Project Plan and Literature Review

Benjamin Turrington

November 2023

Abstract

We propose to investigate the internal structure of different stock portfolios. We are going to go about this on 2 fronts, the first is to analyse stocks within a portfolio and how correlated they are to one another. Next was to use the GARCH process to first analyse the stocks and then attempt to simulate them. All in the aid of investigating varying stock portfolios

1 Introduction

My project is based on GARCH pricing dynamics, there are 2 proposed stages to the project, the first is analysing historical GARCH data, and the second is simulating GARCH pricing for the future. GARCH means Generalized autoregressive conditional heteroskedasticity, a version of which, ARCH, was first theorised in 1982 by Robert F. Engle.[1] It is challenging to simulate GARCH pricing as the amount of volatility is unknown. I also aim to write code that will allow me to compare different stocks and to test if said stocks are correlated in any way.

2 Literature review

Before I began the project I completed a large amount of background research. The first place I started, was recommended by my project supervisor [2]. This book goes through the ARCH and GARCH processes and the taxonomy of stocks. The book gives a very good overview of all econophysics, however, to get a deeper understanding of any specific area other articles are better. GARCH process is all about volatility and a paper that goes through this for S&P 500 was written by Cizeau et al [3]
Firstly I looked into the autoregressive conditional heteroskedasticity theory that was first postulated in an article written by Robert F. Engle in 1982 [1], the article however the article was written specifically about inflation in the United Kingdom, so therefore there are many improvements to be made, some of which are laid out in his later piece from 1993[4]. The equation set out in the

article being:
$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + ... + \alpha_p x_{t-p}^2 \tag{1}$$
Where $\alpha_0, \alpha_1, ..., \alpha_p$ are positive variables and x_t is a random variable with 0 mean and variance σ_t^2. This equation also characterises x_t as a conditional pdf, often given by a Gaussian pdf. The Gaussian pdf was then investigated in 1992 by Baillie and Bollerslev[5].
In an article written in 1986 by Bollerslev [6] the GARCH process was put forward this was done as the ARCH process causes many problems as the value of p has to be very large to be able to describe the evolution over time. This is why the generalised method was introduced. The GARCH modeling and analysis put forward by Bollerslev used the equation:

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + ... + \alpha_q x_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2 \tag{2}$$

This process is called GARCH(p, q) where $\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_q$ are control parameters, with x_t being the same as it was for the ARCH model.
Since the publication of the method and equations of the GARCH process there have been many subsequent articles and reports on different areas of finance for example FX (foreign exchange) rates, returns on stocks, interest rates, and inflation. For example a report by Vedat Akgiray[7] about the "Time Series of Stock Returns" by analysing the "Center for Research in Security Prices (CRSP), and using daily data. This allowed the authors to work out the memory of the stock, in trading days. There are some potential issues with the GARCH process, one of which was analysed by Drost and Nijman, they looked into the different time horizons of the GARCH process specifically for finite variance in GARCH(1,1) process, however, the paper went on to note that the statistics still held true so that "temporal aggregation of a GARCH(1,1) is still a GARCH(1,1) processes."[8]. The GARCH process is still used today and is used to assess the effect of events, for example, a report published in 2023 about the effects of the Ukraine-Russia war on far-reaching places like Egypt.[9]. However, the methods used today use GARCH as the basis and have built the statistical models on that, for example in 2018 trying to predict the volatility of Bitcoin prices[10]. Another event that has been anlysed using the GARCH process was the Covid pandemic for example the one done by Prempeh et al in 2022[11].

I also learned about understanding the correlation between stocks and the taxonomy of stocks. The idea that 2 separate stocks can be correlated in some way has been around for years, as early as 1959[12]. An article written in 1997 by Kadtke and Bulsara [13] completed a review of different stocks using these equations:

$$S_i = lnY_i(t) - lnY_i(t-1) \tag{3}$$

Where $Y_i(t)$ is the price of a stock on day t. This is known as the log return.

$$\rho_{ij} = \frac{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{\langle S_i^2 - \langle S_i \rangle^2 \rangle \langle S_j^2 - \langle S_j \rangle^2 \rangle}} \tag{4}$$

This report analysed 30 stocks in the Dow-Jones and all of the S&P 500. However, this would take a very long time to run the code as this was nearly 26 years ago so the authors could not do much with the data without greater computing power. There have been many other reports and articles that assess the factors that affect stock prices, they often use the distance between stocks to help their analysis, for example in 1993 with Connor and Korajczyk[14]. Some papers also go through the difference between market data and the expected values from various predictive methods, for example, a paper by Galluccio and Bouchaud written in 1998.[15] One report released in 1990 by Lo and Machinlay[16] revealed, using the correlation ρ_{ij} values and d_{ij}, that increase in stock prices from larger companies often resulted in increases from those of smaller companies. Showing that there is a large area of possible analysis from these statistical equations and ideas. The equation for the distance between 2 stocks.

