24369

FM321 – Risk Management and Modelling

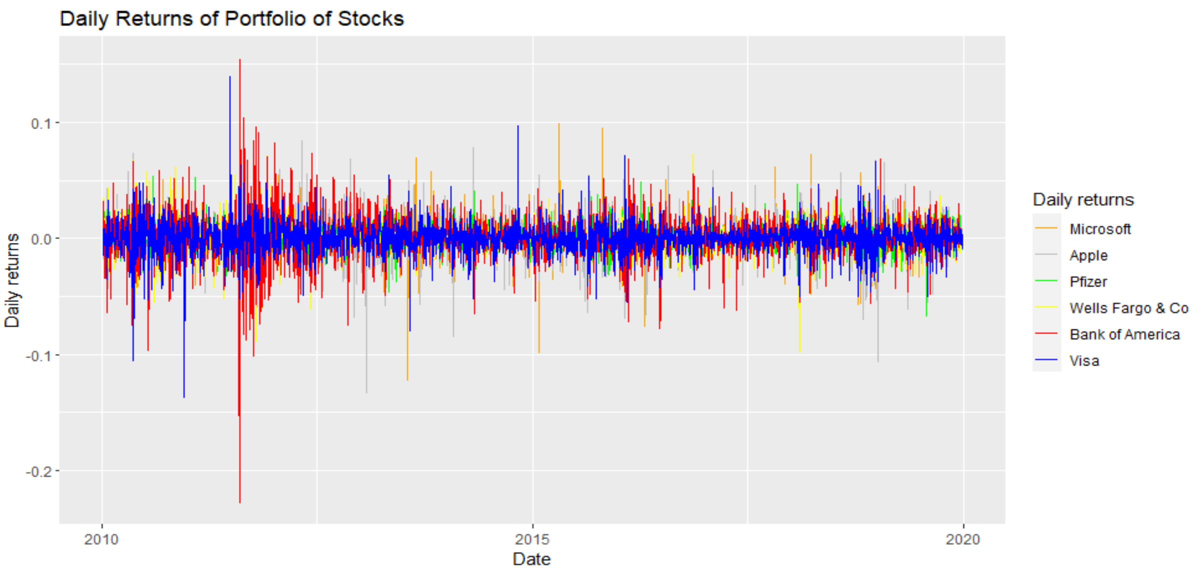
Course Project

Part 1: Importing Stock Data and Portfolio Formation

Starting the analysis in R Studio with the required libraries being downloaded which will allow us to use built-in commands for data cleaning, data importing and statistical modelling purposes, an environment is created to store all our objects. These objects encompass the stock data that will be imported from Yahoo Finance.

Given their commonalities along with their distinctive industries, Microsoft (MSFT), Apple (AAPL), Pfizer (PFE), Wells Fargo & Co (WFC), Bank of America (BAC) and Visa (V) were the stocks chosen for analysis. The *getSymbols()* command under the *quantmode* which is one of the libraries downloaded, allows us to retrieve historical stock data. Given the nature of our analysis which will be built upon forecasting and its performance evaluation, the sampling period plays a significant role as it may include some extreme periods such as a highly volatile financial crises period or non-crisis low market volatility. We particularly want to analyse this as for forecast quality, stationarity in time series is a vital aspect. As well as stationarity, it is indeed important to assess if there exist any structural breaks in the series or not.

Given the availability on Yahoo Finance, including more observations could be optimal as larger sampling will yield a better performance and accuracy. As well as thinking that as the sample size increases, we could possibly observe the Aggregational Gaussianity property of financial returns. Based on these parameters, first a sampling period from 01/01/1995 to 31/12/2019 was chosen. However, due to the financial crisis period between 2007-2009 which causes a high-volatility period, the starting point of sampling period from 01/01/2010 to keep the sampling period as large as possible while giving an extensive care for stationarity.



Plot 1: Daily Returns of Portfolio of Stocks

Then the adjusted stock prices are extracted as the definitions of returns in different sources could naturally vary. This, throughout the analysis, allows us to get an accurate understanding of stocks' performances. Following the extraction of adjusted prices, log-transformation is applied as the continuous returns are theoretically superior to simple returns with their symmetry property. As the last step, the returns are demeaned as they are at very small magnitude at daily frequency. Thus, setting them equal to zero yields a good approximation, by not introducing a huge bias, allowing us to adjust the financial returns for computational purposes.

With our securities and their returns prepared, we form a portfolio of stocks in which all are equally-weighted. We assign equal weights to each of our securities which constitute the portfolio of stocks. Having six stocks in our portfolio, we assign weight of 1/6 to each.

Portfolio Weight	**Tickers**
0.167	MSFT
0.167	AAPL
0.167	PFE
0.167	WFC
0.167	BAC
0.167	V

Table 1: Portfolio Weights

I would also like to note that the return of portfolio calculation with log-returns is not identical as simple returns given their mathematical properties.

