

# STAT 443: Assignment 3

Saksham Sudershan (Student #31339427)

21 March, 2022

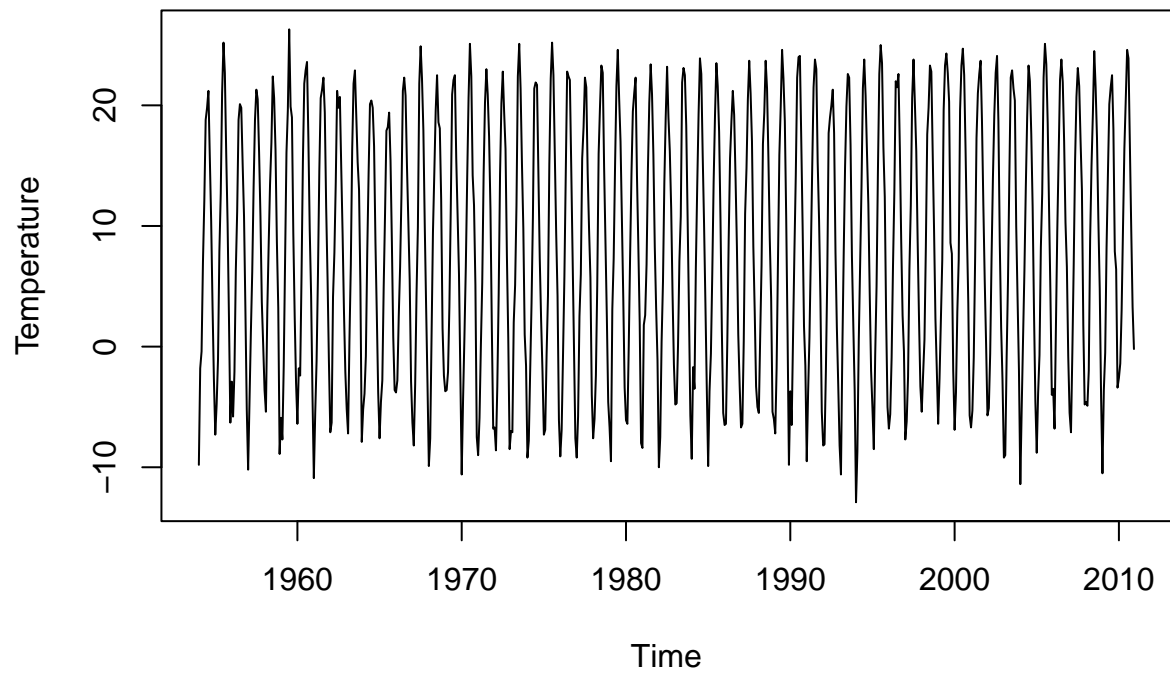
## Problem 1

### Part (a)

```
# Reading data, and making datasets
data1 <- read.csv("rimouski.csv")
rim.data <- ts(data1$Mean.Max.Temp, start = c(1954, 1), frequency = 12, end = c(2017, 8))
# i. Training data from Jan 1954-Dec 2010
rim.train <- window(rim.data, start = c(1954, 1), end = c(2010,12))
# ii. Testing data from Jan 2011-Dec 2016
rim.test <- window(rim.data, start = c(2011,1), end = c(2016, 12))

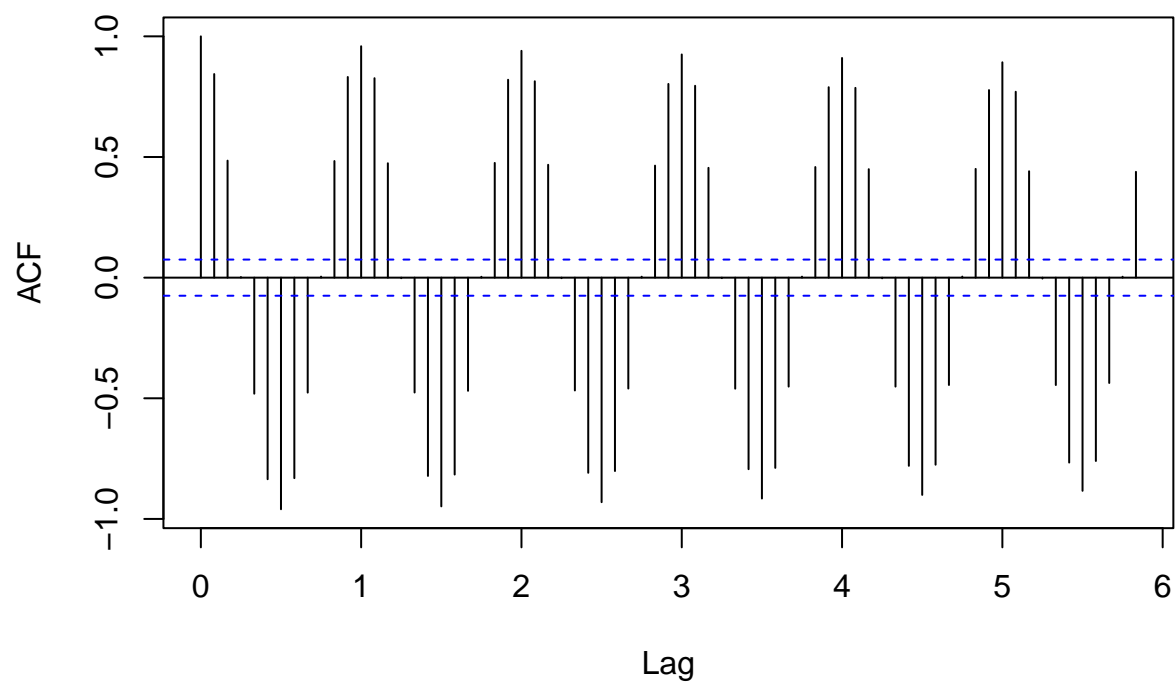
# Plotting the training data, acf and pacf
plot(rim.train, ylab = "Temperature", main = "Mean Max. Temperature 1954-2010")
```

