Stockholms universitet

Forecasting the average global temperature anomaly for January 2020 using an ARIMA-ARCH model

Mathias Carlsson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2020:3
Matematisk statistik
Juni 2020

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2020:3**
http://www.math.su.se

Forecasting the average global temperature anomaly for January 2020 using an ARIMA-ARCH model

Mathias Carlsson*

June 2020

Abstract

In this thesis we are interested in forecasting the monthly average global temperature anomaly (in °C) for January 2020 using data from the *National Oceanic and Atmospheric Administration* (NOAA). The data contains 1680 temperature anomalies from January 1880 to December 2019 and are given in relation to the 20th century average. By first differencing the data, to obtain weakly stationary behaviour, three ARIMA models are chosen as candidates to forecast the temperature anomaly for January 2020: one with AR components, one with MA components and one including both types of components. Their respective forecasting capabilities are then compared, wherepon the winner's residuals are used to model the conditional variance using an ARCH model. Lastly, a forecast is made and the conditional variance is used to give a forecast interval. The result is a forecasted temperature anomaly of 0.989 °C, which is a slight decrease compared to the previous month. However, the forecast interval shows that an increase is also possible.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: mathe_97_carlsson@hotmail.com. Supervisor: Mathias Millberg Lindholm, Kristoffer Lindensjö and Felix Wahl.

Acknowledgements

I would like to thank my supervisors Mathias Millberg Lindholm, Kristoffer Lindensjö and Felix Wahl for their ideas, feedback and guidance. I would also like to thank my family for their support. A special thanks is directed to my dear friend Hilding Köhler, with whom I have had many valuable discussions.

Contents

1 Introduction

"Climate change has taken on political dimensions. That's odd because I don't see people choosing sides over $E = mc^2$ or other fundamental facts of science."

−Neil deGrasse Tyson

As a person living in the 21st century it is almost impossible to not come across the concept of global warming. With the fast progress and development of social medias it has certainly popped up on many radars. This has initiated discussions and caused gatherings, whether the people involved believe that humans have an impact on the global warming or not. Regardless the cause of global warming, there is still a scientific consensus that global temperatures are rising. See ([1]) for further reading about this.

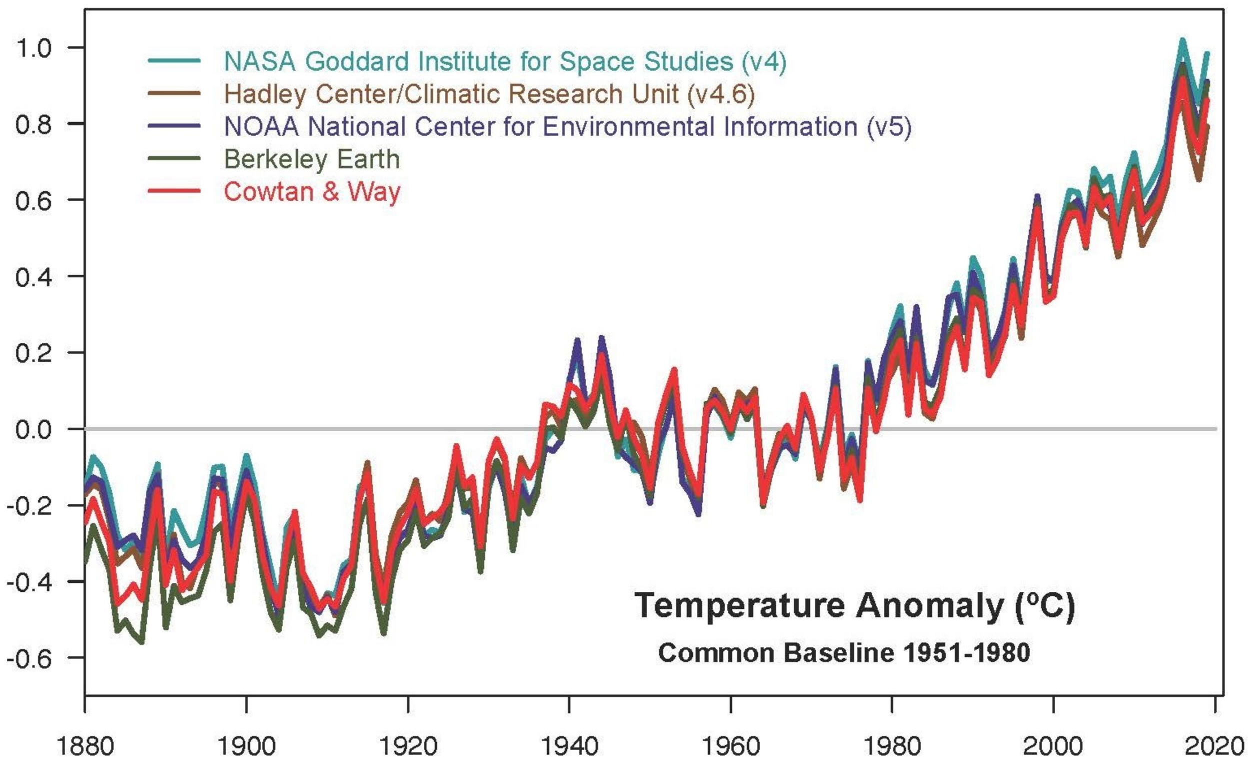

Figure 1: Illustration of the scientific consensus regarding global warming. Source: [1].

In this thesis we are interested in modelling the evolution of the global temperature. More specifically, we will try to forecast the average global temperature anomaly for January 2020 using monthly data from January 1880 to December 2019. The data set is taken from the *National Oceanic and Atmospheric Administration* organisation (NOAA) and consist of monthly average global temperature anomalies provided with respect to the 20th century average (1901-2000).

To start of we will make a mean model using an autoregressive integrated moving average model, where the time series consisting of the average global temperature anomalies is differenced once to create a time series with time-invariant mean. Three models will be compared: one with only autoregressive parameters (AR), one with only moving average parameters (MA) and one including both of the preceding. The purpose of the comparison is to determine if the differenced average global temperature anomaly is best described by an AR, MA or ARMA process. The forecast accuracy of these models is then measured using a rolling window of 1-step-ahead forecasts combined with the root mean squared error (RMSE) and the mean absolute error (MAE). We will see that forecasting with ARIMA(1,1,2) results in the smallest RMSE and MAE, and is therefore the model used to forecast the average global temperature anomaly for January 2020.

Lastly, a forecast interval for the temperature anomaly is given. With a McLeod-Li test showing signs of conditional variance an ARCH(26) model is fitted to the residuals of the ARIMA(1,1,2) model. The forecast interval is then calculated by using the conditional variance estimation. The result is a forecasted temperature anomaly of 0.989°C on January 2020, which is smaller than the temperature anomaly for December 2019, having measured 1.05°C. However, the forecast interval shows that the true anomaly on January 2020 may in fact be larger than 1.05°C.

2 Data description

The data set used in this thesis is taken from NOAA ([2]), which is an american scientific agency within the U.S. department of commerce whose goal is to be a trustworthy authority on climate and historical weather information. The set consists of monthly average global temperature anomalies (in degrees Celsius) between January 1880 and December 2019, provided with respect to the 20th century average (1901-2000). Each anomaly is calculated by using a data set containing the monthly average temperature anomaly for $5° \times 5°$ grids across land and ocean surfaces. Averaging these grids for each month provides the monthly average global temperature anomalies ([2]).

The temperature anomalies between January 1880 and December 2019 constitute 1680 observations. For convenience we will henceforth let $\{1, 2, .., 1680\}$ be the index set denoting each month, where 1 denotes January 1880 and 1680 denotes December 2019. Furthermore, all hypothesis tests will be performed at a five percent significance level.

3 Theoretical background

The following section goes through the essential theoretical parts needed to perform a time series analysis of the temperature anomalies. We would like to point out that this section largely consists of parts from the book *Analysis of Financial Time Series*, written by Tsay ([3]).

3.1 Theoretical aspects of time series modelling

3.1.1 Mathematical definition of a time series

A time series is a time-ordered series obtained by observing a variable over time, usually over equally spaced time intervals. Mathematically, a real valued time series is a stochastic process defined by a probability space $(\Omega, \mathcal{F}, P)$ and an index set T, where

- Ω is the set of all possible individual outcomes, ω.
- $\mathcal{F}$ is a sigma-algebra containing events given as subsets of Ω.
- P is a function defined on $\mathcal{F}$, assigning probabilities to events.

Consider the function $Y(t, \omega) : T \times \Omega \to \mathbb{R}$. For each t, this function constitutes a stochastic variable defined on a probability space $(\Omega, \mathcal{F}, P)$. A time series can be seen as a collection of such stochastic variables, $\{Y_t, t \in T\}$, where Y_t is just an abbreviation of $Y(t, \omega)$ ([4], p. 1-3) .

3.1.2 Linear time series and white noise

A time series $\{Y_t\}$ is called linear if it can be written as

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \tag{1}$$

where μ is the expected value of the Y_t and $\{\varepsilon_t\}$ is a *white noise series*, meaning that it is a sequence of independent and identically distributed stochastic variables with expected value zero and variance σ^2. It must hold that ψ_i^2 goes to 0 as i goes to infinity in order for the variance of Y_t to be finite. This is important in order for the time series to be *weakly stationary*, which is a concept defined in the next section ([3], p. 36).

In this thesis we are only concerned with linear time series and will examine if white noise series are normally distributed. Henceforth, an index set T will be on the form $\{1, 2, ..., N\}$, where N is the number of observations.

3.1.3 Stationarity

Let $\{Y_t, t \in T\}$ be a real valued time series. An important aspect of analysing time series is the assumption of stationarity. The time series $\{Y_t\}$ is called *strictly stationary* if the joint distribution of $(Y_{t_1}, Y_{t_2}, ..., Y_{t_k})$ is time invariant. This means that, for an arbitrary k, the equality

$$F_{Y_{t_1}, Y_{t_2}, ..., Y_{t_k}}(y_{t_1}, y_{t_2}, ..., y_{t_k}) = F_{Y_{t_1+t}, Y_{t_2+t}, ..., Y_{t_k+t}}(y_{t_1}, y_{t_2}, ..., y_{t_k}),$$

must hold for all indices $t_1, t_2, ..., t_k$, not necessarily consecutive. Strict stationarity implies that the distribution of Y_t is the same for all t in the index set T, which tells us that the expected value and the variance of Y_t are the same for all t under the assumptions that $E(|Y_t|) < \infty$ and $E(Y_t^2) < \infty$. Since the distribution function is rarely known in real applications a weaker version of stationarity is often used ([4], p. 3-4).

The time series $\{Y_t\}$ is called *weakly stationary* if the expected value of Y_t and the covariance between Y_t and Y_{t-l} are time invariant. That is if

- $\mathrm{E}(Y_t) = \mu$ is constant,
- $\mathrm{Cov}(Y_t, Y_{t-l}) = \gamma_l$ depends only on l,

with l being an arbitrary integer.

Note that a strictly stationary time series with finite first and second moment is also weakly stationary, but the converse is not necessarily true. The only exception is when the k-dimensional random vector $Y_{t1}, Y_{t2}, ..., Y_{tk}$ follows a multivariate normal distribution. In this thesis we will make use of properties obtained by assuming weak stationarity. For this assumption to hold, it is important that the data fluctuate with constant variance around the fixed level μ when plotted against time ([3], p. 30). Many time series are clearly not stationary in practice. Economic indicators often show trends through time, which means that the mean is time-dependent ([4], p. 4).

3.1.4 Autocorrelation function - ACF

Suppose that we have observed N data points and have the index set $T = \{1, 2, ..., N\}$. Recall from the previous section that the value of $\gamma_l = \mathrm{Cov}(Y_t, Y_{t-l})$ only depends on the lag, l. The covariance γ_l is called the *lag-l autocovariance*. Similarly, the correlation between Y_t and Y_{t-l} is called the *lag-l autocorrelation* and is given by

$$\rho_l = \frac{\mathrm{Cov}(Y_t, Y_{t-l})}{\sqrt{\mathrm{Var}(Y_t)\,\mathrm{Var}(Y_{t-l})}} = \frac{\mathrm{E}[(Y_t - \mathrm{E}(Y_t))(Y_{t-l} - \mathrm{E}(Y_{t-l}))]}{\sqrt{\mathrm{E}[(Y_t - \mathrm{E}(Y_t))^2]\,\mathrm{E}[(Y_{t-l} - \mathrm{E}(Y_{t-l}))^2]}} = \frac{\mathrm{E}[(Y_t - \mu)(Y_{t-l} - \mu)]}{\sqrt{\mathrm{E}[(Y_t - \mu)^2]\,\mathrm{E}[(Y_{t-l} - \mu)^2]}},$$

where we have used the assumption that $\{Y_t\}$ is a weakly stationary time series such that $E(Y_t) = \mu$ for all t in the index set. Here, ρ_l as function of l is called the *autocorrelation function* (ACF). Weak stationarity also implies that $\mathrm{Var}(Y_t) = \mathrm{Cov}(Y_t, Y_t) = \gamma_0$. Thus we can write the ACF as

$$\rho_l = \frac{\gamma_l}{\gamma_0}.$$

Since the true values of μ and γ_0 are not known, the ACF must be approximated with observed data. This leads to the definition of the *lag-l sample autocorrelation*,

$$\hat{\rho}_l = \frac{\hat{\gamma}_l}{\hat{\gamma}_0} \text{ for } 0 \leq l \leq N-1, \tag{2}$$

where $\hat{\gamma}_l = \frac{1}{N} \sum_{t=l+1}^{N} (y_t - \bar{y})(y_{t-l} - \bar{y})$ is the lag-l sample covariance, $\bar{y} = \frac{1}{N} \sum_{t=1}^{N} y_t$ is the sample mean and N is the sample size ([3], p. 30-31).

3.1.5 Partial autocorrelation function - PACF

Consider a weakly stationary time series $\{Y_t\}$. The *lag-l partial autocorrelation* of Y_t, ϕ_l, is the correlation between Y_t and Y_{t-l} after removing the linear dependence of the intermediate variables $Y_{t-1}, Y_{t-2}, ..., Y_{t-l+1}$. Thus, the lag-$l$ partial autocorrelation is given by the correlation between the two following variables.

Variable 1: The variance of Y_t that is not explained by $Y_{t-1}, Y_{t-2}, ..., Y_{t-l+1}$.
Variable 2: The variance of Y_t that is not explained by $Y_{t-1}, Y_{t-2}, ..., Y_{t-l+1}$.

Using ordinary least squares, we can acquire the best linear predictor of Y_t and Y_{t-l} by regressing both variables on $Y_{t-1}, Y_{t-2}, ..., Y_{t-l+1}$. Denoting these predictors with $\hat{Y}_t$ respectively $\hat{Y}_{t-l}$, the *lag-l partial autocorrelation* of $\{Y_t\}$ can then be defined as

$$\phi_l = \text{Cor}(\textbf{Variable 1}, \textbf{Variable 2}) = \text{Cor}(Y_t - \hat{Y}_t, Y_{t-l} - \hat{Y}_{t-l}), \tag{3}$$

with $\phi_0 = 1$ and $\phi_1 = \rho_1 = \text{cor}(Y_t, Y_{t-1})$. Here, ϕ_l as function of l is called the *partial autocorrelation function* (PACF) ([5], p. 68-69).

