Time Series Assignment 2

16343261

5/12/2022
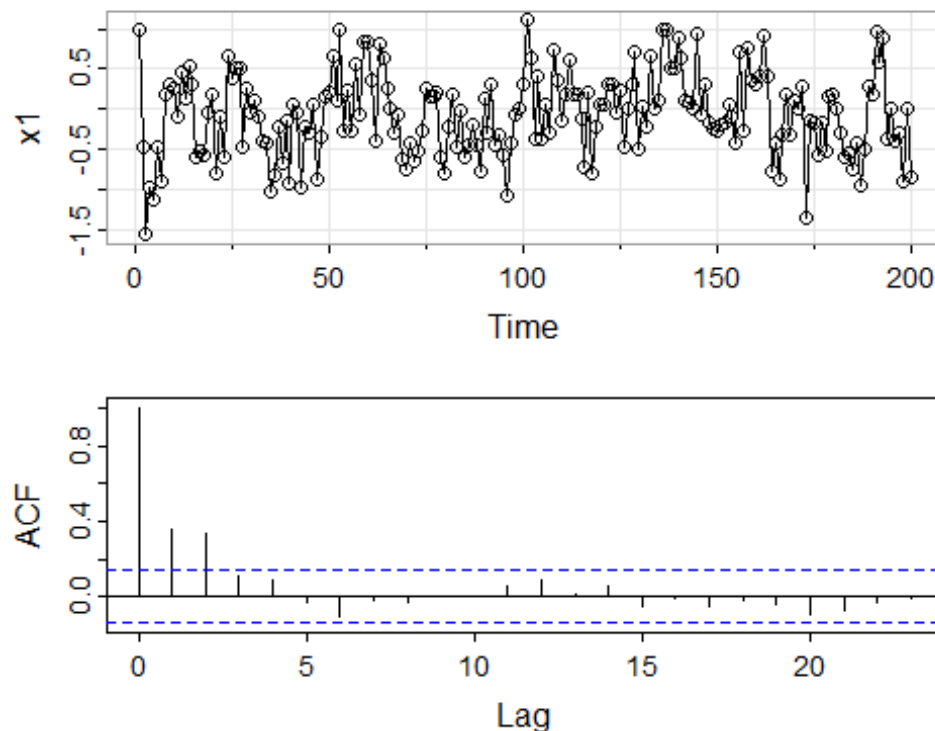
Before we begin our analysis, let it be noted that I have not included loading in our data in this pdf for ease of brevity.

Question 1

(A)

Below is the plot for our time series and our ACF.

```
library(astsa)
par(mfrow=c(2,1))
tsplot(x1,type="o")
acf(x1)
```


From the time series plot, we can observe neither a positive nor negative trend.Therefore, the mean seems to be constant over time. Further, the ACF has a minor decay overtime, this is evidence for stationarity. Therefore, this time series appears stationary.

(B)
```
pacf(x1)
pacf(x1)$acf[1:5]
```

Series x1
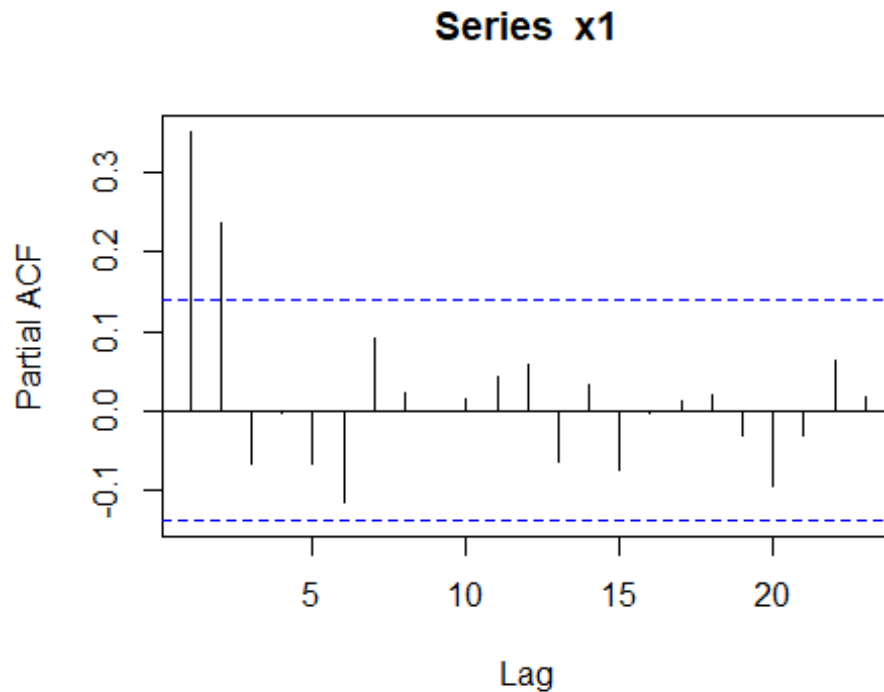

```
## [1]  0.352263166  0.237441762 -0.068017687 -0.004283268 -0.067581858
```

```
sd(x1)
```

```
## [1] 0.5154293
```

