

# STAT 443 Assignment 3

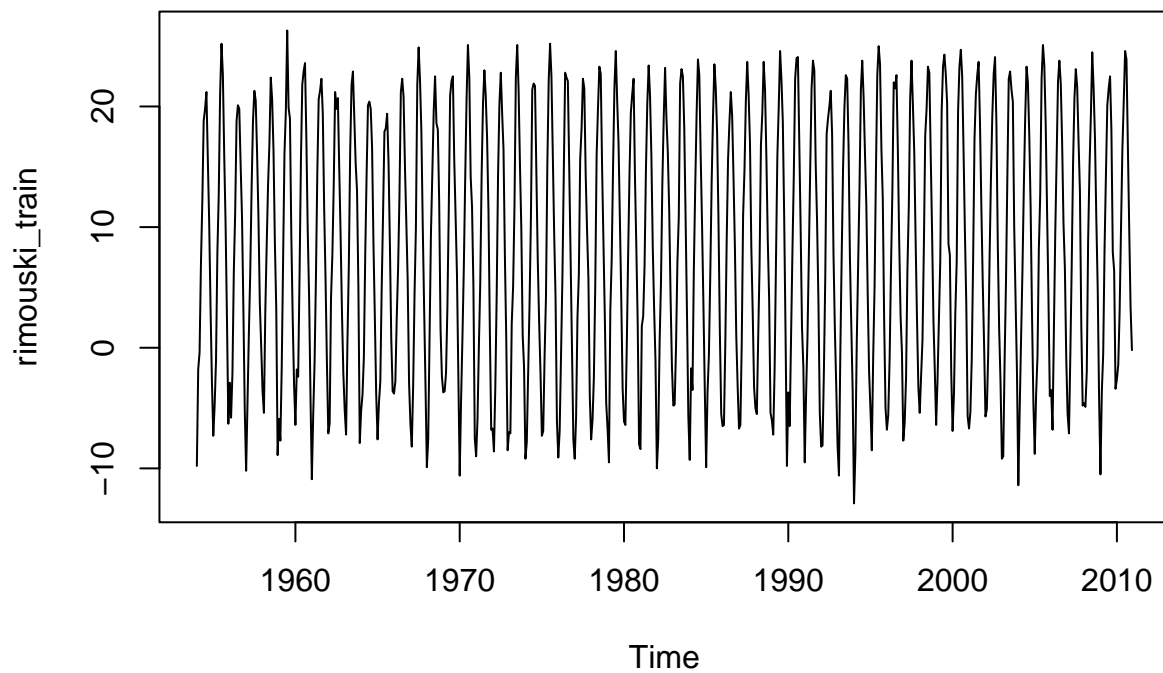
Derek Situ (62222807)

