Time Series Analysis

Name: **CHEA Makara** Academic Year: 2024-2025

1. If a time series displays a mean function that is not constant, name of two general approaches you could use to achieve a series that is stationary. States an advantage of each approach.

   There are many approaches that can use to get stationary process such as

   - Detrending: In case we find its trend easily. Distracting the trend to process have constant mean and we can consider it is stationary process which mean it is not depend on time. In addition, variance or autocovariance is constant, too.
   - Differencing: Differencing seems like this process will lead to ARIMA model which used differencing of order dth. It can apply if the series has the trends or seasonality.
   - Transforming: by using some transforming approach like log-transformed or power transform. A log transformation helps stabilize variance in a time series, especially when there is exponential growth or heteroscedasticity. This transformed is also reduce the lag scale. Assume the process is Y_t then

     $$E[\log(Y_t)] \approx \log(\mu_t) \qquad \& \qquad V(\log(Y_t)) \approx \sigma^2$$

     For power transformed can also apply but clearly the process consists the positive value (if negative or zero, use the additive with constant terms) and the process can be formed to

     $$g(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log x & \text{for } \lambda = 0 \end{cases}$$

     If the value $\lambda \to 1/2$, this is useful with Poisson distribution.

2. The analysis of time series data must account for the fact that data measured close in time are very often (choose one)

   1. independent
   2. discrete
   3. identical
   4. correlated
   5. uncorrelated

   $\Longrightarrow$ It consider often correlated in time process.

3. A simple trend model for a time series might be specified as$Y_t = \mu_t + W_t$.
   Explain briefly in words what each of and signify in this model.

   - Meaning of μ_t : it means that the process have trend.
   - Meaning of W_t : This refers to the white noise of the process. The white noise must be uncorrelated over the lag k.

4. Given the time series Y_t and denote $e_t \sim iid\,\mathcal{N}(0, \sigma^2)$. Which of the following model is harmonic (or quasi-periodic)? Choose all that apply.
   Note: A harmonic model refers to a time series model that describes or captures oscillatory (cyclical) behavior

a. $Y_t = \mu + a\cos(\omega t) + b + \sin(\omega t) + e_t$
b. $Y_t = Y_{t-1} + e_t$
c. $Y_t = (a + bt)\sin(\omega t + \varphi) + e_t$
d. $Y_t = a + bt + ct^2 + e_t$

$\Longrightarrow$ Answer: 1 & 2

5. A data analyst fit a linear trend with as the response, and AIC of the linear model was 453.7. The analyst fit a quadratic trend model with the same as the response, and the AIC of the quadratic model was 446.2. What conclusions can you draw? Explain.

   The AIC formula is represeted by:

   $AIC = 2k - 2\log(L)$, L is Likelihood function and k is number of parameter

   Low AIC score means the model fitting is better.
   $\Longrightarrow$ The quadratics trend with AIC score is 446.2 is better fit to the response.

6. A data analyst fit a linear trend with Y_t as the response, and AIC of the linear model was 624.3. Because a residual analysis showed possible non constant spread of the residuals, the analyst fit a linear trend model with the logarithm of Y_t as the response, and AIC of the log-transformed linear model was 342.8. What conclusion can you draw? Explain.

   $\Longrightarrow$ She can not compare those linear model with AIC score 624.3 to the linear trend with after use logarithm transformed with AIC score 342.8.

7. A student plotted a time series data and she observed that the series has a constant mean function over time. Then she reached the conclusion that the time series is stationary. Do you think her conclusion is valid ? Why or why not ? Explain.

   $\Longrightarrow$ Her conclusion is not enough information to assume model is stationary. She should have evidence on variance or autocovariance then clearly this variance is not depends on time.

8. For a discrete time series $\{Y_t\}$, the series of the first differences is define as $\nabla Y_t = Y_t - Y_{t-1}$. What is the series of second differences $\nabla^2 Y_t =$?

   a. $Y_t - 2Y_{t-1}$
   b. $Y_t - 2Y_{t-1} + Y_{t-2}$
   c. $Y_t - 2Y_{t-1} - Y_{t-2}$
   d. $Y_t - Y_{t-2}$
   e. $Y_t - Y_{t-1} - Y_{t-2}$

   Consider $\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla Y_t - \nabla Y_{t-1} = Y_t - Y_{t-1} - Y_{t-1} + Y_{t+2} = Y_t - 2Y_{t-1} + Y_{t+2}$

   Answer is (b)

9. Let $\{Y_t\}$ be a stationary process. For such a process, how does $V(Y_t)$ compare to $V(Y_{t-l})$ for $l > 0$? Explain.

