FACULTY OF SCIENCE AND TECHNOLOGY

BACHELOR'S THESIS

Study programme / specialisation: Bachelor in Mathematics and Physics	The *spring* semester, *2023* Open
Author: Benjamin Vaag Miller	
Supervisor at UiS: Jan Terje Kvaløy	
Thesis title: Time Series Analysis of Financial Data	
Credits (ECTS): 20	
Keywords: Statistics, Time Series, Econometrics	Pages: 37 + appendix: 3 Stavanger, *15.05.2023*

Time Series Analysis of Financial Data

Benjamin Vaag Miller

May 15. 2023

Abstract

The motivation for this thesis is to grasp general theory and principles for modeling time series, for the application to financial data. By going through the theory that makes up this field and bringing that perception to the implementation of modeling. To then evaluate the applicated models in their sense, and further draw conclusions from the data. The basis for this thesis will be the application of the ARMA model and the GARCH model to a transformation of the stock price for each day. The data is gathered from the closing price of Orkla AS, a company listed on the Oslo Stock Exchange (Oslo Børs).

In the first example, we first determine the order of the ARMA model by looking at significant sample partial autocorrelations and sample autocorrelations. After the order is determined we evaluate the ϕ_i and θ_j with maximum likelihood estimation. Furthermore evaluating the ARMA model with residual plots and forecasting it. The plot of predicted values for the ARMA model varies largely in amplitude compared to actual returns.

We then apply the GARCH model that can measure volatility. The evaluation of the GARCH and ARMA parameters is also done by maximum likelihood estimation. As the predicted values of the GARCH model behave more similarly to the returns than what the ARMA model does, and the diagnostics are accepted we can use the forecasted values to determine a good investment strategy for this financial asset.

Preface

The following bachelor thesis is a part of my Bachelor's in Mathematics and Physics, at the Faculty of Science and Technology of the University of Stavanger.

I want to thank my supervisor, Jan Terje Kvaløy, for tremendous support, guidance, and insightful discussions, through my writing.

Stavanger, May 15. 2023.

Benjamin Vaag Miller

Contents

1 Introduction

This thesis is inspired by the book 'Introduction to Time Series and Forecasting' by Peter J. Brockwell and Richard A. Davis.

Any process that is occurring over a period of time can be modeled as a time series. The objective of time series is to generate an understanding of the underlying system of time-varying data. Time series is generally used in fields such as engineering, finance, and economics. We will specifically look at the log-return of the Orkla AS stock in this thesis. Further evaluating time series models with the diagnostics checkers available. Accompanying the theory with the manually calculated models and then forecasting these models. We can draw conclusions and make strategies for the financial asset we are evaluating over the course of a year.

Through the first two chapters, we will cover the basic theory of statistics which is fundamental to understanding the following topics. Thoroughly explaining random variables, stochastic processes, basic statistical notation, and basics of probability distributions. Furthermore covering the basics of time series before actually applying the theory to the models.

In the third chapter, we cover the ARMA model. With the fundamentals of time series describing the structure and traits of the ARMA model. The first example of time series modeling is the ARMA model. We then use the residuals as the main checker in various ways such as the autocorrelation function, and partial autocorrelation function. These are very important for evaluating the order of model and for evaluating the fit of the model.

In the fourth chapter, we discuss the use of GARCH models and the advantages of this model when handling financial data, and the reason for it being a very popular model for financial data. While evaluating the volatility and forecasting in contrast to the ARMA model.

2 Theory

2.1 Random Variables

This section will cover the basic principles of statistics used for time series analysis, such as the specifications of random variables and probability distributions.

2.1.1 Discrete Probability Distributions

The probability mass function of a random variable X, denoted as $f(x)$, has these attributes

$$\begin{aligned} &1. \quad f(x) \geq 0, \\ &2. \quad \sum_x f(x) = 1, \\ &3. \quad P(X = x) = f(x) \end{aligned}$$

This in turn helps define the cumulative distribution function F of a random variable X can be defined as

$$F(x) = P[X \leq x], \qquad \forall x \in \mathbb{R}$$

For situations when the possible values of X are not continuous, as such $\{x_0, x_1, ...\}$, the pmf denoted as F can usually be written as

$$F(x) = \sum_{t \leq x} f(t)$$

2.1.2 Continuous Probability Distributions

As for a continuous probability distribution the density function $f(x)$ has similar conditions as the discrete distributions.

$$\begin{aligned} &1. \quad f(x) \geq 0, \qquad \forall x \in \mathbb{R} \\ &2. \quad \int_{-\infty}^{\infty} f(x) = 1, \\ &3. \quad P(a < X < b) = \int_a^b f(x)\, dx \end{aligned}$$

Thus, the cumulative distribution function, $F(x)$, of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \forall x \in \mathbb{R}$$

2.1.3 Expectation Value

When analyzing a probability distribution such as above one important thing to compute is the expectation value. The expectation can be viewed as the "center" value of the distribution.

$$\mu = E\left(X\right) = \begin{cases} \sum_{x} xf(x), & \text{for discrete distributions} \\ \int_{-\infty}^{\infty} xf(x), & \text{for continuous distributions} \end{cases}$$

This μ can be understood as the limit of the sample mean as n $\rightarrow \infty$. The sample mean and other sample values we will use a hat over as the notation, as will later will be shown.

2.1.4 Variance and Covariance

We have seen that when analyzing probability distributions the expectation is an important value. Another quantity to look at is the variance.

$$\sigma^2 = \mathrm{Var}\left(X\right) = E[(X-\mu)^2] = \sum_{x} \left(x-\mu\right)^2 f(x)$$

The variance is a measure of the deviation from the expectation as viewed in this term, $(x-\mu)^2$. Another property of the variance, which is a simplification to determine the variance is such

$$\sigma^2 = E(X^2) - \mu^2$$

The relational behavior between two random variables, X and Y, can be measured as a number between 0 and 1, called covariance. From here on a random variable will be abbreviated to "r.v.". For r.v.s X and Y with joint distribution f(x,y), the definition of the covariance

$$\sigma_{X,Y} = Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_{x}\sum_{y} (x-\mu_x)(y-\mu_y)\, f(x,y)$$

2.1.5 Correlation

Correlation is a measure of the linear relationship with the value of correlation varying in values between -1 to 1. A value of zero means no linear relationship.

Whilst values of 1 and -1 means full linear relationship and full negative linear relationship repsectfully. Correlation is denoted as such.

$$\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad \text{if } \sigma_X \sigma_Y > 0 \tag{1}$$

2.2 Concepts

In this chapter we will go through some of the important concepts that use the fundamentals of statistical analysis to provide an understanding of how random variables behave by looking at their traits and their relationship to one another. Furthermore how we use these calculations to predict the future values and critically conclude the outcome.

2.2.1 Basic Notation

A stochastic process is a set of random variables indexed, in our case by time. The random variables X_t is the value of the stochastic process at time t, it represents the random variables of an observation x_t. As for time series, you have continuous-time time series and discrete-time time series. We will mainly focus on discrete-time time series which is a discrete-time stochastic process.

$\{x_t\}$, are the realized values

$$\{x_t\} = \ \{x_1,\ x_2,\ x_3,\ x_4, ...\}$$

This is a general representation of observation data used to make time series models. Whereas $\{X_t\}$ is referred to the whole process of the random variables $\{X_1, X_2, ...\}$. Without curly brackets, X_t, this means the value of the random variable at time t.

In a time series model the means and covariances of the set of r.v.s, $\{X_t\}$, are among the components we try to estimate. For any t.s.m. $\{X_t\}$ we have the mean function with $E(X_t^2) < \infty$

$$\mu_X(t) = E(X_t)$$

and the covariance function

$$\gamma_X(t,h) = Cov(X_t, X_h) = E[(X_t - \mu_X(t))(X_h - \mu_X(h)]$$

No as we now from 1, that the correlation function must be as follows

$$\rho_X(t,h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)}$$

2.2.2 Weakly Stationary Models

A weakly stationary process has

$$\begin{aligned} &\mu_X(t) = \mu_X && \text{not dependent of t} \\ &\gamma_X(t, t+h) = \gamma_X(0, h) && \text{not dependent of t for each h} \\ &E(X^2) < \infty \end{aligned} \tag{2}$$

The first equation means that the mean is independent of time. Furthermore that the auto-covariance is independent of time, and finite variance of the process.

For a time series $\{X_t\}$ to be strictly stationary $(X_1, ..., X_n)$ and $(X_{1+h}, ..., X_{n+h})$ has to have the same joint distributions $\forall$ h and $n > 0$, where $h, n \in \mathbb{Z}$. There are different types of stationarities in processes other than weakly and strict, see [2]. When referring to a stationary time series, *stationary* will mean weakly stationary.

In time series the covariances between the random variables at different times are the ones we pay attention to. Hence we look at $Cov(X_t, X_{t+h})$, which actually just refers to the difference h, when the process is stationary. The covariance in a time series is measured for each variable over different lags, and the function that encapsulates the covariances of all the r.v.s is called the autocovariance function and has the symbol γ. I will refer to this function as ACVF.

$$\begin{aligned} &\gamma_X(h) := \gamma_X(0, h) = \gamma_X(t, t+h) \\ &\text{s.t.} \\ &\gamma_X(h) = Cov(X_t, X_{t+h}) \end{aligned}$$

This "h" is what we call lag and refers to the index-shift or time-shift between the variables. E.g. The variable X_t has its one-lag difference variable X_{t+h}.

Another function used is the autocorrelation function of $\{X_t\}$ at lag h, its abbreviation ACF will be used.

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_t, X_{t+h})$$

In practice a time series model with data used from a real world would have properties that depend on time. These properties are called trend and seasonality. A stationary time series model doesn't have properties that depend on time. Thus, a stationary t.s.m. doesn't have trend or seasonality. Usually when we have a t.s.m. we can remove the trend and seasonality, then what we are left with are the *stationary* residuals.

