Modelling the stock returns and volatility of Equinor using ARIMA and GARCH

candidate nr. 170609

May 2023

Abstract

This paper focuses on investigating the predictive accuracy of Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models in forecasting future returns and volatility of Equinor. Equinor is a predominant company listed on the Euronext Growth market and majority-owned by the Norwegian government. The study aims to assess the models' performance in capturing the underlying dynamics of the company's stock returns and examine their potential application for investment and risk management purposes. The paper provides a procedure, influenced by the Box-Jenkins approach, that proposes fitting ARIMA and GARCH models and then evaluates the models´ prediction accuracy.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Research question

The objective of this research paper is to evaluate the performance of Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models in predicting future stock returns and volatility for Equinor. Accurate forecasts of stock returns and volatility can be essential for investors, financial analysts, and portfolio managers to make more informed decisions about investments and risk management. Equinor is the largest company on the Euronext Growth market and is majority-owned by the Norwegian government (Equinor.com). Equinor's performance has implications for the Norwegian economy, the government's revenues, and the overall performance of the Norwegian stock market due to its size. Understanding the behavior of its stock price and volatility, therefore, becomes even more financially significant. Accurate stock price and volatility predictions for Equinor can help stakeholders develop better trading strategies, improve portfolio performance, and manage risk more effectively.

1.2 Litterateur review

Box and Jenkins, 1970, introduced the ARIMA model, which combines autoregressive (AR), moving average (MA), and differencing components to account for different patterns in time series data. ARIMA models have been applied to various financial time series, such as stock prices, exchange rates, and interest rates. One notable study that applied ARIMA models in the field of finance is by Tsay, 2005. Tsay demonstrated the utility of ARIMA models for predicting stock returns. Furthermore, Gooijer and Hyndman, 2006, provided a comprehensive review of time series forecasting methods, including ARIMA, and their applications in finance and economics. Another study by Adhikari and Agrawal, 2013, applied ARIMA models to forecast stock prices in the Indian market and demonstrated their ability in predicting price movements.

GARCH models, introduced by Bollerslev, 1986, are an extension of the ARCH model originally proposed by Engle, 1982. GARCH models address the shortcomings of ARIMA models in capturing the volatility clustering that can often be observed in financial time series. These models have been used in finance for modeling and forecasting financial market volatility. One influential study by Bollerslev et al., 1992, applied GARCH models to model and forecast exchange rate volatility. Additionally, Engle and Patton, 2001, provided a comprehensive review of volatility forecasting techniques, including GARCH models, and their applications in finance. Nelson, 1991, introduced the Exponential GARCH (EGARCH) model, which has since been used in various studies for modeling asymmetric volatility in financial time series, such as the work by GLOSTEN et al., 1993 and Ding et al., 1993.

The combination of ARIMA and GARCH models, often referred to as ARIMA-GARCH models, has become a popular approach for analyzing financial time series. This combined model can capture both the conditional mean and volatility dynamics in the data. One notable study employing this approach is by Poon and Granger, 2003, which investigated the forecasting performance of ARIMA-GARCH models on stock market returns. Their findings suggested that ARIMA-GARCH models outperform other linear models in predicting stock market volatility.

Although these models have been applied to numerous stocks and financial instruments, there has been limited research specifically focused on the Equinor stock. This study contributes to the existing literature by applying ARIMA and GARCH models to the Equinor stock, analyzing their forecasting performance, and provides insights into whether they can be used as a reliable tool for accurate predictions.

1.3 Data and methods

The dataset used consists of daily closing prices of Equinor stock obtained from Yahoo finance. The dataset covers a time frame spanning over to decades. The study employs the Box-Jenkins methodology to fit ARIMA and GARCH models to the historical stock returns of Equinor. Model selection will be based on information criteria. The models' performance will be evaluated using various error metrics and accuracy measures such as Log-likelihood, Mean Absolute Error and Root Mean Square Error. Once the models have been developed, the residuals of the fitted models will be analyzed. The models will also be used to generate out-of-sample forecasts. The predicted values from the out-of-sample forecast will then be compared to the actual values from the test set to assess the forecasting performance of the models on unseen data.

1.4 Summary and conclusion

This study's findings highlight the limitations of the ARIMA model in producing accurate forecasts for Equinor stock returns, while acknowledging the GARCH model's ability to more effectively capture the stock's volatility dynamics. This research contributes to existing literature on ARIMA and GARCH modelling of financial time series and provides valuable insights for investors and financial analysts regarding Equinor. Incorporating these models into a broader trading strategy could offer insights into portfolio diversification and risk management. However, alternative GARCH variants or machine learning-based approaches may yield better results. Additionally, since Equinor is an oil company, an Autoregressive Distributed Lag (ADL) model that takes oil prices into account might be more appropriate. Future research could investigate other GARCH variants, machine learning models, or ADL models to further enhance forecasting Equinors stock returns or volatility.

2 Data

2.1 Description of data

The data used for this study is sourced from Yahoo Finance. The data consists of the adjusted closing prices of the Equinor stock, because it offers the most accurate reflection of the stock's value over time. The adjusted closing prices account for corporate actions such as stock splits, dividends, and new stock offerings. This allows for a consistent and accurate comparison of the stocks performance throughout time. The data set has a daily frequency, meaning approximately 252 data points per year on average. This frequency provides a detailed view of the stock's behavior and captures short-term fluctuations while still allowing for the analysis of long-term trends. The daily frequency also offers a larger sample size, which improves the statistical robustness of the results. The time range of the data spans from the date when Equinor's stock was first listed on the stock exchange, June 18th, 2001 until May 4th, 2023. This comprehensive time range enables the examination of the stock's entire trading history, providing insights into its historical performance, trends, and any structural changes that may have occurred over the years.

Figure 1: Equinor stock price from 2001-06-18 to 2023-05-04

2.2 Examination and log transforming

Upon visually examining the Equinor stock price, the general trend appears to be upward, indicating a positive long-term performance. However, there are several noticeable dips throughout, suggesting that the stock has experienced periods of decline and increased volatility. It does not appear to be any structural breaks as the stock rises to previous levels after having fallen and then continue the upward trend. There does however seem to be presence of a possible seasonal component, which could imply that there are recurring fluctuations tied to specific times of the year. This might be driven by factors such as industry cycles, seasonal demand, or other oil related influences.

