515-50-B-2021/Summer

Final project report

Aug, 7 2021

Team:
Saurav Mawandia
Sougandh Kohli

Contents

1. Introduction

Where to invest money??
Is investing in financial institutions still risky??
Are tech stocks the new "it" stocks??

These are some of the questions we have trying to answer in this paper. We are attempting to build an algorithm for assessing the past performance of the stocks, calculating the best possible portfolio weights and predicting the future volatilities of our portfolio return. We also plan to calculate the diversification benefits of including such varied sectors and stocks using standard deviation, diversification ratio and concentration ratio.

Initially in this paper we will be analyzing industries to see which has a better performance, stocks to determine which has a higher average return and lower volatility before, during the Pandemic, for us to built an optimized portfolio. We will then fit some of the known distributions such as Generalized Hyperbolic distribution and Normal Inverse Gaussian distribution on our portfolio returns, in order to predict the Value at Risk and Expected Shortfall of our portfolio. Generalized Autoregressive Conditional Heteroskedasticity, also known as GARCH model will be used to predict the future volatilities of stock prices included in our portfolio.

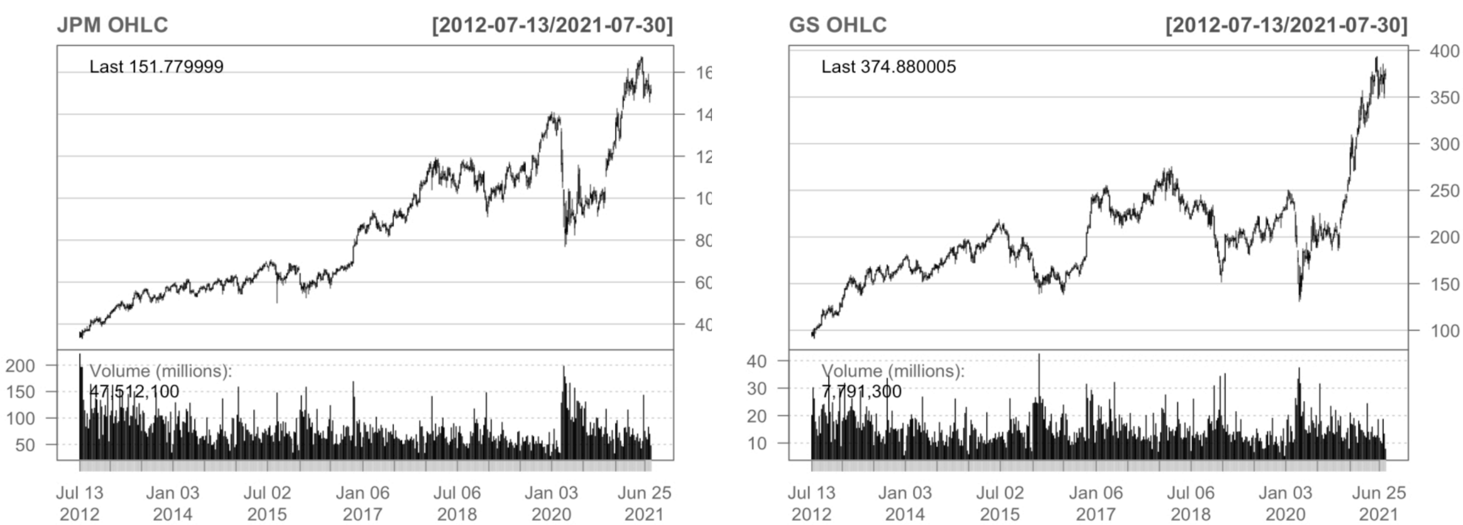
While a successful prediction of stock can yield significant profits, the Random Walk Theory claims that stock prices cannot be predicted using the historical prices. It is however observed that against these fallacies, trading companies like JP Morgan, Berkshire Hathaway and Goldman Sachs have consistently made profits based on investment predictions.

2. Portfolio Description

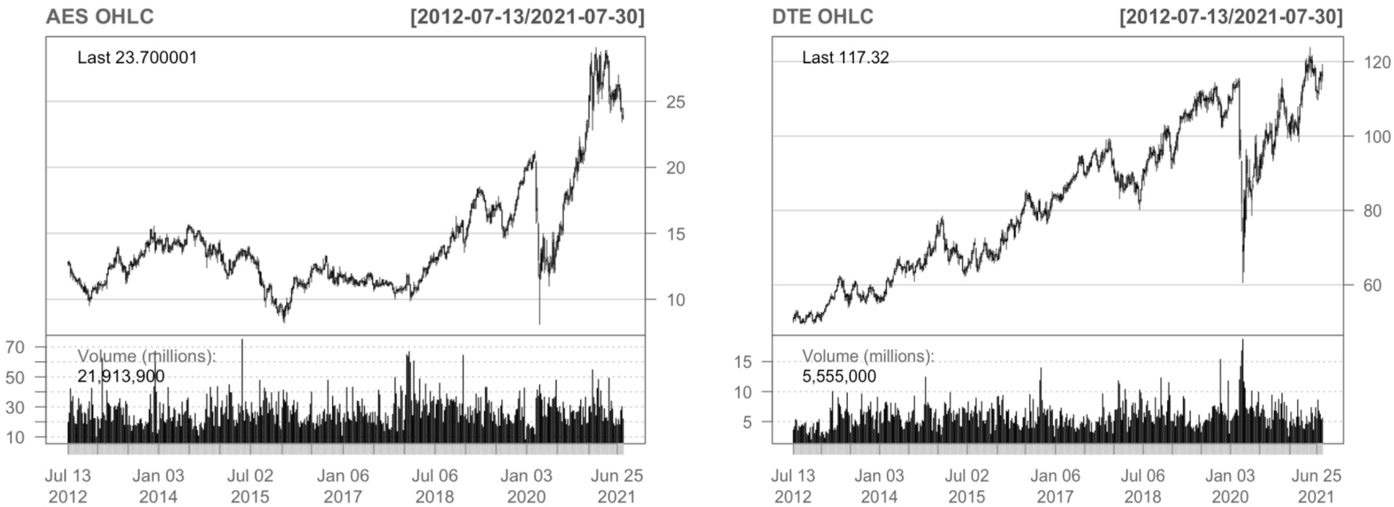
Since this algorithm is attempting to predict performance of any kind of portfolio with varied underlying stocks, not much attention has been paid to selection criteria or correlation between the stocks. We have randomly selected three sectors and two stocks within each sector for our analysis. They have been sourced from yahoo finance. For our paper, risk-free rate is the 3-month treasury bill, which has been sourced from the US department of Treasury website.
For the information purpose, the stocks have listed as below:

1. Financial Services - JP Morgan (JPM), Goldman Sachs (GS)
2. Utilities- DTE energy co(DTE) and AES corp.(AES)
3. Tech: Amazon(AMZN) and Apple(AAPL)

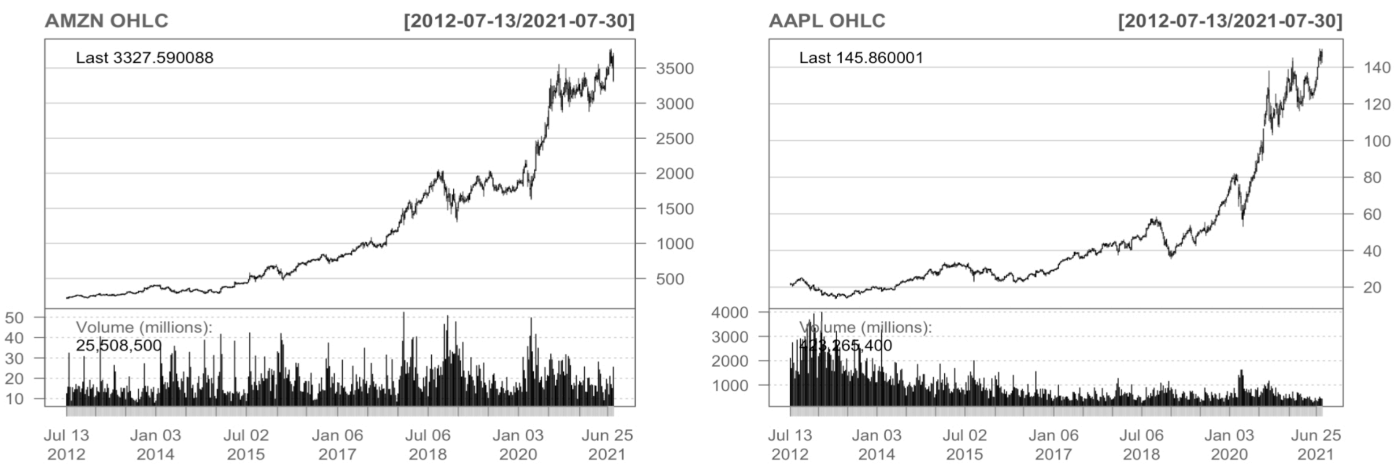
The initial time period selected for this analysis was July 2012-July 2021. However, for our observation period, we have only focussed on the pre and during Pandemic times i.e. July 2019-July 2021, since it captures an economic event. With this observation period, we have a more balanced data since we capture around 9 months of pre-pandemic, 6-months of pandemic and then 9-months of post-pandemic behavior. Please note, lockdown has been in place even in post pandemic times, however, we have seen the performance getting back to pre-covid times after the 6-month observation period.
We have also included the Open High Low Close plots for each of our stocks included in this paper.