### Mean Max. Temperature 1954–2010



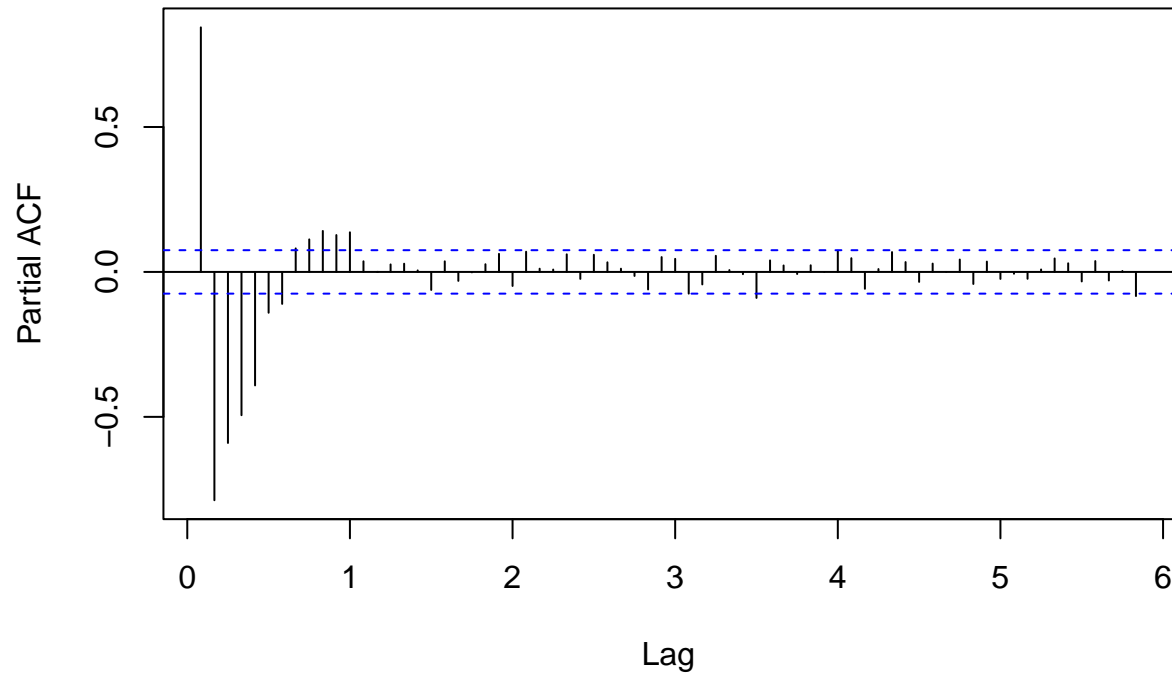
```
acf(rim.train, lag.max = 70)
```

### Series rim.train



```
pacf(rim.train, lag.max = 70)
```

## Series rim.train



The ACF shows high correlation between at year lags of 1,2... or month lags  $h = 12, 24 \dots$ . The period seen from the ACF plot seems to be 1 year, or 12 months. The ACF tails off like a damped sine wave very slowly, and the autocorrelations remain high for a long time. The PACF is almost random till cutting off at time  $t = 1 \text{ year}$ .

### Part (b)

```
modell1 <- arima(rim.train, order = c(0,0,0), seasonal = list(order = c(0,1,0), period = 12))
modell1
```

```
##
## Call:
## arima(x = rim.train, order = c(0, 0, 0), seasonal = list(order = c(0, 1, 0),
##     period = 12))
##
##
## sigma^2 estimated as 6.122:  log likelihood = -1562.31,  aic = 3126.62
```

i.

The fitted  $(0,0,0) \times (0,1,0)_{12}$  model is given by:

$$\nabla_{12}X_t = Z_t$$

$$\Rightarrow X_t - X_{t-12} = Z_t$$

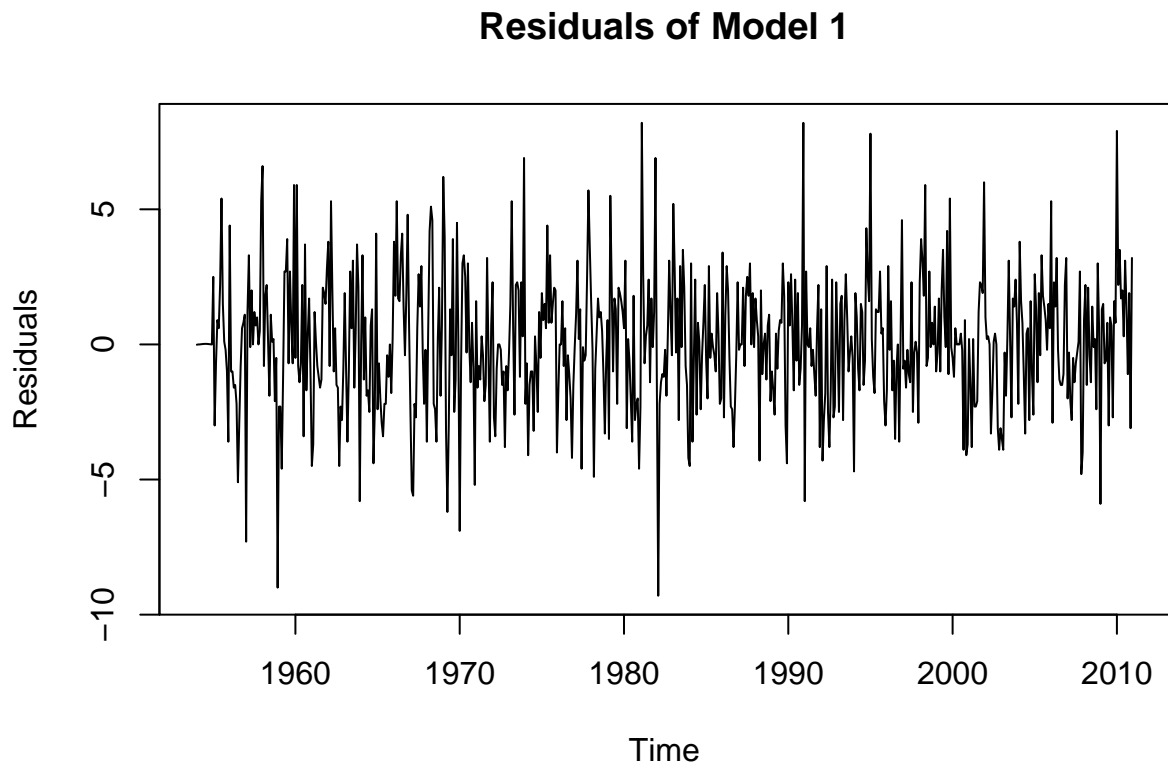
So it can be said that this is an ARMA(12,0) process with  $\alpha_{12} = 1$ .

ii.

The variance of the  $Z_t$  terms is estimated as  $\hat{\sigma}_z^2 = 6.122$  and  $\alpha_{12} = 1$ . All  $\alpha_i = 0$  for  $i \in [1, 11]$ .

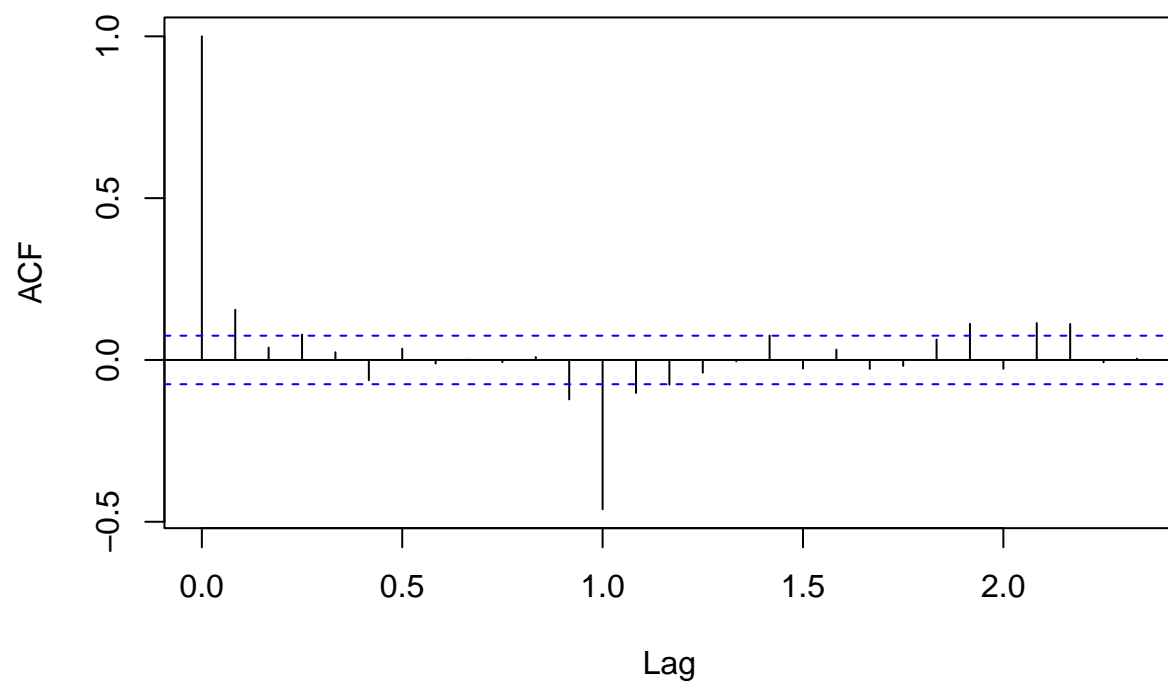
iii.

```
plot(model1$residuals, ylab = "Residuals", main = "Residuals of Model 1")
```

