Analysis of Time Series 1MS014
Assignment 1

Viktor Ågren

April 2022

1 Problem 1

Let $\{w_t\}$, $t = 0, \pm 1, \pm 2, \ldots$, be a white noise process where the wt are simultaneously independent with variance $\sigma_w^2 = 4$, and let

$$x_t = w_t w_{t-1} + 0.4 w_{t-2} w_{t-3}.$$

Calculate the mean and autocovariance function of x_t and state whether it is stationary.

Solution We define the mean function of x_t

Definition 1 (Definition 1.1 in [1]). The mean function is defined as

$$\mu_t = \mathrm{E}(x_t) = \int_{-\infty}^{\infty} x ft(x) dx.$$

provided it exists, where E denotes the usual expected value operator.

By Definition 1 we compute the mean function as

$$\begin{aligned}
\mu_t &= \mathrm{E}[x_t] \\
&= \mathrm{E}[w_t w_{t-1} + 0.4 w_{t-2} w_{t-3}] \\
&= \mathrm{E}[w_t w_{t-1}] + 0.4\mathrm{E}[w_{t-2} w_{t-3}] \\
&= \mathrm{E}[w_t] E[w_{t-1}] + 0.4\mathrm{E}[w_{t-2}]\mathrm{E}[w_{t-3}],
\end{aligned}$$

where we utilize the property that the white noise processes are independent. Since the mean of a white noise process is zero we get that

$$\mu_t = 0.$$

Definition 2 (Definition 1.2 in [1]). The autocovariance function is defined as the second moment product

$$\gamma(s, t) = \mathrm{Cov}(x_s, x_t) = \mathrm{E}[(x_s - \mu_s)(x_t - \mu_t)],$$

for all s and t.

Finding the autocovariance of the function x_t by Definition 2

$$\begin{aligned}\gamma(s,t) &= \mathrm{Cov}(x_s, x_t)\\ &= \mathrm{Cov}(w_s w_{s-1} + 0.4 w_{s-2} w_{s-3}, w_t w_{t-1} + 0.4 w_{t-2} w_{t-3}),\end{aligned}$$

when $s = t$

$$\begin{aligned}\gamma(t,t) &= \mathrm{Cov}(x_t, x_t)\\ &= \mathrm{Cov}(w_t w_{t-1} + 0.4 w_{t-2} w_{t-3}, w_t w_{t-1} + 0.4 w_{t-2} w_{t-3})\\ &= \mathrm{Cov}(w_t, w_t)\mathrm{Cov}(w_{t-1}, w_{t-1}) + 0.4^2 \mathrm{Cov}(w_{t-2}, w_{t-2})\mathrm{Cov}(w_{t-3}, w_{t-3})\\ &= \sigma_w^2 \sigma_w^2 + 0.4^2 \sigma_w^2 \sigma_w^2 = 1.16\sigma^4 = 1.16 \times 4^2 = 18.56,\end{aligned}$$

when $s = t + 1$

$$\begin{aligned}\gamma(t+1,t) &= \mathrm{Cov}(x_{t+1}, x_t)\\ &= \mathrm{Cov}(w_{t+1} w_t + 0.4 w_{t-1} w_{t-2}, w_t w_{t-1} + 0.4 w_{t-2} w_{t-3})\\ &= 0,\end{aligned}$$

when $s = t - 1$

$$\begin{aligned}\gamma(t-1,t) &= \mathrm{Cov}(x_{t-1}, x_t)\\ &= \mathrm{Cov}(w_{t-1} w_{t-2} + 0.4 w_{t-3} w_{t-4}, w_t w_{t-1} + 0.4 w_{t-2} w_{t-3})\\ &= 0,\end{aligned}$$

when $s = t + 2$

$$\begin{aligned}\gamma(t+2,t) &= \mathrm{Cov}(x_{t+2}, x_t)\\ &= \mathrm{Cov}(w_{t+2} w_{t+1} + 0.4 w_t w_{t-1}, w_t w_{t-1} + 0.4 w_{t-2} w_{t-3})\\ &= 0.4\mathrm{Cov}(w_t, w_t)\mathrm{Cov}(w_{t-1}, w_{t-1}) = 0.4\sigma_w^4 = 6.4,\end{aligned}$$

when $s = t - 2$

$$\begin{aligned}\gamma(t-2,t) &= \mathrm{Cov}(x_{t-2}, x_t)\\ &= \mathrm{Cov}(w_{t-2} w_{t-3} + 0.4 w_{t-4} w_{t-5}, w_t w_{t-1} + 0.4 w_{t-2} w_{t-3})\\ &= 0.4\mathrm{Cov}(w_{t-2}, w_{t-2})\mathrm{Cov}(w_{t-3}, w_{t-3}) = 0.4\sigma_w^4 = 6.4.\end{aligned}$$