Part 2: Statistical Analysis of Financial Returns

Upon observing the data generating mechanism and our stock data, it is essential to observe statistical properties of our data set to obtain the building blocks for estimation and model selection procedure.

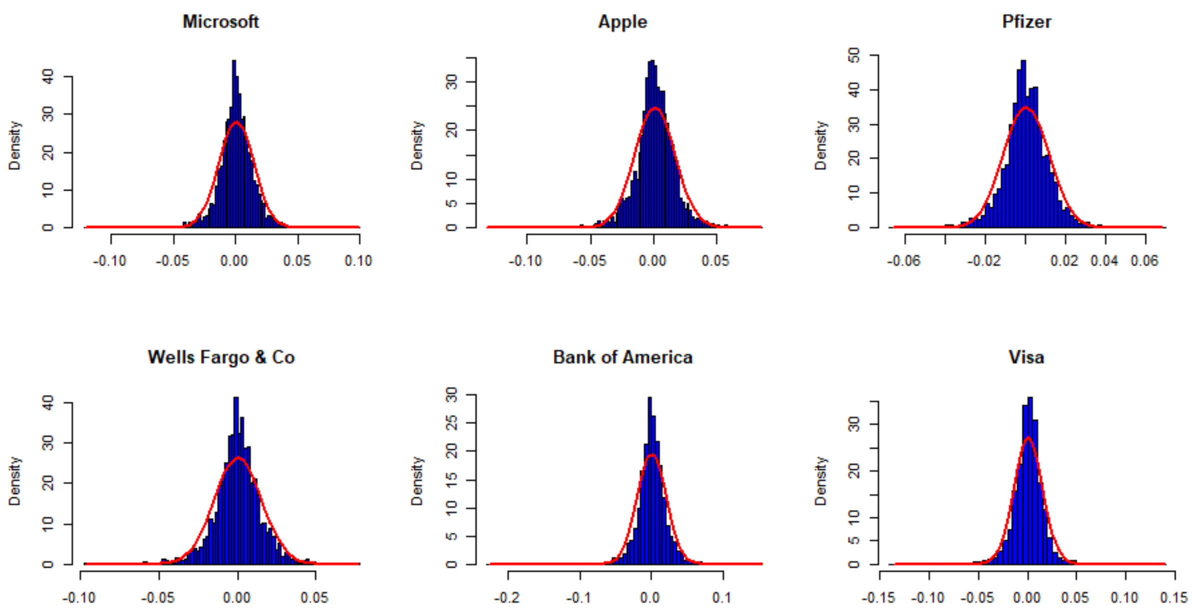
Starting the analysis with descriptive statistics, we observe the mean, standard deviation, maximum, minimum observations along with the skewness and kurtosis coefficients.

Taking *Table 2* as reference, our first observation is the skewness in the distribution of each stock return. Normal distribution has the skewness coefficient of 0 which indicates a symmetrical distribution. Each stock return we have is either negatively or positively skewed. This is also in line with the properties of financial returns. In fact, in majority the financial returns are negatively-skewed given that the investors react more to negative news or downturns than recoveries. Additionally, the Kurtosis coefficients which indicate the fat-tailedness of our distributions are also significantly larger than 3, being the Gaussian Kurtosis coefficient. Overall, we can conclude that the stock returns follow a leptokurtic distribution.

	Microsoft	Apple	Pfizer	WFC	BAC	Visa
Avg.	0.000743	0.000952	0.000433	0.00037199	0.000363	0.00088118
Std.	0.0143	0.0162	0.0115	0.0152	0.0205	0.0147
Max.	0.099	0.0850	0.0683	0.0776	0.1548	0.1397
Min.	-0.121	-0.132	-0.0663	-0.0967	-0.2271	-0.136
Skewness	-0.1084	-0.34089	-0.0219	-0.1717	-0.4746	-0.1443
Kurtosis	6.44016	4.8211	2.838	3.981	9.824	9.203