Taking the natural logarithm of the stock prices helps to stabilize the variance, and makes the data more appropriate for analysis. Taking the logarithm effectively compresses the higher values of the series and expand the lower values. This has the effect reducing the variability of the data (Tsay, 2005). The log-transformed series becomes more stable across, but still seem to be non-stationary. The log transformed data is shown in figure 2.

2.3 Stationary testing and differenceing

To evaluate if the log-transformed series is stationary, the Augmented Dickey-Fuller (ADF) test is employed. ADF test is used to test for the presence of a unit root. The null hypothesis (H0) of the ADF test is that the time series has a unit root, indicating it is non-stationary. The alternative hypothesis (H1) is that the time series does not have

Figure 2: Equinor stock price log-transformed from 2001-06-18 to 2023-05-04

a unit root, suggesting it is stationary. The ADF test computes a t-test statistic for the estimated coefficient in an autoregressive model and its corresponding p-value (Enders, 2014).

	TestStatistic	PValue	Lags
ADF Test	-2.44	0.39	17.00

Table 1: ADF Test Results

Based on the results of the ADF test, we obtain a p-value of 0.39. Since this value exceeds the 0.05 significance threshold for a 95% confidence level, we cannot reject the null hypothesis. The data is therefore differnced. Differencing is a technique that involves subtracting the previous observation from the current observation, thereby creating a new series of differences. This process helps remove trends and other non-stationary components (Enders, 2014).

	TestStatistic	PValue	Lags
ADF Test	-18.92	0.01	17.00

Table 2: ADF Test Results 1. Difference

The results from the Augmented Dickey-Fuller (ADF) test on the differenced data gives a p-value of 0.01 and we can reject the null hypothesis. The time series should now be stationary.

After differencing the log-transformed data, the time series now represents the stock returns, as shown in Figure 3. The transformation is as follows:

$$r_t \approx \Delta y_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1})$$

Here, r_t is the continuously compounded return, and $\frac{P_t}{P_{t-1}}$ represents the gross return on the stock from time $t-1$ to time t.

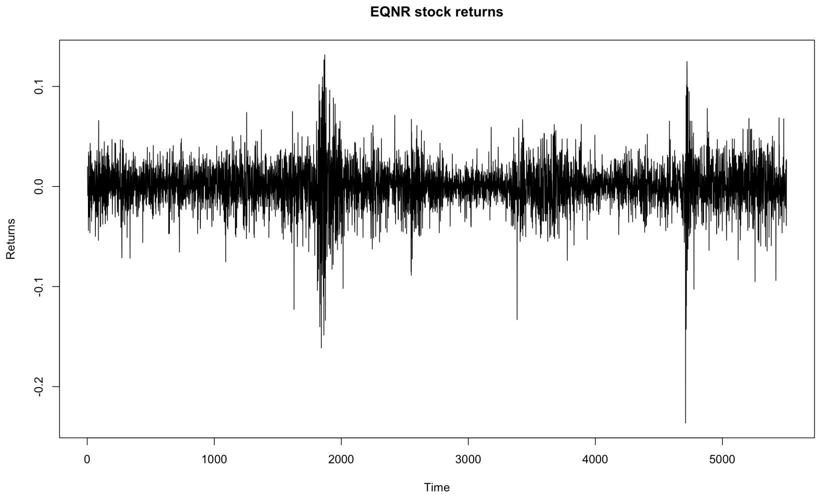

Figure 3: Equinor stock returns 2001-06-18 to 2023-05-04

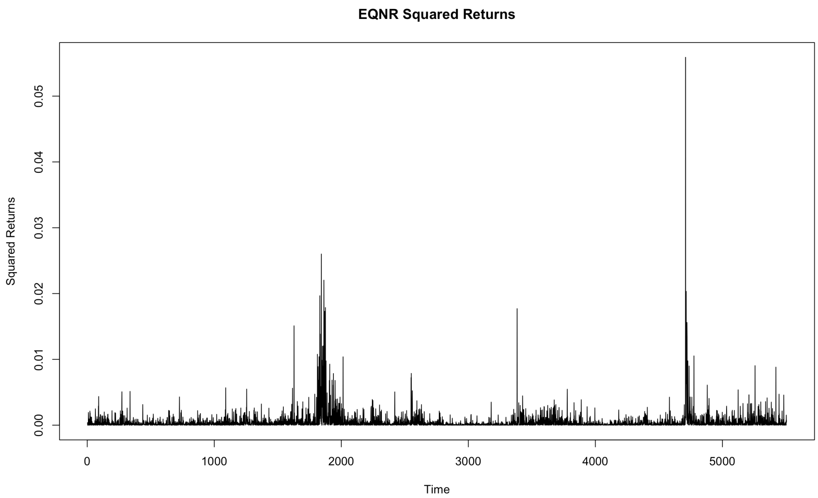

Figure 4: Equinor stock returns squared 2001-06-18 to 2023-05-04

Figure 3 do seem to showcase volatility clustering. Volatility clustering is a common characteristic in financial time series data and refers to the phenomenon where periods of high volatility tend to be followed by more high volatility and similarly, low volatility periods are followed by more low volatility (Enders, 2014). Squaring the returns highlights this phenomenon, as shown in figure 4.

3 Methodology

This study uses the Box-Jenkins approach as the primary methodology for fitting the ARIMA and GARCH models to the Equinor stock returns. The Box-Jenkins approach is widely used in time series analysis and involve three main stages: identification, estimation, and diagnostic checking (Enders, 2014).

3.1 ARIMA model

ARIMA models are a popular choice for time series forecasting due to their ability to handle various types of time series data, such as those with trends and seasonality. ARIMA models can efficiently account for past values (autoregressive component), past forecast errors (moving average component), and non-stationary data by differencing (integrated component) (Enders, 2014).

The autoregressive integrated moving-average (ARIMA) models join the concepts of AR and MA models in the following matter:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{1}$$

where y_t is the observed time series at time t, c is a constant, ϕ_i and θ_j are the autoregressive and moving average coefficients, respectively, and ϵ_t is the error term at time t.

3.2 GARCH model

GARCH models are used for modeling volatility in time series due to their ability to capture characteristics of such volatility clustering. GARCH models effectively capture this by allowing the conditional variance to depend on both past conditional variances (GARCH term) and past errors (ARCH term). In contrast, ARIMA models focus on modeling the mean of a time series, not its volatility. By accounting for time-varying volatility and the dynamics of the conditional variance, GARCH models generally provide more accurate forecasts of financial time series than ARIMA models (Enders, 2014). The GARCH(p, q) model is given as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{2}$$

where ϵ_t is the error term at time t, σ_t^2 is the conditional variance of the error term at time t, ω is a constant term, α_i and β_j are the autoregressive and moving average coefficients, respectively.