It is easy to see that for any $|s-t| > 2$ we get zero. Hence, the autocovariance function for x_t is

$$\gamma(s,t) = \begin{cases} 18.56 & \text{if } s - t = 0,\\ 6.4 & \text{if } |s-t| = 2,\\ 0 & \text{otherwise.}\end{cases}$$

The process has mean zero and its covariance is independent of time, therefore x_t is stationary and we can write the autocovariance function (Definition 1.8 in [1]) of x_t as

$$\gamma(h) = \begin{cases} 18.56 & \text{if } h = 0,\\ 6.4 & \text{if } |h| = 2,\\ 0 & \text{otherwise.}\end{cases}$$

Answer The mean of x_t is $\mu_t = 0$, the autocovariance function is

$$\gamma(h) = \begin{cases} 18.56 & \text{if } h = 0, \\ 6.4 & \text{if } |h| = 2, \\ 0 & \text{otherwise,} \end{cases}$$

and it is stationary.

2 Problem 2

Let $\{x_t\}$ be a time series that may be described by

$$x_t = w_t + 0.3w_{t-1} - 0.1w_{t-2},$$

where w_t is normally distributed white noise with variance $\sigma_w^2 = 0.5$.
Calculate the autocorrelation function $\rho(h)$ for all integer h.

Solution First, we need to calculate the autocovariance to obtain the necessary variables to solve for the autocorrelation function, therefore we solve

$$\gamma(h) = \text{Cov}(x_{t+h}, x_t) = \text{Cov}(w_{t+h} + 0.3w_{t+h-1} - 0.1w_{t+h-2}, w_t + 0.3w_{t-1} - 0.1w_{t-2}).$$

When $h = 0$

$$\begin{aligned}\gamma(0) &= \text{Cov}(x_{t+0}, x_t)\\ &= \text{Cov}(w_t + 0.3w_{t-1} - 0.1w_{t-2}, w_t + 0.3w_{t-1} - 0.1w_{t-2})\\ &= \text{Cov}(w_t, w_t) + 0.3^2\text{Cov}(w_{t-1}, w_{t-1}) + (-0.1)^2\text{Cov}(w_{t-2}, w_{t-2})\\ &= \sigma_w^2 + 0.09\sigma_w^2 + 0.01\sigma_w^2 = 1.1\sigma_w^2 = 0.55,\end{aligned}$$

when $h = 1$

$$\begin{aligned}\gamma(1) &= \text{Cov}(x_{t+1}, x_t)\\ &= \text{Cov}(w_{t+1} + 0.3w_t - 0.1w_{t-1}, w_t + 0.3w_{t-1} - 0.1w_{t-2})\\ &= 0.3\text{Cov}(w_t, w_t) - 0.03\text{Cov}(w_{t-1}, w_{t-1})\\ &= 0.3\sigma_w^2 - 0.03\sigma_w^2 = 0.27\sigma_w^2 = 0.135,\end{aligned}$$

when $h = -1$

$$\begin{aligned}\gamma(-1) &= \text{Cov}(x_{t-1}, x_t)\\ &= \text{Cov}(w_{t-1} + 0.3w_{t-2} - 0.1w_{t-3}, w_t + 0.3w_{t-1} - 0.1w_{t-2})\\ &= 0.3\text{Cov}(w_{t-1}, w_{t-1}) - 0.03\text{Cov}(w_{t-2}, w_{t-2})\\ &= 0.3\sigma_w^2 - 0.03\sigma_w^2 = 0.27\sigma_w^2 = 0.135,\end{aligned}$$

when $h = 2$

$$\begin{aligned}\gamma(2) &= \text{Cov}(x_{t+2}, x_t)\\ &= \text{Cov}(w_{t+2} + 0.3w_{t+1} - 0.1w_t, w_t + 0.3w_{t-1} - 0.1w_{t-2})\\ &= -0.1\text{Cov}(w_t, w_t) = -0.1\sigma_w^2 = -0.05,\end{aligned}$$

when $h = -2$

$$\begin{aligned}\gamma(-2) &= \text{Cov}(x_{t-2}, x_t)\\ &= \text{Cov}(w_{t-2} + 0.3w_{t-3} - 0.1w_{t-4}, w_t + 0.3w_{t-1} - 0.1w_{t-2})\\ &= -0.1\text{Cov}(w_{t-2}, w_{t-2}) = -0.1\sigma_w^2 = -0.05.\end{aligned}$$