Finance sector stocks



Utility sector stocks



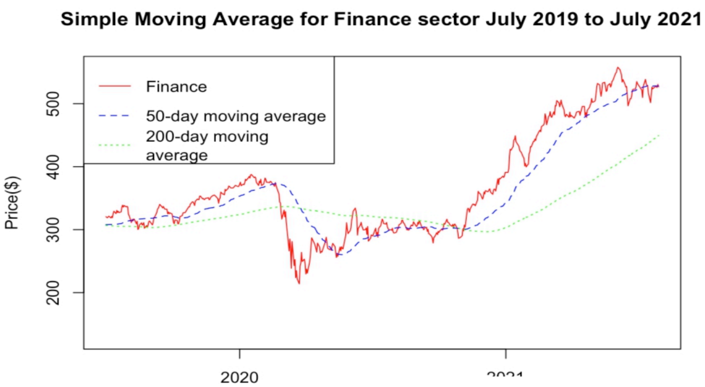
Technology sector stocks

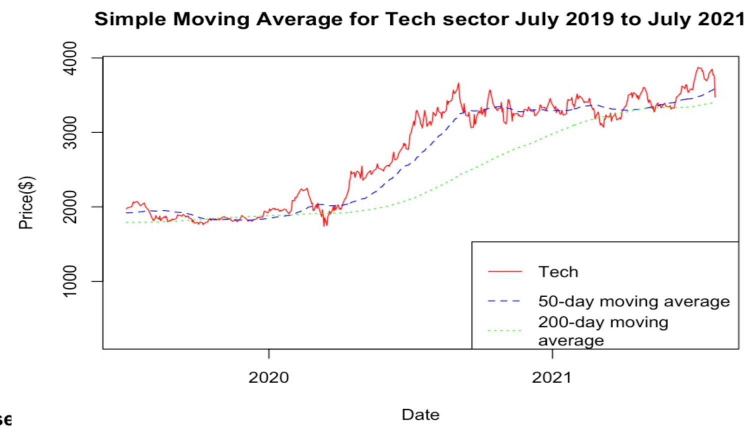
We can see a dip in all the plots above around March 2020, right when pandemic hit and since then the price returns have been higher than even before the pandemic period hence making an interesting observation window for our paper.

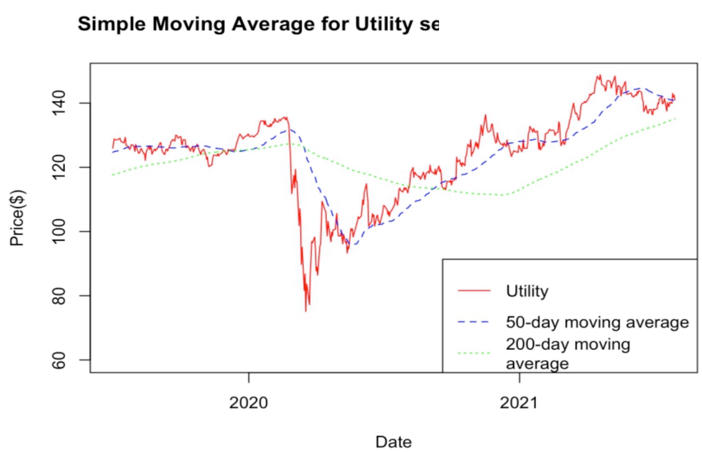
We also tried to capture the benefit from investing $1 at the start of our observation period until July 2021. Clearly we can see the investor would have earned the most from investing in Tech industries, where even during the Pandemic the return had the lowest dip. This in line with the impact observed with this economic event, since more reliability on work from home technology resulted in minimal impact on business as usual for this sector. Highest impact was observed in Utility where we was the lowest earnings compared to other two sectors.

2.1 Moving Average

Next we used the concept of moving average to evaluate our sector performance and assess the future outlook. Moving average is an arithmetic mean of a certain number of data points. The difference between a 50-day moving average and a 200-day moving average is the number of time periods used in the calculation. When the 50 or 200 day average line is below the actual price of the stock then there is a downward trend in the future, implying that our stock prices are inflated and they will go down in the coming days. Both of our sectors, except Technology, saw Moving average trending upwards just after covid period and then back to downward trend in post pandemic period, implying that Technology was still behaving strongly even during the pandemic.