3/21/2022

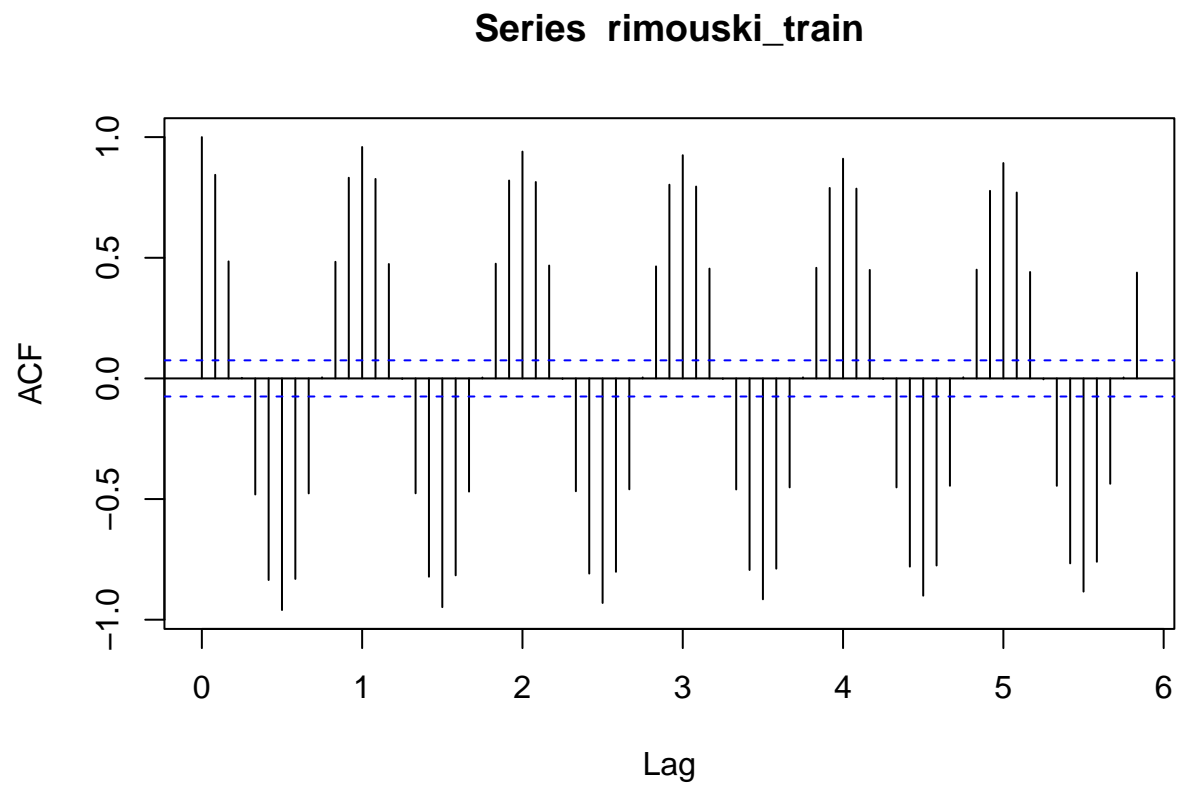
## Question 1

(1a)

```
rimouski <- read.csv("rimouski.csv", header = TRUE)
mean_max <- ts(rimouski$Mean.Max.Temp, start = c(1954, 1), frequency = 12)
rimouski_train <- window(mean_max, start = c(1954, 1), end = c(2010, 12))
rimouski_test <- window(mean_max, start = c(2011, 1), end = c(2016, 12))
plot(rimouski_train)
```

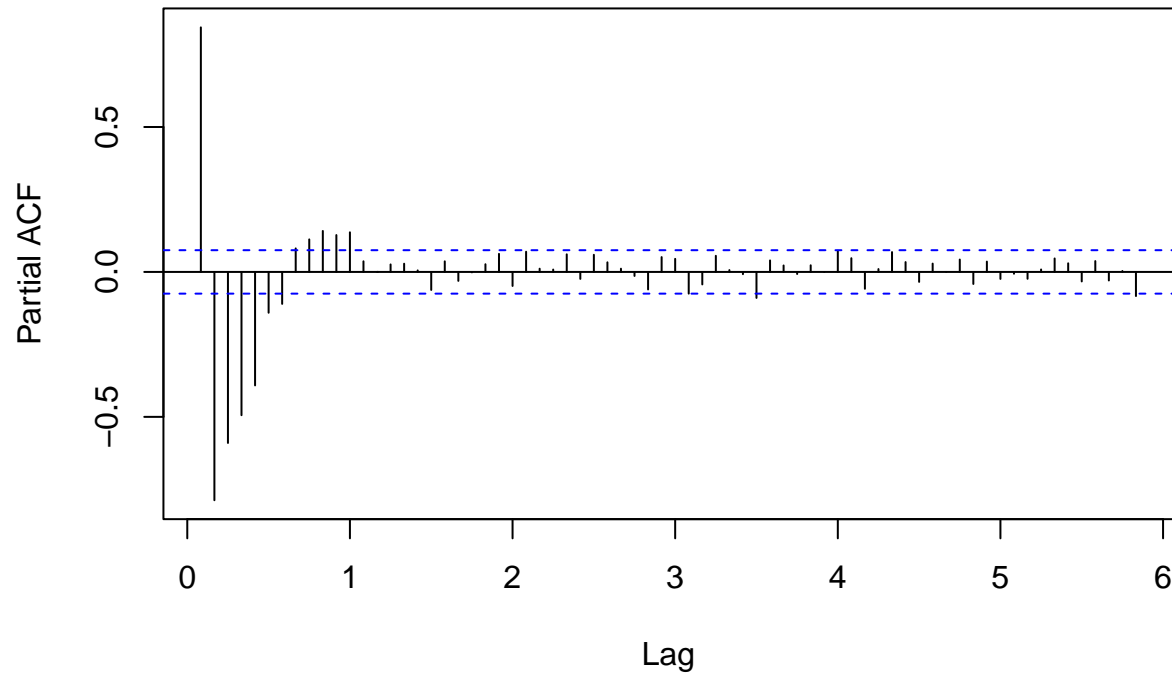


```
acf(rimouski_train, lag.max = 70)
```



```
pacf(rimouski_train, lag.max = 70)
```