3.1.6 Back-shift operator

The back-shift operator, denoted by B, is an operator that shifts back the time index one unit. Applying it on a time series $\{Y_t, t \in T\}$ yields the lagged series

$$B\{Y_t, t \in T\} = \{Y_{t-1}, t \in T\}.$$

Since a constant, c, can be seen as a time series with $Y_t = c$ for all t, it holds that $Bc = c$. It is easily verified that the back-shift operator is a linear operator. Consider a time series $\{Y_t = aX_t + bZ_t + c\}$, where $\{X_t\}$ and $\{Z_t\}$ are two arbitrary time series and a, b and c are constants. Applying the back-shift operator gives

$$B\{aX_t + bZ_t + c\} = B\{Y_t\} = \{Y_{t-1}\} = \{aX_{t-1} + bZ_{t-1} + c\}.$$

Thus it holds that $B(aX_t + bZ_t + c) = aBX_t + bBZ_t + Bc$, showing that it is a linear operator ([6], p. 202).

3.1.7 Autoregressive model - AR(p)

An autoregressive model, denoted $AR(p)$, is defined by

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + ... + \varphi_p Y_{t-p} + \varepsilon_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t, \tag{4}$$

where p is a positive integer denoting the amount of lags Y_t depends on and $\{\varepsilon_t\}$ follows a white noise series. Thus, the current value in this model is seen as a linear combination of the previous p lags and a random shock ε_t. Using the weak stationarity condition we conclude that

$$\mathrm{E}(Y_t) = \frac{\varphi_0}{1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p}. \tag{5}$$

Using the back-shift operator, B, we can rewrite (4) as

$$\phi(B)Y_t = \varphi_0 + \varepsilon_t, \tag{6}$$

where $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 ... - \varphi_p B^p$ is called the *AR(p) polynomial*. The equation given by $\phi(B) = 0$ is called the *characteristic equation*. In order for the process to be a weakly stationary we must be able to write it as a linear combination of the previous shocks where the coefficients constitute a convergent series, that is on the form of (1). Using (5) to get that $\varphi_0 = \mu(1 - \varphi_1 - ... - \varphi_n)$ we can rewrite (6) as

$$\tilde{Y}_t = \phi(B)^{-1}\varepsilon_t, \tag{7}$$

where $\tilde{Y}_t = Y_t - \mu$. Thus (7) is a weakly stationary linear time series if $\phi(B)^{-1}$ has a convergent series expression in powers of B. This is possible if the solutions of the characteristic equation are greater than one in modulus. The proof of this is however left out. If a solution to the characteristic equation lies on the unit circle, the AR(p) process is said to be *unit-root nonstationary* ([3], p. 37-46).

Note that the autocovariances of a stationary AR process can be obtained through tedious calculations. These expressions are, however, not of great value to our analysis, but are merely a way to demonstrate that the series is covariance stationary. Therefore we choose not to write them out.

3.1.8 Moving average model - MA(q)

A time series defined by

$$Y_t = c_0 + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ... - \theta_q\varepsilon_{t-q} = c_0 + \sum_{i=0}^{q} \theta_i\varepsilon_{t-i} \tag{8}$$

is called a *moving average process of order q* and the model is denoted with MA(q). Here, c_0 is a constant and $\{\varepsilon_t\}$ a white noise series. Just by definition, it is easily seen that the MA(q) model is a linear time series, since the form of the model follows the one in (1) with $\psi_i = \theta_i$ for $i > 0$ and $\psi_0 = 1$.

MA processes, being finite linear combinations of a white noise series, are always weakly stationary. Thus the first two moments are time invariant. Taking expectation of (8), we get

$$\mathrm{E}(Y_t) = c_0 + \sum_{i=0}^{q} \theta_i E(\varepsilon_{t-i}) = c_0. \tag{9}$$

Taking the variance of the same equation gives

$$\mathrm{Var}(\mathrm{Y_t}) = \sum_{i=0}^{q} \theta_i^2 \,\mathrm{Var}(\varepsilon_{t-i}) = (1 + \theta_1^2 + ... + \theta_q^2)\sigma^2. \tag{10}$$

Clearly (9) and (10) show that the first and second moment of the time series are time invariant. The same also applies to $\gamma_l = \text{Cov}(Y_t, Y_{t-l})$, but we refer to ([3], p. 57-60) for more reading on this topic.

Using the back-shift operator once again, such that $B\varepsilon_t = \varepsilon_{t-1}$, we can rewrite (8) as

$$Y_t = c_0 + \theta(B)\varepsilon_t,$$

where $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$ is called the *MA(q) polynomial* ([3], p. 57-59).

An MA(q) process is said to be *invertible* if the current error term, ε_t, can be written as an infinite linear combination of its past values where the coefficients constitute a convergent series. That is, $Y_t = c_0 + \theta(B)\varepsilon_t$ can be written as

$$\varepsilon_t = \theta(B)^{-1}\tilde{Y}_t,$$

where $\tilde{Y}_t = Y_t - c_0$. An MA(q) process is thus invertible if $\theta(B)^{-1}$ can be represented as a convergent series in powers of B. This is possible if the roots of the MA(q) polynomial are larger than one in modulus ([5], p. 71-72).

3.1.9 Autoregressive moving average model - ARMA(p,q)

An ARMA(p,q) model includes both autoregressive and moving average terms, giving more flexibility when fitting the time series. It has the form

$$Y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

where p and q are non-negative and $\{\varepsilon_t\}$ is a white noise series. Using the back-shift operator, we can write it as

$$\phi(B)Y_t = \varphi_0 + \theta(B)\varepsilon_t, \tag{11}$$

with $\phi(B)$ and $\theta(B)$ once again denoting the AR(p) and MA(q) polynomial. It is required that there are no common factors between the two polynomials, since this implies that the order of p and q can be reduced ([3], p. 66). This is not easily verified in reality but will be assumed to hold throughout the analysis. A weakly stationary ARMA process has the same expected value as (5), meaning that φ_0 can be written as $\mu(1 - \varphi_1 - \cdots - \varphi_p)$. Plugging this in to (11) gives

$$\phi(B)\tilde{Y}_t = \theta(B)\varepsilon_t, \tag{12}$$

where $\{\tilde{Y}_t\} = \{Y_t - \mu\}$. As for the AR($p$) process, (12) defines a stationary ARMA(p,q) process if the solutions to the characteristic equation, (7), all lie outside the unit circle. If the equation has the unit as a solution, the process is said to be unit-root nonstationary. Correspondingly, the process is called invertible if the roots of the MA(q) polynomial all lie outside the unit circle ([5], p. 79-80). An important aspect of the stationary and invertible ARMA(p,q) process is that it can be represented by an infinite AR process by writing (12) as

$$\frac{\phi(B)}{\theta(B)}\tilde{Y}_t = \varepsilon_t, \tag{13}$$

which implies that it can be approximated by an AR(p) model process for some p ([5], p. 79-80). We will return to this when testing for unit-root nonstationarity.

3.1.10 Order of integration - I(d)

To define the *order of integration*, it is necessary to return to unit-root nonstationarity. Recall that unit-root nonstationarity is connected to the roots of the AR(p) polynomial. With one as a root to the polynomial, $\phi(B)$ can be written as

$$\phi(B) = \varphi(B)(1-B),$$

where $\varphi(B)$ is the polynomial of a stationary AR($p-1$) process. Note that the ARMA process

$$\varphi(B)(1-B)\tilde{Y}_t = \theta(B)\varepsilon_t,$$

where $\tilde{Y}_t = Y_t - \mu$, is unit-root nonstationary. But defining the time series $W_t = (1-B)\tilde{Y}_{t+1} = \tilde{Y}_{t+1} - \tilde{Y}_t = Y_{t+1} - Y_t$ gives us the stationary AR($p-1$,q) process

$$\varphi(B)W_t = \theta(B)\varepsilon_t.$$

Thus, stationarity is obtained by *differencing* the series $\{Y_t\}$ once. Similarly, the problem of d roots being equal to one, $\phi(B) = \varphi(B)(1-B)^d$, is solved by differencing the series d times, that is by letting $W_t = (1-B)^d Y_t$. The minimum number of differences required for the time series $\{Y_t\}$ to become weakly stationary is called the *order of integration* and is denoted $I(d)$ ([5], p. 11-12).

3.1.11 Autoregressive integrated moving average model - ARIMA(p,d,q)

The ARIMA(p,d,q) model is a combination of the ARMA(p,q) model and I(d). A time series $\{Y_t\}$ is said to follow an ARIMA(p,d,q) process if

$$W_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i W_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \tag{14}$$

where p and q specify the order of the ARMA process and $W_t = (1-B)^d Y_{t+1}$. The inverse of this differencing is given by

$$Y_t = (1-B)^{-d} W_t = S^d W_t,$$

where $S = (1-B)^{-1} = 1 + B + B^2 + ...$ and $SW_t = \sum_{i=0}^{\infty} W_{t-i}$ ([5], p. 11-12).

3.1.12 Autoregressive conditional heteroscedastic model - ARCH(m)

Conditional heteroscedasticity is present in a time series $\{Y_t\}$ if the conditional variance (or standard deviation) changes with time. In mathematical terms, we consider

$$\sigma_t^2 = \mathrm{Var}(Y_t|F_{t-1}),$$

where F_{t-1} is all the information that is available at time $t-1$. By modelling the conditional variance one can improve the accuracy of forecast intervals. Consider an arbitrary ARMA model with conditional heteroscedasticity. The conditional variance of this model is given by $\sigma_t^2 = \mathrm{Var}(Y_t|F_{t-1}) = \mathrm{Var}(\varepsilon_t|F_{t-1})$. Thus, when modelling the conditional variance of such a model, one makes use of the error terms series instead of the normal data set. An important aspect of conditional heteroscedasticity being present is that the conditional variance varies within some fixed range, since it in that case can be seen as a stationary process([3], p. 109-113).

One model that tries to explain the evolution of the conditional variance is the *autoregressive conditional heteroscedastic* model (ARCH). The idea behind ARCH is that the error terms ε_t are uncorrelated but dependent through a function of their squared values. The ARCH(m) model is given by

$$\begin{aligned} &\varepsilon_t = \sigma_t u_t, \\ &\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_{t-m}\varepsilon_{t-m}^2 = \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2, \end{aligned} \tag{15}$$

where u_t is a white noise series with variance equal to one and σ_t^2 is the conditional variance. To guarantee non-negativity of σ_t^2 the restrictions $\alpha_i \geq 0$ are imposed. Consider an ARCH(1) process. The unconditional mean is given by

$$\mathrm{E}(\varepsilon_t) = \mathrm{E}(\mathrm{E}(\varepsilon_t|F_{t-1})) = \mathrm{E}(\sigma_t\,\mathrm{E}(u_t|F_{t-1})) = 0,$$

and assuming stationarity of ε_t the unconditional variance is given by

$$\mathrm{Var}(\varepsilon_t) = \frac{\alpha_0}{1-\alpha_1},$$

where we have used $\mathrm{Var}(\varepsilon_t) = \mathrm{E}(\varepsilon_t^2) = \mathrm{E}(\mathrm{E}(\varepsilon_t^2|F_{t-1})) = \mathrm{E}(\sigma_t^2 E(u_t^2)) = \mathrm{E}(\alpha_0 + \alpha_1\varepsilon_{t-1}^2) = \alpha_0 + \alpha_1\,\mathrm{Var}(\varepsilon_t)$ and required that $0 \leq \alpha_1 < 1$ in order to have positive variance. Thus, the ARCH(1) process has a constant unconditional variance while having a time-varying conditional variance ([6], p. 280-283).

3.1.13 ARIMA-ARCH model

The purpose of ARIMA models is to capture the evolution of the conditional mean, while the purpose of ARCH models is to do the same with the conditional variance. This gives rise to a two-way model analysing both the conditional mean and variance, called the ARIMA-ARCH model. It is jointly defined by (14) and (15), where the error term from the ARIMA model follows an ARCH model. If all the roots of the AR polynomial lie outside the unit circle, the ARIMA-ARCH process $\{Y_t\}$ is weakly stationary if $\{\varepsilon_t\}$ is weakly stationary.

Estimation of the parameters in an ARMA-ARCH model can be done in two steps. First, remove the linear dependence by estimating the parameters of an ARMA model. Then, using the residuals from the previous estimation, fit an ARCH model ([6], p. 293-294).

3.2 Model fitting and selection rules

3.2.1 Maximum likelihood estimate - MLE

Maximum likelihood estimation is the method that will be used to estimate coefficients in our ARIMA models. The *likelihood function* plays a central role in this and is defined by

$$L(\theta; x) = f(x; \theta). \tag{16}$$

The likelihood function is seen as a function of the unknown parameters θ and is estimated from a realisation x of a stochastic variable X. For a shorter notation, we will denote the likelihood function with $L(\theta)$.

The maximum likelihood estimate (MLE) $\hat{\theta}_{ML}$ of θ is given by

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta)$$

It is often more convenient to look at the *log-likelihood function*

$$l(\theta) = \log(L(\theta)),$$

which is the natural logarithm of the likelihood function. Since the natural logarithm is a strictly monotone function, it holds that

$$\hat{\theta}_{ML} = \arg\max_{\theta} l(\theta).$$

Note that multiplicative constants in the likelihood function turn into additive constants when applying the natural logarithm, which leads to more straightforward calculations ([7], p. 13-15).

Some statistical packages used in this thesis calculate the likelihood function of ARIMA and ARCH models using conditional maximum likelihood, and one important note is that the packages use state-space representations of ARIMA processes. For interested readers, conditional maximum likelihood is explained in Appendix A.2. We will not, however, go further into state space representations of ARIMA processes.

3.2.2 Akaike's information criterion - AIC

There are multiple selection criteria used to determine the order of an ARIMA model. A common choice is *Akaike's information criterion*

$$AIC = -2l(\hat{\theta}_{ML}) + 2k.$$

Here, $l(\hat{\theta}_{ML})$ is the log-likelihood function evaluated in the MLE and k is the number of estimated parameters. The model minimising the AIC is chosen ([7], p. 224).

A downside to AIC is that it may perform worse when the number of parameters is large in relation to the sample size, leading to the selection of a model with too many parameters. This problem gave birth to an

AIC variant called AICC, which further penalizes models with a large number of parameters. The AICC is given by

$$AICC = AIC + \frac{2k(k+1)}{N-k-1},$$

with N being the sample size. Note that AICC is asymptotically equal to AIC as the sample size goes to infinity, which means that the AIC and AICC will tend to select the same model for sufficiently large sample sizes. To avoid the problem of selecting a model with more parameters than necessary we will only make use of the AICC ([8], p. 66).

3.2.3 Identifying the order of AR(p)

Since the order of a real stationary AR process is unknown, empirical methods can be used to determine it. It can be shown that the ACF of an AR process goes towards zero as the lag, l, goes to infinity. Thus, the ACF is not a good way of deciding the order.

The PACF is on the other hand useful. Let

$$\begin{aligned}
&Y_t = \varphi_{0,1} + \varphi_{1,1}Y_{t-1} + \varepsilon_{1,t} \\
&Y_t = \varphi_{0,2} + \varphi_{1,2}Y_{t-1} + \varphi_{2,2}Y_{t-2} + \varepsilon_{2,t} \\
&Y_t = \varphi_{0,3} + \varphi_{1,3}Y_{t-1} + \varphi_{2,3}Y_{t-2} + \varphi_{3,3}Y_{t-3} + \varepsilon_{3,t} \\
&\vdots
\end{aligned}$$

be the AR models with an order larger than zero, where $\varphi_{i,j}$ is the ith coefficient in an AR(j) model. It turns out that the lag-l partial autocorrelation, ϕ_l, of an AR(p) process is given by $\varphi_{l,l}$. The lag-l partial autocorrelation, according to the definition in (3), is interpreted as the correlation between Y_{t-l} on Y_t after removing the linear dependence of $Y_{t-l+1}, ..., Y_{t-1}$. Thus, the PACF for an AR(p) process should in theory be zero for $l > p$.