2.2.3 Sample Autocovariance function

In the pursuit of finding the best time series model in a practical setting we are not handed the solution of the most accurate fitted model. Usually, we start out with a set of observed data. Among other things, we use the sample ACF as a pointer for which model is of the best fit. The sample covariance is

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \overline{x})(x_t - \overline{x}), \quad h < |n|$$

Furthermore the sample autocorrelation function is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

2.2.4 White Noise

Now that we have discussed fundamentals of time series the most basic t.s.m. is the one called white noise. The r.v.s of a white noise process are independent of each other with zero mean and constant variance. This is the notation we use for white noise

$$\{X_t\} \sim WN(0, \sigma^2)$$

Since white noise is not dependent on time, it is stationary as in 2 with ACF

$$\gamma_X(h) = \begin{cases} \sigma^2, & \text{if h} = 0 \\ 0, & \text{if h} \neq 0 \end{cases}$$

White noise is one of many types of stationary processes.

2.2.5 Components of a Time Series

In a time series model, there are usually time-dependent components of real data. Trend is an occurrence over the long term, theoretically, it can be represented by any function (linear, quadratic, exponential). Seasonality is the changes over time that occur at specific frequencies. These cycles have frequencies that can be evaluated. In real terms, they are hours, days, weeks, years, etc.

The residuals are the errors or deviations that exist between the actual values of a time series and the predicted values 3.6 of a model. As for any real-world situation, the imperfectness is shown in the differences between the model to the data. They represent the portion of the variation in the data that is not accounted for by the model. Residuals can be positive or negative, depending on whether the actual values are higher or lower than the predicted values. The residuals are very useful as they can be used to measure the accuracy of the model. Hence, we use residuals to assess the goodness of fit of the model and to identify any patterns or trends in the data that are not captured by the model.

The way we assess the goodness of a model is by looking at the residuals plotted over time. This plot should be concentrated around a mean and have a constant variation around this mean. This is the assumption of constant mean and variance.

When representing the seasonal component and trend component of a t.s.m. we can separate them from each other as such

$$X_t = m_t + s_t + Y_t$$

where m_t is the trend, s_t is the seasonal, and Y_t is the random noise which is stationary as in 2. It is called the classical decomposition model.

The objective is to estimate m_t and s_t, such that when extracting them such that the residual or noise component, Y_t, is stationary.

2.2.6 Differencing

We define a operator ∇ as the 1-lag difference

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

B is the backward shift operator,

$$BX_t = X_{t-1}$$

It then follows that,

$$\begin{aligned} &B^j(X_t) = X_{t-j} \\ &\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t)),\ j \geq 1 \qquad \text{with} \quad \nabla^0(X_t) = X_t \end{aligned}$$

Differencing is described here and is of importance to the concept of forecasting. It is used for the predicted values 3.6 of models.

3 Autoregressive Moving Average

The autoregressive moving average is a time series model composed of two other models, the moving average and the autoregressive model. At this point, we will review these models further and apply them to a data sample. This chapter is heavily influenced by the book [2].

3.1 Autoregressive model

The autoregressive (AR) model assumes that any current value is a function of previous values, with an added stochastic term, usually white noise $WN(0, \sigma^2)$. The stochastic term is also called the error term, due to its trait of fluctuations. Also referred to as the stationary term. The first-order autoregressive model expresses the current value as a linear function of the previous value with the stationary term.

$$X_t = \phi X_{t-1} + Z_t$$

where $|\phi| < 1$, $\{Z_t\}$ is a white noise process $WN(0, \sigma^2)$ and Z_t is uncorrelated with the random variables of X for every previous value of Z_t.

The general autoregressive model is denoted as AR(p), with p being the number of lags that the model is using to make the linear function for forward values. The generalized model for order p, looks like.

$$X_t = \beta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t$$

3.2 Moving Average

As opposed to the AR-model the moving average (MA) model does not look at the past value, instead, it looks at the error term of the previous value. It is a linear function of past error terms. An MA(q) model is generalized as such

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$$

Z_t is again a $WN(0, \sigma^2)$ and θ_i are constants.

3.3 ARMA model

The autoregressive moving average (ARMA) model brings both the AR and MA model into one model. Denoted ARMA(p, q), with p autoregressive terms and q moving average terms. An ARMA model has the general equation

$$X_t = Z_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j}$$

Where all the parameters ϕ_i and θ_j have no common factors. The MA and AR models usually need a large number of terms to accurately depict a set of data as a time series. This is one of the benefits of the ARMA model as it produces a model with far fewer terms, this is illustrated in chapter 3.8. ARMA(p, q) process has the characteristics of being causal and invertible.

3.4 Partial Autocorrelation function

When determining an ARMA models fit to the data, we use the sample ACF to see the correlations at different lags for the MA(q) part of the ARMA(p, q). For the AR(p) part we use the partial autocorrelation function, $\alpha(h)$. We will refer to it as PACF.

$$\alpha(0) = 1$$

$$\alpha(h) = \phi_{hh} \qquad h \geq 1$$

The PACF at lag h is the last component of

$$\boldsymbol{\phi}_h = \boldsymbol{\Gamma}_h^{-1} \, \boldsymbol{\gamma}_h$$

Here $\boldsymbol{\Gamma}_h$ the autocovariance-matrix $\boldsymbol{\Gamma}_h = [\gamma(i-j)]_{i,j=1}^{h}$ and $\gamma_h = [\gamma(1), \gamma(2), ..., \gamma(h)]$

The difference between the ACF and PACF is that the PACF measures the correlation between two observations after removing the effects of all the other observations between them, whereas the ACF measures the correlation of the variables and all of their lagged values. The PACF is used to determine the order of the autoregressive terms in the ARMA model. While ACF is used for the order of moving average terms. This will be visualized further in the example of the ARMA model.

3.5 Estimation and Forecasting

As previously seen, a lot of theory goes into just understanding how time series operate. Now we have come to the point of how we use the theory to acquire the attributes and be equipped to utilize them. In order to determine the best ARMA(p,q)-model of a stationary time series, we must look at the decision of p and q, the estimation of the mean, the white noise variance σ^2, and the coefficients of ϕ_i and θ_j where i ranges from 1 to p and j from 1 to q.

3.6 Predictors

To predict a future variable of a time series we introduce a linear predictor operator P_n. Its purpose is to forecast the value X_{n+h} based on observations up to time n, with minimum squared error and its behavior is such that

$$P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1$$

When applying the predictor to predict a variable X_{n+h} the following properties apply

$$1. P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu), \quad \boldsymbol{a}_n = (a_1, ..., a_n)^T$$
$$2. E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \boldsymbol{a}_n', \quad \boldsymbol{\gamma}_n(h) = (\gamma(h), ..., \gamma(h+n-1))^T$$
$$3. E(X_{n+h} - P_n X_{n+h}) = 0$$
$$4. E[(X_{n+h} - P_n X_{n+h}) X_j] = 0, \quad j = 1, ..., n$$

The predictor is crucial when understanding the theory behind forecasting. For proofs of these properties see [2], Section 2.5.

3.7 Maximum Likelihood Estimation

We define the likelihood of $\boldsymbol{X}_n$ if $\boldsymbol{X}_n$ is the vector $(X_1, ..., X_n)^T$ with $E(\boldsymbol{X}_i) = 0$ and where $\boldsymbol{X}_n$ is multivariate normally distributed. The covariance matrix as such $\Gamma_n = E(\boldsymbol{X}_n \boldsymbol{X}_n^T)$

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n}} \frac{1}{det(\Gamma_n)} \exp\left(-\frac{1}{2} \boldsymbol{X}_n^T \Gamma_n^{-1} \boldsymbol{X}_n\right)$$

This is the generalized likelihood function for a time series model. When applying this likelihood function to an ARMA process we can express the Γ_n^{-1} by only the $p+q$ parameters.

Now to determine the coefficients ϕ_i and θ_j we first assume $\{Z_t\} \sim \text{IID}(0, \sigma^2)$. Then by the innovations algorithm ([2], Section 2.5.4) we have that the one-step predictor of the time series $\{X_t\}$ is

$$\hat{X}_n = \begin{cases} \sum_{j=1}^{t-1} \theta_{ij} (X_{t-j} - \hat{X}_{t-j}), & 1 \leq t < \text{max(p,q)}, \\ \sum_{j=1}^{t-p} \phi_j X_j + \sum_{j=1}^{q} \theta_{ij} (X_{t-j} - \hat{X}_{t-j}), & t \geq \text{max(p,q)}, \end{cases}$$

Then by $E(X_{n+1} - \hat{X}_{n+1})^2 = \sigma^2 r_n$ we reach the desired likelihood function of the parameters

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = \frac{1}{\sqrt{(2\sigma^2)^n \ \prod_{j=1}^{n} r_{j-1}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}\right)$$

It is called the Gaussian Likelihood Function of an ARMA model. In particular it can be shown that $r_1, r_2, ..., r_n$ can be expressed in terms of $\phi_1, ..., \phi_p$ and $\theta_1, ..., \theta_q$. A more detailed review can be seen in [2], Section 5.2.

To simplify calculations we next take the logarithm:

$$l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = \ln\, L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\,\ln\, a - \frac{1}{2}\,\ln\left[(2\pi)^2 \prod_{j=1}^{n} r_{j-1}\right] - \frac{1}{2\sigma^2}\; S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$$

with

$$S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

then partially differentiating by σ^2,

$$\frac{\partial l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$$

$$0 = \frac{1}{2\sigma^2}\left(-n + \frac{1}{\sigma^2} S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})\right)$$

$$\sigma^2 = \frac{1}{n} S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$$

This σ^2 is the MLE estimator and then will be referred to as such $\hat{\sigma}^2$. With $\hat{\phi}$ and $\hat{\theta}$ being the values that minimize the simplified

$$l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = \ln\left(\frac{1}{n} S(\boldsymbol{\phi}, \boldsymbol{\theta})\right) + \frac{1}{n}\sum_{j=1}^{n} \ln\, r_{j-1}$$

3.8 Estimating and Forecasting ARMA

Let's look at an example of the ARMA model. When modeling using time series it is natural not to compute the coefficients and all the sample values for every time index. This would require a lot of time to calculate without a computer program. As it happens we have tools that can make the calculations for us. The only thing we have to do is supply the values we want to model, then evaluate the model. There are many computer programs that are capable of modeling time series, we will use R-Studio which runs on the programming language R. To visualize the process of modeling and forecasting we will use a set of functions from different packages in R that calculates and computes the models with the use of all theory we have explained.