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{5}$$

In 1999 Mantegna [17]wrote a paper that analysed the distance between different stocks and used Minimal spanning Trees (MST) and indexed hierarchical trees (IHT). Both of these different graph types were discussed in a book about displaying data by West published in 2001[18]. The MST's are able to be built using a method called the "Kruskal algorithm" outlined in many papers, one of them being written in 2019 by Naing and Aye [19]. This allows the user to see which stocks are correlated, or in the case of FX to see which currency appears to drive the other currencies. One of the methods of creating IHT's is to create an ultrametric space and ultrametric distance matrix, a method of doing this was put forward in 1986 by Rammal et at. [20].

3 Case for support and work plan

Based on the Gantt chart I laid out my work schedule over the coming months, it clearly shows what work I am going to be doing and have done up until this point. The work I have done so far involves code writing a lot of which has been around getting a greater understanding for the topic and evaluating the correlation between different stocks.
I wrote code that creates a matrix of distances between stocks, the GARCH process also corresponds to FX data, as there is volatility and the amount of such volatility is unknown. I therefore ran the code for various currencies with respect to the pound, over that last year. As can be seen in the table, the USD is closely linked to the CAD and the EUR. The Japanese Yen has a large distance between itself and all other currencies.

Dij	USD	EUR	AUD	CAD	JPY
USD	0	0.964	1.335	0.742	1.157
EUR	0.964	0	1.178	1.006	1.195
AUD	1.335	1.178	0	1.025	1.235
CAD	0.742	1.006	1.025	0	1.204
JPY	1.157	1.195	1.235	1.204	0

For example, I investigated 26 stocks from large companies from 2019-2022, where ρ_{ij} goes from 1 to -1, with 1 being completely correlated and -1 being anti-correlated

year	ρ_{ij} max	ρ_{ij} min	ρ_{ij} ave
2022	0.890	-0.043	0.407
2021	0.906	-0.164	0.224
2020	0.935	-0.105	0.422
2019	0.864	-0.075	0.278

This is a recreation of something similar from[13, 2]. I code I wrote was designed to recreate this so I could test it with updated data on whichever companies I wanted.

year	ρ_{ij} min	ρ_{ij} max
1990	0.02	0.73
1991	-0.01	0.63
1992	-0.10	0.63
1993	-0.16	0.63
1997	-0.06	0.51

I then went about plotting the probability plots these can be seen in figure 1 and figure 2. Figure 3 shows the outcome when the probability of each ρ_{ij} value is displayed. Figure 1 depicts the data for 26 countries for the years 2019 to 2022, clearly, there is some form of normal distribution. Figure 2 displays each of the years on top of one another, the same 26 companies were used.

I am now going on to work on the next steps of my project which is to write code and understand the links between stocks better. Using MST's and IHT's I am going to be able to see which stocks are more closely linked to which, after this, I will have a greater understanding of the field. Using the MST's I will be able to map out which stocks are reliant on which, it will also enable anyone who reads the report to be able to create MST's and LHT's of their own stocks or proposed portfolios. Figure 4. depicts the MST and an IHT, the MST shows the connection between the stocks. The MST's are created using Kruskal's algorithm. The IHT's can be build by creating a matrix of distances in an ultrametric space. For systems with more stocks there are higher numbers of links between the stocks, this can lead to MST's such as the one depicted in Figure 5. where one stock closely links to many, for example, GE is connected to 12 different stocks. With many stocks however Figure 6. shows that it can become confusing although it shows the same information as Figure 5. just it has values assigned to the points.

After I gain a greater understanding of the field I will be able to move onto coding elements of the GARCH and ARCH processes. There are 2 stages I would like to achieve in this project the first is to use the GARCH processes to analyse data on stock prices, once I have code that uses the statistical GARCH process I can use it to analysis the effects of certain events. The second stage is to use the GARCH process to attempt to simulate the volatility of certain stocks.

