Financial Data Analysis – W5261 Final Project Report

Wen Ren (wr2279), Xuan Zhang (xz2463), Yifei Lang (yl3365)
Dec. 10th, 2016 Columbia University

Summary

With the development of financial industry, more and more complicated financial instruments and derivatives have been sparking the market. The increasingly complex market gives rise to various challenges to traditional assumptions. It has been apparent that the linear correlation among assets cannot reveal the real dependency among assets, and the traditional Value at Risk (VaR) with normal distribution assumption may sometimes underestimate the true risk. Therefore, in our project, we will fit our assets into different distributions, construct portfolios based on investors' preferences, use nonparametric methods to estimate the VaR and apply copula to find the dependency among different assets.

Key Words: Descriptive Statistics, Portfolio Management, Asset Allocation, Risk Management, Value at Risk, Expected Shortfall, Principal Component Analysis, Factor Analysis, Copula

1. Data Collection

We collected the monthly prices of 15 assets and S&P500 index price over the time period of Jan. 2010 to Dec. 2015 from Yahoo Finance. For the complete company/asset information, please refer to the Appendix Table 9. The 15 assets come from 7 industries, including Basic Materials (FCX, MRO, X, VALE); Consumer Goods (AAPL, COKE, F); Financial (BAC); Healthcare (MDT, PFE); Industrial Goods (GE); Services (SIRI, SBUX); and Technology (AMD, VZ). For each asset, there are 72 monthly prices, and all of them are after the adjustments of dividend and stock split.

We transformed the assets' adjusted close prices into net returns and mainly researched on net returns' properties. Each asset has 71 net returns in total. In this report, the net return or net returns will be called "return" or "returns" for short.

We collected the three-month T-Bill rate monthly data over the time period of Jan. 2010 to Dec. 2015, and then take the average as our risk free rate. The risk free rate is around 0.07% per year, and data was downloaded from Federal Reserve official website[1].

2. Descriptive Statistics

In this part, we report sample descriptive statistics of net returns, including Means, Standard Deviations (SD), Skewness Coefficients, Kurtosis Coefficients, Betas, Sharpe Ratio, Stationarity Test, Normality Test and Distribution Estimation, etc. Please refer to the Appendix Table 10 for the overall summary of the descriptive statistics. Now, we will discuss these statistics into details.

(1) Monthly Prices

The monthly prices plots are showed in Appendix Figure 3 (3 graphs in total). Each plot includes monthly closing price for five assets with a separate S&P500 curve. We can see from the first graph that AMD, AAPL, BAC, FCX are relatively flat, while COKE has a greater volatility compared to other four assets. The range is approximately 0 to 200, and they are under S&P index. In the second graph, MDT has an increasing trend compared to the flatness of F, GE, MRO, and PFE. All of them are under S&P 500 index and MDT is the closest one to the index. The range is about 5 to 70. In addition, in third graph, SIRI, SBUX, VALE, VZ are below S&P

[1] Federal Reserve Official Website link: http://www.federalreserve.gov/releases/h15/data.htm

500 index. X is above the index at the beginning and it is more volatile than others. The range is about 0 to 60.

(2) Monthly Returns

The monthly return plots are showed in Appendix Figure 4. From the plots, we can see that the change for each asset is similar to the change of S&P500. However, compared with S&P 500, the 15 assets are more volatile. Most of them are relatively stationary over time, and most of the assets fluctuate about the means. We will have outlier test in part (6) with detailed investigation and analysis about the outliers and reasons that may explain some fluctuations.

(3) Equity Curve

We draw equity curve here for all the 15 assets to show the growth of $1 over the time period of 2010-2015. The equity curve of all the assets and S&P 500 is plotted in Appendix Figure 5, and the separate plots are in Appendix Figure 6. From this plot, we can compare the change in value of the assets over a period of time. We can see that approximately seven curves are below S&P 500 curve, including AMD, BAC, FCX, F, MDT, X, and VALE.

(4) Histogram and Density Estimation

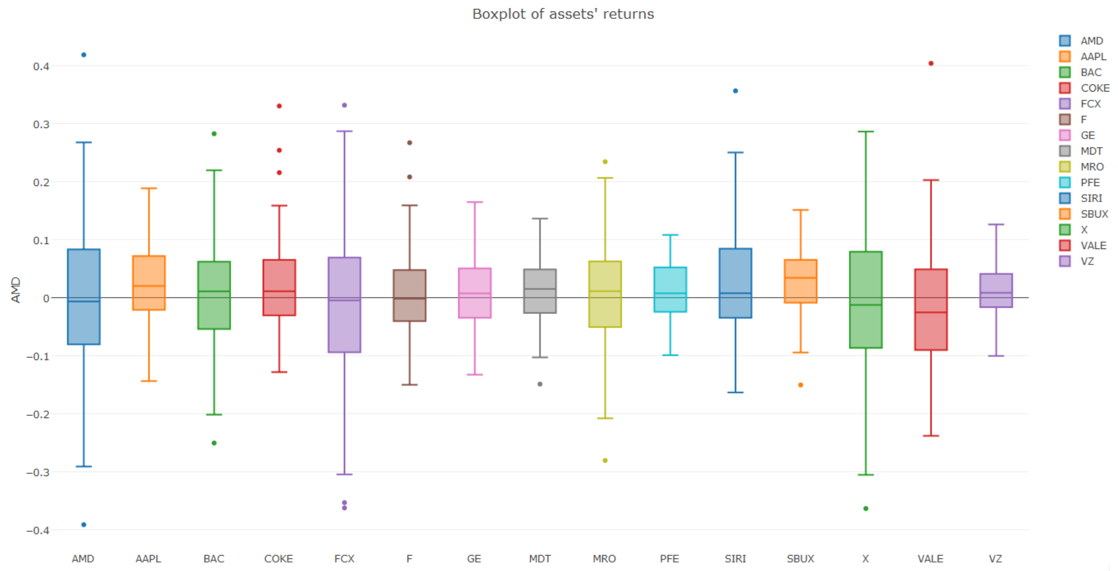
The histograms and density estimation plots are in Appendix Figure 7 and Figure 8. From the histograms, we can see that most of them have bell shape curve. COKE, VALE, SIRI and F are skew to left, SBUX is skew to right. MRO and VZ does not have smooth curve as others.

In the Density Estimation part, we use the Gaussian kernel to fit the density of each asset, and use different bandwidth (0.3, 0.7, 1, 3, and 7 times of the default bandwidth nrd0 in r) to adjust the density estimation with the aim to see how the bandwidth of kernel function affect the density estimation. From the Appendix Figure 8, we find out that the bigger the bandwidth, the smother the density estimation and the bigger the bias, and vice versa.

(5) Box Plot

From the box plot below, we can see that the outliers exist and we will closely look at them in the next part. The medians are similar with each other and SBUX has the highest median.

Figure 1: Boxplot of Assets' Net Returns



(6) Outliers

The outliers' date and corresponding returns are outlined in Appendix Table 11. We can find out that the outliers are mostly clustered in 2010-2012 and 2015, and only 2 outliers happened in 2013 and 2014. This phenomenon can also be found in the net return plots in which we can see that most of the assets are relatively volatile during 2010-2012 and 2015. Moreover, during 2010-2012, there are lots of lower outliers; while in 2015, the upper outliers are in dominant.

The volatile return and lower outliers during 2010-2012 are mainly because of the sharp drop in stock price in Aug 2011 across United States, Middle East, Europe and Asia. The Eurozone debt crisis was dramatically worsened in 2011. Nations such as Greece, Ireland and Portugal were especially in deep recession, United States and Frances suffered from AAA rating concerns. Overall, the global economy was generally gloomy. After the mid of 2012, the global economy was on the mend and stock price keep going up. In 2015, the global stock had a relatively big downturn. It is mainly because of the fluctuation of Chinese stock market, which results in more outliers in our assets.

We can also find out that the companies in Basic Material industry have more low outliers than technology companies in 2015. Due to the oil price falling from middle of 2014, Basic Material companies suffer from revenue declining, which results in the lower outliers of returns for these companies.

(7) Pairwise Scatter Plots and Covariance Matrix

The covariance matrix (Appendix Table 12), correlations matrix (Appendix Table 13) and pairwise scatter plot between the returns of different assets (Appendix Figure 11) can give us a direct impression of the relationships among the assets. Please also find the correlation matrix heatmap and a virtualization plot of the correlation in the Appendix Figure 9 and Figure 10. From the covariance and correlation matrix, we can see that almost all the assets have positive covariance, except for COKE and VZ. We can also find that "Cock-Cola" has no significant correlation with any other company. This can tell us that the stock price of Coca-Cola would not fluctuate with other companies, which means Coca-Cola maybe come from a different industry compared with other companies. Although Coca-Cola and Apple, as well as Ford belong to Consumer Goods section together, this kind of Beverage would not be influenced by the others. What's more, Verizon is another company which was not influenced by other companies. All the correlations of Verizon are less than 0.5, and the highest correlation is 0.48 with Pfizer which is a Drug Manufacturers company.

When focus on some high correlations, we can find that most of correlations between Freeport-McMoRan (FCX), a copper company, and other companies are relatively high. The highest correlation appears between FCX and Vale, a metals and minerals company. The FCX also has high correlations with the Marathon Oil which is an Oil & Gas company, the United States Steel which is obviously a Steel & Iron company, as well as Ford which is an Auto Manufactures company. This phenomenon indicates that the Basic Materials industry has a relatively large influence to other industries, including other metals' price.

Moreover, the correlation between General Electric and Medtronic is 0.65. It seems that an electronic company should not have such a high relationship with a healthcare company. But one reason is that GE is a diversified machinery company, which involves in various industries including the medical equipment industry. Another pair of high correlation is Medtronic and Pfizer, a Medical Appliance & Equipment company and a drug manufacturers company. These two companies come from two related industry. So that it is reasonable to have a high correlation.

Based on all the analysis above, we can find that most of the correlations among these 15 companies are less than 0.5, which implies a low correlation. So buying a portfolio of these stocks can obtain a lower risk than just buying one of them. On the other hand, since there is nearly no negative correlation, it means that any two of these companies do not come from contrary industries. Thus, the reduction of risk will be limited, which means that there are better portfolios if we add some other stocks.

(8) Stationarity Test

We use the Priestley-Subba Rao (PSR) test for non-stationarity. The result showed below. We can see that only five of them are nonstationary, which includes COKE, FCX, F, MRO and SIRI.

Table 1: Stationarity Test Results

	AMD	AAPL	BAC	COKE	FCX
T	0.305	0.102	0.305	0.0008	0.012
Stationary	YES	YES	YES	NO	NO
	F	**GE**	**MDT**	**MRO**	**PFE**
T	0.017	0.07	0.149	0.0022	0.711
Stationary	NO	YES	YES	NO	YES
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
T	2.59e-5	0.138	0.264	0.053	0.067
Stationary	NO	YES	YES	YES	YES

(9) QQ Plot and Normality Test

The QQ plots of all the assets are in Appendix Figure 12. The plots show that the lines are close to linear, especially after deleting outliers. This means that it is reasonable to assume normality of our assets. VZ may be the one that most likely nonlinear.

We also use Kolmogorov-Smirnov Tests to test normality. All the p-values are really small. They all pass the normality test.

(10) Fitting Distributions

In this part, we use the returns to fit distributions. We want to see which one of normal distribution or t distribution is a better fit to the data. The result showed below, with AIC values. We can see that 9 assets fit t distribution well (with lower AIC value), while the other six fit normal better.

Table 2: Fit Distribution Results

	AMD	AAPL	BAC	COKE	FCX
Normal	-73.23	-170.50	-124.59	-156.54	-77.52
t	-74.00	-169.55	-124.96	-161.96	-78.53
Best One	T	Normal	T	T	T
	F	**GE**	**MDT**	**MRO**	**PFE**
Normal	-147.57	-182.99	-208.99	-129.59	-229.25
T	-147.95	-181.99	-208.47	-131.16	-227.13
Best One	T	Normal	Normal	T	Normal
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
Normal	-117.15	-200.44	-83.71	-109.56	-229.30
t	-117.97	-200.82	-83.40	-112.50	-228.75
Best One	T	T	Normal	T	Normal

(11) Sharpe Ratio and Beta

The Sharpe Ratio is showed below, and the Sharpe ratio for S&P 500 is 0.896. We can see that SBUX has the largest Sharpe Ratio which is 1.638, and X has the lowest Sharpe Ratio which is -0.37.

The betas are showed below, and the beta for S&P 500 is 1 since it is the market risk. We can see that beta for AMD, BAC, FCX, F, GE, MDT, MRO, SIRI, X, and VALE are greater than 1, and beta for AAPL, COKE, PFE, SBUX, and VZ are smaller than 1. Since beta is a measure of how aggressive the asset and how sensitive it is to market movement, we can conclude that AMD, BAC, FCX, F, GE, MDT, MRO, SIRI, X, and VALE are aggressive assets, and AAPL, COKE, PFE, SBUX, and VZ are non-aggressive assets.

