STAT 4290 Final Project Report

Statistical and Data Analysis For Financial Engineering

Jiaqi Chen(jc4260), Po Hu(ph2438), Yunjuan Liu(yl3322), Yuwen Zhai(yz2702)

Jiaqi Chen(Descriptive Statistics, Portfolio Theory and Asset Allocation)
Po Hu(Classifying Technical Indicators and analysis, Principal Component Analysis, Risk Management)
Yunjuan Liu(Descriptive Statistics, Portfolio Theory and Asset Allocation)
Yuwen Zhai(Principal Component Analysis, Risk management, Copulas)

1. Introduction

Any investment has to go through two main steps. The first step begins with observation, predicting, and analyzing securities and ends by making certain beliefs about the future performance of these securities. The second step starts with the creation of trust about future performance and ends with the allocation of a portfolio.
In the current project, we initially selected 10 stocks covered fields in financial industrial, actuarial, internet and manufacture from Yahoo Finance and looked for their monthly performance from January 1st, 2004 to December 31th, 2014. Our assets include AXP, GE, AAPL, IBM, INTC, MSFT, JPM, PG, AIG and GS. In order to construct efficient portfolio, we use 4-week Treasury bill of secondary market as our risk free asset.

2. Summary

We applied clustering method to solve the technical indicators classification problem. Through conducting K-Medoid clustering method, we classified 12 indicators into 3 and 8 groups respectively, and we used the results to construct two portfolios. The first portfolio contains three assets: GE, INTC and PG. The second portfolio contains four assets: GE, AAPL, MSFT and PG. By using portfolio theory (Markowitz), we got the sharp ratio for the first portfolio is 0.17, and sharp ratio for the second portfolio is 0.37. We find the efficient frontier, settle down the minimum variance portfolio and the tangency portfolio under the constraint of allowing short-selling or no short-selling. We also construct the certain portfolio under a given expect return. Then we make a comparison of different portfolio and find that the tangency portfolio with short-sale has the highest sharp ratio, but its risk is also relatively high. From the risk management analysis, the PG asset is the least risky asset with the smallest expected loss and the AIG is the most risky asset with largest expected loss for our portfolio. Using the copulas to model the joint distribution of return, the meta t-distribution copulas with empirical distribution of data is the best fit for our data based on AIC.

3. Descriptive Statistics

	AXP	**GE**	**AAPL**	**IBM**	**INTC**
mean	1.136	0.384	3.856	0.649	0.620
variance	112.000	61.400	106.900	28.700	53.100
	MSFT	**JPM**	**PG**	**AIG**	**GS**
Mean	0.855	0.927	0.762	0.919	0.987
Variance	42.400	68.000	18.300	836.000	77.200

TABLE 3.1 MEAN AND VARIANCE.

From table 3.1, it can be seen that the apple stock have the highest return, which is 3.856%, while stock AIG have the highest variance, which is 8.36%.

	PG	**MSFT**	**JPM**	**INTC**	**IBM**
skewness	-0.1468	0.0876	-0.1926	-0.2021	-0.6674
	GS	**GE**	**AXP**	**AIG**	**AAPL**
skewness	-0.1845	-0.3369	0.1194	5.0685	-0.3455

TABLE 3.2 THE SKEWNESS OF 10 ASSETS

From the table 3.2, the skewness of the stocks PG, MSFT, JPM, INTC, IBM, GS, GE, AXP, AAPL is around zero , which is a good evidence of following normal distribution. The skewness of stock AIG is 5.06852118, which indicates that there is a right skew in the data.

	PG	MSFT	JPM	INTC	IBM
kurtosis	0.2127	0.9630	0.9101	0.0390	1.7210
	GS	**GE**	**AXP**	**AIG**	**AAPL**
kurtosis	0.5146	2.1236	7.8046	40.6426	1.3608

TABLE 3.3 THE KURTOSIS OF 10 ASSETS

From Table 3.3, the kurtosis of the stock PG, MSFT, JPM, INTC, IBM, GS, GE,AAPL is between 0 and 3, which indicates that this is a slight peak in the distribution, while for AXP and AIG, the skewness is 7.804 and 40.64262, which shows a high peak in the distribution of returns of stocks, shows an evidence that AXP and AIG do not follow a normal distribution well.

From Figure 3.1 in Appendix, we can see that most of the returns of stocks fluctuate significantly during the time of financial crisis. And most of the returns achieved their lowest return during that time. It can also be seen that the returns of AXP and AIG stay very stable on the none global financial crisis period. And for other stocks, the returns fluctuate more randomly and intensely compared with AXP and AIG.

Figure 3.2 in Appendix is the QQ plot of the returns of stock. It can be seen that the plot for PG, MSFT, JPM INTC, IBM, GS, and AAPL is nearly a straight line, which follows the assumption of normal distribution. While for GE, the line is light tailed, and for AXP and AIG, especially for AIG, the line is of “S” shape, which is strong evidence against normality.

From Figure 3.3 in Appendix, it can be seen that returns of stocks MSFT and INTC have least outliers. And for stock AXP and AIG, they have the most outliers.

	PG	MSFT	JPM	INTC	IBM
DF TEST	-5.22	-4.92	-5.67	-5.52	-5.46
	GS	**GE**	**AXP**	**AIG**	**AAPL**
DF TEST	-5.14	-4.17	-5.11	-3.98	-5.13

TABLE 3.3 THE DF TEST OF 10 ASSETS

After doing the dickey-fuller test, we find that all of the stocks are stationary. From Figure 3.4 in Appendix, the equity curve plot for the stock is PG, MSFT, JPM, INTC, IBM, GS, GE , AX, AIG and AAPL separately. It can be seen that most of their revenue is of a downward shape. However, AAPL has an upward trend of equity curve, which means that it is profitable to invest in AAPL stock.

From Figure 3.5 in the Appendix, we can see that the covariance between AXP and AIG is quite high, which is 0.0121, and the covariance between AIG and GS is high as well.

From Figure 3.6 in the Appendix, the correlation between GE and AXP, JPM and GS, JPM and GE are the highest three correlations of all. Both JPM and GS and GE are financial institution so they are highly correlated, also GE and AXP are financial companies. So that's the reason why they are highly correlated.

4. Classifying Technical Indicators and analysis

In this section, we begin by constructing portfolios using the 10 stocks, and then we combine Technical Indicators and Clustering method to make filtration of our 10 stocks, and invested on the most promising assets. Finally, we evaluated and compared the performances of these portfolios.

4.1. Research Methodology

FIGURE 4.1.1 RESEARCH STAGES

4.2. Technical Analysis Indicators

Technical indicators are classified into leading and lagging indicators. Leading indicators, lead the market price. They try to predict the next price movements. In contrast, lagging indicators follow the market, they do not predict the market, and instead they indicate what prices are doing. In this project, we choose 12 indicators from these two categories.

Simple Moving Average (SMA)	Exponential Moving Average (EMA)	Weighted Moving Average (WMA)
Two Simple Moving Average (DSMA)	Two Exponential Moving Average (DEMA)	Two Weighted Moving Average (DWMA)
Relative Strength Index (RSI)	Bollinger Bands (Bollinger)	Commodity Channel Index (CCI)
Money Flow Index (MFI)	Moving Average Convergence/Divergence (MACD)	Price Rate of Change (ROC)

TABLE 4.3.1 THE TECHNICAL INDICATORS

From Table 4.3.1, we calculate these 12 indicators using R after we specify which indicators to use,

4.3. Clustering Implementation

Cluster analysis aims to segment data into different groups based on the similarity of the properties that they share. Here, we use K-Medoids clustering method; it is an optimization problem that aims to minimize a sum of pairwise dissimilarities.

1. Read signal matrices of 10 stocks into R. The dimension of each matrix is 132*12, constituting of 12 indicators and 132 transaction days.
2. Calculate the correlation between the indicators for each asset using formula:

$$\rho u_j = \sum_{i=1}^{m} \frac{(S_{ijl} - \bar{s}_{jl})(S_{ijh} - \bar{s}_{jh})}{\sqrt{\sum_{i=1}^{m}(S_{ijl} - \bar{s}_{jl})(S_{ijl} - \bar{s}_{jl})}} \;\forall j = 1,2,\dots,n, l = 1,2,\dots,k, h = 1,2,\dots,k$$

3. Calculate the average correlation of the correlation matrices using:

$$\rho_{lh} = \sum_{j=1}^{n} \frac{\rho_{lhj}}{n} \;\forall h = 1,2,\dots,k, l = 1,2,\dots,k$$

4. Using the correlation matrix as input in the K-Medoid algorithm and define the dissimilarity (distance) matrix using:

$$D_{lh} = |1 - \rho_{lh}| \; h = 1,2,\dots,k, l = 1,2,\dots,k$$

5. Perform K-Medoid algorithm to classify Technical Indicators

Through Silhouette algorithm, we find that the optimal number of cluster is 8, and second largest silhouette width correspond to 3 clusters.