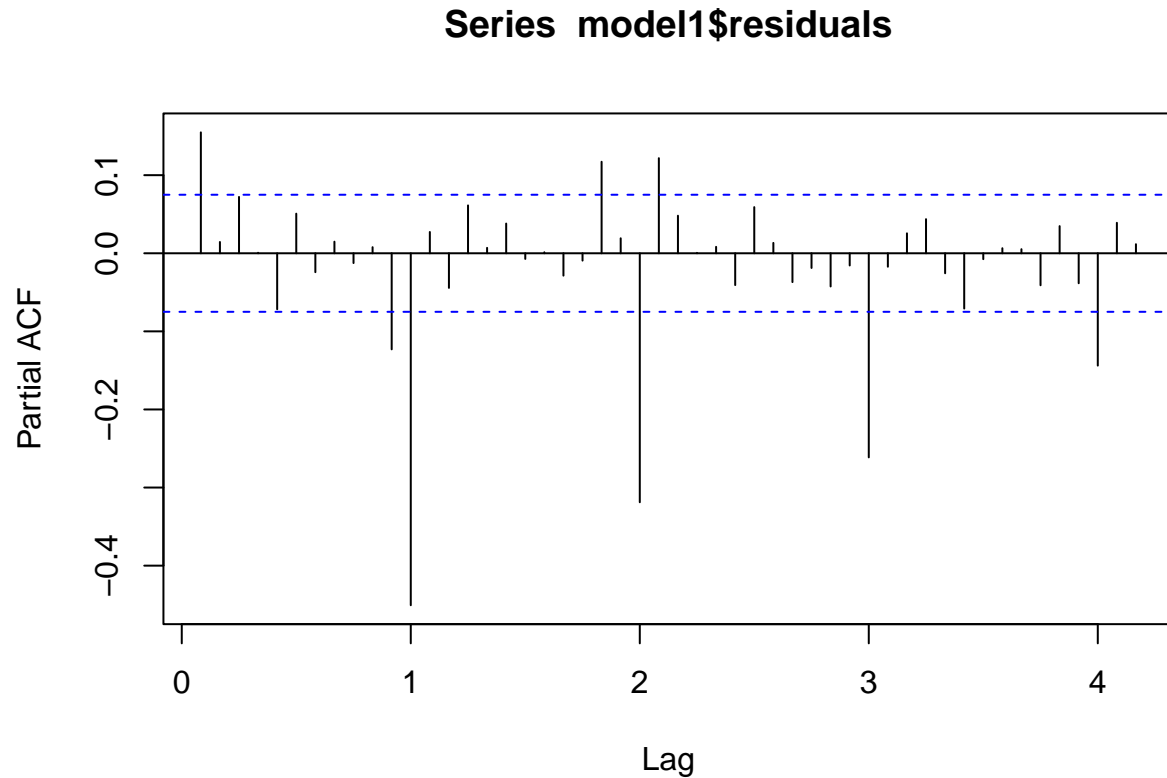


```
acf(model1$residuals)
```

### Series model1\$residuals



```
pacf(model1$residuals, lag.max = 50)
```



The ACF plot shows a negative spike at lag  $h = 1$ , or at the 1 year mark. This means that there is still correlation between the residuals at lags of 12 months that are not captured by our SARIMA model. This is corroborated with the PACF plot which shows spikes at lag  $h = 1, 2, 3 \dots$ , meaning there is negative partial ACF at each year mark between the residuals, not captured by our model.

### Part (c)

i.

Even after seasonal differencing for a period of  $h = 12$ , we can see a significant negative autocorrelation in the ACF plot at year 1, and negative partial autocorrelations at each year mark seen in the PACF plot. This points to our model being unable to capture something seasonal. This significant correlation also cuts off in the ACF plot, and to fix this, We can add a seasonal MA ( $Q=1$ ) component.

ii.

The fitted  $(0, 0, 0) \times (0, 1, 1)_{12}$  model is given by:

$$\begin{aligned}\nabla_{12}X_t &= (1 + \Theta B^{12})Z_t \\ \Rightarrow X_t - X_{t-12} &= Z_t + \Theta Z_{t-12}\end{aligned}$$

So this would be an ARMA(12,12) process.

iii.

```
model12 <- arima(rim.train, order = c(0,0,0), seasonal = list(order = c(0,1,1), period = 12))
model12
```

```
##
## Call:
## arima(x = rim.train, order = c(0, 0, 0), seasonal = list(order = c(0, 1, 1),
##     period = 12))
##
## Coefficients:
##          sma1
##        -0.9206
## s.e.    0.0191
##
## sigma^2 estimated as 3.4:  log likelihood = -1375.98,  aic = 2755.96
```

The parameters are estimated as  $\hat{\Theta} = -0.9206$ ,  $\hat{\sigma}_z^2 = 3.4$  and  $\alpha_{12} = 1$ . All  $\alpha_i = 0$  for  $i \in [1, 11]$ , and all  $\beta_j = 0$  for  $j \in [1, 11]$ .

iv.

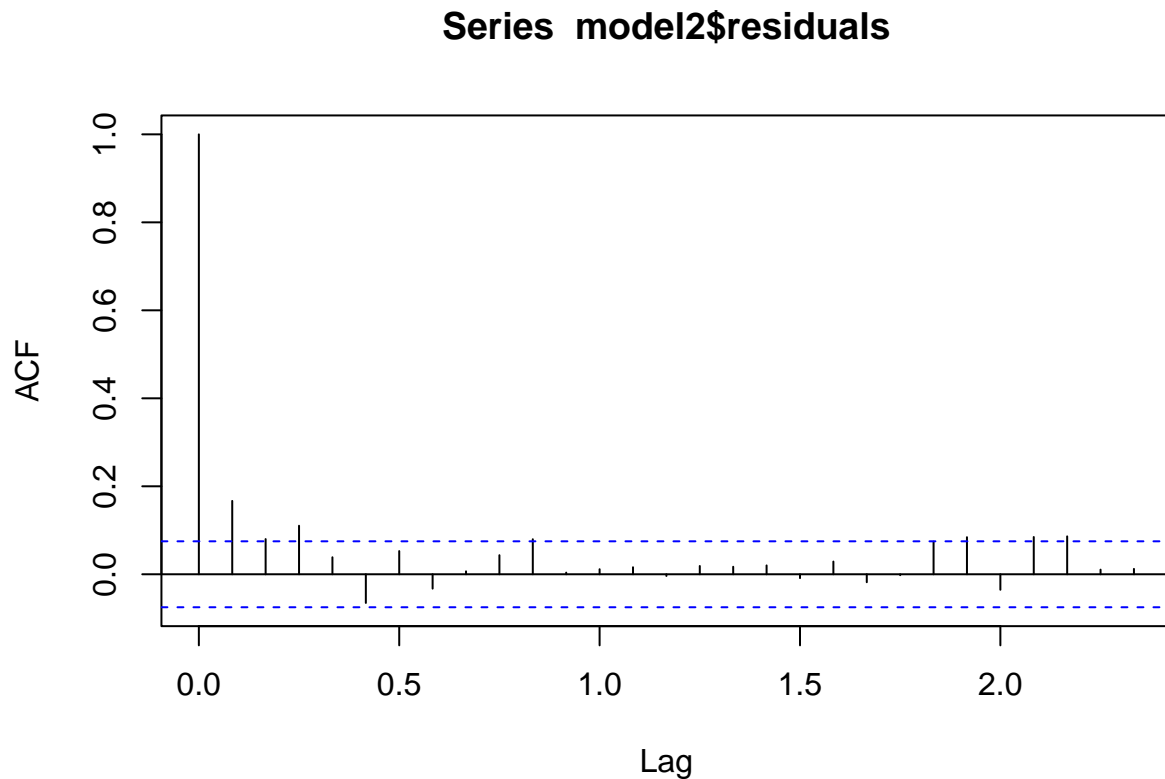
The AIC of  $(0, 0, 0) \times (0, 1, 0)_{12}$  model is 3126.62. The AIC of  $(0, 0, 0) \times (0, 1, 1)_{12}$  model is 2755.96.

