.0439	0.5876
MRK	0.0769	0.1622	0.1808	0.4667	0.4214	1.0000	0.1453	-0.0248	-0.0039	0.3115
CIE	0.2417	0.2730	0.0113	0.1170	0.3645	0.1453	1.0000	0.2871	0.1186	0.4819
CVI	0.1894	0.3806	0.2793	0.0590	0.0697	-0.0248	0.2871	1.0000	0.0886	0.4758
BMP	0.1136	0.1998	0.0321	0.1363	0.0439	-0.0039	0.1186	0.0886	1.0000	0.2307
S&P 500	0.5200	0.5200	0.3820	0.5051	0.5876	0.3115	0.4819	0.4758	0.2307	1.0000

Figure 5a Return series of nine stocks

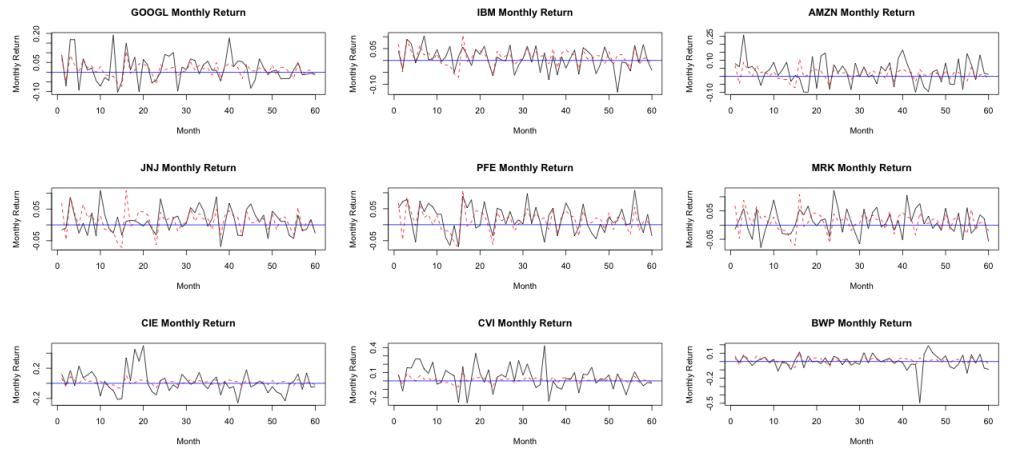


Figure 4: Figure 5a Return series of nine stocks

Figure 5b. Stock price of nine stocks

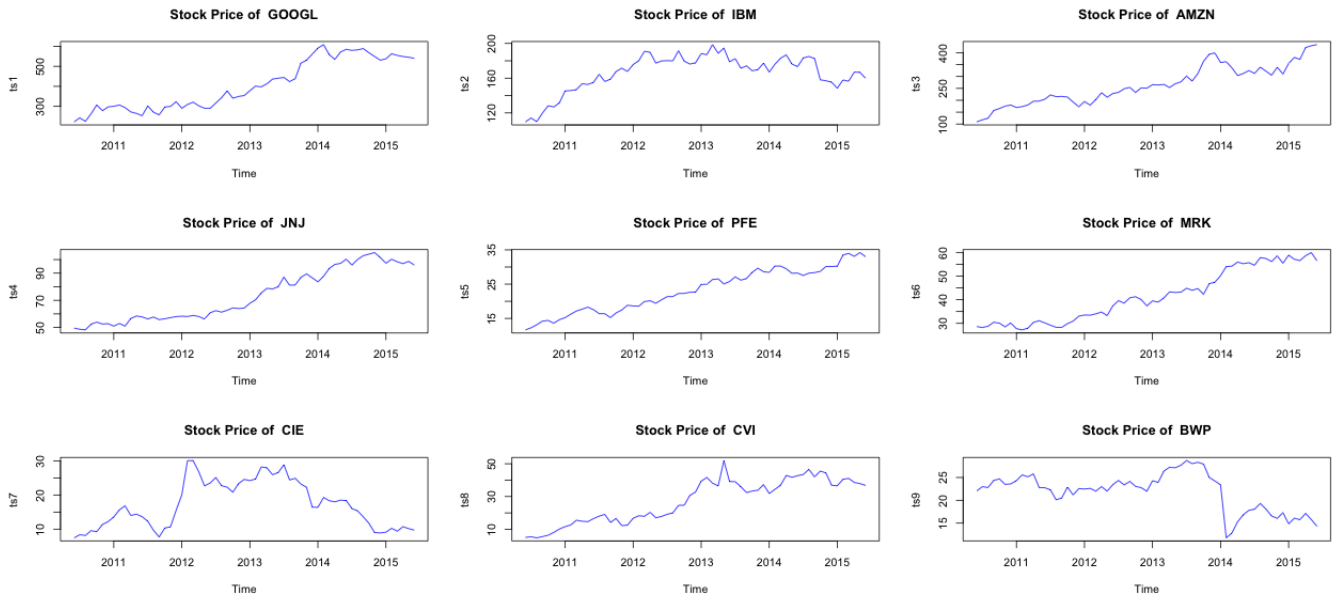


Figure 5: Figure 5b.Stock price of nine stocks

Figure 6. Scatter plot of nine stocks and S&P 500

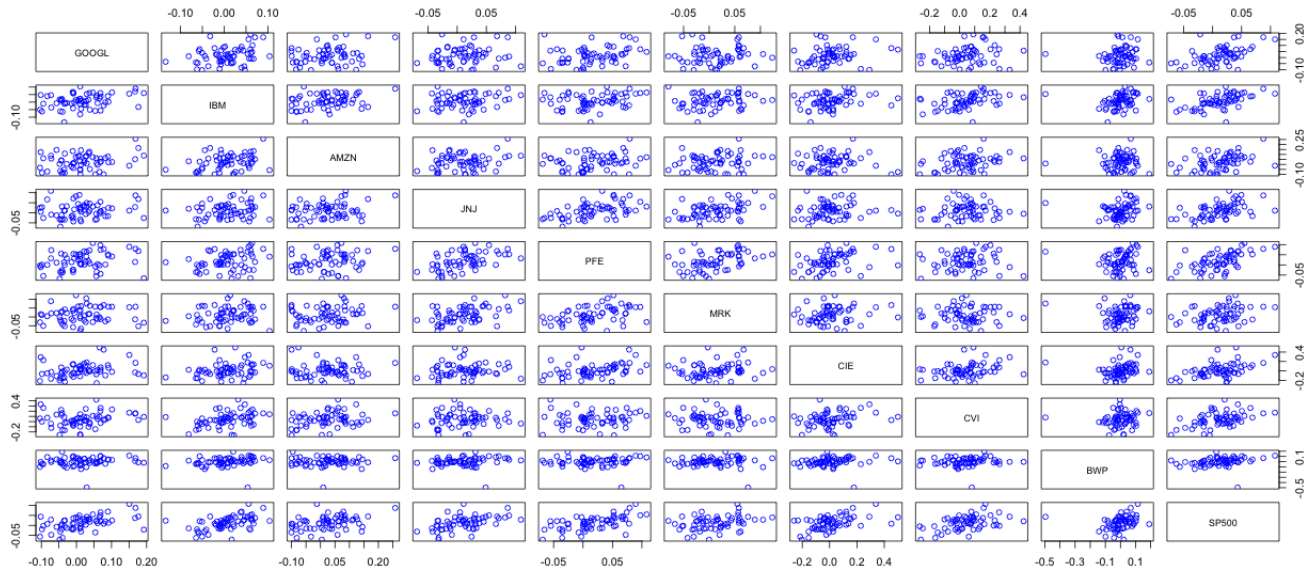


Figure 6: Figure 6. Scatter plot of nine stocks and S&P 500

Boxplot for Outliers

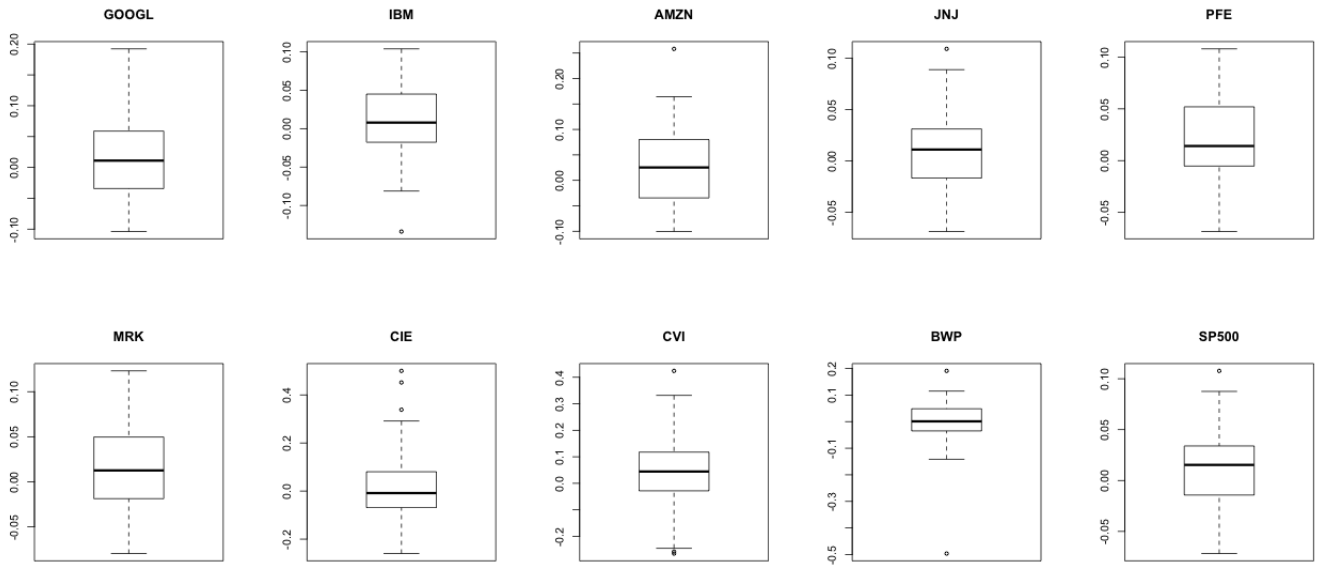


Figure 7: Boxplot for Outliers