UNIVERSITY OF MALAYA

Master of Data Science

Faculty of Computer Science & Information Technology

WQD7011 Numerical Optimization

Lecturer: Dr Lim Chee Kau

OPTIMIZATION OF STOCK PORTFOLIO – COMPARATIVE ANALYSIS

Group: Optimisers

Name	Matric Number
Soo Kar Lok	S2000961
Hong Zi Shen	S2114600
Jonathan Kiew Weng Kiat	S2043163
Tay Shi Hui	17170153

Contents

CHAPTER 1 INTRODUCTION

1.1 Background

Gone are the days when people relied solely on their savings to ensure their financial security in the future. Savings may not be enough to provide financial security in today's environment. It is also possible that money sitting in the savings account is not serving the goal. This is due to two factors: first, the idle cash in the bank account is an opportunity cost because it cannot make more money; second, it lacks the ability to beat inflation. Investing gives financial security in the present and the future as it enables one to increase his wealth while also generating inflation-beating returns. With the advancement of time, investing options have amplified manifold. Today, there is no shortage of financial instruments, such as stocks, ETFs, bonds mutual funds and REITs. However, with multiple options, the complexity and confusion have also heightened. Stocks are one of the most attractive options from a comparative earning capacity point of view, in addition to their relatively low barriers to entry. A strong and well-balanced stock portfolio helps in generating higher returns, mitigating risk, and substantially accumulating wealth over time. Portfolio optimization is a crucial part of managing risk and maximizing returns from stock investments.

There are three sets of performance measurement tools to assist with portfolio evaluations, which are the Treynor, Sharpe, and Jensen ratios (Troy, 2021). The Sharpe ratio was developed by Nobel laureate William F. Sharpe and is used to help investors understand the return of an investment compared to its risk (Fernando, 2021). The Sharpe ratio adjusts a portfolio's past performance—or expected future performance—for the excess risk that was taken by the investor. A high Sharpe ratio is good when compared to similar portfolios or funds with lower returns. However, the Sharpe ratio has several weaknesses, including an assumption that investment returns are normally distributed. The Sharpe ratio is calculated as follows:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p}$$

where:

R_p = return of portfolio
R_f = risk-free rate
σ_p = standard deviation of the portfolio's excess return

We subtract the risk-free rate from the return of the portfolio. We use average Malaysia bond yield rate as the baseline for the risk-free rate: http://www.worldgovernmentbonds.com/bond-historical-data/malaysia/10-years/. We use the rate of year 2020 and 2021.

Divide the result by the standard deviation of the portfolio's excess return. The standard deviation helps to show how much the portfolio's return deviates from the expected return. The standard deviation also sheds light on the portfolio's volatility.

1.2 Problem Statement

Even the most experienced stock pickers, who have access to millions of dollars in research, frequently underperform basic index funds. According to the performance data from Dalbar's 2018 Quantitative Analysis of Investor Returns, actual returns to investors are significantly below the corresponding index returns. This performance gap leads to staggering differences in one's portfolio over time. The reason the average stock investor underperforms the market so consistently is that human is hardwired to make irrational decisions when it comes to money.

Besides, important investment decisions are frequently made by the average investor based on a limited quantity of data available, without considering other relevant data (Why, 2019).

1.3 Objectives

The objectives of this project are:

1. To formulate an objective function that maximizes the Sharpe Ratio by optimizing the weightage of the average rate of return of the stock investments.
2. To optimize the objective function based on the defined constraints using multiple optimization methods.
3. To compare the performance of multiple optimizations methods on optimizing the objective function.

1.4 Objective Function

The objective function used in this project is:

$$\max\left(\frac{\sum_{i=1}^{N} W_i\mu_i - R_f}{\sqrt{\sum_i \sum_j W_i W_j \sigma_{ij}}}\right)$$

subject to:

$$\sum_{i=1}^{N} W_i = 1$$

$$0 \leq W_i \leq 1$$

where:

W_i = weightages of the stocks
μ_i = average return of the stock investments
R_f = risk-free rate

1.5 Dataset

In this project, we would like to investigate the performance of multiple optimization methods on optimizing the portfolio consisting of stocks listed in Bursa Malaysia. The historical daily stock prices (from 2020 January to 2021 December) of top 20 FTSE Bursa Malaysia KLCI companies were downloaded (in the form of CSV files) from Yahoo Finance[1]. The companies are listed as follows:

- Dialog Group Berhad
- Top Glove Corp Berhad
- Digi Berhad
- Petronas Berhad
- Maxis Berhad
- Petronas Dagangan Berhad
- Tenaga Nasional Berhad
- IHH Healthcare Berhad
- Petronas Chemicals Group Berhad
- Telekom Berhad
- Genting Malaysia Berhad
- Nestle Berhad
- Sime Darby Berhad
- Public Bank Berhad
- MISC Berhad
- Genting Berhad
- Kuala Lumpur Kepong Berhad
- IOI Corp Berhad
- Maybank Berhad
- Hong Leong Financial Group Berhad

[1] https://finance.yahoo.com/

CHAPTER 2 METHODOLOGY

2.1 Methodology

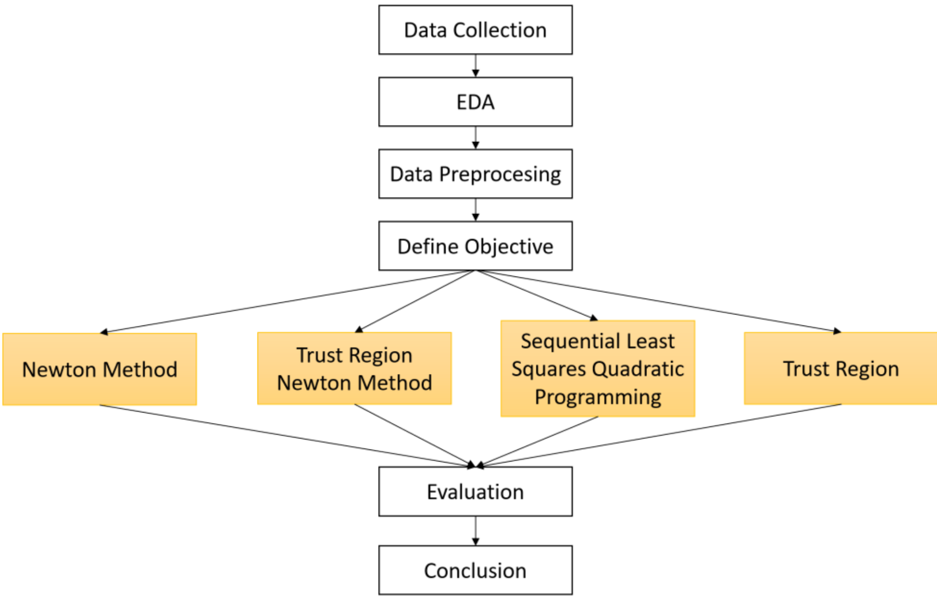

Figure 2.1 Flowchart of Methodology

Above figure shows the overview of the methodology workflow. Firstly, the stock price of 20 listed companies in KLSE from 2020 January to 2021 December were collected. Then we performed exploratory data analysis to see how the stocks perform over the past 2 years. Then we preprocessed the data and convert them into the right data frames. We also calculated their risk-free rate, standard deviation, average historical returns, and variance covariance returns which will be used as the parameters of the Sharpe Ratio formula. To have the best portfolio of the given stocks, we define an objective function to maximizes the Sharpe Ratio by optimizing the weightage of the average rate of return of the stock investments. In this study, we have tried 4 different algorithms to solve the objective function. After that, we evaluate these 4 methods by using Iteration, Convergence, Value of Objective Function, Max Sharpe Ratio, Annualized Risk and Annualized Expected Portfolio Return.

2.2 Newton Method (Newton-CG)

The Newton method approach will be the initial method for portfolio optimization. Newton's method (sometimes called Newton-Raphson method) uses first and second derivatives and indeed performs better in searching suitable direction. Given a starting point, construct a quadratic approximation to the objective function that matches the first and second derivative values at that point. We can obtain a quadratic approximation to the twice continuously differentiable function using the Taylor series expansion of about the current point, neglecting terms of order three and higher.

$$f(x) \approx f\left(x^{(k)}\right) + \left(x - x^{(k)}\right)^T g^k + \frac{1}{2}\left(x - x^{(k)}\right)^T F\left(x^{(k)}\right)\left(x - x^{(k)}\right) \triangleq q(x)$$

Where, for simplicity, we use the notation

$$g^{(k)} = \nabla f\left(x^{(k)}\right)$$

Applying the FONC to q yields

$$0 = \nabla q(x) = g^{(k)} + F(x^{(k)})(x - x^{(k)})$$

Hence, if Hessian of the function is positive infinite. In this case we take steps in direction of the gradient ∇f to increase the function (to find maxima) and in the direction of the negative gradient −∇f to decrease the function (to find minima).