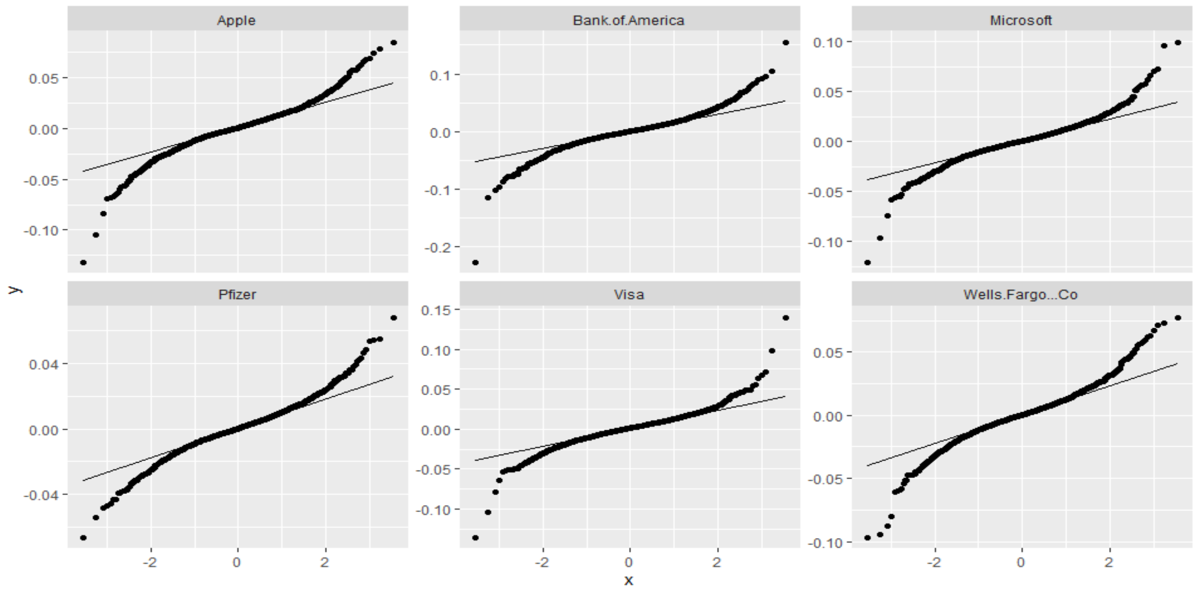
Table 2: Descriptive Statistics

We can also clearly observe the distribution of the stock returns graphically. Imposing the Gaussian distribution allows us to visually understand that the returns do not follow Normal distribution. It also allows us to observe the leptokurtic distribution rather than a mesokurtic.



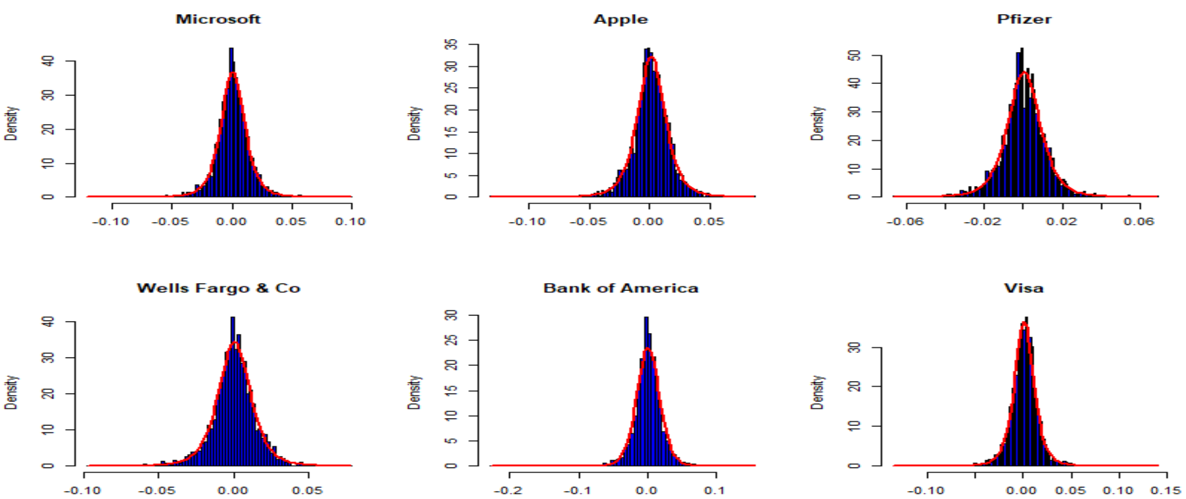
Plot 2: Histogram of Stock Returns with Normal Distribution.

I would like to extend our analysis with the QQ-Plots where the flat lines indicate the standard normal distribution. The sigmoid shape which each return exhibits explain the fat-tailedness of returns.