We believe this new structure may have interesting properties such as a new way of investigating differing stock portfolios which could have applications in the field of economics and could give a better understanding of how to diversify a portfolio depending on the user's wants and needs, for example, if they want "high risk, high reward."

After I use the GARCH process to simulate and analyse the stocks, I hope to use this information to then be tested for correlation between the simulated stocks, so that if the user wants to have stocks that would appear to be correlated in the same portfolio or stocks that aren't correlated to lead to diversification. Investigating a portfolio in this way is going to be a multi-step process.

We aim to be able to use the GARCH process on portfolios of up to 6 stocks, as directed by my supervisor due to the very large computing power expected to be needed. I will simulate the GARCH process on multiple stocks then compare the stocks and attempt to change the coupling parameters which will enable us to assess the stocks over a large selection of variables leading to a deeper understanding of the stocks.

I will use the historical data after the investigation of each stock using the GARCH processes, this will allow me to refine the parameters of the statistical equations and therefore better simulate each of the stocks to attempt to predict and simulate the correlation. I hope to be able to analyse portfolios of up to 6 stocks, due to the computing power that is expected to be required from running the code, if more stocks were analysed this would cause the run time to be too high. Although I hope that my code is able to support more stocks so with access to a computer with more computing power it could be run.

By the end of the project, I plan to write a report going over my findings using my code. The aim is for the report to allow anyone who reads it to be able to form their own judgment on stocks that would best suit them. Furthermore, they should be able to use the report to better understand the relationships between different stocks and how they link together.

The research that I collate can also be used to determine the impact of various events, be they political or environmental, for example, assessing the correlation between different pharmaceutical companies during the COVID-19 pandemic. The report and processes could also be used to investigate the effects of the Ukraine-Russia war on the stock price of oil and gas companies, or any such future events.

The critical path analysis displayed in Figure 8 shows the stages that I believe the project will go through. Each stage has 3 numbers assigned, the max, min and Exp, in that order. The max is the maximum amount a task should take, the min is the minimum amount of time an event should take if everything goes right first time, and the Exp is the amount of time I expect a task to take. For the small tasks that I believe they will not take longer than a week even for the max time if everything goes wrong. The max time for the project to be completed is 32 weeks, with the minimum being 21, with the expected time being 24 weeks. The critical time path was then used to create the Gantt chart.

The Gantt chart lays out all of the stages of the project, the Gantt chart in Figure 7. The Green involves work that involves coding and testing the code, the blue represents the writing of the various reports, and the purple represents both the Christmas and Easter holidays. The Gantt chart breaks down all of the non-GARCH-related sections of the project as we know how that part of the project is going to progress, however, the GARCH investigation and simulation section of the project is currently unknown. The GARCH aspect of the project as of now cannot be broken down into smaller pieces as that information is not yet known, however, the timings were created with my project supervisor who is experienced in this field.

References

[1] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

[2] Rosario N Mantegna. *An introduction to econophysics : correlations and complexity in finance.* Cambridge University Press, 2007.

[3] Pierre Cizeau, Yanhui Liu, Martin Meyer, C.-K. Peng, and H. Eugene Stanley. Volatility distribution in the sp500 stock index. *Physica A: Statistical Mechanics and its Applications*, 245(3):441–445, 1997.

[4] Robert F Engle. Statistical models for financial volatility. *Financial Analysts Journal*, 49(1):72–78, 1993.

[5] Richard T Baillie and Tim Bollerslev. Conditional forecast densities from dynamic models with garch innovations. *J. Econometrics*, 52:91–113, 1992.

[6] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[7] Vedat Akgiray. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *Journal of business*, pages 55–80, 1989.

[8] Feike C Drost and Theo E Nijman. Temporal aggregation of garch processes. *Econometrica: Journal of the Econometric Society*, pages 909–927, 1993.

[9] Hisham Abdeltawab Mahran. The impact of the russia–ukraine conflict (2022) on volatility connectedness between the egyptian stock market sectors: evidence from the dcc-garch-connectedness approach. *The Journal of Risk Finance*, 24(1):105–121, 2022.

[10] Raffaele Mattera, Massimiliano Giacalone, et al. Alternative distribution based garch models for bitcoin volatility estimation. *The Empirical Economics Letters*, 17(11):1283–1288, 2018.