Table 3: Sharpe Ratio and Beta

	AMD	AAPL	BAC	COKE	FCX
Sharpe Ratio	-0.08	1.109	0.235	0.985	-0.237
Beta	2.078	0.928	1.541	0.477	2.273
	F	**GE**	**MDT**	**MRO**	**PFE**
Sharpe Ratio	0.357	0.753	0.732	0.066	0.876
Beta	1.465	1.387	1.103	1.579	0.790
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
Sharpe Ratio	0.924	1.638	-0.37	-0.6	0.919
Beta	1.621	0.789	1.861	1.493	0.483

3. Portfolio Theory

In portfolio theory part, we are going to construct minimum variance portfolio (MVP), tangent portfolio and compute the efficient frontier under the conditions when the short sales are allowed and when they are not allowed. Under each circumstance, we will compute the portfolio mean, standard deviation, Value at Risk, expected shortfall and Sharpe ratio. Then, we will compare the portfolios with one another with the aim to find the best or most suitable portfolios for potential investors.

For each portfolio, we estimate 5% Value at Risk of the returns over a month, and estimate related expected shortfall on $100,000 investment over a month investment horizon. Because normal distribution doesn't fit our data well, we use nonparametric method to calculate VaR and expected shortfall here.

(1) Short Sales are allowed

When short sales are allowed, it is possible for some assets to have negative weights. It means that the investor can sell the asset he or she does not currently owned and subsequently repurchase them. You can find the efficient portfolio frontier plot when short sales are allowed in the Appendix Figure 13.

- **Minimum Variance Portfolio (MVP)**

The MVP descriptive statistics and the MVP weights can be found in Appendix Table 14. You may also find the comparison of the portfolio and 15 assets in the Appendix Table 20. In accordance with the definition of MVP, the portfolio has the lowest standard deviation compared with all other 15 single assets. Although MVP doesn't have the highest mean, it has the highest Sharpe ratio among all 15 single assets. According to the MVP weights, in order to minimize the overall variance (standard deviation), the assets with high standard deviation will be shorted (e.g. FCX, GE and X) or take relatively small amount (e.g. AMD and SIRI) in the portfolio; while the assets with low standard deviation will have relatively more weight (e.g. AAPL, COKE and VZ). The MVP also has the lowest VaR and expected shortfall among all other 15 single assets which means the portfolio diversifies the overall risk.

- **Tangency Portfolio**

The tangency portfolio descriptive statistics and weights are in Appendix Table 15. You may also find the comparison of the portfolio and 15 assets in the Appendix Table 20. In accordance with the definition of tangency portfolio, the Sharpe ratio of tangency portfolio is much higher than all other 15 single assets. According to the tangency portfolio, the asset with low Sharpe ratio will be shorted (e.g. AMD, FCX, X and VALE); while the asset with high Sharpe ratio will take a relatively bigger weight (e.g. AAPL, SBUX and VZ). The tangency portfolio also has the lowest VaR and expected shortfall among all 15 single assets. Compared with MVP, tangency portfolio has lower VaR and expected shortfall as well.

(2) Short sales are not allowed

When short sales are not allowed, all the weights should be either positive or zero. You may find the efficient portfolio frontier plot when short sales are not allowed in the Appendix Figure 14.

- **Minimum Variance Portfolio (MVP)**

The MVP statistics and MPV weights when there are no short sales are outlined in Appendix Table 16. You may also find the comparison of the portfolio and 15 assets in the Appendix Table 20. The standard deviation of MPV is also the smallest among all other 15 single assets, according to MVP definition. The VaR and expected shortfall are also far less than those of the single assets. Because the short sales are not allowed, all weights are non-negative. The current portfolio only contains 8 assets (AAPL, BAC, COKE, MDT, PFE, SIRI, SBUX and VZ), while other assets' weights are zero.

Because we add a restriction (no short sale) to the portfolio, the standard deviation of current portfolio is higher than that of the MVP with no restriction, while the Sharpe ratio is lower. It means that to some extent, the restriction make the portfolio less efficient.

- **Tangency Portfolio**

The tangency portfolio statistics and weights when there are no short sales are outlined in Appendix Table 17. You may also find the comparison of the portfolio and 15 assets in the Appendix Table 20. The Sharpe ratio of the portfolio is the highest compared with all other 15 single assets. Because no short sales allowed, the portfolio only contains 5 assets (AAPL, COKE, SIRI, SBUX and VZ) which have relatively high Sharpe ratios. Those assets with low Sharpe ratios are not needed in this portfolio.

Compared with tangency portfolio when the short sales are allowed, the Sharpe ratio of current asset is much lower, while the standard deviation is much higher. This also resonates with our conclusion that the restriction makes the portfolio less effective. The VaR and expected shortfall of the tangency portfolio are far less than those of the single assets.

In summary, we can see that the tangency portfolio with short sales has the highest Sharpe ratio and mean, and the lowest VaR and expected shortfall. The MVP with short sales has the lowest standard deviation. For all the portfolios, no short sale means an additional restriction to the portfolio constructions which exert some negative impact to the portfolio structure. The restriction increase the portfolio's risk and somehow decrease the return. Whether to choose MVP or tangency portfolio or other portfolio structures depends on investors' risk and return preference.

Table 4: Results Summary of MVP and Tangency Portfolio with and without Short Sales (Annualized)

	Mean	**SD**	**VaR**	**ES**	**Sharpe Ratio**
MVP (no short sale)	0.203	0.116	3178	4582	1.74
Tangent (no short sale)	0.261	0.129	3995	4495	2.02
MVP (with short sale)	0.203	0.107	3232	4078	1.89
Tangent (with short sale)	0.332	0.121	2708	3162	2.73

4. Asset Allocation

Assume we have $100,000 to invest into the 15 assets and our target expected return is 6% per year (which corresponds to an expected return of 0.5% per month). No short sales allowed in this part. We use nonparametric method to calculate VaR and expected shortfall.

(1) No T-Bills and No Short Sales Allowed

When there are no T-Bills, we will only use the 15 risky assets to construct a portfolio to achieve the target expected return of 6% annually. The portfolio statistics and weights are outlined in Appendix Table 18. The portfolio contains 7 assets (AMD, COKE, MDT, PFE, X, VALE and VZ).

The 5% VaR of the portfolio is 6009, which means that there is 5% probability that the portfolio will loss

$6009 over a month. The expected shortfall of the portfolio is 7187, which means the expected loss when VaR is exceeded is $7187.

(2) With T-Bills and No Short Sales Allowed

When the T-Bills are allowed in the portfolio, we will use both the 15 risky assets and the risk free asset (T-Bills) to construct a portfolio to achieve the target expected return. We already know that the tangency portfolio is the most efficient one when there are only risky assets. Therefore, when with T-Bills, the efficient portfolio will be the combination of tangency portfolio and risk free asset with proper weights to be assigned to both of them in order to meet the target expected return requirement.

The portfolio statistics and weights are outlined in Appendix Table 19. The portfolio only contains 5 risky assets (AAPL, COKE, SIRI, SBUX and VZ) and the risk free asset (T-bill). Because the returns of some assets are negative, some are much greater than 6%; the dominant part of the portfolio is the risk free asset, T-Bill, which accounts for 77.2%.

Compared with the allocation of only risky assets, the standard deviation, VaR and expected shortfall of this portfolio are much lower, which means that current portfolio has much better diversification effect and much less risk.

5. Risk Management

In order to have a comprehensive grasp of assets' risk and have a better understanding of different risk measurements, here we use three methods to conduct risk management for the 15 assets. First, we calculate Value at Risk (VaR) and expected shortfall (ES) based on the assumption that the assets follow normal distribution with estimated means and variances. Second, we use nonparametric method to calculate VaR and expected shortfall. Last, we use bootstrap to compute the estimated standard errors and compute 95% confidence interval (CI) for our VaR estimates and expected shortfall estimates.

For each asset, we estimate 5% Value at Risk of the net returns over a month, and estimate related expected shortfall on $100,000 investment over a month investment horizon.

(1) Based on normal distribution

From the estimation of 5% VaR and expected shortfall, we find out that the first asset, AMD, has the highest VaR and expected shortfall; while the last asset, VZ, has the lowest VaR and expected shortfall.

Table 5: VaR and Expected Shortfall based on Normal Distribution

	AMD	**AAPL**	**BAC**	**COKE**	**FCX**
VaR	23590	9442	15531	10699	23510
ES	29502	12421	19648	13986	29246
	F	**GE**	**MDT**	**MRO**	**PFE**
VaR	12914	9316	7790	15460	6557
ES	16416	12045	10062	19435	8527
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
VaR	14303	6764	23012	19909	6496
ES	18641	9177	28503	24486	8465

(2) Based on Nonparametric Method

Using nonparametric method, the 13th asset, X, has the highest VaR, while the 5th asset, FCX, has the highest expected shortfall. The last asset, VZ, has the lowest VaR and expected shortfall.

From the result, we find out that the asset with the highest VaR does not necessarily have the highest expected shortfall. VaR and expected shortfall depends on the asset's distribution, and expected shortfall is even more sensitive to the distribution shape. Sometimes, the asset with greater VaR may have a relatively smaller expected shortfall when compared with other assets, and vice versa.

Table 6: VaR and Expected Shortfall based on Nonparametric Method

	AMD	**AAPL**	**BAC**	**COKE**	**FCX**
VaR	22342	10021	15286	8057	22143
ES	29759	11942	19090	9722	32299
	F	**GE**	**MDT**	**MRO**	**PFE**
VaR	11524	9554	7162	17236	6461
ES	13436	11268	10116	21636	7826
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
VaR	13702	6590	25747	17746	5414
ES	15470	10219	30519	20948	7190

(3) Based on Bootstrap

We are also interested in the confidence interval of the VaR and expected shortfall. Because normal distribution does not fit our data very well, we prefer nonparametric VaR and expected shortfall. Due to the small amount of data (each asset only has 71 returns), we use bootstrap to generate more samples of each asset and then calculate the 95% confidence interval based on bootstrap samples. Please refer to the Appendix Table 21 and Table 22 for the bootstrap VaR and expected shortfall confidence interval results.

6. Principal Component Analysis and Factor Analysis

(1) Principal Component Analysis

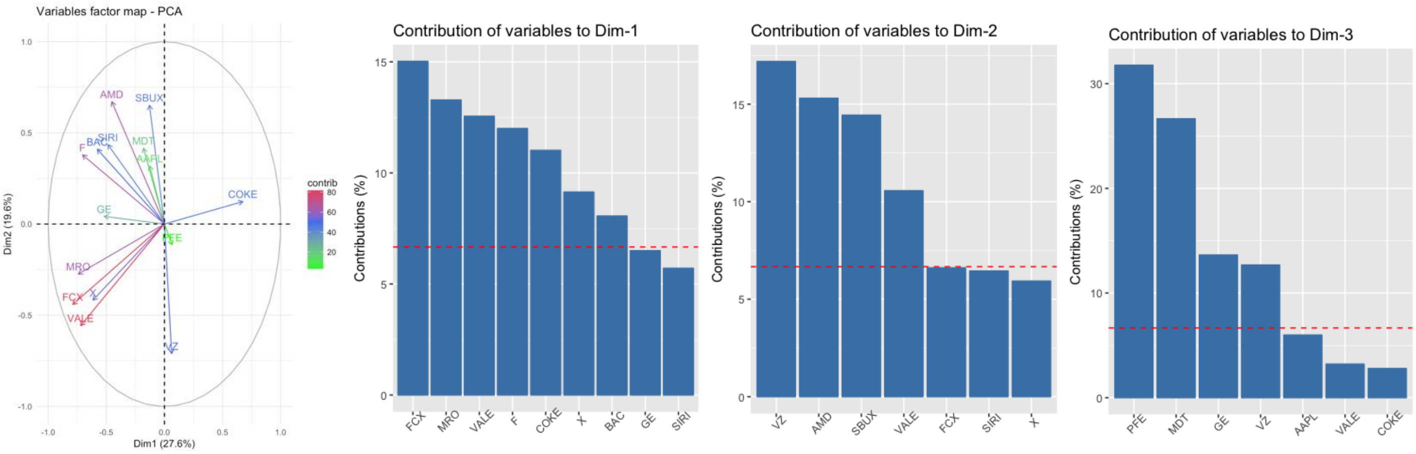
When we run Principal Component Analysis, our goal should be reducing the dimension of the variable. So every principal component is a linear combination of all variables and explains a part of the variance of the whole data. Then we should decide how many PCs we should keep.

From the Appendix Table 23 and Figure 15, we can simply find that the first three principal components explained 64.3% variance and contributed most to the variance explanation. And variance explanation of the 4th PC decreased significantly. But containing the 4th PC may increase the total variance explanation to 72.1%. So, it will be reasonable to keep the first three or four principal components and drop other principal components. Now, we can have more analysis to the first three principal components.