Additionally, we see that for any $|h| > 2$ implies $\gamma(h) = 0$, therefore we can show that the autocovariance function is

$$\gamma(h) = \begin{cases} 0.55 & \text{if } h = 0, \\ 0.135 & \text{if } |h| = 1, \\ -0.05 & \text{if } |h| = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Since the time series can easily be shown to be mean zero and the autocovariance function does not depend on time, then the series is stationary and we can compute the autocorrelation function as per definition

Definition 3 (Definition 1.9 in [1])**.** The autocorrelation function (ACF) of a stationary time series will be written using (1.15, see [1]) as

$$\rho(h) = \frac{\gamma(t+h,t)}{\gamma(t+h,t+h)\gamma(t,t)} = \frac{\gamma(h)}{\gamma(0)}$$

$-1 \leq \rho(h) \leq 1, \forall h.$

When $h = 0$, the solution is $\rho(0) = 1$, when $h = \pm 1$

$$\rho(\pm 1) = \frac{\gamma(1)}{\gamma(0)} = \frac{0.135}{0.55} = \approx 0.2455,$$

when $h = \pm 2$

$$\rho(\pm 2) = \frac{\gamma(1)}{\gamma(0)} = \frac{-0.05}{0.55} \approx -0.091.$$

Thus, the ACF is

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0, \\ 0.2455 & \text{if } |h| = 1, \\ -0.091 & \text{if } |h| = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Answer The Autocorrelation function for all integers h is

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0, \\ 0.2455 & \text{if } |h| = 1, \\ -0.091 & \text{if } |h| = 2, \\ 0 & \text{otherwise.} \end{cases}$$

3 Problem 3

Let $\{x_t\}$ be a time series that may be described by

$$x_t = 0.5x_{t-1} + w_t + 0.3w_{t-1} - 0.1w_{t-2}$$

where w_t is normally distributed white noise with variance $\sigma_w^2 = 0.5$.
We observe $x_1, x_2, \dots, x_{100}$, where the last five observations are $x_{96} = 0.1, x_{97} = 0.2, x_{98} = 0.3, x_{99} = 0.2$ and $x_{100} = 0.1$.

(a) Is $\{x_t\}$ invertible and casual?

(b) Predict the values of x_{101} and x_{102}. Approximations are permitted.

(c) Calculate prediction intervals for x_{101} and x_{102}, each with (approximate) confidence level 95%.

(d) What would be the 95% prediction interval of x_{200}, and why?

Solution (a)

Definition 4 (Definition 3.5 in [1])**.** A time series $\{x_t;\ t = 0, \pm 1, \pm 2, \dots\}$ is **ARMA**(p, q) if it is stationary and

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

with $\phi_p \neq 0, \theta_q \neq 0$, and $\sigma_w^2 > 0$. The parameters p and q are called the autoregressive and the moving average orders, respectively, where $w_t \sim wn(0, \sigma_w^2)$.

By definition 4, the time series $\{x_t\}$ is ARMA(p, q), we write it in operator form

$$(1 - 0.5B)x_t = (1 + 0.3B - 0.1B^2)w_t.$$

We observe that this is an ARMA(1,2) process, but first we check for parameter redundancy and notice that

$$\begin{aligned}\phi(B) &= (1 - 0.5B),\\ \theta(B) &= (1 + 0.3B - 0.1B^2) = (1 - 0.2B)(1 + 0.5B),\end{aligned}$$

and therefore no common factors that can be cancelled.

Definition 5 (Definition 3.6 in [1])**.** The AR and MA polynomials are defined as

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z_p, \quad \phi_p \neq 0,$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z_q, \quad \theta_q \neq 0,$$

respectively, where z is a complex number.