[11] Kwadwo Boateng Prempeh, Joseph Magnus Frimpong, and Newman Amaning. Determining the return volatility of the ghana stock exchange before and during the covid-19 pandemic using the exponential garch model. *SN Business & Economics*, 3(1):21, 2022.

[12] H. Markowitz. Portfolio selection: Efficient diversification of investment. *J. Wiley*, 1959.

[13] Rosario Nunzio Mantegna. Degree of correlation inside a financial market. In *AIP Conference Proceedings*, volume 411, pages 197–202. American Institute of Physics, 1997.

[14] Gregory Connor and Robert A Korajczyk. A test for the number of factors in an approximate factor model. *the Journal of Finance*, 48(4):1263–1291, 1993.

[15] Stefano Galluccio, Jean-Philippe Bouchaud, and Marc Potters. Rational decisions, random matrices and spin glasses. *Physica A: Statistical Mechanics and its Applications*, 259(3):449–456, 1998.

[16] A.W.Lo and A.C.Mackinlay. Awhen are contrarian profits due to stock market overreaction? *Rev. Financial Studies*, 3:175–205, 1990.

[17] Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11:193–197, 1999.

[18] Douglas Brent West et al. *Introduction to graph theory.* Prentice hall, Inglewood Cliffs NJ, 2001.

[19] Aye Aye Naing and Soe Moe Aye. A study for kruskal's mst algorithm based on design and analysis of computer algorithms courses. *University Journal of Research and Innovation*, page 39, 2019.

[20] Rammal Rammal, Gérard Toulouse, and Miguel Angel Virasoro. Ultrametricity for physicists. *Reviews of Modern Physics*, 58(3):765, 1986.

[21] J B Kadtke and A Bulsara. Applied nonlinear dynamics and stochastic systems near the millenium. proceedings. 12 1997.