3.3 ACF and PACF analysis

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) display the relationships between observations at various lags. These tools will be utilized to determine the suitable AR and moving and MA orders for the ARIMA model. Additionally, the ACF and PACF plots for the squared returns are examined to better understand the presence of heteroskedasticity in the volatility.

3.4 ARIMA and GARCH model selection

To determine the most suitable ARIMA and GARCH models, a loop in R that iterates through various combinations of model parameters and selects the models with the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, is implemented. Both AIC and BIC are designed to balance the goodness of fit of a model with its complexity, where a lower value indicates a better trade-off. The AIC and BIC is defined as:

$$AIC = 2k - 2\ln(L)$$

$$BIC = k\ln(n) - 2\ln(L)$$

where k is the number of estimated parameters in the model, L is the maximum value of the likelihood function for the model, and n is the number of observations in the dataset. BIC imposes a stronger penalty on model complexity, which makes it more suitable given the large sample size of the data (Enders, 2014). The ARIMA model will be tested against both information criteria, while the GARCH model is only tested against BIC. This comprehensive approach to model selection should ensure that the most appropriate ARIMA and GARCH models are chosen, providing the best overall fit to the data as measured by the respective information criteria.

3.5 Diagnostic checks

3.5.1 Testing for autocorrelation in the resiudals

The Ljung Box test is used to test for autocorrelation in the residuals. The residuals of a well-specified ARIMA time series model and the standarized squared residuals of a well-specified GARCH model should be independently and identically distributed (i.i.d) with no autocorrelation. The null hypothesis (H0) is that there is no autocorrelation in the residuals, while the alternative hypothesis (H1) is that autocorrelation is present in the residuals. A p-value below 0.05 suggests that the model has not adequately captured the underlying dynamics of the series, as there is still significant autocorrelation present in the residuals. For the GARCH model, the Ljung-Box test is conducted on the squared standardized residuals. Absence of significant autocorrelations in the squared standardized residuals implies that the GARCH model has successfully captured the conditional heteroskedasticity in the data (Enders, 2014). For financial time series with daily data, researchers have commonly used lags from 5 (one trading week) and up to 20, as this covers roughly one trading month of daily observations (Wei, 2005). To ensure robust results, the Ljung-Box test is applied for multiple lag lengths, ranging from 5 and up to 40. If the p-values are consistently above 0.05, it would indicate that the model has successfully captured the underlying dynamics of the time series. Additionally, the ARCH-LM test is employed to investigate the presence of ARCH effects in the ARIMA model. This test is specifically designed to identify ARCH patterns in time series data and is based on the principle that the squared residuals from an autoregressive model should exhibit no significant autocorrelation under the null hypothesis of no ARCH effect (Enders, 2014).

3.5.2 Out-of-sample forecast

To assess the model's predictive accuracy on unseen data and help identify potential overfitting, an out-of-sample forecast is conducted. It involves generating predictions for a time period not used during the model training process by splitting the data into a training and testing set. The model is estimated using the training set, and its predictions are compared against the actual values from the test set. In this context, the ARMA model is employed to forecast the stock returns, while the GARCH model forecasts the volatility. Visual plots as well as error metrics will be used to asses the models out of sample prediction accuracy.

4 Results

4.1 ACF and PACF

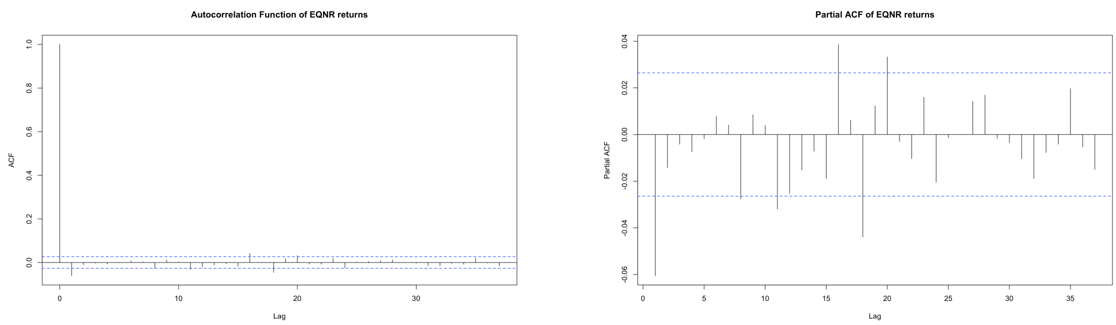

Figure 5: ACF and PACF of the returns

The ACF shows a significant spike at lag 1 followed by a quick decay, indicating that the series may have a significant moving average component. The PACF shows more of an oscillating pattern, suggesting the presence of a more complex structure. Given the observed characteristics of the ACF and PACF plots, an ARMA(0, 1) model, which captures the moving average aspect of the time series, is considered as an initial choice. However, an ARMA(1, 1) model, which incorporates both autoregressive and moving average components, will also be tested as it could account for the more complex structure suggested by the PACF plot.

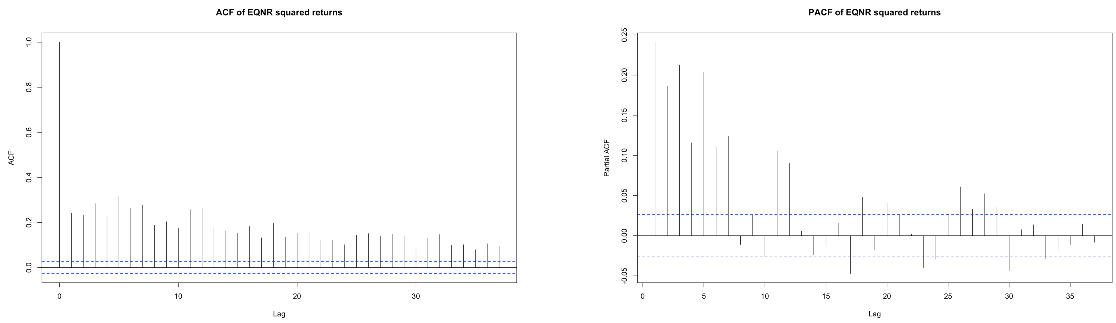

Figure 6: ACF and PACF plot squared returns

The ACF and PACF for the squared returns seem to confirm the presence of volatility clustering by showcasing a strong autocorrelation structure in the volatility of the returns. Several significant spikes for multible lags can be observed. The presence of this patterns suggests that the ARIMA model may not be sufficient in capturing the underlying dynamics of the series with regards to volatility. This indicates that the GARCH model are likely to provide more accurate forecasts.