After viewing both our ACF and PACF, we can observe the following properties.

- Our PACF declines after 2 lags.

- Our ACF declines gradually.

Therefore I believe an AR(2) model to be appropriate.

```
ar2<-arima(x1, order = c(2, 0, 0),include.mean = FALSE)
```

(C)

Below, we have the codes used to estimate the parameters of an AR(2) process via Yule-Walker method and the output as well.

```
x1.yw <- ar.yw(x1, order = 2)
x1.yw
```

```
##
## Call:
## ar.yw.default(x = x1, order.max = 2)
##
## Coefficients:
##      1       2
## 0.2686  0.2374
##
## Order selected 2  sigma^2 estimated as  0.2218
```

Hence, we have that $\phi\hat{}1 = 0.2686$, $\phi\hat{}2 = 0.2374$, and $(\sigma)\hat{}2 = 0.2218$.

Therefore, our AR Equation is: $Xt = 0.2686Xt-1 + 0.2374Xt-1 + Wt$ with $Wt \sim N(0, 0.2218)$.

(D)

Below we compute the inferior and superior limits of the confidence intervals.

```
inf<- x1.yw$ar-qnorm(1-0.05/2)*sqrt(diag(x1.yw$asy.var.coef))
sup<- x1.yw$ar+qnorm(1-0.05/2)*sqrt(diag(x1.yw$asy.var.coef))
inf

## [1] 0.1329730 0.1017936

sup

## [1] 0.4042693 0.3730899
```

Therefore, approximate 95% confidence intervals for $\phi 1$ and $\phi 2$ are respectively given by (0.1329730 , 0.4042693) and (0.1017936, 0.3730899). As zero is not included in our confidence intervals, we can say there is evidence that our two parameters are statistically significant to be included in our model.