Plot 3: QQ-Plots of Stock Returns

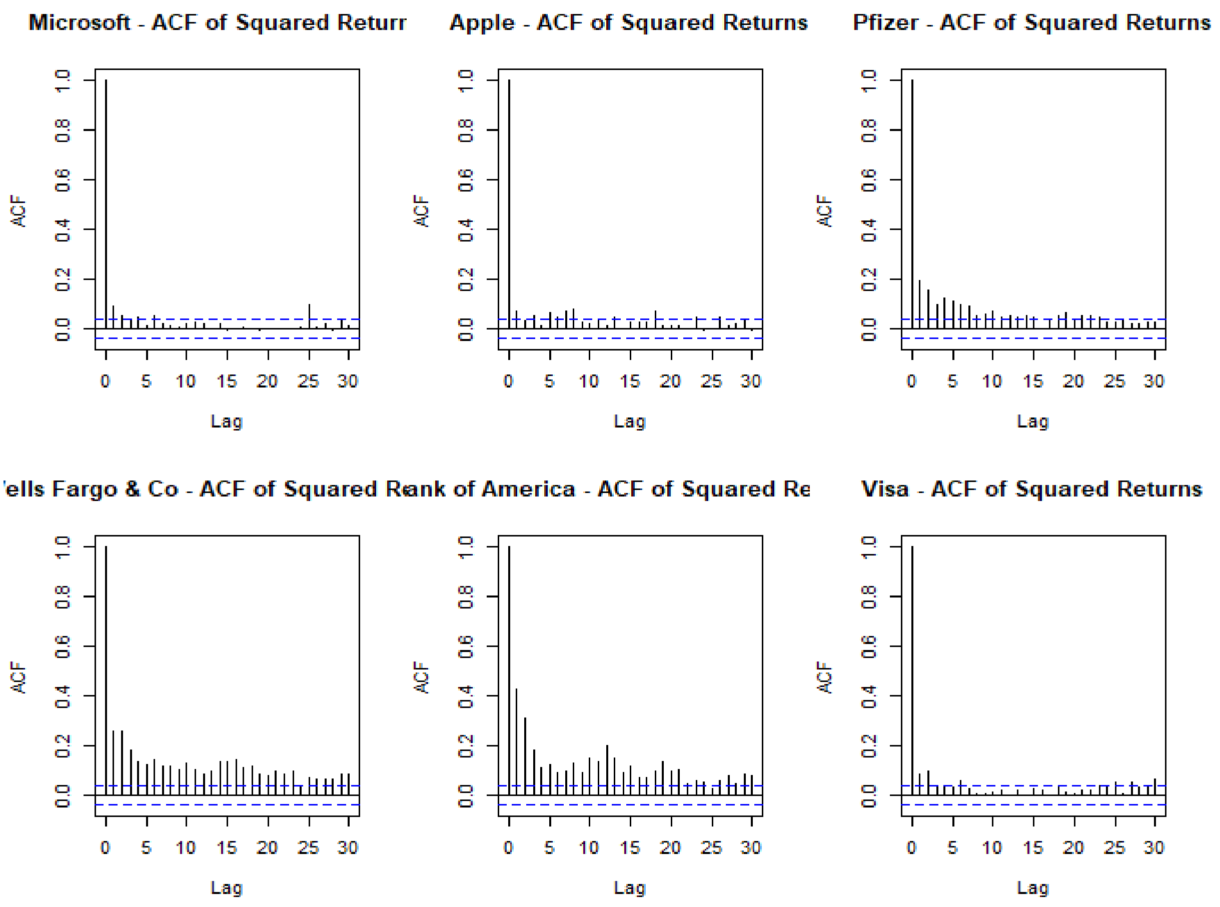
Lastly, to mathematically understand and prove non-normality of stock returns we can implement the Jarque-Bera test. Alternatively, we could use the Kolmogorov-Smirnov test which estimates the minimum distance with a reference distribution. Jarque-Bera Test of normality, given its mathematical specification requires a degrees of freedom of 2 which then enables us to obtain the critical value. Having H0: Normality vs. H1: Non-normality, we obtain a critical value of 5.991465. Based on the test statistics obtained, the null hypothesis is rejected for each iteration as they are higher than the critical value. Additionally, the p-value is lower than 0.05. Hence each stock return follows a non-normal distribution. We could use another distribution that exhibits fat tails as it might be more suitable to work with, fitting the returns better.



Plot 4: Histogram of Stock Returns with t-distribution.

Fitting the t-distribution, we can state that the stock returns show a better fit. Along with the analysis of Jarque-Bera Test of Normality and QQ-Plots, we can understand that the stock returns distribution is better approximated with t-distribution.