We will make a transformation of the data to simulate the return on an investment. The transformation is done as such:

$$X_t = \ln \frac{P_t}{P_{t-1}} \tag{3}$$

Where P_t is the price of a stock at day t. This tranformation is also called *log-return*, we will just call it return.

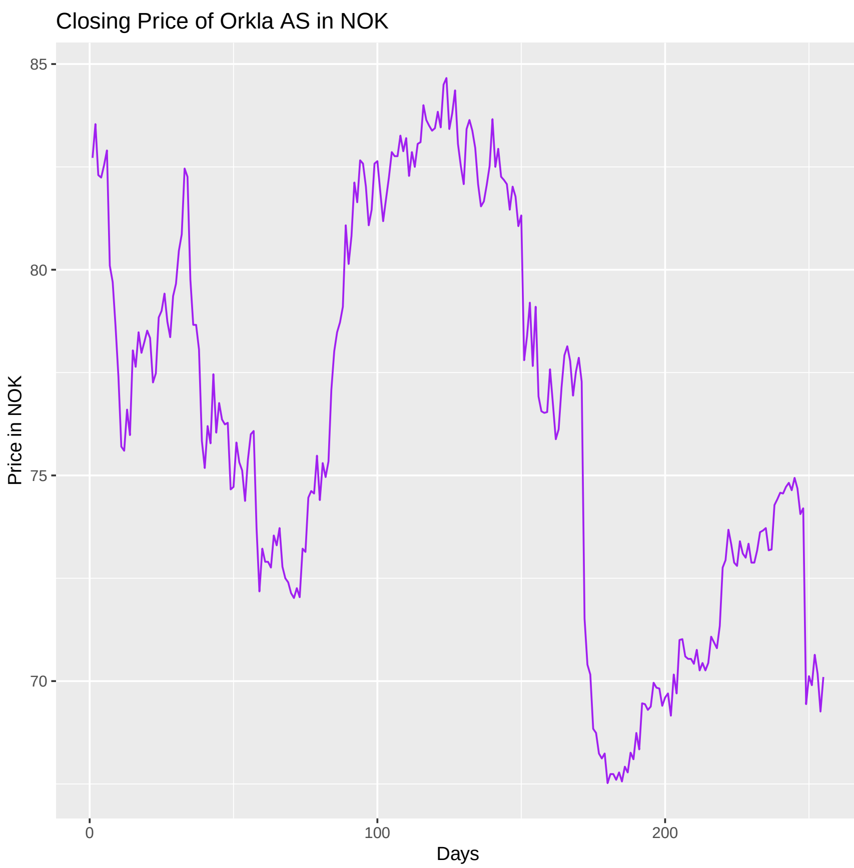

Figure 1: Closing price of Orkla AS throughout period [4].

The chosen stock is Orkla AS which is a grocery retail company mostly directed to the Norwegian market with smaller shares in other markets around Norway.

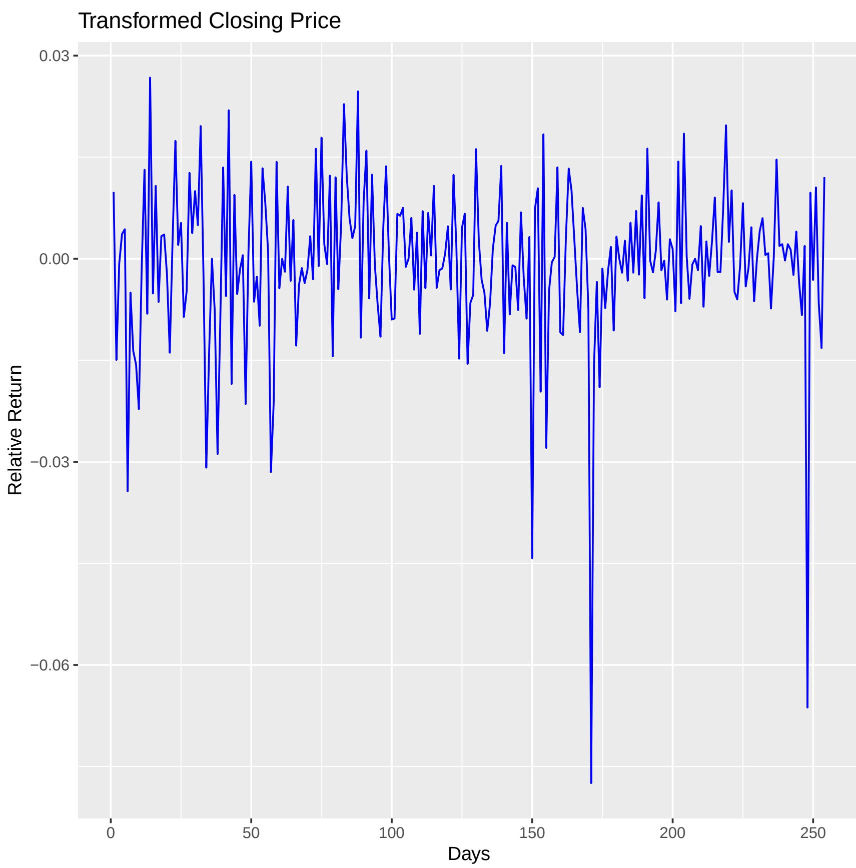

Figure 2: Logarithmic returns of Orkla AS of the period [4].

The data are collected from Yahoo Finance [4] between the dates "22.02.2022 – 22.02.2023". The raw data can be viewed in figure 1.

First, to determine the order of p and q we look at the ACF and PACF of the ARMA respectively.

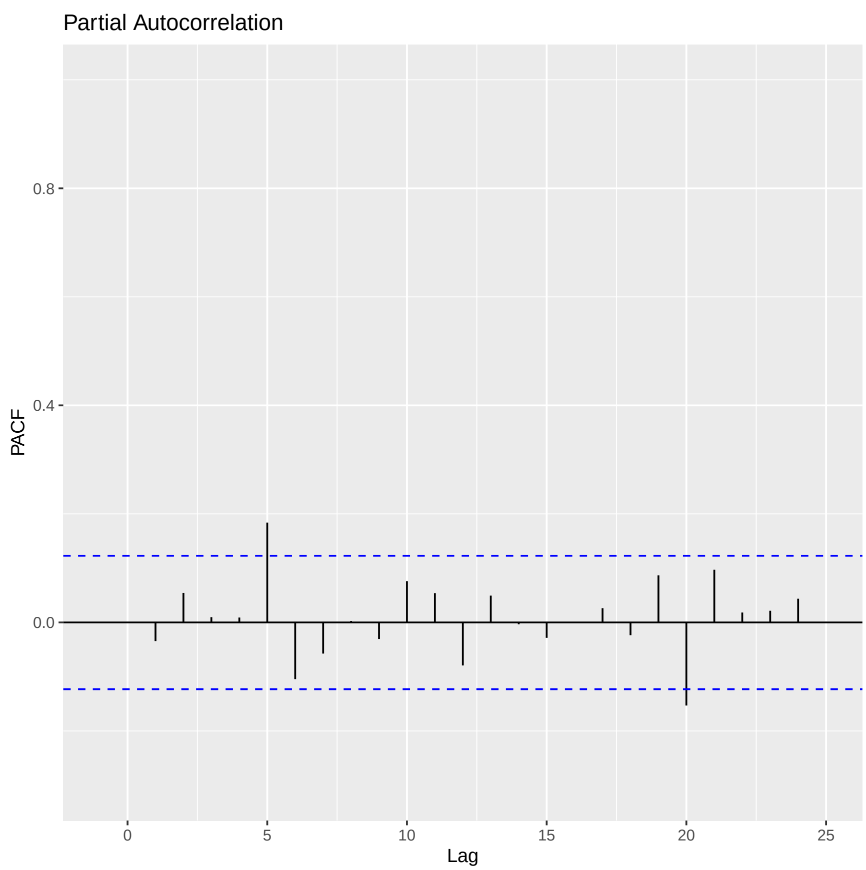

Figure 3: PACF of the transformed data of Orkla AS.

In the PACF of figure 3 we see that the PACF has a significant correlation at lag 5 and lag 20. Since the absolute value of the correlation is larger at lag 5. This would be a good choice for order p.

In 4, we observe that the ACF has similar, but not identical values. The choice of q would therefore be 5.