Question 2

(A)
```
mean(x2)

## [1] 0.0068

par(mfrow=c(3,1))
tsplot(x2,type="o")
acf(x2)
pacf(x2)
```
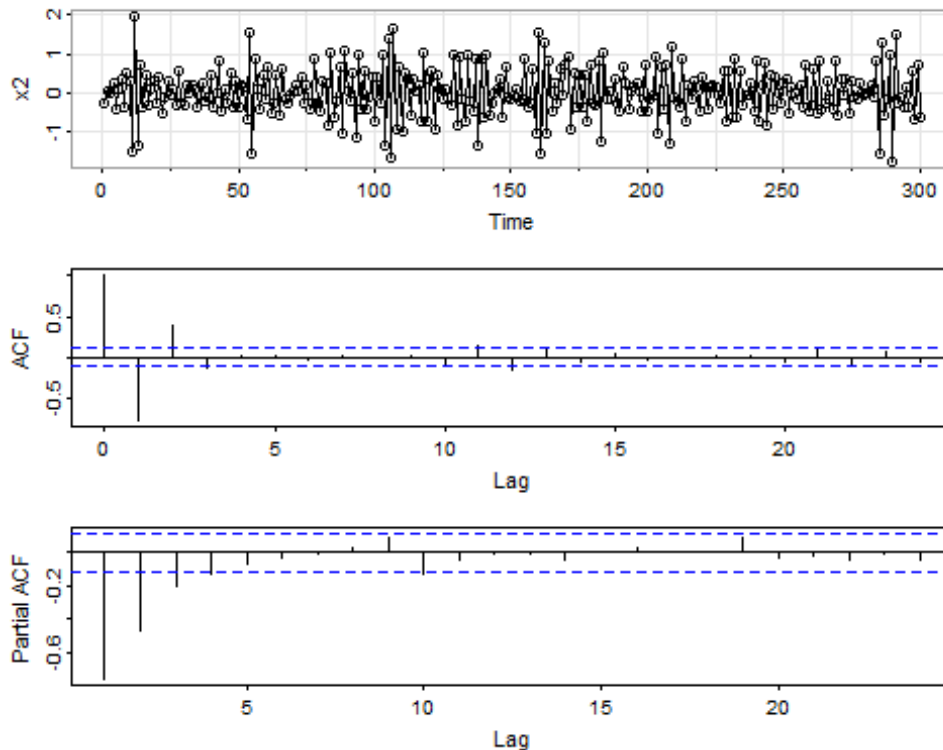
Here we find both the ACF and the PACF are exhibiting a gradual decrease. Here we observe our PACF gradually declining from negative values to zero. For our ACF, we find that it moves between negative and positive values.

(B)

Both ACF and PACF show slow decay (gradual decrease). Hence, the ARMA (1,1) model would be appropriate for the series. However, observing the ACF plot: it sharply drops after two significant lags which indicates that an MA (2) would be a good candidate model for the process. Therefore, we should experiment with both ARMA (1,1) and MA (2) for the process and later select the optimal model based on a performance metric like AIC (Akaike Information Criteria).

```
set.seed(12345)
arma11<-arima(x2, order = c(1, 0, 1), include.mean = TRUE)
ma2<-arima(x2, order = c(0, 0, 2), include.mean = TRUE)
```

We will use the AIC and BIC to determine which model to pick. The selected model is that one that minimises the considered information criterion quantity.

```
AIC(arma11, ma2)

##         df      AIC
## arma11   4 229.5612
## ma2      4 242.3655

BIC(arma11, ma2)
```

```
##        df      BIC
## arma11  4 244.3763
## ma2     4 257.1806
```

Therefore, ARMA(1,1) should be selected as it has the minimum AIC and BIC.

(C)

Maximum likelihood estimation
```
arma11
```

```
##
## Call:
## arima(x = x2, order = c(1, 0, 1), include.mean = TRUE)
##
## Coefficients:
##            ar1      ma1  intercept
##        -0.5753  -0.6431     0.0086
## s.e.    0.0523   0.0468     0.0046
##
## sigma^2 estimated as 0.1219:  log likelihood = -110.78,  aic = 229.56
```

The fitted model is given by $X_t = -0.5753 X_{t-1} + W_t + -0.6431 W_{t-1}$, with $W_t \sim N(0, 0.1219)$.

We want to test the hypotheses $H0 : \phi_j = 0$ against $H1 : \phi_j \neq 0$, for $j = 1, 2$ and also $H0 : \theta_1 = 0$ against $H1 : \theta_1 \neq 0$. We can do this using a confidence interval.

```
fit <- arima(x2, order = c(1, 0, 1), include.mean = TRUE)
# standard errors
s.e<-sqrt(diag(fit$var.coef))
# inferior bound
fit$coef-qnorm(1-0.05/2)*s.e
```

```
##            ar1            ma1      intercept
## -0.6777600980 -0.7347100764 -0.0004536853
```

```
# superior bound
fit$coef+qnorm(1-0.05/2)*s.e
```

```
##          ar1          ma1    intercept
## -0.47282548  -0.55145002   0.01758361
```

As zero is not included in our confidence intervals, we can say there is evidence that our two parameters are statistically significant to be included in our model.

