MSc in Mathematical Engineering - Statistical Learning
Course of Insurance and Econometrics (8 CFU)

GARCH modelling of farmaceutical stocks

Authors :
Mirko Giovio
Daniele Venturini
Paolo Vergottini

Group 18

Course Lecturers:
Prof. Carlo Sgarra
Dr. Silvia Dell'Acqua
Dr. Francesca Grassetti
Dr. Luca Gonzato

POLITECNICO
MILANO 1863

Milan - June 2021

Abstract

AstraZeneca, Johnson & Johnson, Moderna, and Pfizer were unknown names to the wider public and gained extreme importance in the last months due to the production of COVID vaccines. Our goal is to correctly model the value of their stocks and capture the meaningful volatility, following a statistically sound approach. Firstly we performed some exploratory analysis to identify relevant trends and transformed our data to reach stationarity, in order to describe the means of the data series through ARMA models. Then we tried to capture the conditional volatilities with GARCH and EGARCH models. We focused on EGARCH model to investigate the presence of leverage effects. We compared the forecasting powers of GARCH and EGARCH models over a 10 working days horizon with MAE and MSE as indexes, obtaining comparable results. Then we explored possible development directions for this work, by accounting for correlation among the four stocks series and treating their modelling as a joint effort. In order to obtain a reliable estimate of Pearson correlation we built a Bootstrap method over regression residuals.

Contents

1 Introduction

The work is based on a paper, submitted by Dr.Linna Hu from Shanghai University and published in Atlantis Press, entitled "*Research on Stock Returns and Volatility-Based on ARCH - GARCH Model*". The author analyzed time series data for real estate stocks. Firstly, she established a reasonable ARMA model to predict and analyze the stock price. At the same time, to prevent the uncertainty and risk of the stock market, and measure the volatility of the stock yield effectively, a reasonable GARCH model was modeled to study the volatility of the stock return rate. Specifically, the author tackles the problem of conditional heteroskedasticity, not only with a GARCH model but also with a more advanced asymmetric model to account for the leverage effect (the observed tendency of an asset's volatility to be negatively correlated with the asset's returns). Finally, the model was used to forecast future the future trend of stock price, providing reliable information service and decision guidance for investors and decision-makers.
On the other hand, our paper models the stock value of the four leading companies in the COVID vaccine production business. Indeed, the project data is composed of the market close prices of AstraZeneca, Johnson & Johnson, Moderna, and Pfizer on the yearly period ending on April 25th, 2021. We shall see whether these models are suited for farmaceutical stocks. Indeed, we strongly expect the presence of leverage effects, since we directly experience the mediatic wave of vaccine doubts and concerns. In this sense, we believe the EGARCH models to be a fitting choice to assess our hypothesis.
In the following paragraphs, we will detail all the steps of the analysis we performed exploiting the MATLAB software packages.

2 Graphical Exploration and Stationarity

We started by analyzing the Closing Price series. A first graphical exploration quickly highlights the difference in magnitude and evolution between the four companies' values (Figure 1). While AstraZeneca, Johnson & Johnson, and Pfizer seem to have a pretty stable behavior (with some minor increasing or decreasing trends), this is not the case for Moderna, where a growing value pattern is undeniable.

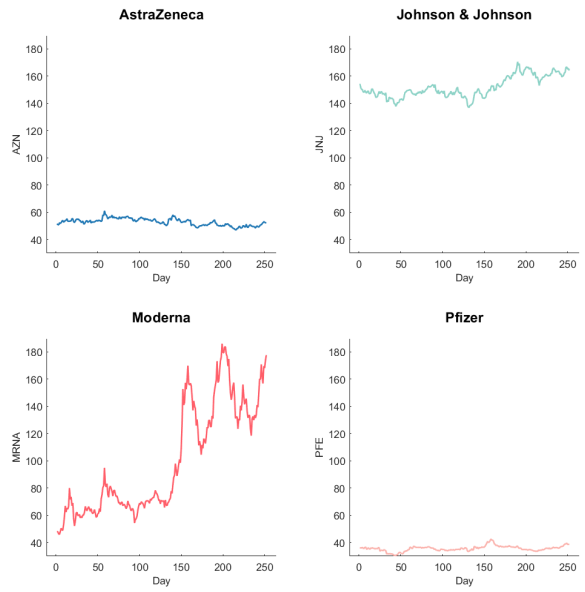

Figure 1: Close prices of the four stocks

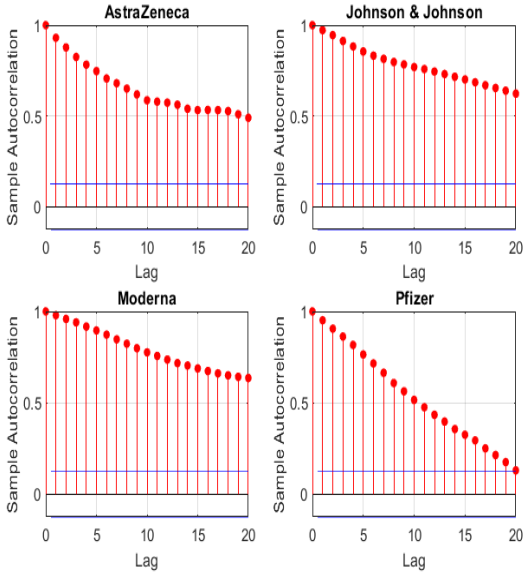

Figure 2: Autocorrelation function of the four stocks

The essential hypothesis of the study is the stationarity of the data series. Indeed, we plotted in Figure 2 the Autocorrelation Function (ACF) to check for the stationarity of the time series . In particular, if the time series contains a unit root then it is non-stationary. From the plot it is clear that the ACF decays to zero very slowly in all cases: this was the first hint at non-stationarity.

We used a statistical hypothesis unit root test to objectively determine whether the series requires differencing. In our analysis, we used the Augmented Dickey Fuller test. The test is done by using the built-in function of MatLab *adftest*. We performerd diffent types of ADF tests:

1. We tested for a restriction of the Augmented Dickey Fuller test, just a simple Dickey-Fuller test with 0 lag and no trend stationary or drift model:

$$H_0: \quad \Delta y(t) = \epsilon(t)$$

$$H_1: \quad \Delta y(t) = \gamma * y(t-1) + \epsilon(t) \quad \gamma \neq 0$$

As we can see in Table 1 the results are very poor. If we assume a threshold for the p-value at 5% we shall reject the null hypothesis of stationarity in all the four cases.

Type	**AstraZeneca**	**Johnson & Johnson**	**Moderna**	**Pfizer**
No TS	0.620	0.763	0.925	0.701
With TS	0.008	0.095	0.511	0.318

Table 1: P-values of the tests with and withouth Trend-Stationarity (TS)

2. We, then, tested for a Dickey-Fuller test with 0 lag, with drift and a trend stationary model (TS).

$$H_0: \quad \Delta y(t) = a_0 + \epsilon(t)$$
$$H_1: \quad \Delta y(t) = a_0 + a_1 * t + \gamma * y(t-1) + \epsilon(t) \quad \gamma \neq 0 \; and \; a_1 \neq 0$$

In this case only the p-value of AstraZeneca is beyond the 5% level.

Although AstraZeneca's p-value decreased dramatically in the TS case, we have ultimately strong statistical evidence each of the closing prices' series contains a unit root; hence the series are not stationary. However, we also know that if a series has l unitary roots, then the difference of l^{th} order will be a stationary series. Therefore we transformed our data series, firstly applying the log operator and then computing the first differences (Figure 7).

$$\Delta P(t) = log(p(t)) - log(p(t-1))$$

Furthermore, the log operator is widely used in financial analysis, since most of the stock's prices usually are log-normally distributed. Indeed, the histograms and the QQ-plots of the transformed prices suggest in most of the cases normality is guaranteed (Figure 3 and 4) .

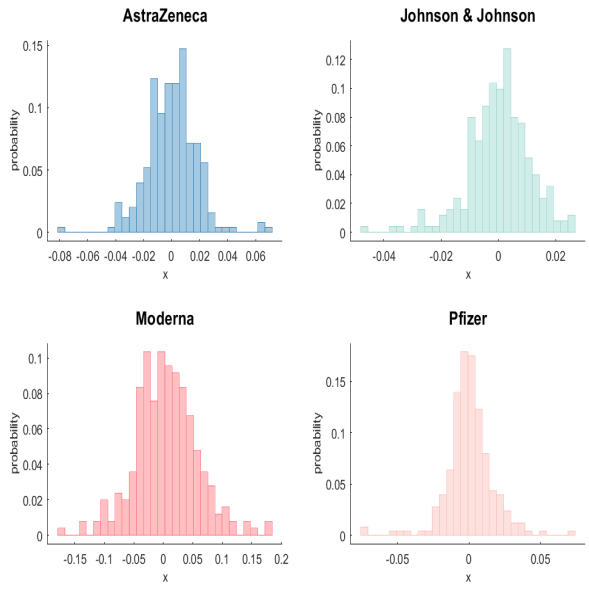

Figure 3: Histograms of the transformed data

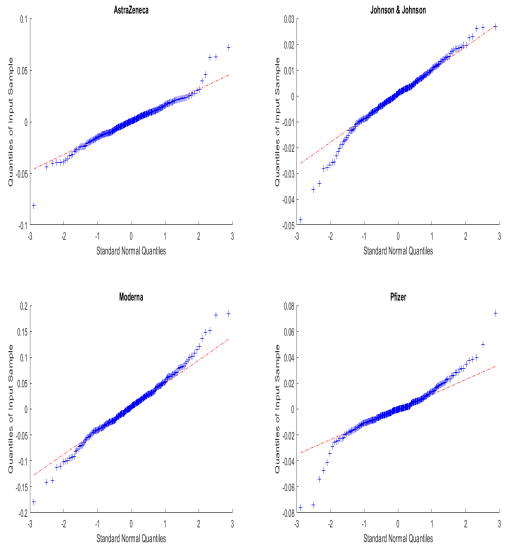

Figure 4: QQ-plots of the transformed data

We then plotted the autocorrelation and partial autocorrelation functions of the difference of log prices. From a first glance, the ACF is definitely better behaved and does not hint at non-stationarity.