Group 1	Group 2	Group 3
SMA, WMA, Bollinger, ROC	EMA, DSMA, DEMA, DWMA, MACD	RSI, CCI, MFI

TABLE 4.4.1 THE 3 CLUSTER NUMBER RESULTS

From table 4.4.1, when k = 3, we choose ROC, EMA, MACD and MFI as Technical Indicators for future stock selection.

Group 1	Group 2	Group 3	Group 4
SMA, WMA	EMA, ESMA, DEMA, DWMA	RSI	Bollinger
Group 5	**Group 6**	**Group 7**	**Group 8**
CCI	MFI	MACD	ROC

TABLE 4.4.2 THE 8 CLUSTER NUMBER RESULTS

From Table 4.4.2, when k = 8, we choose WMA, EMA, RSI, Bollinger, CCI, MFI, MACD and ROC as Technical Indicators for future stock selection.

4.4. Stock Selection

1. When k = 3, for each stock, we sum the ROC column, EMA column MACD column and MFI column for each asset.

	ROC	EMA	MACD	MFI	ROC+EMA+MACD+MFI
1	0	0	0	0	0
2	0	1	0	0	2
3	0	-1	0	0	-2
4	0	-1	0	0	-2
5	0	1	0	1	4

2. If ROC+EMA+MACD+MFI is bigger than 1, which indicate that more than 1 indicators show the price of this stock will go up, and vice versa. If the number of (ROC+EMA+MACD+MFI >=1) is bigger than the number of (ROC+EMA+MACD+MFI<=1), we move to the next step, otherwise we discard this stock.
3. If the number of (ROC+EMA+MACD+MFI >=2) is bigger than the number of (ROC+EMA+MACD+MFI<=2), we invest this stock, otherwise, we discard this stock.

By using this method, we conclude that GE, INTC and PG are most promising, and we need to invest on these 3 assets. We then repeat the same steps when k=8, and we conclude that GE, AAPL, MSFT and PG is best combination to choose. After using clustering method, we construct 2 portfolios:

Portfolio 1	Portfolio 2
GE, INTC, PG	GE, AAPL, MSFT, PG

5. Portfolio Theory and Asset Allocation

In this part we use Matlab program to find the efficient frontier, settle down the minimum variance portfolio and the tangency portfolio under the constraint of allowing short-selling or no short-selling. We compute the sharp ratio of each asset and portfolio and try to find the best portfolio allocation. We also construct the certain portfolio under a given expect return.
Using the Matlab tools, we find the efficient frontier when the short-sell is allowed and when it is not allowed. The graphs are shown as follows.
From the Figure 5.1.1 in Appendix, we can see that if short-selling is allowed, the variance will be smaller for any given expect return. This tells us that short-selling is useful for diversifying the assets and decreasing risk.

5.1. Minimum Variance Portfolio

We find the minimum variance portfolio when there is short-sell and where there is not. The result is as follows.

	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	Variance	Expect return
no short	0.000	0.000	0.000	0.306	0.000	0.136	0.000	0.558	12.522	0.740
short	-0.039	-0.117	-0.016	0.334	0.021	0.173	0.032	0.619	11.782	0.721

TABLE 5.1.1 THE MINIMUM VARIANCE PORTFOLIO WITH SHORT-SELL AND WITHOUT SHORT-SELL

We can conclude the result that the annually mean return for the MVP without short-sell is 8.8812 and it's standard deviation is 12.25832.The annually mean return for the MVP allowing short-sell is 8.6532 and the standard deviation is 11.89055.

Comparing with the variance and expect return of the 10 original assets, we can see that the variance of the minimum variance portfolio is highly below any variance of the original portfolio. This shows that, by diversifying the assets, investors can decrease their risk a lot without decreasing the expect return too much.

Compare the MVP without short-sell and with short-sell, we can see that if there is short-sell, the minimum variance is lower. This can be explained when we refer to the covariance matrix. We notice that IBM, MSFT and PG have lower variance than the other. So the more weight they have, the lower the variance will be. When there is no short-sell, we can only set the weight of the other asset to be 0.But if there is short-sell, we can short-sell the other assets and put more weight on these 3 assets, thus making the minimum variance smaller.

5.2. Tangency Portfolio

Now we use Matlab to find the tangency portfolio when there is short-sell and when there is not. The result is as follows.

	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS	variance	Expect return
no short	0.000	0.000	0.635	0.000	0.000	0.000	0.028	0.337	0.000	0.000	49.967	2.732
short	-0.048	-0.484	0.842	0.099	-0.276	0.041	0.430	0.642	0.001	-0.248	67.181	3.586

TABLE 5.2.1 THE TANGENCY PORTFOLIO

We can see that if there is short-sell, the tangency portfolio has a higher return as well as variance.

5.3. Certain Portfolios allocation

Suppose we want to achieve a target expected return of 2% per month, now we construct the efficient portfolio only with risky asset. The result is shown as follows.

	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS
weight	0	0	0.4001	0.05888	0	0	0.03974	0.50128	0	0
variance	27.8393									

TABLE 5.3.1 THE EFFICIENT PORTFOLIO WITH FIXED EXPECTED RETURN.

Then we construct another portfolio using both risky assets and riskless asset to achieve the same target expected return. Since we have already found the tangency portfolio, it is obvious that we should use tangency portfolio without short-sale and the riskless asset to construct the portfolio.

The formula should be:

W1*riskless rate+(1-W1)*tangency expreturn=2

	riskfree	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS
weight	0.279	0.000	0.000	0.000	0.221	0.000	0.098	0.000	0.402	0.000	0.000

TABLE 5.3.1 THE EFFICIENT PORTFOLIO WITH RISK-FREE ASSET

5.4. Portfolio Comparison

We have constructed various portfolios in the previous parts. In this part we try to compare them using sharp ratio. We will also look at the VaR for the portfolios.

In the previous sections, we constructed the minimum variance portfolio with and without short-sale, and the tangency portfolio with and without short-sale. We also constructed 2 portfolios with given expect return of 2%.

	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS
Sharp	0.097	0.035	0.363	0.1	0.07	0.114	0.099	0.152	0.028	0.1

TABLE 5.4.1 THE SHARP RATIO OF THE 10 ORIGINAL ASSETS

	MVP without short	MVP with short
Sharp	0.177	0.123
VaR	505828.000	490333.000
ES	653129	633216.000

TABLE 5.4.2 THE MINIMUM VARIANCE PORTFOLIO OF THE 10 ASSETS

From Table 5.4.2, we can see that the MVP without short-sale has a lower VaR and ES, which indicates that the risk is lower. Meanwhile we can see that minimum variance portfolio, while having a higher risk, the sharp ratio is higher, which means investors will prefer it.

	Tangency portfolio without short	Tangency portfolio with short
Sharp	0.370461	0.423733
VaR	885087	985030
ES	1179331	1326216

TABLE 5.4.3 THE TANGENCY PORTFOLIO WITH AND WITHOUT SHORT-SALE

From Table 5.4.3, we can see that the tangency portfolio with short has a higher sharp ratio than the tangency portfolio without short. To explain this, we can go back to the sharp ratio of the 10 original stocks and their weights in these 2 portfolios. We can find out that the tangency portfolio with short-sale generally short those stocks with lower sharp ratio and put higher weight on those stocks with higher sharp ratio. Meanwhile, the VaR and ES of the tangency portfolio without short-sale also have a higher VaR and ES, which means that its risk is higher.

	2% portfolio with risky assets	**2% portfolio with risky assets and riskless asset**
Sharp	0.357535	0.370461
VaR	336475	664563
Shortfall	445055	884197

TABLE 5.4.4 THE PORTFOLIOS CONSTRUCTED UNDER A GIVEN EXPECT RETURN OF 2% PER MONTH

It is quite strange that the portfolio constructed only with risky assets has a lower VaR and ES. To explain this, I guess that, as riskless asset has a far lower expected return than the risky assets, then to reach the same expected return, we have to put more weight on the risky assets, thus making the risk higher. But of course, the portfolio contains both risky and riskless assets have the highest sharp ratio as it is using tangency portfolio to construct the whole portfolio.