4 Appendices

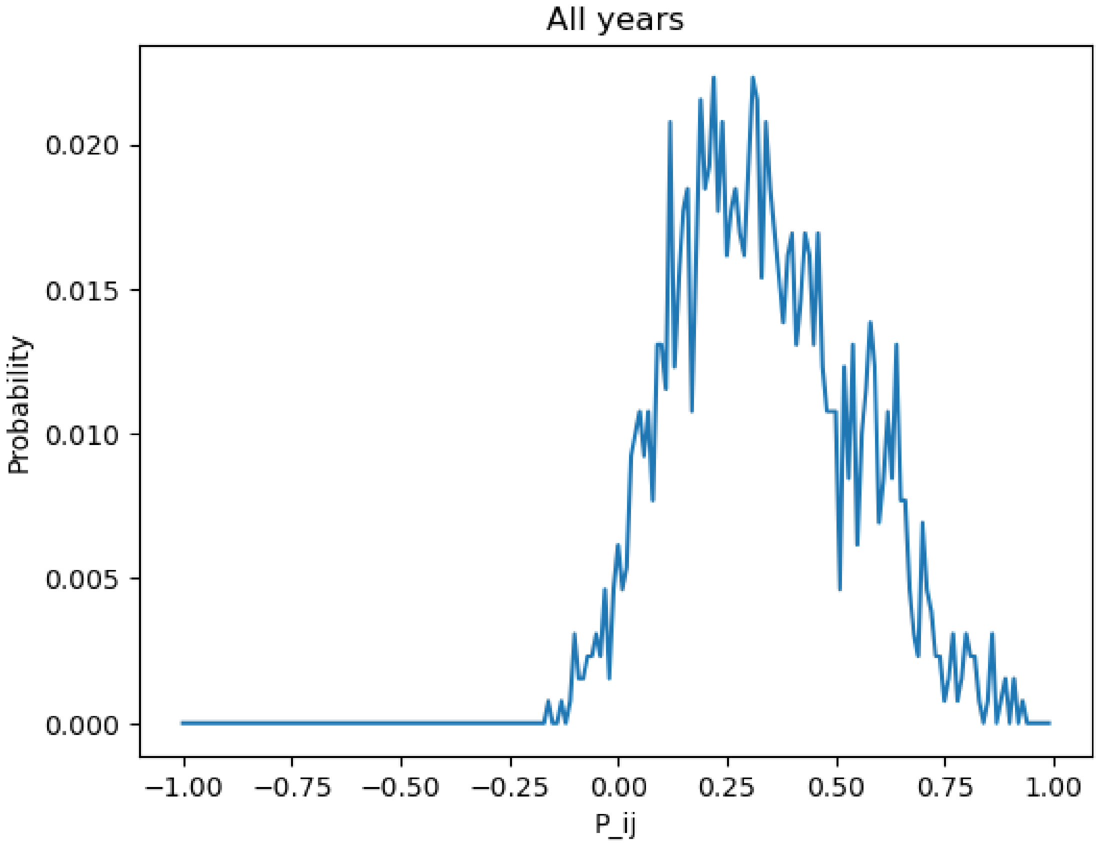

Figure 1: Probability plot of the 26 companies over the years 2019-2022

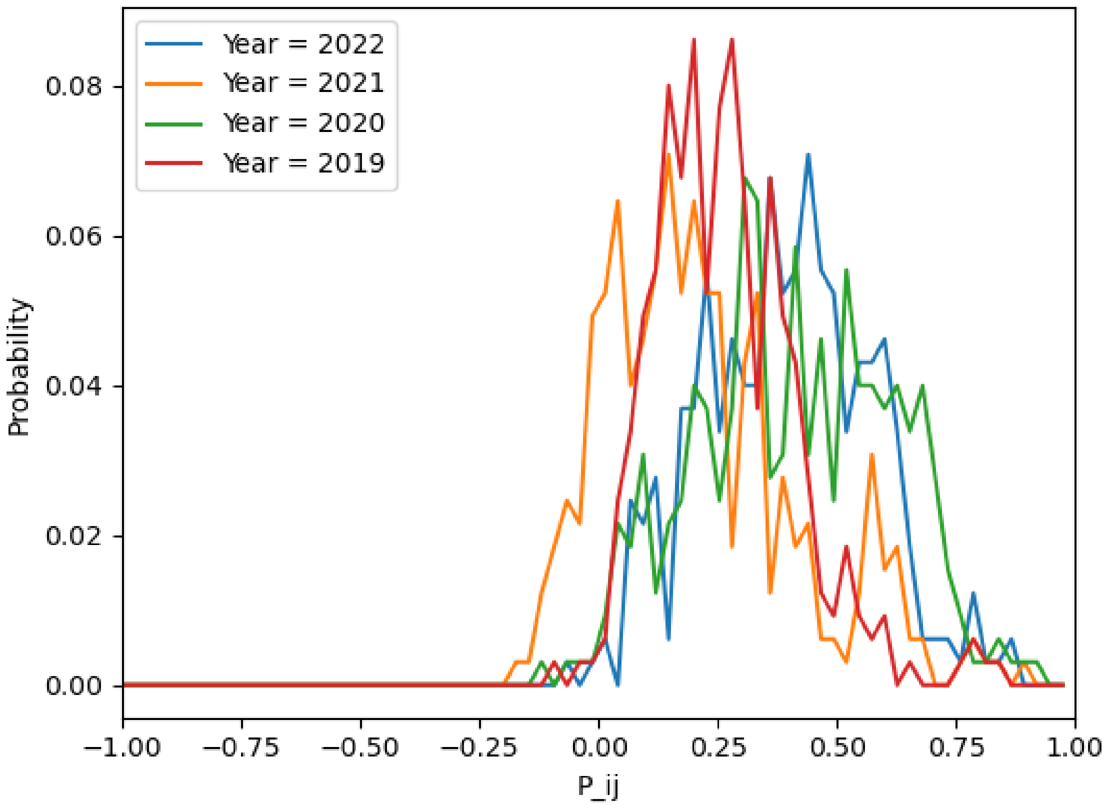