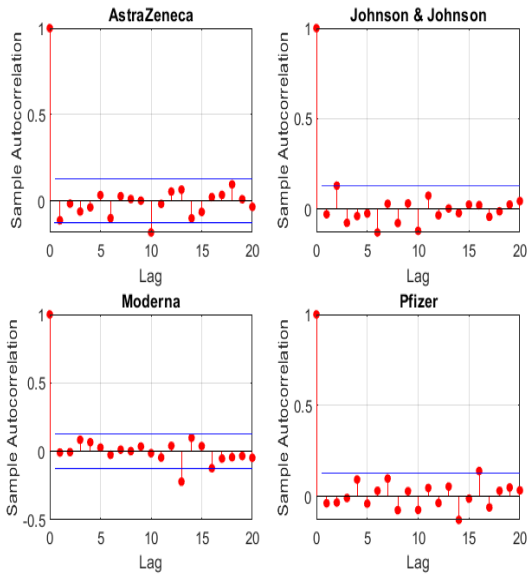

Figure 5: Autocorrelation function of the four transformed stocks values

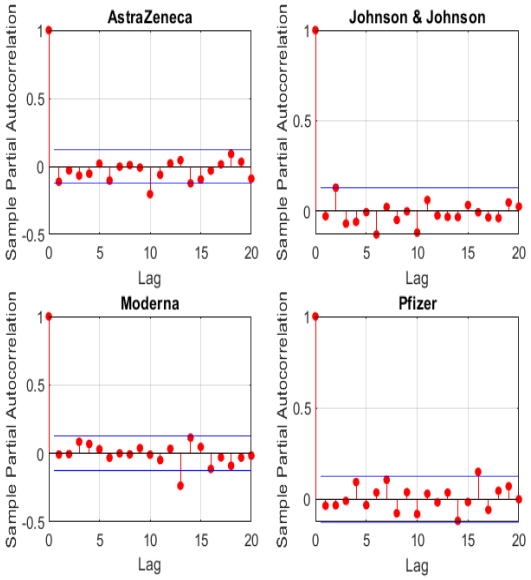

Figure 6: Partial-Autocorrelation function of the four transformed stocks values

Once again we repeated the procedure to test for stationarity with the Augmented Dickey Fuller test:

1. We tested for a restriction of the Augmented Dickey Fuller test, just a simple Dickey-Fuller test with 0 lag and no trend stationary or drift model:

   As shown in Table 2, in all cases MATLAB returns an extremely low p-value of 0.1%, well below the conventional 5% threshold.

2. We, then, tested for a Dickey-Fuller test with 0 lag, with drift and a trend stationary model (TS). Once again the returned p-values are all equal to 0.1%.

Since all the Augmented Dickey-Fuller tests on the log prices gave us the same extremely clear result, we can conclude that all the time series are stationary.

Type	**AstraZeneca**	**Johnson & Johnson**	**Moderna**	**Pfizer**
No TS	$1*10^{-3}$	$1*10^{-3}$	$1*10^{-3}$	$1*10^{-3}$
With TS	$1*10^{-3}$	$1*10^{-3}$	$1*10^{-3}$	$1*10^{-3}$

Table 2: P-values of the tests with and withouth Trend-Stationarity (TS) with transformed data

3 ARMA modelling

Now that we have stationary time series, we can go deeper into the analysis. The conditional mean equation of the stationary time series must always be represented by an AR, MA, or ARMA model.
To select the best model we followed the Box and Jenkins model selection approach.
In the first stage, the identification stage, we examined the time plot of the series, the autocorrelation function, and the partial correlation function (Figure 7,8,9).

In the second stage, the goal was to select a stationary and parsimonious model that has a good fit. Therefore, after the graphical exploration of ACFs and PACFs we estimated ARMA models with AR and MA orders ranging from 0 to 4, considering all possible combinations, with a local search algorithm. At each iteration, we also computed the loglikelihood of each model and obtained the relative Akaike and Bayes Information Criterions (AIC and BIC). These indices aim at finding a parsimonious model which is able to properly fit data, minimizing a trade-off function between number of calibrated parameters (model complexity) and loglikelihood value.
A crucial concept in the second stage of the BoxJenkins approach is the principle of parsimony. Incorporating additional coefficients will necessarily increase fit at a cost of reducing degrees of freedom. Box and Jenkins argue that parsimonious models produce better forecasts than overparameterized models, preventing the risk of overfitting. A parsimonious model fits the data well without incorporating any needless coefficients, which may be the effect of white noise. Indeed, forecasters do not want to project poorly estimated coefficients into the future. The aim is to approximate the true data-generating process carefully but not to restrict too much the model freedom (the well-known bias-variance trade-off).
The third and last stage involved diagnostic checking to ensure that the residuals from the estimated model mimic a white-noise process. Consequently, we decided, for each stock, to plot the residuals autocorrelation function and a Q-Q plot of residuals against the normal distribution. This procedure will allow us to visually detect the normality of the residuals and the presence of significant lags in the autocorrelation. Indeed, the sample autocorrelation function (ACF) is an useful qualitative tools to assess the presence of autocorrelation at individual lags. Nevertheless, to get a reliable quantitative result, we exploited three statistical tests:

1. we used the *one-sample Kolmogorov-Smirnov* non-parametric test to check whether the residuals are normally distributed or not (the non-parametric alternative to the parametric but extremely sensitive Shapiro-Wilk test). Here we used the MATLAB function *kstest*.

$$H0 : res \sim N(\mu, \sigma^2)$$

$$H1 : res \not\sim N(\mu, \sigma^2)$$

2. we carried out the *Ljung-Box Q-test* to verify the presence of significant autocorrelation lags over the residuals. The Ljung-Box Q-test is a more quantitative way to test for autocorrelation at multiple lags jointly. The null hypothesis for this test is that the first m autocorrelations are jointly zero. We ran the *lbqtest* function in MATLAB, with a default number of lags m.

$$H0 : \rho_1 = \rho_2 = ... = \rho_m = 0$$
$$H1 : \exists i \ : \ \rho_i \neq 0$$

3. A time series exhibiting conditional heteroscedasticity -or autocorrelation in the squared series-is said to have autoregressive conditional heteroscedastic (ARCH) effects. So we performed the Engle's ARCH test to account for residuals heteroscedasticity and the existence of ARCH effects. Engle assumes the following form for the squared residuals:

$$\epsilon_t = y_t - \hat{\mu}_t$$
$$\epsilon_t^2 = \alpha_0 + \alpha_1 * \epsilon_{t-1}^2 + ... + \alpha_p * \epsilon_{t-p}^2 + v_t$$

$$\text{H0: } \alpha_1 = \alpha_2 = ... = \alpha_p = 0$$
$$\text{H1: } \alpha_k \neq 0 \text{ for at least one } k = 1, ..., p.$$

Notice that the number of lags p has to be specified. In our work we just ran the function *archtest* with the number of lags selected by default from MATLAB.

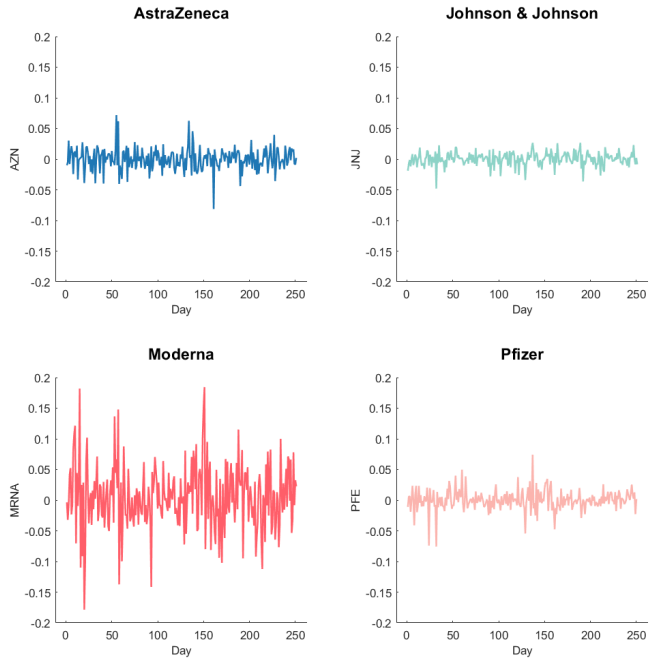

Figure 7: Close prices of the four transformed stocks values

3.1 AstraZeneca

Plotting ACF and PACF we observed a clear drop in both after the first lag. So by exploratory analysis (Figure 5 and 6), we are attracted by an ARMA(1,1) or ARMA(0,1) or ARMA(1,0) model.
The model with the lowest BIC is the ARMA(0,1) while the model with the lowest AIC is the ARMA(3,3). Respecting the parsimony concept, we selected the AR(1) model. This means that the log returns of AstraZeneca's time series are described by a Moving Average model of order 1.

	Value	StandardError	TStatistic	PValue
Constant	8.0048e-05	0.0011496	0.069628	0.94449
AR{1}	-0.11298	0.066239	-1.7056	0.088084
Variance	0.00031092	1.7756e-05	17.511	1.1817e-68

Figure 8: ARMA summary table (AstraZeneca)

Therefore we can build this kind of model obtaining the following p-values. Notice that, despite being the selected model, none of the coefficients (Figure 8) is deemed to be significant except maybe for the AR(1) whose value is below 10% yet above the threshold of 5%. In addition, also the contribution of the constant term is negligible. This suggests that the conditional mean could be well approximated by a straight line with value zero.