The *lag-l sample partial autocorrelation* is given by $\hat{\varphi}_{l,l}$, and is an estimation of $\varphi_{l,l}$ obtained by fitting an AR model of order l. Since the PACF is zero for $l > p$, the sample PACF should be close to zero for all $l > p$. In fact, for a weakly stationary Gaussian AR(p) model it holds that

- $\hat{\varphi}_{p,p}$ converges to φ_p when the sample size goes to infinity
- $\hat{\varphi}_{l,l}$ converges to zero for all $l > p$,

which are useful points for deciding the order ([3], p. 46-47).

On the hypothesis that the data follow an AR(p) process, the lag-l sample partial autocorrelation, for $l > p$, approximately follows a normal distribution with expected value zero and variance $\frac{1}{N}$. Thus, individual PACF values can be tested with $H_0 : \varphi_{l,l} = 0$ against $H_1 : \varphi_{l,l} \neq 0$ ([5], p. 70).

3.2.4 Identifying the order of MA(q)

Unlike an AR process, the PACF of an MA process does not equal zero for $l > q$. Instead, the PACF goes to zero as the number of lags goes to infinity. The ACF, on the other hand, is a useful tool for identifying the order. For a time series following an MA(q) process it holds that $\rho_q \neq 0$ and $\rho_l = 0$ for $l > q$. Thus, the sample ACF, given by (2), should not be equal to zero for $l = q$, but be close to zero for all $l > q$. We demonstrate this property for an MA(1) model,

$$Y_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1}. \tag{17}$$

Multiplying (17) by Y_{t-l} gives

$$Y_{t-l} Y_t = Y_{t-l} c_0 + Y_{t-l} \varepsilon_t - \theta_1 Y_{t-l} \varepsilon_{t-1}. \tag{18}$$

Using the stationarity condition together with (9) it is easily shown that $\gamma_l = \mathrm{Cov}(Y_t, Y_{t-l}) = \mathrm{E}(Y_t, Y_{t-l}) - \mu^2 = \mathrm{E}(Y_t, Y_{t-l}) - c_0^2$. Now, taking the expectation of (18) and using the fact that $\gamma_l = E(Y_t, Y_{t-l}) - c_0^2$, we get

$$\gamma_0 = (1 + \theta_1^2)\sigma^2, \qquad \gamma_1 = -\theta_1 \sigma^2 \qquad \text{and} \qquad \gamma_l = 0 \qquad \text{for} \qquad l > 1. \tag{19}$$

Since the ACF is given by dividing (19) by γ_0, this shows that the ACF is equal to zero for l larger than one ([3], 59-60).

A useful fact when looking at the ACF is that with the assumption of $\{Y_t\}$ being an independent and identically distributed sequence with finite second moment, $\hat{\rho}_l$ is asymptotically normal with expected value zero and variance $\frac{1}{N}$ for all positive lags. This can be used to test for significant autocorrelations using $H_0 : \rho_l = 0$ against the alternative $H_1 : \rho_l \neq 0$ ([3], p. 31).

3.2.5 Identifying the order of ARMA

First, we would like to point out that this section also applies to ARIMA models where the order of integration has been chosen, since the differenced data follows a stationary ARMA process. Choosing the right value of d is another topic discussed later.

Both the ACF and PACF are non-informative when it comes to determining the order of an ARMA model. It can namely be shown that the ACF and PACF of an ARMA process do not cut off at any lag. Instead, both go to zero as the lag goes to infinity. This is, however, no problem since the ARMA model with lowest AICC will be chosen.

3.3 Model checking

This section covers methods and tests to check whether a fitted model satisfies the model conditions. Since weakly stationary processes are sought, it includes tests for autocorrelation, conditional heteroscedasticity and unit-root nonstationarity.

3.3.1 Ljung-Box test

The *Ljung-Box test* tests if several autocorrelations of a time series $\{Y_t\}$ are jointly equal zero. The Ljung-Box test considers the null hypothesis $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$ against the alternative hypothesis $H_1 : \rho_i \neq 0$ for some $i \in \{1, 2, \ldots, m\}$. The test statistic is given by

$$Q(m) = N(N+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{N - l},$$

where m is the number of joint tests to be performed and $\hat{\rho}_l$ is the lag-l ACF as given in (2). Under the null hypothesis and the assumption that $\{Y_t\}$ is independent and identically distributed with finite second moment, $Q(m)$ is asymptotically χ^2-distributed with m degrees of freedom. Thus the null hypothesis is rejected if $Q(m) > \chi^2_\alpha(m)$, where $\chi^2_\alpha(m)$ denotes the α-quantile of the chi-squared distribution with m degrees of freedom.

The choice of m is not easily made in practice, and it may affect the performance of the test statistic. Tsay ([3], p. 33) proposes the choice of $m \approx log(N)$ for a higher power of the test. However, this needs to be reviewed in the case of seasonal time series since autocorrelations of a certain multiplicity is important. Since the monthly average global temperature anomalies have monthly seasonality, we have to take this into consideration.

When testing the residual series belonging to a fitted model, the degrees of freedom have to be adjusted to $m - g$, where g is the number of AR and MA coefficients in the model. This is because of the constraints added to the residual series from fitting the model. The adjustment can be made provided that $m \geq g$ ([3], p. 32-33, 50-51).

3.3.2 McLeod-Li test

The McLeod-Li test uses the squared residual series to check for autoregressive conditional heteroscedasticity, called *the ARCH effect*. This means that the conditional variance at time t is thought to be dependent on the previous squared residuals. The test is performed by using the Ljung-Box statistic on the squared residual series and following the procedure given in section 3.3.1. Here, the test corresponds to having the null hypothesis H_0 : "No conditional heteroscedasticity" against the alternative hypothesis H_1 : "Conditional heteroskedasticity" ([3], p. 114).

3.3.3 Augmented Dickey-Fuller test

Previously we have considered the case of unit-root nonstationarity, meaning that the AR polynomial has got at least one root on the unit circle. Unit-root nonstationarity implies that the mean and autocovariances of the series are time-dependent, which is why ways of detecting this are needed. In reality, unit-root nonstationarity behaviour is exhibited in stocks, meaning that they for example do not vary around a fixed level ([5], p. 11). To check if a unit root is present in a time series $\{Y_t\}$ one can use the *augmented Dickey-Fuller test*.

To explain the test procedure, consider an AR(p) model $Y_t = g(t) + \varphi_1 Y_{t-1} + ... + \varphi_p Y_{t-p} + \varepsilon_y$, where $g(t)$ is a deterministic function of time. Earlier we have only considered the case of $g(t) = \varphi_0$ being constant. Subtracting Y_{t-1} from both sides, we obtain

$$Y_t - Y_{t-1} = g(t) + (\varphi_1 - 1)Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t,$$

which can be rewritten as

$$Y_t - Y_{t-1} = g(t) + (-\phi(1))Y_{t-1} + \sum_{i=1}^{p-1} a_i (Y_{t-i} - Y_{t-i-1}) + \varepsilon_t, \tag{20}$$

where $\phi(1)$ denotes the AR(p) polynomial evaluated at $B = 1$ and $a_i = -\sum_{k=i+1}^{p} \varphi_k$. The existence of a unit root is equivalent to $-\phi(1) = 0$. Thus, by obtaining the ordinary least squares estimation of $-\phi(1)$, a t-test with $H_0 : -\phi(1) = 0$ against the one-sided alternative $H_1 : -\phi(1) < 0$ can be performed. These hypotheses correspond to H_0 : "The series is unit-root nonstationary" and H_1 : "The series is stationary". If the null hypothesis can not be rejected, we assume that Y_t is $I(d)$ for some $d \geq 1$ and perform the same test until H_0 is rejected. Note that an insufficient amount of lags leads to autocorrelation being present among the residuals, which is why having the right AR(p) model is essential for the test.

The test is performed using the t-statistic. However, since the series is nonstationary under the null hypothesis, the t-statistic will converge to a nonstandard distribution. To make things more tedious, this asymptotic distribution changes depending on different cases of $g(t)$. Dickey and Fuller provided three different cases:

Case 1 : The model includes no constant and no trend, meaning that $g(t) = 0$
Case 2 : The model includes a drift-term but no trend, meaning that $g(t) = c_0$
Case 3 : The model includes a drift-term and a trend, meaning that $g(t) = c_0 + c_1 t$.

The asymptotic distributions for these cases have been simulated by Dickey and Fuller and are available in statistical software ([6], p. 246-250). One important observation is that, according to (13), an ARMA model can be approximated by an AR model, meaning that the test can be applied on ARMA models. However, the condition for an AR approximation is that the ARMA model is invertible.

For clarity, we finish this section off by giving an example of how an AR(3) model can be rewritten on the form of (20). Subtracting by Y_{t-1} gives

$$Y_t - Y_{t-1} = g(t) + (\varphi_1 - 1)Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \varepsilon_t.$$

Using that $0 = \varphi_3 Y_{t-2} - \varphi_3 Y_{t-2}$ we rewrite the above as

$$Y_t - Y_{t-1} = g(t) + (\varphi_1 - 1)Y_{t-1} + (\varphi_2 + \varphi_3)Y_{t-2} - \varphi_3(Y_{t-2} - Y_{t-3}) + \varepsilon_t.$$

Now, using that $0 = (\varphi_2 + \varphi_3)Y_{t-1} - (\varphi_2 + \varphi_3)Y_{t-1}$ we get

$$Y_t - Y_{t-1} = g(t) + (\varphi_1 + \varphi_2 + \varphi_3 - 1)Y_{t-1} - (\varphi_2 + \varphi_3)(Y_{t-1} - Y_{t-2}) - \varphi_3(Y_{t-2} - Y_{t-3}) + \varepsilon_t,$$

which is on the form of (20).

3.4 Forecasting

Suppose that we are at time point h and that we are interested in forecasting the value Y_{h+l} for some positive integer l. The index h is called the *forecast origin* and l the *forecast horizon*. Denote the forecast of Y_{h+l} with $\hat{Y}_h(l)$ and let F_h be the information available at the forecast origin. Choosing $\hat{Y}_h(l)$ as the value that minimizes the squared error loss function $\mathrm{E}((Y_{h+l} - \hat{Y}_h(l))^2 | F_h)$ results in the l-step-ahead forecast of Y_t at forecast origin h,

$$\hat{Y}_h(l) = \mathrm{E}(Y_{h+l} | F_h).$$

The associated *forecast error* is given by $e_h(l) = Y_{h+l} - \hat{Y}_h(l)$ ([3], p. 54-55).

3.4.1 1-step-ahead forecast for ARMA models

Recall that $\hat{Y}_h(l)$ is the l-step-ahead forecast of Y_t at forecast origin h. In this thesis, we are only concerned with the 1-step-ahead forecast since the average global temperature anomaly for January 2020 is to be predicted with data available from December 2019. For an ARMA(p,q) model, the 1-step-ahead forecast is given by

$$\hat{Y}_h(1) = \mathrm{E}(Y_{h+1}|F_h) = \varphi_0 + \sum_{i=1}^{p} \varphi_i Y_{h+1-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{h+1-i}. \tag{21}$$

The associated forecast error is $e_h(1) = Y_{h+1} - \hat{Y}_h(1) = \varepsilon_{h+1}$, which implies that $\mathrm{Var}(e_h(1)) = \mathrm{Var}(\varepsilon_{h+1}) = \sigma^2$ is the variance of the 1-step-ahead forecast error. If p or q is zero the forecast is given by removing either the AR or MA terms from (21). The associated forecast error and its variance will however remain the same ([3], p. 55, 62,68)

3.4.2 Measures of forecast accuracy

The evaluation of forecast accuracy for a model requires a summary statistic, which the goal is to minimise. We will use statistics that measure the errors of the forecast based directly on the deviation of the forecast from the real value. There are several suggestions on how this statistic should be chosen, and two commonly used ones are the *root mean squared error* (RMSE) and the *mean absolute error* (MAE). Since this thesis only uses 1-step-ahead forecasts, the RMSE and MAE are given by

$$\begin{aligned} RMSE &= \frac{1}{n}\sqrt{\sum_{i=0}^{n-1}(Y_{t+1+i} - \hat{Y}_{t+i}(1))^2} = \frac{1}{n}\sqrt{\sum_{i=0}^{n-1} e_{t+i}(1)^2} \quad \text{and} \\ MAE &= \frac{1}{n}\sqrt{\sum_{i=0}^{n-1} |e_{t+i}(1)^2|}, \end{aligned} \tag{22}$$

where n is the number of 1-step-ahead forecasts made ([9], section 3.3.4). There is no consensus on which statistic is to be preferred, but simulations have shown that the RMSE is more appropriate to use when errors follow a normal distribution. However, the downside to the RMSE is its sensitivity to outliers, since these errors have larger impact as a result of being squared ([10]).

3.4.3 Training and test set

An critical aspect of forecasting is the use of genuine forecasts. This means that data which has been used to fit the model should not be used to determine the accuracy of a forecast. Instead, the accuracy of forecasts is determined solely on how well the model performs on new data that have not been used in the fitting process. This brings us to the concepts of *training and test sets.*

When presented with a data set it is common to divide it into two parts: the training set and the test set. The training set is used to fit the model, that is to estimate the parameters, and the test set is used to evaluate the accuracy of the model. Since the test set is not used when determining forecasts, it provides a better picture of how well the model forecasts new data.([9], section 3.4).

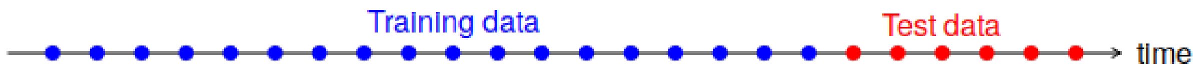

Figure 2: Illustration of training and test sets. Source: [9].

3.4.4 Rolling window forecast

A common way to test the predictive accuracy of a model is to use a *rolling window* procedure. This procedure starts by initially splitting the data into a training set and a test set, with the training set consisting of the oldest observations. The training set constitutes the first window and is used to fit the model. After the model has been fitted, the l-step-ahead forecast is calculated. Since the forecasted data is already known, a prediction error is also calculated. The window is then rolled forward a fixed increment and the same procedure is performed until an l-step-ahead forecast can no longer be performed. A measure of forecast accuracy is then used together with the forecast errors to evaluate the predictive accuracy of the model ([11], p. 313-315, 349-350).

Denote the size of the window with m. This means that the test set consists of $N - m$ observations. Thus, the use of the rolling window method with 1-step-ahead forecasts leads to $N - m$ forecast errors. In this thesis, the forecast accuracy is evaluated using the RMSE and MAE given in (22). See Figure 2 for an illustration of the rolling window forecast, where $N = 26$, $m = 6$ and the forecast horizon is equal to one.

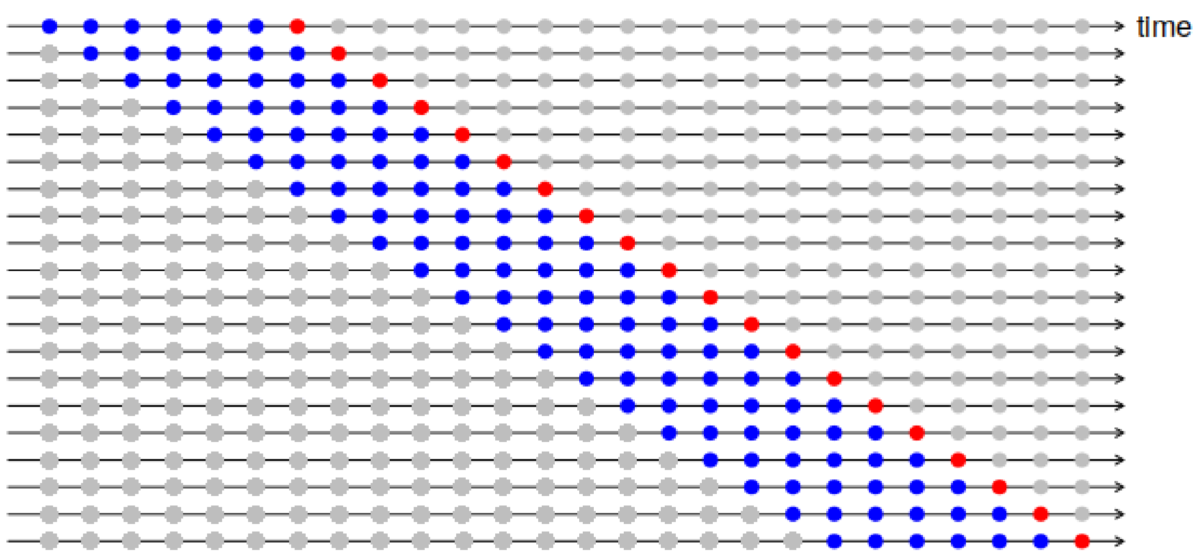

Figure 3: Illustration of a rolling window forecast. Windows in blue and forecasted values in red. Source: [9].