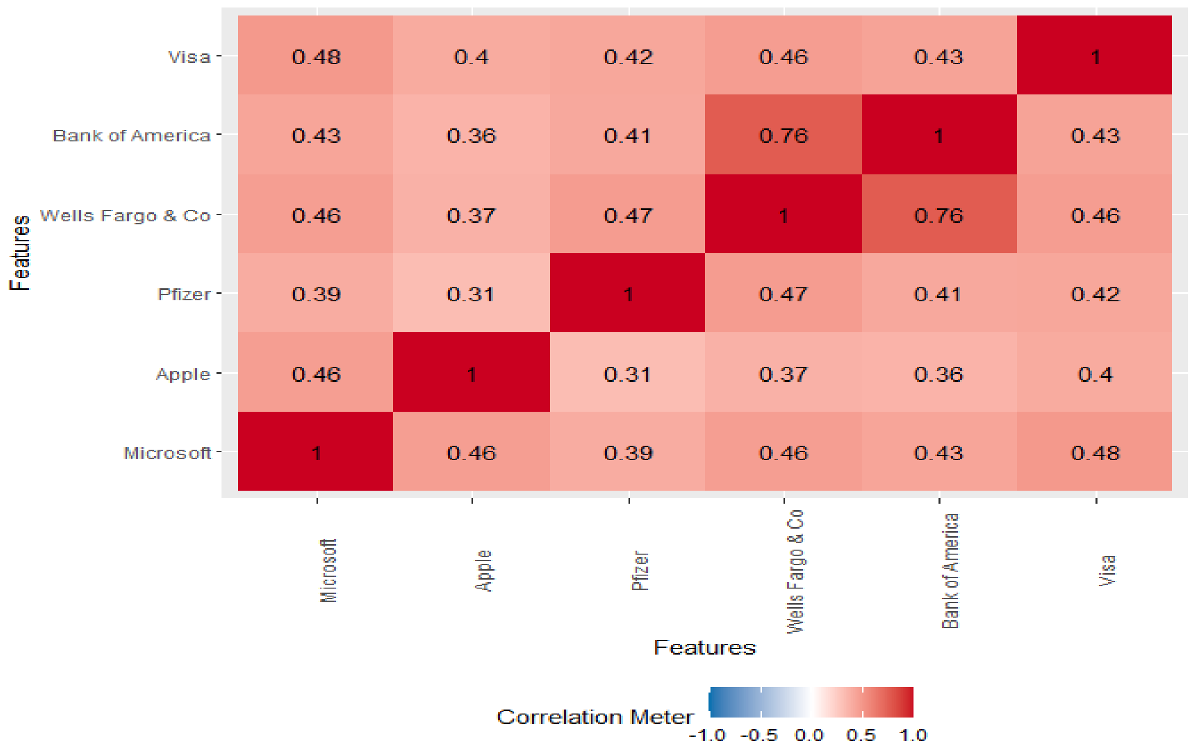
Moving on with the volatility cluster analysis, we compute the Autocorrelation Function of our series. We obtain a 95% confidence interval constructed as $1.96/\sqrt{T}$, where T is the sample size. The lags on the correlogram are determined through the Akaike-Schwarz Information Criterion or as $\sqrt{T}$. The serial autocorrelation analysis is particularly important as it shows the predictability of quantities. This is determined by evaluating the autocorrelation coefficients' significany which is yielded from exceeding the interval. First analysing the daily returns, we observe that the autocorrelation coefficients are not significant in general. Daily returns are always within the 95% boundary which indicates that the daily returns are not predictable from their past values. However, when the daily-squared returns, the autocorrelation coefficients are highly significant over the lags indicating high predictability from the past values. Thus volatility is highly predictable given persistent and significant autocorrelations. I would like to mention that this does not violate the Efficient Market Hypothesis.



Plot 5: Autocorrelation Functions of Squared Returns

For the mathematical analysis of the presence of autocorrelation, we use Ljung-Box portmanteau Test having H0: rho0=rho1=...=0 (White Noise) vs. H1: any of the rho is not equal to 0. We specify the number of lags as 20, obtaining the critical value of 31.41043, we then compare our critical value with test statistic values for each stocks. For each iteration, our test statistics values exceed the critical value, hence residing in the rejection region indicating significant autocorrelation coefficients. Hence, the information is passed from one time point to another.

As the last step, I wanted to produce a correlation matrix between the stocks to observe the unconditional sample correlation between 6 stocks and to fully understand the relationship and cover up the initial Exploratory Data Analysis and statistical analysis phase.