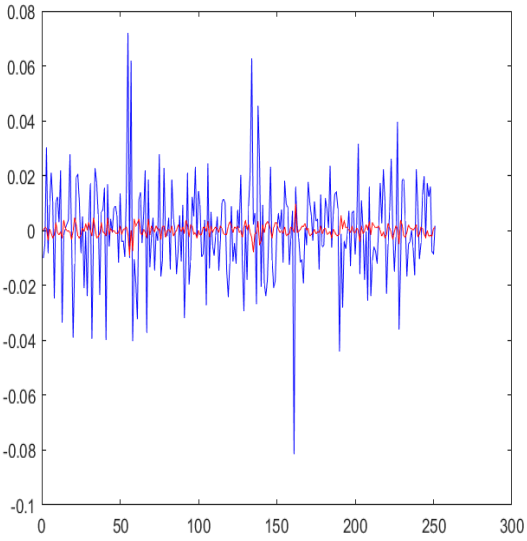

Figure 9: True data series vs estimated one (AstraZeneca)

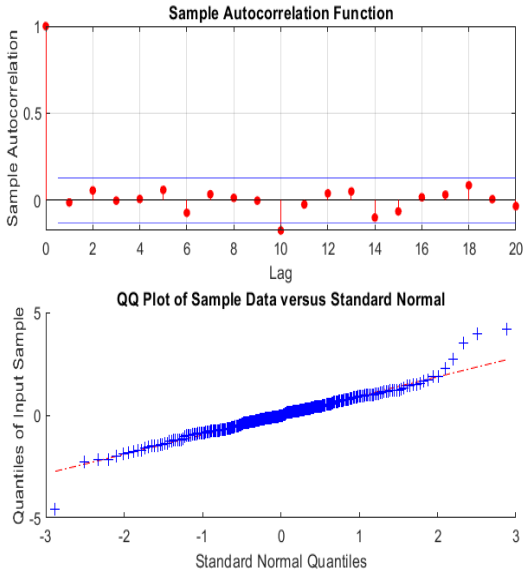

Figure 10: ACF and Q-Q plot of residuals (AstraZeneca)

We, then, proceed with diagnostic checking, plotting the autocorrelation function and a QQ-plot of residuals. Graphically we do not identify noticeable lags in the autocorrelation,

while the Q-Q plot seems to behave nicely (Figure 10), even though there may be a slight problem with the upper tail.
In the AstraZeneca's case all the three p-values of the tests are well above 5 %; therefore in all cases I will not have sufficient evidence to reject the null hypothesis. Indeed, the residuals are assumed to be gaussian, there is no significant autocorrelation lags as well as no remarkable ARCH effects.

p-value K-S test : 0.67
p-value L-B test: 0.23
p-value E-A test: 0.82

3.2 Johnson & Johnson

Plotting ACF and PACF we cannot observe only a one clear drop after a certain lag. So with a simple exploratory analysis it is difficult to assess which is the right model (Figure 5 and 6).
As a consequence we had to determine the correct model order exploiting the value of the BIC and AIC coefficients: they both attain their minimal value in (3,3), thus we selected an ARMA(3,3) model.

	Value	StandardError	TStatistic	PValue
Constant	0.00013651	1.5457e-05	8.8317	1.0309e-18
AR{1}	-0.24535	0.022928	-10.7	1.0127e-26
AR{2}	0.19717	0.018418	10.705	9.5977e-27
AR{3}	0.84371	0.017251	48.907	0
MA{1}	0.17985	0.021299	8.4444	3.0554e-17
MA{2}	-0.17985	0.021043	-8.547	1.2632e-17
MA{3}	-1	0.019472	-51.354	0
Variance	0.00010506	8.7308e-06	12.034	2.3673e-33

Figure 11: ARMA summary table (Johnson & Johnson)

The p-value table highlights how all the coefficients are significant, since all their p-values are very close to zero (Figure 7). Hence, we can assume this model to be meaninful to explain the conditional mean of the log returns for Johnson & Johnson.

Through graphical exploration,we cannot identify remarkable lags in the autocorrelation, while the Q-Q plot ibehaves quite nicely (Figure 13), in spite of a slight problem with the lower tail.

p-value K-S test : 0.31
p-value L-B test : 0.88
p-value E-A test : 0.88

As long as the diagnostic is concerned, in Johnson & Johnson's case all the three p-values of the tests are above 0.3; therefore we will not have sufficient evidence to reject the null hypothesis in any of the tests. Indeed, the residuals are assumed to be gaussian, there appear not to be significant autocorrelation lags as well as no remarkable ARCH effects.

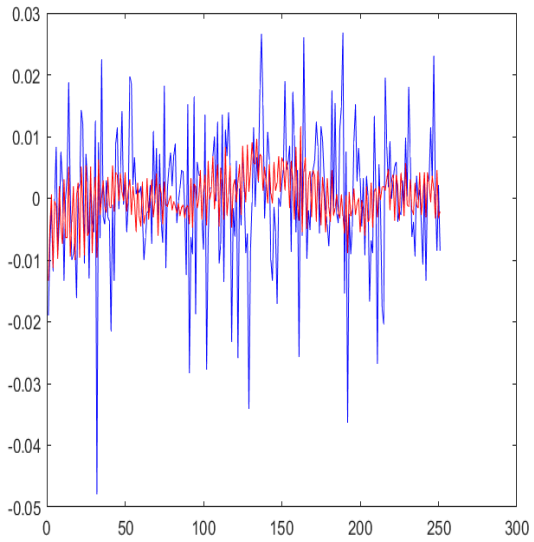

Figure 12: True data series vs estimated one (Johnson & Johnson)

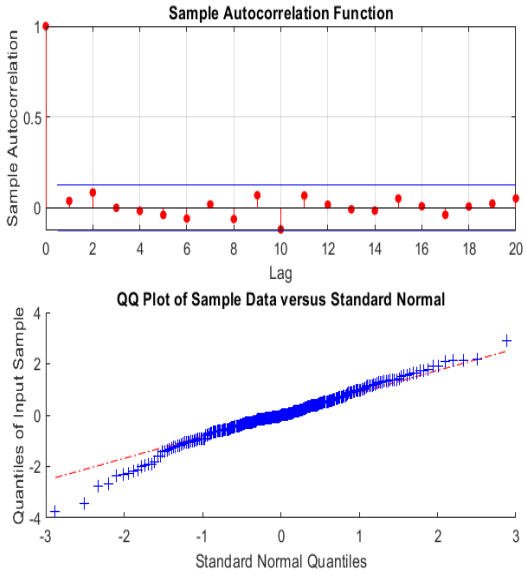

Figure 13: ACF and Q-Q plot of residuals (Johnson & Johnson)

3.3 ARMA order for Moderna

From the plot of the ACF and the PACF and we do not notice a clear behaviour after the first lag (Figure 5 and 6). Luckly, AIC and BIC are concordant, since they both attain their minimal value in (2,2). So we select an ARMA(2,2) model to explain the conditional mean for Moderna.

	Value	StandardError	TStatistic	PValue
Constant	0.011268	0.0068764	1.6387	0.10128
AR{1}	-0.19454	0.0061554	-31.604	3.2424e-219
AR{2}	-0.96823	0.0070132	-138.06	0
MA{1}	0.19638	0.018424	10.659	1.5825e-26
MA{2}	1	0.019973	50.068	0
Variance	0.0026938	0.00021445	12.562	3.4274e-36

Figure 14: ARMA summary table (Moderna)

The summary (Figure 14) displays overall negligible p-values, except for the constant term, which may as well be put to zero. Consequently, we can deem the model to be statistically sound.

To assess the validity of the model we proceeded with plotting the ACF and he QQ-plot of residuals (Figure 16). The plots do not suggest the existence of remarkable lags in the ACF, while we may suppose the normality of the data.

p-value K-S test : 0.63
p-value L-B test : 0.21
p-value E-A test : 1.7 e-05

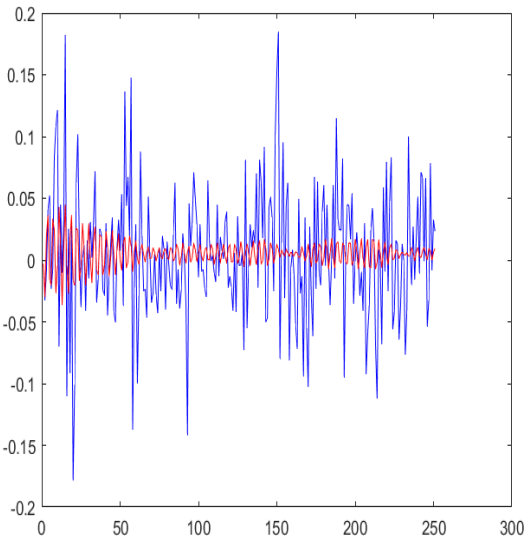

Figure 15: True data series vs estimated one (Moderna)

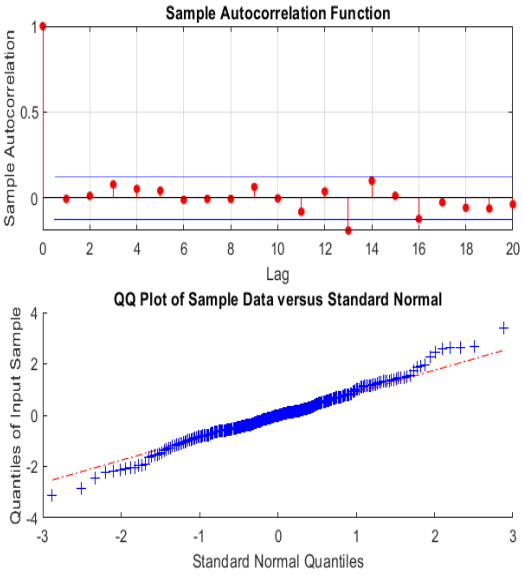

Figure 16: ACF and Q-Q plot of residuals (Moderna)

The Kolmogorov-Smirnov normality test has a p-value over 60%, providing more than enough evidence towards the gaussianity of residuals. The autocorrelation test accepts the null hypothesis with a p-value around 20%. Finally, Engle's test has an extremely little p-value, almost negligible. Thus, we can assume the presence of ARCH effects. This will lead to the analysis in the paragraph 4.3.

3.4 ARMA order for Pfizer

Plotting ACF and PACF of log returns the parsimonious ARMA(1,1) seems a suitable option.
On the other hand, with the local search algorithm AIC selects a ARMA(4,4) model as well as BIC. Due to the concordance of the two indexes, we chose an ARMA(4,4) model.

	Value	StandardError	TStatistic	PValue
Constant	2.3963e-05	0.00081354	0.029455	0.9765
AR{1}	0.65836	0.048029	13.708	9.1522e-43
AR{2}	-0.072875	0.066088	-1.1027	0.27016
AR{3}	0.4726	0.062214	7.5964	3.0453e-14
AR{4}	-0.80583	0.041694	-19.327	3.1719e-83
MA{1}	-0.62719	0.034063	-18.413	1.0335e-75
MA{2}	0.049876	0.03428	1.4549	0.14568
MA{3}	-0.56563	0.032061	-17.642	1.1636e-69
MA{4}	0.96447	0.027811	34.679	1.6362e-263
Variance	0.00024906	1.5294e-05	16.285	1.2701e-59

Figure 17: ARMA summary table (Pfizer)

Despite a few not significant coefficients (e.g. AR(2) , MA(2) and the constant term), on the whole, we can consider the model statistically sound (Figure 17). In particular the MA(4) and the AR(4) are meaningful, beyond any reasonable doubt.