   $\Longrightarrow$ For stationary process, the variance of process does not depend on time. It means that

$$V(Y_t) = V(Y_{t-1})$$
$$V(Y_{t-1}) = \ldots = V(Y_{t-l}), \quad l > 0$$
$$\Longleftrightarrow V(Y_t) = V(Y_{t-l}) = \sigma^2$$

10. Consider two random variables X and Y. Suppose $E(X) = 3, V(X) = 9, E(Y) = 0, V(Y) = 4$, and $corr(X, Y) = -0.2$. Find the following, showing all your steps:

   a. $V(X + Y)$

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) - [E(X+Y)]^2 \\ &= E[X^2] + 2E[XY] + E[Y^2] - [E[X]^2 + 2E[X]E[Y] + E[Y]^2] \\ &= V(X) + V(Y) + 2Cov(X,Y) \\ &= V(X) + V(Y) + 2\sqrt{V(X)V(Y)}Corr(X,Y) \\ &= 9 + 4 + 2\sqrt{9 \times 4} \times (-0.2) \\ &= 13 - 2.4 = 10.6 \end{aligned}$$

   b. $Corr(X + Y, 2X - Y)$

$$\begin{aligned} Corr(X+Y, 2X-Y) &= \frac{Cov(X+Y, 2X-Y)}{\sqrt{V(X+Y)V(2X-Y)}} \\ \text{Consider} \quad V(2X-Y) &= 4V(X) + V(Y) + 4Cov(X,Y) \\ &= 4 \times 9 + 4 + 2 \times (2.4) = 44.8 \\ Cov(X+Y, 2X-Y) &= E[(X+Y)(2X-Y)] - [E(X+Y)]E[(2X-Y)] \\ &= E[2X^2 + XY - Y^2] - \Big[\big(E[X] + E[Y]\big)\big(2E[X] - E[Y]\big)\Big] \\ &= 2E[X^2] + E[XY] - E[Y^2] - 2E[X]^2 \\ &= 2V[X] + \sqrt{V(X)V(Y)}Cov(X,Y) - \big(V(Y) + E[Y]^2\big) \\ &= 2 \times 9 - 6 \times 0.2 - 4 = 18 - 5.2 = 12.8 \\ \Longrightarrow Corr(X+Y, 2X-Y) &= \frac{12.8}{\sqrt{10.6 \times 44.8}} = 0.5873 \end{aligned}$$

11. Let $\{e_t\}$ be a normal white noise process with mean 0 and variance 1. Let $\{Y_t\}$ be defined as $Y_t = e_t + 0.5e_{t-2}$. Answer the following.

   (a) The process $\{Y_t\}$ is

      i. AR(1)
      ii. AR(2)
      iii. MA(1)
      iv. MA(2)
      v. ARIMA(0,1,2)

   $\Longrightarrow$ The process is Moving Average of order 2, that is MA(2)

   (b) Find the autocovaraince function of this process. Write the autocovariance function as a piecewise function for various values of the lags, specifically for lags $k = 0, 1, 2$ and $k > 2$. Show all your working process.

   Proof. We have MA(2): $Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$, where $\theta_1 = 0, \theta_2 = 0.5$
   Moreover, $\{e_t\} \sim iid \;\; \mathcal{N}(\mu, \sigma_e^2)$,where $\mu = 0, \sigma_e^2 = 1$ (Standard normal).