By definition 5 we write the AR and MA polynomials as

$$\phi(z) = 1 - 0.5z,$$

and

$$\theta(z) = (1 - 0.2z)(1 + 0.5z).$$

Theorem 1 (Theorem (Property 3.1) in Lecture 4)**.** *An ARMA(p, q) model is causal if and only if $\phi(z) \neq 0$ for all $|z| \leq 1$.*

Theorem 2 (Theorem (Property 3.2) in Lecture 4)**.** *An ARMA(p, q) model is invertible if and only if $\theta(z) \neq 0$ for all $|z| \leq 1$.*

The ARMA(1,2) model is causal since the AR polynomial $\phi(z) = (1 - 0.5z) = 0$ when $z = 2$. The model is also invertible since the roots of $\theta(z) = (1-0.2z)(1+0.5z) = 0$ when $z_1 = 5$ and $z_2 = -2$.

Solution (b)-(d) No solution

Answer (a) The ARMA(1,2) model is invertible and causal

4 Problem 4

Swedish fuel consumption for power plants producing electricity, gas or heating, using peat fuels ('trädbränsle'), for 2009:1-2020:4 (quarterly data) is given in the file bransle.dat at Studium.
Find a suitable ARIMA (or SARIMA) model for these data, or a transformation thereof. Analyze the model residuals carefully, in order to make sure that the model provides a good description of the data.

Solution Initially we are going to construct a time series and plot for the fuel consumption data. Notice that the data is quarterly, therefore we create a time series with frequency 4, starting in Q1 2009. A simple command to plot a time series with the autocorrelation function and partial auto-correlation function is "ggtsdisplay" from the forecast package.

```
library(forecast)
library(lmtest)
x<-read.table("bransle.dat")$V1
timeseries<-ts(x, frequency=4,start=c(2009,1))
ggtsdisplay(timeseries)
```

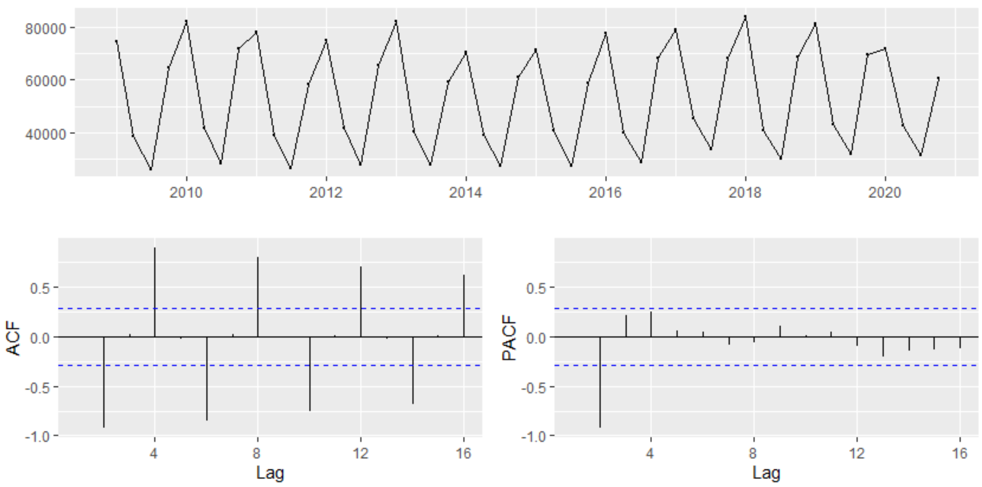

Figure 1: Swedish fuel consumption (time series, AFC, and PACF)

The next step is to inspect Figure 1 to check for any anomalies. Straight away we notice that the time series has obvious seasonality, i.e., the characteristic of the time series experiences regular and predictable changes that recur every calendar year. In this case, it reaches a high during the winter months and a low during the summer months, which seems to make sense regarding fuel consumption for power plants producing electricity, gas or heating. We do not however see any noticeable change in variance throughout time, but just to make sure we will apply the Box-Cox transformation

$$y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases}$$

In R we utilize the BoxCox.lambda function, which chooses the value of λ that minimizes the coefficient of variation.

```
lambda <- BoxCox.lambda(temp.global)
lambda

> [1] -0.1287132
```

i.e. we get a $\lambda = -0.1287132$, and then it is possible to compare the original time series Figure 1 to the time series Figure 2 that uses Box-Cox transformation.

```
timeseries.BoxCox<-BoxCox(timeseries, lambda=lambda)
ggtsdisplay(timeseries.BoxCox)
```