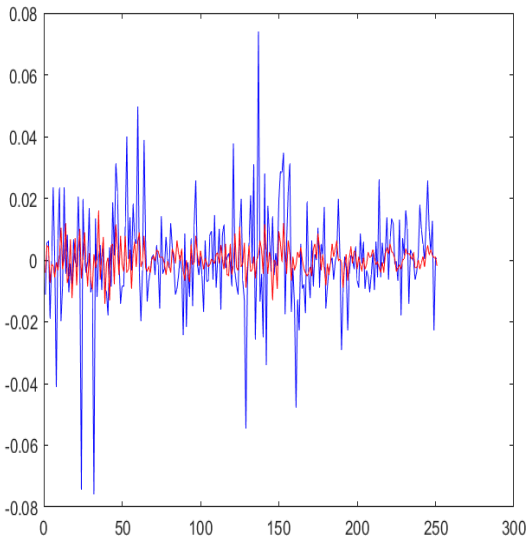

Figure 18: True data series vs estimated one (Pfizer)

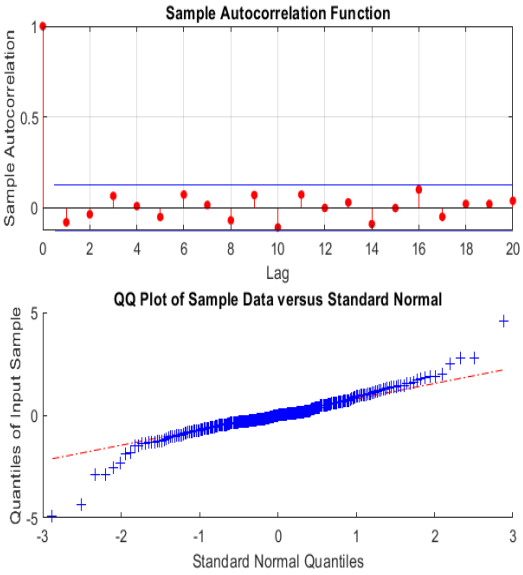

Figure 19: ACF and Q-Q plot of residuals (Pfizer)

In Pfizers case, the Q-Q plot is quite ill-behaved with abnormalities over both tails. On the flip side, the sample autocorrelation function does not hint at the existence of significant lags after zero (Figure 19).

p-value K-S test : 0.02
p-value L-B test : 0.39
p-value E-A test : 0.61

Indeed, the one-sample Kolmogorov-Smirnov test rejects the gaussianity of residuals at 5% (with a p-value around 2%), while the Ljung-Box Q-test accepts the null hypothesis (with a p-value around 50%). Moreover Engle's ARCH test hints at the absence of ARCH effects (with a p-value of 80%).

4 GARCH modelling

From the previous analysis we can deduce that for all the conditional mean models there is no residual autocorrelation. Moreover the residual sequences could all be considered white noise sequences, except for the case of Moderna. Nevertheless, our primary objective was to capture the volatility of the stock returns effectively and to verify the existence of leverage effects. GARCH models may still be an effective tool in this direction. So, following the suggestion of the reference paper, we still proceeded exploring the possible GARCH models. We will follow the same scheme used in chapter 3 for ARMA modelling:

1. graphical analysis of the plot autocorrelation and partial autocorrelation of squared residuals
2. loop simulation of GARCH(p,q) with Grid-Search
3. model selection based on lowest BIC / AIC
4. p-value analysis

It would be, however, meaningless to test for the presence of ARCH effects in the residuals of the GARCH models, since we already know that not to be the case. Hence, we will perform diagnostic only in chapter 6, by confronting the predictive power of GARCH models against the EGARCH models, that we will estimate in the following chapter.

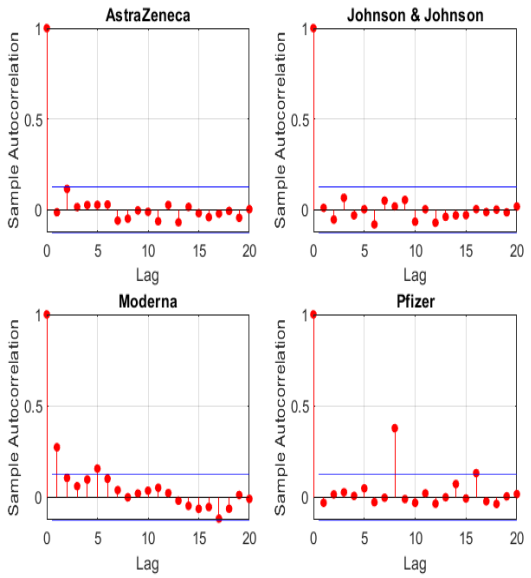

Figure 20: Autocorrelation function of the squared residuals

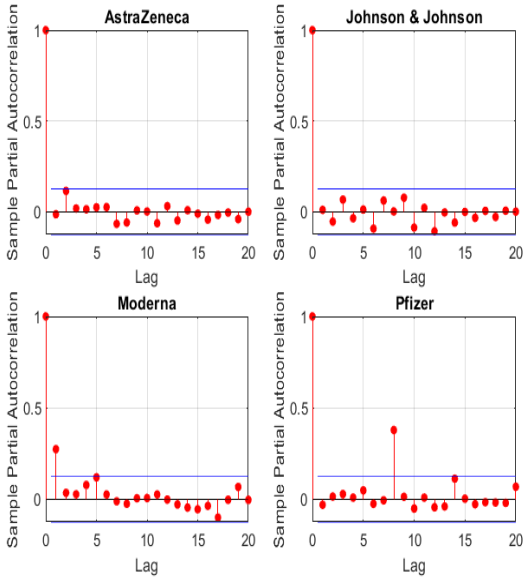

Figure 21: Partial-Autocorrelation function of the squared residuals

4.1 AstraZeneca

Both the squared residuals autocorrelation and partial autocorrelation functions are not significant to graphically detect the order of the GARCH model. Therefore we performed an extensive search over a parameter grid.AIC selected a GARCH(0,2) model, while BIC opted for a GARCH(0,1). At first we considered the simpler ARCH(1) model, but the obtain p-value was so high (almost 1) we had to reject it in favour of an ARCH(2) model.

	Value	StandardError	TStatistic	PValue
Constant	0.00029176	1.9867e-05	14.685	8.0129e-49
ARCH{1}	0	0.0411	0	1
ARCH{2}	0.056059	0.039992	1.4018	0.16098

Figure 22: GARCH summary table (AstraZeneca)

As a matter of fact, from the summary table (Figure 22) we observe that the coefficient of the ARCH(1) lag is set to zero by the model, while the coefficient for the ARCH(2) term is significant at 15%. Although the model's result may be poor, we still retain it, in line with the idea presented in the beginning of this chapter.

4.2 Johnson & Johnson

As we did with AstraZeneca, the graphical analysis happened to be unresolving, so we performed a grid analysis on the AIC and BIC indicators. This time we deduced that the best model was a GARCH(2,1).

+

	Value	StandardError	TStatistic	PValue
Constant	2.7706e-05	3.1578e-05	0.87738	0.38028
GARCH{1}	0	0.22381	0	1
GARCH{2}	0.6695	0.23793	2.8138	0.0048955
ARCH{1}	0.089373	0.073082	1.2229	0.22136

Figure 23: GARCH summary table (Johnson & Johnson)

While the ARCH(1) coefficient is not meaninful at the 5% threshold, the GARCH part is definitely statistically sound. The GARCH(2) term has a coefficient with p-value less than 5‰. On the whole, we have strong evidence in favour of this model.

4.3 Moderna

In this case we notice from the autocorrelation and partial autocorrelation functions that a GARCH(2,2) would be a wise choice. However, both the BIC and the AIC indexes lean towards a GARCH(0,2) model and at the same time the p-values obtained with the GARCH(2,2) model were dramatically high. Finally,we chose an ARCH(2) model.

	Value	StandardError	TStatistic	PValue
Constant	0.0016182	0.00026283	6.1569	7.4173e-10
ARCH{1}	0.22333	0.10572	2.1124	0.03465
ARCH{2}	0.24274	0.095777	2.5344	0.011263

Figure 24: GARCH summary table (Moderna)

In this case, the p-values are quite satisfying, so we can consider ARCH(2) as a solid model for the conditional variance of Moderna. This is consistent with the result of the Engle's ARCH test.

4.4 Pfizer

Here we observed a significant at the 8^{th} lag in the ACF and PACF of squared residuals. Therefore we ran the loop for p=0,..9 and q=1,...,9. At the end AIC selected an ARCH(9) while BIC selected an ARCH(8). We opted for the simpler ARCH(8).

The summary table shows that the 8^{th} lag is surely significant, while all the others are not meaningful at 5%. On the whole we consider the model valid, since all the other terms have coefficient at least 10 times lower in value.

GARCH(0,8) Conditional Variance Model (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	6.0608e-05	1.8778e-05	3.2276	0.0012482
ARCH{1}	0.0020038	0.043632	0.045926	0.96337
ARCH{2}	0.089737	0.049159	1.8254	0.067934
ARCH{3}	0	0.021586	0	1
ARCH{4}	0.051635	0.048008	1.0755	0.28214
ARCH{5}	0.048747	0.059049	0.82553	0.40907
ARCH{6}	0.005082	0.041468	0.12255	0.90246
ARCH{7}	0.022994	0.045969	0.5002	0.61694
ARCH{8}	0.65007	0.15238	4.2661	1.9888e-05

Figure 25: GARCH summary table (Pfizer)

5 EGARCH modelling

To capture the leverage effect for each of the times series of the 4 companies, we developed an EGARCH model (exponential-GARCH model) The model is structured as follow:

$$\ln(h_t) = \alpha_0 + \sum_{k=1}^{q} \alpha_k * (\epsilon_{t-k}/h_{t-k}^{0.5}) + \sum_{w=1}^{q} \lambda_w * |\epsilon_{t-w}/h_{t-w}^{0.5}| + \sum_{j=1}^{p} \beta_j * \ln(h_{t-j})$$

The equation for the conditional variance is in log-linear form. Regardless of the magnitude of ln(ht), the implied value of ht can never be negative. Hence, it is permissible for the coefficients to be negative. Most importantly, the EGARCH model allows for leverage effects.
Let's consider an EGARCH(1,1) model. If $\epsilon_{t-1}/h_{t-1}^{0.5}$ is positive, the effect of the shock on the log of the conditional variance is $\alpha_1 + \lambda_1$. If $\epsilon_{t-1}/h_{t-1}^{0.5}$ is negative, the effect of the shock on the log of the conditional variance is -$\alpha_1 + \lambda_1$. Therefore the model treats differently the impact of negative shocks (bad news) and of positive shocks (good news).
Using this model we try to improve our model estimation for the condition volatility. Here we decided not to estimate the EGARCH order with a loop search, but rather to use in all cases the simple and parsimonious EGARCH(1,1), widely and commonly used in the literature.