To try to understand the reason for a 5-day correlation. The data collected is from a stock that is listed on Oslo Stock Exchange (Oslo Børs). The stock exchange is only open Monday through Friday, which suggests some form of relationship between the closing price on a weekly basis (since a week on the stock exchange is five days). Furthermore, we see that the second-largest correlation at a 20-day difference occurs on what we might suggest being an approximately monthly basis. As a week on the stock exchange is 5 days then a month is ap-

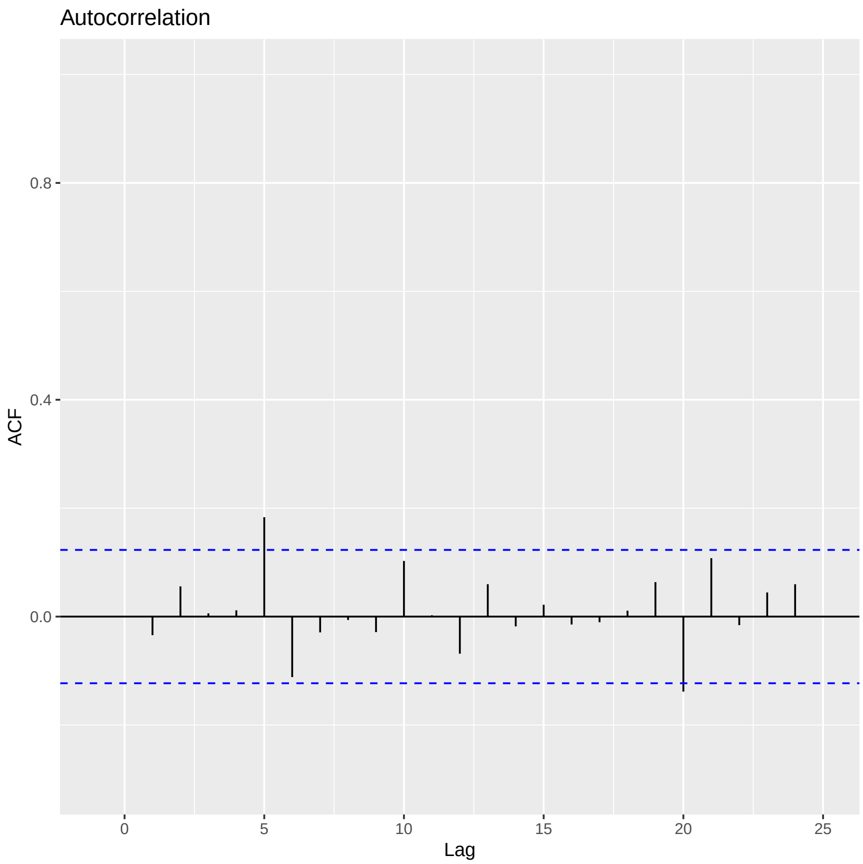

Figure 4: ACF of the transformed data of Orkla AS.

proximately 20 days. With the knowledge of this, we choose our ARMA model to account for these correlations.

There are other ways to determine the best order of an ARMA model called the AIC, FPE, and BCE criterions. These will not be discussed, but are diagnostic checks that are mentioned in [2].

The applied ARMA model has the fitted values as such

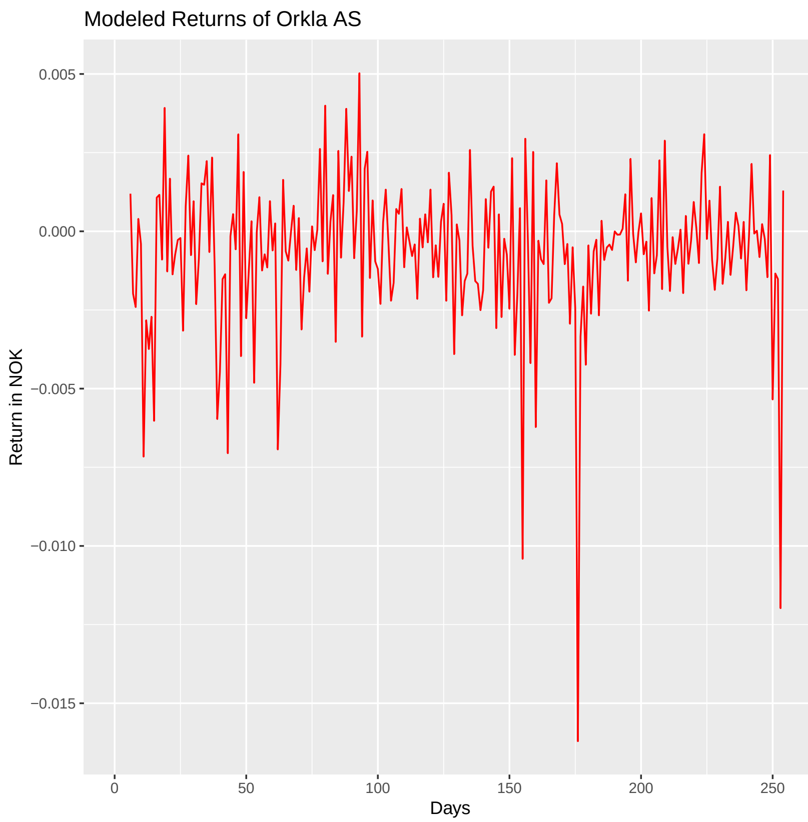

Figure 5: ARMA model of returns

The fitted values of the ARMA model are the predicted values for each of the values. By inspection, the similarities are substantial to the values before they have been run through the ARMA model. To visually inspect we plot them over each other.

$$
\begin{aligned}
X_t = & -0.376X_{t-1} - 0.414X_{t-2} - 0.312X_{t-3} - 0.831X_{t-4} - 0.043X_{t-5} \\
& + 0.354Z_{t-1} + 0.49Z_{t-2} + 0.347Z_{t-3} + 0.95Z_{t-4} + 0.212Z_{t-5}
\end{aligned}
$$

These are the values of ϕ_i and θ_j that were calculated by maximum likelihood by the function in R.

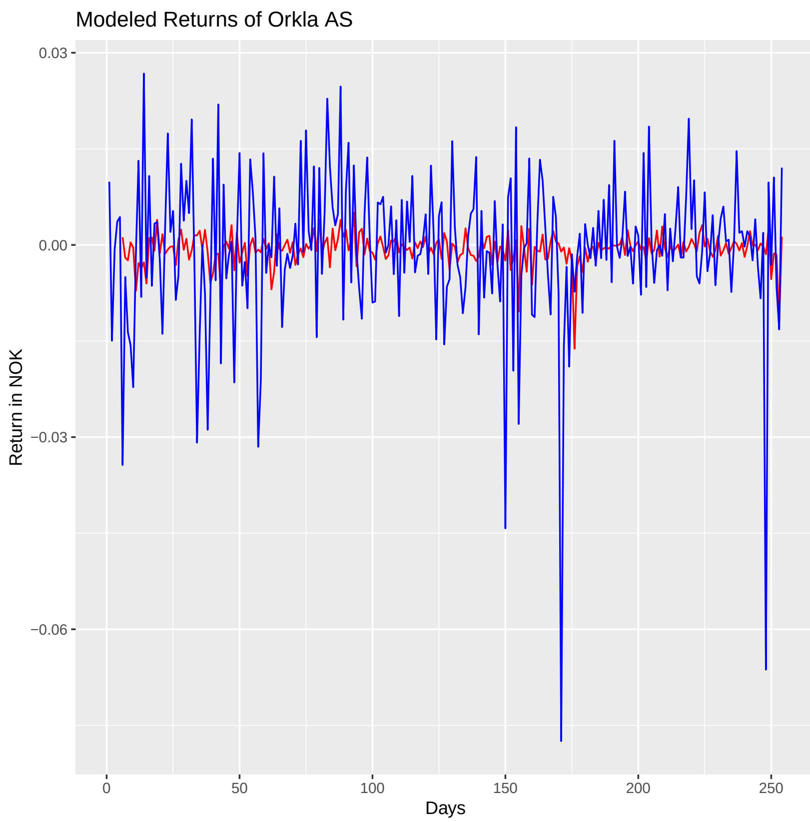

Figure 6: The predicted returns of the ARMA model over the returns.

In perspective, we can now see the predicted returns are closer to 0 than the returns of the stock. One diagnostic check is looking at the predicted values and the observed values should be similar. As we saw they follow a very common structure however the predicted values are one order of magnitude smaller. This is a sign of not generally being the best fit.

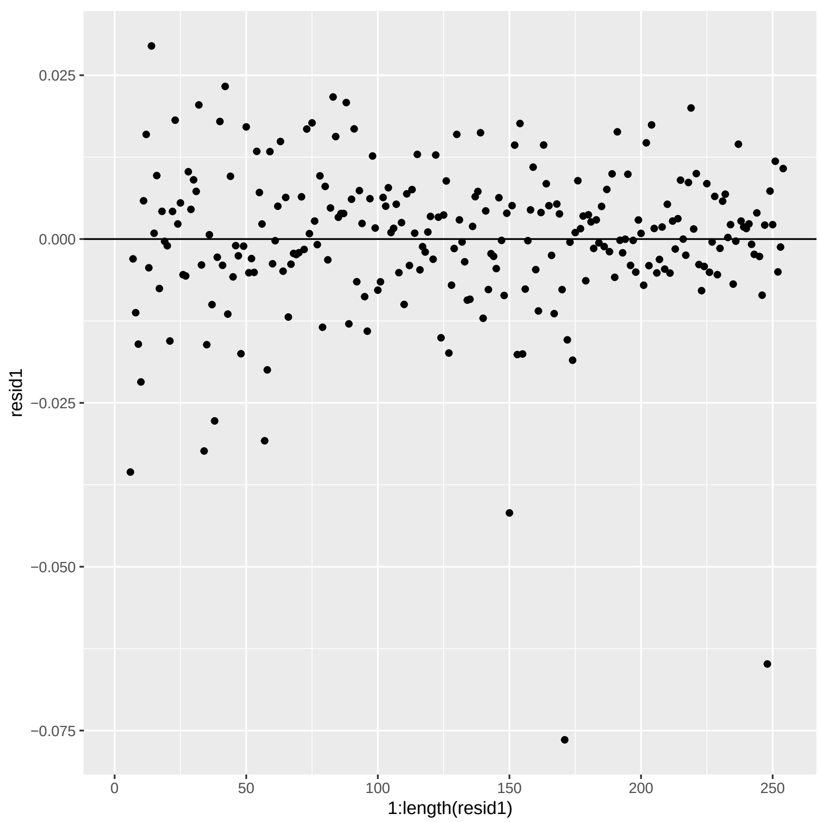

Figure 7: The residuals plotted over the days.

From 2.2.5 we need to see that the constant variance assumption is upheld. By examining figure 7 the residuals are concentrated around the mean $\mu = 0$. This is testing for stationarity and for randomness. Assessing the data we see that the residuals are to some degree at a constant distance from the mean. Although there should not be any trend in this residual plot there may be a pattern of the variance decreasing.