Based on the AIC, we would choose the one with  $Q = 1$ , that is, the  $(0, 0, 0) \times (0, 1, 1)_{12}$  model as the AIC is lower.

## Part (d)

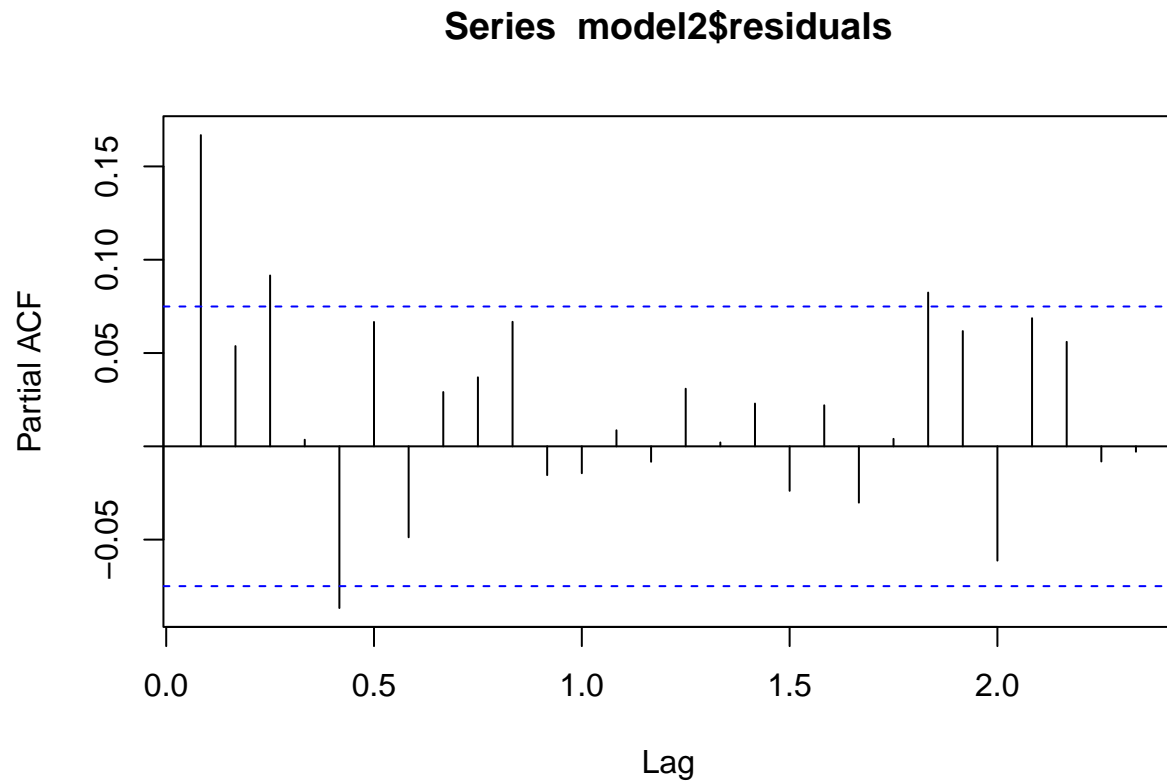
i.

```
acf(model2$residuals)
```





```
pacf(model2$residuals)
```



The ACF and PACF for lag  $h = 1$  are significant, as seen in both the plots. The PACF also seems to cut off at lag  $h = 1$  so in order to fix this, we can add an AR ( $p=1$ ) component.

ii.

The fitted  $(1, 0, 0) \times (0, 1, 1)_{12}$  is given by:

$$(1 - \phi B)\nabla_{12}X_t = (1 + \Theta B^{12})Z_t$$

$$(1 - \phi B)(X_t - X_{t-12}) = (1 + \Theta B^{12})Z_t$$

$$X_t - X_{t-12} - \phi X_{t-1} + \phi X_{t-13} = Z_t + \Theta Z_{t-12}$$

So this would be an ARMA(13,12) process.

iii.

```
model3 <- arima(rim.train, order = c(1,0,0), seasonal = list(order = c(0,1,1), period = 12))
model3
```

```
##
```

```
## Call:
```

```
## arima(x = rim.train, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1),
##   period = 12))
```

```
##
```

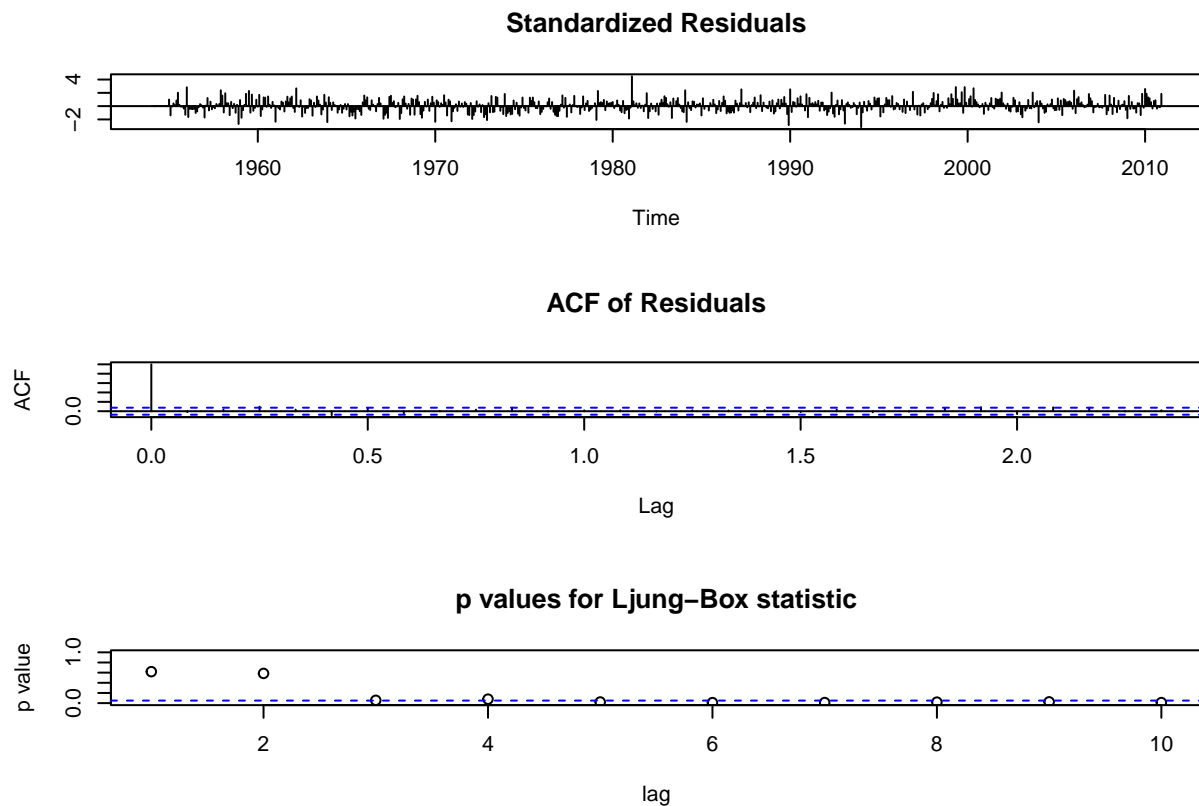
```
## Coefficients:
```

```
##          ar1      sma1
##      0.1808 -0.9354
## s.e.  0.0387  0.0190
##
## sigma^2 estimated as 3.281:  log likelihood = -1365.21,  aic = 2736.43
```