For portfolios constructed using Technical Indicators combined with clustering algorithm, we calculated their sharp ratios and MVPs.

	Tangency Portfolio			**MVP**		
Portfolio 1	**GE**	**INTC**	**PG**	**GE**	**INTC**	**PG**
Weight	-0.0793	0.2109	0.8683	-0.049	0.21	0.839
Sharp ratio	0.162			0.159		
Mu	0.744			0.751		
std	4.01			4.01		

TABLE 5.4.5 SUMMARY OF PORTFOLIO 1

	Tangency Portfolio				**MVP**			
Portfolio 2	**GE**	**AAPL**	**MSFT**	**PG**	**GE**	**AAPL**	**MSFT**	**PG**
Weight	-0.398	0.733	-0.038	0.703	-0.043	-0.02	0.2686	0.794
Sharp ratio	0.395				0.159			
mu	3.18				0.742			
std	7.76				3.94			

TABLE 5.4.6 SUMMARY OF PORTFOLIO 2

From Table 5.4.5 and Table 5.4.6, we find that the sharp ratios of two portfolios using the combination of Technical Indicators and clustering method are lower than the sharp ratio of the portfolio using original 10 stocks. There are several explanations for this phenomenon. First, we only used the returns of 10 stocks, it is possible that the number of stocks is not large enough to support for K-medoid clustering method. Since we used the average correlation matrix of the 12 indicators from 10 stocks as input for the K-medoid clustering algorithm, the second reason that might lead to this result is that we did not download enough historical data, since we used monthly data range from Jan. 1st, 2004 to Dec. 31th, 2014, therefore there are total 132 historical data for each stock. As a result, the limited sample size might cause this outcome. The third explanation is that after conducting the clustering algorithm, we only choose a simple combining strategy to construct the portfolio, and if we could craft a more sophisticated portfolio, we might get some positive result.

6. Principal Component Analysis

From Figure 6.1 and 6.2 in Appendix, we can see that AXP and GE are most highly correlated. AIG and PG are least correlated.

From the correlation matrix (Figure 6.2 in Appendix) and Expected return versus risk (Figure 6.3 in Appendix), we can see that our 10 assets are somewhat correlated, since they are mainly come from manufactory and financial industry. Compared with assets with $\rho=0$, we can see that diversification will reduce risk.

We will use PCA to study how the curves change from month to month.

First, we will look at the 10 eigenvalues. The results from R 's function prcomp are as follow.

```
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4     Comp.5     Comp.6
Standard deviation     0.2997125 0.1456663 0.09201618 0.07470897 0.06320932 0.05422281
Proportion of Variance 0.6447774 0.1523065 0.06077547 0.04006319 0.02867888 0.02110397
Cumulative Proportion  0.6447774 0.7970838 0.85785931 0.89792250 0.92660138 0.94770535
                           Comp.7     Comp.8     Comp.9     Comp.10
Standard deviation     0.05038038 0.04516459 0.03934900 0.034045530
Proportion of Variance 0.01821893 0.01464186 0.01111392 0.008319939
Cumulative Proportion  0.96592428 0.98056614 0.99168006 1.000000000
```

FIGURE 5.5 PCA OF 10 ASSETS

One can see, for example, that the standard deviation of the first component is 0.300 and represents 64.5% of the total variance. Also the first six principal components have 94.7% of the variance, and this increase to 98.1% for the first eight principal components. The variances are plotted in Figure 6.5. We will concentrate on the first six principal components since approximately 95% of the variation in the changes in yields is in the space they span.

7. Risk Management

In this section, we will use historic data to estimate the distribution of returns. First, we will assume that the returns for each asset are i.i.d. with normal distribution. For each asset, the estimated 5% value-at-risk and the expected shortfall on $100,000 investment over a one month investment horizon are as follow.

	AAPL	**AIG**	**GE**	**AXP**	**GS**
VaR(0.05)	13148.357	**46638.394**	12510.056	16267.856	13460.846
	IBM	**INTC**	**JPM**	**MSFT**	**PG**
VaR(0.05)	8158.206	11367.933	12636.396	9850.406	**6281.350**

TABLE 7.1 THE ESTIMATED 5% VALUE-AT-RISK ON $100,000 INVESTMENT WITH NORMAL DISTRIBUTION.

	AAPL	AIG	GE	AXP	GS
ES(0.05)	17468.216	**58719.978**	15785.617	20689.207	17131.260
	IBM	**INTC**	**JPM**	**MSFT**	**PG**
ES(0.05)	10395.549	14413.352	16081.948	12569.933	**8070.697**

TABLE 7.2 THE EXPECTED SHORTFALL ON $100,000 INVESTMENT WITH NORMAL DISTRIBUTION.

From Table 7.1, we can see that the AIG asset the highest VaR value and the PG asset has the lowest VaR value. For AIG asset, a monthly return of -46.638% or less occurred only 5% of the time in the historic data, so we estimate that there is 5% chance of a return of that size occurring during the next month. A return of -46.638% on a $100,000 investment yields a revenue of -$46,638.394, and therefore there is 5% chance of a loss exceeding $46,638.394 over the next month. While for PG asset, there is 5% chance of a loss exceeding $628.135 on a $100,0000 investment over the next month. Therefore, the PG asset has the smallest loss and the AIG has the largest loss with 5% chance. And we prefer the PG asset to the AIG asset because we always expect the smaller loss for our portfolio.

From Table 7.2, we can see that the AIG asset has the highest ES value and the PG asset has the lowest ES value. For AIG asset, the expected loss given that the loss exceeds VaR is $58,719.978. While the expected loss given that the loss exceeds VaR for PG asset is $8,070.697. Therefore the PG asset has the smallest expected loss and the AIG asset has the largest expected loss. And we preferred the PG asset to the AIG asset because we expect smaller expected loss for our portfolio.

And the following table gives the estimated 5% value-at-risk and the expected shortfall on $100,000 investment using nonparametric method.

	AAPL	AIG	GE	AXP	GS
VaR(0.05)	11555.704	**22921.458**	12095.590	12664.329	14117.924
	IBM	**INTC**	**JPM**	**MSFT**	**PG**
VaR(0.05)	7834.576	12265.079	10944.851	8723.955	**5260.252**

TABLE 7.3 THE NONPARAMETRIC ESTIMATED 5% VALUE-AT-RISK ON $100,000 INVESTMENT.

	AAPL	AIG	GE	AXP	GS
ES(0.05)	19055.017	**45535.642**	18198.378	18819.565	18442.081
	IBM	**INTC**	**JPM**	**MSFT**	**PG**
ES(0.05)	12530.658	15356.407	18494.710	12948.132	**8599.769**

TABLE 7.4 THE NONPARAMETRIC EXPECTED SHORTFALL ON $100,000 INVESTMENT.

From Table 7.3, we can see that the AIG asset has the highest VaR value and the PG asset has the lowest VaR value. For AIG asset, there is 5% chance of a loss exceeding $22,921.458 over the next month. While for PG asset, there is 5% chance of a loss exceeding $5,260.252 on a $100,0000 investment over the next month.
From the table 6.2, we can see that the AIG asset has the highest ES value and the PG asset has the lowest ES value. For AIG asset, the expected loss given that the loss exceeds VaR is $45,535.642. While the expected loss given that the loss exceeds VaR for PG asset is $8,599.769.

For now, we will assume an i.i.d. sample of historic returns and use model-free resampling to compute a 95% confidence interval for VaR by bootstrapping. And the standard error and 95% confidence interval are provided as follow.

	AAPL	AIG	GE	AXP	GS	IBM	INTC	JPM	MSFT	PG
Standard error	1486.795	**5505.816**	1248.582	2611.869	2157.647	1343.610	1740.580	3843.811	1525.120	**1111.708**
2.5% lower bound	8762.238	**16373.842**	8984.802	7195.522	8831.673	5914.917	8674.154	8989.532	8273.501	**4430.586**
97.5% upper bound	14416.033	**36988.547**	13299.875	19167.053	17569.116	11735.941	14220.499	20122.660	12843.609	**7940.354**

TABLE 7.5 THE STATDARD ERROR AND 95% CONFIDENCE INTERVAL FOR VAR(0.05)

From Table 7.5, we can see that the PG asset has the smallest standard error and the AIG asset has the largest standard error. And we can see that the PG asset has the narrowest 95% confidence interval for the VaR(0.05) and the AIG asset has the widest 95% confidence interval. We can conclude that the PG asset is the least risky asset and the AIG is the most risky asset for our portfolio.