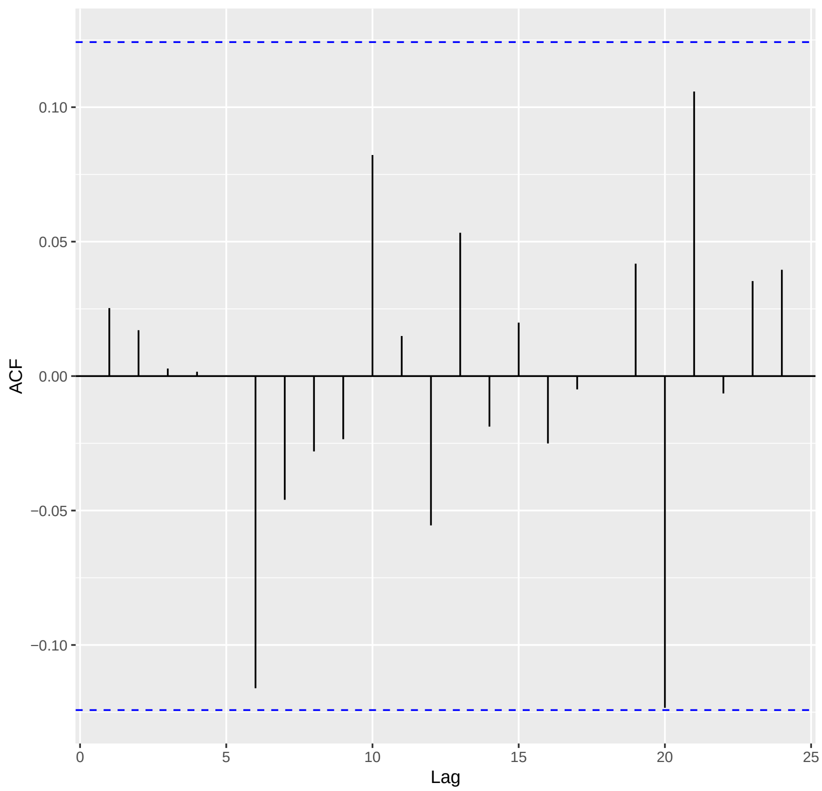

Figure 8: ACF of residuals with the upper and lower bounds of $\sigma \cdot 1.96/\sqrt{n}$

The diagnostics of residuals involves also looking at the PACF and ACF of the residuals. For large number of observations n, the sample autocorrelations of an iid process $\{X_t\}$ with finite variance are approximately normally distributed $N(0, 1/n)$ ([2], p.146). To verify this we have to count how many of the residuals fall outside the lower and upper bounds of the ACF or PACF. In figure 3 only one falls outside the bounds which we determine to be not enough to reject the hypothesis of normal distribution.

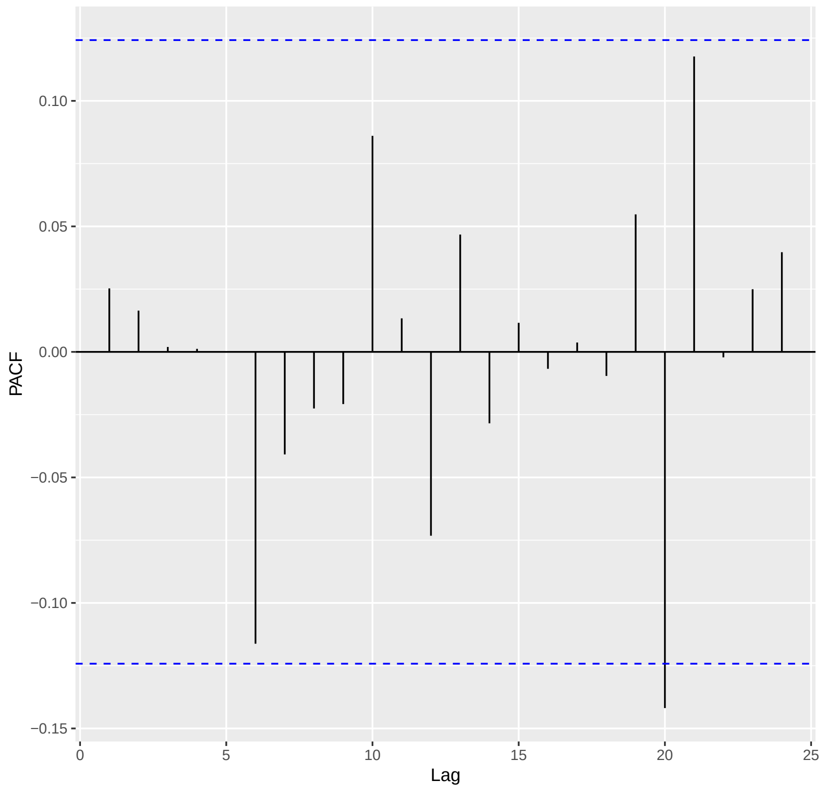

Figure 9: PACF of residuals with the upper and lower bounds of $\sigma \cdot 1.96/\sqrt{n}$

The main motives of time series analysis are to model a process as accurately as possible to describe the behavior of such processes. The other main motive is forecasting which will be covered in the last part of this chapter.

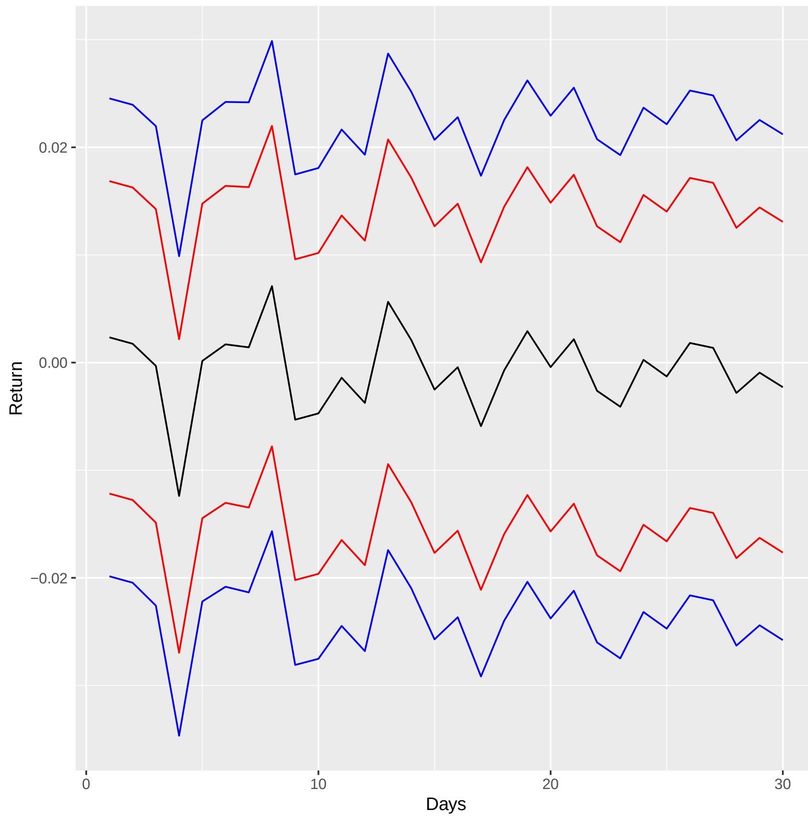

Figure 10: Forecasted ARMA model 30 days forward. The red area is the 80% prediction interval, while blue area is 95% prediction interval.

The forecast is using the value of ϕ_i and θ_j and the same equations to determine predicted values to calculate forecasted values with uncertainty in the form of prediction intervals. In figure 10 we see that for $n \to \infty$ $E(X_t) = 0$, which also is a component of an ARMA model.

In the mind of an asset manager, the way to earn the most money is to sell the stock when the return is at its highest. In our model, this is at 8 days after the last day of the observed price. As investment strategy varies vastly on what outcome you would like to accomplish with an investment. We will discuss this in greater detail in the discussion.

We will see that there is a time series model that is better to predict the fluctuations, since up til now we have assumed stationarity in the process of returns. We will try to change our assumption to view a time series in a different manner.

4 GARCH Models

The following equations and explanations are interpreted as explained by [2] and [5].

Now the representation of financial data is not best modeled with an ARMA model mostly because the returns on a stock price aren't entirely a stationary process. The way we improve our modeling to account for this serial dependence is that we incorporate a second term in the stochastic term. The conditional variance, which is the main part of this model, as shown in figure 7, is denoted h_t and was postulated by Engle [3]

$$X_t = \sqrt{h_t} e_t \quad \text{where } \{e_t\} \sim IID\ N(0,1) \tag{4}$$

The ARCH model that Engle [3] presented as in [2] was

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2$$

A generalization of the ARCH model was introduced by Bollerslev [1] called GARCH, where the conditional variance h_t is

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{5}$$

with $\alpha > 0$ and $\alpha_j, \beta \geq 0$ and j is an integer larger than 1.

A very common expression in economics is volatility. Volatility describes the change in the variance compared to the mean. A period with high volatility means that the deviation from the mean is large in amplitude. Volatility in the GARCH model is actually just h_t. From equation4 we see that h_t is related to X_t^2. For a more economic explanation of volatility see [7].

We call this model the GARCH(p, q)-model which stands for general autoregressive conditional heteroscedasticity. Heteroscedasticity means that the finite variance of the process $\{X_t\}$ is not constant and the term can be interpreted as volatility. Conditional means that heteroscedasticity is not independent of time, hence where at a different time t the volatility is dependent on the volatility of the previous observed values. In this way, the GARCH model takes into account the previous change in errors.

The reason why we use the GARCH model as opposed to the ARCH model is due to a trait in the ARCH model that it may tend to burst as opposed to modeling a process with persistent fluctuations (volatility), where the GARCH model will model the fluctuations greater for longer than the ARCH model would. The extra term that differentiates the GARCH from the ARCH does contribute when modeling certain types of data.

When referring to the GARCH model we only refer to the X_t term that is now an error term as in equation 5. We will apply this error term in an ARMA model.

By choosing the GARCH$(1, 1)$ order model we obtain the following model using the R package "rugarch", [6].