2.3 Trust Region (trust-constr)

Trust region strategy is based on the definition of an area that extends from the current search point and allows the model to predict the goal function with some precision. To proceed from the beginning, point to the minimizer in the trust zone, a step is calculated. The assumption is that a quadratic model will suffice to approximate the original functions in this way.

In general, the derivation of the trust region radius, Δk, has a significant impact on the trust region method. This radius can't be too big because the anticipated minimizer will be too far away from the actual minimizer, and it can't be too small because the steps will be too little, causing delayed iteration toward the solution.

As previously stated, the quadratic model drives the trust region method:

$$m_k(p) = f_k + {g_k}^T p + \frac{1}{2} p^T B_k p$$

The f_k is the function value at the beginning point x_k, g_k is the gradient of f_k, which is $\nabla f(x_k)$, and B_k is the approximation of Hessian.

For the solution of the subproblem, the general mathematical definition is:

$$\min m_k(p) = f_k + {g_k}^T p + \frac{1}{2} p^T B_k p \qquad \text{subject to: } ||p|| \leq \Delta_k$$

$\Delta_k > 0$ is the trust-region radius.

The ratio of must be determined to see if the model m_k is a fair approximation of the real function. The formula for p is as follows:

$$\rho = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

The actual decrease is in the numerator, while the expected reduction is in the denominator. The model will decide whether to adjust the radius of the trust-region based on the value of p. The model will lower the radius for the following iteration if the value of p is negative or extremely close to zero. If, on the other hand, p is almost equal to 1, the approximation is satisfactory, and the model will consider expanding the trust-region in the next iteration.

2.4 Trust Region Newton Method (trust-ncg)

Newton's method with a trust region is designed to take advantage of the second-order information in a function's Hessian, but with more stability that Newton's method when functions are not globally well-approximated by a quadratic. This is achieved by repeatedly minimizing quadratic approximations within a dynamically sized "trust region" in which the function is assumed to be locally quadratic.

2.5 Sequential Least Squares Quadratic Programming (SLSQP)

The sequential least squares quadratic programming approach is the next portfolio optimization strategy. This is a non-linear optimization approach for minimization that is subject to an equation or a vector of non-linear functions that act as inequality restrictions since it is a constrained optimization method. To put it another way, this method solves a sequence of quadratic algorithms by iterating over several approximations. The problem is stated in its most basic mathematical form as follows:

$$\min_{x \in R^n} f(x) \qquad \begin{array}{l} \text{subject to:} \\ c_i(x) \leq 0 \\ i = 1, 2, \dots, m \end{array}$$

There are various aspects to consider when programming the function. For the constraints, those in the equation forms should be translated into two opposing inequality constraints, such as $c(x) = 0$ being processed into $c(x) \leq 0$and $-c(x) \leq 0$. Because such transformation is inconvenient, a practical alternative that takes into consideration the equation, linear restrictions, and boundaries is the insertion of slack variables, which results in the construction of a better objective function:

$$\min_{x \in R^n} f(x) \qquad \begin{array}{l} \text{subject to:} \\ A^T x = b \\ c(x) = 0 \\ l \leq x \leq u \end{array}$$

In this case, a condition that has no bounds allows the variables of u_i and -l_i to be large numbers. The objective function can also be written in a different representation:

$$\min_{x \in R^n} f(x) \qquad \begin{array}{l} \text{subject to:} \\ \begin{pmatrix} x \\ A^T x \\ c(x) \end{pmatrix} \leq u \end{array}$$

According to the function, it also enables users to use the equation of $l_i = u_i$ for optimization.

CHAPTER 3 RESULTS & DISCUSSION

Experiments are conducted to gain insights into the performance of each optimization method. For each optimization method, the resulting visualization are thoroughly display and discuss. Other than that, the results are combined and aggregated to be shown with their specific details in Table 1. As a whole, the research aims to identify the allocation of funds in a portfolio to gain maximum profit.