From Figure 2 below, we can find out which companies contribute to the first two PCs more, which means more important than other companies. It is clear that FCX and VALE contribute most to the variance of first two PCs. And Ford, AMD as well as MRO contribute a little bit less. And next are SBUX, SIRI, COKE, VZ and BAC. These phenomena can provide us a sense of the financial market. First, the fluctuation of basic materials industry may cause significant change to other industries. So, when we do research or find opportunities in financial market, the price of basic materials like copper, steel and oil may be concerned prior. Second, the fluctuation of basic materials may influence the manufacturer industry first because they use basic material in production.

On the other hand, the companies like PFE and APPLE have little contribution to the first two PCs. But as we can find from the results above, PFE contributed most to PC3 and APPLE contributes most to PC4. This situation told us that healthcare and consumer goods also have huge influence to the financial market, but they influence the market from different aspect.

Figure 2: Projection of the first two PCs (1), Contribution of each variable to the PC1-PC3 (2-4)



(2) Factor Analysis

Factor analysis, to some extent, is like the PCA, but provides a more direct way to explain the results. So, for the convenience of explanation, it is important to choose how many factors to use. Thus, we did factor analysis for the number of factors equals to 2, 3, 4 and 5. Here is a brief summary of the results.

Table 7: Results Summary for Factor Analysis

Number of Factors	2	3	4	5
Cumulative Variance	0.428	0.511	0.538	0.59
p-value for the test of factor number sufficiency	0.0309	0.58	0.724	0.748

From the table above, we can choose the number of factors. Firstly, the null hypothesis of the test is: the number of factors is sufficient. So, only when the p-value of the test exceeds 0.05, the number of factors will be reasonable. Thus, 2 factors are not sufficient in this situation. Secondly, in order to explain the results better, it is necessary to choose a number of factors that can explain more variance of the data. Therefore, we decide to choose 4 factors as our result. This is because 4 factors explain more variance than 3 factors and the p-value of the test increase significantly from 0.58 to 0.724. On the other hand, although 5 factors can explain more variance than 4 factors, it increased the p-value by only 0.02. Thus, 4 factors seem to be the best choice. Now, let’s take a look at the results of the factor analysis.

From the Appendix Table 24 and Appendix Figure 16, we can find the loadings of FCX, MRO, X and VALE are relatively high in factor 1 than any other factors. As we mentioned before, these companies belong to copper, oil, steel, and metals industries respectively, this tells us the first factor is an index of basic materials companies. The prices of the stocks of these companies influence the financial market most. At the same time, factor2 has similar influence to the financial market. And the loadings of AMD, BAC, GE, MDT, APPLE, COKE, FORD, as well as SBUX are all relatively high than any other factors. These companies involve in all kinds of industries. But the common property of these companies is that they all relate to consumption, which means the factor2 can be regarded as a consumption index.

Then, let's take a look at factor 3; the loadings of VZ and PFE are much larger than other factors. These two companies are in the Drug Manufacturers and Telecom Services industries. Obviously, drug and telecom are two necessary things in our life, so the factor 3 can be seen as an index of rigid demands, which means people have to pay for it sometime. Finally, the factor 4, the only one loading which is larger than others is the loading of SIRI, which is a broadcasting company. We cannot define it in a clear way because we have no more relevant companies. But as a guess, we may consider it as an index of entertainment which also plays a role in people's life.

7. Copula

We compared 5 kinds of copulas: Normal copula, T copula, Clayton copula, Gumbel copula and Frank copula to find out the best model of the joint distribution of the returns. And the results can be compared by various criteria like AIC, BIC and Likelihood which are showed below:

Table 8: Copula Result

Copula	Normal	T	Clayton	Gumbel	Frank
AIC	-297	-300	-236	-244	-236
BIC	-295	-296	-234	-242	-234
Likelihood	150	152	119	123	119

From the values of AIC, we find T copula has lowest value, and BIC also provides the same result. For AIC and BIC, lower values mean that that copula fits better. And the larger the likelihood is, the better the model is. Since the likelihood is 152 for the T copula, we conclude that T copula fits better than others.

Just resonates with the fact that our data shows a slightly heavy tail correlation, t copula fits the best compared with other copulas. The t-copula emphasizes extreme results which indicate that it is good for modeling data where there is high correlation in extreme values. Also in t copula, the tail distribution is symmetric which can be seen in the density plot in Appendix Figure 17. From Appendix Figure 18, we can see that the low correlation make the random samples a little bit close to the independence case, but in the two tails we can see the correlation effects

8. Conclusions

(1) **Descriptive Statistics:** From our analysis, we find out that most of the assets are stationary. Although all of the assets pass the normality test, most of them fit t distribution better. This is because their distributions all have the bell shape and also have a fat tail. Among the correlations of 15 assets, most of them are positive and less than 0.5. The fact of low correlations indicates that constructing a portfolio of these assets can reduce overall risk.

(2) **Portfolio Theory:** The tangency portfolio without any restrictions has the highest Sharpe ratio and mean, and the lowest VaR and expected shortfall. The MVP without any restrictions has the lowest standard deviation. When added no short sales restriction, the portfolio's risk increases and return decreases, which make the portfolio less effective.

(3) **Asset Allocation:** If we want to achieve the target expected return of 0.5% per month, we need to invest around 77.2% into risk free asset, because our assets returns are relatively much larger than 0.5%. Invest in both risky assets and risk free asset will results in higher return and lower risk than just invest in risky assets.

(4) **Risk Management:** Using nonparametric method will get more accurate VaR and expected shortfall because most of our assets fit t distribution better. What's more, the asset with the highest VaR does not necessarily have the highest expected shortfall. VaR and expected shortfall depends on the asset's distribution, and expected shortfall is even more sensitive to the distribution shape.

(5) **Principal Component Analysis:** The first 3 PCs explain the 64.3% of the total variance, while the first 4 PCs explain 72.1%. Thus, it is reasonable to keep the first 3 or 4 principal components. From PCA, we can also get some sense of financial market. For example, the fluctuation of basic materials industry may cause significant change to other industries. Also healthcare and consumer goods also have huge influence to the financial market, but they influence the market from different aspect.

(6) **Copula:** Our data fits t copula best. This is mainly because t copula not only keeps the symmetric shape just like normal copula, but also emphasizes the tail extreme values, which is in accordance with the characteristics of our assets data.

APPENDIX

1. Data Collection

Table 9: Company / Asset Information

Ticker	Company Name	Industry Sector
AMD	Advanced Micro Devices	Technology
AAPL	Apple	Consumer Goods
BAC	Bank of America	Financial
COKE	Coca-Cola	Consumer Goods
FCX	Freeport-McMoRan	Basic Materials
F	Ford Motor	Consumer Goods
GE	General Electric	Industrial Goods
MDT	Medtronic plc	Healthcare
MRO	Marathon Oil Corporation	Basic Materials
PFE	Pfizer	Healthcare
SIRI	Sirius XM Holdings	Services
SBUX	Starbucks	Services
X	united states steel corporation	Basic Materials
VALE	Vale S.A.	Basic Materials
VZ	Verizon Communications	Technology

2. Descriptive Statistics

Table 10: Descriptive Statistics Summary (Not Annualized)

Assets	Mean	SD	Skewness	Kurtosis	Stationarity	Distribution	Beta	Sharpe Ratio
AMD	-0.003	0.141	0.061	0.542	Yes	T	2.078	-0.08
AAPL	0.023	0.071	0.034	-0.263	Yes	Normal	0.928	1.11
BAC	0.007	0.099	0.032	0.325	Yes	T	1.541	0.24
COKE	0.224	0.079	1.253	2.835	No	T	0.477	0.98
FCX	-0.009	0.137	-0.022	0.51	No	T	2.273	-0.24
F	0.009	0.084	0.55	0.348	No	T	1.465	0.36
GE	0.014	0.065	0.178	-0.305	Yes	Normal	1.387	0.75
MDT	0.011	0.054	-0.227	0.026	Yes	Normal	1.103	0.73
MRO	0.002	0.095	-0.318	0.76	No	T	1.579	0.07
PFE	0.012	0.047	-0.092	-0.713	Yes	Normal	0.79	0.88
SIRI	0.028	0.104	0.649	0.519	No	T	1.621	0.92
SBUX	0.027	0.058	-0.359	0.376	Yes	T	0.789	1.64
X	-0.014	0.131	-0.245	0.005	Yes	Normal	1.861	-0.37
VALE	-0.019	0.11	0.816	1.885	Yes	T	1.493	-0.60
VZ	0.013	0.047	0.34	-0.121	Yes	Normal	0.483	0.92
SP500	0.01	0.038	-0.135	0.142	Yes	T	1	0.90