As previously mentioned there is a criterion used to evaluate orders of the types of time series models. The one that is vastly used for GARCH is called AIC (Akaike Information Criterion). It has to be calculated after the estimation of parameters for a model with order p and q. Therefore I have manually computed the AIC for GARCH models with the use of R Studio. The model with the lowest AIC value is determined to be the best fit for the observed values that are at hand. For the mathematical equation of the AIC see [2] p.149.

After computing the AIC value for GARCH models we found the lowest AIC value to be obtained by the GARCH model with the order $(1, 1)$ and ARMA order $(5, 5)$. As we can see the ARMA order is the same as in the previous example 3.8.

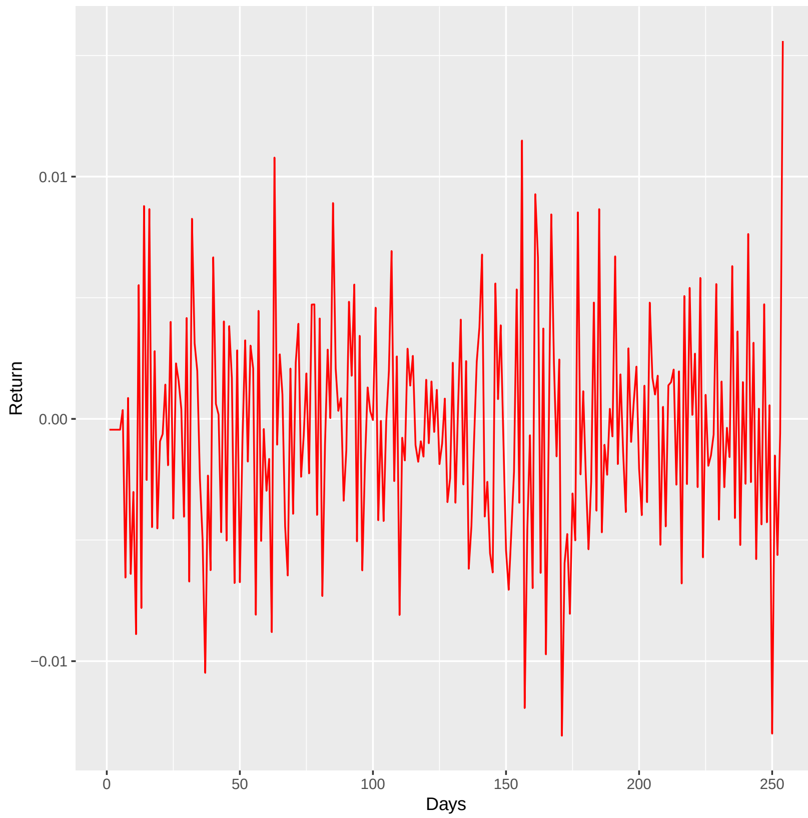

Figure 11: The predicted values of GARCH model.

The GARCH model does give more fluctuated data as we know from equation 5. We can also see that the fluctuations are heavy due to the dependence of the β_j. Thus, fluctuating patterns are followed by fluctuations over many days. The volatility is apparent. As the fluctuations roam heavily the dependence on the volatility for each day preserves the volatility even further.

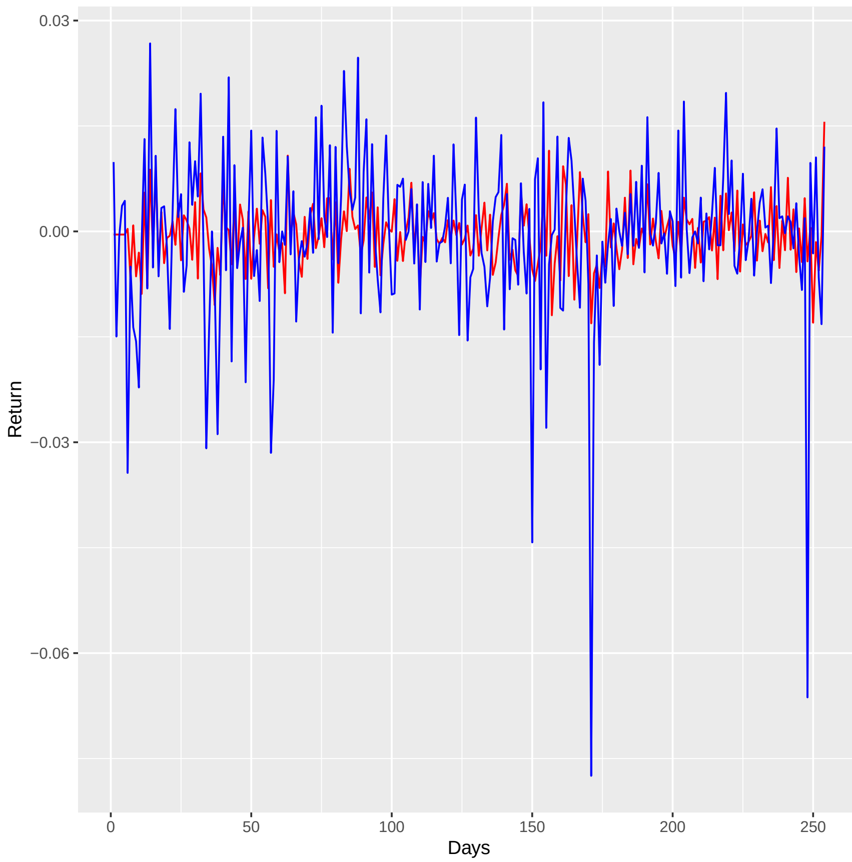

Figure 12: The predicted values of the GARCH model (red) vs. observed values of returns.

In figure 12 we see the predicted values of the GARCH model. The predicted values behave in a similar pattern with lower amplitude than observed returns. The amplitude of predicted values is following the volatility of the observed returns. We see also that the extreme values of observed returns are not followed with as high amplitude in predicted values this is due to the short interval of days that the extreme values occur.

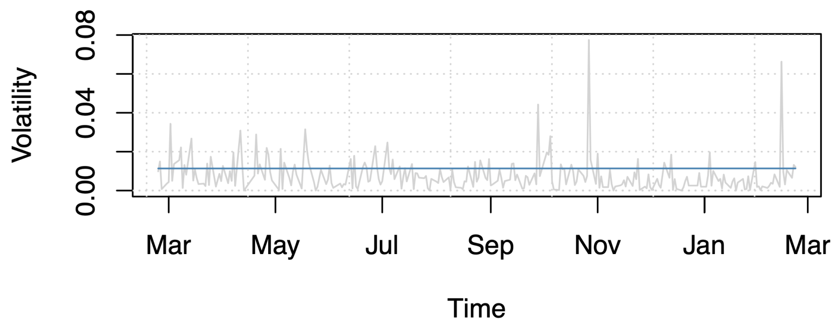

Figure 13: Conditional variance (h_t) plot of returns in GARCH model.

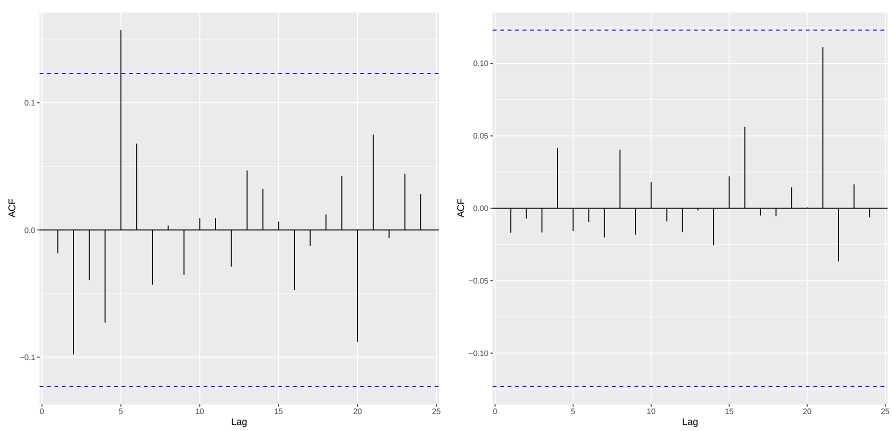

Figure 14: ACF of residuals (left), and ACF of squared residuals (right).

In our case we see in figure 14 in the ACF of residuals that we may have significant autocorrelation at lag 5. The significance of residuals points to a dependence between the residuals which in turn might make the extrapolated white noise sequence not entirely stationary (the extrapolated white noise sequence can be seen in figure 17). Hence the model might indicate that this not necessarily is the best model, for such a process. This can also indicate that the model has a systematic failure in capturing this change at lag 5. What we saw from the previous example with ARMA model is that the stock return exhibits a correlation on a weekly basis. This might be something that can be represented as a seasonal component.

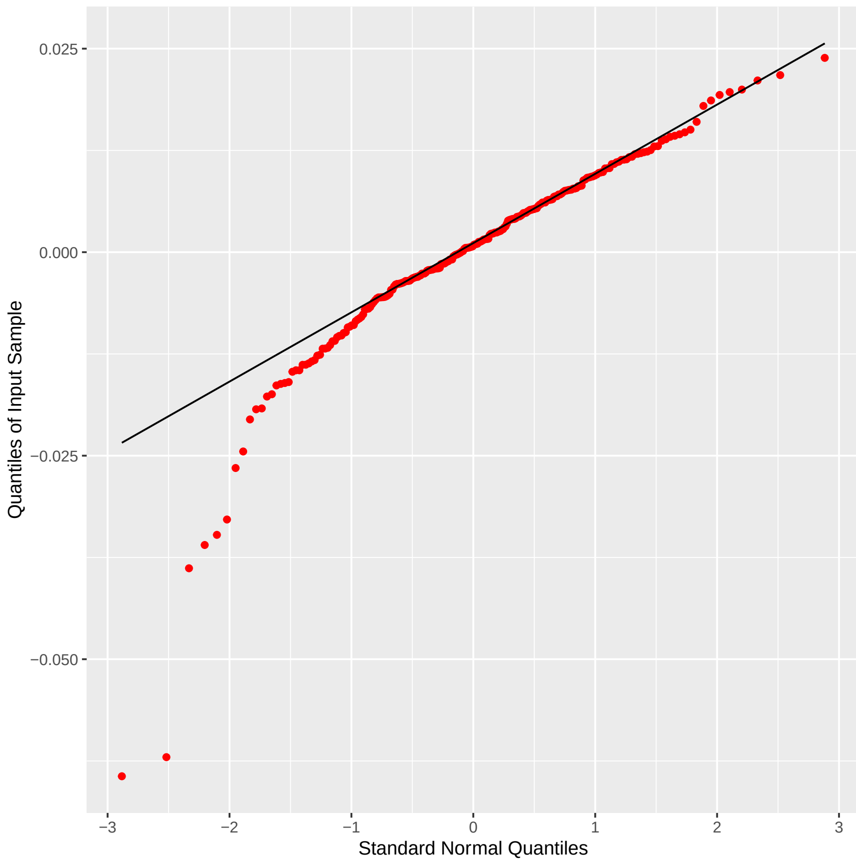

Figure 15: Sample quantiles plotted over theoretical standard normal distribution quantiles. A perfect normal distribution would see all the points on the line.