(D)

We will use the AIC and BIC to determine which model to pick. The selected model is that one that minimises the considered information criterion quantity.

```
fit1 <- arima(x2, order = c(1, 0, 0), include.mean = TRUE)
fit2 <- arima(x2, order = c(0, 0, 2), include.mean = TRUE)
```

```
fit3 <- arima(x2, order = c(1, 0, 1), include.mean = TRUE)
fit4 <- arima(x2, order = c(2, 0, 1), include.mean = TRUE)
fit5 <- arima(x2, order = c(1, 0, 2), include.mean = TRUE)
fit6 <- arima(x2, order = c(2, 0, 2), include.mean = TRUE)

AIC(fit1, fit2, fit3, fit4, fit5, fit6)

##      df      AIC
## fit1  3 317.4942
## fit2  4 242.3655
## fit3  4 229.5612
## fit4  5 228.0349
## fit5  5 228.5918
## fit6  6 229.8298

BIC(fit1, fit2, fit3, fit4, fit5, fit6)

##      df      BIC
## fit1  3 328.6055
## fit2  4 257.1806
## fit3  4 244.3763
## fit4  5 246.5538
## fit5  5 247.1107
## fit6  6 252.0524
```

Here we see two different models are minimising our criteria with AIC selecting ARMA(2,1) and BIC selecting ARMA(1,1).

BIC penalizes model complexity more heavily. The only way they should disagree is when AIC chooses a larger model than BIC.In this case, it should be logical to select ARMA(1,1) to minimise the complexity of our chosen model.
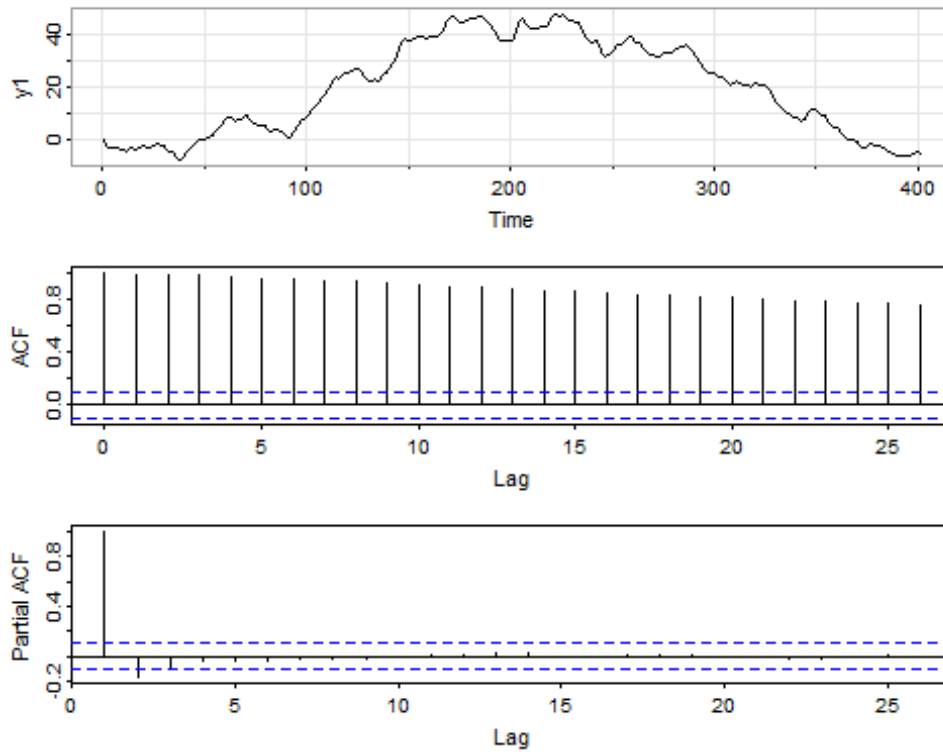
Question 3