8. Copulas

In this section, we will fit copulas to the multivariate data set of returns on our index. First, we fit univariate normal distribution to the ten variables, and then we will fit the normal copulas to the data. Moreover, we fit meta t-distribution, Clayton and Gumbel copulas to the data using parametric and empirical distribution methods. By comparing the AIC for the normal, t, Clayton and Gumbel copulas, we can know which copulas fits the data best.

	Normal	t-distribution	Clayton	Gumbel
Empirical	248.7	**270.4**	177.7	144.4
Parametric	243.6	252	145.9	145.4

TABLE 8.1 THE MAXIMUM LOG-LIKELIHOOD VALUE

Because the AIC's value is a decreasing function of log-likelihood value, we choose the model with the largest log-likelihood value as our best fit of data. From Table 8.1, we know that the meta t-distribution copulas with empirical distribution of data is the best fit for our data since its log-likelihood value is maximum.

9. Future work

Just as we analyzed before, several reasons lead to the result that our strategy cannot generate more returns. However Technical Analysis and Fundamental Analysis are used by many traders, brokerages, trading groups, and financial institutions which justify the usefulness of Indicators. Therefore, there are lots of works for us to do in the future. First, we only choose a simple combining strategy in constructing portfolios. Are there other strategies to perform well in stock market? Second, we simply combine the signal using method in portfolio strategy part. Are there any more interesting patterns that give more stock exchanges?

After this research, we honed our strategic thinking by constructing portfolios and evaluating their performances. Besides, we acquired deep understanding of the family of indicators and we will still conduct the further research regarding to this topic.