3.5 Software and packages

The statistical computations in this thesis are performed using the programming language *R* ([12]), version 3.6.2, together with the integrated development environment *R studio* ([13]), version 1.2.5033. To facilitate the coding, the package *Tidyverse* ([14]) is used throughout. Other used packages, which implement the theory of model fitting and test the conditions of weakly stationarity, are *TSA* ([15]), *aTSA* ([16]), *Forecast* ([17]) and *Rugarch* ([18]).

4 Analysis of the average global temperature anomalies

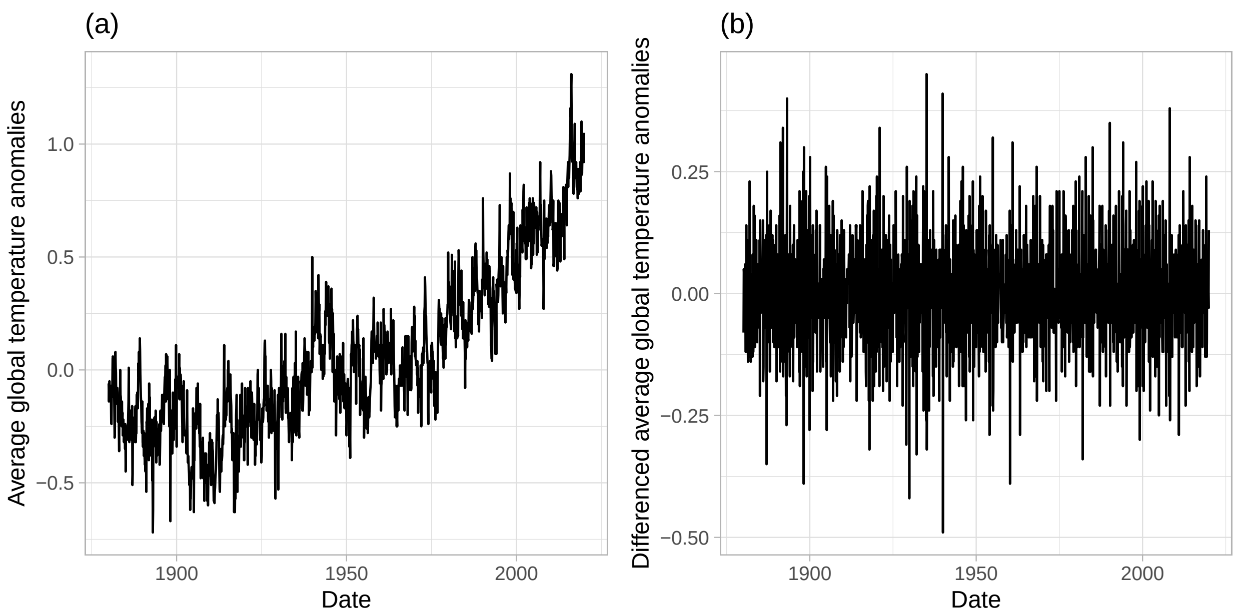

Figure 4: (a) The average global temperature anomalies and (b) the the differenced anomalies.

Plot (a) in Figure 4 does not have the look of a stationary process, whereas (b) does. The anomalies in (a) may in fact be trend stationary, but this is nothing we will go further into in this thesis. It becomes quite clear that the differenced anomalies do not include either a drift or trend. Augmented Dickey-Fuller tests of the differenced anomalies can be found in Appendix B.1 and will be used in the upcoming section, 4.1. However, in Appendix B.1 we clearly see that the augmented Dickey-Fuller test is significant for every lag between 1 and 60, which makes it quite reasonable to think that the differenced temperature anomalies constitute a weakly stationary time series. The differenced data set does at least seem to solve the problem of a time-dependent mean. Therefore, the differenced temperature anomalies will be used in the analysis ahead.

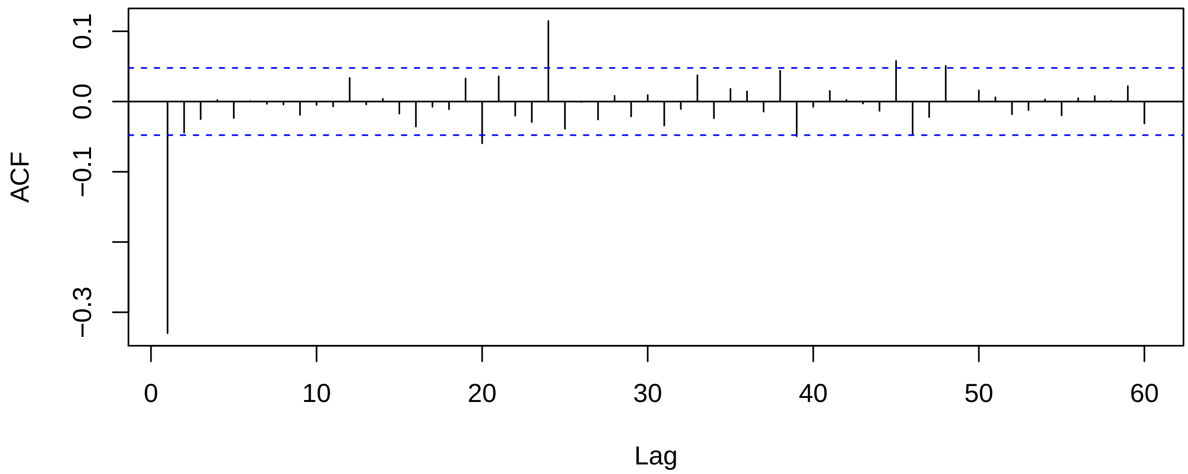

Figure 5: ACF of the differenced temperature anomalies. The blue dashed lines represent a 95% confidence interval.

Figure 5 above shows that the differenced anomalies have significant autocorrelations at lags 1, 20, 24, 39, 45 and 48. The differenced temperature anomalies' dependence of previous observations suggests that a time series model may be suitable.

4.1 Three model candidates

4.1.1 Integrated AR model

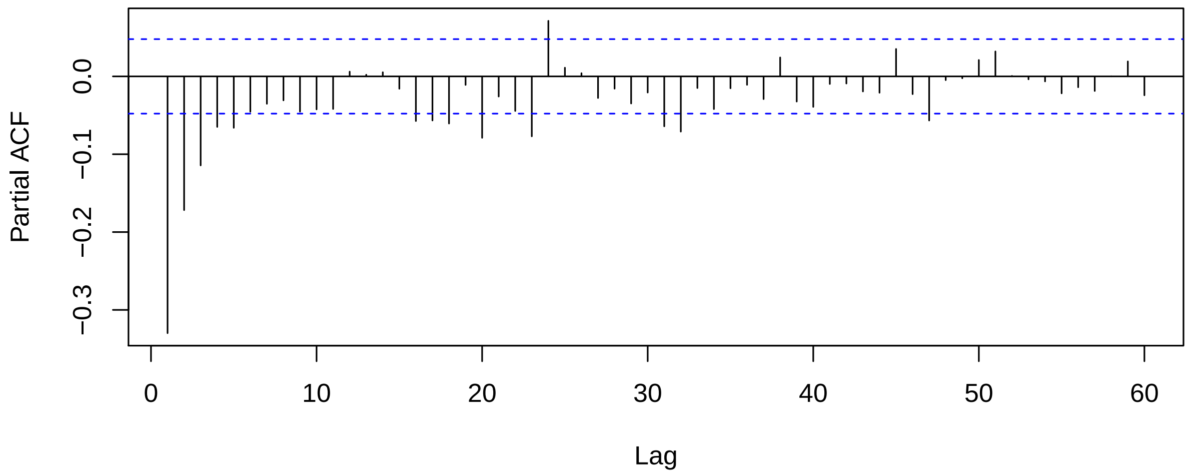

Figure 6: PACF of the differenced temperature anomalies. The blue dashed lines represent a 95% confidence interval.

The PACF plot shows significance on several lags and suggests that, for instance, AR(5), AR(20), AR(32) or AR(47) may be suitable models. Note however, that the significance of some of these PACF values could be the result of small randomness or measurement errors. The PACF plot indicates a suitable AR model but is ultimately chosen by AICC since it penalizes models with a large amount of parameters.

The resulting model is an ARIMA(32,1,0) model. See Appendix C.1 for parameter estimations. In order for this model to be a good fit we need to check whether the conditions on the error terms hold, which translates to check that the residuals are independent and $N(0, \sigma^2)$ distributed. Recall that this is important for the augmented Dickey-Fuller test, which requires non-autocorrelated error terms. Figure 7 below shows that the residuals are not autocorrelated on any of the 60 first lags. This is also confirmed by the Ljung-Box tests in Figure 18 (Appendix B.2), which show that less than five percent of the tests rejected the null hypothesis of no autocorrelation. This is to be expected if the null hypothesis is true, and therefore we assume that the residuals are not autocorrelated.

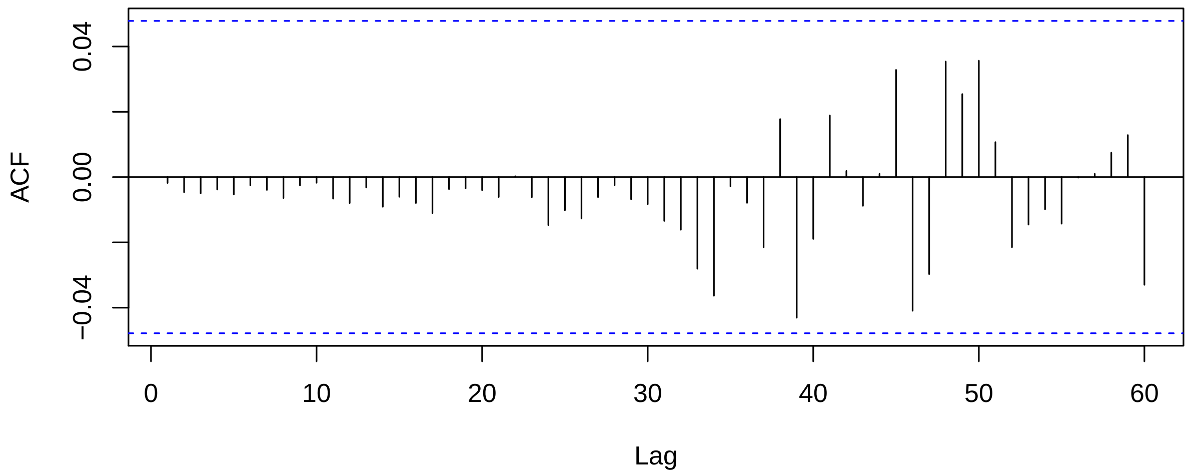

Figure 7: ACF of the residuals from ARIMA(32,1,0). The blue dashed lines represent a 95% confidence interval.

Now it remains to see whether it is reasonable to assume that the residuals are normally distributed. The theoretical normal distribution plotted on the histogram in Figure 8 is obtained by maximum likelihood estimation of the residuals obtained from the fitted ARIMA(32,1,0) model.

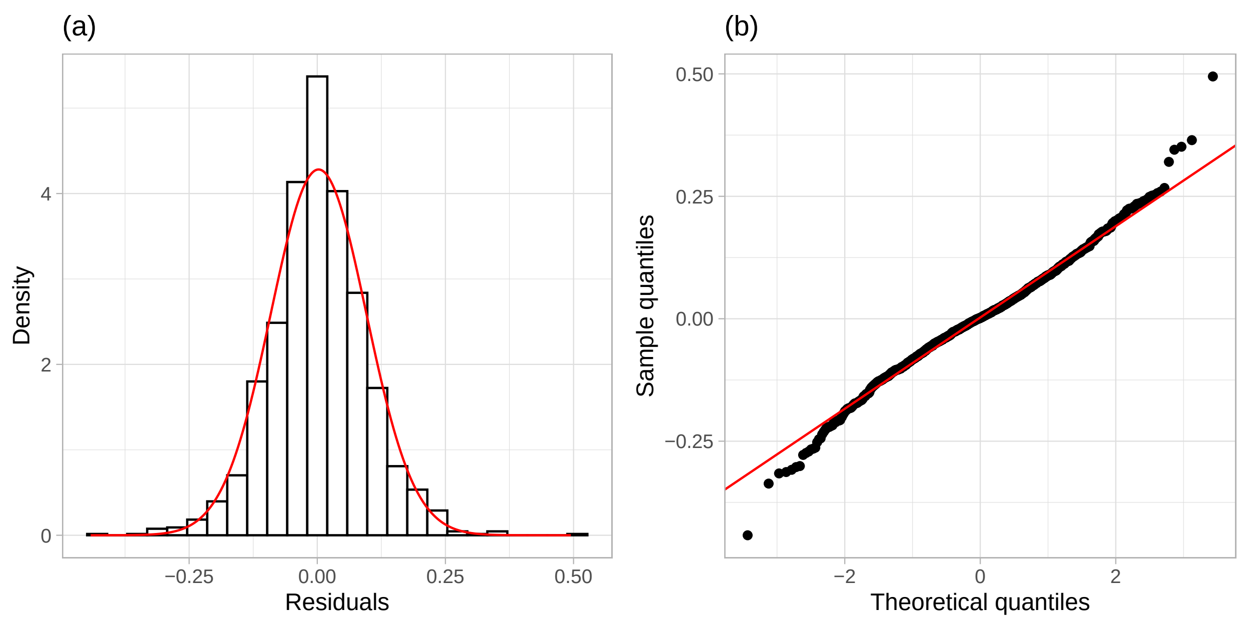

Figure 8: (a) Histogram with N(0.00258, 0.00868) density function and (b) standard normal Q-Q plot with residuals from fitted ARIMA(32,1,0).

Figure 8 shows that it is reasonable to assume that the residuals follow a normal distribution. The histogram follows the theoretical density function nicely and the Q-Q plot shows a linear trend with a few outliers.

To end this section, we return to the stationarity assumption of the differenced anomalies. Since the residuals of the AR(32) model were uncorrelated, we perform the augmented Dickey-Fuller test using 32 lags. Figure 17 in Appendix B.1 shows that the p-value is 0.01 or smaller, meaning that the null hypothesis of unit-root nonstationarity is rejected and stationarity is assumed.

4.1.2 Integrated MA model

To visually detect the order of an MA process, one can look at the ACF. Figure 5 shows that, for instance, MA(1), MA(24), MA(39) and MA(44) may be suitable models for the differenced temperature anomalies. However, some of these significant ACFs may be the result of randomness in data. Using AICC an ARIMA(0,1,26) is chosen. See Appendix C.2 for parameter estimations.

We now check if the residuals satisfy the model conditions. Figure 9 shows insignificant ACFs for all lags except 58. This could be the result of some randomness in the data. The Ljung-Box tests in Figure 19 (Appendix B.3) show that the null hypothesis of no autocorrelation can not be rejected. Therefore, the residuals are assumed to be uncorrelated.