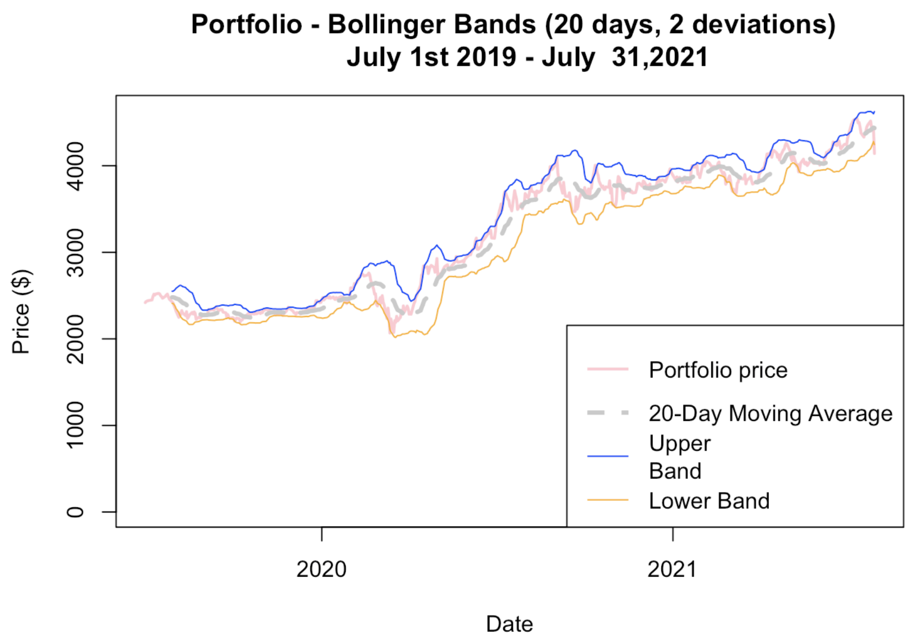






2.2 Bollinger Bands

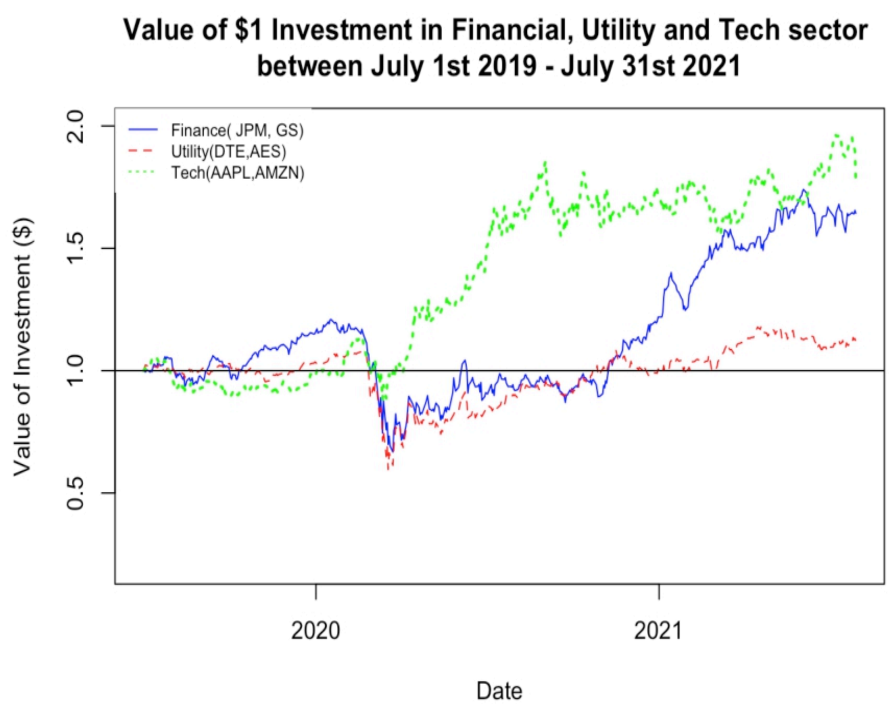
In the plot below, Bollinger Bands bracket the 20-day Simply Moving Average of the stock with an upper and lower band, along with the daily movements of the stock's price included in the portfolio. Because standard deviation is a measure of volatility, when the markets become more volatile the bands widen (during March 2020-Aug 2021); during less volatile periods, the bands contract.



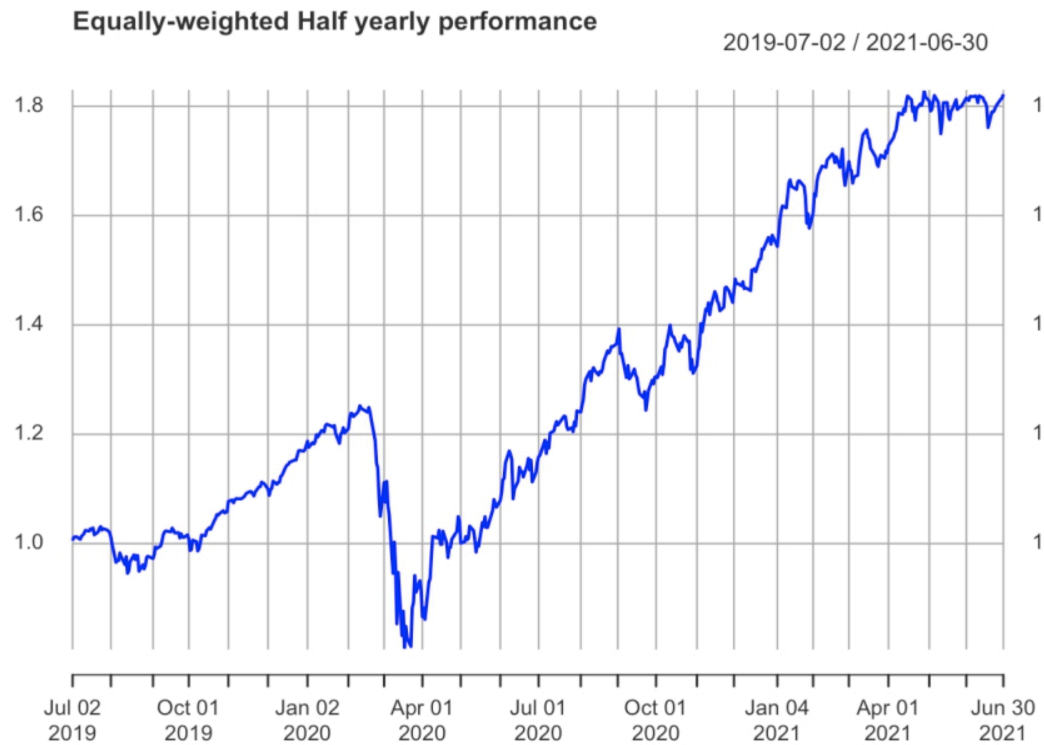
Lastly, we also created a plot to capture the price returns vs Total returns (which includes the impact of price returns and dividends) and observe that the total return on our portfolio(gray line) is slightly higher than just price returns from our stocks included in the portfolio.

3. Creating a portfolio

Portfolio return is simply weighted average of return from each asset included in the portfolio. For the purpose of this paper, we are focussing on equally weighted portfolio. This is because we are attempting to make user understand how to construct an algorithm for stock investing and the important aspects that should be assessed before anyone starts investing. This tool can be easily modified to user's expectation.
An equally weighted portfolio invests equal amount of money in each stock irrespective of return expectation or risk factor. No preference is given to firm's market share or future outlook. It is the most straight forward approach for portfolio building.



We have captured the half yearly performance of our EW portfolio during our observation period. This plot shows that our portfolio performance dipped just at the start of the pandemic (March 2020) and since then has picked up even better than before the pandemic.