## Series rimouski\_train



From the plot of the data we notice a clear, consistent seasonal pattern. The acf oscillates around 0 with a period of 12 months and decays very slowly. The pacf cuts off after lag = 12 months.

(1b)

(1b.i)

```
rimouski_fit <- arima(rimouski_train, order = c(0, 0, 0),  
                      seasonal = list(order = c(0, 1, 0), period = 12))
```

The model equation is

$$Y_t = Z_t$$

where  $Y_t = X_t - X_{t-12}$ . Thus we have

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

or

$$X_t = X_{t-12} + Z_t.$$

which is an ARMA(12,0) process.

(1b.ii)

```
rimouski_fit$sigma2
```

```
## [1] 6.121756
```

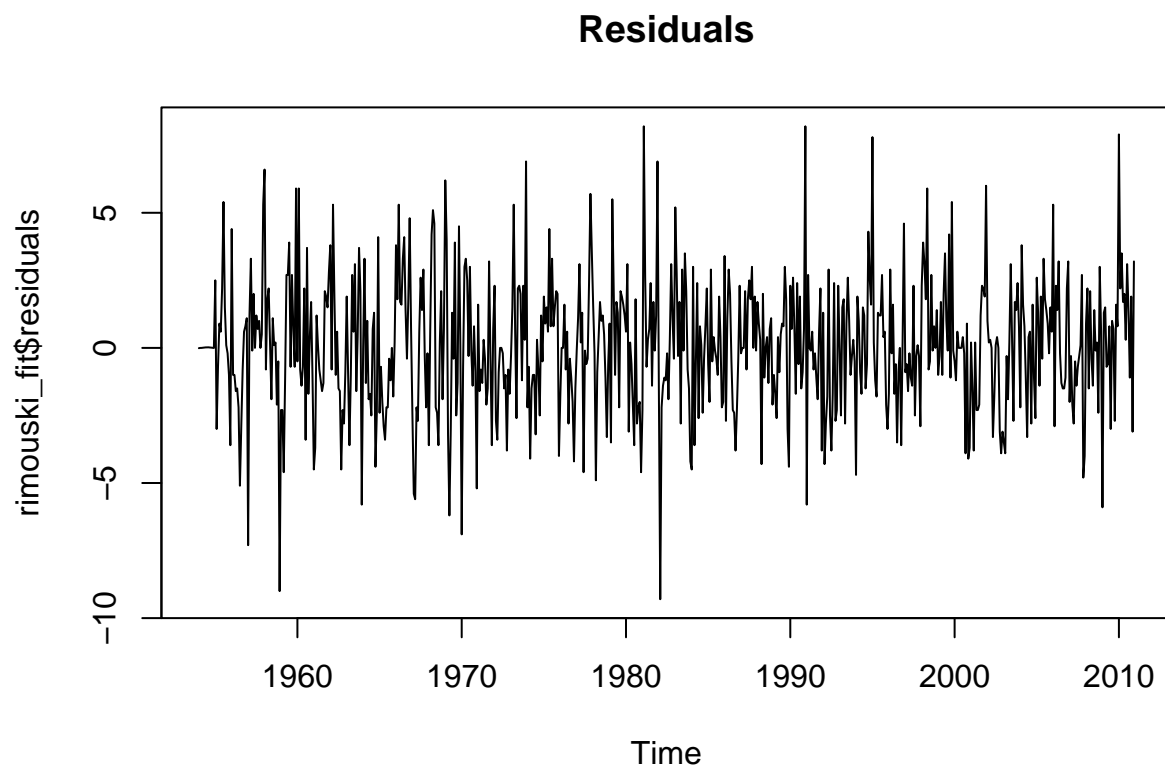
The estimates for the parameters are

$$\hat{\alpha}_i = 0, \hat{\alpha}_{12} = 1, \hat{\sigma}^2 = 6.122$$

for  $i = 1, 2, \dots, 11$ .

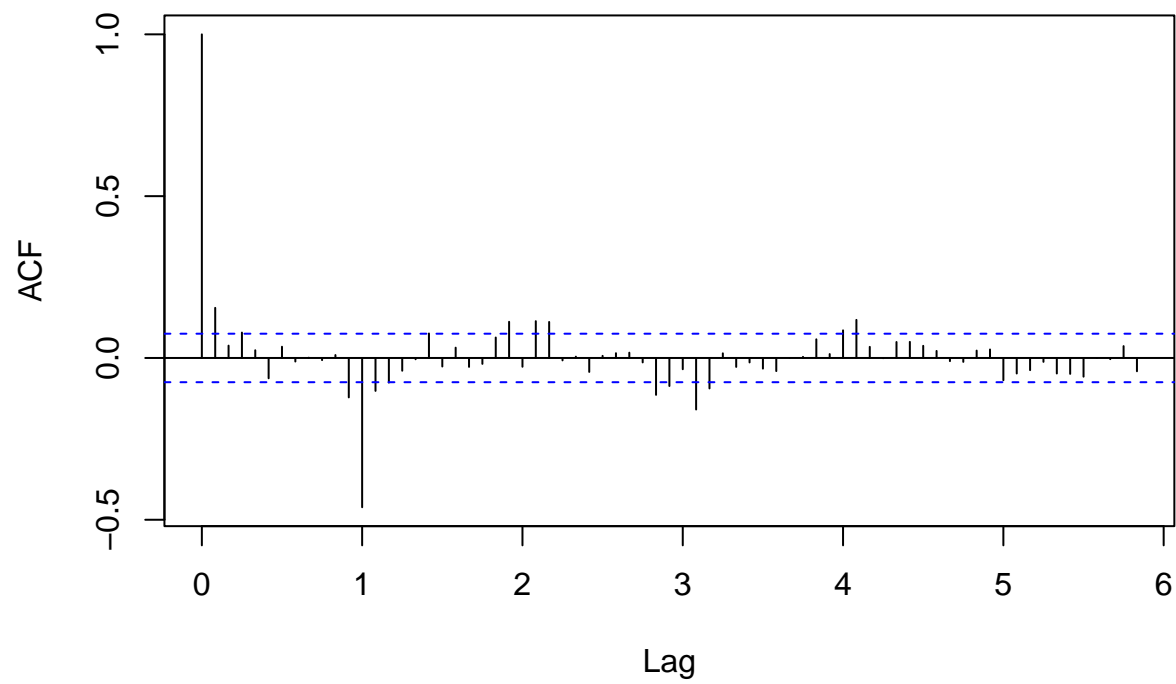
(1b.iii)

```
plot(rimouski_fit$residuals, main = "Residuals")
```