The diagnostics for GARCH model depends on the normality of residuals. We see here that although there are a few residuals that differ from the quantile-line the data appear to be approximately normally distributed.

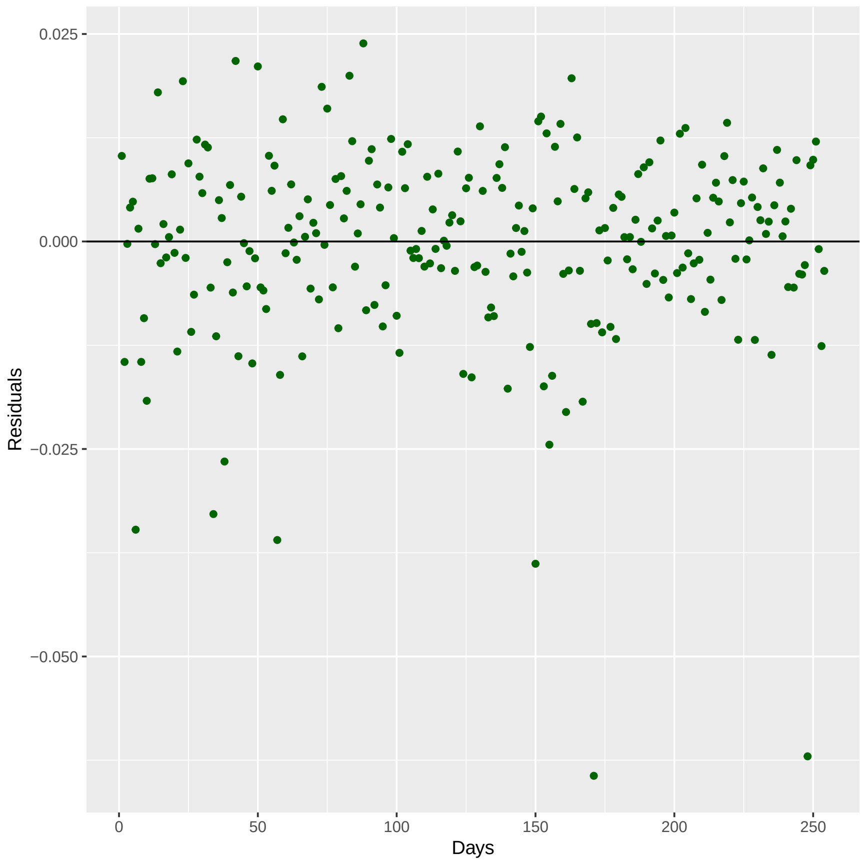

Figure 16: The residuals plotted over each day in index.

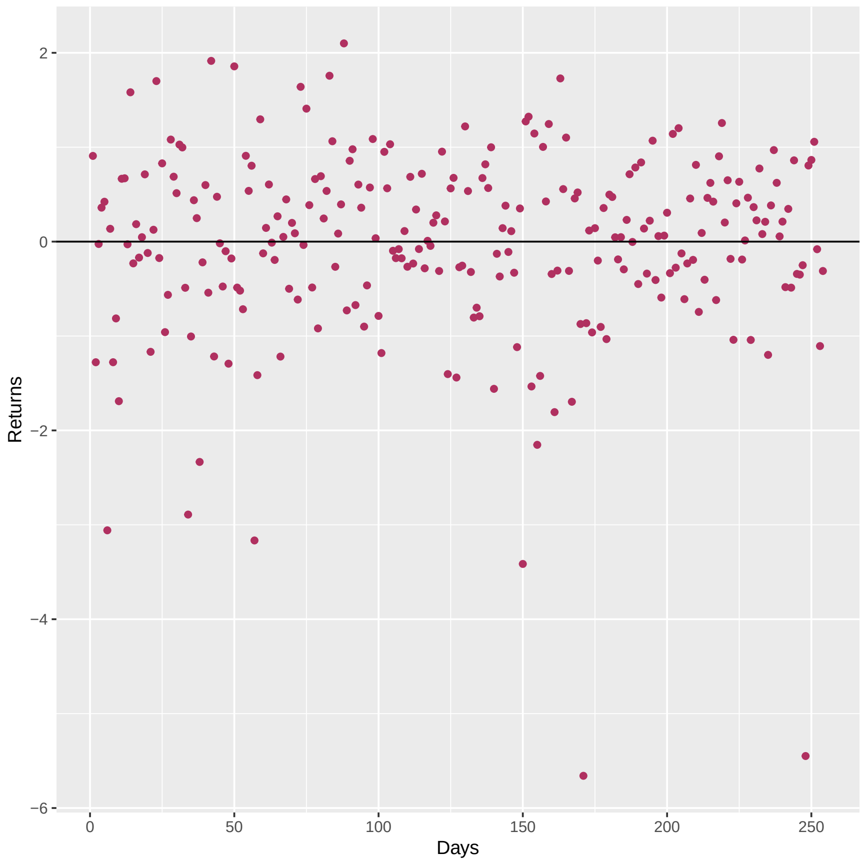

Figure 17: The extrapolated white noise sequence from the GARCH model.

As we know white noise is a stationary sequence, that would obey the properties from equation 2. While the residuals shouldn't exhibit any trend with constant variance. They appear to be as such.

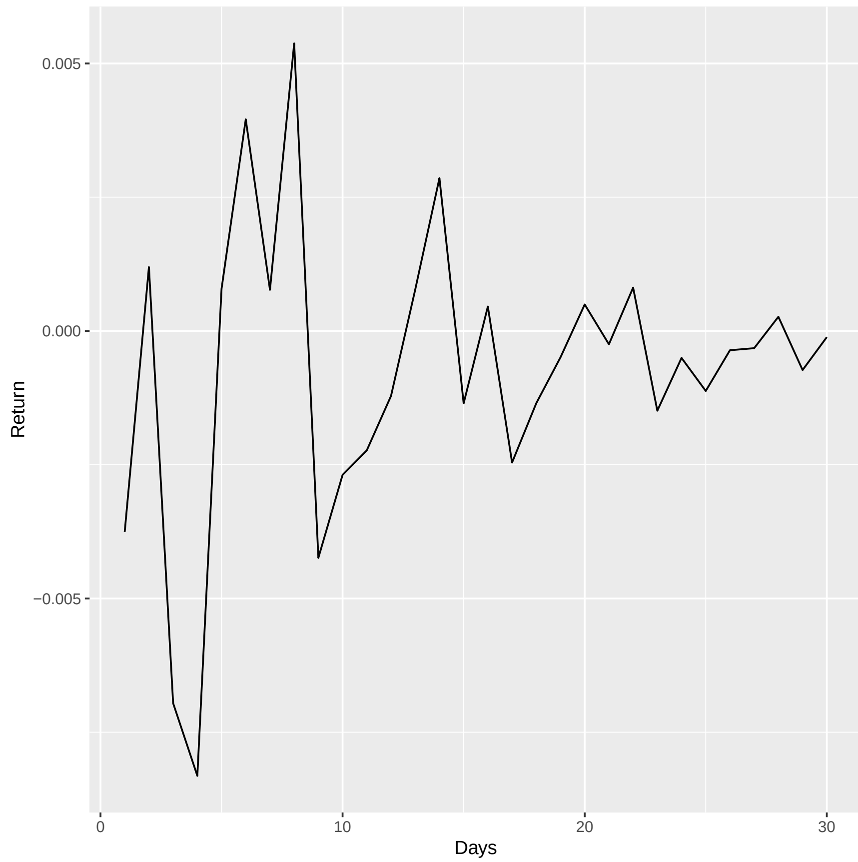

Figure 18: Forecasted values of relative return in GARCH model.

The forecasted values have a dependence on volatility but we see that the volatility is not upheld in the long term. A reason for that is the GARCH model will eventually 'flatten out' as the volatility is decreasing. It might be useful to only forecast in a short-term perspective after collected data. Although it may be more useful to increase the collection of days for the stock price.