Figure 3: Monthly Closing Price for 15 Assets with S&P500

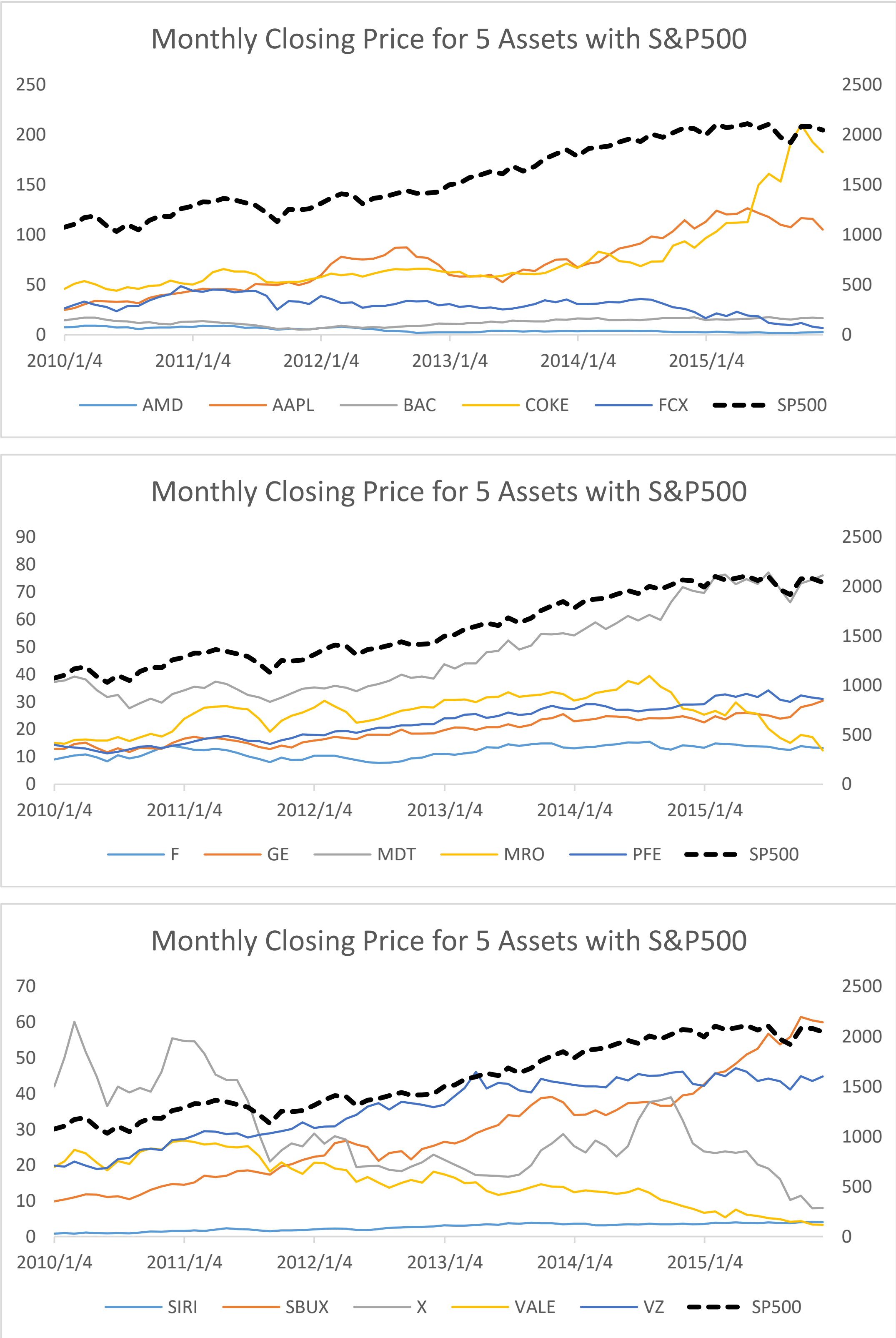


Figure 4: Monthly Net Return Plots for 15 Assets with S&P500

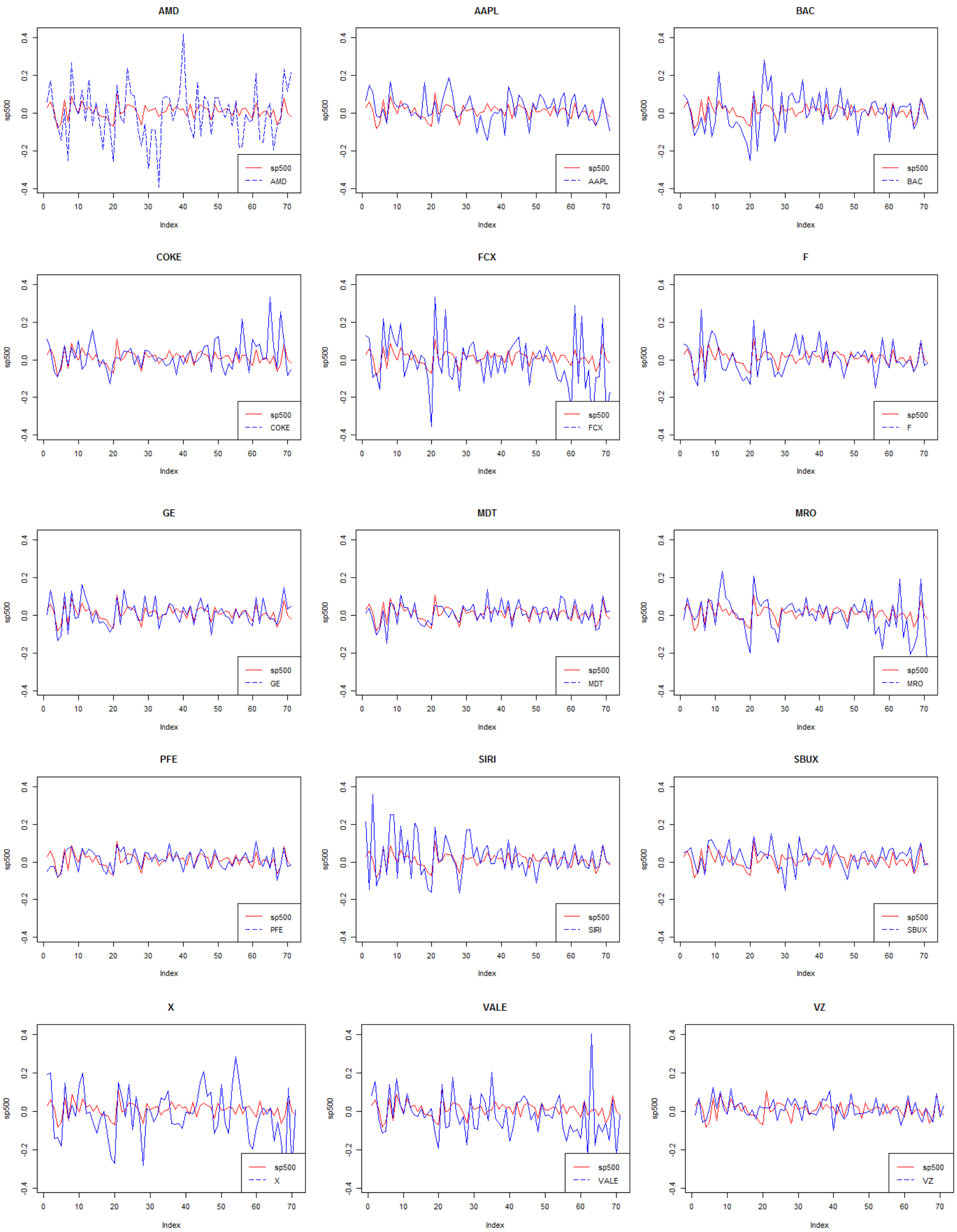


Figure 5: Equity Curve of all 15 assets with S&P500

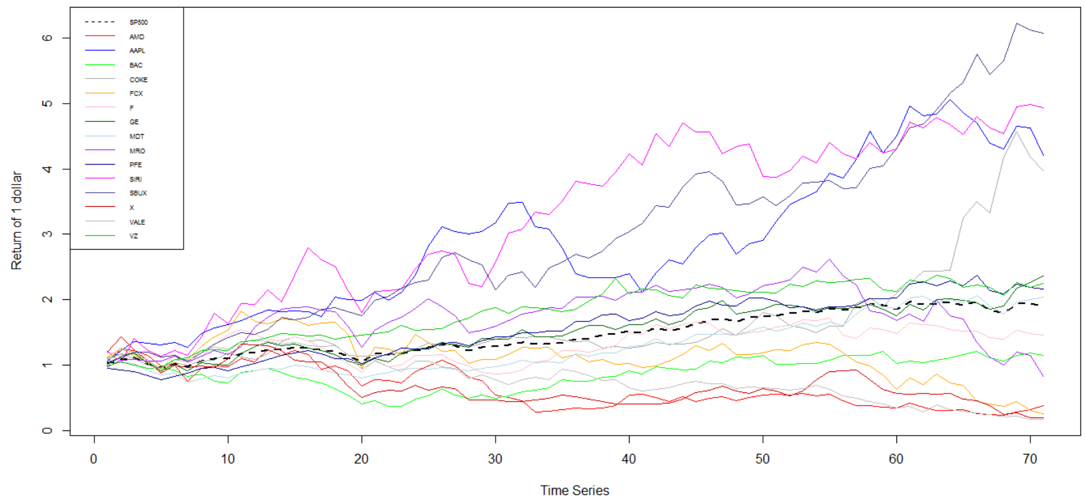
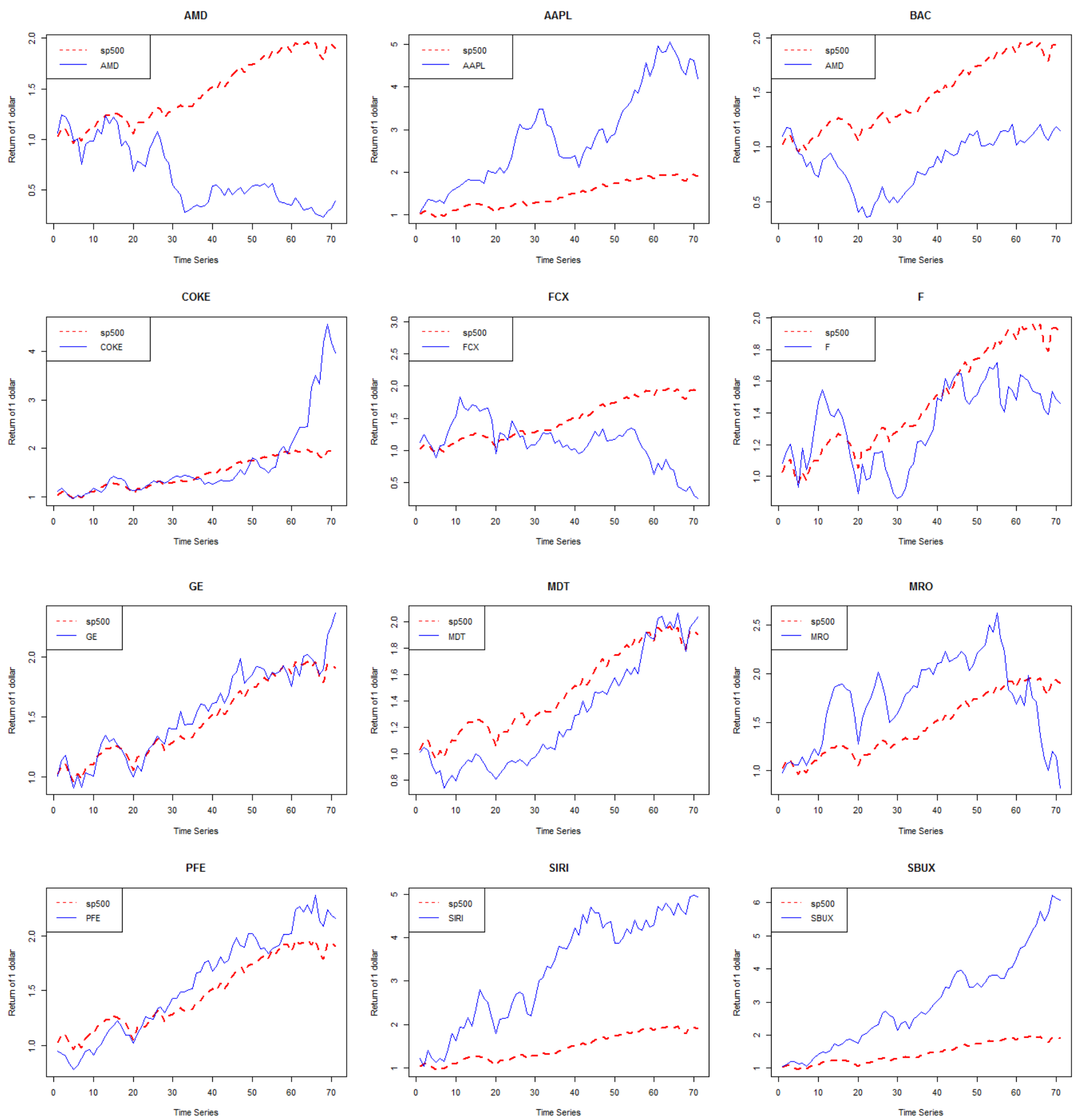