Figure 2: Probability plot of the 26 companies over the years 2019-2022

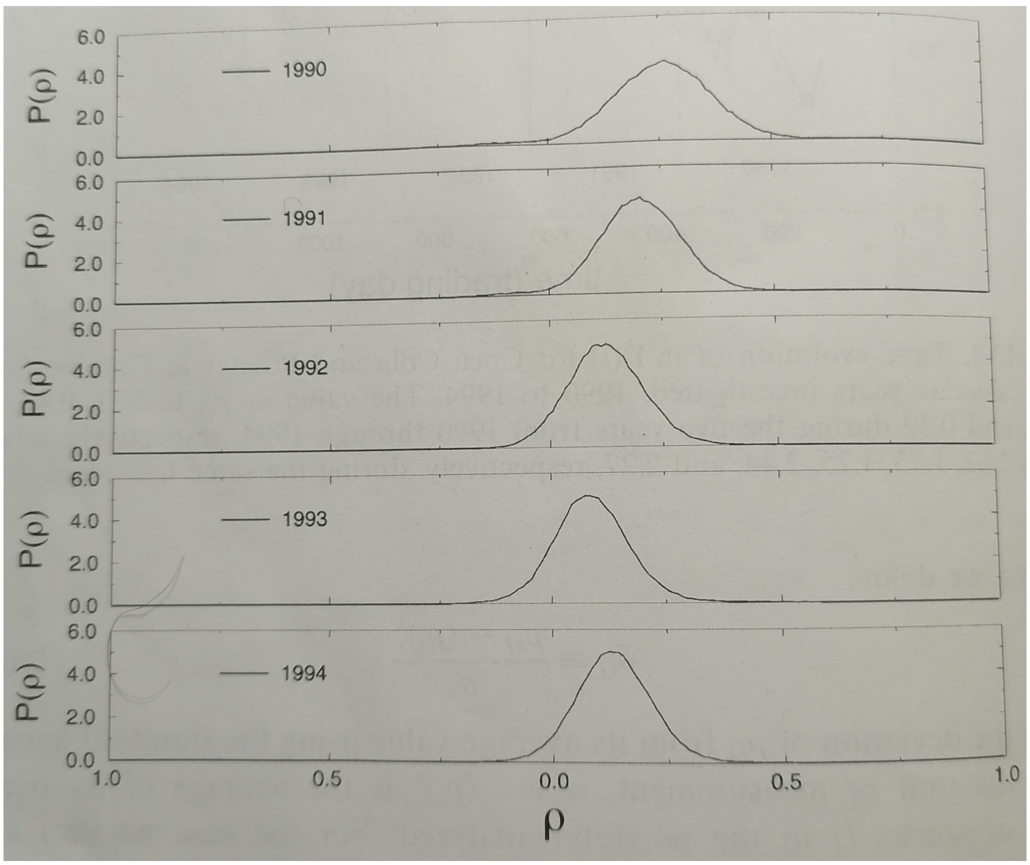

Figure 3: Here is a graph of probability of the ρ_{ij} from the years 1990 to 1994 for S&P 500 [13, 2]