	Value	StandardError	TStatistic	PValue
Constant	1.4842	0.42779	3.4694	0.00052171
GARCH{1}	0.81766	0.051995	15.726	1.0081e-55
ARCH{1}	-0.13945	0.059334	-2.3503	0.018759
Leverage{1}	0.28365	0.074997	3.7821	0.00015551

Figure 26: AstraZeneca

	Value	StandardError	TStatistic	PValue
Constant	10	3.3697	2.9676	0.0030015
GARCH{1}	-0.093976	0.3668	-0.2562	0.79779
ARCH{1}	0.28104	0.14139	1.9877	0.046847
Leverage{1}	0.2224	0.080744	2.7544	0.0058807

Figure 27: Johnson & Johnson

In all the cases, but Moderna, we can see that the leverage coefficient is absolutely significant, hinting at the presence of leverage effect. Overall, our initial study hypothesis is confirmed.

	Value	StandardError	TStatistic	PValue
Constant	-1.0085	0.52595	-1.9174	0.055182
GARCH{1}	0.83125	0.088006	9.4454	3.5418e-21
ARCH{1}	0.29837	0.14141	2.1099	0.034864
Leverage{1}	0.098357	0.073068	1.3461	0.17827

Figure 28: Moderna

	Value	StandardError	TStatistic	PValue
Constant	-15.812	0.50543	-31.285	7.3699e-215
GARCH{1}	0.90148	0.052165	17.281	6.5028e-67
ARCH{1}	0.11668	0.086449	1.3497	0.17712
Leverage{1}	0.15996	0.0603	2.6527	0.0079857

Figure 29: Pfizer

6 Forecasting

To assess the performance of the GARCH and EGARCH models we decided to confront their forecasted values over an horizon of 10 working days from the 26th of April 2021 to the 5th of May 2021. We computed the Mean Absolute Error (MAE) as well as the Mean Squared Error (MSE) taking as base value the real value of log returns. The definition of the two indexes are the following:

$$MAE = \frac{\sum_{i=0}^{n} \mid \hat{y}_i - y_i \mid}{n}$$

$$MSE = \frac{\sum_{i=0}^{n} (\hat{y}_i - y_i)^2}{n}$$

where y_i and $\hat{y}_i$ are respectively the true value and the estimated one.

In Table 3 are reported the values of MAE and MSE for the two types of models for each company. First of all we can notice that in AstraZeneca's and Johnson Johnson's case the MAE and MSE values are pretty close for the GARCH and the EGARCH model. Therefore we may assume the predictive power of the two model families to be similar.

	AstraZeneca	**Johnson & Johnson**	**Moderna**	**Pfizer**
MAE GARCH	0.0029	0.0091	0.0335	0.0134
MAE EGARCH	0.0028	0.0090	0.0553	0.0088
MSE GARCH	$1.25*10^{-5}$	$1.12*10^{-4}$	0.0015	$3.4*10^{-4}$
MSE EGARCH	$1.17*10^{-5}$	$1.14*10^{-4}$	0.0043	$1.5*10^{-4}$

Table 3: MAE and MSE values for GARCH and EGARCH models

This is not the case for Moderna and Pfizer: for the former the GARCH model achieves better predictive performances, while for the latter the converse conclusion holds.

7 Correlation

In our work we treated the four stocks as independent, performing four split-off analysis. However if there were a meaningful correlation among the stocks' values, we could exploit this valuable property in the modelling. Indeed, we can easily compute all the correlations coefficients and observe they are positive. This empirical conclusion is fully consistent with the nature of information: whenever one of the four vaccines was at the center of the public debate due to some negative news, all of other companies were affected.
In order to obtain a more robust result, if compared to an overly simplistic point estimate, we decided to opt for a bootstrap approach to retrieve an empirical, yet theoretically valid, distribution of correlations. Given two stocks, we built an elementary linear regression among the two and computed the fitted values, the residuals and the Pearson correlation coefficient between the two vector. We, then, sampled with replacement the residuals vector, summed the new residual to its fitted value and computed once again the correlation coefficient between the original independent variable vector and the newly built one. By repeating this procedure a great amount of times (10000 repetitions), we managed to extract a sample of correlations. Finally, we plotted the histograms and built confidence intervals for the correlation values.

	Inf	**Mean**	**Sup**
AZN-JNJ	0.268	0.376	0.484
AZN-MRNA	0.081	0.199	0.319
AZN-PFE	0.166	0.300	0.393
JNJ-MRNA	0.003	0.118	0.240
JNJ-PFE	0.374	0.468	0.561
MRNA-PFE	0.064	0.183	0.303

Figure 30: 95% Confidence Intervals for Pearson Correlations

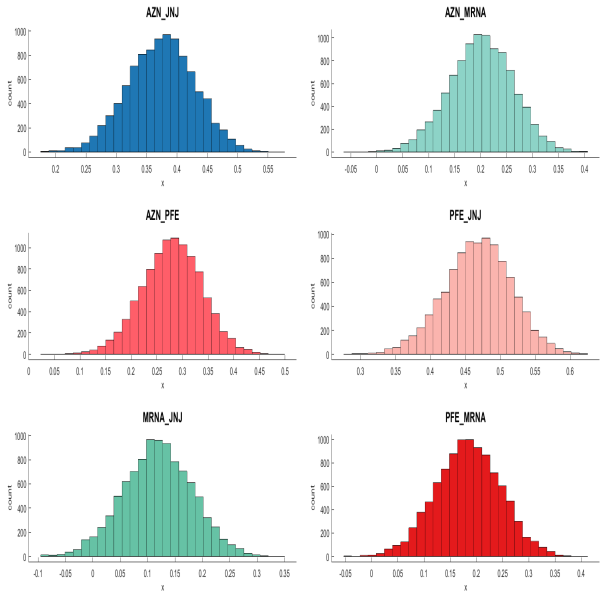

Figure 31: Histograms for correlation obtained by Bootstrap

All the correlation are contained into strictly positive confidence interval at 95%. We can, then, reject an initial hypothesis of independence (since 0 is not contained in the confidence interval) and state the existence of positive correlation among the stocks.

8 Conclusion

In this paper, the time series of farmaceutical stocks are modeled and analyzed. Through MATLAB, initial ARMA models are established for the data, and the advantages and disadvantages of each order are compared. Finally, the ARMA models are determined. While the ARCH effects are found in the residual test. Therefore, the final models are combined with the ARMA models and the GARCH/EGARCH models, and the two are combined to estimate. The ARCH model can eliminate the conditional heteroscedasticity for stock and other volatility data. The combination of the two can make the model more realistic and provide investors with better reference. We ultimately draw the following conclusions:

- The EGARCH model is a good way to describe the asymmetric phenomenon of shocks in the stock market, that is, the response of investors to the same amount of bad news and good news is different. Thanks to the EGARCH model, we managed to verify the presence of Leverage effects in our data, but Moderna's log returns.
- The forecasting power of the GARCH and EGARCH models is quite comparable, even though there are specific cases in which one model behaves better in predicting the stock's value if compared to the other.
- We rejected the hypothesis of independence and built confidence intervals for the correlations of the log returns. Since all the correlations are positive and significant, a possible evolution of the work may concern a joint modelling of the stocks' values.

References

[1] Lina Hu. *Research on Stock Returns and Volatility-Based on ARCH - GARCH Model* . Atlantis Press, October 2017.

[2] Walter Enders. *Applied Econometrics Time Series, 4^{th} edition.* Wiley, University of Alabama, 2014.