Equally Weighted Vs Half Yearly Performance

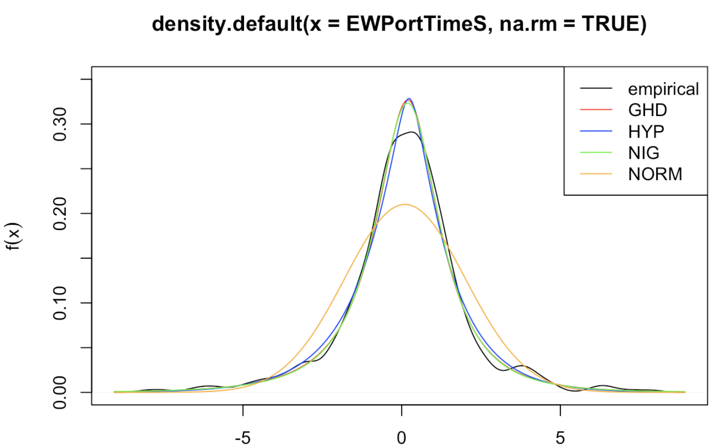
4. Distribution fitting

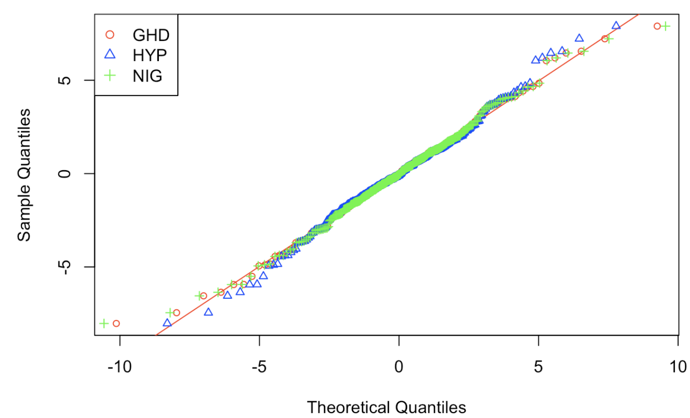
The purpose to fit a distribution is to predict the behavior of stock returns and capture the tail probabilities accurately since as an investor we would want to know our future losses in our investments. And if our stock returns fit a distribution, then we can easily model these two.

In our paper, we have focussed on three known distributions to fit our stock returns, namely:

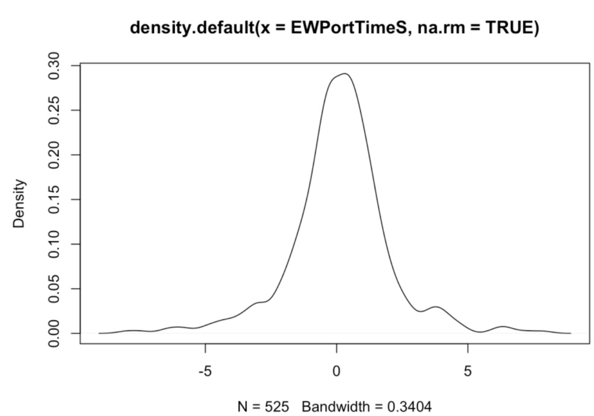
- Generalized Hyperbolic distribution
- Hyperbolic distribution
- Normal Inverse Gauss distribution

In general, stock returns are usually not independently distributed and show a leptokurtic behavior (kurtosis > 3). This means an investor may see broader fluctuations in returns, as shown below for our portfolio in consideration





We also created a density plot and Q-Q plot for our distributions and the empirical data, to see and contrast the behavior . By just looking at the plot below, we can see that NIG distribution seems to be the best fit for our portfolio returns.



5. Best fit model

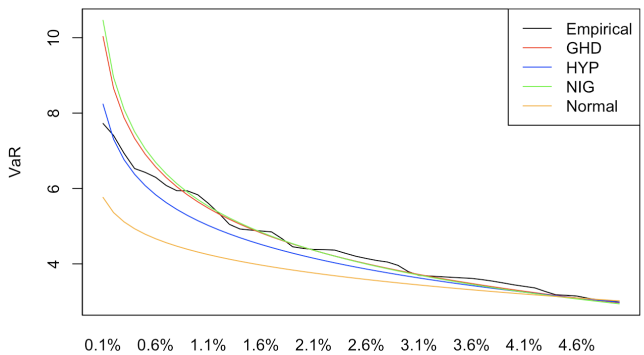
In order to see which model works best for our returns, we also performed some statistical analysis. We used the Akaike information criterion (AIC) in the scope of the generalized hyperbolic distribution class to assess which model is the best fit.

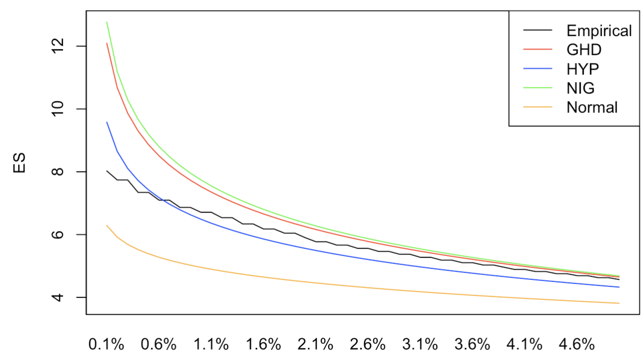
The lower the AIC value, better is the model, and smaller is the LLM statistic, the better is the fit. We also used the likelihood ratio test (readily available in Rpackage) to perform hypothesis testing (hypothesis being that our dataset is significantly different) and used the p-value to see if our H0 is true or not. The results have been summarized in the table below.