(A)
```
par(mfrow=c(3,1))
tsplot(x3,type="o")
acf(x3)
pacf(x3)
```
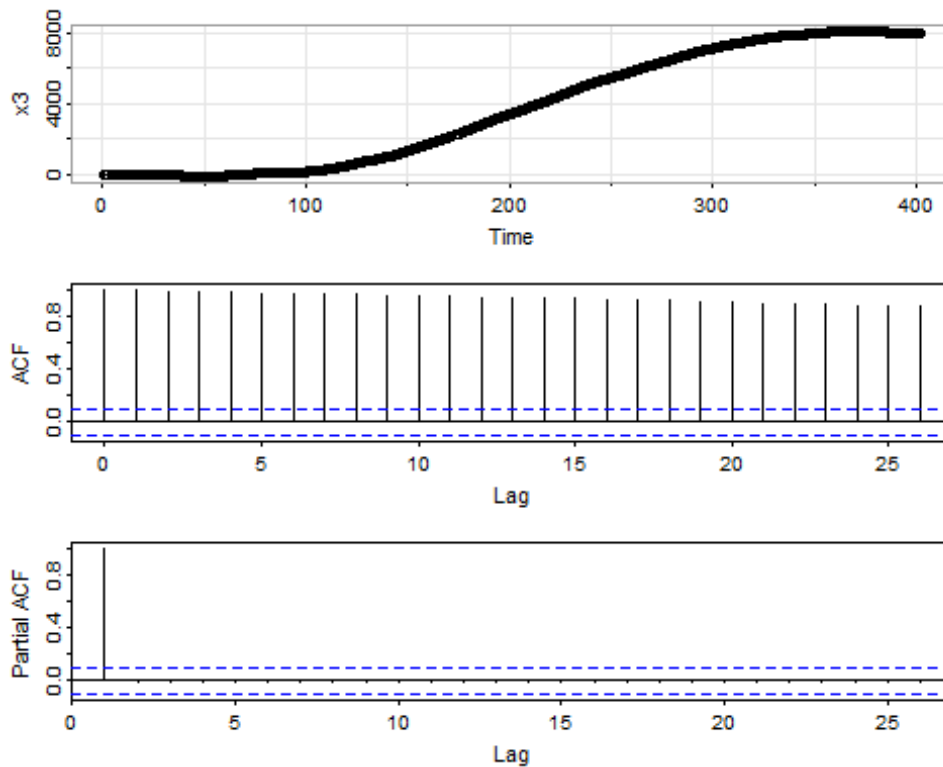
No it does not appear to be stationary. We will now check each difference until it is stationary starting with differences = 1.
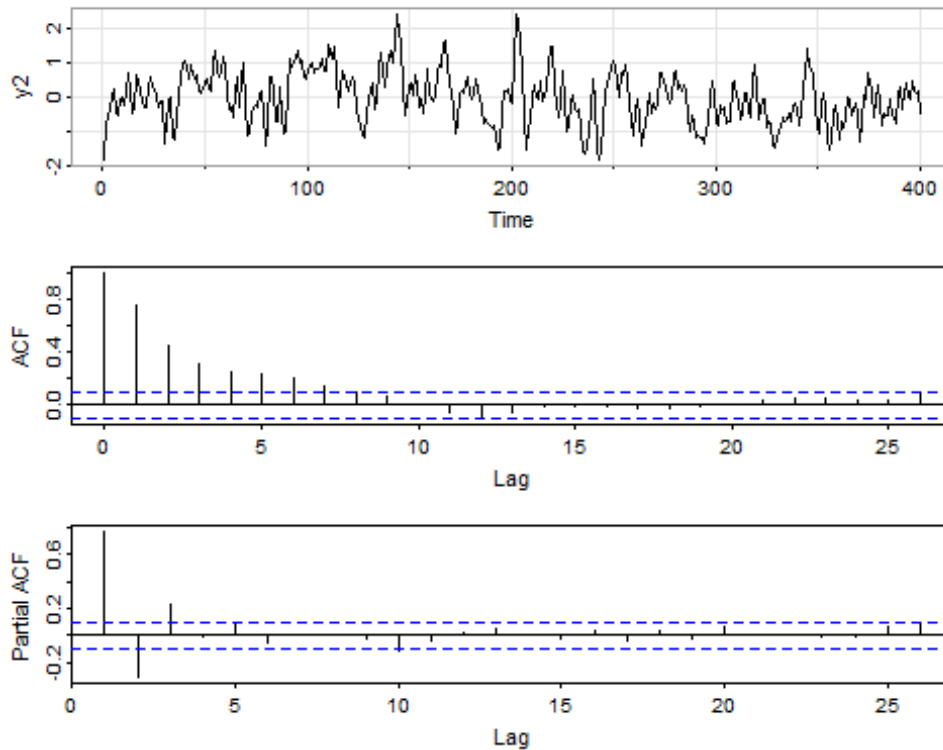
```
y1<-diff(x3,differences = 1)
par(mfrow=c(3,1))
tsplot(y1)
acf(y1)
pacf(y1)
```