4.2 ARIMA models

4.2.1 ARIMA(0,0,1)

The best fitted ARIMA model given both of the Information Criteria (BIC & AIC) is the ARIMA(0,0,1):

$$y_t = 0.0004 - 0.0621\varepsilon_{t-1} + \varepsilon_t \tag{3}$$

where y_t is the EQNR return at time t, ε_t is the error term at time t, ε_{t-1} is the lagged error term.

Best Fitted ARIMA Model			
Model	ARIMA(0,0,1)		
Parameter	**Estimate**	**Standard Error**	**P-Value**
MA1	-0.0621	0.0136	6.62e-06
Intercept	0.0004	0.0003	0.133
Goodness-of-Fit Measures			
σ^2	0.0005		
Log Likelihood	13099.72		
AIC	-26193.43		
AICc	-26193.43		
BIC	-26173.59		
Error Metrics			
ME	-2.780922e-07		
RMSE	0.02239372		
MAE	0.01590641		
MASE	0.6883269		
ACF1	0.0006734087		

Table 3: Best fitted ARIMA model and its summary statistics

The ARIMA(0,0,1) model has an estimated MA1 coefficient of -0.0621 which statistically significant. The negative coefficient implies that when the previous period's error term was positive (actual return was higher than the predicted return), the current period's return is likely to be lower than the predicted return, and vice versa. The small magnitude of the coefficient implies this relationship is weak.

The estimated intercept value is close to zero and not statistically significant. This can be seen as the average return across the entire period. Finding intercept values near zero is a typical occurrence in financial time series, as suggested by the Random Walk Hypothesis (Enders, 2014). These findings are consistent with the existing literature on ARIMA models applied to financial time series. Several studies have reported small coefficients for their ARIMA models, which is attributed to the inherent difficulty in predicting financial market movements due to their complex and dynamic nature.

4.2.2 ARIMA(1,0,1)

For the other model suggested we get these summary statistics:

ARIMA(1,0,1) Model			
Model	ARIMA(1,0,1)		
Parameter	**Estimate**	**Standard Error**	**P-value**
AR1	0.1896	0.2230	0.3953
MA1	-0.2506	0.2203	0.2553
Intercept	0.0004	0.0003	0.133
Goodness-of-Fit Measures			
σ^2	0.0005		
Log Likelihood	13100.09		
AIC	-26192.17		
AICc	-26192.17		
BIC	-26165.72		
Error Metrics			
ME	-6.722003e-07		
RMSE	0.02239222		
MAE	0.01590482		
MASE	0.688258		
ACF1	1.64247e-05		

Table 4: ARIMA(1,0,1) model and its summary statistics

The ARIMA(1,0,1) model has an estimated AR1 coefficient of 0.1810 and an estimated MA1 coefficient of -0.2425. Neither of the coefficients are statistically significant. The positive AR1 suggests that when the previous period's return was positive, the current period's return is likely to be higher, and vice versa.

The goodness-of-fit measures are quite similar for both models, with the ARIMA(0,0,1) model having marginally lower AIC and BIC values, suggesting a slightly better model fit. The error metrics however are quite close for both models, indicating similar performance in terms of prediction accuracy.

4.3 GARCH model

The best fitted GARCH model, with the ARMA(0,1) mean equation, is the ARMA(0,1)-GARCH(1,1) model. The mathematical expression for this model is as follows:

Mean Equation (ARMA(0,1)):

$$y_t = 7.2358 \times 10^{-4} + \varepsilon_t - 2.0900 \times 10^{-2} \cdot \varepsilon_{t-1} \tag{4}$$

Variance Equation (GARCH(1,1)):

$$\sigma_t^2 = 3.7672 \times 10^{-6} + 5.3342 \times 10^{-2} \cdot \varepsilon_{t-1}^2 + 9.3954 \times 10^{-1} \cdot \sigma_{t-1}^2 \tag{5}$$

Where:

- y_t represents the return at time t

- ε_t is the error term (residual) at time t, with $\varepsilon_t = y_t - \mu$
- σ_t^2 is the conditional variance at time t

The GARCH(1,1) model demonstrates a good fit to the data, as shown by the log likelihood value of 13794.38. The model's coefficients exhibit statistical significance. However, the 'MA1' coefficient is not statistically significant at the 5% level, implying that the MA(1) term may not be crucial in explaining the mean return and will therefore be excluded in the out-of-sample forecast. The positive alpha1 coefficient suggests that the current period's volatility is influenced by the previous period's shocks or errors, while the positive beta1 coefficient implies that the current period's volatility is influenced by the previous period's volatility.The presence of both positive alpha1 and beta1 coefficients indicates that the series exhibits volatility clustering. This observation aligns with the existing literature on financial time series analysis (Enders, 2014)

(Enders, 2014).

GARCH(1,1) Model Output				
Coefficient	**Estimate**	**Std. Error**	**t-value**	**Pr($>|t|$)**
μ	7.236e-04	2.362e-04	3.064	0.00219**
MA1	-2.090e-02	1.437e-02	-1.454	0.14582
ω	3.767e-06	8.328e-07	4.523	6.09e-06***
α_1	5.334e-02	4.895e-03	10.897	$<$ 2e-16***
β_1	9.395e-01	5.448e-03	172.467	$<$ 2e-16***

Table 5: GARCH(1,1) model output and its summary statistics

5 Diagnostics checks and model evaluation

5.1 Residual analysis

The ARCH-LM test on the ARIMA model resulted in a p-value significantly smaller than the significance thresholds of 0.05 suggesting the presence of ARCH effects within the time series.