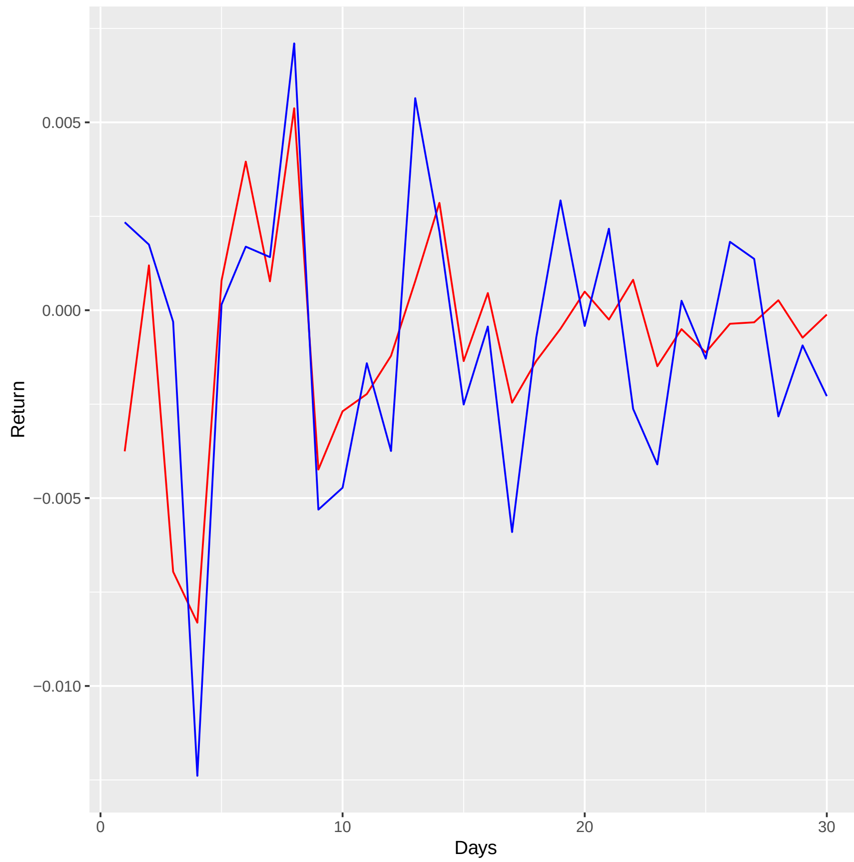

Figure 19: Forecasted values of GARCH model (red) compared to the ARMA (blue).

In figure 19 we see that actually the ARMA model will larger amplitudes at the extreme but they have ”bursty” tendencies. This means that when a fluctuated value is forecasted the ARMA model will ”usually” predict a value in the opposite direction very quickly as is seen for day 4 in figure 19. While the GARCH model has the capability to withstand the volatility after a day with high amplitude. On day 6 to 8, we see the volatility is upheld largely in the GARCH model, where the ARMA ”bursts” back.

5 Discussion

5.1 Choice of Data

The chosen data set is the stock price of a company called Orkla ASA. We will use the stock price for each day over a year. Orkla AS is a Norwegian grocery retail company. Although its biggest consumer is the Norwegian population it has market shares in neighbouring countries. For our purpose, this is a stock with very little volatility (4) over time. This means that the trend of this stock is small since the gradual increase or decrease of the stock price is low in contrast to other stocks. Yet we still postulate that we can model this as a stationary process. This is a very important aspect as it affects the credibility of the use of time series analysis.

The data is downloaded as a CSV file from Yahoo Finance [4]. From each day in the data set, we get an opening price and a closing price. We only use the closing price. In similar analyses of the price of a stock the closing price is used, or an average between the opening and closing price.

The chosen stock may not be volatile enough to see the effects of the GARCH model in the way volatility is exploited in some investment strategies. This is discussed further.

5.2 Evaluating the Models

In figure 3 and 4 there was a significant autocorrelation of residuals. This may be due to a seasonal component that could be differences, thus may be taken care of and improve the accuracy of the model.

5.3 Usage as Investment Tool

As a trader you try to take advantage of fluctuations, assuming that the returns of stock prices can be viewed as a stationary process. Modeling the fluctuations with previous values you can make an investment on a certain day and sell when your model is predicting an apex then selling. As seen by the figures of modeled returns (figure 12, 6) the returns are very small. Since fluctuations are very small in percentage terms. The way a trader can make money is by buying with a large position (amount). In this way, you could make a decent yield with small fluctuations in stock price. The downside of this investment strategy is that it is very risky. When investing a large amount, the loss will be much more disastrous if the price then doesn't act as predicted.

Another strategy that is more long-term is that we can take a larger amount of stocks and model them by comparing the fluctuations and at what value of t where X_t can be approximated to zero. This is what's called risk aversion, and securing assets for the long-term. Where you would, due to macroeconomic factors, invest in stocks that would give you the most predictable investment. Compared to the market the loss is less, than the downfall in the market.

6 Conclusion

The purpose of modeling with time series is to try and understand the underlying data. If using time series models can help make investment decisions then it should be used. The main weakness in time series is that the models may not represent the data, in the instance of ARMA models that the data are not stationary. Although the logic behind GARCH models, that a period with high volatility will be followed by a period with high volatility, does make sense. There is always a possibility that it doesn't. We therefore represent the data with models with confidence intervals, to account for this possibility with an uncertainty that the next observed value may actually differ from what the forecasted value actually predicts. This may be one of the limits to time series analysis since it is purely theoretical. Therefore it may not be the only tool to evaluate an investment but it, with other economic tools can be very effective to evaluate an investment.

7 Appendix

7.1 *Appendix I*

```
Call:
arima(x = returns, order = c(5, 0, 5))

Coefficients:
          ar1      ar2      ar3      ar4      ar5     ma1     ma2     ma3     ma4     ma5  intercept
      -0.3765  -0.4142  -0.3116  -0.8309  -0.0430  0.3538  0.4900  0.3468  0.9504  0.2124     -6e-04
s.e.   0.4611   0.1541   0.1963   0.1255   0.4123  0.4553  0.1073  0.2095  0.1098  0.4477      8e-04

sigma^2 estimated as 0.0001274:  log likelihood = 775.95,  aic = -1527.9
```

Figure 20: The coefficients of evaluated ARMA model in R.

7.2 *Appendix II*

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(5,0,5)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error     t value Pr(>|t|)
mu     -0.000446    0.000008 -5.3106e+01  0.00000
ar1    -0.970801    0.000273 -3.5499e+03  0.00000
ar2    -0.334251    0.000132 -2.5275e+03  0.00000
ar3    -0.328678    0.000100 -3.2849e+03  0.00000
ar4    -0.653168    0.000169 -3.8561e+03  0.00000
ar5    -0.573736    0.001286 -4.4626e+02  0.00000
ma1     1.055567    0.000242  4.3551e+03  0.00000
ma2     0.517465    0.000138  3.7565e+03  0.00000
ma3     0.550988    0.000162  3.3946e+03  0.00000
ma4     0.873316    0.000196  4.4516e+03  0.00000
ma5     0.856237    0.000214  4.0027e+03  0.00000
omega   0.000000    0.000001  2.0021e-01  0.84132
alpha1  0.000020    0.000709  2.8882e-02  0.97696
beta1   0.998854    0.000153  6.5349e+03  0.00000

Robust Standard Errors:
        Estimate  Std. Error     t value Pr(>|t|)
mu     -0.000446    0.000073 -6.1095e+00  0.00000
ar1    -0.970801    0.000342 -2.8378e+03  0.00000
ar2    -0.334251    0.000218 -1.5356e+03  0.00000
ar3    -0.328678    0.000389 -8.4520e+02  0.00000
ar4    -0.653168    0.001478 -4.4192e+02  0.00000
ar5    -0.573736    0.012253 -4.6826e+01  0.00000
ma1     1.055567    0.001493  7.0694e+02  0.00000
ma2     0.517465    0.000188  2.7556e+03  0.00000
ma3     0.550988    0.000514  1.0720e+03  0.00000
ma4     0.873316    0.000251  3.4727e+03  0.00000
ma5     0.856237    0.001462  5.8560e+02  0.00000
omega   0.000000    0.000017  8.9150e-03  0.99289
alpha1  0.000020    0.001981  1.0336e-02  0.99175
beta1   0.998854    0.001677  5.9552e+02  0.00000

LogLikelihood : 777.0266

Information Criteria
------------------------------------

Akaike       -6.0081
Bayes        -5.8131
Shibata      -6.0137
Hannan-Quinn -5.9296

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                    0.08624  0.7690
Lag[2*(p+q)+(p+q)-1][29]  14.40493  0.8447
Lag[4*(p+q)+(p+q)-1][49]  21.53615  0.8248
d.o.f=10
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
```

```
------------------------------------
                        statistic p-value
Lag[1]                    0.07399  0.7856
Lag[2*(p+q)+(p+q)-1][5]   0.32413  0.9811
Lag[4*(p+q)+(p+q)-1][9]   0.63598  0.9966
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]   0.07153 0.500 2.000  0.7891
ARCH Lag[5]   0.46704 1.440 1.667  0.8933
ARCH Lag[7]   0.55746 2.315 1.543  0.9729

Nyblom stability test
------------------------------------
Joint Statistic:  49.467
Individual Statistics:
mu     0.03585
ar1    0.03539
ar2    0.03519
ar3    0.03537
ar4    0.03421
ar5    0.03820
ma1    0.03622
ma2    0.03637
ma3    0.03628
ma4    0.03617
ma5    0.03603
omega  8.11632
alpha1 0.06823
beta1  0.05104

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         3.08 3.34 3.9
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
```

```
------------------------------------
                   t-value   prob sig
Sign Bias           0.7891 0.4308
Negative Sign Bias  0.4153 0.6783
Positive Sign Bias  0.1854 0.8531
Joint Effect        1.4274 0.6991


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     31.20      0.03841
2    30     49.31      0.01071
3    40     46.47      0.19165
4    50     69.23      0.03003
```

Figure 22: The coefficients of evaluated GARCH model.

References

[1] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of econometrics* 31.3 (1986), pp. 307–327.

[2] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. 3rd ed. Springer Texts in Statistics. Springer, 2016. ISBN: 978-3-319-29852-8.

[3] Robert F Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the econometric society* (1982), pp. 987–1007.

[4] Yahoo Finance. Retrieved from: https://finance.yahoo.com/quote/ORK.OL/history?p=ORK.OL. Downloaded: 23/02/2023, Date of stock prices: 22/02/2022 - 22/02/2023. 2023.

[5] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

[6] Alexios Ghalanos. *rugarch: Univariate GARCH models.* R package version 1.4-0. 2014.

[7] Adam Hayes. *Volatility: Meaning In Finance and How it Works with Stocks*. URL: `https://www.investopedia.com/terms/v/volatility.asp#toc-the-bottom-line`. (accessed: 01.04.2023).