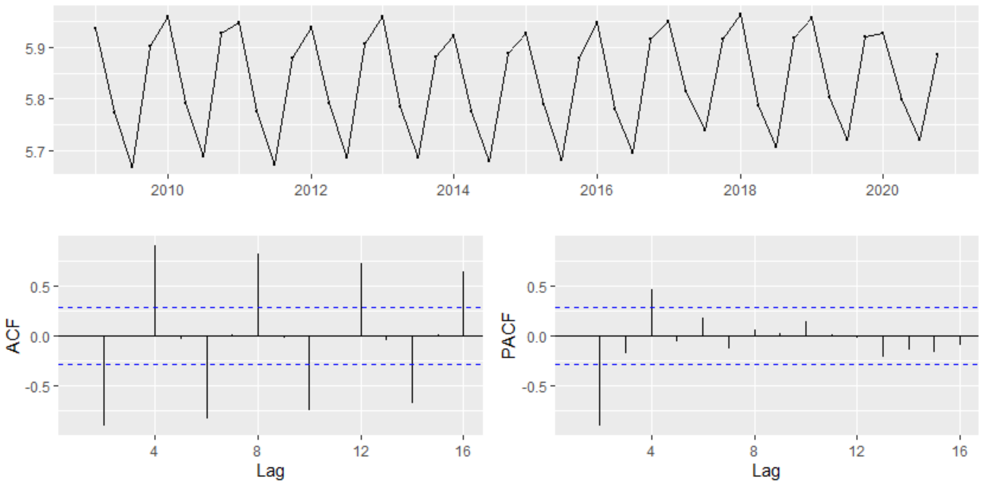

Figure 2: Swedish fuel consumption (time series, AFC, and PACF) with Box-Cox transformation

While we cannot see a significant difference the λ seems to be different enough to warrant a fit using Box-Cox in the final model.
Next is to verify that the time series is stationary, since to fit a SARIMA model we need to make sure the mean, variance, and autocovariance of the series are time invariant. To correct the seasonal trend, we apply a differencing with lag of 4 which indicates quarterly data.

```
timeseries.sdiff<-diff(timeseries,lag=4,differences=1)
ggtsdisplay(timeseries.sdiff)
```

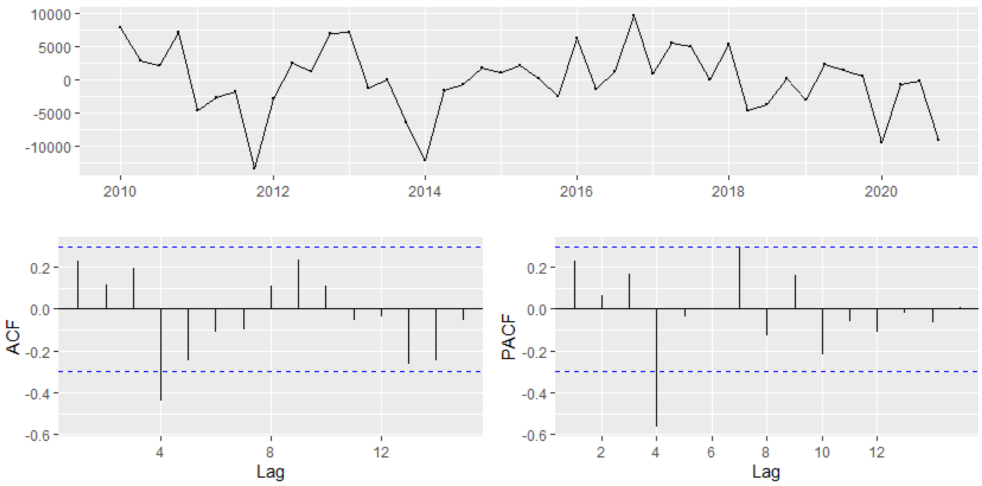

Figure 3: time series with a difference of lag 4

As we can infer by the ACF of Figure 3, the seasonal trend seems to be insignificant, and only one lag at 4 is significant. We could also have applied the Augmented Dickey-Fuller Test to statistically analyze if the differenced time series is stationary. Another step that could have been made is to difference once without any lags, to see if that would remove the regular trend.
In the next step, we will identify a time series model that is appropriate for the data set, and it is already assumed that a SARIMA model with seasonal differencing and Box-Cox transformation will be applied. Let fit 1 be the first SARIMA model, then from Figure 3 we can see that the ACF somewhat resembles a sine wave that has a significant lag spike at 4. This could suggest a AR(p) model in the seasonal. Let us try $\text{ARIMA}(1,0,0)(1,1,0)_4$, i.e. an AR model for the regular and seasonal ARIMA, with seasonal difference.

```
fit1<-Arima(timeseries,order=c(1,0,0)
            ,seasonal=c(1,1,0),lambda=lambda)
ggtsdisplay(fit1$residuals)
```