Consider

$\gamma_0 = Cov(Y_t, Y_t) = V(Y_t) = V(e_t) + \theta_1^2 V(e_{t-1}) + \theta_2^2 e_{t-2} = (1 + \theta_1^2 + \theta_2^2)\sigma_e^2$
For $\quad \gamma_k = Cov(Y_t, Y_{t-k)} = Cov\left(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-k} - \theta_1 e_{t-k-1} - \theta_2 e_{t-k-2}\right)$
In case $\quad k \geq 3 \rightarrow \gamma_k = 0$
We will consider for the case $k = 0, 1, 2$

- For $k = 0, \quad \gamma_0$
- For $\quad k = 1 \Longrightarrow \gamma_1 = Cov(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3})$
  $= Cov(-\theta_1 e_{t-1}, e_{t-1}) + Cov(-\theta_2 e_{t-2}, -\theta_1 e_{t-2})$
  $= -\theta_1\sigma_e^2 + \theta_1\theta_2\sigma_e^2 = (-\theta_1 + \theta_1\theta_2)\sigma_e^2$
- For $k = 2 \Longrightarrow \gamma_2 = Cov(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-2} - \theta_1 e_{t-3} - \theta_2 e_{t-4})$
  $= Cov(-\theta_2 e_{t-2}, e_{t-2}) = -\theta_2\sigma_e^2$

We obtain

$$\gamma_k = \begin{cases} (1+\theta_1^2+\theta_2^2)\sigma_e^2 & \text{if} \quad k = 0, \\ (-\theta_1 + \theta_1\theta_2)\sigma_e^2 & \text{if} \quad k = 1, \\ -\theta_2\sigma_e^2 & \text{if} \quad k = 2 \\ 0 & k \geq 3 \end{cases}$$

$$\text{Therefore,} \quad \gamma_k = \begin{cases} 1.25 & \text{if} \quad k = 0, \\ 0 & \text{if} \quad k = 1, \\ -0.5 & \text{if} \quad k = 2 \\ 0 & k \geq 3 \end{cases}$$

□

(c) Find the autocorrelation function of this process. Write the autocorrelation function as a piecewise function for various values of the lags as you considered in part(b)

Proof. For autocorrelation function we have $\rho_k = \dfrac{\gamma_k}{\gamma_0}$, in general,by using the model from part (b) we get $\rho_k = \begin{cases} 1 & \text{if} \quad k = 0, \\ \dfrac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & \text{if} \quad k = 1, \\ -\dfrac{\theta_2}{1+\theta_1^2+\theta_2^2} & \text{if} \quad k = 2, \\ 0 & \text{if} \quad k \geq 3 \end{cases}$

$$\Longrightarrow \rho_k = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k = 1, \\ -\dfrac{0.5}{1.25} = -0.4 & \text{if } k = 2, \\ 0 & \text{if } k \geq 3 \end{cases}$$

□

12. Let $e_t \sim iid \quad \mathcal{N}(0, \sigma^2)$ and $Y_t = e_t e_{t-1}$. Calculate mean and covariance functions of Y_t. Conclude on stationary of the process Y_t.

Proof. Calculate mean and covaraince
We have $e_t \sim iid \quad \mathcal{N}(0, \sigma^2)$

$$\begin{aligned} Y_t &= e_t e_{t-1} \\ \Longrightarrow E[Y_t] &= E[e_t e_{t-1}] = E[e_t]E[e_{t-1}] = 0 \qquad (1) \\ \text{covariance} \quad Cov(Y_t, Y_{t-k}) &= Cov(e_t e_{t-1}, e_{t-k} e_{t-k-1}) \\ \text{If } k = 0, \text{that is variance}, V(Y_t) &= V(e_t e_{t-1}) \\ &= E[(e_t e_{t-1})^2] - [E(e_t e_{t-1})]^2 \\ &= E[e_t^2]E[e_{t-1}^2] = V(e_t)V(e_{t-1}) \\ &= \sigma_e^2 \sigma_e^2 = \sigma_e^4 \quad \text{variance does not depend on time } t \quad (2) \end{aligned}$$

From (1) & (2) We can conclude this process is stationary. □

13. Consider the MA(2) model given by $Y_t = e_t - \frac{2}{3}e_{t-1} + \frac{1}{4}e_{t-2}$. Is this model invertible? Explain.