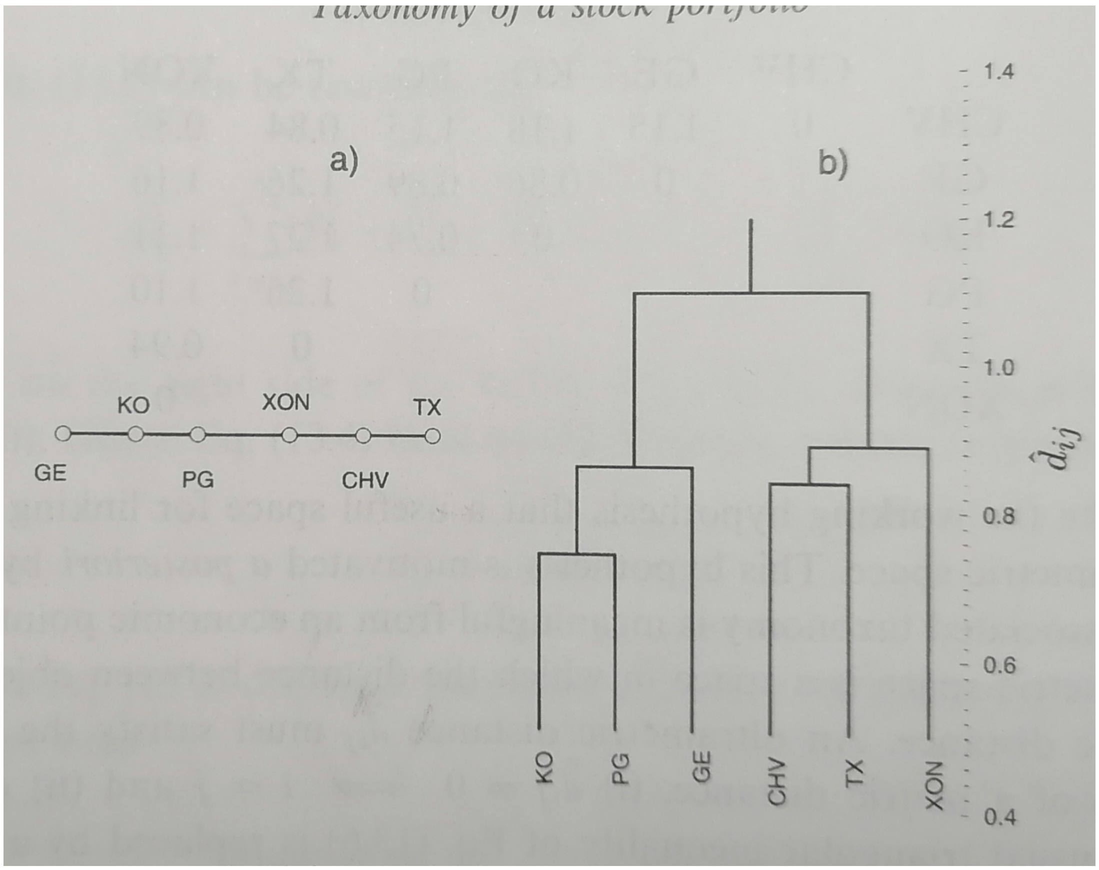

Figure 4: This is an MST (minimum spanning tree) and IHT (indexed hierarchical tree) [2]

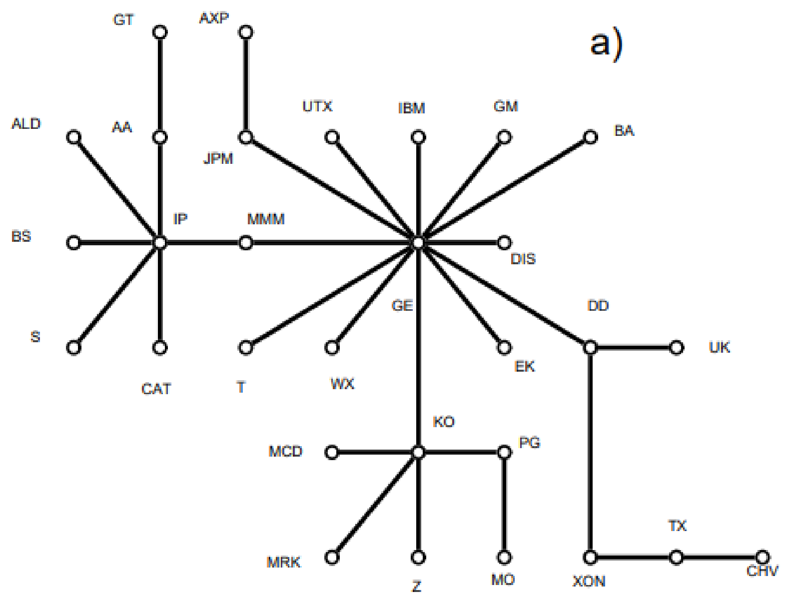

Figure 5: This is an MST (minimum spanning tree) for a large range of companies [2, 17]

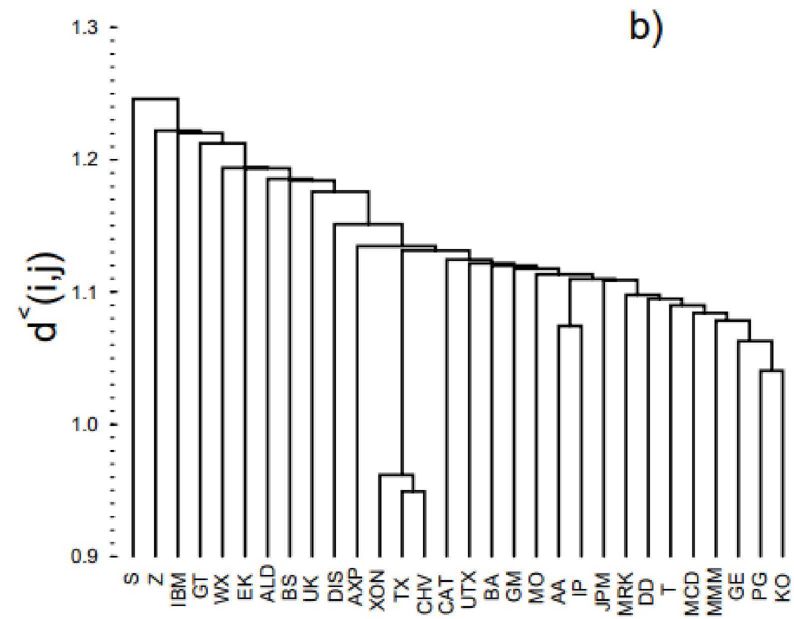

Figure 6: This is an IHT (indexed hierarchical tree) for a large range of companies [2, 17]