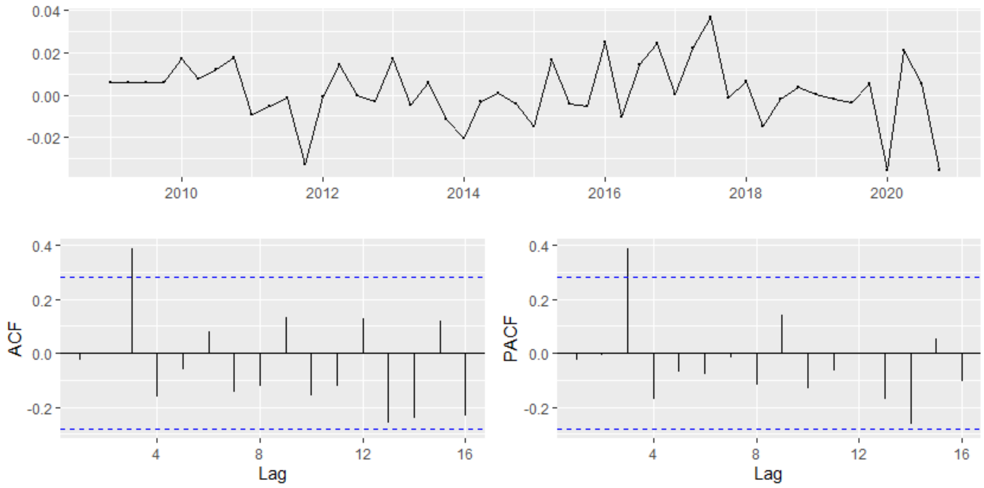

Figure 4: Fit 1, ARIMA$(1,0,0)(1,1,0)_4$

Figure 4 seems to suggest that we do not capture all of the necessary variables to have a good fit for the data set. For fit 2, we increase the AR(p) for the seasonal part.

```
fit2<-Arima(timeseries,order=c(0,1,0)
            ,seasonal=c(2,1,0),lambda=lambda)
ggtsdisplay(fit2$residuals)
```