Table 6: ARCH LM-test Result

Statistic	Df	P-value
1197.9	12	2.2e-16

Table 7: Ljung-Box Test Results

ARIMA(0,0,1) Residuals		**ARIMA(1,0,1) Residuals**		**GARCH Sqrd Std Residuals**	
Lag	P-value	Lag	P-value	Lag	P-value
5	0.913006	5	0.971225	5	0.3587447
10	0.7340367	10	0.7311282	10	0.07753371
15	0.2086301	15	0.1851786	15	0.2416659
20	0.0005125716	20	0.0004919973	20	0.3972739
25	0.00106637	25	0.001054196	25	0.4830115
30	0.004323005	30	0.004414367	30	0.6023088
35	0.005512954	35	0.005689448	35	0.7446565
40	0.0002394122	40	0.0002579035	40	0.7690355

Reagridng the Ljung-Box test, for shorter lag lengths (5, 10, and 15), the p-values are quite high for the ARIMA models, indicating that the model has successfully accounted for temporal dependence in the data up to these lags, as there is no significant autocorrelation in the residuals. However, for higher lag lengths (20, 25, 30, 35, and 40), the p-values are substantially lower, suggesting the presence of significant autocorrelation in the residuals. This implies that the ARIMA models might not have adequately captured all the underlying dynamics of the time series at higher lag lengths. This means its accuracy and reliability for longer-term forecasts might be limited compared to its performance in the short term. The test results for the Ljung-Box test on the squared standardized residuals of the GARCH model, shows that the model has effectively captured the non-linear dependencies and volatility clustering in the data. More residuals analysis on both of the ARIMA models is included in the appendix.

5.2 Out-of-sample forecasts

In the out-of-sample forecast for both of the ARIMA models (figure 7), we observe that the predicted values differ substantially from the actual values of the test set. The forecasts values oscillates around zero in a relatively small magnitude, whereas the actual values exhibit more pronounced fluctuations around zero. This discrepancy indicates that the ARIMA models do not accurately capture the underlying patterns of the time series. The models' inability to track the actual values' movements could be attributed to the inherently complex and unpredictable nature of financial time series data, which often exhibit non-linearities, fat-tailed distributions, and volatility clustering (Enders, 2014). These observation highlights the limitations of the ARIMA model in forecasting financial time series and why alternative models, such as GARCH produce more accurate forecasts. The reported root mean

Model	RMSE	MAE
ARIMA(0,0,1)	0.02331371	0.01776268
ARIMA(1,0,1)	0.02330221	0.01774196

Table 8: RMSE and MAE for ARIMA(0,0,1) and ARIMA(1,0,1) forecasts

squared error (RMSE) and mean absolute error (MAE) values for both models have similar performances, with ARIMA(1,01) being slightly better than ARIMA(0,0,1) in terms of forecasting accuracy. However the differences in RMSE and MAE values between the two models are small which implies that the improvement in forecasting performance from the ARIMA(1,0,1) to the ARIMA(0,0,1) should not be significant enough to impact decision-making

The out-of-sample forecasts for the GARCH model reveals a more satisfactory result (figure 8). The predicted and actual squared residuals shows a somewhat decent alignment, meaning the model is able to predict volatility patterns. Additionally, the model seems to capture both high and low volatility periods, as evidenced by the forecasts during both calm and turbulent conditions. The RMSE and MAE values is also substantially lower compared to the ARIMA models. However, the largest volatility peaks of the test set are considerably larger than the predictions of the model. It suggests that the model may be underestimating extreme events or may not be capturing all of the underlying dynamics in the data. It could indicate that the model may need further refinement or adjustments to better capture the extreme events in the data. Overall, the GARCH model's out-of-sample performance suggests that it could be a somewhat reliable tool for predicting volatility in financial time series data.

	RMSE	MAE
GARCH(1,1) Out-of-Sample	0.0009400939	0.0005364151

Table 9: RMSE and MAE for GARCH(1,1) Out-of-Sample Forecasts

5.3 Conclusive findings

The ARIMA models evaluated in this analysis, specifically the ARIMA(0,0,1) and ARIMA(1,0,1), exhibit limitations in capturing the underlying dynamics of the Equinor stock returns, particularly at higher lag lengths. Consequently, their predictive capabilities are restricted, particularly for long-term forecasting. In contrast, the GARCH model demonstrates a higher degree of success in accounting for the volatility clustering and non-linear dependencies inherent in the data, resulting in more accurate out-of-sample predictions. Despite the potential for further refinement to enhance the model's ability to capture extreme events, the GARCH model does indeed emerge as a more reliable and robust tool for forecasting financial time series data compared to the ARIMA models.

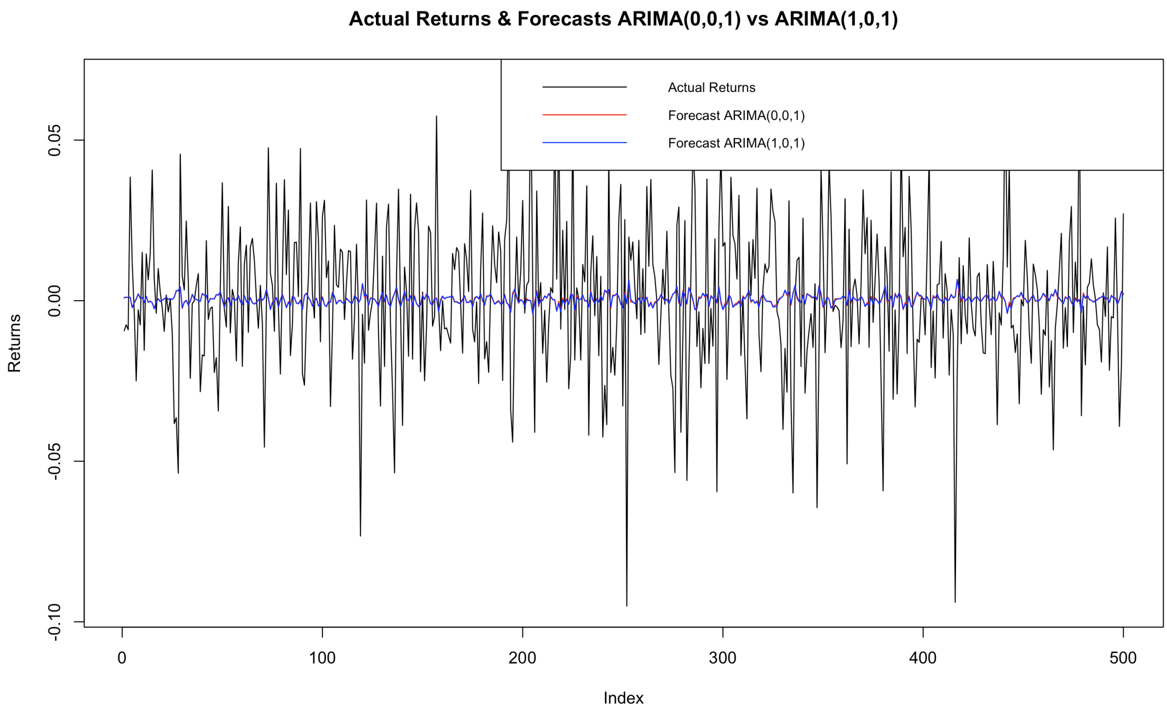

Figure 7: Plot of the actual returns and the predicted returns of the ARIMA(0,0,1) model for the last 500 values of the dataset