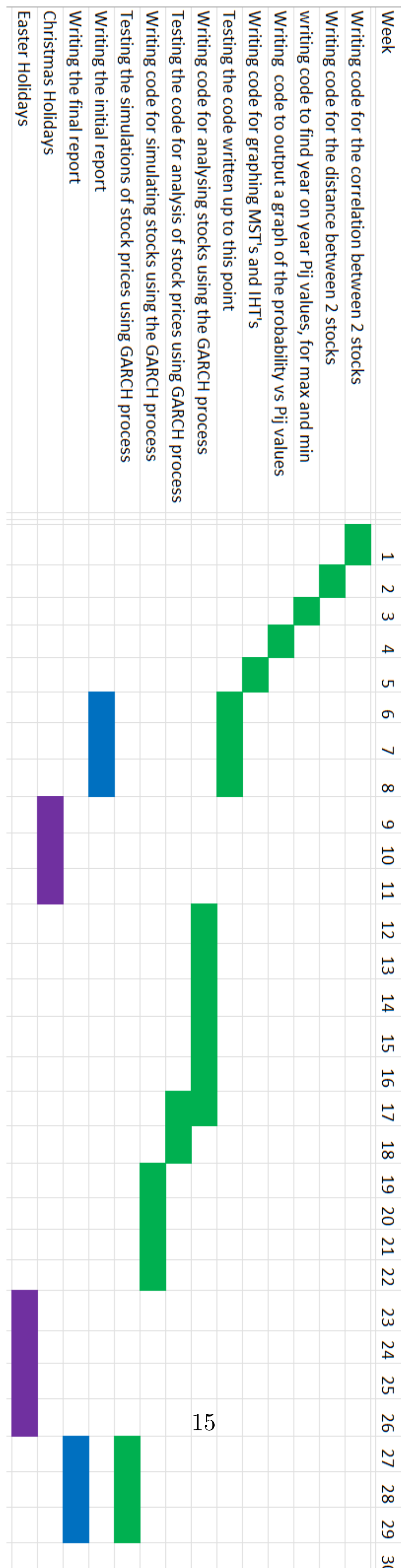

Figure 7: This is the Gantt chart that I created for this project. The first week is the week beginning 23/10/2023

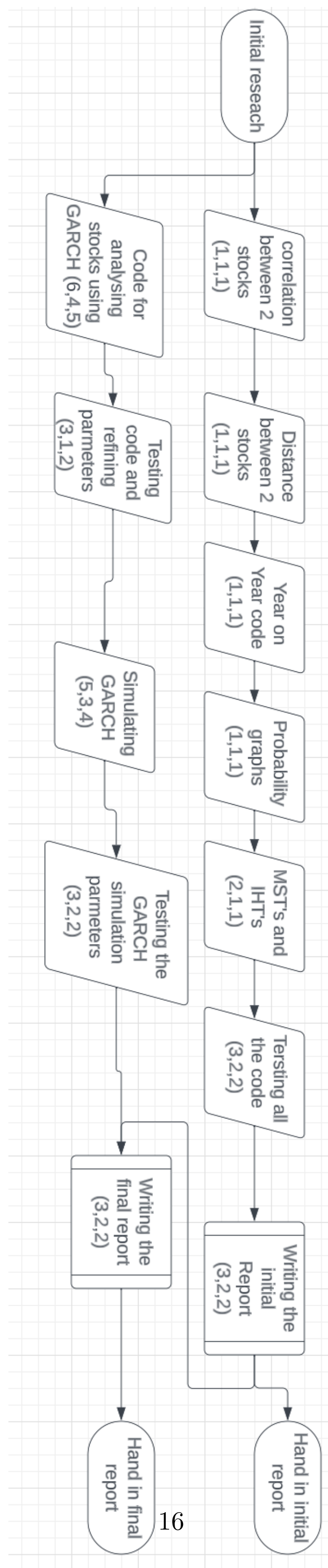

Figure 8: This is the critical path analysis chart that I created for this project. This was used to form the Gantt chart