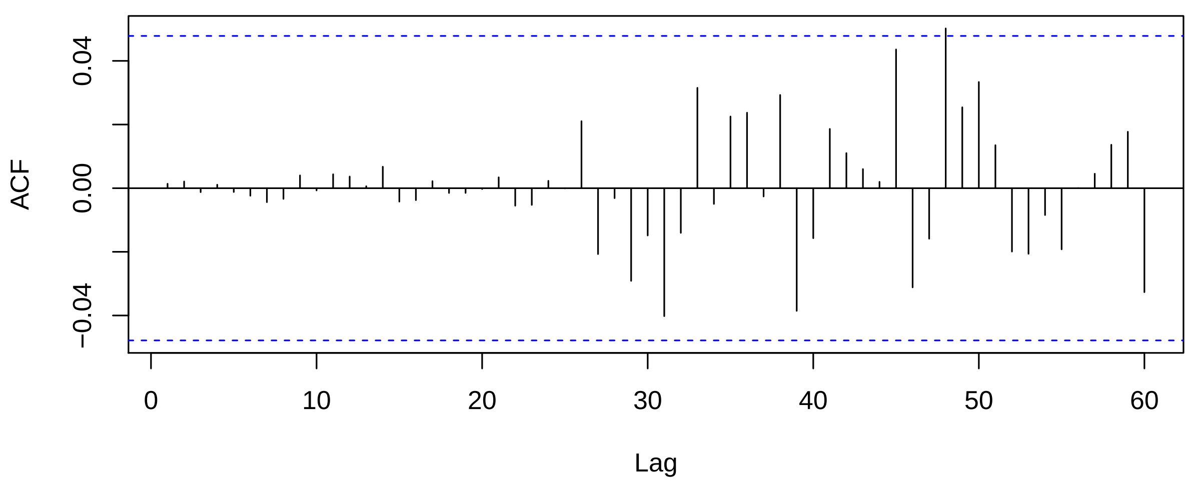

Figure 9: ACF of residuals from ARIMA(0,1,26). The blue dashed lines represent a 95% confidence interval.

To see whether the residuals follow a normal distribution, a histogram with a theoretical normal distribution density, acquired by fitting the residuals using maximum likelihood, and a normal Q-Q plot is plotted in Figure 10.

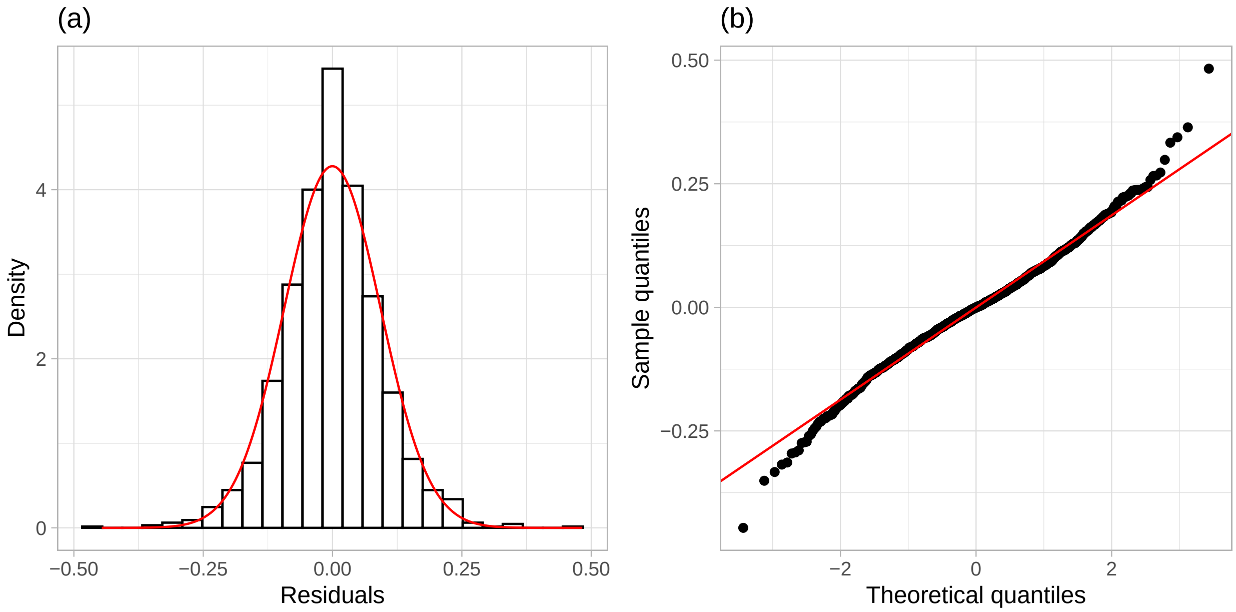

Figure 10: (a) Histogram with N(-0.00030, 0.00870) density function and (b) standard normal Q-Q plot with residuals from fitted ARIMA(0,1,26).

The residuals of the ARIMA(0,1,26) model may follow a normal distribution according to Figure 10. The histogram closely follows the theoretical normal distribution and the Q-Q plot exhibits a linear trend with only a few outliers (slightly heavy tails).

Recall that an MA process, by definition, is stationary. No test is therefore needed to confirm this condition.

4.1.3 Integrated ARMA model

The order of an ARMA process can not be identified by using ACF and PACF, since neither of them cuts off at any lag. Solely relying on the AICC results in an ARIMA(1,1,2) model for the temperature anomalies. See Appendix C.3 for parameter estimations.

One can in Figure 11 see that there may be autocorrelation present among the residuals. This is verified by the Ljung-Box tests in Appendix B.4 (see Figure 20), where more than fifty percent of the tests rejected the null hypothesis of no autocorrelation. Autocorrelation among the residuals shows that the mean model is not able to adequately capture correlation in the data. This may, however, not affect the models' ability of forecasting, which is why we will continue using it. An ARCH model may perhaps fix the problem of autocorrelation if it turns out that the autocorrelation largely can be explained by the autocorrelation of the residuals' squared values.

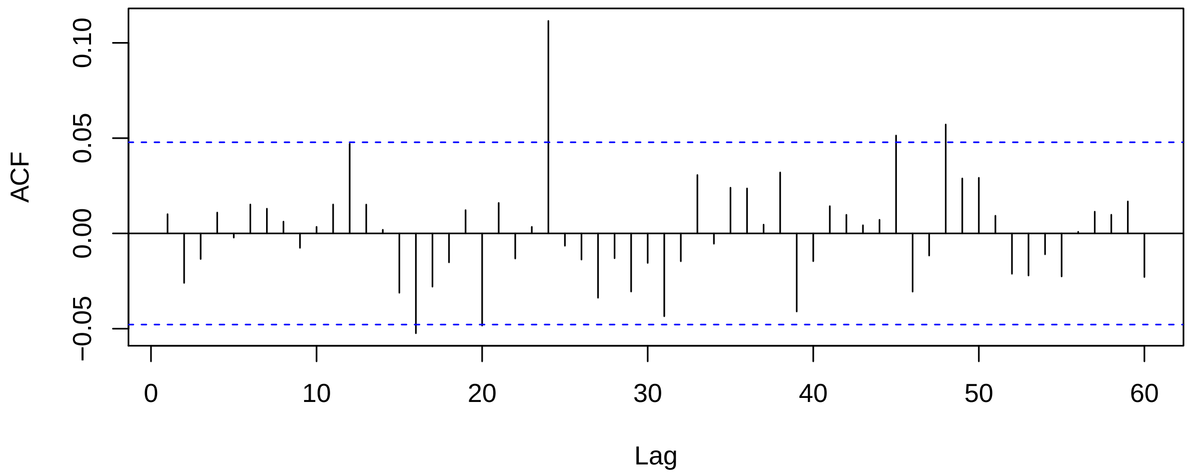

Figure 11: ACF of residuals from ARIMA(1,1,2). The blue dashed lines represent a 95% confidence interval.

Figure 12 shows that the residuals may follow a normal distribution. As in the two previous sections we see that the histogram follows the fitted theoretical normal distribution and that the Q-Q plot shows a nice linear trend, with a few values causing the sample distribution to have slightly heavier tails.

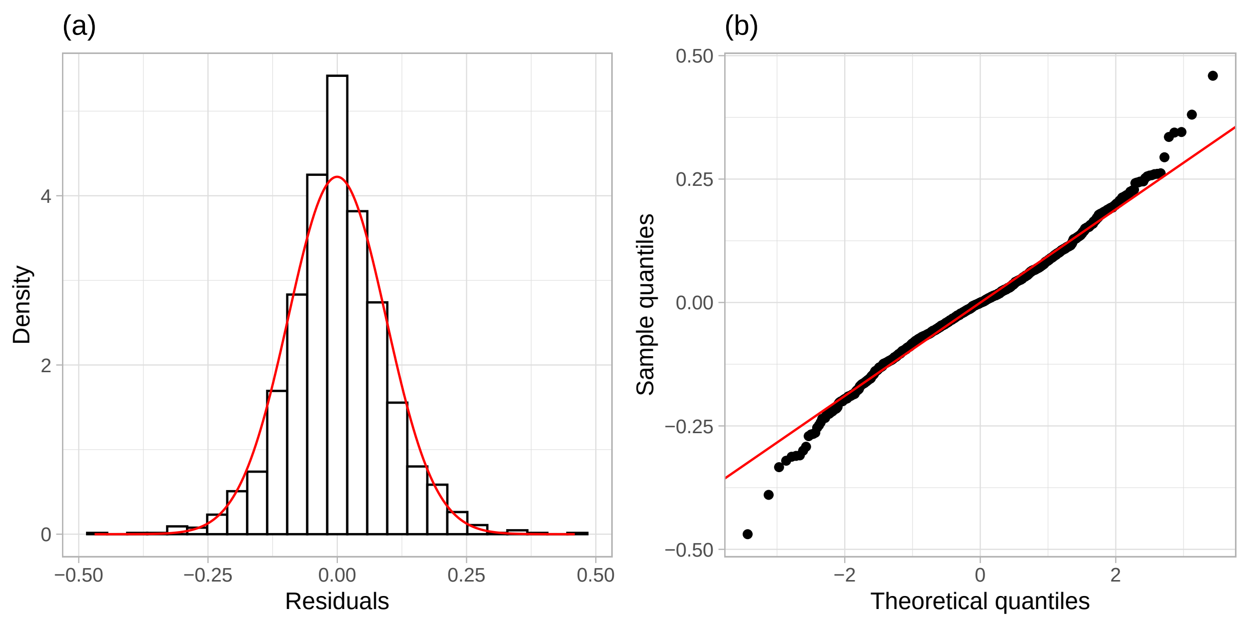

Figure 12: (a) Histogram with N(-0.00031 , 0.00892) density function and (b) standard normal Q-Q plot with residuals from fitted ARIMA(1,1,2).

Important aspects of an ARMA model is that both the stationary and invertible conditions hold, meaning that the roots of both the AR and MA polynomial are larger than one in modulus. The reader can easily verify, using the estimations in Appendix C.3, that all real solutions to the MA polynomial are larger than 1 or smaller than -1, both for the ARIMA(1,1,2) and ARIMA(0,1,26) model. Note, however, that the R package represents the MA(q) polynomial by $\theta(B) = c_0 + \theta_t + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q$ instead of the representation given in (8). In the same manner, it is easily verified that the ARIMA(1,1,2) is stationary, since the single solution to the characteristic equation is larger than 1. The stationarity condition can also be argued for by noting that the ARIMA(1,1,2) model can be represented as an infinite AR model and can thus be approximated by an AR model of larger order, as described in (13). Since the ADF tests in Figure 17 (Appendix B.1) all reject unit-root nonstationarity for large orders of p, the stationarity condition should hold for the ARIMA(1,1,2) model.

4.2 Selection of the final model

To select which model to use for forecasting of the average global temperature anomaly for January 2020, we use a rolling window forecast with RMSE and MAE as measures of forecast accuracy. The differenced temperature anomalies data set contains 1679 data points, of which the last 679 (around 40%) are used as test data. This implies that the size of each window is 1000 observations. After the model is fitted on a window, the 1-step-ahead forecast is calculated and the forecast error measured. When the rolling window has forecasted the 1679th data point, the RMSE and MAE are calculated, and the model minimising these measurements is chosen as the final model.

The results of the rolling window forecast are found in Table 1.

Table 1: Measures of forecast accuracy for all three candidate models

Model	RMSE	MAE
ARIMA(1,1,2)	0.08996	0.26199
ARIMA(0,1,26)	0.09036	0.26385
ARIMA(32,1,0)	0.09088	0.26480

Although the differences are not large, Table 1 shows that the ARIMA(1,1,2) model has both the smallest RMSE and MAE. It is thus selected as the final model.

The RMSE and MAE are measures that compress much information into a single number. It may perhaps be the case that the ARIMA(32,1,0) model has better forecast accuracy than the chosen model but with one single forecast being very far off the real value, causing the RMSE and MAE to be larger. For the interested we have plotted the 1-step-ahead forecasts for each of the three models, see Figure 16 in Appendix A.2, which shows that the overall performance is similar for all models.

4.3 Forecast of the average global temperature anomaly for January 2020

We have now arrived at the last section of the analysis, which consists of section 4.3.1 and 4.3.2. The first mentioned section gives a point forecast while the latter gives a forecast interval for the temperature anomaly for January 2020.

4.3.1 Point forecast

Since the ARIMA(1,1,2) won the rolling window forecast in 4.2, it will now be used to forecast the temperature anomaly for January 2020. Recall that the whole analysis has been carried out using the differenced temperature anomalies, which is a data set of 1679 observations. This causes us to forecast the 1680th difference instead of directly forecasting the 1681st temperature anomaly. Now, let $\{W_t = Y_{t+1} - Y_t, t \geq 0\}$

denote the differenced temperature anomalies, where $\{Y_t, t \geq 0\}$ are the temperature anomalies and the 1680th difference is given by $W_{1680} = Y_{1681} - Y_{1680}$. Given all information up until the 1680th month , F_{1680}, the 1-step-ahead forecast of W_{1680} can be written as

$$\hat{W}_{1679}(1) = \hat{Y}_{1680}(1) - Y_{1680}.$$

The forecast of the average global temperature anomaly can thus be calculated by

$$\hat{Y}_{1680}(1) = \hat{W}_{1679}(1) + Y_{1680}, \tag{23}$$

where $\hat{Y}_{1680}(1)$ is the 1-step-ahead forecast of the average global temperature anomaly for January 2020, Y_{1681}. Note that, given F_{1680}, both Y_{1680} and $\hat{W}_{1679}(1)$ are known. Equation (23) can be generalised to $\hat{Y}_t(1) = \hat{W}_{t-1}(1) + Y_t$ with F_t given.

By using (21) to forecast the 1680th differenced temperature anomaly, we get that the forecasted average global temperature anomaly for January 2020 is given by 0.987.