Appendix A

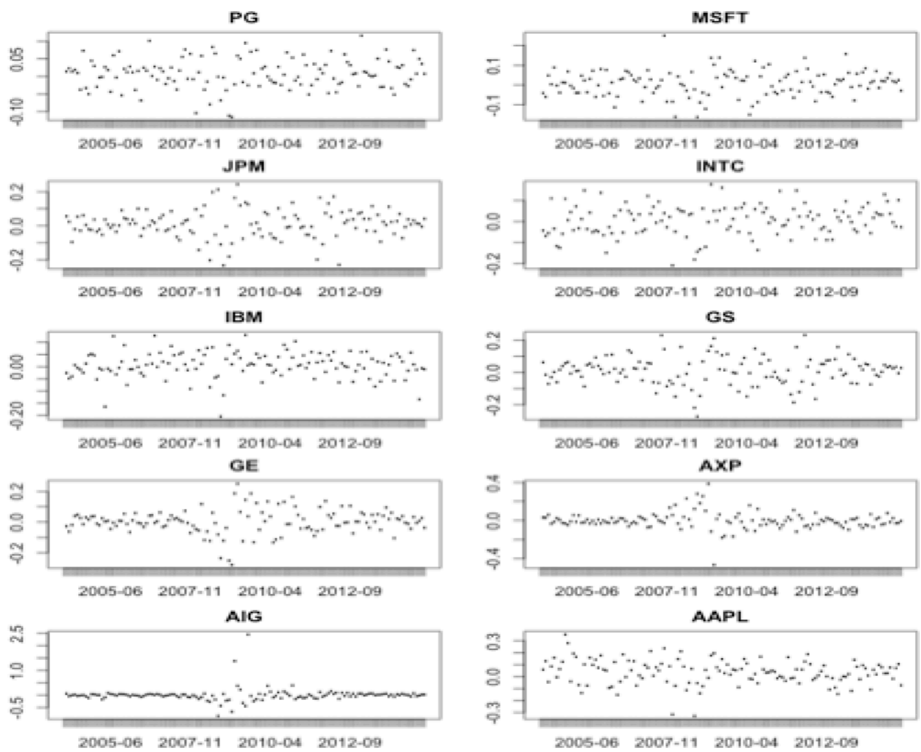

FIGURE 3.1 RETURNS OF STOCKS

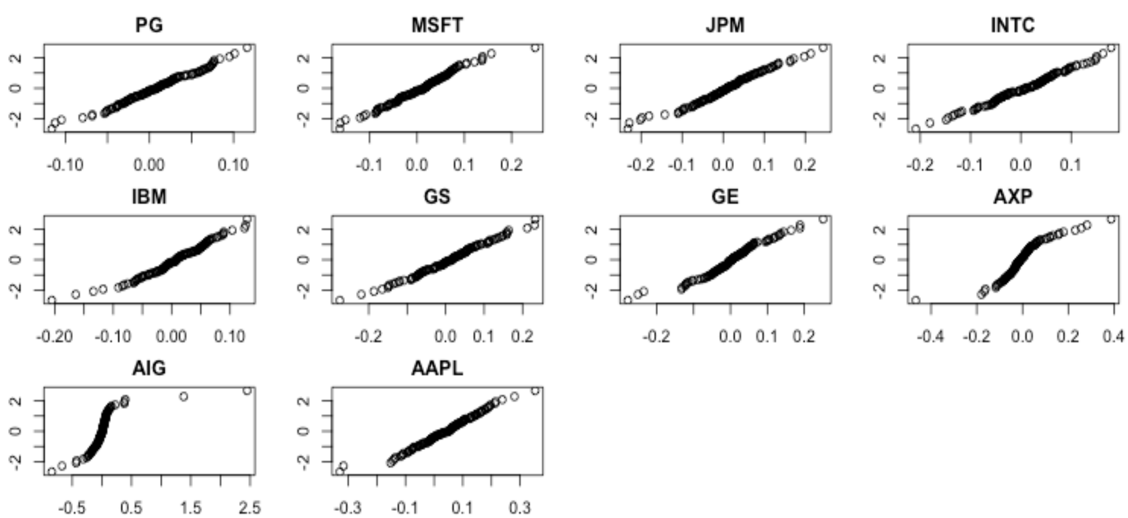

FIGURE 3.2 QQ-PLOT OF STOCKS

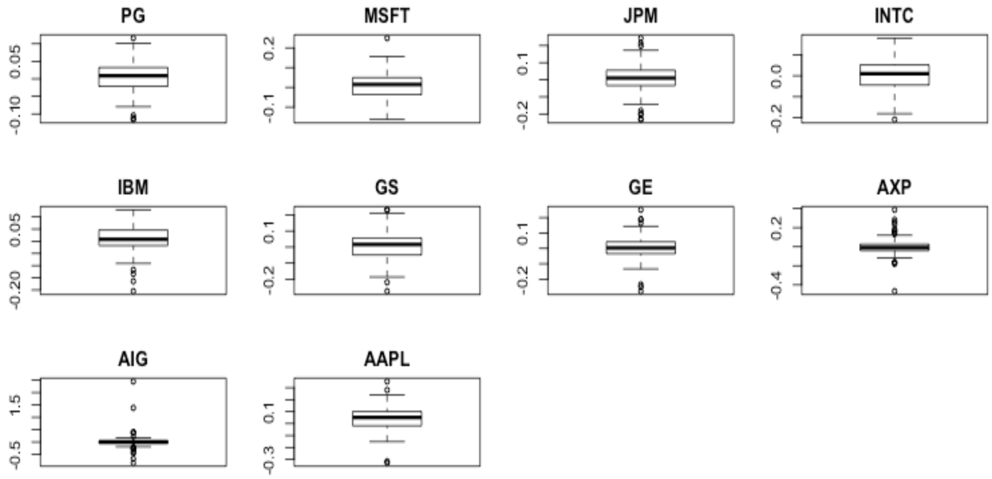

FIGURE3.3 BOX PLOT OF STOCKS

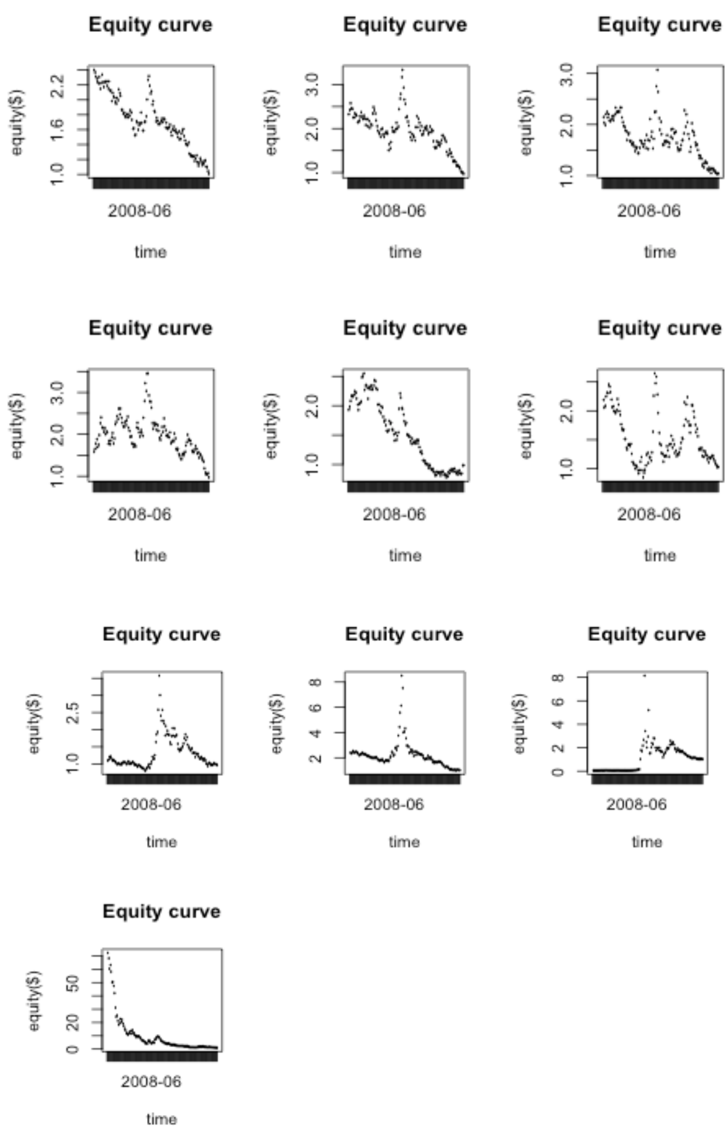

FIGURE3.4 EQUITY CURVE

	X	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS
X	1441.000	26.8	18.9	-65.03	0.812	40.4	20.91	15.9	7.97	49.61	-0.566
AXP	26.804	112.0	51.6	40.52	17.839	28.7	29.07	47.8	15.32	121.13	38.542
GE	18.904	51.6	61.4	25.73	16.172	24.2	21.85	38.3	16.73	86.25	29.665
AAPL	-65.032	40.5	25.7	106.88	19.517	33.5	27.31	15.3	9.27	65.75	38.251
IBM	0.812	17.8	16.2	19.52	28.669	18.2	9.49	15.4	4.40	19.71	20.368
INTC	40.435	28.7	24.2	33.54	18.181	53.1	22.91	17.5	7.30	59.06	22.966
MSFT	20.915	29.1	21.9	27.31	9.491	22.9	42.36	21.3	6.92	53.40	26.979
JPM	15.906	47.8	38.3	15.31	15.394	17.5	21.32	68.0	11.52	74.14	42.760
PG	7.967	15.3	16.7	9.27	4.404	7.3	6.92	11.5	18.34	7.09	7.474
AIG	49.607	121.1	86.2	65.75	19.711	59.1	53.40	74.1	7.09	835.96	69.741
GS	-0.566	38.5	29.7	38.25	20.368	23.0	26.98	42.8	7.47	69.74	77.156

FIGURE3.5COVRIANCE TABLE OF THE TEN STOCKS

```
> cor(dat)
```

	X	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS
X	1.0000	0.0667	0.0635	-0.166	0.004	0.146	0.0847	0.0508	0.0490	0.0452	-0.0017
AXP	0.0667	1.0000	0.6221	0.370	0.315	0.373	0.4222	0.5483	0.3382	0.3960	0.4147
GE	0.0635	0.6221	1.0000	0.317	0.385	0.424	0.4283	0.5927	0.4984	0.3805	0.4308
AAPL	-0.1657	0.3704	0.3175	1.000	0.353	0.445	0.4059	0.1796	0.2094	0.2200	0.4212
IBM	0.0040	0.3149	0.3853	0.353	1.000	0.466	0.2724	0.3487	0.1921	0.1273	0.4331
INTC	0.1462	0.3727	0.4237	0.445	0.466	1.000	0.4829	0.2913	0.2338	0.2803	0.3587
MSFT	0.0847	0.4222	0.4283	0.406	0.272	0.483	1.0000	0.3973	0.2485	0.2838	0.4719
JPM	0.0508	0.5483	0.5927	0.180	0.349	0.291	0.3973	1.0000	0.3263	0.3110	0.5904
PG	0.0490	0.3382	0.4984	0.209	0.192	0.234	0.2485	0.3263	1.0000	0.0573	0.1987
AIG	0.0452	0.3960	0.3805	0.220	0.127	0.280	0.2838	0.3110	0.0573	1.0000	0.2746
GS	-0.0017	0.4147	0.4308	0.421	0.433	0.359	0.4719	0.5904	0.1987	0.2746	1.0000

```
>
```

FIGURE3.6 CORRELATION TABLE OF THE TEN STOCKS

```
> print(cbind(my.k.choices,avg.sil.width))
```

	my.k.choices	avg.sil.width
[1,]	2	0.2064893
[2,]	3	0.2710469
[3,]	4	0.2412409
[4,]	5	0.2074639
[5,]	6	0.2040445
[6,]	7	0.2007712
[7,]	8	0.2751512
[8,]	9	0.2376430
[9,]	10	0.2033573

FIGURE 4.4.1

Through Silhouette algorithm, we find that the optimal number of cluster is 8, and second largest silhouette width correspond to 3 clusters.

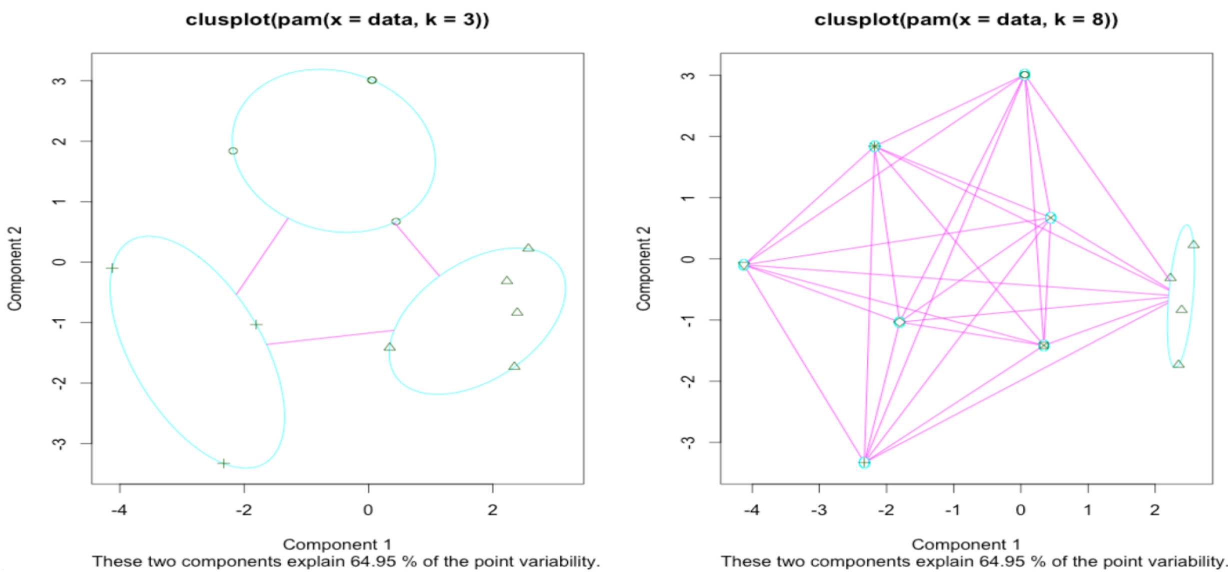

FIGURE 4.4.2 CLUSTER PLOTS FOR 3 AND 8 CLUSTERS.