model symmetric lambda alpha.bar mu sigma gamma aic llh ## 8 NIG TRUE -0.50000 0.47277 0.15079 1.9311 0.00000 2069.0 -1031.5 ## 3 NIG ## 6 ghyp ## 1 ghyp ## 10t ## 7 hyp. TRUE 1.00000 0.26353 0.15048 1.8407 0.00000 2074.6 -1034.3 ## converged n.iter FALSE -0.50000 0.47592 0.21013 1.9263 -0.10686 2070.3 -1031.2 TRUE -0.26886 0.48899 0.15181 1.9200 0.00000 2070.9 -1031.4 FALSE -0.25085 0.49023 0.21364 1.9146 -0.11024 2072.1 -1031.0 TRUE -1.37365 0.00000 0.14747 2.2544 0.00000 2074.1 -1034.1 ## 8 TRUE 104 ## 3 TRUE 121 ## 6 TRUE 189 ## 1 TRUE 510 ## 10 TRUE 106 ## 7 TRUE 106	## $statistic ## L ## 0.90068 ## ## $p.value ## [1] 0.64738 ## ## $df ## [1] 1 ## ## $H0 ## [1] TRUE

6. Risk calculations

One of the most important aspect of fitting these distributions is to model the losses. In this paper, we have modeled Value at Risk also known as VaR, which tells us with any confidence level of 1% or 5%, our portfolio will not have losses more than the VaR amount. However, Expected Shortfall is fitting a distribution in tail losses, further giving us the losses expected from any portfolio in worst case scenario. Expected Shortfall is also known as Tail VaR.

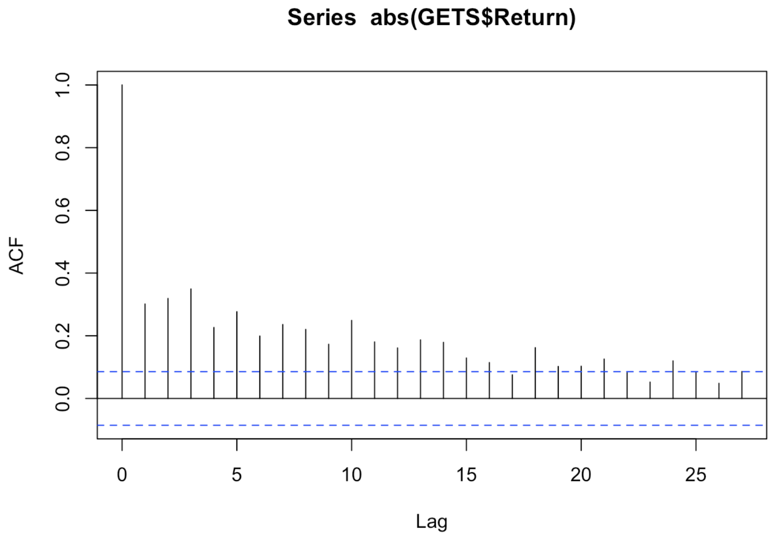


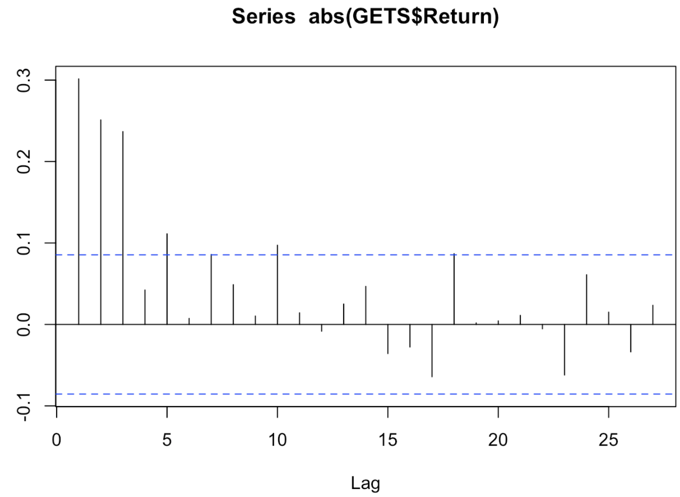


In the charts above, we have plotted the predicted VaR and ES for various confidence levels. Compared to empirical performance, NIG distribution is more conservative, since at 0.1% VaR and ES is much higher than observed. However, they converge towards each other as our confidence level reduces, i.e. we move towards the right of the x-axis.