Here,  $\alpha_1 = -\alpha_{13} = \hat{\phi} = 0.1808$ ,  $\hat{\theta} = -0.9354$ , and  $\hat{\sigma}_z^2 = 3.281$ . All  $\alpha_i = 0$  for  $i \in [2, 11]$ , and all  $\beta_j = 0$  for  $j \in [1, 11]$ .

iv.

```
tsdiag(model3)
```



The ACF of Residual plot shows that none of the residuals are significantly correlated, which means that our model captures the correlation between the data well. However, the Ljung-Box statistic is high only for lags  $h = 1, 2$ , and falls below the limits, meaning we can reject the null hypothesis, and so we can say that the residuals are not independently distributed.

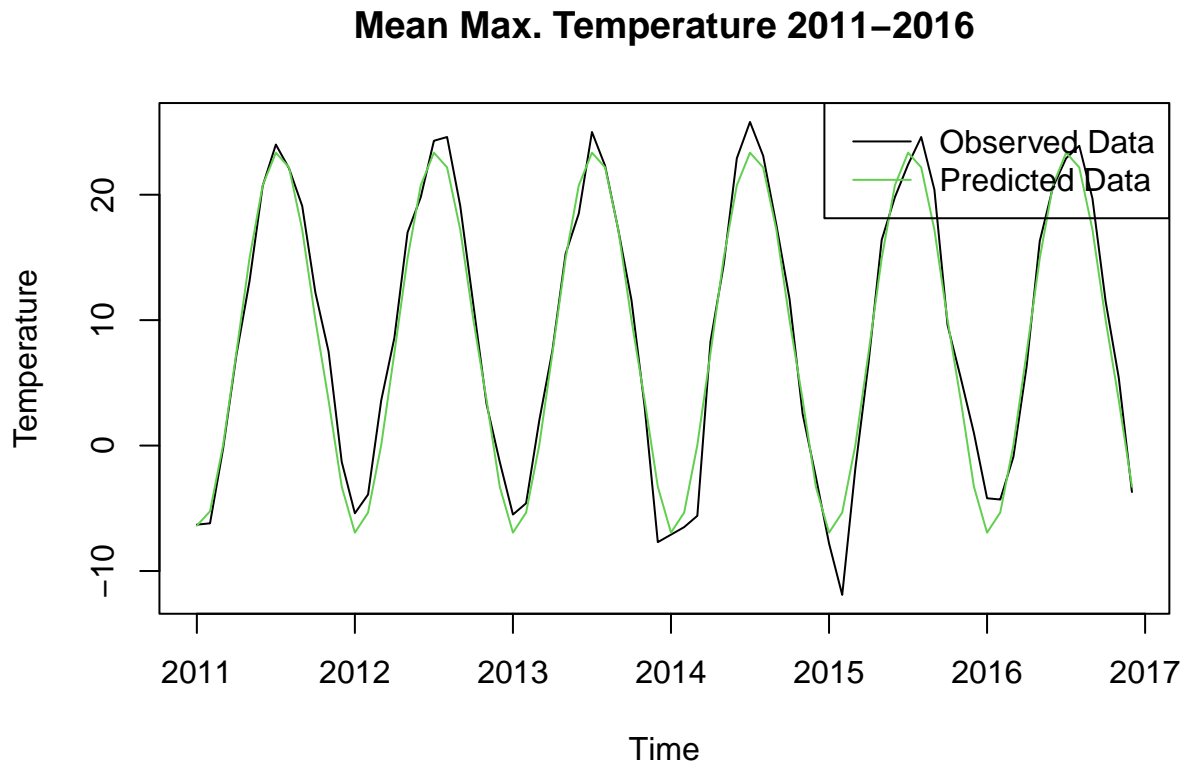
v.

The AIC of  $(0, 0, 0) \times (0, 1, 1)_{12}$  model is 2755.96. The AIC of  $(1, 0, 0) \times (0, 1, 1)_{12}$  model is 2736.43.

Based on the AIC, we would choose the one with  $p = 1$ , that is, the  $(1, 0, 0) \times (0, 1, 1)_{12}$  model, as it has the lower AIC.

### Part (e)

```
plot(rim.test, xlab = "Time", ylab = "Temperature", main = "Mean Max. Temperature 2011-2016")
lines(predict(model3, n.ahead = 12*6)$pred, col = 3)
legend("topright", legend = c("Observed Data", "Predicted Data"), col = c(1,3), lty = c(1,1))
```



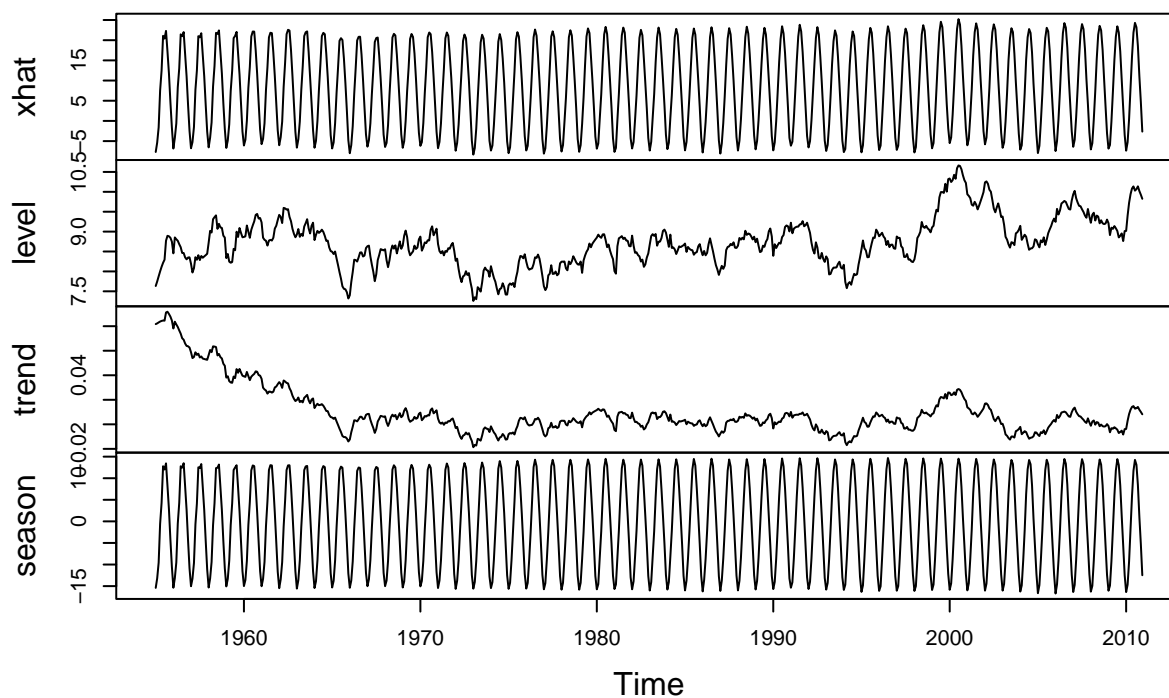
The prediction fit quite well to the actual data, only differing nears the peaks or the troughs. So our model can predict data well for all months other than peak summer or peak winter.