A Matlab Appendix

A.1 Main

```
%% Load all the needed data!

warning('OFF', 'MATLAB:table:ModifiedAndSavedVarnames');
AZN = readtable('AZN.csv');
AZN = AZN(:,5);
AZN = table2array(AZN);
JNJ = readtable('JNJ.csv');
JNJ = JNJ(:,5);
JNJ = table2array(JNJ);
MRNA = readtable('MRNA.csv');
MRNA = MRNA(:,5);
MRNA = table2array(MRNA);
PFE = readtable('PFE.csv');
PFE = PFE(:,5);
PFE = table2array(PFE);

%Plot autocorrelation
figure();
subplot(2,2,1);
autocorr(AZN);
title('AstraZeneca');
subplot(2,2,2);
autocorr(JNJ);
title('Johnson & Johnson');
subplot(2,2,3);
autocorr(MRNA);
title('Moderna');
subplot(2,2,4);
autocorr(PFE);
title('Pfizer');


% Plot all the time series

clear g;
g(1,1)=gramm('x',1:252,'y',AZN);
g(1,1).geom_line();
g(1,1).set_names('x','Day','y','AZN');
g(1,1).axe_property('YLim',[30 190]);
g(1,1).set_title('AstraZeneca');
g(1,2)=gramm('x',1:252,'y',JNJ);
g(1,2).geom_line();
g(1,2).axe_property('YLim',[30 190]);
g(1,2).set_names('x','Day','y','JNJ');
g(1,2).set_title('Johnson & Johnson');
g(2,1)=gramm('x',1:252,'y',MRNA);
g(2,1).geom_line();
g(2,1).axe_property('YLim',[30 190]);
g(2,1).set_names('x','Day','y','MRNA');
```

```
g(2,1).set_title('Moderna');
g(2,2)=gramm('x',1:252,'y',PFE);
g(2,2).geom_line();
g(2,2).set_names('x','Day','y','PFE');
g(2,2).set_title('Pfizer');
g(2,2).axe_property('YLim',[30 190]);
figure('Position',[100 100 800 800]);
g(1,1).set_color_options('map','d3_10');
g(1,2).set_color_options('map','brewer3');
g(2,1).set_color_options('map','winter');
g(2,2).set_color_options('map','brewer_pastel');
g.draw();

% TEST STATIONARITY
% Exploit the Augmented Dickey Fuller to check for the presence of ...
    unit root.
% Overall we observe high p-values; therefore we cannot reject the ...
    null hypothesis
% of unit root. Stationarity is not guaranteed.

%No TS
[¬,pValue] = adftest(AZN);
pValue
[¬,pValue] = adftest(JNJ);
pValue
[¬,pValue] = adftest(MRNA);
pValue
[¬,pValue] = adftest(PFE);
pValue

%TS
[¬,pValue] = adftest(AZN,'model','TS');
pValue
[¬,pValue] = adftest(JNJ,'model','TS');
pValue
[¬,pValue] = adftest(MRNA,'model','TS');
pValue
[¬,pValue] = adftest(PFE,'model','TS');
pValue

%% LOG RETURNS

AZN=log(AZN);
laggedAZN = lagmatrix(AZN,1);
diff=AZN-laggedAZN;
AZN_pct=diff;
AZN_pct(1,:)=[];

JNJ=log(JNJ);
laggedJNJ = lagmatrix(JNJ,1);
diff=JNJ-laggedJNJ;
JNJ_pct=diff;
JNJ_pct(1,:)=[];

MRNA=log(MRNA);
```

```
laggedMRNA = lagmatrix(MRNA,1);
diff=MRNA-laggedMRNA;
MRNA_pct=diff;
MRNA_pct(1,:)=[];

PFE=log(PFE);
laggedPFE = lagmatrix(PFE,1);
diff=PFE-laggedPFE;
PFE_pct=diff;
PFE_pct(1,:)=[];

%plot ACF and PACF and QQ_plot

figure();
subplot(2,2,1);
autocorr(AZN_pct);
title('AstraZeneca');
subplot(2,2,2);
autocorr(JNJ_pct);
title('Johnson & Johnson');
subplot(2,2,3);
autocorr(MRNA_pct);
title('Moderna');
subplot(2,2,4);
autocorr(PFE_pct);
title('Pfizer');

figure();
subplot(2,2,1);
parcorr(AZN_pct);
title('AstraZeneca');
subplot(2,2,2);
parcorr(JNJ_pct);
title('Johnson & Johnson');
subplot(2,2,3);
parcorr(MRNA_pct);
title('Moderna');
subplot(2,2,4);
parcorr(PFE_pct);
title('Pfizer');

figure();
subplot(2,2,1);
qqplot(AZN_pct);
title('AstraZeneca');
subplot(2,2,2);
qqplot(JNJ_pct);
title('Johnson & Johnson');
subplot(2,2,3);
qqplot(MRNA_pct);
title('Moderna');
subplot(2,2,4);
qqplot(PFE_pct);
title('Pfizer');

```

```
% Plot first differences

clear g
g(1,1)=gramm('x',1:251,'y',AZN_pct);
g(1,1).geom_line();
g(1,1).set_names('x','Day','y','AZN');
g(1,1).axe_property('YLim',[-0.2 0.2]);
g(1,1).set_title('AstraZeneca');
g(1,2)=gramm('x',1:251,'y',JNJ_pct);
g(1,2).geom_line();
g(1,2).axe_property('YLim',[-0.2 0.2]);
g(1,2).set_names('x','Day','y','JNJ');
g(1,2).set_title('Johnson & Johnson');
g(2,1)=gramm('x',1:251,'y',MRNA_pct);
g(2,1).geom_line();
g(2,1).axe_property('YLim',[-0.2 0.2]);
g(2,1).set_names('x','Day','y','MRNA');
g(2,1).set_title('Moderna');
g(2,2)=gramm('x',1:251,'y',PFE_pct);
g(2,2).geom_line();
g(2,2).set_names('x','Day','y','PFE');
g(2,2).set_title('Pfizer');
g(2,2).axe_property('YLim',[-0.2 0.2]);
figure('Position',[100 100 800 800]);
g(1,1).set_color_options('map','d3_10');
g(1,2).set_color_options('map','brewer3');
g(2,1).set_color_options('map','winter');
g(2,2).set_color_options('map','brewer_pastel');
g.draw();

% histograms
clear g;
g2(1,1)=gramm('x',AZN_pct);
g2(1,2)=gramm('x',JNJ_pct);
g2(2,1)=gramm('x',MRNA_pct);
g2(2,2)=gramm('x',PFE_pct);
g2(1,1).stat_bin('normalization','probability','geom','overlaid_bar');
g2(1,2).stat_bin('normalization','probability','geom','overlaid_bar');
g2(2,1).stat_bin('normalization','probability','geom','overlaid_bar');
g2(2,2).stat_bin('normalization','probability','geom','overlaid_bar');
g2(1,1).set_title('AstraZeneca');
g2(1,2).set_title('Johnson & Johnson');
g2(2,1).set_title('Moderna');
g2(2,2).set_title('Pfizer');
g2(1,1).set_color_options('map','d3_10');
g2(1,2).set_color_options('map','brewer3');
g2(2,1).set_color_options('map','winter');
g2(2,2).set_color_options('map','brewer_pastel');
figure('Position',[100 100 800 600]);
g2.draw();

%% TEST STATIONARITY
% As before I test for unit root. Since all the p-values are low I can ...
    assume
% stationarity of the time series.
```

```
212
213 %No TS
214
215 [¬,pValue] = adftest(AZN_pct);
216 pValue
217 [¬,pValue] = adftest(JNJ_pct);
218 pValue
219 [¬,pValue] = adftest(MRNA_pct);
220 pValue
221 [¬,pValue] = adftest(PFE_pct);
222 pValue
223
224
225 %TS
226 [¬,pValue] = adftest(AZN_pct,'model','TS');
227 pValue
228 [¬,pValue] = adftest(JNJ_pct,'model','TS');
229 pValue
230 [¬,pValue] = adftest(MRNA_pct,'model','TS');
231 pValue
232 [¬,pValue] = adftest(PFE_pct,'model','TS');
233 pValue
234
235 %% ARMA order for AZN
236 % We plot ACF and PACF and I observe a clear drop in both after the ...
      first lag.
237 % So by exploratory analysis we would choose an ARMA(1,1). Then we ...
      proceed by
238 % simulating the BIC and AIC coefficients. BIC identifies the best ...
      model in (1,1),
239 % while AIC prefers the more complex (3,2) model. Overall we chose the ...
      (1,1) model.
240
241 LOGL = zeros(5,5); % Initialize
242 PQ = zeros(5,5);
243
244  for p = 0:4
245      for q = 0:4
246         Mdl = arima(p,0,q);
247         [EstMdl,¬,logL] = estimate(Mdl,AZN_pct,'Display','off');
248         LOGL(p+1,q+1) = logL;
249         PQ(p+1,q+1) = p + q;
250      end
251 end
252
253 LOGL = reshape(LOGL,25,1);
254 PQ = reshape(PQ,25,1);
255 [aic,bic] = aicbic(LOGL,PQ+1,100);
256 bicm=reshape(bic,5,5); %minimal value in (1,1)
257 aicm=reshape(aic,5,5); %minimal value in (3,3), but also (1,1) good value
258
259 %%
260
261 T=length(AZN_pct);
262 mod=arima(1,0,0);
```

```
AZN_mod=estimate(mod,AZN_pct);

AZN_res= infer(AZN_mod,AZN_pct);
AZN_pred = AZN_pct-AZN_res;

figure()
plot(AZN_pct,'-b')
hold on
plot(AZN_pred,'-r')

dy=AZN_pct;
Mdl = mod; % define the model
EstMdl = AZN_mod; % perform estimation
[dyF,dyMSE] = forecast(EstMdl,1,dy(end-10:end)); % forecast h step ...
    ahead %try last 10 days
dyF_ci = [dyF - 1.96*sqrt(dyMSE) dyF + 1.96*sqrt(dyMSE)]
figure
plot(dy,'Color',[.7,.7,.7]);
hold on
plot(T+1,dyF,'bo','LineWidth',2);
plot(T+1,dyF + 1.96*sqrt(dyMSE),'rd',...
        'LineWidth',2);
plot(T+1,dyF - 1.96*sqrt(dyMSE),'rd','LineWidth',2);
legend('Observed','Forecast',...
        '95\% Confidence ...
           Interval','Location','Best','interpreter','latex');
title('1-Step Ahead','interpreter','latex')
grid on
set(gca,'FontSize',20)

%% ARMA order for JNJ
% We plot ACF and PACF and I observe a clear drop in both after the ...
    first lag.
% So by exploratory analysis we would choose an ARMA(1,1).  BIC and ...
    AIC both identify
% the best model in (3,3).Overall we chose the (3,3) model.
LOGL = zeros(5,5); % Initialize
PQ = zeros(5,5);

 for p = 0:4
     for q = 0:4
        Mdl = arima(p,0,q);
        [EstMdl,¬,logL] = estimate(Mdl,JNJ_pct,'Display','off');
        LOGL(p+1,q+1) = logL;
        PQ(p+1,q+1) = p + q;
     end
end

LOGL = reshape(LOGL,25,1);
PQ = reshape(PQ,25,1);
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicm=reshape(bic,5,5);
aicm=reshape(aic,5,5);

%%
```

```
mod=arima(3,0,3);
JNJ_mod=estimate(mod,JNJ_pct);

JNJ_res= infer(JNJ_mod,JNJ_pct);
JNJ_pred = JNJ_pct-JNJ_res;
figure()
plot(JNJ_pct,'-b')
hold on
plot(JNJ_pred,'-r')

dy=JNJ_pct;
Mdl = mod; % define the model
EstMdl = JNJ_mod; % perform estimation
[dyF,dyMSE] = forecast(EstMdl,1,dy(end-10:end)); % forecast h step ahead
dyF_ci = [dyF - 1.96*sqrt(dyMSE) dyF + 1.96*sqrt(dyMSE)]
figure
plot(dy,'Color',[.7,.7,.7]);
hold on
plot(T+1,dyF,'bo','LineWidth',2);
plot(T+1,dyF + 1.96*sqrt(dyMSE),'rd',...
        'LineWidth',2);
plot(T+1,dyF - 1.96*sqrt(dyMSE),'rd','LineWidth',2);
legend('Observed','Forecast',...
        '95\% Confidence ...
           Interval','Location','Best','interpreter','latex');
title('1-Step Ahead','interpreter','latex')
grid on
set(gca,'FontSize',20)

%% ARMA order for MRNA
% We plot ACF and PACF and we observe a quite clear behaviour after ...
   the first
% lag. BIC and AIC both identify the best model in (2,2).Overall we ...
   chose the
% (2,2).


LOGL = zeros(5,5); % Initialize
PQ = zeros(5,5);

 for p = 0:4
     for q = 0:4
        Mdl = arima(p,0,q);
        [EstMdl,¬,logL] = estimate(Mdl,MRNA_pct,'Display','off');
        LOGL(p+1,q+1) = logL;
        PQ(p+1,q+1) = p + q;
     end
end

LOGL = reshape(LOGL,25,1);
PQ = reshape(PQ,25,1);
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicm=reshape(bic,5,5);
aicm=reshape(aic,5,5);