To cross reference, we used Malaysia bond yield rate as the base rate for the comparison. According to the website, http://www.worldgovernmentbonds.com/bond-historical-data/malaysia/10-years/, a trusted source to obtain information about bond yields of all countries, it was found that the average bond yield rate in Malaysia is 3.148% (2.685% in 2020 and 3.611% in 2021). Therefore, it can be concluded that any investment with a return below 3.148% is considered a bad investment. If any analysis results in a negative Sharpe ratio, it either means the risk-free rate is greater than the portfolio’s return, or the portfolio's return is expected to be negative. Thus, based on this assumption, any return below 3.148% should generate a negative Sharpe Ratio in our study.

3.1 Newton Method

Based on the chart in Figure 3.1, the chart suggests putting more weight (% of funds) in buying TM, IHH, MAYBANK, GENM and PCHEM shares. It can be observed the annualized expected portfolio return computed by Newton Method is only 0.165% which is much lower than the bond yield rate. The maximum Sharpe Ratio computed by the Newton Method is -25.965. Therefore, we can conclude for the year 2020 and 2021, it was not a right time to invest in the KLSE. This is because the best return in the top 20 stocks in the KLSE was only 0.165%. Investing in the bond market would have gotten a better return.

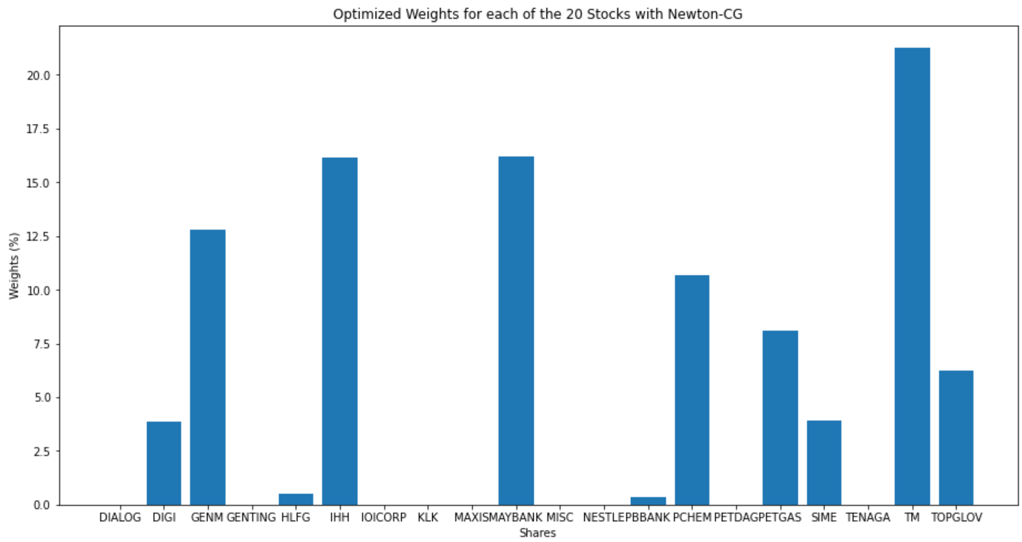

Figure 3.1 Optimized Weights with Newton Method

3.2 Trust Region Newton Method

Based on the chart in Figure 3.2, the chart shows approximately the same weights suggestion as the Newton Method from the previous section. Other than that, it can be observed that the maximum Sharpe Ratio and the expected portfolio return of this method are also almost the same as Newton Method, with values of -25.968 for Sharpe Ratio and 0.165% for the expected return. Similarly, we arrived at the same conclusion as before, the year 2020 and 2021 is not a period to invest in the KLSE stock market based on results from the top 20 KLSE stocks. Investing in the bond market would have yield better returns.