Proof. We have

$$Y_t = e_t - \frac{2}{3}e_{t-1} + \frac{1}{4}e_{t-2}$$

Associate characteristic polynomial equation:

$$\Theta(x) = 1 - \frac{2}{3}x + \frac{1}{4}x^2$$

$$\text{Set} \quad \Theta(x) = 0$$

$$\Longleftrightarrow 1 - \frac{2}{3}x + \frac{1}{4}x^2 = 0$$

$$\Longleftrightarrow 3x^2 - 8x + 12 = 0$$

$$\Longrightarrow x_{1,2} = \frac{-b \pm \sqrt{\Delta}}{2a} = \frac{4 \pm 2i\sqrt{5}}{3}$$

$$\text{We have} \quad |x_1| = |x_2| = |x_{1,2}| = \sqrt{(4/3)^2 + (2\sqrt{5}/3)^2} = 2 > 1$$

Therefore this model of moving average MA(2) is invertible. □

14. This question concerns the empirical identification of ARIMA models.

(a) The sample ACFs for series ans its first difference are given below (n=100)

	lag						
ACF	1	2	3	4	5	6	7
Z_t	0.55	0.58	0.53	0.53	0.52	0.53	0.43
ΔZ_t	-0.55	0.063	0.026	0.041	0.073	0.15	-0.18

Based on this information, which ARIMA model(s) would you consider for the series. Explain your reasoning.

Proof. For ARIMA model(s),
We need to consider 3 parameters such as ARIMA(p, d, q)

- Parameter p is the order of autoregressive model, based on ACF plot for Z_t, we observe that the graph is slightly tails off then we are note that it is not stationary and this behavior seems consist the AR model with order $p \geq 1$.
- Parameter d, the graph is illustrated only 1 time differencing, then ACF plot show the cut off behavior after lag 1. We get exactly information that $d = 1$, but parameter q seems not clearly. However, we can assume $q \geq 1$ from ACF plot at ∇Z_t.
- Due to the lack information of process like PACF and other criteria, the beginning we can take ARIMA$(p \geq 1, d = 1, q \geq 1)$

□

(b) The time series plot of 100 observations of variable Z_t and its ACF and PACF plot are shown in Figure 1. Also shown in the figure are similar plots for the log-transformed data. Based on this information, state whether the data should be analyzed to the original scale or the logarithmic scale, and which ARIMA model(s) would you consider for the series. Explain.

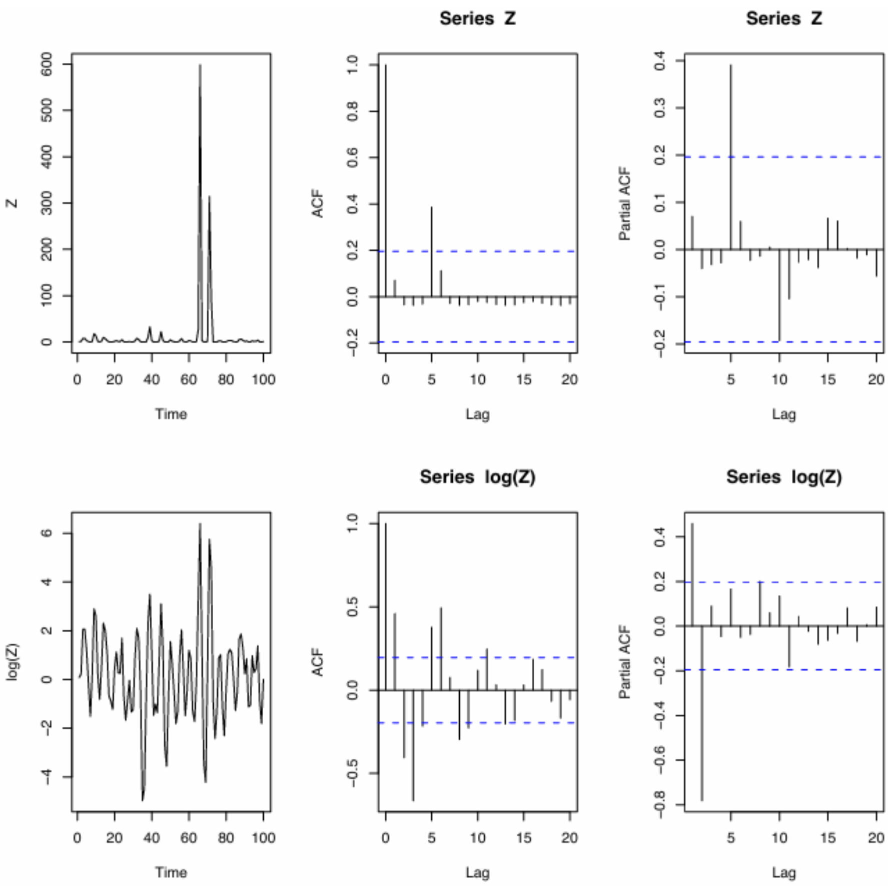

Figure 1:

Proof. Consider the ARIMA Model(s).