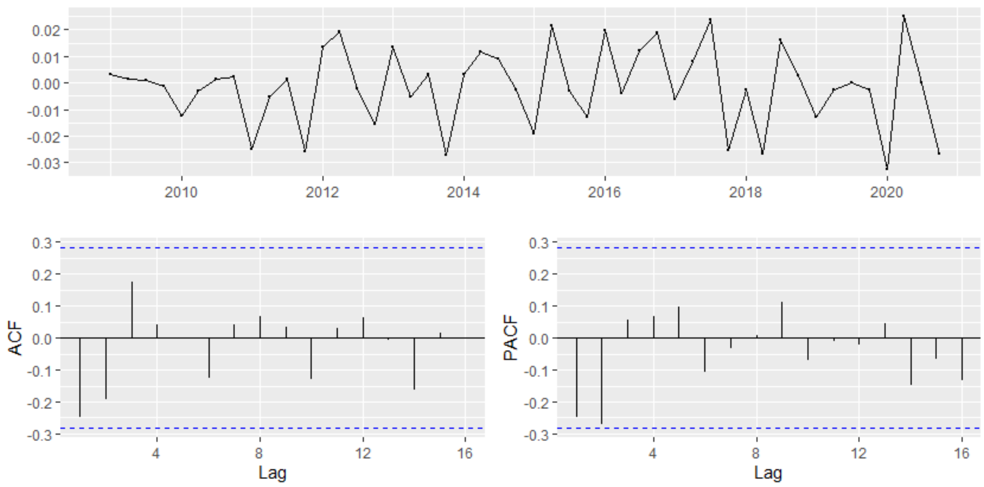

Figure 5: Fit 2, ARIMA$(1,0,0)(2,1,0)_4$

Fit 2 captures the data set better and produces a time series with no significant spikes in the ACF or PACF. The next step is to compare the Akaike information criterion (AICc) of fit 2 compared to similar models and see if we have the best fit.

```
Summary(fit2)
>  ARIMA(0,1,0)(2,1,0)[4], AICc=-225.71

fit3<-Arima(timeseries,order=c(0,1,0)
             ,seasonal=c(2,1,1),lambda=lambda)
Summary(fit3)
>  ARIMA(0,1,0)(2,1,1)[4], AICc=-223.82

fit4<-Arima(timeseries,order=c(0,1,1)
             ,seasonal=c(2,1,0),lambda=lambda)
Summary(fit4)
>  ARIMA(0,1,1)(2,1,0)[4], AICc=-228.27

fit5<-Arima(timeseries,order=c(0,1,1)
             ,seasonal=c(2,1,0),lambda=lambda)
Summary(fit5)
>  ARIMA(0,1,2)(2,1,0)[4], AICc=-225.86
```

Of the models tested, fit 5 the ARIMA$(0, 1, 1)(2, 1, 0)_4$ has the smallest AICc value, and that would then be the best model. The last step is to analyze the residuals to make sure the model provides a good description of the model.

```
qqnorm(fit5)
hist(fit5$residuals)
qqnorm(fit5$residuals)
tsdiag(fit5)
```