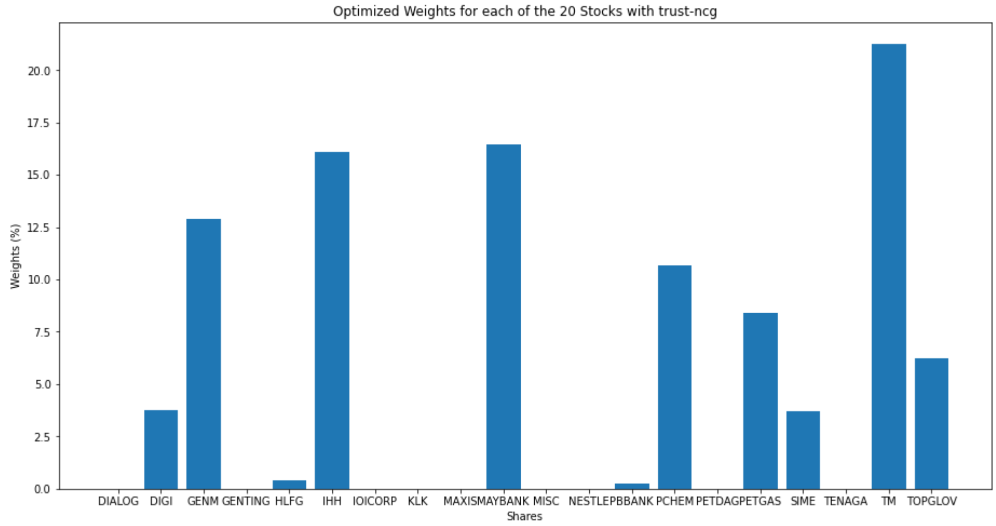

Figure 3.2 Optimized Weights with trust-ncg

3.3 Sequential Least Squares Quadratic Programming Method

Based on the chart in Figure 3.3, it is observed that the maximum Sharpe Ratio and expected returns from the portfolio slightly differs when compared with Newton Method and Trust Region Newton Method, with values of -20.761 for Sharpe Ratio and 0.263% for the expected return. However, the weightage suggestions are different from the first two algorithms. Based on Figure 3.2, the chart suggests the inclusion of 5 shares (e.g.TM, IHH, TOPGLOV, PETGAS and PCHEM in the portfolio.)

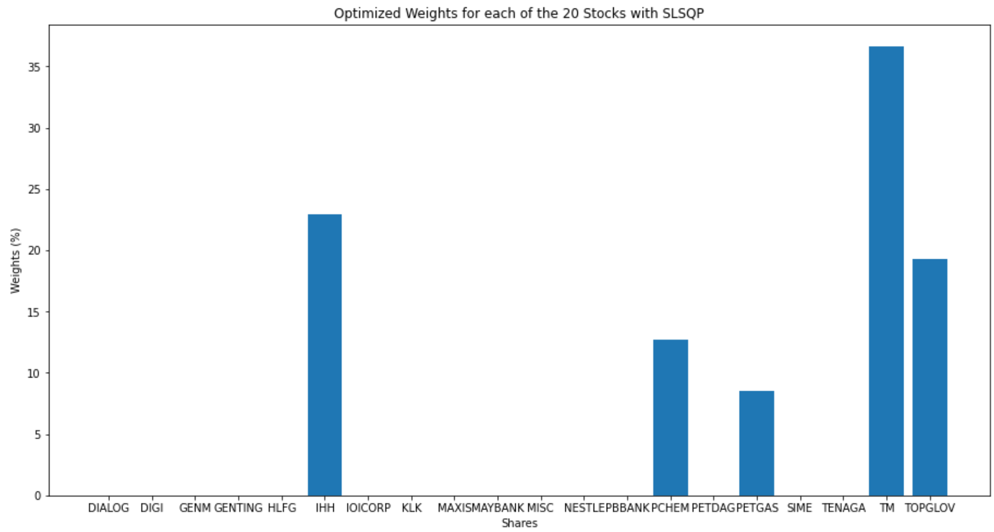

Figure 3.3 Optimized Weights with SLSQP

3.4 Trust Region Method

Based on the visualization in Figure 3.4, it can be observed that the maximum Sharpe Ratio and expected portfolio return are the lowest when compared with all the methods. The values for the Sharpe Ratio was -27.226 and the expected return was 0.127%. Other than that, we noted that the weightage suggestion for this method is quite interesting as it is very different from the other three algorithms. This is because the weightage for each stock is quite evenly distributed.