Figure 6: Equity Curve of each asset with S&P500



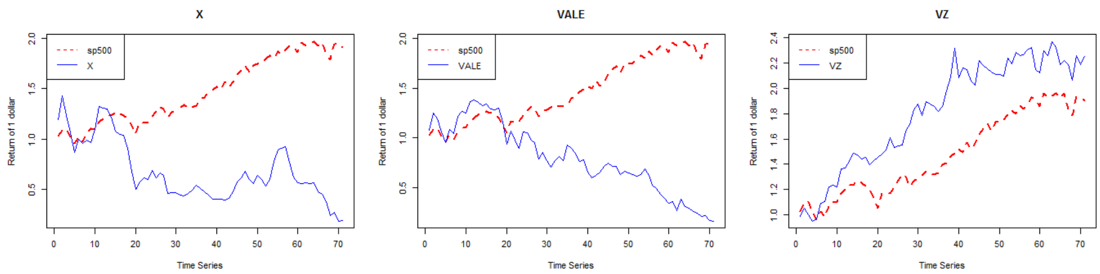


Figure 7: Histogram of 15 assets

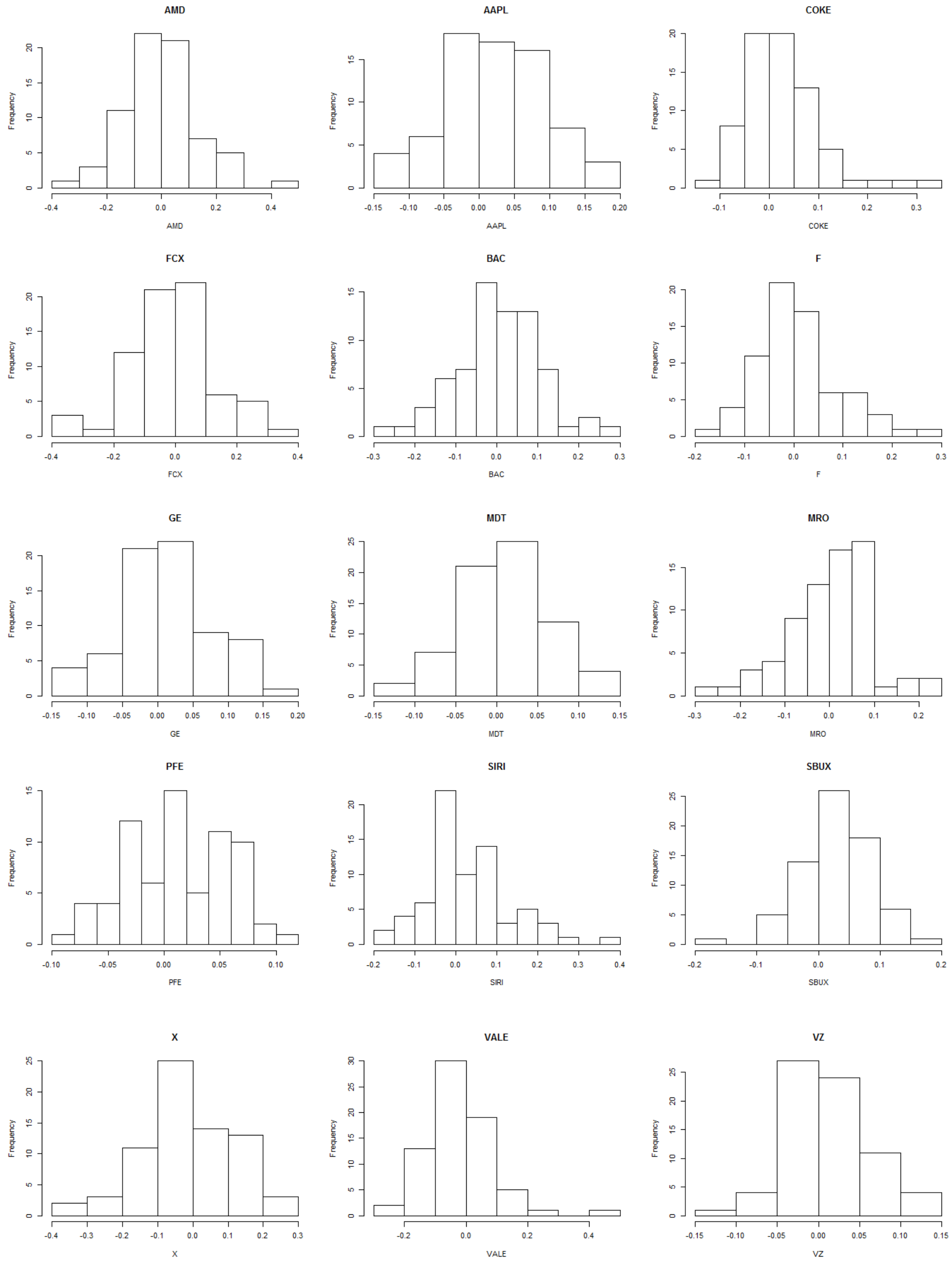
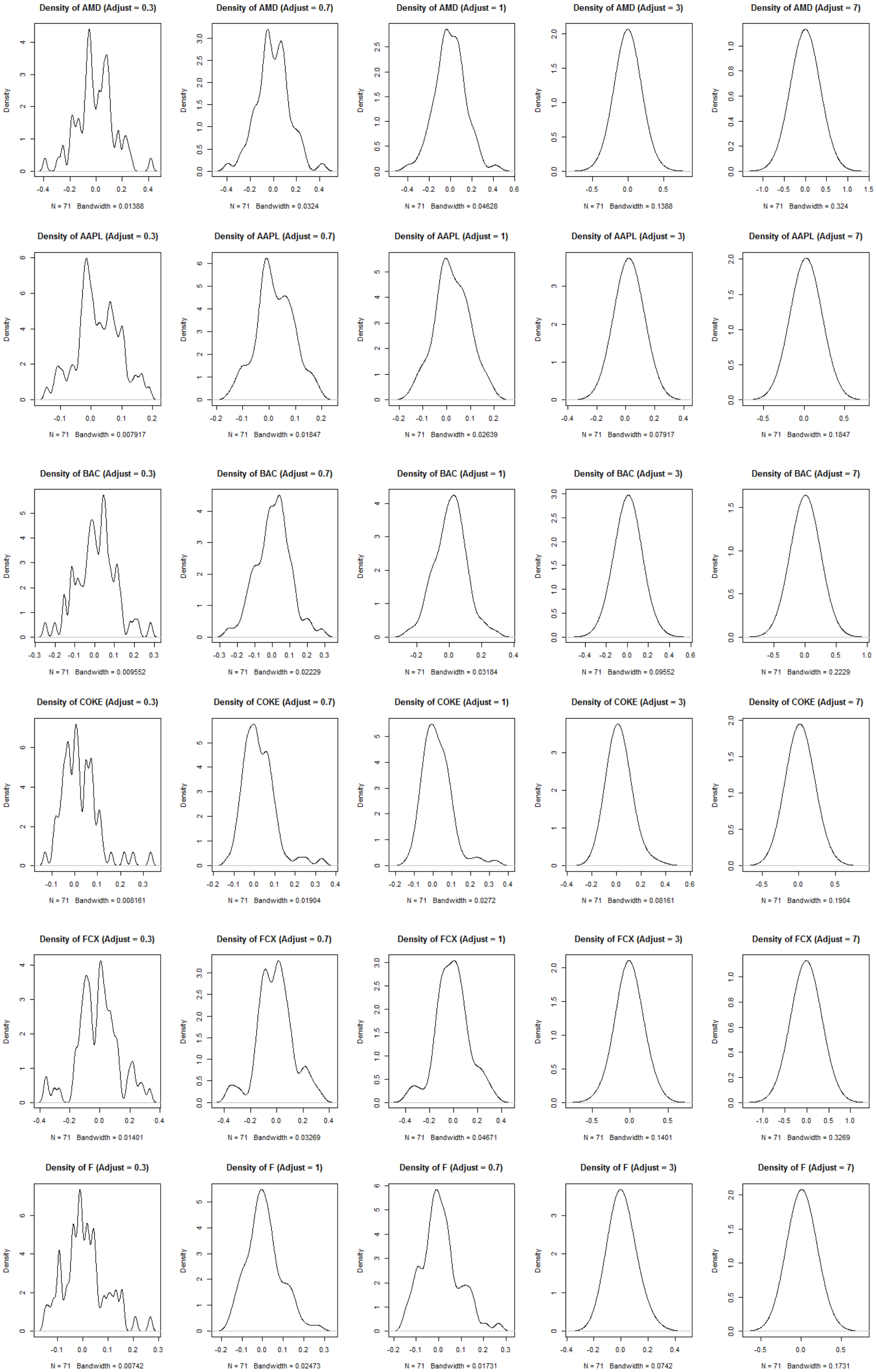
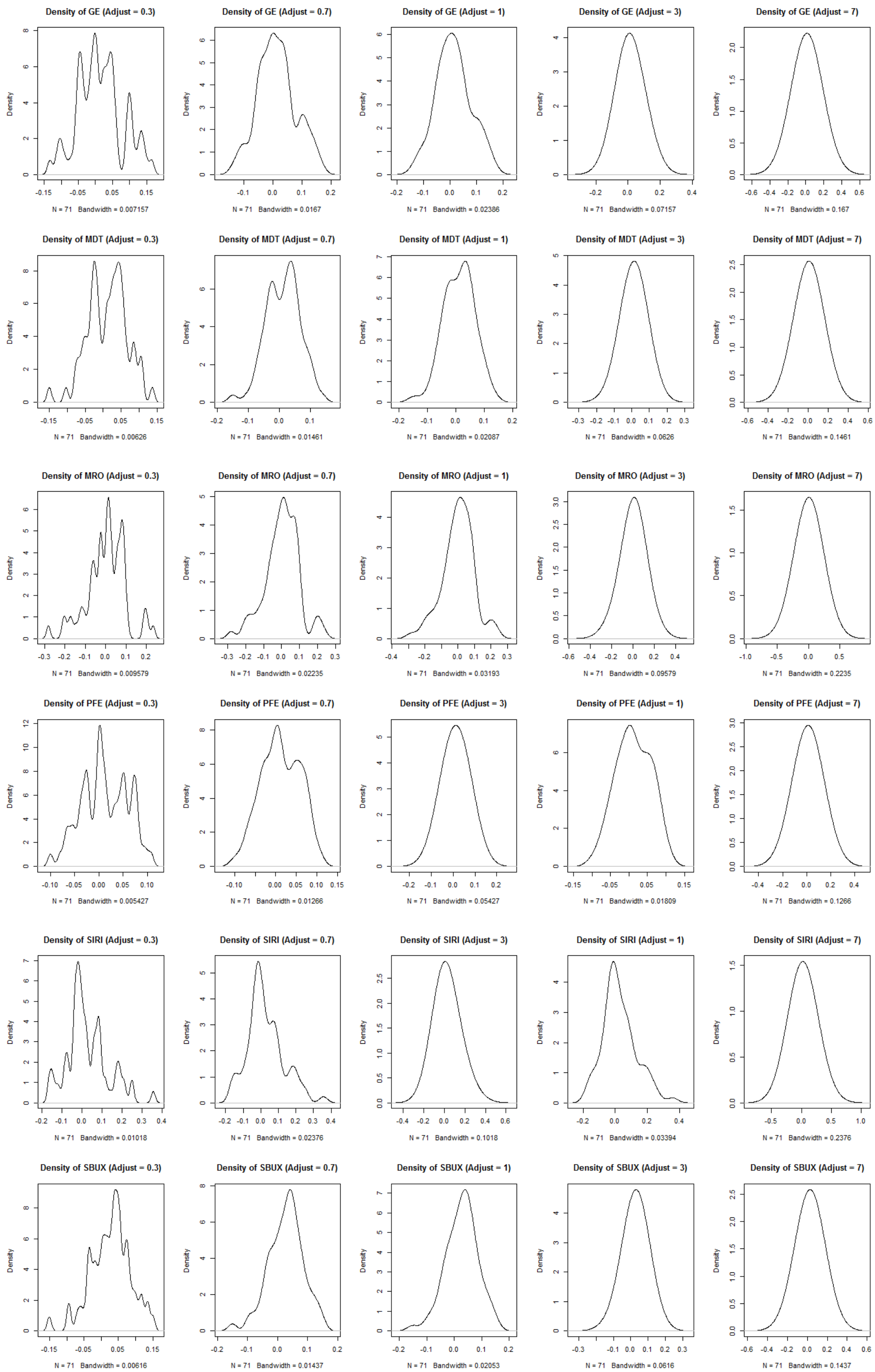


Figure 8: Density Estimation of 15 assets with 5 different bandwidths respectively




Density of GE (Adjust = 0.3)
N = 71 Bandwidth = 0.007157
Density of GE (Adjust = 0.7)
N = 71 Bandwidth = 0.0167
Density of GE (Adjust = 1)
N = 71 Bandwidth = 0.02386
Density of GE (Adjust = 3)
N = 71 Bandwidth = 0.07157
Density of GE (Adjust = 7)
N = 71 Bandwidth = 0.167
Density of MDT (Adjust = 0.3)
N = 71 Bandwidth = 0.00626
Density of MDT (Adjust = 0.7)
N = 71 Bandwidth = 0.01461
Density of MDT (Adjust = 1)
N = 71 Bandwidth = 0.02087
Density of MDT (Adjust = 3)
N = 71 Bandwidth = 0.0626
Density of MDT (Adjust = 7)
N = 71 Bandwidth = 0.1461
Density of MRO (Adjust = 0.3)
N = 71 Bandwidth = 0.009579
Density of MRO (Adjust = 0.7)
N = 71 Bandwidth = 0.02235
Density of MRO (Adjust = 1)
N = 71 Bandwidth = 0.03193
Density of MRO (Adjust = 3)
N = 71 Bandwidth = 0.09579
Density of MRO (Adjust = 7)
N = 71 Bandwidth = 0.2235
Density of PFE (Adjust = 0.3)
N = 71 Bandwidth = 0.005427
Density of PFE (Adjust = 0.7)
N = 71 Bandwidth = 0.01266
Density of PFE (Adjust = 3)
N = 71 Bandwidth = 0.05427
Density of PFE (Adjust = 1)
N = 71 Bandwidth = 0.01809
Density of PFE (Adjust = 7)
N = 71 Bandwidth = 0.1266
Density of SIRI (Adjust = 0.3)
N = 71 Bandwidth = 0.01018
Density of SIRI (Adjust = 0.7)
N = 71 Bandwidth = 0.02376
Density of SIRI (Adjust = 3)
N = 71 Bandwidth = 0.1018
Density of SIRI (Adjust = 1)
N = 71 Bandwidth = 0.03394
Density of SIRI (Adjust = 7)
N = 71 Bandwidth = 0.2376
Density of SBUX (Adjust = 0.3)
N = 71 Bandwidth = 0.00616
Density of SBUX (Adjust = 0.7)
N = 71 Bandwidth = 0.01437
Density of SBUX (Adjust = 1)
N = 71 Bandwidth = 0.02053
Density of SBUX (Adjust = 3)
N = 71 Bandwidth = 0.0616
Density of SBUX (Adjust = 7)
N = 71 Bandwidth = 0.1437
Density