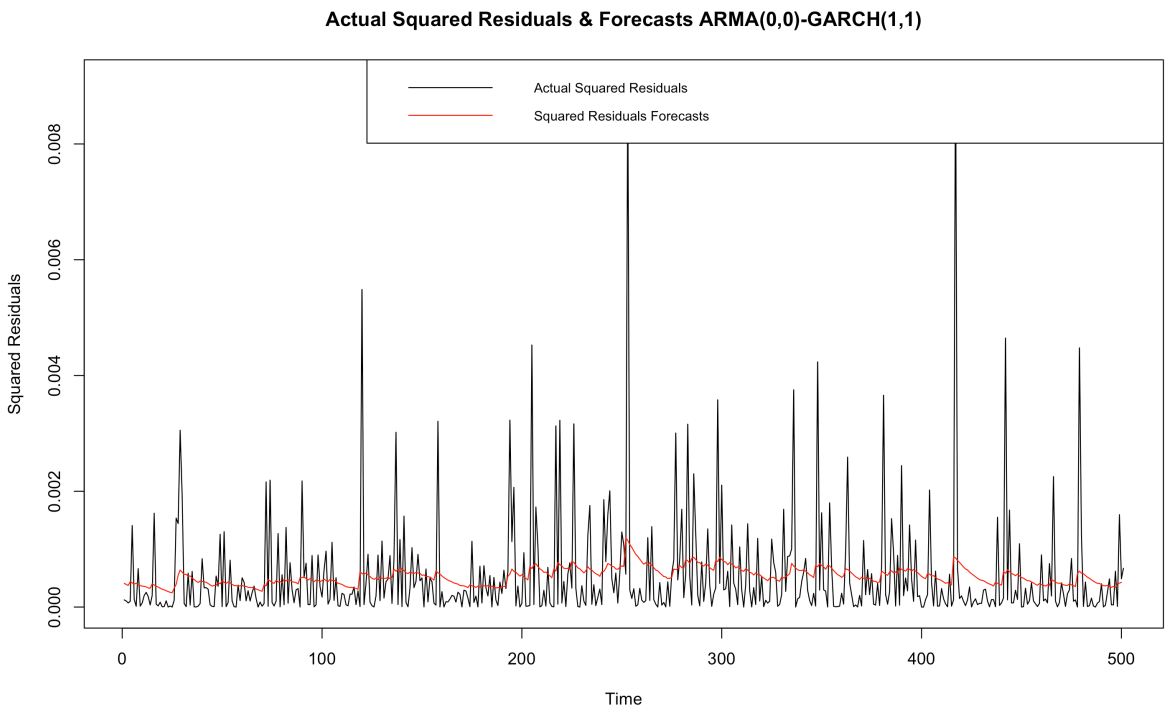

Figure 8: Plot of the actual squared residuals and the predicted squared residuals of the ARMA(0,0)-GARCH(1,1) model for the last 500 values of the dataset

6 Conclusion

In conclusion, this research paper evaluated the performance of ARIMA and GARCH models in predicting future returns and volatility of the Equinor stock. The findings reveal that the ARIMA model falls short in producing accurate forecasts for Equinor stock returns, while the GARCH model effectively captures some of the stock's volatility dynamics. By demonstrating the utility and limitations of ARIMA and GARCH models in forecasting Equinor stock prices and volatility, this study contributes to the existing literature and serves as a resource for investors and financial analysts wanting to develop better trading strategies, improve portfolio performance, and manage risk more effectively regarding the Equinor stock. The performance of ARIMA and GARCH models in this research suggests that incorporating GARCH models into a broader trading strategy could provide some valuable insights into portfolio diversification and risk management. However, this study also acknowledges the potential for more advanced models, such as alternative GARCH variants or machine learning-based approaches, to yield even better results. Another thing to consider is that Equinor is an oil company, and its stock performance could be heavily influenced by oil prices. An Autoregressive Distributed Lag (ADL) model might therefore be more appropriate. Incorporating oil price data into the analysis may provide a more comprehensive understanding of Equinor's stock behavior and improve the forecasting accuracy of the models. Considering Equinor's critical role in the Norwegian economy and its significant presence on the Euronext Growth market, further exploration of these advanced forecasting techniques, including ADL models, could be highly beneficial for stakeholders seeking to better understand this influential stock. Future research directions could include the investigation of other GARCH variants, machine learning models, or ADL models to enhance the forecasting performance for Equinor stock returns and volatility further.

References

Adhikari, R., & Agrawal, R. (2013). *An introductory study on time series modeling and forecasting*. https://doi.org/10.13140/2.1.2771.8084

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *52*, 5–59. https://doi.org/10.1016/j.jeconom.2023.02.001

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*. https://doi.org/https://doi.org/10.1016/0304-4076(92)90064-X

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*. https://www.sciencedirect.com/science/article/pii/092753989390006D

Enders, W. (2014). *Applied econometric time series*. John Wiley Sons Inc.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*. http://www.jstor.org/stable/1912773

Engle, R. F., & Patton, A. (2001). What good is a volatility model? *Quantitative Finance*. https://doi.org/10.1088/1469-7688/1/2/305

GLOSTEN, L. R., JAGANNATHAN, R., & RUNKLE, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*. https://doi.org/https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Gooijer, J. G., & Hyndman, R. (2006). 25 years of time series forecasting. *International Journal of Forecasting*.

Matange, S., & Heath, D. (2011). *Statistical graphics procedures by example: Effective graphs using SAS*. SAS Institute.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*. http://www.jstor.org/stable/2938260

Poon, S.-H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*. https://www.aeaweb.org/articles?id=10.1257/002205103765762743

Tsay, R. S. (2005). *Analysis of financial time series*. John Wiley & Sons.

Wei, W. W. S. (2005). *Time series analysis: Univariate and multivariate methods* (2nd ed.). Pearson/Addison-Wesley.

ref.bib

A Supplementary tables and figures

This is Appendix A, which contains supplementary data analysis and test results.