```

```
%%
mod=arima(2,0,2);
MRNA_mod=estimate(mod,MRNA_pct); %low p-values for AR

MRNA_res= infer(MRNA_mod,MRNA_pct);
MRNA_pred = MRNA_pct-MRNA_res;
figure();
plot(MRNA_pct,'-b');
hold on;
plot(MRNA_pred,'-r');

dy=MRNA_pct;
Mdl = mod; % define the model
EstMdl = MRNA_mod; % perform estimation
[dyF,dyMSE] = forecast(EstMdl,1,dy(end-10:end)); % forecast h step ahead

dyF
dyF_ci = [dyF - 1.96*sqrt(dyMSE) dyF + 1.96*sqrt(dyMSE)]

figure
plot(dy,'Color',[.7,.7,.7]);
hold on
plot(T+1,dyF,'bo','LineWidth',2);
plot(T+1,dyF + 1.96*sqrt(dyMSE),'rd',...
        'LineWidth',2);
plot(T+1,dyF - 1.96*sqrt(dyMSE),'rd','LineWidth',2);
legend('Observed','Forecast',...
        '95\% Confidence ...
           Interval','Location','Best','interpreter','latex');
title('1-Step Ahead','interpreter','latex')
grid on
set(gca,'FontSize',20)
%% ARMA order for PFE
% We plot ACF and PACF and I observe a not so clear behaviour after ...
   the first
% lag. So by exploratory analysis we would choose an ARMA(1,1).
% BIC identifies the best model in (1,1), while AIC chooses ...
   (4,4).Overall we
% chose the (1,1) model.

LOGL = zeros(4,4); % Initialize
LOGL = zeros(5,5); % Initialize
PQ = zeros(5,5);

 for p = 0:4
     for q = 0:4
        Mdl = arima(p,0,q);
        [EstMdl,¬,logL] = estimate(Mdl,PFE_pct,'Display','off');
        LOGL(p+1,q+1) = logL;
        PQ(p+1,q+1) = p + q;
     end
end

LOGL = reshape(LOGL,25,1);
PQ = reshape(PQ,25,1);
```

```
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicm=reshape(bic,5,5);
aicm=reshape(aic,5,5);

%%
mod=arima(4,0,4); %better choice than (1,1)
PFE_mod=estimate(mod,PFE_pct); %low p-values

PFE_res= infer(PFE_mod,PFE_pct);
PFE_pred = PFE_pct-PFE_res;

figure()
plot(PFE_pct,'-b')
hold on
plot(PFE_pred,'-r')

dy=PFE_pct;
Mdl = mod; % define the model
EstMdl = PFE_mod; % perform estimation
[dyF,dyMSE] = forecast(EstMdl,1,dy(end-10:end)); % forecast h step ahead

dyF
dyF_ci = [dyF - 1.96*sqrt(dyMSE) dyF + 1.96*sqrt(dyMSE)]

figure
plot(dy,'Color',[.7,.7,.7]);
hold on
plot(T+1,dyF,'bo','LineWidth',2);
plot(T+1,dyF + 1.96*sqrt(dyMSE),'rd',...
        'LineWidth',2);
plot(T+1,dyF - 1.96*sqrt(dyMSE),'rd','LineWidth',2);
legend('Observed','Forecast',...
        '95\% Confidence ...
           Interval','Location','Best','interpreter','latex');
title('1-Step Ahead','interpreter','latex')
grid on
set(gca,'FontSize',20)
%% MEAN MODEL DIAGNOSTIC for AZN

AZN_stdr = AZN_res/sqrt(AZN_mod.Variance);
figure
subplot(2,2,1)
plot(AZN_stdr)
title('Standardized Residuals')
subplot(2,2,2)
histogram(AZN_stdr,10)
title('Standardized Residuals')
subplot(2,2,3)
autocorr(AZN_stdr)
subplot(2,2,4)
parcorr(AZN_stdr)
qqplot(AZN_stdr);
[h,p] = kstest(AZN_stdr);
p
[h,pValue] = lbqtest(AZN_res);
```

```
pValue
[h,pValue]=archtest(AZN_res);
pValue
% This tests the null hypothesis of no ARCH effects against the ...
    alternative ARCH model with k lags

%% MEAN MODEL DIAGNOSTIC for JNJ

JNJ_stdr = JNJ_res/sqrt(JNJ_mod.Variance);
figure
subplot(2,2,1)
plot(JNJ_stdr)
title('Standardized Residuals')
subplot(2,2,2)
histogram(JNJ_stdr,10) %skewed
title('Standardized Residuals')
subplot(2,2,3)
autocorr(JNJ_stdr)
subplot(2,2,4)
parcorr(JNJ_stdr)
qqplot(JNJ_stdr);
[h,p] = kstest(JNJ_stdr);
p
[h,pValue] = lbqtest(JNJ_res);
pValue
[h,pValue]=archtest(JNJ_res);
pValue

%% MEAN MODEL DIAGNOSTIC for MRNA

MRNA_stdr = MRNA_res/sqrt(MRNA_mod.Variance);
figure
subplot(2,2,1)
plot(MRNA_stdr)
title('Standardized Residuals')
subplot(2,2,2)
histogram(MRNA_stdr,10)
title('Standardized Residuals')
subplot(2,2,3)
autocorr(MRNA_stdr)
subplot(2,2,4)
parcorr(MRNA_stdr)
qqplot(MRNA_stdr);
[h,p] = kstest(MRNA_stdr);
p
[h,pValue] = lbqtest(MRNA_res);
pValue
[h,pValue]=archtest(MRNA_res);
pValue
%% MEAN MODEL DIAGNOSTIC for PFE

PFE_stdr = PFE_res/sqrt(PFE_mod.Variance);
figure
subplot(2,2,1)
plot(PFE_stdr)
```

```
526 title('Standardized Residuals')
527 subplot(2,2,2)
528 histogram(PFE_stdr,10)
529 title('Standardized Residuals')
530 subplot(2,2,3)
531 autocorr(PFE_stdr)
532 subplot(2,2,4)
533 parcorr(PFE_stdr)
534 qqplot(PFE_stdr);
535 [h,p] = kstest(PFE_stdr);
536 p
537 [h,pValue] = lbqtest(PFE_res);
538 pValue
539 [h,pValue]=archtest(PFE_res);
540 pValue
541
542 %% PLOTS res sq
543
544 figure();
545 subplot(2,2,1);
546 autocorr(AZN_res.^2);
547 title('AstraZeneca');
548 subplot(2,2,2);
549 autocorr(JNJ_res.^2);
550 title('Johnson & Johnson');
551 subplot(2,2,3);
552 autocorr(MRNA_res.^2);
553 title('Moderna');
554 subplot(2,2,4);
555 autocorr(PFE_res.^2);
556 title('Pfizer');
557
558 figure();
559 subplot(2,2,1);
560 parcorr(AZN_res.^2);
561 title('AstraZeneca');
562 subplot(2,2,2);
563 parcorr(JNJ_res.^2);
564 title('Johnson & Johnson');
565 subplot(2,2,3);
566 parcorr(MRNA_res.^2);
567 title('Moderna');
568 subplot(2,2,4);
569 parcorr(PFE_res.^2);
570 title('Pfizer');
571
572
573 %% GARCH model for AZN
574 % We observed the ARCH effect test in 3 cases has very high p-value, ...
      however
575 % we still chose to model the conditional volatility.
576
577
578 LOGL = zeros(5,4);
579 PQ = zeros(5,4);
```

```
for p = 0:4
    for q = 1:4
        Mdl3 = arima(1,0,0);
        CVarMdl3 = garch(p,q);
        Mdl3.Variance = CVarMdl3;
        [EstMdl,¬,logL] = estimate(Mdl3,AZN_pct,'Display','off');
        LOGL(p+1,q) = logL;
        PQ(p+1,q) = p + q;
     end
end

LOGL = reshape(LOGL,20,1);
PQ = reshape(PQ,20,1);
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicm=reshape(bic,5,4);
aicm=reshape(aic,5,4);

%%

dy=AZN_pct;
Mdl3 = arima(1,0,0);
CVarMdl3 = garch(0,2);
Mdl3.Variance = CVarMdl3;
[GARCH_A,¬,logL] = estimate(Mdl3,dy);

%% GARCH model for JNJ

figure
subplot(1,2,1)
autocorr(JNJ_res.^2);
subplot(1,2,2)
parcorr(JNJ_res.^2);


LOGL = zeros(4,4);
PQ = zeros(4,4);
for p = 1:4
    for q = 1:4
        Mdl3 = arima(3,0,3);
        CVarMdl3 = garch(p,q);
        Mdl3.Variance = CVarMdl3;
        [EstMdl,¬,logL] = estimate(Mdl3,JNJ_pct,'Display','off');
        LOGL(p,q) = logL;
        PQ(p,q) = p + q;
     end
end

LOGL = reshape(LOGL,16,1);
PQ = reshape(PQ,16,1);
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicm=reshape(bic,4,4);
aicm=reshape(aic,4,4);
%%