- The first graph are shown the original series Z over time t, ACF plot, and PACF plot respectively. In addition, the second graph are illustrated the series Z after applied log-transformed.
- We know clearly that the series is not stationary process that is ARIMA(p, d, q) model is considered to fit this series.
- The ACF and PCAF plot are oscillating, we might set $p \geq 1, q \geq 0$.
- After applied log-transformed as shown on the graph, we can assume that the process can be stationarity. Clearly, $d = 0$. If the log-transformed of series Z is not stationary, we might need more criteria for testing stationary, that is Augmented Dickey-Fuller (ADF).
- One more thing to get parameter p and q, have a look to ACF and PACF plot. Nevertheless, both ACF and PACF shows the behavior of AR model. If we look at the PACF plot, we see the process is not significantly after log 2 or we say that it is cut off after lag 2. We can conclude $p \geq 2$ and for parameter q is still need more observation because of the complexity of the series.
- All in all, we can set the model ARIMA$(p \geq 2, d = 0, q \geq 0)$

□

15. An AR(1) model was fitted to a time series with 100 observations.

```
> fit1

Call:
arima0(x = Z, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.4788     0.0966

Approx standard errors:
         ar1  intercept
      0.0898     0.2495

sigma^2 estimated as 1.722:  log likelihood = -169.2,  aic = 342.41
```

Figure 2 and 3 show some of the diagnostics check with the fitted AR(1) model. Comment on goodness of for of the AR(1) model. In particular, comment on whether the residuals (i) have any outliers, (ii) conform to the idd normal assumption, and (iii) are of constant variance. If you judge the fiited AR(1) model inadequate, that is, fit is poor, what model would you consider for the data, in view of the information from the residuals. You must explain your comments to get credits.

Proof. Based on graphs,

- Based on given result, we fit ARIMA $(1, 0, 0)$ or just AR(1).
- From figure 2 and 3, we see that the standardized residuals plot and ACF plot and PACF plot respectively. Those plot we can draw a conclusion that outlier is not recognized or less significantly after lag 0. AS we can see the standardized residuals plot consists the stability. This information is also provided us that residuals have a constant variance.