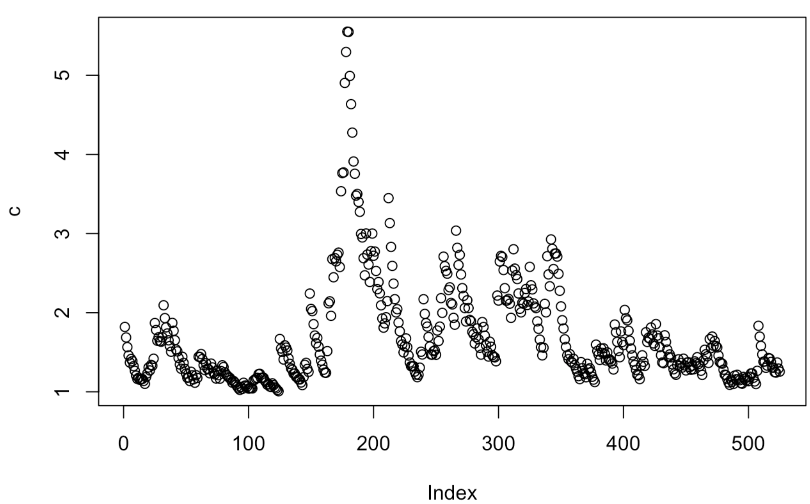
6.1 GARCH Model

The next step in our algorithm construction after distribution fitting is to predict the price volatilities of our underlying stocks included in the portfolio. For this exercise we had multiple options to select. For purpose of this paper however, we will only be focussing on GARCH model. GARCH model was selected since it offers an additional element when compared to the ARCH model. This model also includes the lagged endogenous variables in the variance equation—that is, now the conditional variance will not only depend on past squared errors but also on lagged conditional variances, thus providing more accurate predicted values.
In order to select the lag period for the GARCH model input, we created a plot for Autocorrelation and Partial autocorrelation to select the lag period, and we can clearly see that by selecting lag period as 1, we are capturing the maximum variance in our returns data.





In the below plot, we can see the volatility in our portfolio price returns fluctuating more when Pandemic started and then falling back to prePandemic behavior. However, the volatility is more concentrated within c=3, thus highlighting that our portfolio returns will not be extremely volatile.



The VaR from our GARCH model simulation for our portfolio is at 1.0018.
Further, the predicted Expected Shortfall for each our stocks is

GS	JPM	AAPL	AMZN	DTE	AES
-7.0036	-7.5542	-8.1439	-8.7308	-9.3042	-9.005

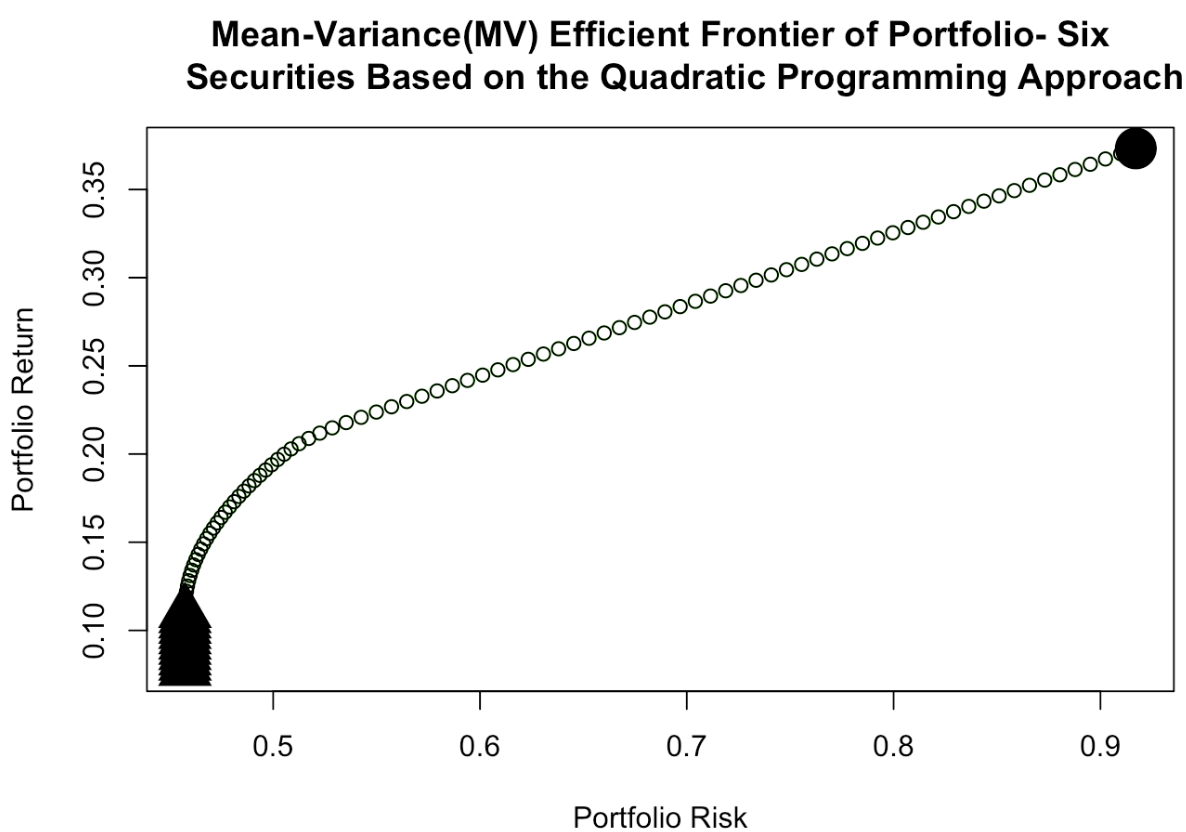
7. Portfolio optimization

The final step in our algorithm is to optimize our portfolio returns. For us the goal is to minimize risk and maximize return. We have used the Markovitz Mean variance Efficient Frontier theory to create an efficient frontier using all possible combinations of weights in our portfolio. We then used a tangent portfolio using risk-free asset and find an optimized portfolio for ourselves.

From our selected stocks, below are the corresponding average returns in our observation period:

```
  Avg.Ret
## JPM   0.0958
## GS    0.0773
## AES   0.1305
## DTE   0.1472
## AAPL  0.0983
## AMZN  0.3733
```

Using the minimum and maximum return from this, we created an efficient portfolio. Below is the plot, where the black dot is our resulting most efficient portfolio. If however, we have a risk free asset included in our portfolio, the results would have been different. This strategy makes us riskier, but only for purpose of this paper we have been selected this strategy.