We see at differences = 1, we have still not achieved stationarity.

```
y2<-diff(x3,differences = 2)
par(mfrow=c(3,1))
tsplot(y2)
acf(y2)
pacf(y2)
```

The number of differences necessary to achieve stationarity is 2.

(B)

Judging from our PACF and ACF, there are 3 potential models. Using the same reasoning from Question 2d, we can test which should be our selected model using our AIC and BIC criteria.

```
set.seed(12345)
fit1<-arima(y2, order = c(0, 0, 2), include.mean = FALSE)
fit2<-arima(y2, order = c(1, 0, 0), include.mean = FALSE)
fit3<-arima(y2, order = c(1, 0, 1), include.mean = FALSE)

AIC(fit1,fit2,fit3)

##        df      AIC
## fit1   3 526.1152
## fit2   2 565.4711
## fit3   3 497.8557

BIC(fit1,fit2,fit3)

##        df      BIC
## fit1   3 538.0896
## fit2   2 573.4540
## fit3   3 509.8301
```

Here we find it clear that ARMA(1,1) should be our selected model.

```
fit3

##
## Call:
## arima(x = y2, order = c(1, 0, 1), include.mean = FALSE)
##
## Coefficients:
##           ar1     ma1
##        0.5839  0.5550
## s.e.   0.0479  0.0494
##
## sigma^2 estimated as 0.1996:  log likelihood = -245.93,  aic = 497.86
```

The fitted model is given by Xt = 0.5839(Xt-1) + Wt + 0.555(Wt-1).

(C)

A 95% confidence interval will be the equivalent of testing the statistical significance of our lags using a significance level of 5%.

We want to test the hypotheses H0 : ϕj = 0 against H1 : ϕj/= 0, for j = 1, 2 and also H0 : $\theta 1$ = 0 against H1 : $\theta 1$/= 0. We can do this using a confidence interval.

```
s.e_fit3<-sqrt(diag(fit3$var.coef))
# inferior bound
fit3$coef-qnorm(1-0.05/2)*s.e_fit3

##        ar1       ma1
## 0.4899460 0.4581137

# superior bound
fit3$coef+qnorm(1-0.05/2)*s.e_fit3

##        ar1       ma1
## 0.6777838 0.6519247
```

Therefore, approximate 95% confidence intervals for $\phi 1$ and $\phi 2$ are respectively given by (0.4899460, 0.6777838) and (0.4581137, 0.6519247). As zero is not included in our confidence intervals, we can say there is evidence that our two parameters are statistically significant to be included in our model.

(D)