A.1 Tables of all computed AIC and BIC values

A lower AIC and BIC value indicates a better fitted model. See table 10, 11 and 12.

	p	d	q	AIC
1	0	0.00	0	-26141.23
2	0	0.00	1	-26160.03
3	0	0.00	2	-26158.60
4	0	0.00	3	-26156.65
5	0	0.00	4	-26154.96
6	0	0.00	5	-26152.96
7	1	0.00	0	-26159.54
8	1	0.00	1	-26158.65
9	1	0.00	2	-26156.59
10	1	0.00	3	-26154.65
11	1	0.00	4	-26152.96
12	1	0.00	5	-26150.96
13	2	0.00	0	-26158.53
14	2	0.00	1	-26156.55
15	2	0.00	2	-26154.57
16	2	0.00	3	-26152.64
17	2	0.00	4	-26150.96
18	2	0.00	5	-26148.96
19	3	0.00	0	-26156.61
20	3	0.00	1	-26154.62
21	3	0.00	2	-26152.62
22	3	0.00	3	-26150.93
23	3	0.00	4	-26148.96
24	3	0.00	5	-26146.96
25	4	0.00	0	-26154.94
26	4	0.00	1	-26152.95
27	4	0.00	2	-26150.95
28	4	0.00	3	-26148.95
29	4	0.00	4	-26146.95
30	4	0.00	5	-26144.96
31	5	0.00	0	-26152.97
32	5	0.00	1	-26150.97
33	5	0.00	2	-26148.96
34	5	0.00	3	-26146.96
35	5	0.00	4	-26144.97
36	5	0.00	5	-26156.20

Table 10: All computed AIC values for different ARIMA models

	p	d	q	BIC
1	0	0.00	0	-26128.00
2	0	0.00	1	-26140.19
3	0	0.00	2	-26132.15
4	0	0.00	3	-26123.59
5	0	0.00	4	-26115.29
6	0	0.00	5	-26106.68
7	1	0.00	0	-26139.70
8	1	0.00	1	-26132.21
9	1	0.00	2	-26123.53
10	1	0.00	3	-26114.97
11	1	0.00	4	-26106.67
12	1	0.00	5	-26098.06
13	2	0.00	0	-26132.08
14	2	0.00	1	-26123.49
15	2	0.00	2	-26114.90
16	2	0.00	3	-26106.36
17	2	0.00	4	-26098.07
18	2	0.00	5	-26089.46
19	3	0.00	0	-26123.55
20	3	0.00	1	-26114.95
21	3	0.00	2	-26106.34
22	3	0.00	3	-26098.04
23	3	0.00	4	-26089.45
24	3	0.00	5	-26080.85
25	4	0.00	0	-26115.27
26	4	0.00	1	-26106.67
27	4	0.00	2	-26098.05
28	4	0.00	3	-26089.44
29	4	0.00	4	-26080.83
30	4	0.00	5	-26072.23
31	5	0.00	0	-26106.69
32	5	0.00	1	-26098.07
33	5	0.00	2	-26089.45
34	5	0.00	3	-26080.84
35	5	0.00	4	-26072.23
36	5	0.00	5	-26076.86

Table 11: All computed BIC values for different ARIMA models

A.2 QQ-plots for the residuals

A QQ plot (Quantile-Quantile plot) for the residuals is a graphical tool used to assess whether the residuals of the model follow a normal distribution. In other words, it helps evaluate the goodness of fit of the model by checking the underlying assumptions about the error term. If the residuals follow a normal distribution, the points in the

Model	BIC
m- 1 - 1	-5.003145
m- 1 - 2	-5.001838
m- 1 - 3	-5.001106
m- 1 - 4	-5.000314
m- 2 - 1	-5.001578
m- 2 - 2	-5.000274
m- 2 - 3	-4.999802
m- 2 - 4	-4.999936
m- 3 - 1	-5.000016
m- 3 - 2	-4.998709
m- 3 - 3	-4.998237
m- 3 - 4	-4.998515
m- 4 - 1	-4.998456
m- 4 - 2	-4.997148
m- 4 - 3	-4.996650
m- 4 - 4	-4.997121

Table 12: All computed BIC values for different GARCH models

QQ plot should form a roughly straight line along the diagonal (Matange and Heath, 2011). In figure 9 and 10, we observe that there is some deviation from the diagonal line, which indicate that the residuals are not normally distributed.

A.3 ACF and PACF plots for the ARIMA models residuals

A well specified model should have no autocorrelation in the reisudals. Figure 11 and 12 show that there is some significant autocorrelation in the residuals for both models at different lags.