4.3.2 Forecast interval

It is quite reasonable to think that the forecasted temperature anomaly in the previous section will differ from the real temperature anomaly by a small amount. To make room for forecast errors, a 95 percent forecast interval will therefore be given. Using the notation in the previous section we see that the forecast error of the 1680th differenced temperature anomaly is given by $e_{1679}(1) = W_{1680} - \hat{W}_{1679}(1)$. But from section 3.4.1 it holds that $e_{1679}(1) = \varepsilon_{1680}$, and thus $\mathrm{Var}(e_{1670}(1)) = \sigma^2$. Using the assumption of normality and that $\mathrm{E}(e_{1679}(1)) = 0$ we get that

$$\mathrm{P}(-z_{0.025} < \frac{W_{1680} - \hat{W}_{1679}(1)}{\sigma} < z_{0.025} | F_{1680}) = 0.95 \iff$$

$$\mathrm{P}(\hat{W}_{1679}(1) - z_{0.025}\sigma < W_{1680} < \hat{W}_{1679}(1) + z_{0.025}\sigma | F_{1680}) = 0.95,$$

where $z_{0.025}$ is the 0.025-quantile of a standard normal distribution. We conditioned on F_{1680} since it consists of all the information available when forecasting Y_{1681}. By writing W_{1680} as $Y_{1681} - Y_{1680}$ and using the fact that Y_{1680} given F_{1680} is known, results in the following forecast interval for the average global temperature anomaly for January 2020.

$$Y_{1681} = Y_{1680} + \hat{W}_{1679}(1) \pm z_{0.025}\sigma. \tag{24}$$

However, Figure 13 (a) shows that the squared residuals have seasonal autocorrelation, which means that conditional heteroscedasticity is present. This is also confirmed by the McLeod-Li test found in Appendix B.5. The presence of conditional heteroscedasticity makes it reasonable to model the conditional distribution of $\{W_t\}$ to achieve greater accuracy of the forecast interval. Using an ARCH model, we replace the unconditional variance, σ, in (24) by the estimated conditional variance, $\hat{\sigma}_{1680}$.

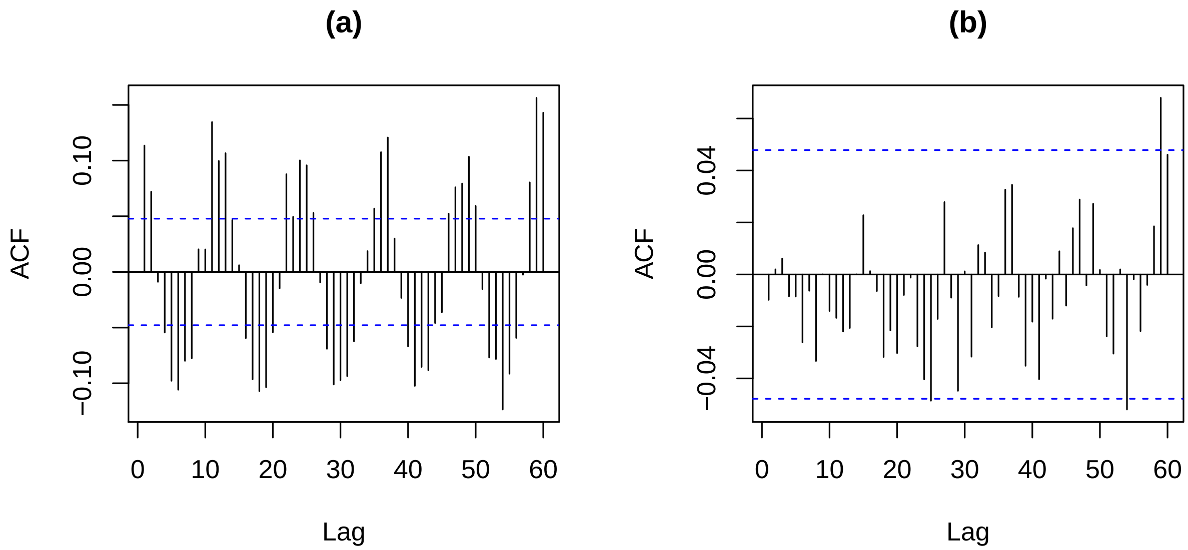

Figure 13: (a) ACF of the squared residuals and (b) ACF of the squared adjusted residuals from ARIMA(1,1,2). The blue dashed lines represent a 95% confidence interval.

Model selection using AICC results in an ARCH(26) model. The series $\{\frac{\varepsilon_t}{\sigma_t} = u_t\}$ corresponds to the adjusted residuals of the ARIMA(1,1,2) model. The ACF plot of the adjusted residuals are given by (b) in Figure 13. With only a few ACF values being significant, the ARCH model may have been able to explain the conditional variance. McLeod-Li tests of the series $\{\frac{\varepsilon_t^2}{\sigma_t^2}\}$ can be found in Appendix B.6 and gives the result of no conditional variance, meaning that our ARCH(26) model largely explains the conditional variance. However, the problem of autocorrelation still stands, which is seen in the results of the Ljung-Box tests in Appendix B.7.

Recall, from section 3.1.2, that we want to examine if $\{u_t\}$ are $N(0, 1)$ distributed. By the look of Figure 14, the adjusted residuals may very well be $N(0, 1)$ distributed, although not independent according to the Ljung-Box tests in Appendix B.7. But since the aim of this section is to model the conditional variance, we do not look further into the problem of autocorrelation present in the series $\{u_t\}$. The obtained forecast interval is plotted in the next section.

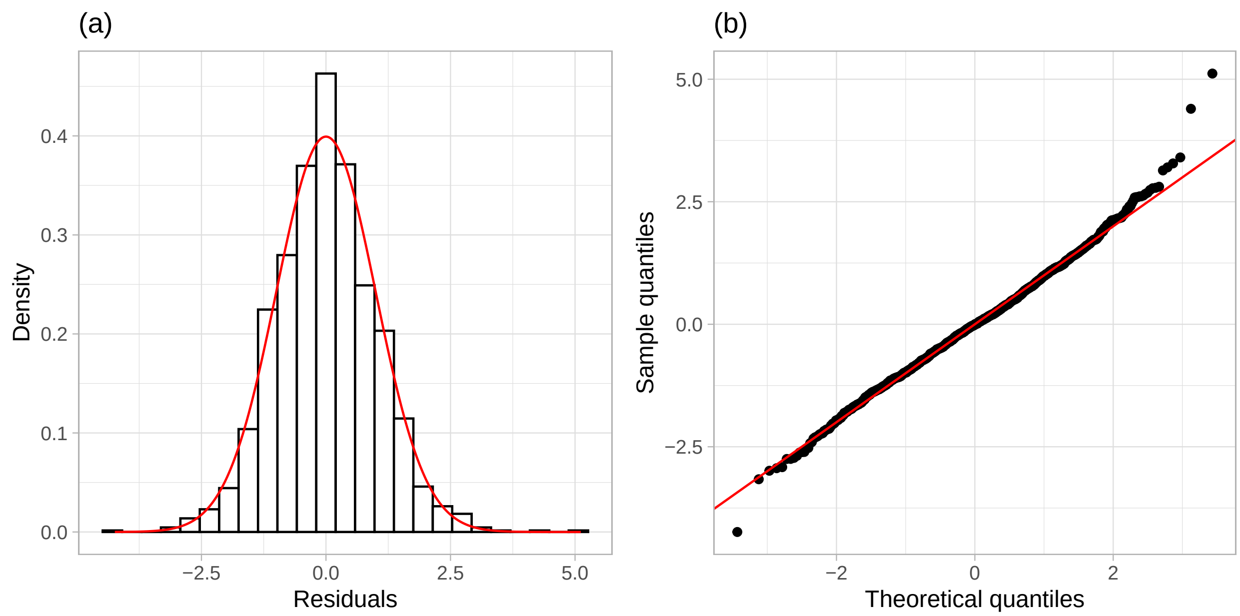

Figure 14: (a) Histogram with N(-0.00004 , 0.99917) density function and (b) standard normal Q-Q plot of the adjusted residuals from ARIMA(1,1,2)

4.3.3 Results

Replacing σ in (24) with $\hat{\sigma}_{1680}$ gives that $Y_{1681} = Y_{1680} + \hat{W}_{1679}(1) \pm z_{0.025}\hat{\sigma}_{1680}$ conditioned on all available info up until December 2019 (month 1680). Thus, we have arrived at the results of our analysis. Figure 15, found below, shows the temperature anomalies for a ten-year period, up until January 2020. The black line shows the real temperature anomalies, the red line shows the 1-step-ahead forecasted anomalies and the blue lines show the forecast interval. The forecasted average global temperature anomaly for January 2020 is given by 0.989°C and the 95 percent forecast interval is given by 0.989 ± 0.163°C. Note that the forecast is smaller than the temperature anomaly measured in December 2019. This does not necessarily mean that the real temperature anomaly for January 2020 will be smaller than the anomaly for December, since the forecast interval extends past that value. However, the anomaly will probably not go past the largest peak reached in March 2016 (month 1635) and will most certainly not be negative.

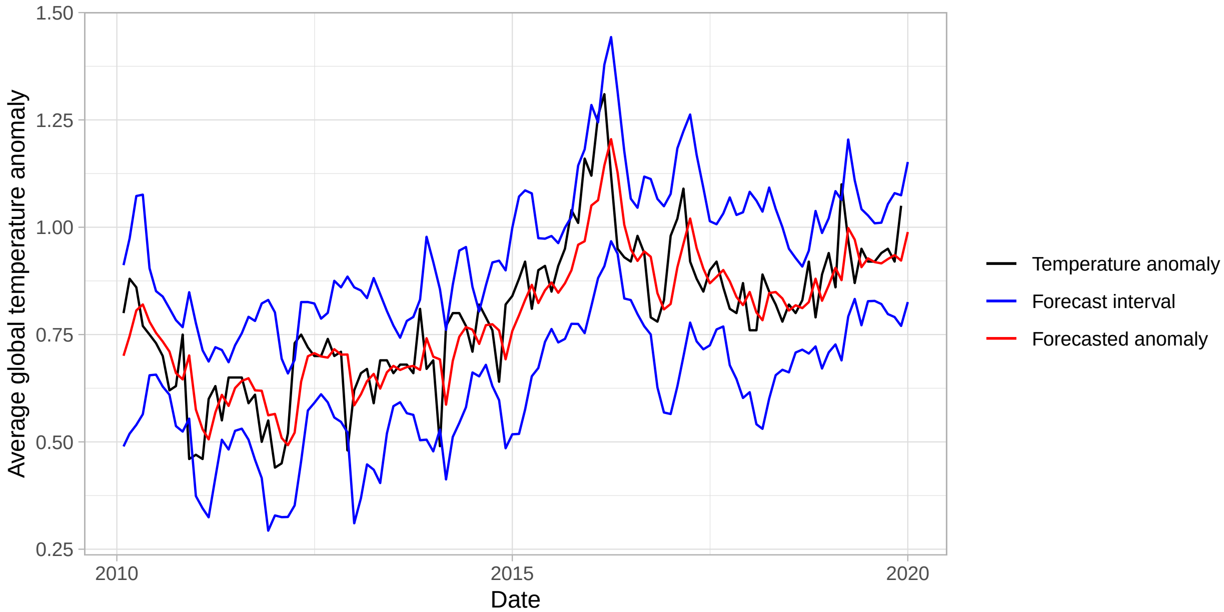

Figure 15: 1-step-ahead rolling window forecasts with corresponding 95% forecast interval for ARIMA(1,1,2)

Discussion

The objective of this thesis was to find an ARIMA-ARCH model that largely describes the movements of the monthly average global temperature anomalies, so that it could be used to predict the temperature anomaly for January 2020. Seeing that the global temperature anomaly is difference stationary, we compared the predictive capabilities of three models: ARIMA(32,1,0), ARIMA(0,1,26) and ARIMA(1,1,2), which were chosen by AICC. Using a rolling window to acquire 679 1-step-ahead forecasts for each model resulted in the ARIMA(1,1,2) model having the lowest RMSE and MAE. One can although discuss whether the results in Table 1 actually show that the forecasting capabilities are different for the three models, since the RMSEs differ less than one thousandth of a degree Celsius. It may be the case that dividing the time period into different parts results in different models having best forecasting accuracy. This is something that can be studied further in the future.

Something that must be taken into account when using the ARIMA(32,1,0) and ARIMA(0,1,26) model is the problem of overfitting. Including too many parameters may result in the model explaining random variation in the data on which it is trained, which can lead to misleading forecasts. The question here is if one of these models overfit the data. It is worth mentioning that we do not have any expertise regarding global temperature, and can therefore not comment on whether the models are reasonable. Intuitively, however, it seems plausible that the global temperature may depend on temperatures as long as 2-3 years back. The fact that the RMSEs for all three models differ by less than one thousandth of a degree shows that the ARIMA(1,1,2) barely outperforms the other two when it comes to forecast accuracy. Also, recall that the ARIMA(1,1,2) model was the only model that did not meet model condition of no autocorrelated error terms. This suggests, once again, that it may be worth looking at the two models even further. Also worth mentioning is that we fitted AR and MA models with fewer parameters, for example MA(1) and AR(4), but these always resulted in autocorrelated residuals that were visibly not normally distributed.

When performing a rolling window forecast, one has to think about parameter constancy. One wishes for the parameters to be constant over time. If that is not the case, the model fitted in each window is not suitable for those observations. Recall that we decided the number of parameters using AICC on the whole data set of 1679 differenced temperature anomalies. This does not imply that the same number of parameters are suitable for each window of 1000 observations. To determine if the parameters are constant over time, one can use statistical tests. No such test was used in this thesis. Instead, parameter constancy was assumed, which leaves room for further improvement. One idea is to divide the time period into subperiods and see whether other models are more suitable.

As a reader, one may notice that every test performed in this thesis uses 60 lags. This is not a conventional number. Section 3.3.1 tells us that letting the number of lags be equal to $\log(N)$ results in higher power of the Ljung-Box test. The same section also states that this needs to be reviewed when working with seasonal time series, which is the case with our data set. Since the data set consists of monthly observations, we have yearly seasonality. It is therefore more important to consider a multiple of the seasonality when choosing lags. We ultimately chose to use 60 lags, which corresponds to five years, but no statistical/mathematical result was used for this selection. A simulation study could be performed to decide what multiple of seasonality to use.

Seasonal autoregressive integrated moving average models (SARIMA) may be of interest since seasonality is an essential part of the average global temperature anomaly. Perhaps this could help fix the problem of the autocorrelated residuals and the seasonally autocorrelated squared residuals of the ARIMA(1,1,2) model (see Figure 13 (a)). An effort to fix these problems with a SARIMA model was made but gave no satisfying result. Also, lack of time made it hard to look further into these models, which is why they were excluded from this thesis.

The rolling window forecast performed used 40 percent (679 data points) of all observations as a test set. There is no convention on what proportion to use as training and test set respectively. But the reasoning to why 40 percent was chosen is there are many observations available to fit the model in each window, but also many observations to calculate the RMSE and MAE. A large number of observations leads to more accurate estimations of the model parameters. In the same way, a large number of observations leads to a stabilized RMSE and MAE. Another thing to discuss regarding the RMSE and MAE is that they were calculated using the differenced temperature anomalies instead of the temperature anomalies. Using the previously used notation, $W_t = Y_{t+1} - Y_t$, where $t = 1000, 1001, ..., 1679$ denotes the forecast origin in the rolling window, we have

$$W_{t+1} - \hat{W}_t(1) = Y_{t+1} - Y_t - (\hat{Y}_t(1) - Y_t) = Y_{t+1} - \hat{Y}_t(1).$$

Here we used that $\hat{W}_t(1) = \hat{Y}_t(1) - Y_t$ since Y_t is known at each step of the rolling window. Hence, by looking at (22), we see that the RMSE and MAE of the differenced temperature anomalies are equal to the RMSE and MAE of the temperature anomalies.

One thing to keep in mind regarding the augmented Dickey-Fuller test is that it only tests if 1 is a root to the AR polynomial. Recall that an ARMA process is unit-root nonstationary if it has a root on the unit circle. This means that the time series can still be unit-root nonstationary even if an augmented Dickey-Fuller test rejects the null hypothesis. The case when $\phi(1) = 0$ is especially interesting in time series analysis since it implies that data can be transformed into stationary data by differencing an appropriate amount of times. We have searched for sources that consider unit-roots that are not equal to 1, but with no luck. Perhaps it may be the case that real-life unit-root nonstationarity takes the form of a root equal to 1, but this is something we have not verified.