### Part (f)

i.

```
model4 <- HoltWinters(rim.train)
plot(model4$fitted, main = "Fitted Values - Holt Winters Model")
```

## Fitted Values – Holt Winters Model



```
model4$beta
```

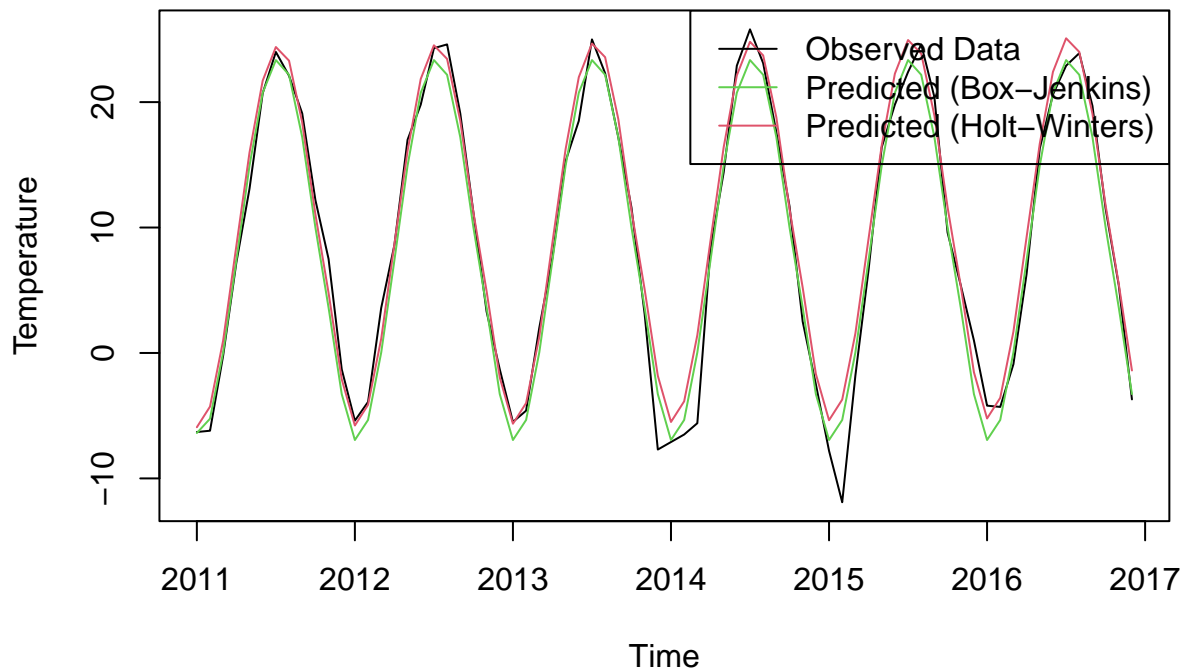
```
##      beta
## 0.01715428
```

The fitted values for the Holt Winters model show a decreasing at the start, with a random trend after approximately 1965. The  $\beta$  in the Holt Winters model is  $\beta = 0.01715 \approx 0$ , so we cannot say for sure that there is a trend.

ii.

```
plot(rim.test, xlab = "Time", ylab = "Temperature", main = "Mean Max. Temperature 2011-2016")
lines(predict(model3, n.ahead = 12*6)$pred, col = 3)
lines(predict(model4, n.ahead = 12*6), col = 2)
legend("topright", legend = c("Observed Data", "Predicted (Box-Jenkins)", "Predicted (Holt-Winters)"),
```

## Mean Max. Temperature 2011–2016



Both the models work quite well, with the Box-Jenkins prediction slightly edging out the Holt-Winters forecast in peaks and we can see the reverse in troughs. However, it is quite difficult to say which model is better given just this graph.

iii.

```
# MSPE of Box-Jenkins model
mean((predict(model3, n.ahead = 12*6)$pred-rim.test)^2)
```

```
## [1] 3.832878
```

```
# MSPE of Holt-Winters model
mean((predict(model4, n.ahead = 12*6) - rim.test)^2)
```

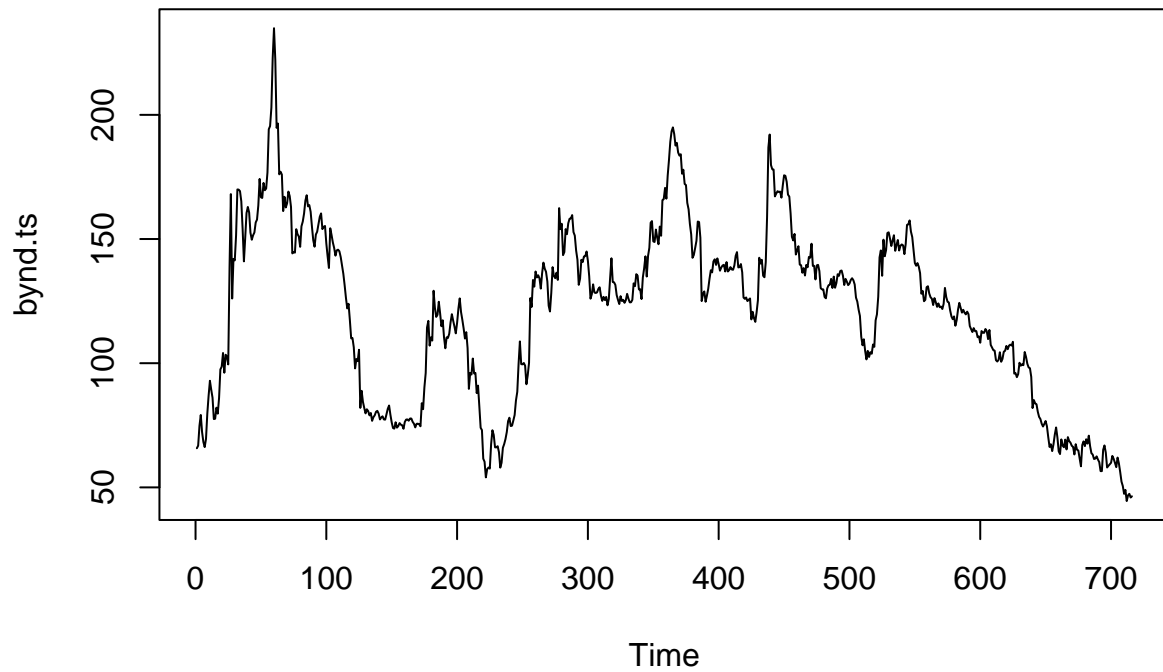
```
## [1] 4.496152
```

The MSPE of Box-Jenkins model is 3.833, while the MSPE of Holt-Winters model is 4.496. Since the MSPE for Box-Jenkins model is lower, we can say that it performs better.