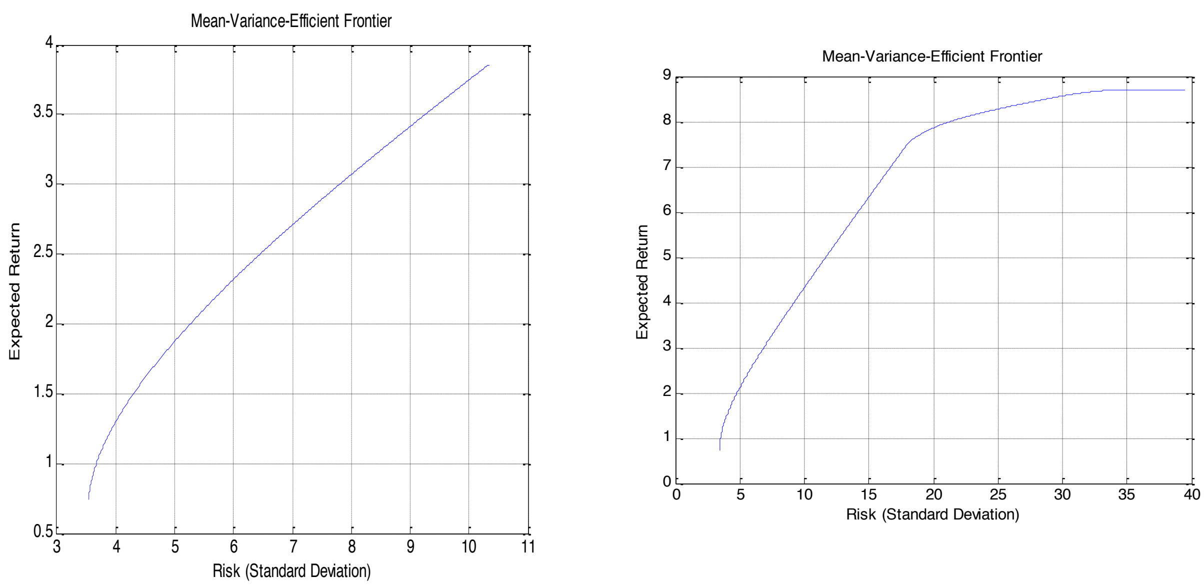

FIGURE 5.1.1 THE EFFICIENT FRONTIER WITH SHORT-SELL AND WITHOUT SHORT-SELL.