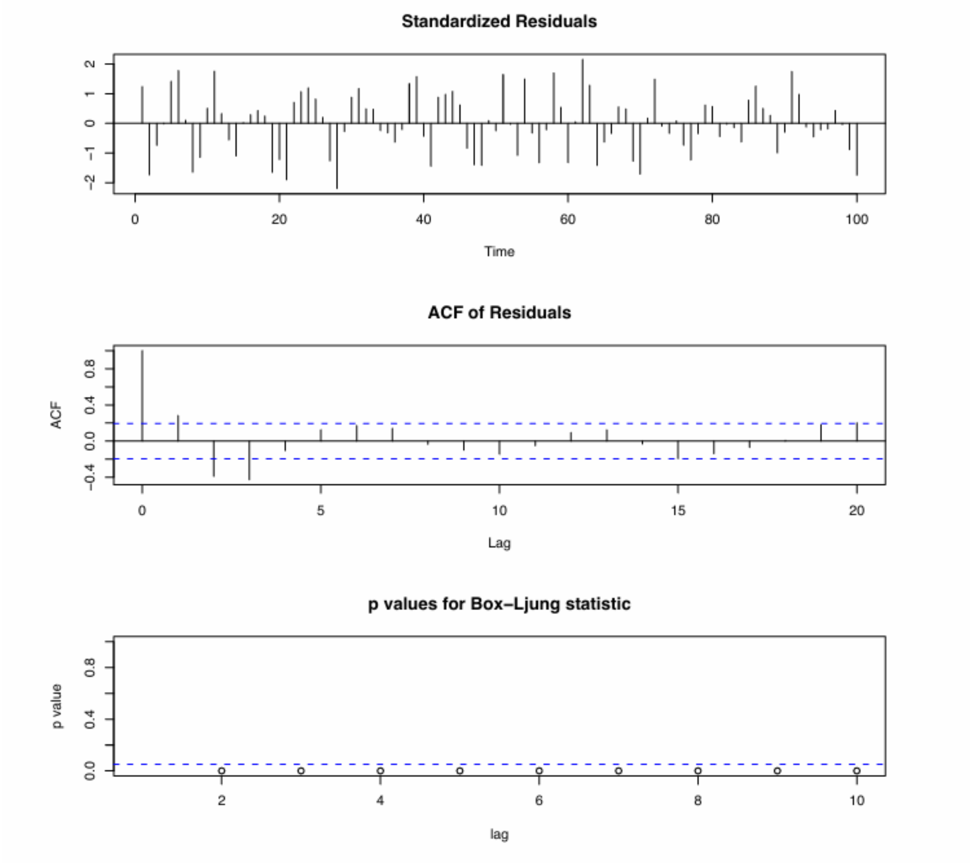

Figure 2: The upper figure shows the standardized residual plot, the middle figure shows the ACF of the residuals, and the bottom figure shows the p-value of the Box-Ljung statitics.

- If we observe the Q-Q plot in figure 3, the residual can be accepted the normality assumption because its correlation is performed on the straight line. Similarity, Ljung-Box plot in figure 2 shown the p-value is around 0.05. We can assume that p-value > 0.05 then H_0 is not rejected (H_0 : residuals is white noise). The Null Hypothesis give us the white noise is uncorrelated. We also proof that its variance is constant. Consequently, residuals are iid Normal distribution.

□

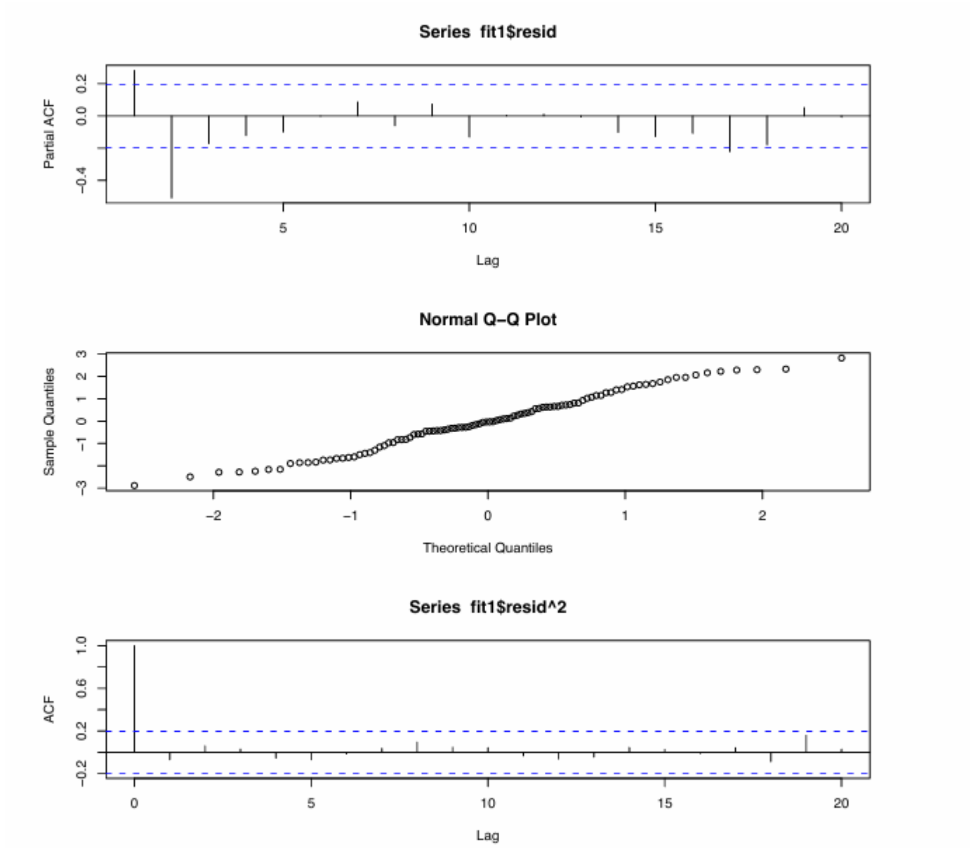

Figure 3: The upper figure shows the pacf of the residuals, the middle figure is the q-q normal plot of the residuals and the bottom figure shows the ACF of the squared residuals.