```

```
dy=JNJ_pct;
Mdl3 = arima(3,0,3);
CVarMdl3 = garch(2,1);
Mdl3.Variance = CVarMdl3;
[GARCH_J,¬,logL] = estimate(Mdl3,dy);

%% GARCH model for MRNA

subplot(1,2,1)
autocorr(MRNA_res.^2);
subplot(1,2,2)
parcorr(MRNA_res.^2);


LOGL = zeros(5,4);
PQ = zeros(5,4);
for p = 0:4
    for q = 1:4
        Mdl3 = arima(2,0,2);
        CVarMdl3 = garch(p,q);
        Mdl3.Variance = CVarMdl3;
        [EstMdl,¬,logL] = estimate(Mdl3,MRNA_pct,'Display','off');
        LOGL(p+1,q) = logL;
        PQ(p+1,q) = p + q;
     end
end

LOGL = reshape(LOGL,16+4,1);
PQ = reshape(PQ,16+4,1);
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicm=reshape(bic,4+1,4);
aicm=reshape(aic,4+1,4);
%%
dy=MRNA_pct;
Mdl3 = arima(2,0,2);
CVarMdl3 = garch(0,2);
Mdl3.Variance = CVarMdl3;
GARCH_M = estimate(Mdl3,dy);


%% GARCH model for PFE

subplot(1,2,1)
autocorr(PFE_res.^2);
subplot(1,2,2)
parcorr(PFE_res.^2);

dimp=9;
dimq=9;
LOGL = zeros(dimp,dimq);
PQ = zeros(dimp,dimq);
for p = 0:dimp-1
    for q = 1:dimq
        display(p);
        display(q);
```

```
        Mdl3 = arima(4,0,4);
        CVarMdl3 = garch(p,q);
        Mdl3.Variance = CVarMdl3;
        [EstMdl,¬,logL] = estimate(Mdl3,PFE_pct,'Display','off');
        LOGL(p+1,q) = logL;
        PQ(p+1,q) = p + q;
    end
end

LOGL = reshape(LOGL,dimp*dimq,1);
PQ = reshape(PQ,dimp*dimq,1);
[aic,bic] = aicbic(LOGL,PQ+1,100);
bicmx=reshape(bic,dimp,dimq);
aicmx=reshape(aic,dimp,dimq);
%%

dy=PFE_pct;
Mdl3 = arima(4,0,4);
CVarMdl3 = garch(0,9);
Mdl3.Variance = CVarMdl3;
GARCH_P = estimate(Mdl3,dy);


%% EGARCH

Mdl3 = arima(1,0,0);
CVarMdl3 = egarch(1,1);
Mdl3.Variance = CVarMdl3;
[EGARCH_A,¬,logL] = estimate(Mdl3,AZN_pct);


Mdl3 = arima(3,0,3);
CVarMdl3 = egarch(1,1); %% bic 3,2 aic 4,2
Mdl3.Variance = CVarMdl3;
[EGARCH_J,¬,logL] = estimate(Mdl3,JNJ_pct);


Mdl3 = arima(2,0,2);
CVarMdl3 = egarch(1,1); %%best for aic and bic
Mdl3.Variance = CVarMdl3;
[EGARCH_M,¬,logL] = estimate(Mdl3,MRNA_pct);


Mdl3 = arima(4,0,4);
CVarMdl3 = egarch(1,1);
Mdl3.Variance = CVarMdl3;
[EGARCH_P,¬,logL] = estimate(Mdl3,PFE_pct);


%% FORECASTING

%Load forecast data
warning('OFF', 'MATLAB:table:ModifiedAndSavedVarnames');
AZNf = readtable('AZN_future.csv');
AZNf = AZNf(:,5);
```

```
AZNf = table2array(AZNf);
JNJf = readtable('JNJ_future.csv');
JNJf = JNJf(:,5);
JNJf = table2array(JNJf);
MRNAf = readtable('MRNA_future.csv');
MRNAf = MRNAf(:,5);
MRNAf = table2array(MRNAf);
PFEf = readtable('PFE_future.csv');
PFEf = PFEf(:,5);
PFEf = table2array(PFEf);


%Go to logs

AZNf=log(AZNf);
laggedAZNf = lagmatrix(AZNf,1);
diff=AZNf-laggedAZNf;
AZNf_pct=diff;
AZNf_pct(1,:)=[];

JNJf=log(JNJf);
laggedJNJ = lagmatrix(JNJf,1);
diff=JNJf-laggedJNJ;
JNJf_pct=diff;
JNJf_pct(1,:)=[];

MRNAf=log(MRNAf);
laggedMRNAf = lagmatrix(MRNAf,1);
diff=MRNAf-laggedMRNAf;
MRNAf_pct=diff;
MRNAf_pct(1,:)=[];

PFEf=log(PFEf);
laggedPFEf = lagmatrix(PFEf,1);
diff=PFEf-laggedPFEf;
PFEf_pct=diff;
PFEf_pct(1,:)=[];

%Forecasting GARCH vs EGARCH

%A
[dy_AG,dyMSE_AG] = forecast(GARCH_A,10,AZNf_pct(1:10));
s=dy_AG-AZNf_pct(1:10);
MAE_AG=sum(abs(s))./10;
MSE_AG=sum(s.^2)./10;

[dy_AE,dyMSE_AE] = forecast(EGARCH_A,10,AZNf_pct(1:10));
s=dy_AE-AZNf_pct(1:10);
MAE_AE=sum(abs(s))./10;
MSE_AE=sum(s.^2)./10;

%J

[dy_JG,dyMSE_JG] = forecast(GARCH_J,10,JNJf_pct(1:10));
s=dy_JG-JNJf_pct(1:10);
```

```
MAE_JG=sum(abs(s))./10;
MSE_JG=sum(s.^2)./10;

[dy_JE,dyMSE_JE] = forecast(EGARCH_J,10,JNJf_pct(1:10));
s=dy_JE-JNJf_pct(1:10);
MAE_JE=sum(abs(s))./10;
MSE_JE=sum(s.^2)./10;

%M
[dy_MG,dyMSE_MG] = forecast(GARCH_M,10,MRNAf_pct(1:10));
s=dy_MG-MRNAf_pct(1:10);
MAE_MG=sum(abs(s))./10;
MSE_MG=sum(s.^2)./10;

[dy_ME,dyMSE_ME] = forecast(EGARCH_M,10,MRNAf_pct(1:10));
s=dy_ME-MRNAf_pct(1:10);
MAE_ME=sum(abs(s))./10;
MSE_ME=sum(s.^2)./10;

%P
[dy_PG,dyMSE_PG] = forecast(GARCH_P,10,PFEf_pct(1:10));
s=dy_PG-PFEf_pct(1:10);
MAE_PG=sum(abs(s))./10;
MSE_PG=sum(s.^2)./10;

[dy_PE,dyMSE_PE] = forecast(EGARCH_P,10,PFEf_pct(1:10));
s=dy_PE-PFEf_pct(1:10);
MAE_PE=sum(abs(s))./10;
MSE_PE=sum(s.^2)./10;



%% correlation
cmap = copper(6);

%subplot(3,2,1);

corr(AZN_pct,JNJ_pct);
[se1,CI] = bootstrap_R2(AZN_pct,JNJ_pct);
%h=hist(se1);
title('AZN_JNJ')
CI

%subplot(3,2,2);

corr(AZN_pct,MRNA_pct);
[se2,CI] = bootstrap_R2(AZN_pct,MRNA_pct);
%hist(se2);
title('AZN_MRNA')

CI

%subplot(3,2,3);

corr(AZN_pct,PFE_pct);
```

```
[se3,CI] = bootstrap_R2(AZN_pct,PFE_pct);
%hist(se3);
%title('AZN_PFE')

CI

%subplot(3,2,4);

corr(JNJ_pct,PFE_pct);
[se4,CI] = bootstrap_R2(JNJ_pct,PFE_pct);
%hist(se4);
%title('PFE_JNJ')

CI


%subplot(3,2,5);

corr(JNJ_pct,MRNA_pct);
[se5,CI] = bootstrap_R2(JNJ_pct,MRNA_pct);
%hist(se5);
CI
title('MRNA_JNJ')



%subplot(3,2,6);

corr(PFE_pct,MRNA_pct);
[se6,CI] = bootstrap_R2(PFE_pct,MRNA_pct);
%hist(se6);
title('PFE_MRNA')

CI

%%
figure();
g6(1,1)=gramm('x',se1);
g6(1,1).stat_bin('fill','face');
g6(1,1).set_title('''face''');
g6(1,1).set_title('AZN_JNJ');

g6(1,2)=gramm('x',se2);
g6(1,2).stat_bin('fill','face');
g6(1,2).set_title('''face''');
g6(1,2).set_title('AZN_MRNA');

g6(2,1)=gramm('x',se3);
g6(2,1).stat_bin('fill','face');
g6(2,1).set_title('''face''');
g6(2,1).set_title('AZN_PFE');

g6(2,2)=gramm('x',se4);
g6(2,2).stat_bin('fill','face');
g6(2,2).set_title('''face''');
```

```
910 g6(2,2).set_title('PFE_JNJ');
911
912 g6(3,1)=gramm('x',se5);
913 g6(3,1).stat_bin('fill','face');
914 g6(3,1).set_title('''face''');
915 g6(3,1).set_title('MRNA_JNJ');
916
917 g6(3,2)=gramm('x',se6);
918 g6(3,2).stat_bin('fill','face');
919 g6(3,2).set_title('''face''');
920 g6(3,2).set_title('PFE_MRNA');
921
922 g6(1,1).set_color_options('map','d3_10');
923 g6(1,2).set_color_options('map','brewer3');
924 g6(2,1).set_color_options('map','winter');
925 g6(2,2).set_color_options('map','brewer_pastel');
926 g6(3,1).set_color_options('map','brewer2');
927 g6(3,2).set_color_options('map','brewer1');
928
929
930 g6.draw();
```

A.2 bootstrap$_r2.m$

```
1  function [se,CI] = bootstrap_R2(v1,v2)
2
3  %INPUT
4  % v1,v2 = vectors of equal length
5  %
6  %OUTPUT
7  %se=vector of correlations obtained via bootstrap
8  %CI= 95\% confidence interval for correlation
9
10 m=fitlm(v2,v1);
11 fitted=m.Fitted;
12 res=table2array(m.Residuals(:,1));
13 R2=m.Rsquared.Ordinary;
14
15 se = (bootstrp(10000,@(bootr)corr(fitted+bootr,v2),res));
16 m=mean(se);
17 std_sol=std(se);
18 CI=[m-1.96*std_sol m m+1.96*std_sol];
19
20 end
```