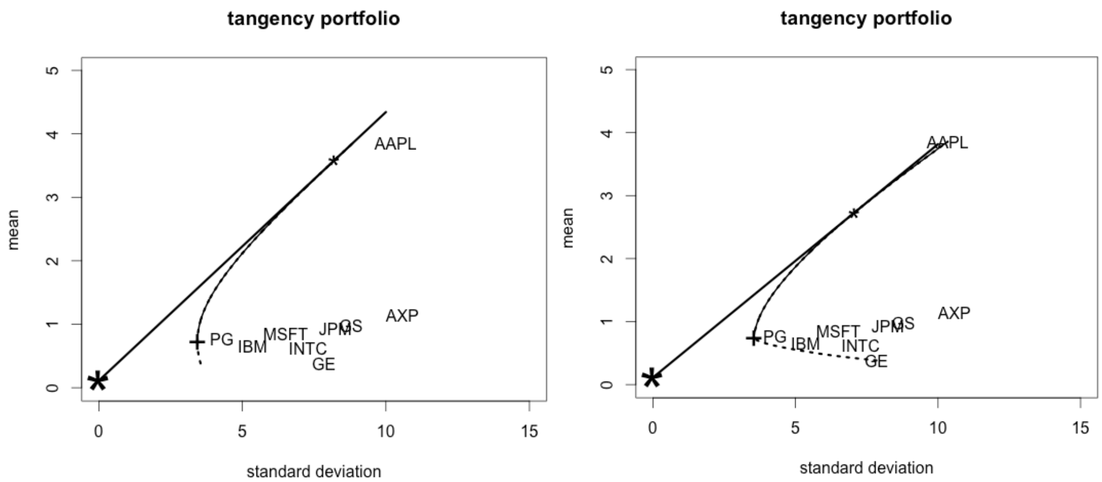

FIGURE 5.1.2 THE TANGENCY PORTFOLIO WITH SHORT-SELL AND WITHOUT SHORT-SELL

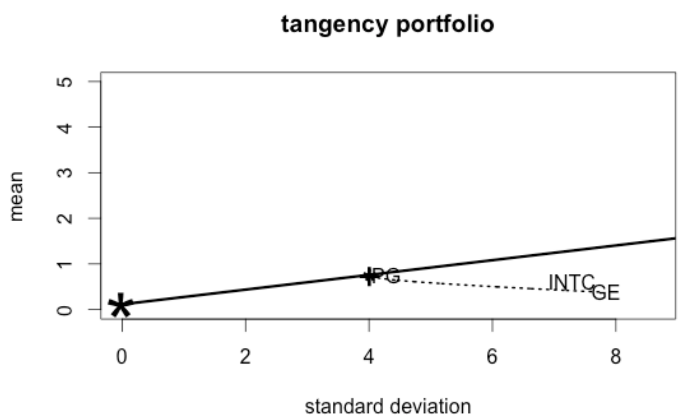

FIGURE 5.4.1 TANGENCY PORTFOLIO OF PORTFOLIO 1

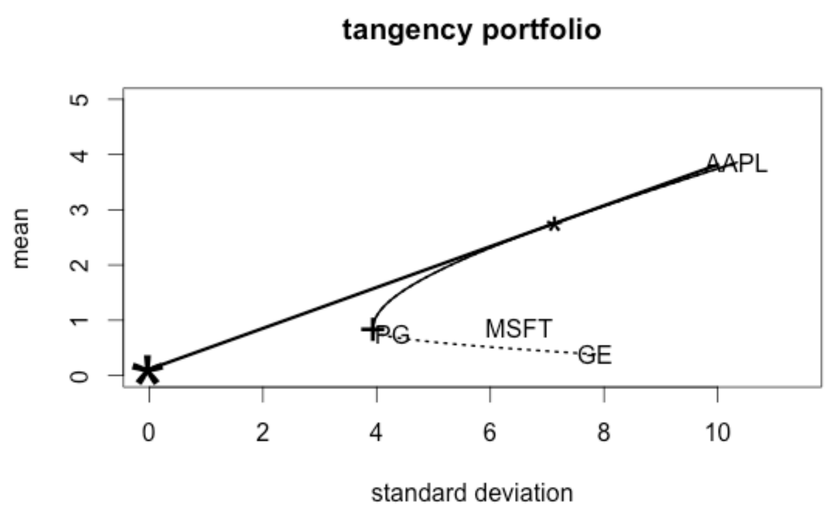

FIGURE 5.4.2 TANGENCY PORTFOLIO OF PORTFOLIO 2

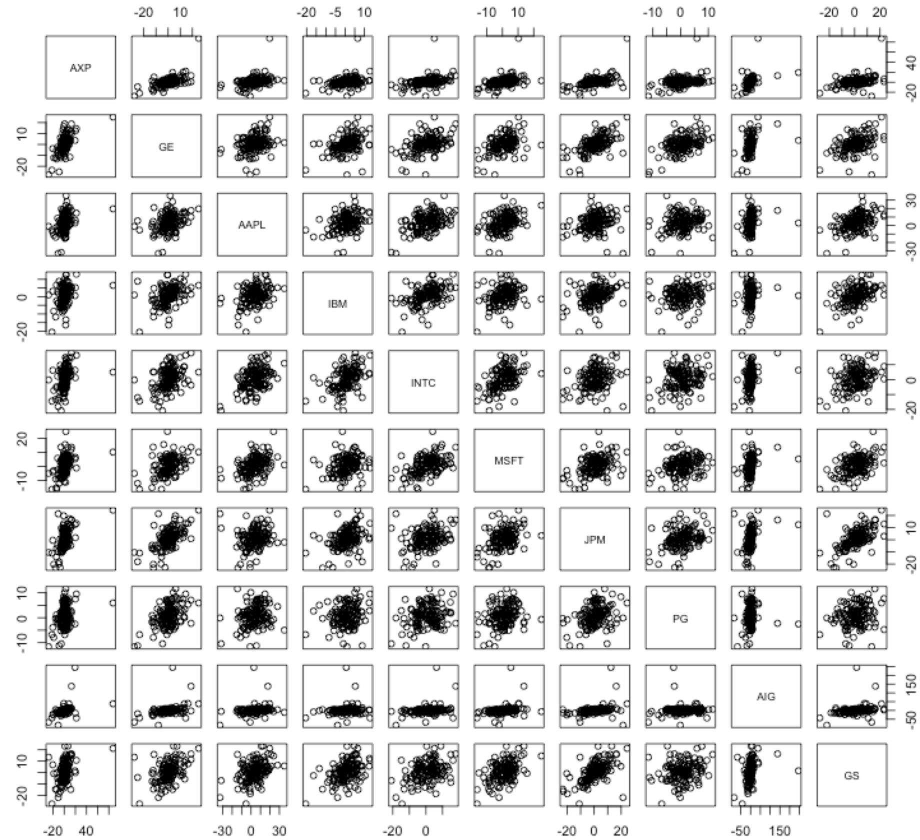

FIGURE 6.1 SCATTER PLOT MATRIX OF 10 ASSETS

	AXP	GE	AAPL	IBM	INTC	MSFT	JPM	PG	AIG	GS
AXP	1.0000000	0.6220898	0.3704399	0.3148863	0.3726658	0.4221898	0.5482944	0.33821747	0.39595822	0.4146940
GE	0.6220898	1.0000000	0.3174961	0.3853003	0.4237149	0.4282889	0.5926735	0.49840816	0.38053120	0.4308285
AAPL	0.3704399	0.3174961	1.0000000	0.3525930	0.4452109	0.4058839	0.1795688	0.20944070	0.21998581	0.4212278
IBM	0.3148863	0.3853003	0.3525930	1.0000000	0.4659090	0.2723540	0.3486816	0.19207975	0.12732358	0.4330792
INTC	0.3726658	0.4237149	0.4452109	0.4659090	1.0000000	0.4829476	0.2913476	0.23375670	0.28029404	0.3587459
MSFT	0.4221898	0.4282889	0.4058839	0.2723540	0.4829476	1.0000000	0.3972870	0.24846302	0.28379938	0.4719264
JPM	0.5482944	0.5926735	0.1795688	0.3486816	0.2913476	0.3972870	1.0000000	0.32626915	0.31099115	0.5903635
PG	0.3382175	0.4984082	0.2094407	0.1920798	0.2337567	0.2484630	0.3262692	1.00000000	0.05726418	0.1986953
AIG	0.3959582	0.3805312	0.2199858	0.1273236	0.2802940	0.2837994	0.3109912	0.05726418	1.00000000	0.2746064
GS	0.4146940	0.4308285	0.4212278	0.4330792	0.3587459	0.4719264	0.5903635	0.19869530	0.27460637	1.0000000

FIGURE 6.2 CORRELATION MATRIX OF 10 ASSETS

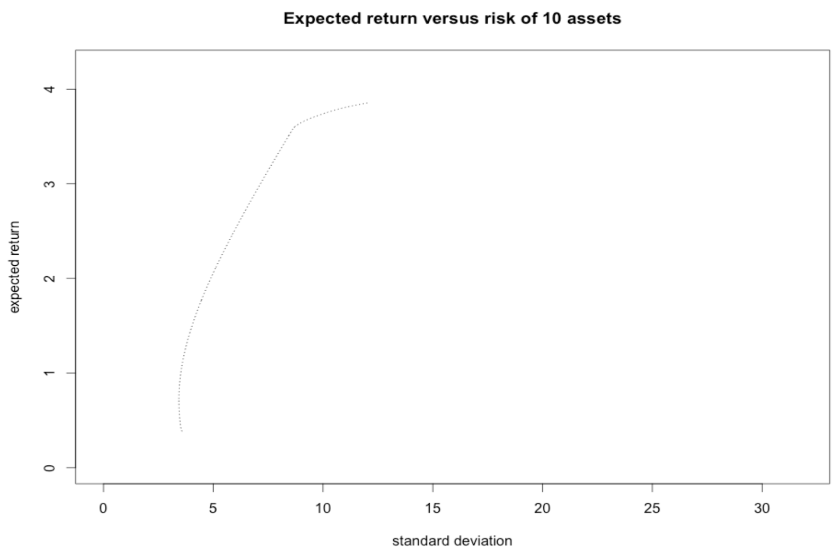

FIGURE 6.3 EXPECTED RETURN VERSUS RISK OF 10 ASSETS

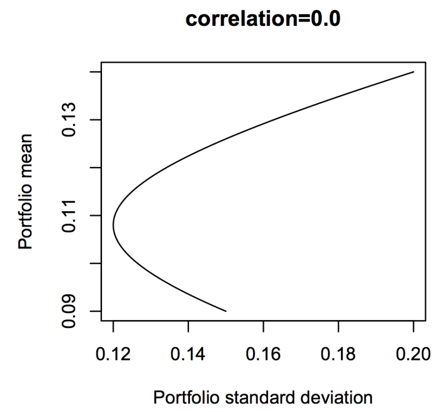

FIGURE 6.4 EXPECTED RETURN VERSUS RISK WITHOUT CORRELATION

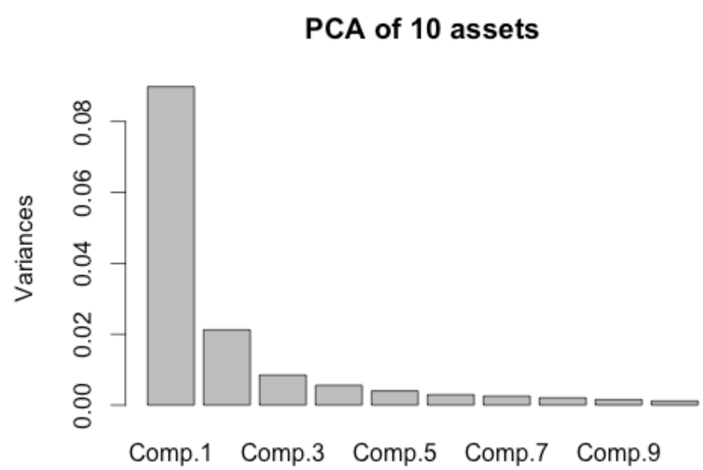

FIGURE 7.5 PCA OF 10 ASSETS

Appendix B
Technical Indicators that used

1. **Simple Moving Average (SMA)**

The simple moving average (SMA) indicator requires closing prices for its calculations (Achehs [2001]). The closing prices for a certain period are summed and then divided by n to obtain the SMA. The parameter n can have any value. In general, a large n is intended for long-term forecasting and small n is for short trends.

$$SMA_i = \frac{\sum_{j=i-n+1}^{i} C_j}{n} \quad i = n, \dots$$

where C. is the closing price for day j , n is the time period (in this study, n = 20), and SMAi. is the simple moving average value for day i. Whenever the asset's price rises above its moving average, a buy signal is generated and the opposite occurs when the asset's price falls below its moving average—a sell signal is generated. Thus we can represent the signal conditions as follows:

$$Signal_i = \begin{cases} +1 & \text{if } (C_i > SMA_i) \text{ and } (C_{i-1} < SMA_{i-1}) \\ -1 & \text{if } (C_i < SMA_i) \text{ and } (C_{i-1} > SMA_{i-1}) \\ 0 & \text{Elsewhere} \end{cases}$$

2. **Exponential Moving Average (EMA)**

Exponential moving average (EMA) is similar to SMA, except that it apphes a percentage of today's closing price to a percentage of yesterday's moving average value (Achelis [2001]). The EMA can be calculated using the following steps:

SMA: 20 period sum / 20
Multiplier: (2 / (Time periods + 1)) = (2 / (20 + 1))
EMA: {Close - EMA(previous day)} x multiplier + EMA(previous day).