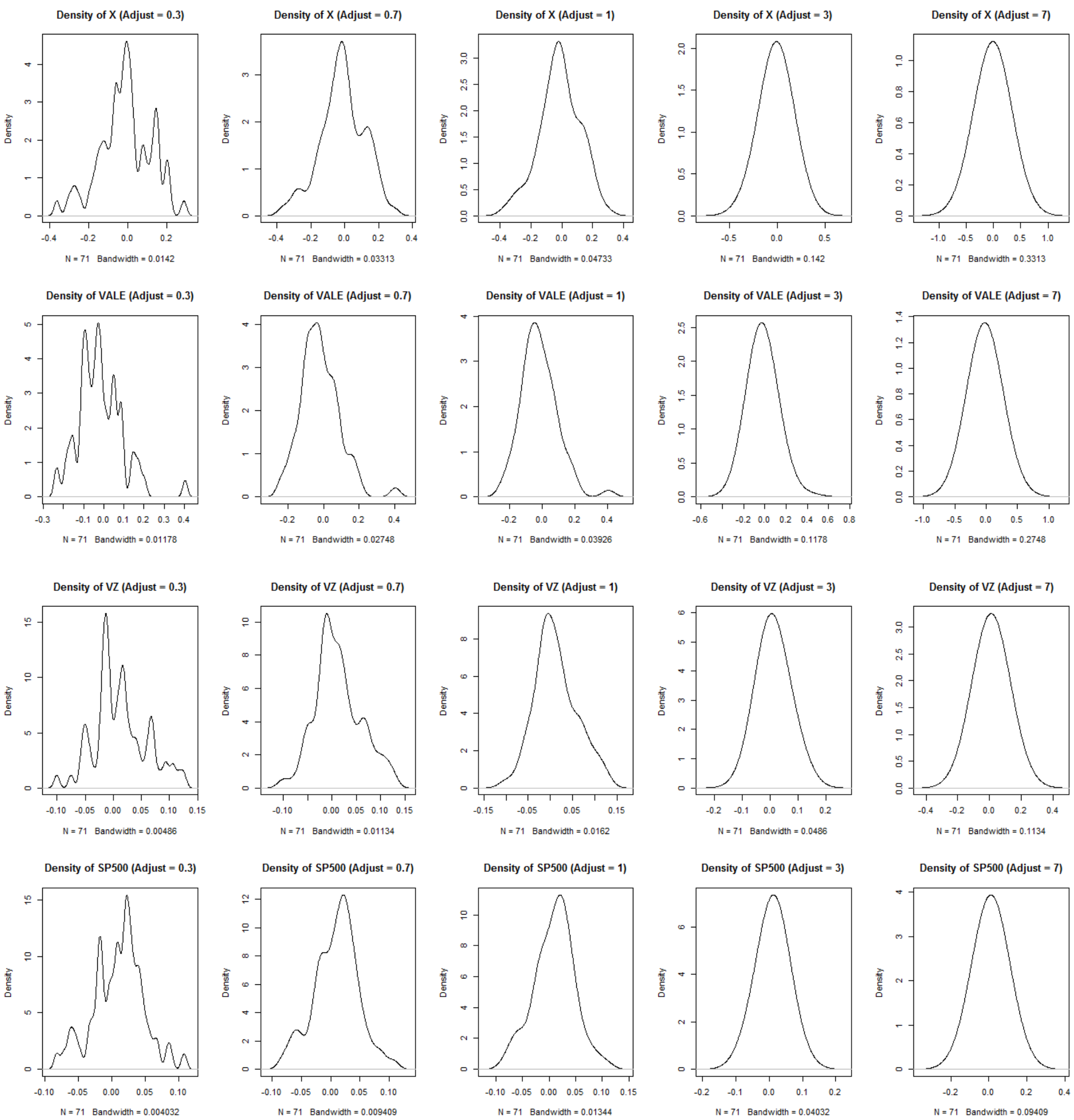


Table 11: Assets Outliers

Company	Date	Return Value	Company	Date	Return Value
AMD	Oct. 2012	-0.392	F	Jul. 2010	0.267
	May. 2013	0.418		Oct. 2011	0.208
BAC	Sep. 2011	-0.251	MDT	Aug. 2010	-0.149
	Jan. 2012	0.282	MRO	Jan. 2011	0.234
COKE	Oct. 2014	0.215		Dec. 2015	-0.281
	Jun. 2015	0.33	SIRI	Apr. 2010	0.356
	Sep. 2015	0.254	SBUX	Jul. 2012	-0.151
FCX	Sep. 2011	-0.354	X	Sep. 2015	-0.364
	Oct. 2011	0.332	VALE	Apr. 2015	0.404
	Jul. 2015	-0.363	VZ	Jul. 2010	0.126

Table 12: Covariance Matrix

	AMD	AAPL	BAC	COKE	FCX	F	GE	MDT	MRO	PFE	SIRI	SBUX	X	VALE	VZ
AMD	0.0200	0.0031	0.0057	0.0007	0.0071	0.0054	0.0046	0.0035	0.0044	0.0010	0.0045	0.0039	0.0049	0.0037	0.0001
AAPL		0.0051	0.0019	0.0009	0.0034	0.0016	0.0019	0.0010	0.0022	0.0003	0.0022	0.0015	0.0020	0.0019	0.0003
BAC			0.0097	0.0006	0.0044	0.0040	0.0033	0.0022	0.0033	0.0013	0.0036	0.0018	0.00055	0.0039	0.0000
COKE				0.0062	0.0012	0.0010	0.0010	0.0007	0.0006	0.0006	0.0002	0.0009	0.0010	0.0003	-0.0001
FCX					0.0188	0.0063	0.0046	0.0023	0.0087	0.0019	0.0052	0.0026	0.0104	0.0109	0.0021
F						0.0070	0.0026	0.0021	0.0032	0.0011	0.0046	0.0022	0.0047	0.0043	0.0005
GE							0.0043	0.0023	0.0032	0.0015	0.0024	0.0014	0.0034	0.0034	0.0014
MDT								0.0030	0.0020	0.0016	0.0024	0.0013	0.0025	0.0009	0.0008
MRO									0.0090	0.0016	0.0039	0.0015	0.0053	0.0057	0.0010
PFE										0.0022	0.0017	0.0009	0.0018	0.0009	0.0011
SIRI											0.0108	0.0019	0.0030	0.0037	0.0000
SBUX												0.0033	0.0021	0.0016	0.0003
X													0.0173	0.0079	0.0016
VALE														0.0120	0.0019
VZ															0.0022

Table 13: Correlation Matrix

	AMD	AAPL	BAC	COKE	FCX	F	GE	MDT	MRO	PFE	SIRI	SBUX	X	VALE	VZ
AMD	1.000	0.311	0.411	0.063	0.365	0.456	0.503	0.452	0.328	0.148	0.304	0.482	0.263	0.239	0.009
AAPL		1.000	0.267	0.160	0.350	0.270	0.410	0.260	0.322	0.099	0.303	0.354	0.211	0.249	0.092
BAC			1.000	0.083	0.325	0.479	0.518	0.411	0.352	0.288	0.356	0.310	0.425	0.365	0.002
COKE				1.000	0.115	0.157	0.194	0.167	0.080	0.153	0.023	0.209	0.099	0.035	-0.018
FCX					1.000	0.550	0.512	0.306	0.664	0.289	0.365	0.332	0.576	0.727	0.320
F						1.000	0.473	0.472	0.396	0.281	0.533	0.459	0.425	0.471	0.123
GE							1.000	0.651	0.512	0.490	0.353	0.377	0.396	0.475	0.460
MDT								1.000	0.380	0.616	0.432	0.406	0.346	0.149	0.322
MRO									1.000	0.356	0.394	0.286	0.422	0.550	0.213
PFE										1.000	0.356	0.325	0.289	0.171	0.476
SIRI											1.000	0.310	0.218	0.326	0.009
SBUX												1.000	0.273	0.246	0.109
X													1.000	0.545	0.262
VALE														1.000	0.361
VZ															1.000

Figure 9: Correlation Matrix Heatmap

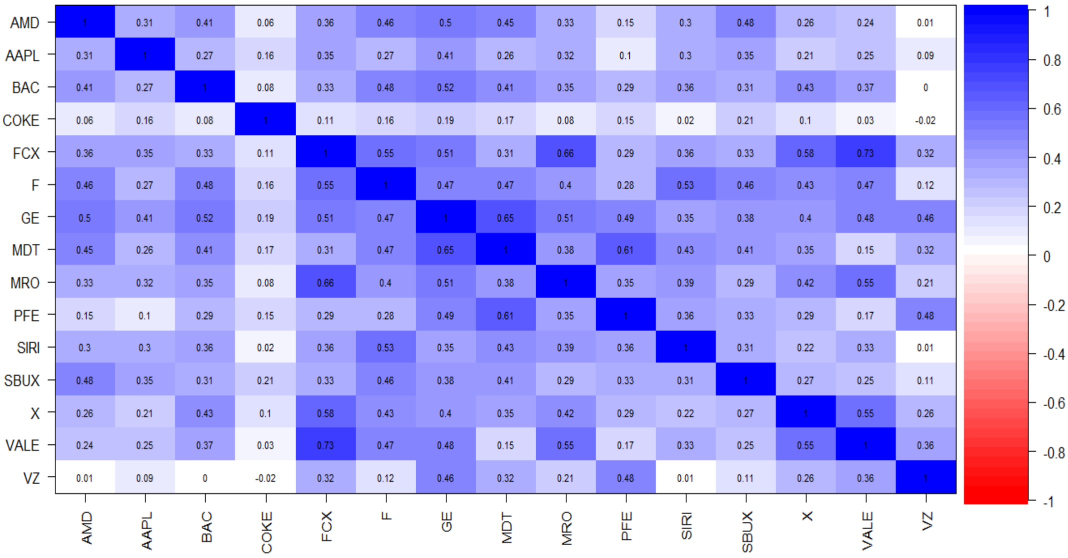


Figure 10: Correlation Plot with Number (left) and Correlation Plot with Significance Test (right)