## Problem 2

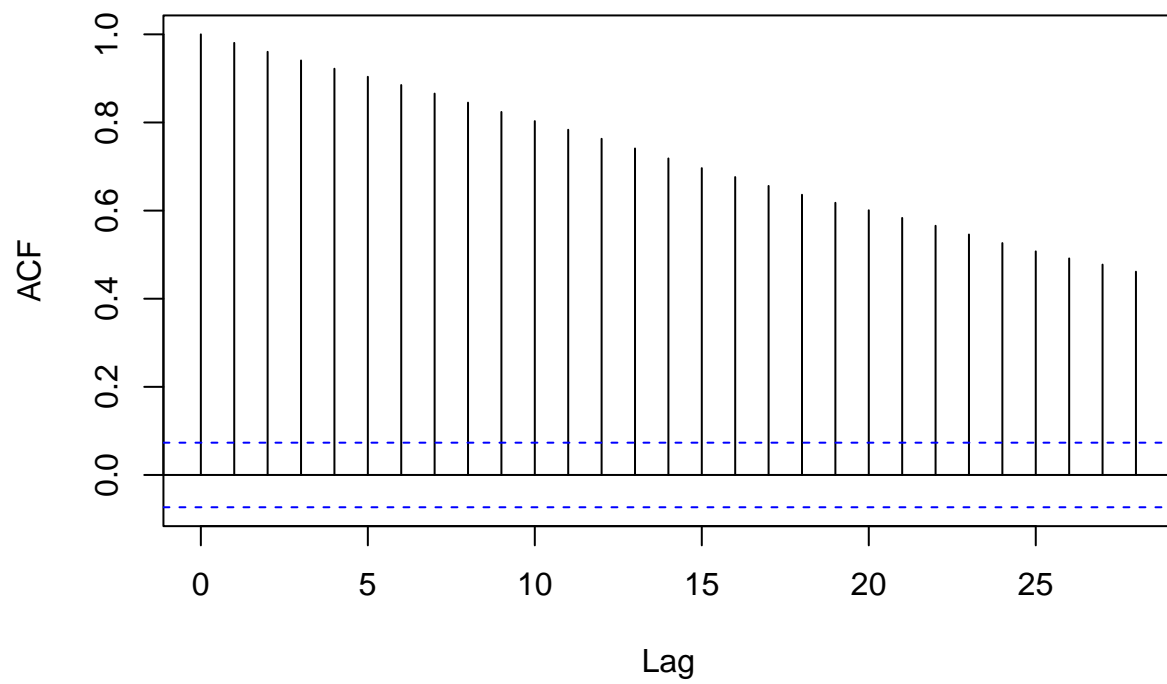
### Part (a)

```
# Reading and coercing into TS  
data2 <- read.table("bynd.txt", sep = " ", header = F)  
bynd.ts <- ts(data2$V2)  
plot(bynd.ts)
```



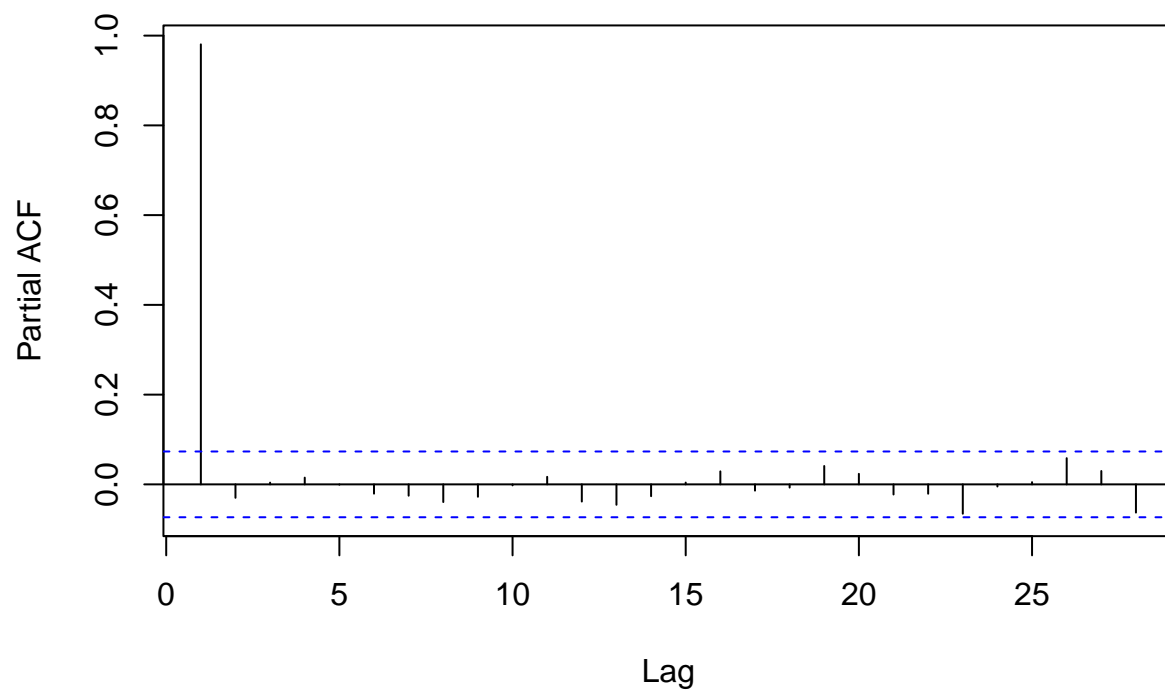
```
acf(bynd.ts)
```

### Series bynd.ts



```
pacf(bynd.ts)
```

## Series bynd.ts



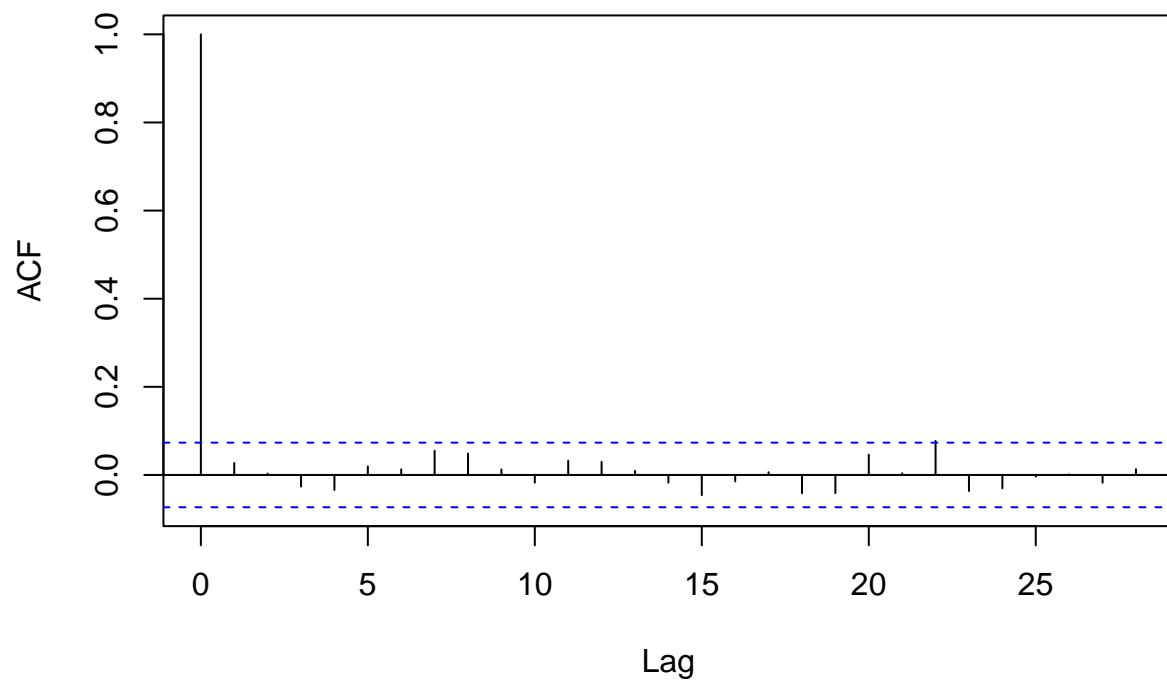
From the ACF, we can easily see that there is very high correlation between subsequent daily closing prices, as is to be expected for any share price time series. The PACF is very high for values with lag  $h = 1$ , and cuts off after that. Thus, we could fit an AR(1) process to this time series.