Each model created in this thesis contains all parameters smaller than and equal to the order of the model. The ARIMA(32,1,0), for instance, consists of 32 estimated AR parameters (see Appendix C.1). Note that some of these are a lot smaller than the other estimates and may be insignificant to the model. A continuation of this thesis could therefore be to remove insignificant parameters and see if it improves the forecast accuracy of the model. This can perhaps be done using likelihood ratio tests. One must, however, be careful, since removing wrong parameters will cause the residuals to be autocorrelated.

The results, shown in Figure 15, are the predictions of our ARIMA(1,1,2)-ARCH(26) for the 120 most recent observations. No backtest of the forecasting has been done and is something to look at in the future, since it would put a number on the model's predictive accuracy. However, Figure 15 shows that the model may be suitable. It is easily verified that the real temperature anomaly violated the forecast intervals eight times. This is 0.067 percent of the anomalies. If the model largely explains the data, we expect violations 5 percent of the times, which is almost the number obtained.

This thesis made use of the ARCH model to model the conditional variance, which resulted in the previously mentioned ARCH(26) model. One can instead use the GARCH model, which is a generalization of the ARCH model. By using GARCH it is possible to compare conditional variance models in the same manner as the comparison of ARIMA models, meaning that one would expect to be able to replace the current ARCH(26) model with a GARCH model having fewer parameters. Lastly, we want to address that the assumption of normally distributed error terms could be switched out for t-distributed error terms. When plotting the residuals of our models against theoretical quantiles of a suitable t-distribution, with degrees of freedom obtained by maximum likelihood, they show a strong linear trend. Similarly, the histogram nicely follows the theoretical t-distribution. Fitting the ARCH model with the assumption of t-distributed error terms may perhaps increase the forecasting accuracy of the models.

References

[1] NASA, Scientific consensus: Earth's climate is warming, (2020). https://climate.nasa.gov/scientific-consensus/ (accessed April 24, 2020).

[2] N.N.C. for Environmental information, Climate at a glance: Global time series, (2020). https://www.ncdc.noaa.gov/cag/global/time-series (accessed January 27, 2020).

[3] R.S. Tsay, Analysis of Financial Time Series, 3rd ed., Wiley, New York, 2010.

[4] W. Fuller, Introduction to statistical time series, 2nd ed., Wiley, New York, 1976.

[5] J. Box George.E. P., Time series analysis: Forecasting and control, 4th ed., Holden-Day, New York, 1976.

[6] T. Rachev, S. Mittnik, F. Fabozzi, S. Focardi, T. Jašić, Financial econometrics : From basics to advanced modeling techniques, John Wiley & Sons, Inc., Hoboken, 2007.

[7] L. Held, D. Sabanés Bové, Applied statistical inference likelihood and bayes, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[8] K.P. Burnham, D.R. Anderson, Model selection and multimodel inference: A practical information-theoretic approach, 2nd ed., Springer, New York, 2002.

[9] R. Hyndman, G. Athanasopoulos, Forecasting: Principles and practice, 2nd ed., OTexts, Australia, 2018.

[10] T. Chai, R.R. Draxler, Root mean square error (rmse) or mean absolute error (mae)? - arguments against avoiding rmse in the literature, Geoscientific Model Development. 7 (2014) 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.

[11] E. Zivot, J. Wang, Modeling financial time series with s-plus® [electronic resource], Second Edition., Springer Science+Business Media, Inc., New York, NY, 2006.

[12] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2019. https://www.R-project.org/.

[13] RStudio Team, RStudio: Integrated development environment for r, RStudio, Inc., Boston, MA, 2019. http://www.rstudio.com/.

[14] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T.L. Pedersen, E. Miller, S.M. Bache, K. Müller, J. Ooms, D. Robinson, D.P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the tidyverse, Journal of Open Source Software. 4 (2019) 1686. https://doi.org/10.21105/joss.01686.

[15] K.-S. Chan, B. Ripley, TSA: Time series analysis, 2018. https://CRAN.R-project.org/package=TSA.

[16] D. Qiu, ATSA: Alternative time series analysis, 2015. https://CRAN.R-project.org/package=aTSA.

[17] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeen, forecast: Forecasting functions for time series and linear models, 2020. http://pkg.robjhyndman.com/forecast.

[18] A. Ghalanos, Rugarch: Univariate garch models., 2020.

[19] E. Zivot, Estimation of arma models, (2005). https://faculty.washington.edu/ezivot/econ584/notes/armaestimation.pdf (accessed February 27, 2020).

Appendix

Appendix A

A.1 - Conditional maximum likelihood

If $y_1, y_2, ..., y_N$ is a realization of the i.i.d random variables $Y_1, Y_2, ..., Y_N$ then (16) can be written as

$$L(\theta) = \prod_{t=1}^{N} f(y_t; \theta),$$

since the joint density function of N i.i.d variables is given by the product of their marginal densities. This gives the log-likelihood

$$l(\theta) = \sum_{t=1}^{N} \log(f(y_t; \theta)).$$

For a weakly stationary time series $\{Y_t, t = 1, 2, ..., N\}$ the above does not work since the random variables are not i.i.d. This can be solved by factorising the joint density as a product of conditional densities and the density of the *initial values*, giving

$$L(\theta) = \Big(\prod_{t=p+1}^{N} f(y_t | y_{t-1}, ..., y_1; \theta) \Big) \cdot f(y_p, ..., y_1; \theta),$$

where $y_p, y_{p-1}, ..., y_1$ are called the initial values. The log-likelihood is thus given by

$$l(\theta) = \Big(\sum_{t=p+1}^{N} \log(f(y_t | y_{t-1}, ..., y_1; \theta)) \Big) + \log(f(y_p, ..., y_1; \theta)). \tag{25}$$

The expression on the right-hand side of (25) is called the *exact log-likelihood*, the first term is called the *conditional log-likelihood* and the second term is called the *marginal log-likelihood*. In time series modelling the marginal log-likelihood is not always known, which is why the conditional log-likelihood is sometimes maximised instead of the exact log-likelihood. Some packages used in this thesis maximise the conditional log-likelihood, thus giving the estimation

$$\hat{\theta}_{mle} = \arg\max_{\theta} \Big(\sum_{t=p+1}^{T} \log(f(y_t | y_{t-1}, ..., y_1; \theta)) \Big).$$

When $\{Y_t\}$ is a stationary process it holds that $\hat{\theta}_{mle}$ and $\hat{\theta}_{cmle}$ are consistent and have the same limiting distribution ([19]).

A.2 - Rolling window forecasts of the three candidate models

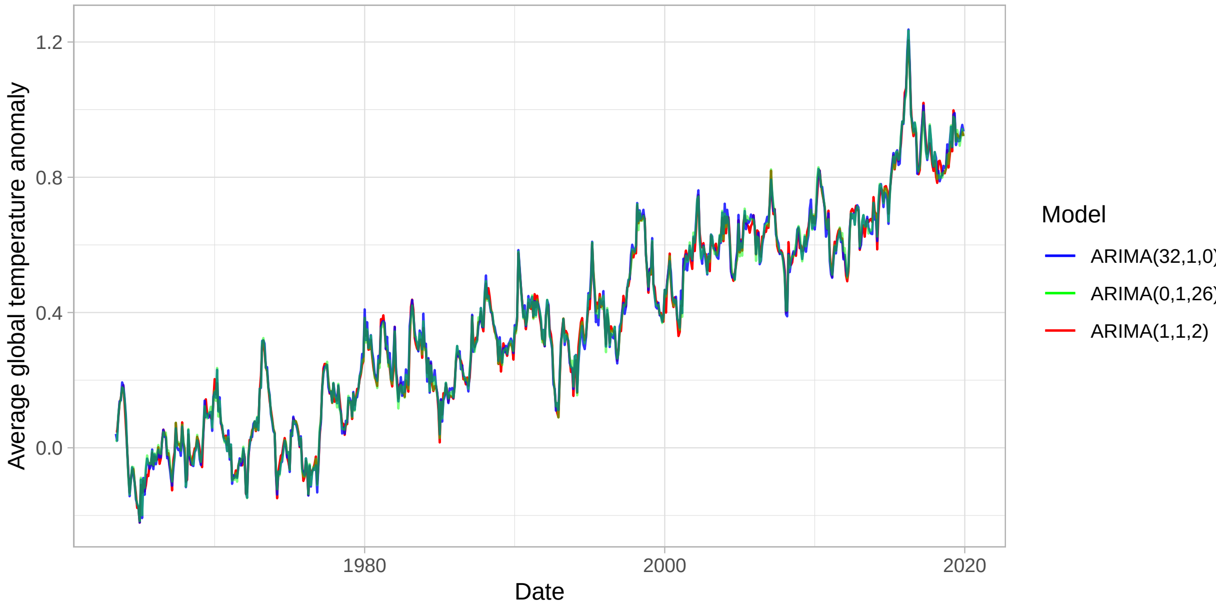

Figure 16: 1-step-ahead rolling window forecasts of the three candidate models

Figure 16 shows that the forecast accuracy is similar for all three models, meaning that choosing ARIMA(1,1,2) as the final model is reasonable.

Note that Figure 16 shows temperature anomalies and not differenced temperature anomalies. Table 1 has been calculated using 1-step-ahead forecasts of the 1001st to the 1679th differenced temperature anomaly. From these, the 1-step-ahead forecasts of the temperature anomalies for months 1002 to 1680 (June 1963 to December 2019) were calculated according to section 4.3.1.

Appendix B - Tests

B.1 Augmented Dickey-Fuller tests

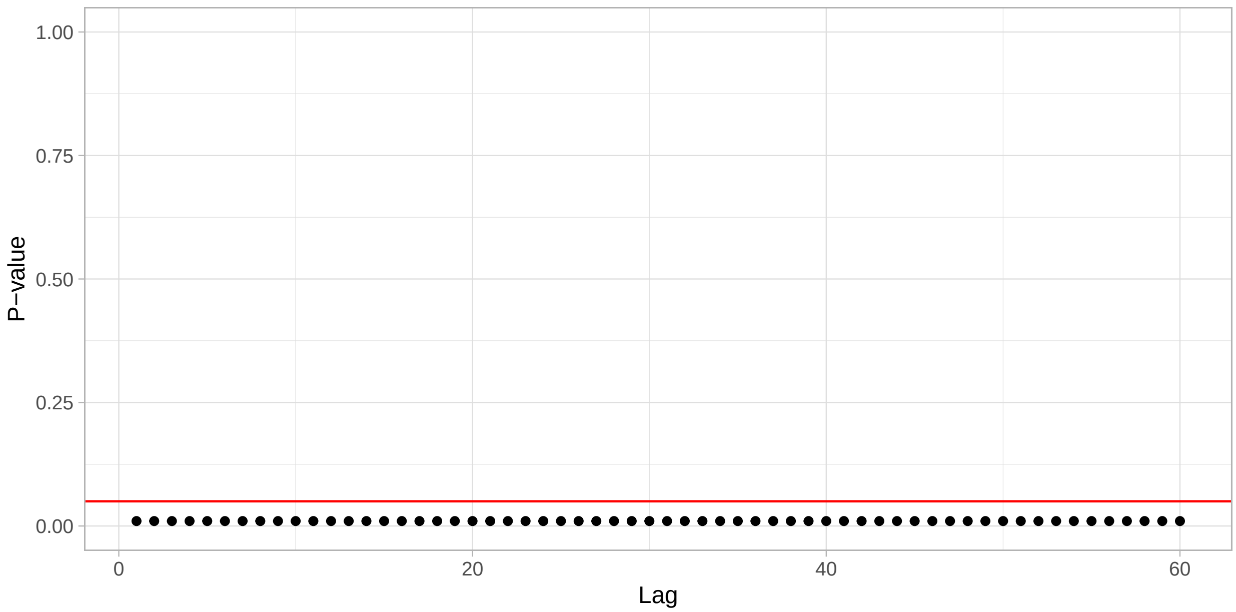

Figure 17: ADF tests on the differenced anomalies. The red line represents the 5% significance level.

B.2 Ljung-Box tests on the residuals from ARIMA(32,1,0)

The first test is on lag 33 because of the adjustment needed on the degrees of freedom as described in section 3.3.1

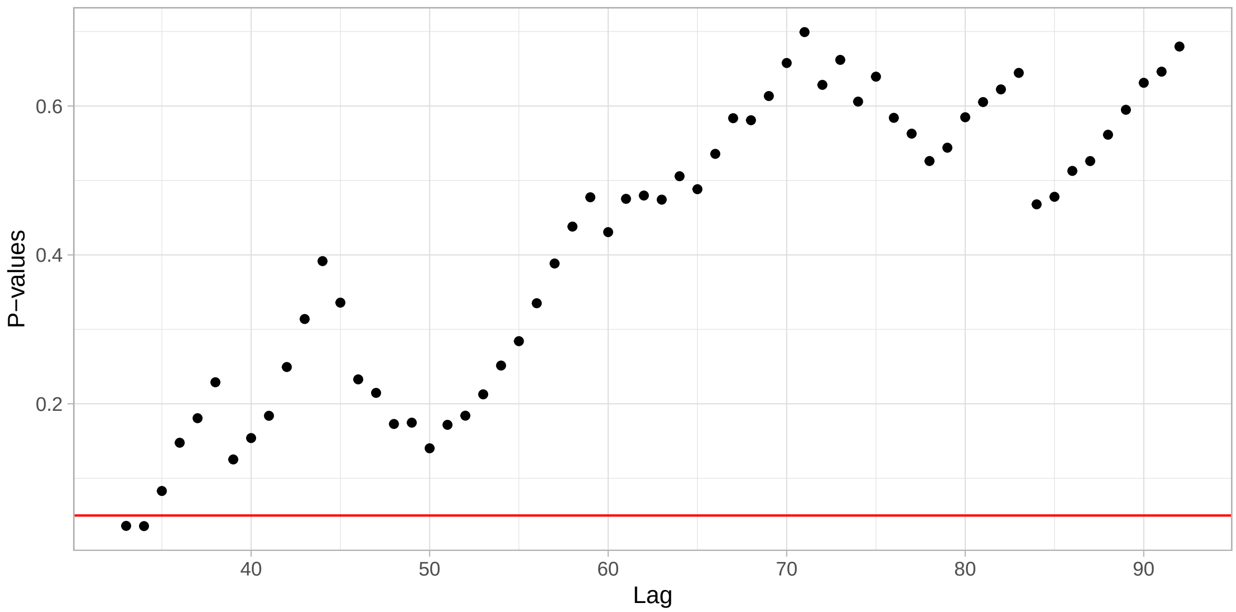

Figure 18: Ljung-Box tests on the residuals from ARIMA(32,1,0). The red line represents the 5% significance level.