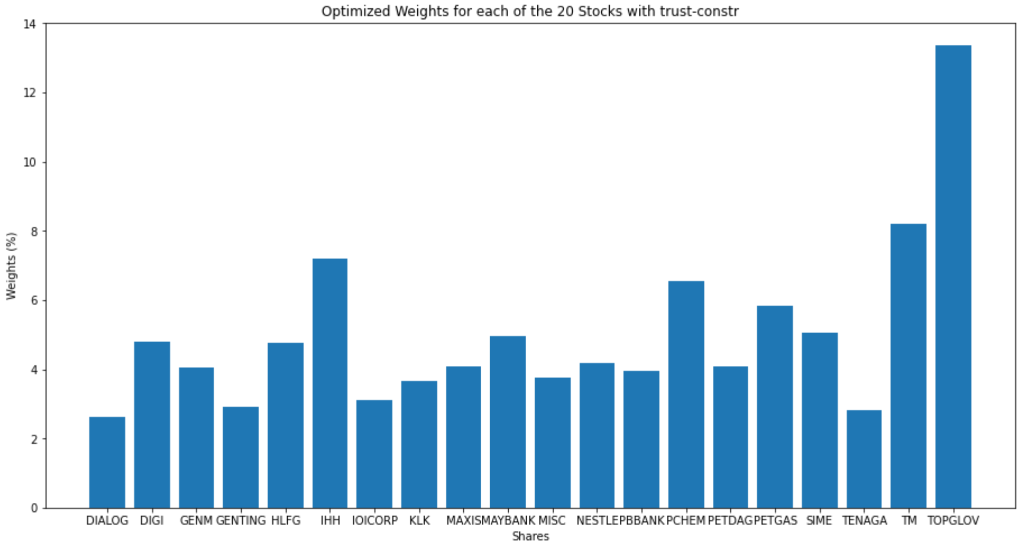

Figure 3.4 Optimized Weights with trust-constr

3.5 Comparative Results from All 4 Optimization Methods

The following Table 1 shows the comparison of four algorithms. The SLSQP takes least iteration to converge which is 6 iterations and the longest is Newton Method. The highest maximum value returned by the objective function is 0.13013 using the Newton Method and the lowest is 0.04303 using the Trust Region method. The expected return computed by these four algorithms are between the ranges of 0.127% to 0.263%. All four algorithms gave negative Sharpe Ratio. This means that the performance of the share is worse than the performance of the Malaysia bond.

	Newton-CG	Trust-NCG	SLSQP	Trust-Constr
Iteration	13	9	6	10
Converge	Yes	Yes	Yes	Yes
Value of Objective Function	0.13013	0.12830	0.07870	0.04303
Max Sharpe Ratio	-25.965	-25.968	-20.761	-27.226
Annualized Risk	0.115%	0.115%	0.139%	0.111%
Annualized Expected Portfolio Return	0.165%	0.165%	0.263%	0.127%

Table 1: Max Sharpe Ratio and Expected Return

The following Table 2 shows the weightages distribution of all the top 20 KLSE stocks with regards to the 4-optimization method used. As discuss previously, SLSQP methods highlights only 5 stocks, meaning the risk-rate of this method is the highest, as the portfolio is not diverse. The opposite could be said with the Trust Region method, as the Trust Region method highlights all stocks, meaning a higher diversification. However, this method provides the least return.

	Newton-CG (%)	**Trust-NCG (%)**	**SLSQP (%)**	**Trust-Constr (%)**
DIALOG	0	0	0	2.624
DIGI	3.843	3.77	0	4.808
GENM	12.777	12.863	0	4.057
GENTING	0	0	0	2.92
HLFG	0.513	0.388	0	4.748
IHH	16.17	16.062	22.902	7.187
IOICORP	0	0	0	3.12
KLK	0	0	0	3.656
MAXIS	0	0	0	4.095
MAYBANK	16.201	16.439	0	4.969
MISC	0	0	0	3.758
NESTLE	0	0	0	4.18
PBBANK	0.335	0.249	0	3.943
PCHEM	10.694	10.655	12.743	6.563
PETDAG	0	0	0	4.097
PETGAS	8.092	8.385	8.507	5.831
SIME	3.907	3.729	0	5.061
TENAGA	0	0	0	2.825
TM	21.242	21.226	36.604	8.21
TOPGLOV	6.226	6.236	19.244	13.349

Table 2: Weightages of each of the 20 Stocks

3.6 Volatility of KLSE stocks

Based on the above section, the annualized expected portfolio rate for each optimization method is quite low. Thus, we plotted the price change of the top 20 stocks to check the reasoning behind our results. It was observed from Figure 5 that the fluctuations in stock prices in the year 2020 and 2021 was high. This was different than our expectation as blue chips stocks were supposed to be more stable with little fluctuations. Upon careful inspection, we observed that this could be due to the pandemic.