### Part (b)

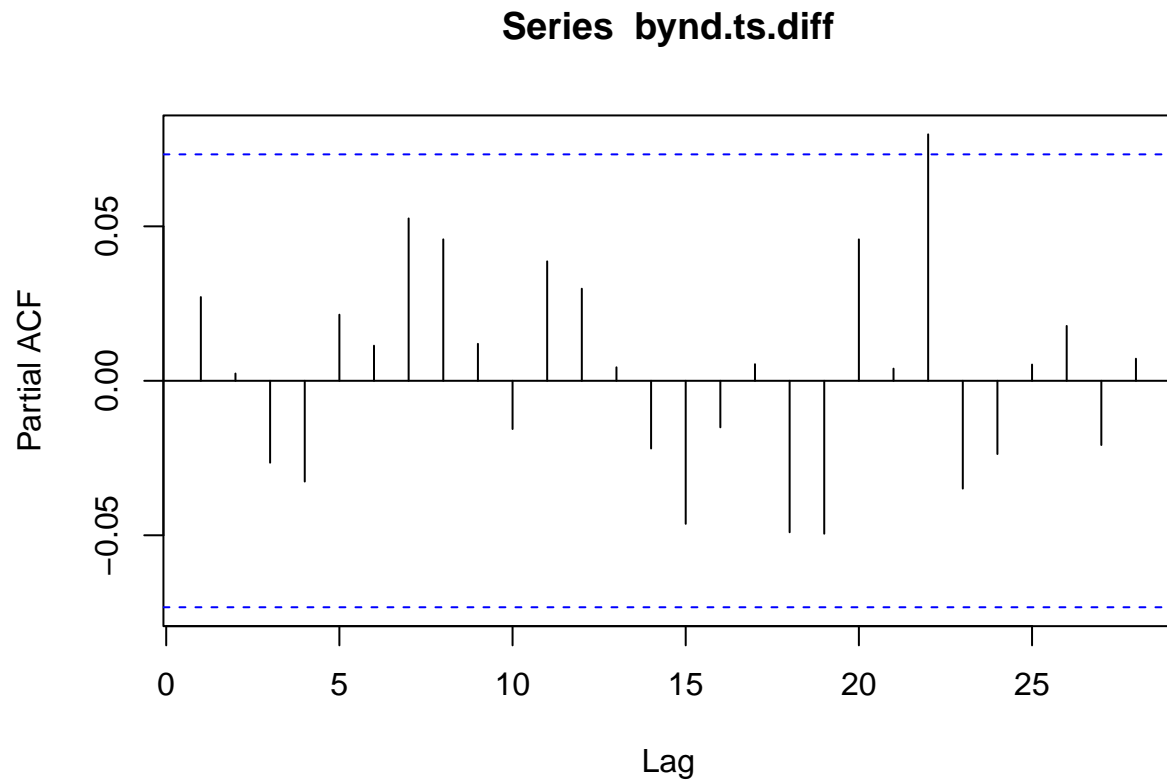
```
bynd.ts.diff <- diff(bynd.ts)
acf(bynd.ts.diff)
```



### Series bynd.ts.diff



```
pacf(bynd.ts.diff)
```



The ACF of the differenced time series resembles that of a white noise, and the PACF values also fall mostly within the confidence intervals, meaning there aren't many that are significant. This differenced time series can thus be modeled by a stationary stochastic process.

### Part (c)

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

This model provides us with a detrended series, which applies here in case of the daily closing price of Beyond Meat Inc. We can remove the trend from many time series by differencing it once, and this detrended series often resembles a stationary process. As seen in part (b), the detrended series resembles a white noise process, so we can say that this is a good model for our data.

### Part (d)

This is an ARIMA (0, 1, 0) model.

### Part (e)

```
bynd.model <- arima(bynd.ts, order = c(0,1,0))
bynd.model
```

```
##
## Call:
```

```
## arima(x = bynd.ts, order = c(0, 1, 0))
##
##
## sigma^2 estimated as 35.98:  log likelihood = -2295.48,  aic = 4592.95
```

The variance of  $Z_t$  is estimated by the ARIMA model we applied to the data. Thus the estimate of variance of  $Z_t$ ,  $\hat{\sigma}_z^2 = 35.98 \approx 36$ .

## Part (f)

We know that  $X_t = X_{t-1} + Z_t$ .

$$\begin{aligned}\hat{X}_t(l) &= \mathbf{E}[X_{t+l}|X_t = x_t, X_{t-1} = x_{t-1} \dots] \\ &= \mathbf{E}[X_{t+l-1} + Z_{t+l}|X_t = x_t, X_{t-1} = x_{t-1} \dots] \\ &= \mathbf{E}[X_{t+l-2} + Z_{t+l-1} + Z_{t+l}|X_t = x_t, X_{t-1} = x_{t-1} \dots]\end{aligned}$$

If we keep expanding this, we get,

$$\hat{X}_t(l) = \mathbf{E}[X_{t+l-l} + Z_{t+l} + Z_{t+l-1} \dots + Z_{t+l-l+1}|X_t = x_t, X_{t-1} = x_{t-1} \dots]$$

Now,  $l > 0$ , so,

$$\begin{aligned}\mathbf{E}[Z_{t+1}] &= \mathbf{E}[Z_{t+2}] = \dots = \mathbf{E}[Z_{t+l}] = 0 \\ \therefore \hat{X}_t(l) &= \mathbf{E}[X_t|X_t = x_t, X_{t-1} = x_{t-1} \dots] \\ &\Rightarrow \hat{X}_t(l) = x_t\end{aligned}$$

So for any lead time  $l > 0$ , we would use  $x_t$  as the forecast.

## Part (g)

Writing our AR(1) process as an infinite MA process,

$$\begin{aligned}X_t &= (1 - B)^{-1}Z_t \\ \Rightarrow X_t &= (1 + B + B^2 \dots)Z_t \\ \therefore \psi_j &= 1 \quad \text{for all } j \in [0, \infty)\end{aligned}$$

Now, our forecast error is given by,

$$e_t(l) = Z_{t+l} + Z_{t+l-1} + Z_{t+l-2} + Z_{t+l-3} \dots Z_{t+1}$$

And,

$$\text{Var}[e_t(l)] = l \times \sigma_z^2$$

So, the 90% prediction interval at lead time  $l$  is given by,

$$\left[ x_t - 1.64 \cdot \sqrt{l \times \sigma_z^2}, x_t + 1.64 \cdot \sqrt{l \times \sigma_z^2} \right]$$

For our prediction of the stock price on March 10, assuming that we are only counting trading days,

$$\hat{x}_{716}(5) = x_{716} = 46.41$$

And the 90% prediction interval is,

$$\left[ 46.41 - 1.64 \cdot \sqrt{5 \times 36}, 46.41 + 1.64 \cdot \sqrt{5 \times 36} \right]$$

So, the 90% prediction interval is

$$[24.41, 68.41]$$

## Part (h)

Our forecasts will all be based on  $x_t = 46.41$ , and the forecasted value at any point for Beyond Meat Inc. share price will be 46.41. It is likely that the price remains quite similar in the short term, so if the price of 46.41 is satisfactory for the director, then it would be a good idea to sell some stocks in the short term. Especially if looking to avoid risk, it would be a good idea to sell in the short run. However, in the long run, the price of the stock behaves quite randomly, in the sense that its as likely to increase as it is to decrease, given this random walk model. So any long run prediction about the Beyond Meat Inc. share price is difficult.