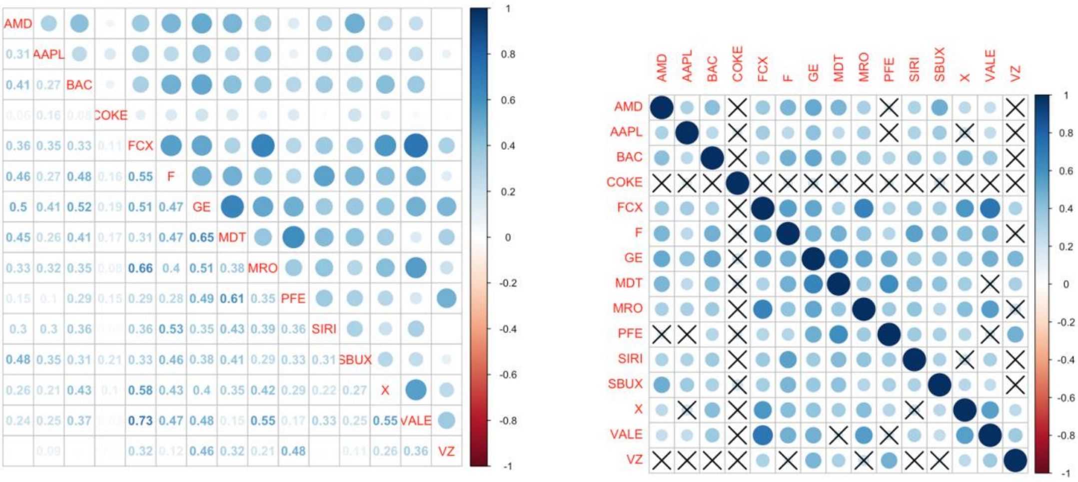


Figure 11: Pairwise Scatter Plots

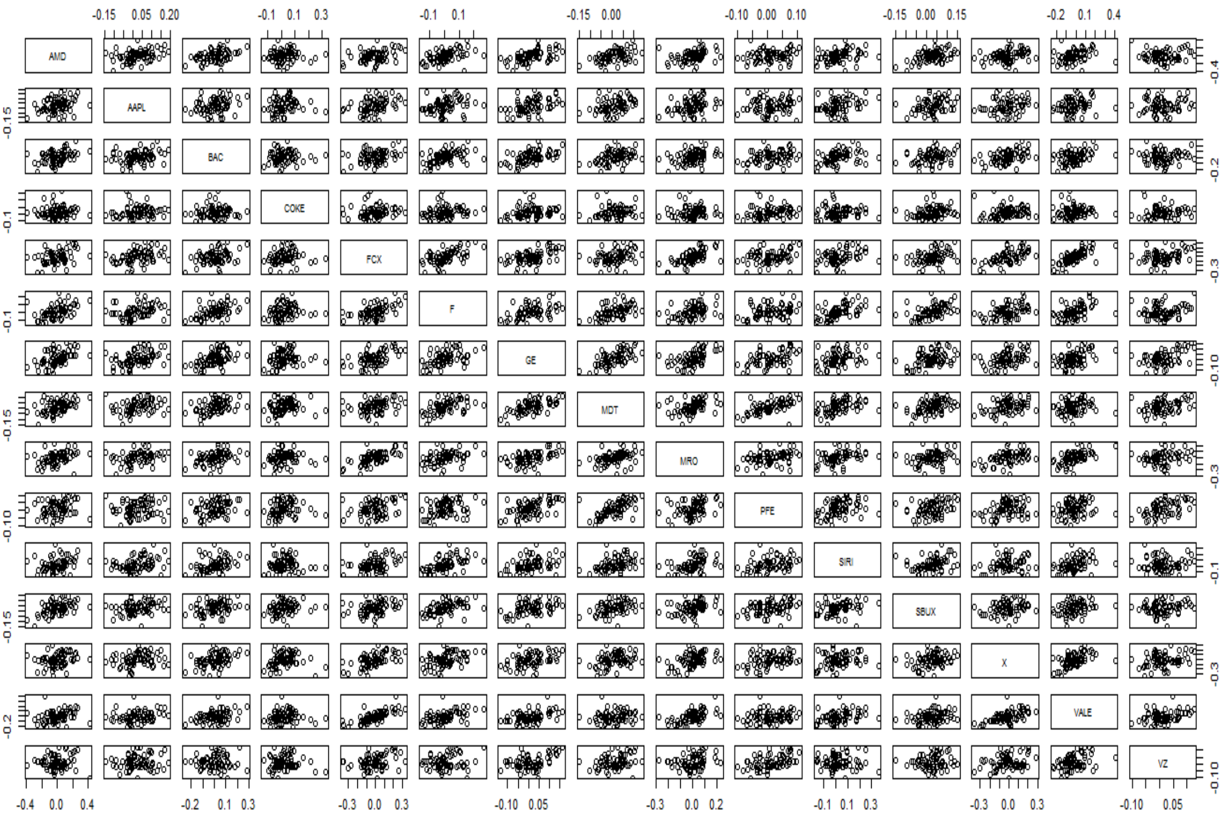
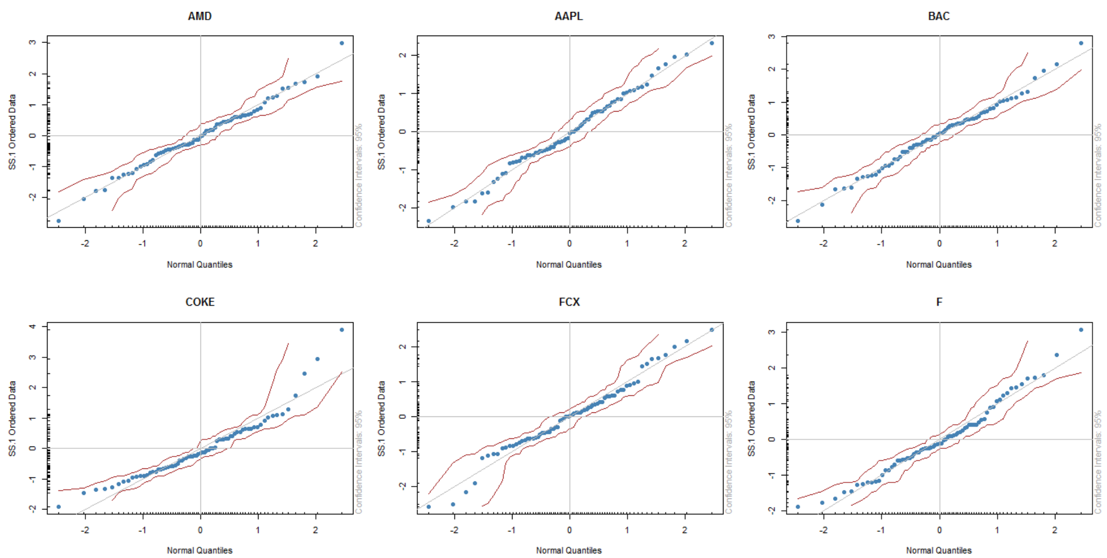
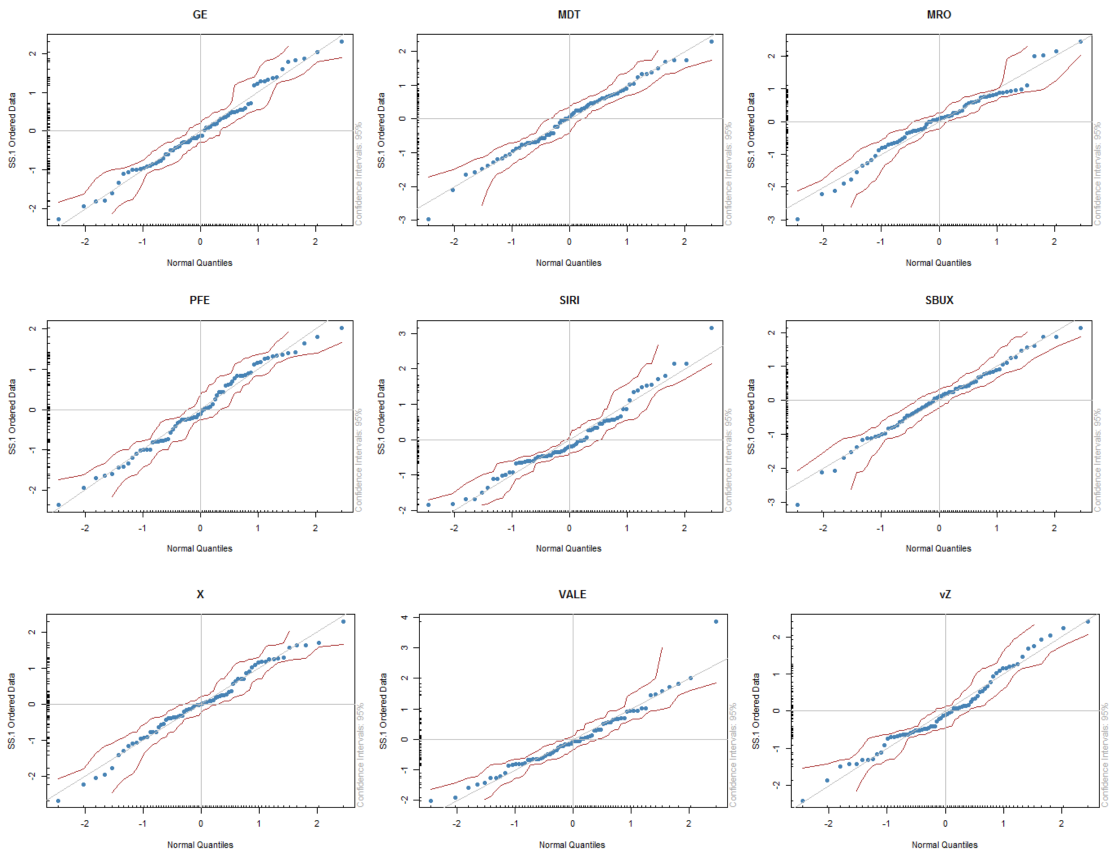


Figure 12: QQ Plot





3. Portfolio Theory

Figure 13: Efficient Portfolio Frontier when Short Sales are allowed

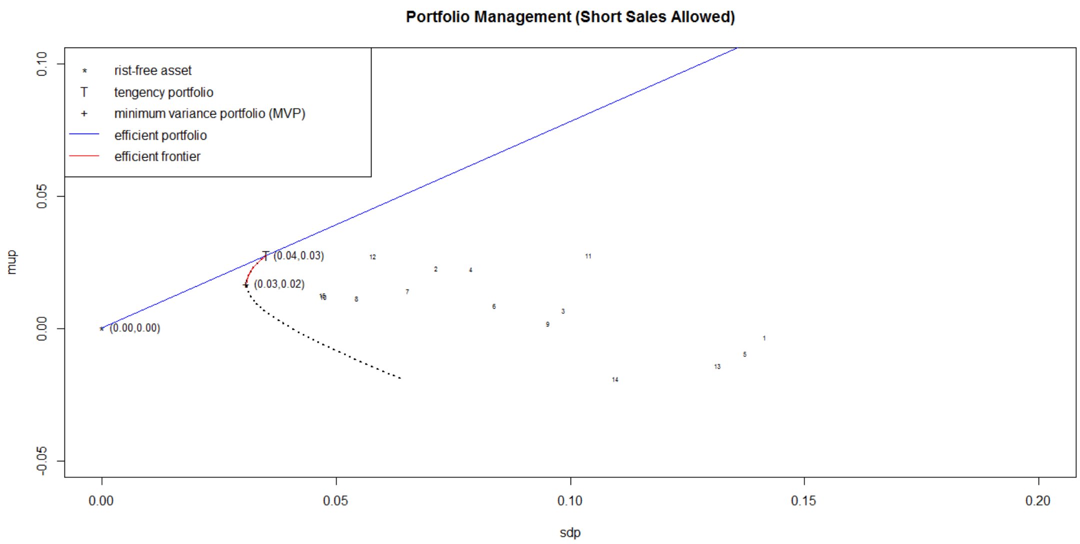


Table 14: Minimum Variance Portfolio with Short Sales (Annualized)

Mean	**SD**	**VaR**	**ES**	**CI of VaR**	**CI of ES**	**Sharpe**
0.203	0.107	3232	4078	(2476, 4955)	(3191, 5382)	1.89
	AMD	**AAPL**	**BAC**	**COKE**	**FCX**	
Weight	0.89%	13.96%	6.74%	12.98%	-9.79%	
SD	0.49	0.25	0.34	0.27	0.48	
	F	**GE**	**MDT**	**MRO**	**PFE**	
Weight	4.08%	-21.51%	9.71%	5.19%	15.94%	
SD	0.29	0.23	0.19	0.33	0.16	
	SIRI	**SBUX**	**X**	**VALE**	**VZ**	
Weight	0.85%	12.97%	-3.14%	5.02%	46.12%	
SD	0.36	0.20	0.46	0.38	0.16	

Table 15: Tangency Portfolio with Short Sales (Annualized)

Mean	**SD**	**VaR**	**ES**	**CI of VaR**	**CI of ES**	**Sharpe**
0.332	0.121	2708	3162	(2101, 3835)	(2938, 3858)	2.73
	AMD	**AAPL**	**BAC**	**COKE**	**FCX**	
Weight	-5.33%	13.56%	7.39%	14.22%	-7.53%	
Sharpe ratio	-0.08	1.11	0.24	0.98	-0.24	
	F	**GE**	**MDT**	**MRO**	**PFE**	
Weight	1.56%	-5.01%	-5.41%	4.63%	-0.87%	
Sharpe ratio	0.36	0.75	0.73	0.07	0.88	
	SIRI	**SBUX**	**X**	**VALE**	**VZ**	
Weight	10.62%	32.83%	-3.83%	-10.54%	53.71%	
Sharpe ratio	0.92	1.64	-0.37	-0.60	0.92	

Figure 14: Efficient Portfolio Frontier when Short Sales are Not Allowed

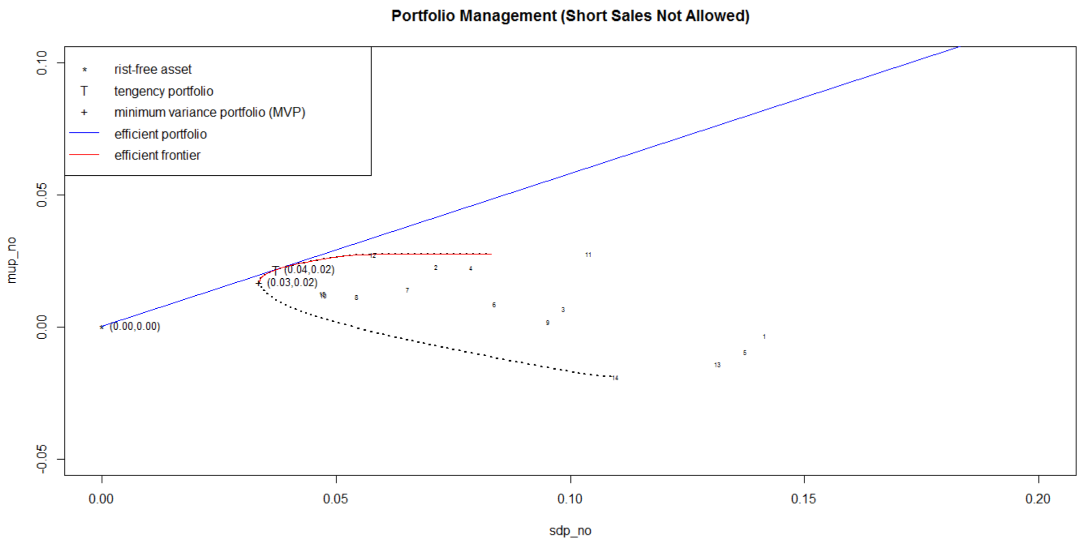


Table 16: Minimum Variance Portfolio without Short Sales (Annualized)

Mean	SD	VaR	ES	CI of VaR	CI of ES	Sharpe
0.203	0.116	3718	4582	(2509, 5302)	(3831, 6051)	1.74

	AMD	AAPL	BAC	COKE	FCX
Weight	0.00%	10.34%	2.75%	12.70%	0.00%
SD	0.49	0.25	0.34	0.27	0.48
	F	**GE**	**MDT**	**MRO**	**PFE**
Weight	0.00%	0.00%	2.09%	0.00%	17.87%
SD	0.29	0.23	0.19	0.33	0.16
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
Weight	1.11%	14.60%	0.00%	0.00%	38.54%
SD	0.36	0.20	0.46	0.38	0.16

Table 17: Tangency Portfolio without Short Sales (Annualized)

Mean	SD	VaR	ES	CI of VaR	CI of ES	Sharpe
0.261	0.129	3995	4495	(3828, 5305)	(3324, 5278)	2.02

	AMD	AAPL	BAC	COKE	FCX
Weight	0.00%	10.24%	0.00%	16.07%	0.00%
Sharpe ratio	-0.08	1.11	0.24	0.98	-0.24
	F	**GE**	**MDT**	**MRO**	**PFE**
Weight	0.00%	0.00%	0.00%	0.00%	0.00%
Sharpe ratio	0.36	0.75	0.73	0.07	0.88
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
Weight	7.61%	36.40%	0.00%	0.00%	29.69%
Sharpe ratio	0.92	1.64	-0.37	-0.60	0.92