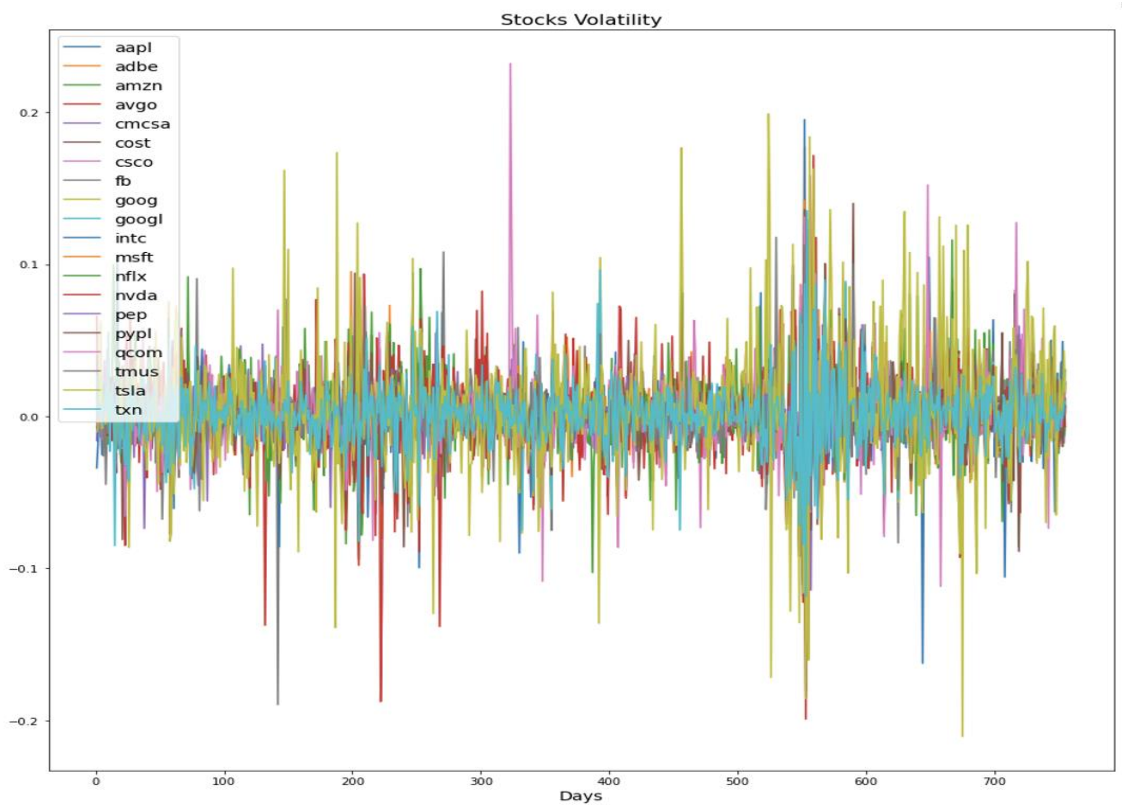

Figure 5: Stock Volatility from Year 2020 to 2021

CHAPTER 4 CONCLUSION

The stock market is an interesting domain to be study. This is due to reasoning such as it provides investors with a method to gain income outside of their salary ranges. However, this is not without risks. Therefore, this research is important to investors as it provides investors with more insights into the stocks, they are investing in. Optimization methods are used in this research to show the varying results. It is noted that the best expected portfolio return is from the SLSQP method. However, it is also concluded that this optimization method is also the riskiest as the weightage suggestions lacks stock diversification. In summary, we concluded that using optimization methods, we showed that we can select various portfolio combinations to improve the expected stock returns. However, as the covid situation in 2020 and 2021 affected the stock market adversely, it was better to invest in the bond market. As for future work, since the 20 companies expected return is not as good as bonds, researchers may build upon our idea and try to include different stocks combinations or even applied this methodology on the top 20 ACE market stocks.

The following are the links to the dataset and the source code for this study:

- https://colab.research.google.com/drive/18dXFKS_32VLGsA_0Z8p8gxzp0R_zIGSR?authuser=1#scrollTo=WS0Mrp7Y58eE
- https://github.com/zsqy/stock-portfolio-optimization

CHAPTER 5 REFERENCES

Fernando, J. (2021, April 10). Sharpe Ratio. Retrieved from Investopedia: https://www.investopedia.com/terms/s/sharperatio.asp

Troy, S. (2021, December 12). Measuring Portfolio Performance. Retrieved from Investopedia: https://www.investopedia.com/articles/08/performance-measure.asp

Why It's So Difficult to Manage Your Own Portfolio. (2019, November 18). Model Investing. Retrieved June 14, 2022, from https://modelinvesting.com/articles/why-its-so-difficult-to-manage-your-own-portfolio/