Plot 6: Correlation Matrix of Stock Returns

Part 3: O-GARCH and Variance-Covariance Matrix Estimation

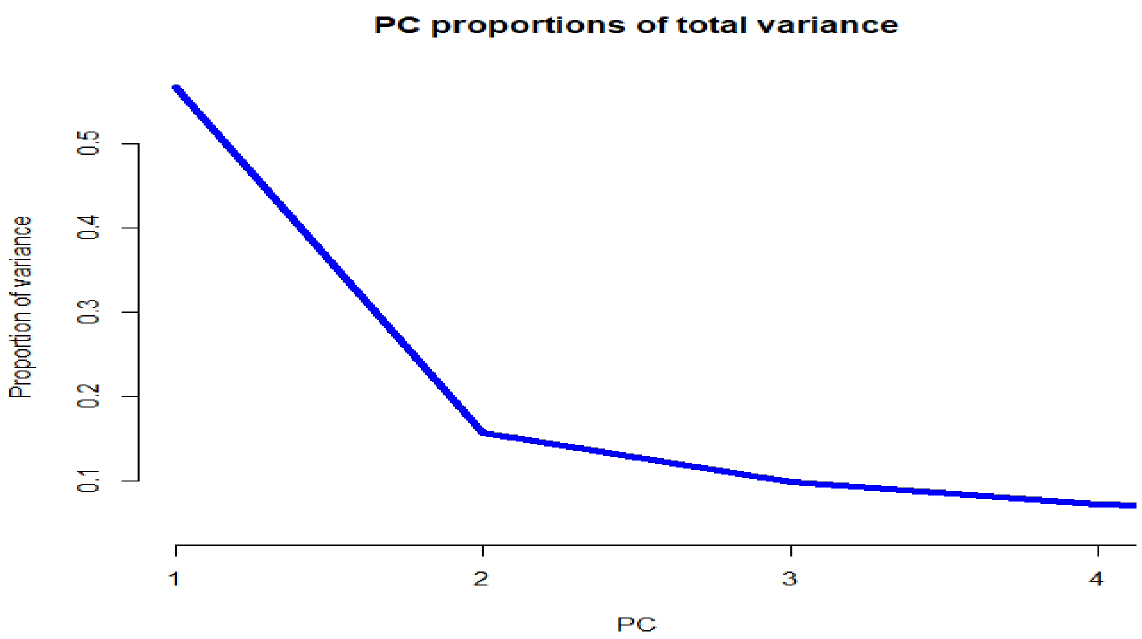
Orthogonal-GARCH implements PCA to estimate the factor structure of the stock returns sequentially and univariate GARCH to model factor variances, ensuring orthogonality. Applying PCA allows us to reduce the dimensionality to order of N, adressing the Curse of Dimensionality problem. We can estimate a GARCH (1,1) for factor variances. Equal weights is needed to forecast conditional variance. Conditional Covariance Matrix is estimated through the multivariate volatility model. We have a large system, and the Orthogonal GARCH is the most feasible option. We aim to replace the initial set of variables with a small number of uncorrelated factors.

We first model the factors with Principal Component Analysis, the fraction of the total variation explained by each principal component can be summarized as the following:

	PC1	PC2	PC3	PC4	PC5	PC6
Std.	0.02884	0.01519	0.01201	0.01031	0.009544	0.008144
Proportion of Variance	0.56556	0.15697	0.09807	0.07233	0.061950	0.045110
Cumulative Proportion	0.56556	0.72254	0.82060	0.89293	0.954890	1.000000

Table 3: Summary of Principle Components

Table 3 gives information regarding the importance of components. The first principal component is the market factor as each factor loading is of the same sign. The critical point is to determine the number of principal components that will be used as that number will affect our variance-covariance matrix formation. We can identify how many principal components to be used through producing a Scree Plot which uses the Elbow Rule. Alternatively, we can also use Bai-Ng Criteria to determine the number of factors that must be used.



Plot 6: Scree Plot

Based on the visual inspection of *Table 6*, we can use the first two factors that explains the variability by 72%. Thus number of factors will be equal to 2.

Modelling the individual asset volatilities for the first 2 factors, we specify a GARCH (1,1) model for each principal component. Using the output from Principal Component Analysis, and GARCH, we construct the variance-covariance matrix for each trading day. Referring back to the *Plot 4* where we have analysed that t-distribution fits better as the distribution of our stock returns, we fit t-GARCH to capture the tail behaviour in a more comprehensive way and increase our model's performance.

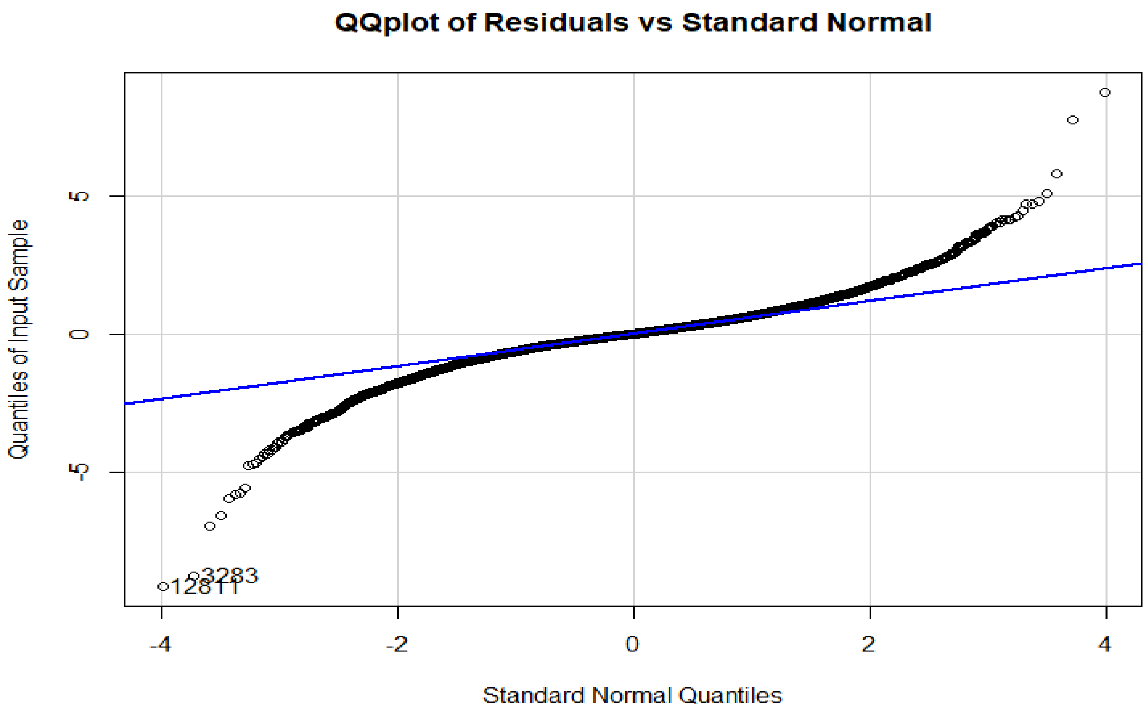
	PC1	**PC2**
Ω	0.00002	0.00002
α	0.12602	0.09186
β	0.84794	0.83939
shape	5.76896	5.34957