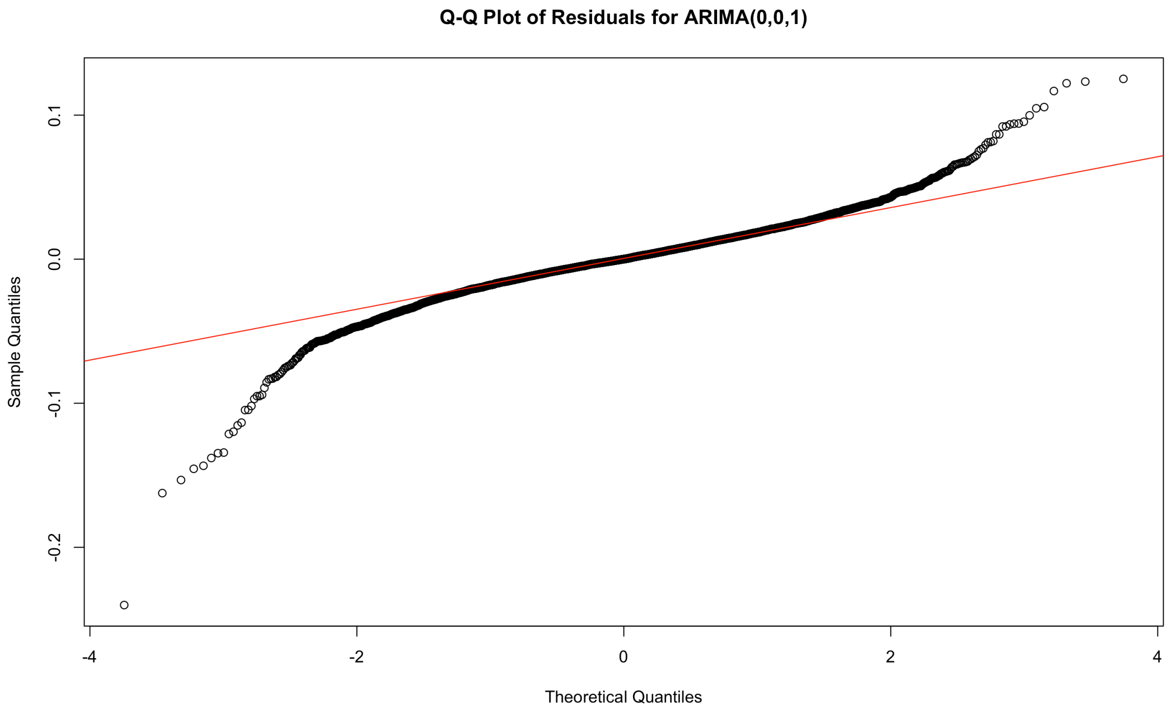

Figure 9: QQ-plot for ARIMA(0,0,1) residuals

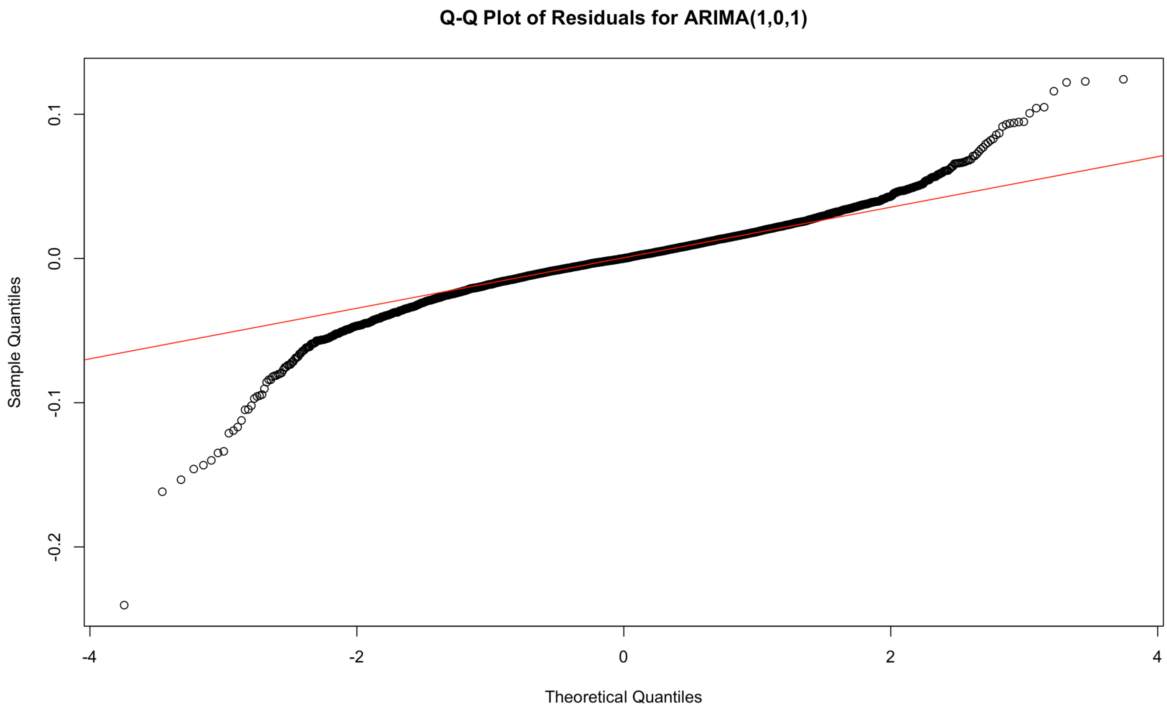

Figure 10: QQ-plot for ARIMA(1,0,1) residuals

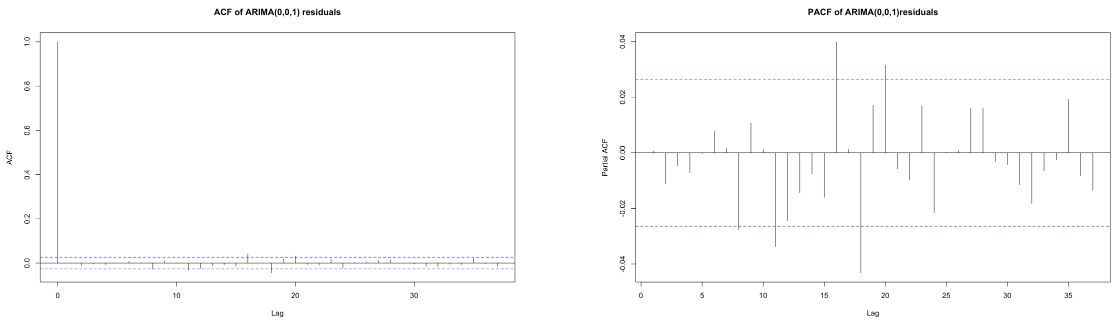

Figure 11: ACF and PACF of the residuals of the ARIMA(0,0,1) model

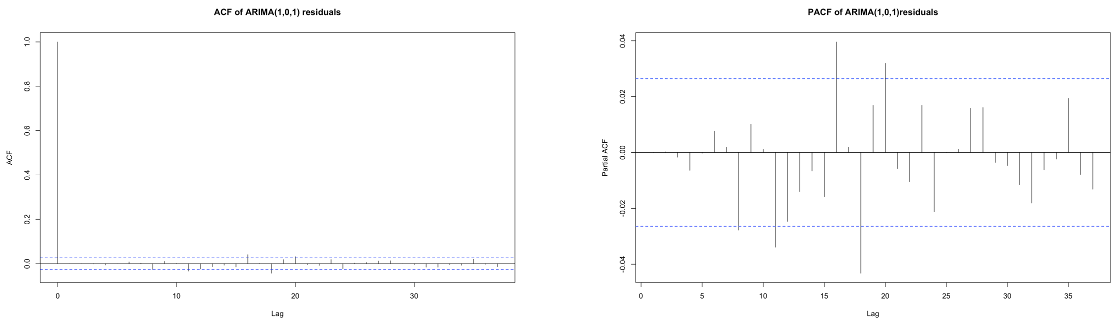

Figure 12: ACF and PACF of the residuals of the ARIMA(1,0,1) model