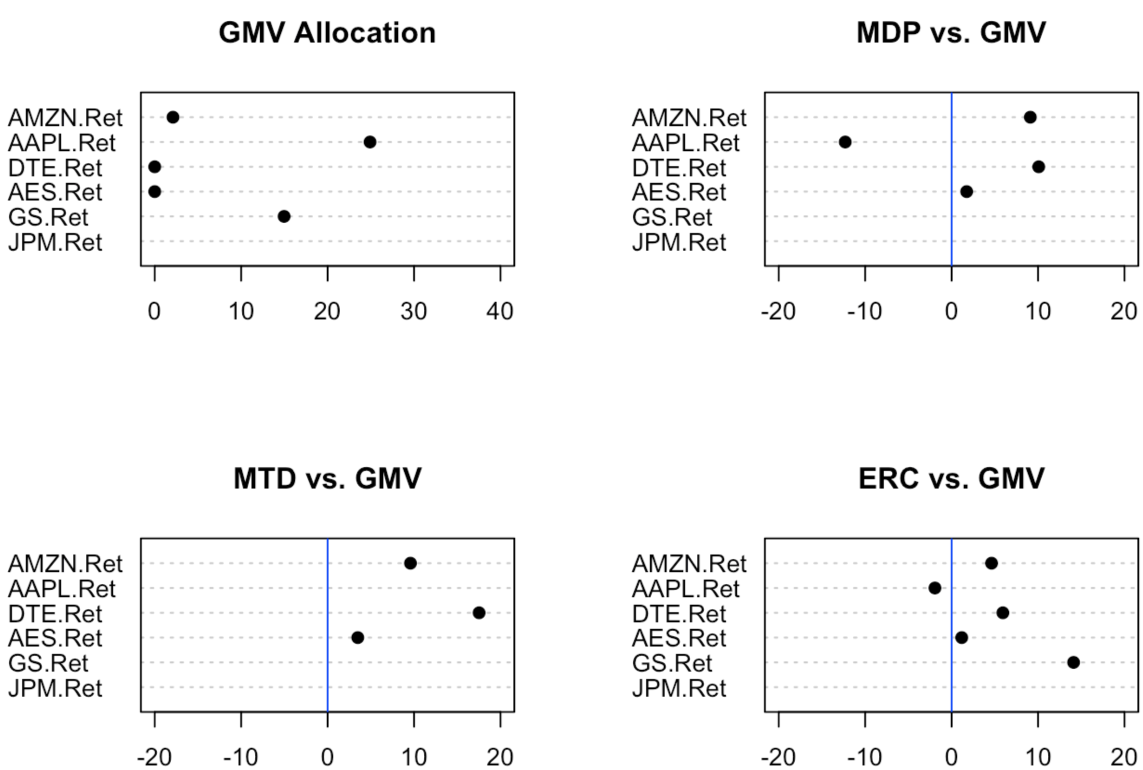
8. Diversification benefit

The utility of an investor is increased when the return is allocated to an asset mix instead of holding a single risky or risk-free asset. Diversification has two dimensions. The first dimension is the question of the underlying common characteristic with respect to which the assets are diverse. The second considers the question of how to measure the degree of diversification with respect to this characteristic. For purpose of this paper, we have used four portfolios to capture the diversification benefit in our selected portfolio, and calculate the appropriate weight at which we get least variance and maximum return.

The portfolios considered are:

- The global-minimum variance portfolio.
- The most diversified Portfolio,
- The equal-risk contributed Portfolio
- Minimum tail-dependent portfolio

The results from our simulation have been consolidated in the plots below, where can see GMV is including all our stocks and is providing good weights to high performing stocks like AAPL and minimal weight to low performing stock such as DTE and AES.



We further calculated statistics such as Standard Deviation, Diversification ratio and Concentration ratio to select which portfolio gives us most diversified results.

```
##       GMVw  MDPw  MTDw  ERCw
## SD    2.28 4.362 5.775 3.354
## ES95 1.00 1.000 1.000 1.000
## DR    1.81 2.343 2.257 2.264
## CR    0.29 0.219 0.236 0.175
```

Clearly from results above we can see GMV is the most diversified portfolio with least standard deviation observed compared to other three portfolios considered.

9. Conclusion

In this paper, we have created an algorithm for the user where he can select any sector or stock and create a portfolio which will have most diversified returns. Few of the observations made while working on this paper has been listed below for the user to assess:

- Pandemic did not had an extreme adverse effect on stock market as credit crisis of 2008.
- Even though no analysis was conducted in selecting our stocks, Technology and Finance did perform well during and after the pandemic period as assessed by the $1 investment done in observation period.
- In order to create an efficient portfolio, if the user is more risk-averse than the author, they may have to include a risk-free asset in the portfolio, since for this purpose we have only included and focussed on stocks as an asset in our portfolio.
- Even with a straight forward approach in portfolio creation, we can still create a diversified portfolio with tools shown in the paper, if the purpose of the investor is to achieve higher return with minimal risk

10. Code

RMD File:
https://github.com/sauravmawandia/515-Project/blob/master/Stock%20Analysis.Rmd

Uploaded the RMD file along with the Report

Output Of the Code:
https://htmlpreview.github.io/?https://github.com/sauravmawandia/515-Project/blob/master/Stock-Analysis.html