B.3 Ljung-Box tests on the residuals from ARIMA(0,1,26)

The first test is on lag 27 because of the adjustments needed on the degrees of freedom as described in section 3.3.1

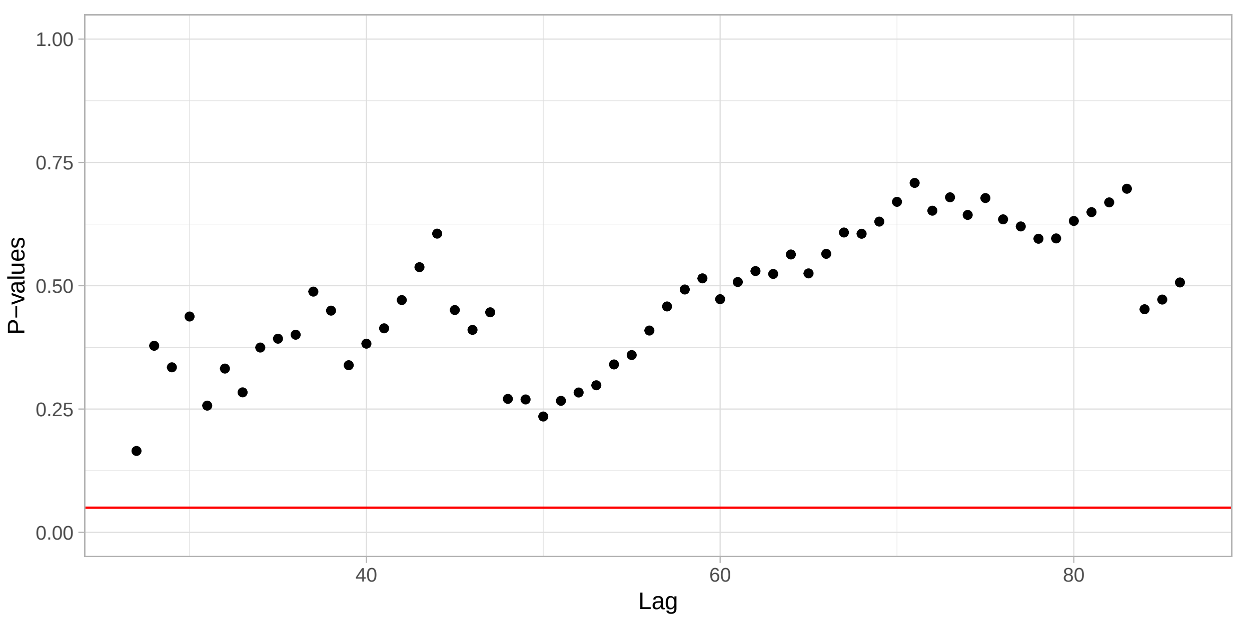

Figure 19: Ljung-Box tests on the residuals from ARIMA(0,1,26). The red line represents the 5% significance level.

B.4 Ljung-Box tests on the residuals from ARIMA(1,1,2)

The first test is on lag 4 because of the adjustments needed on the degrees of freedom as described in section 3.3.1

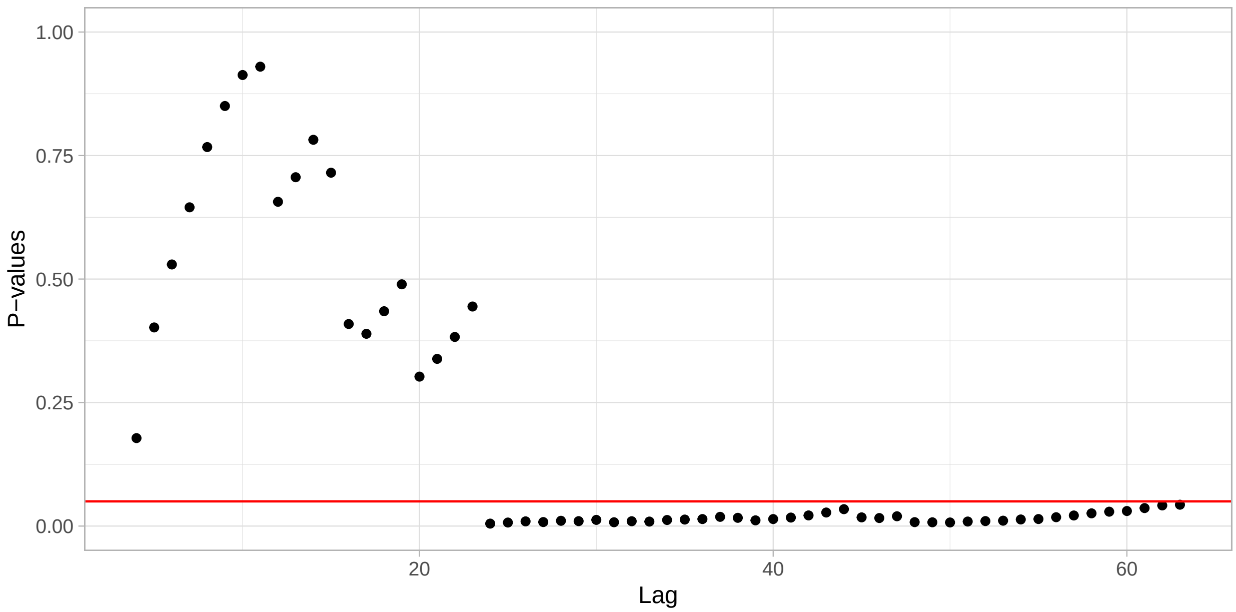

Figure 20: Ljung-Box tests on the residuals from ARIMA(1,1,2). The red line represents the 5% significance level.

B.5 McLeod-Li tests on the residuals from ARIMA(1,1,2)

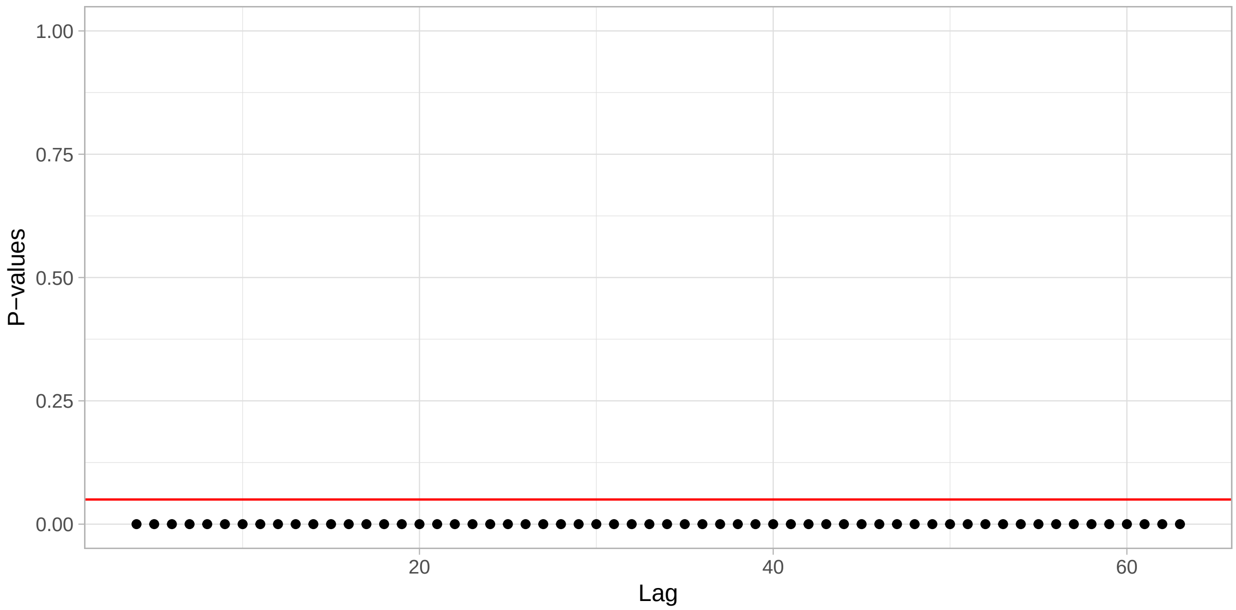

Figure 21: McLeod-Li tests on the residuals from ARIMA(1,1,2). The red line represents the 5% significance level.

B.6 McLeod-Li tests on the adjusted residuals from ARIMA(1,1,2)

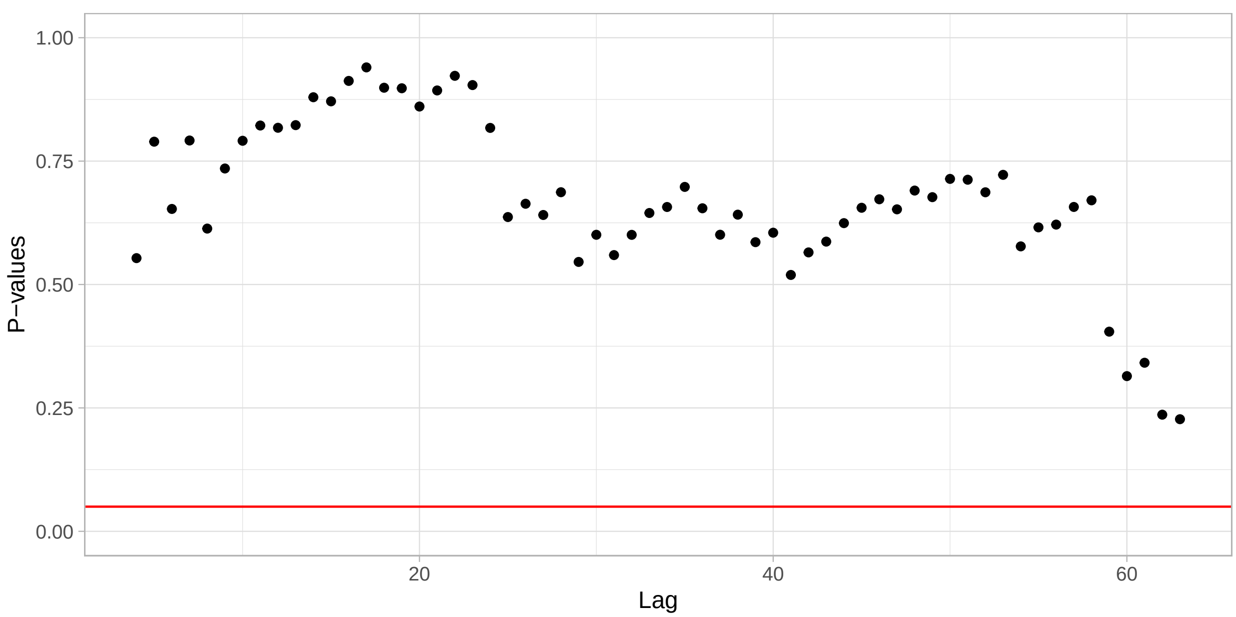

Figure 22: McLeod-Li tests on the adjusted residuals from ARIMA(1,1,2). The red line represents the 5% significance level.

B.7 Ljung-Box tests on the adjusted residuals from ARIMA(1,1,2)

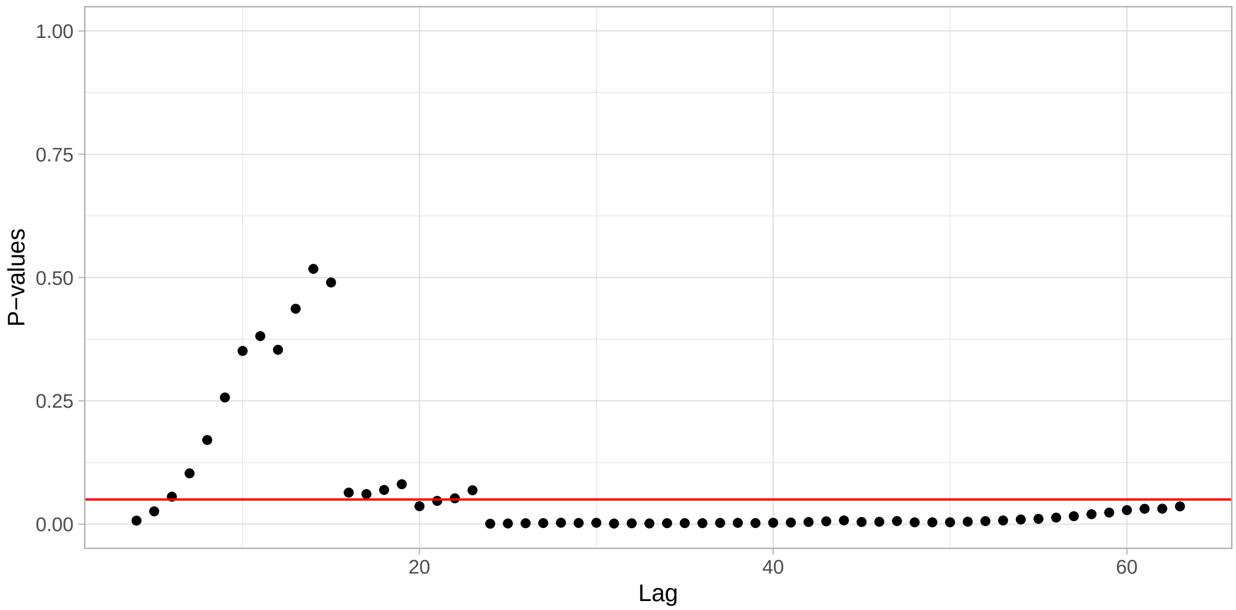

Figure 23: Ljung-Box tests on the adjusted residuals from ARIMA(1,1,2). The red line represents the 5% significance level.

Appendix C - Chosen models

C.1 Chosen integrated AR model

The chosen model is an ARIMA(32,1,0) meaning that the differenced temperature anomalies is thought to follow an AR(32) process. The parameter estimations are given below.

```
        ar1         ar2         ar3         ar4         ar5         ar6 
-0.44776770 -0.27948749 -0.20142705 -0.14164460 -0.13426001 -0.10614102 
        ar7         ar8         ar9        ar10        ar11        ar12 
-0.09167042 -0.08527427 -0.10543660 -0.09887300 -0.08871388 -0.05150687 
       ar13        ar14        ar15        ar16        ar17        ar18 
-0.05828549 -0.06935403 -0.10117541 -0.13469114 -0.12499148 -0.11309393 
       ar19        ar20        ar21        ar22        ar23        ar24 
-0.07594499 -0.12278973 -0.07121306 -0.07903563 -0.06405753  0.05394477 
       ar25        ar26        ar27        ar28        ar29        ar30 
-0.01565507 -0.03695949 -0.06590007 -0.05925833 -0.07530834 -0.06894953 
       ar31        ar32 
-0.09597243 -0.07119898 
```

C.2 Chosen integrated MA model

The chosen model is an ARIMA(0,1,26) meaning that the differenced temperature anomalies is thought to follow an MA(26) process. The parameter estimations are given below.

```
         ma1           ma2           ma3           ma4           ma5 
-0.4550497904 -0.0844425513 -0.0439816479 -0.0181713489 -0.0415750846 
          ma6           ma7           ma8           ma9          ma10 
-0.0094386037 -0.0100315399 -0.0159795754 -0.0436524295 -0.0083558854 
         ma11          ma12          ma13          ma14          ma15 
-0.0180164029  0.0195258970 -0.0184652383 -0.0323338196 -0.0304053605 
         ma16          ma17          ma18          ma19          ma20 
-0.0470220964 -0.0209273278 -0.0020013628  0.0228222623 -0.0604306492 
         ma21          ma22          ma23          ma24          ma25 
 0.0335017085 -0.0033406368  0.0097967535  0.1032919602 -0.0582621548 
         ma26     intercept 
-0.0462323996  0.0006371056 
```

C.3 Chosen integrated ARMA model

The chosen model is an ARIMA(1,1,2) meaning that the differenced temperature anomalies is thought to follow an ARMA(1,2) process. The parameter estimations are given below.

```
         ar1          ma1          ma2    intercept 
 0.874545001 -1.333502962  0.348290782  0.000635113 
```

C.4 Chosen ARCH model

```
       omega       alpha1       alpha2       alpha3       alpha4       alpha5 
8.239251e-04 1.374345e-01 2.409151e-02 1.580401e-02 2.765260e-08 1.354513e-08 
      alpha6       alpha7       alpha8       alpha9      alpha10      alpha11 
1.494467e-09 1.120753e-07 1.403200e-08 2.809388e-02 1.443350e-02 1.081504e-01 
     alpha12      alpha13      alpha14      alpha15      alpha16      alpha17 
1.385458e-01 1.872188e-01 1.802305e-02 2.500918e-07 2.043691e-08 3.054908e-08 
     alpha18      alpha19      alpha20      alpha21      alpha22      alpha23 
1.187249e-08 2.042502e-08 2.707642e-08 1.091579e-03 4.773964e-02 5.636184e-03 
     alpha24      alpha25      alpha26 
1.067252e-01 8.817107e-02 5.372131e-02 
```