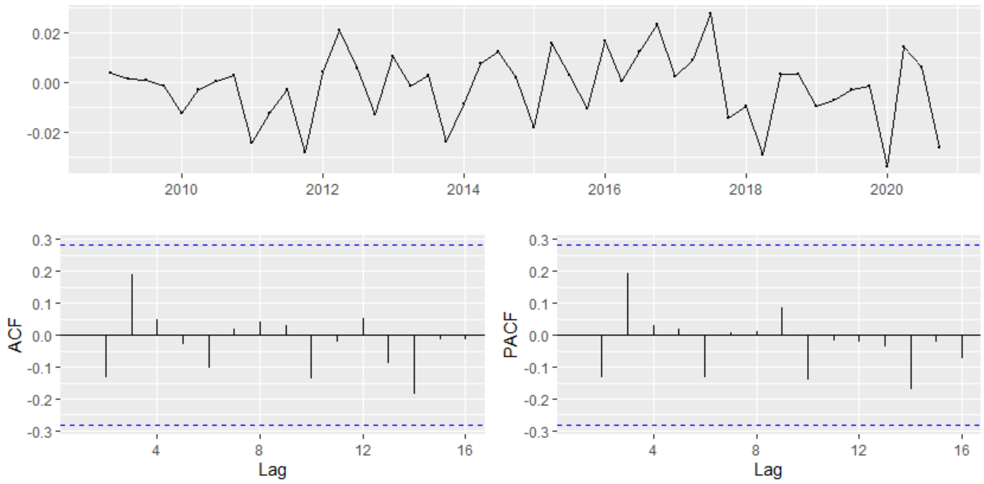

Figure 6: Fit 5, ARIMA$(0,1,1)(2,1,0)_4$

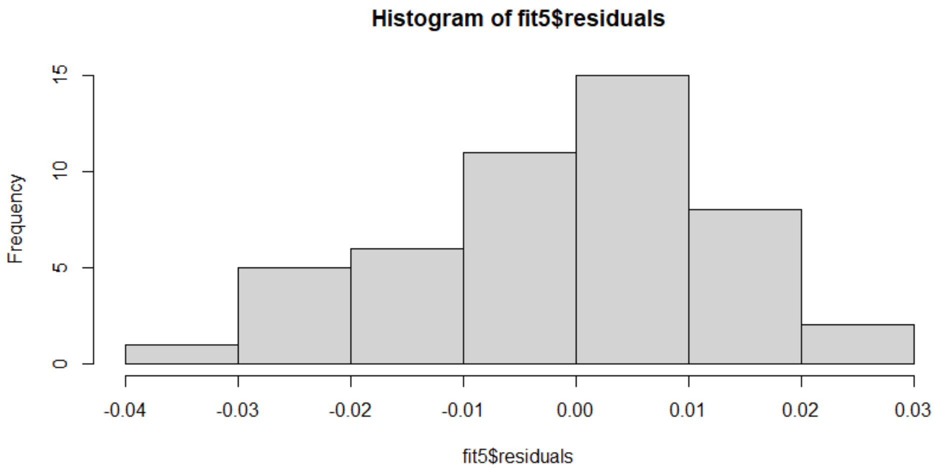

Figure 7: Histogram

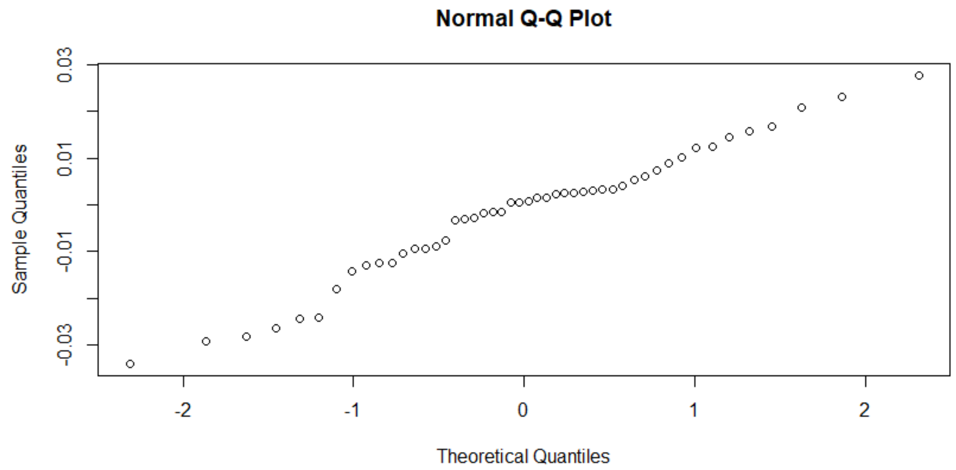

Figure 8: Normal Q-Q plot

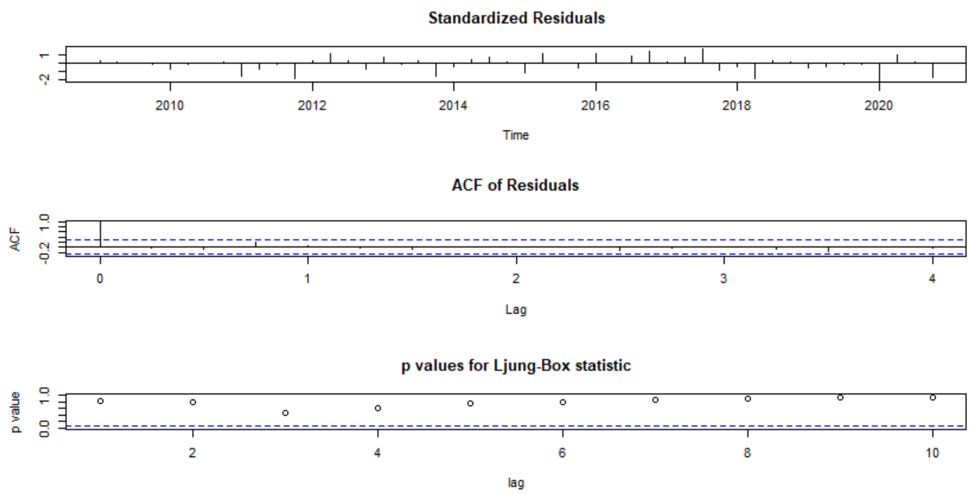

Figure 9: Time-series diagnostics

The ACF plot of the residuals from the $\text{ARIMA}(0,1,1)(2,1,0)_4$ model shows all correlations within the threshold limits indicating that the residuals are behaving like white noise. The histogram of the residuals in Figure 7 are somewhat normally distributed and we have to remember that the data set is quite small of only 48 data points. The same conclusion can be inferred from the Normal Q-Q plot in Figure 8. Additionally, we see from the plot of time-series diagnostics, Figure 9, that the p values are well above the rejection line.

Answer The final model is a seasonal ARIMA$(0,1,1)(2,1,0)_4$ with a Box-Cox transformation and coefficients listed below

```
ARIMA(0,1,1)(2,1,0)[4]
Box Cox transformation: lambda= -0.1287132

Coefficients:
          ma1      sar1     sar2
      -0.3905  -0.9705  -0.4686
s.e.   0.1491   0.1559   0.1617

sigma^2 = 0.0002315:  log likelihood = 118.66
AIC=-229.33   AICc=-228.27   BIC=-222.28

Training set error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE
ACF1
Training set -866.491 3588.487 2508.748 -0.9958983 4.350802 0.6744152 -0.06140681
```

References

[1] Robert Shumway and David Stoffer. *Time Series Analysis and Its Applications: With R Examples*. 01 2017.