Table 4: GARCH Model Fit

GARCH, incorporating all the lags as opposed to ARCH, produces stable volatility estimates. Based on α and β parameters, given that their sum is close to 1, the model's predictability dies out slowly. As β, memory, persistency parameter, is larger than the α, news parameter, we could also state that the volatility majorly stems from the past conditional volatility. Overall, the covariance stationarity condition is fulfilled. This process is covariance stationary and ergodic given that $(\alpha + \beta) < 1$.

To ensure that our selection of model which is Non-Normal GARCH performs better than GARCH, we can compare the two with Likelihood Ratio Test, having GARCH nested in t-GARCH. Multiple likelihood is produced for each factor embedded into the model. Having obtained the values as 124.5174 and 179.9257 assigned to *LR_GARCH* with the critical value of 3.841459, we obtain test statistic values that are significantly higher than the critical value, we can clearly state that the H0 is rejected. Thus the fit definitely improves which indicates that t-GARCH is a better specification under these stock returns. Although t-GARCH requires more data feed, imposing normality restrictions yield a significant loss in fit. Asymptotically, Wald and Lagrange Multiplier tests would also yield the same result.

I would like to extend the analysis with the Residual Analysis of GARCH (1,1). Diagnostic Test allows us to test the distribution of residuals with our assumptions. Additionally, we can test serial correlation of residuals.



Plot 7: QQ-Plot of GARCH Residuals

Residuals (ε) equalling to returns over σ, possess fat tails which is in line with our Student's t-distribution specification. This is visually inspected through *Plot 7* and approved through Jarque-Bera test where H0 is rejected. Additionally, conducting the Ljung-Box test yields that our residuals exhibit serial autocorrelation which is an indication that there exists predictability from some lag values.

Then forming the conditional variance-covariance matrix, *ht*, is constructed with O-GARCH. It is positive-semidefinite. The loop iterates through each observation to form the time-varying conditional covariance matrix. Portfolio's conditional volatility is estimated by using the variance-covariance matrix and portfolio weights. The last observation in our time series is taken from 2019-12-30. Taking 2019-12-30 as the date t, we should then forecast *ht* for the next 1 day, yielding the date t+1. The portfolio weights are held constant but due to the time-varying nature, *ht* will vary for each t. The multiplication of portfolio weights with the conditional variance-covariance matrix yields the portfolio volatility which requires the quadratic optimization solution. We ultimately obtain a conditional volatility estimate of 0.004839598, representing the expected volatility of the portfolio's return on 2 January 2020, and corresponding to a portfolio variance estimate of 2.341924e-05.

Comparing our portfolio variance with the variance-covariance matrix of 2 January 2020, particularly with the variance terms that are residing in the diagonal, the portfolio variance is relatively higher than the individual variances. We might be in such a risky environment by nature where diversification is not optimal given the existence of tail events.

Part 4: Value-at-Risk Computation for 2 January 2020

At this stage, to define our parametric Value-at-Risk estimate, we need three elements:

- **Significance level:** p=0.05
- **Conditional Volatility Model:** t-GARCH
- **Distribution:** Student's t-distribution

Following our analysis in Part 2, as the model is specified for the distribution of returns, we use t-distribution in our parametric VaR estimation.

Choosing the significance level, the volatility model and the type of distribution in the volatility model, we set the equation for our parametric VaR estimate. The same distribution imposed needs to be used to obtain the critical values. Then, for each subsample we will be forecasting the VaR for t+1. The VaR estimation follows an iterative approach, for each point in time, we will observe how our portfolio behaves, whether VaR breach has occurred or not.

The degrees of freedom of our distribution is derived from the shape parameter of t-GARCH. In this case, due to the nature of our 2 factor-implemented model, we obtain 2 shape values. We can assign a single degrees of freedom value by taking the weighted-average of our two principal components, based on their power in the explanation of variance proportion.