4. Asset Allocation

Table 18: Asset Allocation with only Risky Assets (Annualized)

Mean	SD	VaR	ES	CI of VaR	CI of ES	Sharpe
0.06	0.15	6009	7187	(4069, 7320)	(6205, 8610)	0.39

	AMD	AAPL	BAC	COKE	FCX
Weight	4.58%	0.00%	0.00%	5.02%	0.00%
Mean	-0.04	0.27	0.08	0.27	-0.11
	F	**GE**	**MDT**	**MRO**	**PFE**
Weight	0.00%	0.00%	12.34%	0.00%	31.56%
Mean	0.10	0.17	0.14	0.02	0.14
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
Weight	0.00%	0.00%	0.41%	21.96%	24.13%
Mean	0.33	0.33	-0.17	-0.23	0.15

Table 19: Asset Allocation with Risky Assets and Risk Free Asset (Annualized)

Mean	SD	VaR	ES	CI of VaR	CI of ES	Sharpe
0.06	0.03	906	1020	(868, 1204)	(753, 1197)	2.02

	AMD	AAPL	BAC	COKE	FCX
Weight	0.00%	2.33%	0.00%	3.66%	0.00%
Mean	-0.04	0.27	0.08	0.27	-0.11
	F	**GE**	**MDT**	**MRO**	**PFE**
Weight	0.00%	0.00%	0.00%	0.00%	0.00%
Mean	0.10	0.17	0.14	0.02	0.14
	SIRI	**SBUX**	**X**	**VALE**	**VZ**
Weight	1.73%	8.29%	0.00%	0.00%	6.76%
Mean	0.33	0.33	-0.17	-0.23	0.15
	Risk Free				
Weight	77.20%				
Mean	0.0007				

Table 20: Comparison between Different Portfolios and 15 Assets (Annualized)

	Mean	SD	VaR	ES	Sharpe Ratio
AMD	-0.04	0.49	22342	29759	-0.08
AAPL	0.27	0.25	10021	11942	1.11
BAC	0.08	0.34	15286	19090	0.24
COKE	0.27	0.27	8057	9722	0.98
FCX	-0.11	0.48	22143	32299	-0.24
F	0.10	0.29	11524	13436	0.36
GE	0.17	0.23	9554	11268	0.75
MDT	0.14	0.19	7162	10116	0.73
MRO	0.02	0.33	17236	21636	0.07
PFE	0.14	0.16	6461	7826	0.88
SIRI	0.33	0.36	13702	15470	0.92
SBUX	0.33	0.20	6590	10219	1.64
X	-0.17	0.46	25747	30519	-0.37
VALE	-0.23	0.38	17746	20948	-0.60
VZ	0.15	0.16	5414	7190	0.92
MVP (no short sale)	0.203	0.116	3178	4582	1.74
Tangent (no short sale)	0.261	0.129	3995	4495	2.02
MVP (with short sale)	0.203	0.107	3232	4078	1.89
Tangent (with short sale)	0.332	0.121	2708	3162	2.73
No T-bill with 6% return	0.06	0.15	6009	7187	0.39
With T-bill with 6% return	0.06	0.03	906	1020	2.02

5. Risk Management

Table 21: Standard Error, VaR and Confidence Interval based on Bootstrap

	SE	VaR	Lower Bound	Upper Bound
AMD	3978	22342	15539	28152
AAPL	1751	10021	8212	14541
BAC	2336	15286	10378	19242
COKE	1110	8057	6866	10815
FCX	6823	22143	8923	31101
F	1451	11524	8981	13842
GE	1896	9554	7863	14065
MDT	1398	7162	4001	9388
MRO	3116	17236	13636	24089
PFE	910	6461	4971	8726
SIRI	2426	13702	11293	19807
SBUX	1989	6590	3695	9578
X	4487	25747	20917	35916
VALE	2280	17746	12785	22861
VZ	868	5414	3296	6847

Table 22: Standard Error, Expected Shortfall and Confidence Interval based on Bootstrap

	SE	ES	Lower Bound	Upper Bound
AMD	4988	29759	20349	39235
AAPL	1449	11942	9475	15242
BAC	3093	19090	13087	24566
COKE	1478	9722	6584	12355
FCX	5070	32299	28316	45741
F	1196	13436	11823	16313
GE	1408	11268	9228	14884
MDT	2284	10116	5300	13723
MRO	3299	21636	15173	28269
PFE	1083	7826	5702	9843
SIRI	1336	15470	14569	19638
SBUX	2435	10219	5360	14812
X	3671	30519	24652	39340
VALE	2087	20948	18042	25424
VZ	1392	7190	4308	9276

6. Principal Component Analysis and Factor Analysis

Table 23: PCA loadings and explained variance

	PC1	PC2	PC3	PC4	PC5
AMD	-0.223	0.391	0.016	-0.221	-0.278
AAPL	-0.062	0.186	0.245	-0.644	0.263
BAC	-0.284	0.239	-0.027	0.191	-0.313
COKE	0.332	0.072	0.167	0.162	-0.232
FCX	-0.388	-0.257	0.151	-0.050	0.033
F	-0.346	0.221	0.053	0.319	0.009
GE	-0.255	0.025	-0.369	-0.379	-0.219
MDT	-0.089	0.243	-0.516	0.001	-0.035
MRO	-0.365	-0.160	0.015	-0.100	0.222
PFE	0.033	-0.066	-0.564	0.157	0.149
SIRI	-0.239	0.254	-0.045	0.286	0.644
SBUX	-0.063	0.380	0.016	-0.134	-0.134
X	-0.302	-0.243	0.062	-0.384	-0.384
VALE	-0.354	-0.325	0.179	-0.158	-0.016
VZ	0.031	-0.415	-0.356	-0.014	-0.014
Proportion of Variance	**0.276**	**0.196**	**0.171**	**0.078**	**0.067**
Cumulative Proportion	**0.276**	**0.473**	**0.643**	**0.721**	**0.788**

Figure 15: PCA Scree Plot

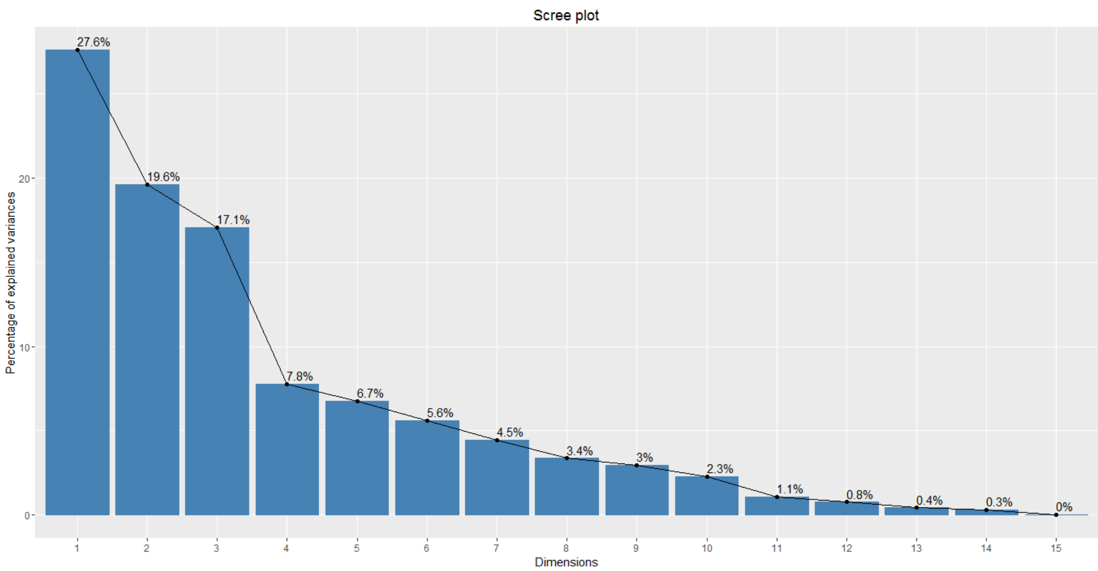
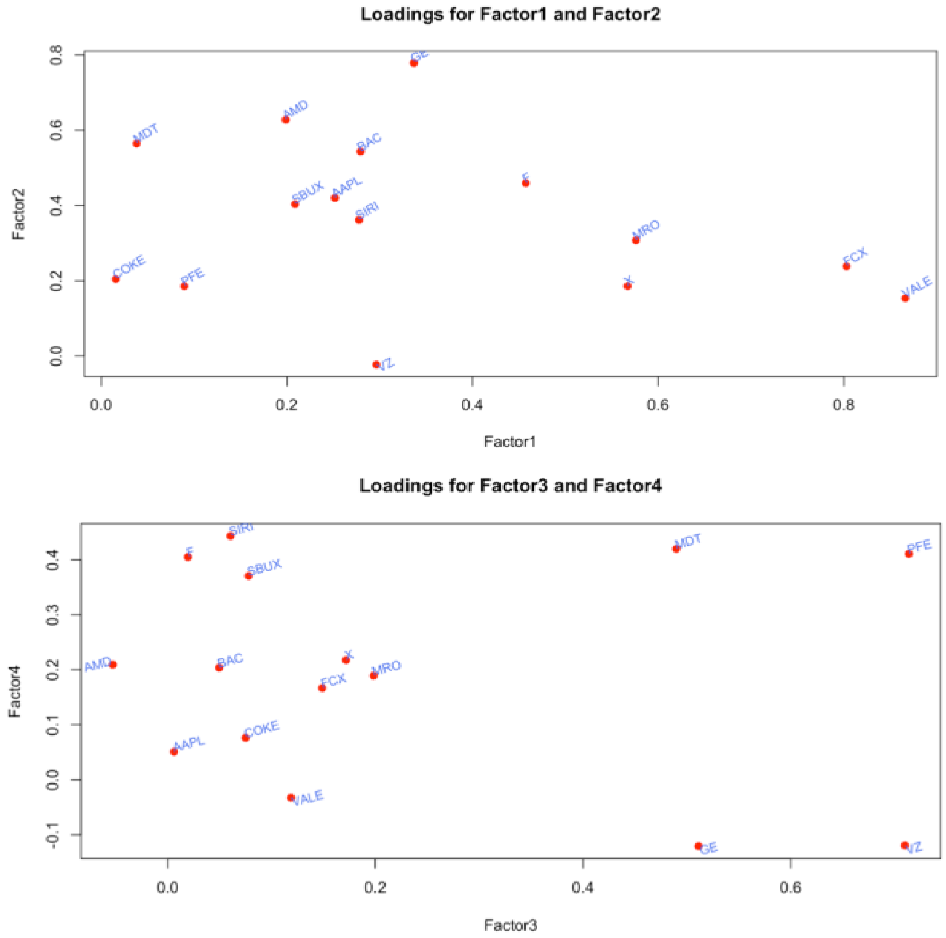


Table 24: Factor Analysis Result

	FA1	FA2	FA3	FA4	FA5
FCX	0.803	0.238	0.149	0.167	-0.278
MRO	0.576	0.307	0.198	0.189	0.263
X	0.567	0.186	0.172	0.218	-0.313
VALE	0.866	0.153	0.119		-0.232
AMD	0.199	0,628		0.209	0.033
BAC	0.279	0.543		0.204	0.009
GE	0.337	0.778	0.511	-0.121	-0.219
MDT		0.564	0.490	0.420	-0.035
PFE		0.185	0.714	0.411	0.222
VZ	0.296		0.710	-0.119	0.149
AAPL	0.252	0.410			0.644
COKE		0.204			-0.134
F	0.457	0.459		0.405	-0.384
SIRI	0.278	0.361		0.443	-0.016
SBUX	0.209	0.403		0.370	-0.014
Proportion of Variance	**0.185**	**0.172**	**0.109**	**0.072**	**0.067**
Cumulative Proportion	**0.185**	**0.357**	**0.466**	**0.538**	**0.788**

Figure 16: Factor Loadings Plot



7. Copula

Figure 17: Density Visualization of t Copula

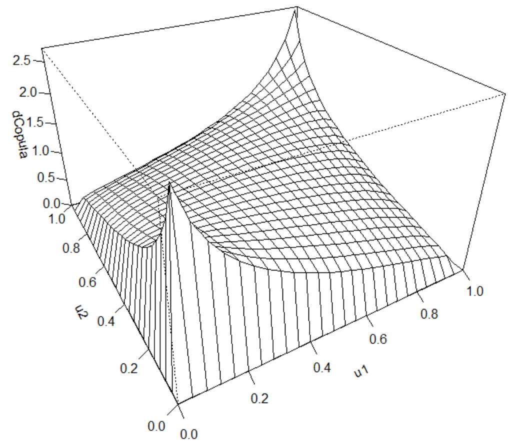


Figure 18: Random Samples derived from t Copula