Standardised Residuals should be normally distributed and the Q-Q Plot will show this. If residuals follow close to a straight line on this plot, it is a good indication they are normally distributed. By using the SARIMA function, we will obtain diagnostic tool plots for checking if the model is well-fitted.

```
sarima(y2,p=1,d=0,q=1,no.constant=TRUE)

## initial  value -0.275941
## iter   2 value -0.704227
```

```
## iter    3 value -0.786847
## iter    4 value -0.804291
## iter    5 value -0.812525
## iter    6 value -0.813648
## iter    7 value -0.813704
## iter    8 value -0.813729
## iter    9 value -0.813729
## iter    9 value -0.813729
## iter    9 value -0.813729
## final   value -0.813729
## converged
## initial  value -0.804039
## iter    2 value -0.804089
## iter    3 value -0.804114
## iter    4 value -0.804119
## iter    4 value -0.804119
## iter    4 value -0.804119
## final   value -0.804119
## converged
```
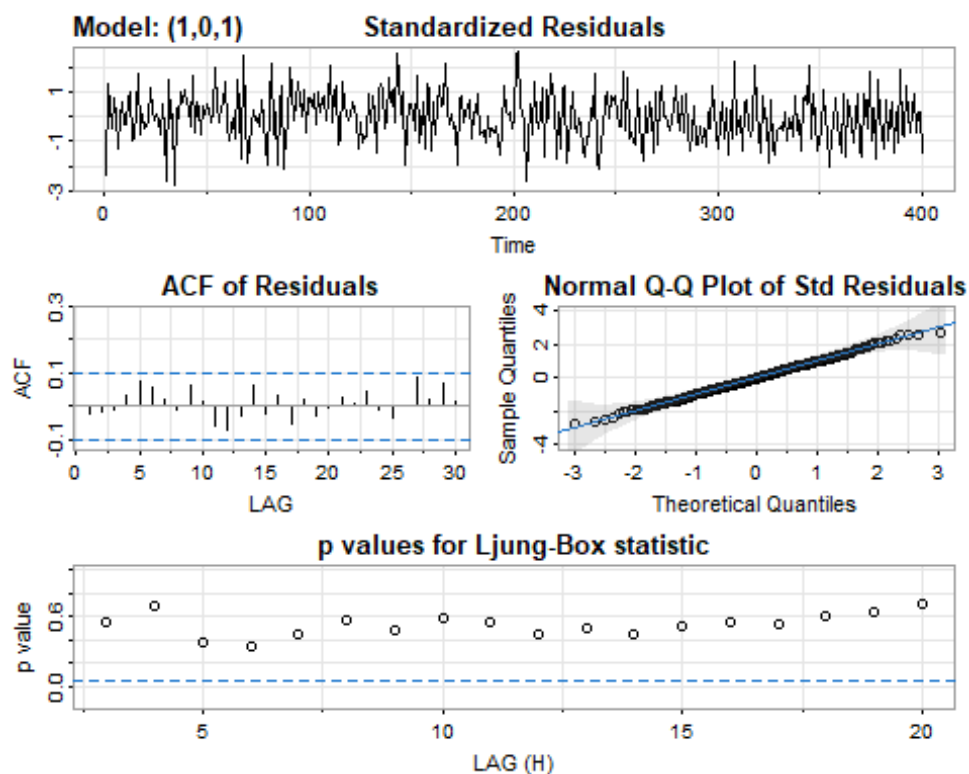


```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
##      xreg = xmean, include.mean = FALSE, transform.pars = trans, fixed =
fixed,
```

```
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1     ma1
##       0.5839  0.5550
## s.e.  0.0479  0.0494
##
## sigma^2 estimated as 0.1996:  log likelihood = -245.93,  aic = 497.86
##
## $degrees_of_freedom
## [1] 398
##
## $ttable
##      Estimate     SE t.value p.value
## ar1    0.5839 0.0479 12.1845       0
## ma1    0.5550 0.0494 11.2255       0
##
## $AIC
## [1] 1.244639
##
## $AICc
## [1] 1.244715
##
## $BIC
## [1] 1.274575
```

From the plot of the standardised residuals and its ACF, we see that the residuals are behaving like a white noise, that is, the mean is zero, the variance seems to be constant over the time, and all the sample autocorrelations are inside the bands. Regarding the normal assumption for the white noise, we can observe from the qq-norm plot a good agreement between the sample quantiles from residuals and the theoretical quantiles from a standard normal distribution. So, the normal assumption seems not to be violated.

The Ljung-Box-Pierce plot shows that the source of autocorrelation from the the data has been captured by the ARMA(1,1) model since most of our p-values are greater than 0.05. Therefore, we have statistical evidence that the model is well-fitted.