3. **Weighted Moving Average (WMA)**

Weighted moving average (WMA) assigns more weight to recent dates (Achelis [2001]). WMA can be calculated as follows:

$$EMA_i = \frac{\sum_{k=1}^{n} \sum_{j=i-n+1}^{i} W_k C_j}{\sum_{k=1}^{n} W_k} \quad i = n, \dots \infty \qquad (18)$$

where W^ is the assigned weight. Whenever the asset's price rises above its moving average, a buy signal is generated, and when the asset's price falls below its moving average, a sell signal is generated. Thus, we can represent the signal conditions as follows:

$$Signal_i = \begin{cases} +1 & \text{if } (C_i > WMA_i) \text{ and } (C_{i-1} < WMA_{i-1}) \\ -1 & \text{if } (C_i < WMA_i) \text{ and } (C_{i-1} > WMA_{i-1}) \\ 0 & \text{Elsewhere} \end{cases}$$

4. Two Simple Moving Average (DSMA)

The two simple moving averages (TSMA) indicator is similar to SMA except that we have to calculate two moving averages instead of one. The first SMA is called the long simple moving average (LSMA) and the second moving average is the short simple moving average (SSMA). LSMA is called long because it is divided by a longer period n and SSMA is called short because it is divided by a shorter period of in. In this study, n = 24 and m = 12.

$$LSMA_i = \frac{\sum_{j=i-n+1}^{i} C_j}{n} \quad i = n, \ldots$$

$$SSMA_k = \frac{\sum_{j=i-n+1}^{i} C_j}{m} \quad k = m, \ldots$$

Whenever SSMA rises above LSMA, a buy signal is generated, and conversely, when SSMA falls below LSMA, a sell signal is generated. Thus, we can represent the signal conditions as follows:

$$Signal_i = \begin{cases} +1 \text{ if} \left(SSMA_i > LSMA_i\right) \text{ and } \left(SSMA_{i-1} < LSMA_{i-1}\right) \\ -1 \text{ if} \left(S\$MA_i < LSMA_i\right) \text{ and } \left(SSMA_{i-1} > LSMA_{i-1}\right) \\ \quad 0 \quad \text{Elsewhere} \end{cases}$$

5. Two Exponential Moving Average (DEMA)

The two exponential moving averages indicator (TEMA) is similar to TSMA in signals assignments and to EMA in the calculation's procedure. SEMA is Short Exponential Moving Average LEMA is Long Exponential Moving Average.

$$Signal_i = \begin{cases} +1 \text{ if} \left(SEMA_i > LEMA_i\right) \text{ and } \left(SEMA_{i-1} < LEMA_{i-1}\right) \\ -1 \text{ if} \left(SEMA_i < LEMA_i\right) \text{ and } \left(SEMA_{i-1} > LEMA_{i-1}\right) \\ \quad 0 | \quad \text{Elsewhere} \end{cases}$$

6. Two Weighted Moving Average (DWMA)

This measure, two weighted moving averages (TWMA), is similar to TSMA and TEMA in the assignments ofthe signals and similar to EMA in the calculation's procedure. SWMA is Short Weighted Moving Average LWMA is Long Weighted Moving Average.

$$Signal_i = \begin{cases} +1 \text{ if} \left(SWMA_i > LWMA_i\right) \text{ and } \left(SWMA_{i-1} < LWMA_{i-1}\right) \\ -1 \text{ if} \left(SWMA_i < LWMA_i\right) \text{ and } \left(SWMA_{i-1} > LWMA_{i-1}\right) \\ \quad 0 \quad \text{Elsewhere} \end{cases}$$

7. Relative Strength Index (RSI)

The relative strength index (RSI) is an oscillator index (Achelis [2001] and Jobman [1994]) that follows the price and grades it between 0 and 100, where the number 100 means that the asset is highly overpriced and 0 means that the asset is highly undervalued:

$$RSI_i = 100 - \left(\frac{100}{1 + \left(\frac{U}{D}\right)} \right) i = n+1, \ldots$$

Where U is the average of upward price change for n given period and D is the average of downward price change for n.

$$Signal_i = \begin{cases} +1 \quad \text{if } (RSI_i > 30) \quad \text{and} \quad (RSI_{i-1} < 30) \\ -1 \quad \text{if } (RSI_i < 70) \quad \text{and} \quad (RSI_{i-1} > 70) \\ \qquad 0 \quad \text{Elsewhere} \end{cases}$$

8. Bollinger Bands (Bollinger)

Bollinger bands consist of a set of three curves drawn in relation to securities prices (Achelis [2001] and Wilder [1978]). The three curves are the middle band(MB), the upper band (UB) and the lower band (LB). The three curves are calculated as follows

$$MB_i = \frac{\sum_{j=i-n+1}^{i} C_j}{n} \; i = n, \ldots \infty$$

$$UB_i = MB_i + \sigma\sqrt{\frac{\sum_{j=i-n+1}^{i} \left(C_j - MB_i\right)^2}{n}} \; i = n, \; \ldots$$

$$LB_i = MB_i - \sigma\sqrt{\frac{\sum_{j=i-n+1}^{i} \left(C_j - MB_i\right)^2}{n}} \; i = n, \; \ldots$$

Where O is the standard deviation of the closing prices for period. However, a buy signal is generated when the asset's price cross the lower band and a sell signal is generated when the asset's price touches the upper band.

$$Signal_i = \begin{cases} +1 \;\; \text{if } (C_i > LB_i) \;\; \text{and} \;\; (C_{i-1} < LB_{i-1}) \\ -1 \;\; \text{if } (C_i < UB_i) \;\; \text{and} \;\; (C_{i-1} > UB_{i-1}) \\ \qquad 0 \quad \text{Elsewhere} \end{cases}$$

9. Commodity Channel Index (CCI)

The commodity channel index (CCI) measures the variation of a security's price from its mean (Achelis [2001] and Lambert [1980]). A buy signal is generated when the CCI is below -70, and a sell signal is generated when the CCI is above 70.

$$Signal_i = \begin{cases} +1 \;\; \text{if } (CCI_i > -70) \;\; \text{and} \;\; (CCI_{i-1} < -70) \\ -1 \;\; \text{if } (CCI_i < +70) \;\; \text{and} \;\; (CCI_{i-1} > +70) \\ \qquad 0 \quad \text{Elsewhere} \end{cases}$$

10. Money Flow Index (MFI)

The money fiow index (MFI) can be calculated by the following steps (Achelis [2001]):

1. Calculate the typical price for period i:

$$TP_i = \frac{H_i + L_i + C_i}{3} \; i = 1, 2, \ldots$$

2. where TP. is the typical price for day i, H is the highest price in day i, L. is the lowest price in day i, and C is the closing price for day i.

$$MF_i = TP_i \times Volume_i \quad i = 1, 2, \; \ldots$$

3. If TP. is greater than TP. j, it is considered positive money; if TP. is less than TP. j , it is considered negative money.

4. Calculate the money ratio MR:

$$MR_i = \frac{\sum_{j=2}^{n} Positive\ Money\ Flow}{\sum_{j=2}^{n} Negative\ Money\ Flow} \quad i = 1, \ldots$$

5. Calculate MFI for period i:

$$MFI_i = 100 - \left[\frac{100}{1 - MR_i}\right] i = 1, 2, \ldots$$

$$Signal_i = \begin{cases} +1 & \text{if } (MFI_i > 30) \text{ and } (MFI_{i-1} < 30) \\ -1 & \text{if } (MFI_i < 70) \text{ and } (MFI_{i-1} > 70) \\ 0 & \text{Elsewhere} \end{cases}$$

11. Moving Average Convergence/ Divergence (MACD)

Moving average convergence/divergence (MACD) can be defined as the difference between the faster and slower exponential moving averages (Achelis [2001]). Usually, it is the difference between EMA for 12 days and EMA for 26 days. It is recommended to buy if M^CD. goes above its moving average and sell it if M^CD. Falls below its moving average.

$$MACD_i = EMA_i^{12} - EMA_i^{26} \quad i = 26, \ldots \qquad (49)$$

$$Signal_i = \begin{cases} +1 \text{ if } (MACD_i > EMA_i) \text{ and } (MACD_{i-1} < EMA_{i-1}) \\ -1 \text{ if } (MACD_i < EMA_i) \text{ and } (MACD_{i-1} > EMA_{i-1}) \\ 0 \quad \text{Elsewhere} \end{cases}$$

12. Price Rate of Change (ROC)

The price rate of change (ROC) can be calculated as follows (Achelis [2001] and Wilder [1978]):

$$ROC_i = \frac{C_i - C_{i-n}}{C_{i-n}} \times 100 \quad i = n+1, n+2, \ldots$$

The higher the ROC, the more overbought the security; the lower ROC means that the security is oversold.

$$Signal_i = \begin{cases} +1 & \text{if } (ROC_i > -6.5\%) \text{ and } (ROC_{i-1} < -6.5\%) \\ -1 & \text{if } (ROC_i < +6.5\%) \text{ and } (ROC_{i-1} > +6.5\%) \\ 0 & \text{Elsewhere} \end{cases}$$