Then we calculate the estimated portfolio loss, *VaR_0.05*, at 5% significance level, as well as applying negative sign to it as VaR represents the loss by nature. We calculate the 5% percentile of t-distribution with our specified degrees of freedom and then multiply it with the conditional portfolio volatility for 2 January 2020.

We again obtain two different VaR estimates which are approximately 0.0077 and 0.00761. This is where principal component weighted-averages will become useful as we form the *portfolio_VaR* by multiplying our VaR values with respective weights. Overall, yielding us a parametric VaR estimation of 0.007680447 for 2 January 2020. Hence, with 95% probability the daily return is expected to be above -0.768%.

Part 5: Backtesting

Having forecasted VaR (0.05), to evaluate the forecast performance we now apply backtesting. The performance will be tested through the understanding of whether the forecast is consistent with the realised actual observations. The backtesting procedure should capture the advantages of the parametric approach versus the historical simulation as theoretically, the parametric method incorporates the advantages of using a well-deployed volatility model, although we put more assumptions.

The estimation window length is mainly determined by the choice of our VaR model and probability level. The size of the testing window needs to increase with the extremity of VaR levels. Particularly, as we implemented GARCH, more observations are required. In this regard, having assigned p=0.05, 20 yields a sample size of 400 to compute VaR(p). We would like to observe 20 observations in the tail. Additionally, as there exists no structural breaks in the data, we could state that the VaR forecast will perform well, as well as capturing the non-linear dependence.

To obtain a hit sequence, VaR is estimated for each respective trading day based on rolling window analysis by selecting sample. We estimate our conditional volatilities and obtain VaR for each time period. Thus, different VaR measures are then obtained for each trading day as the rolling window iterates over time, showing the dynamic, time-varying nature of VaR forecasts. Through here, we obtain our violation ratio which is essentially the ratio of empirical probabilities. It represents the proportion of times that VaR has been breached compared to our significance level. We say that losses will exceed VaR with the probability of p. VR = 1 indicates that our forecast is well performing, so that breaches occur with expected intervals. On the other hand, VR>1 indicates the underestimation of VaR meaning that we are breaching more than we should, and VR<1 specify overestimation of VaR. Obtaining a Violation Ratio of 5.99, we seem to underestimate VaR.

In the case that VR>1 is statistically significant, the computational method for our risk measure, VaR could be alternated. The significancy of this will be tested through the Unconditional Coverage Test (Bernoulli Test) where we construct two different likelihoods from Bernoulli PDF. Which will allow us to test whether VR=1 or not, i.e, p=0.05 or not.

The Bernoulli coverage test is nonparametric as it does not assume a distribution for the returns and provides good benchmarks for the assessment of accuracy of our VaR model. The test states that if the VaR forecasts are correct, a breach should happen with the probability p.

The reason why Bernoulli distribution is used is as it encompasses the binary behaviour that is assigned to breach/non breach, 1/0 respectively. Thus it will yield a boolean value, T: if there is a breach and F: No Breach. And we assign 1 and 0 respectively yielding a Bernoulli variable. Having the Bernoulli PDF, we will obtain the likelihood functions which will be used for Likelihood Ratio Test. We essentially use the out-of-sample data, *(WE+1: Trading Days)*, for further analysis, such as likelihood ratio tests and conditional coverage tests.

The null hypothesis is that the model is correctly specified (VR = 1). With a *picap* value that is approximately 0.05, we form our test statistic from the likelihoods of restricted and unrestricted models and ultimately fail to reject the H0. Failing to reject the H0 shows that VR=1 so that our empirical probability is equal to 0.05. We can assess that the model is well-specified and the observed violations are consistent with the expected behaviour.

Extending the analysis with the Conditional Coverage Test (Independence Test) which states that if VaR forecasts are correct, breach is independent of whether a breach has occurred the day before or not. We determine whether a violation in one period yields a higher likelihood of a violation in the next period. Thus, the corollary of it is that if the risk measures are correct, then probability of breach should be independent of time. The test has the following hypothesis,

H0: P01=P11

H1: P01 is not equal to P11

Where; P01: The probability of observing a breach given that the previous period we have not observed a breach,

P11: The probability of observing a breach given that the previous period we had a breach.

If our null hypothesis holds, the VaR measures are independent across two subsequent periods.

With the R script, the possibility of each 4 outcomes occurring was observed, which resides in our *logical* matrix. Upon the calculation of each respective probability, *hat_p* captures the overall probability of observing a breach. Then obtaining our unrestricted and restricted models, forming test statistic, comparing it to chi-squared critical value, we ultimately fail to reject H0.

Hence, there exists no significant evidence of dependence between VaR forecasts across two subsequent periods, suggesting that our model might actually be capturing enough information.