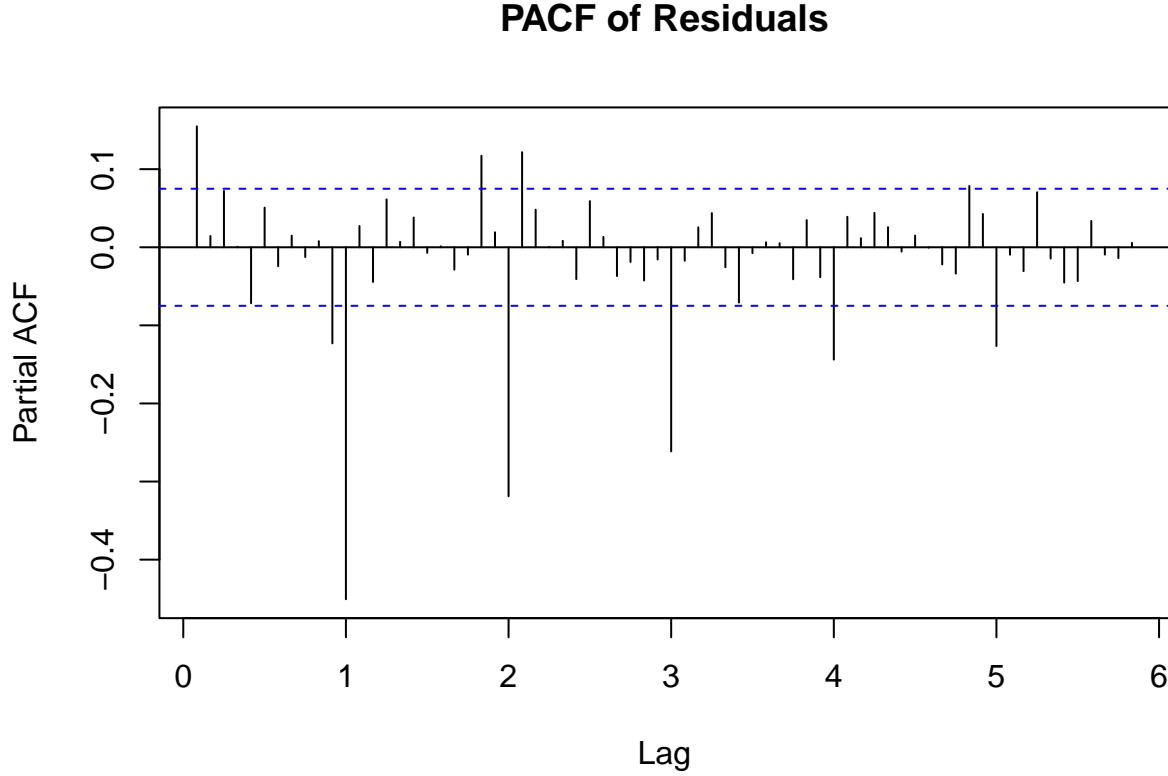


```
acf(rimouski_fit$residuals, lag.max = 70, main = "ACF of Residuals")
```

## ACF of Residuals



```
pacf(rimouski_fit$residuals, lag.max = 70, main = "PACF of Residuals")
```



From the plot of the residuals, it is hard to tell if they have autocorrelation since the data points are condensed very closely together. The plot of the acf of the residuals reveals that at a lag of 12 months, there is a large negative autocorrelation that is significant while the acf is insignificant at other lags. From the plot of the pacf we see it is significant for the first few lags that are multiples of 12.

(1c)

(1c.i)

From the plots made in (b), we see that there are significant negative autocorrelations in the residuals at lag = 12 months, and significant negative partial autocorrelations for the first few lags that are multiples of 12 months. This is a sign that there may be something we are not accounting for that has a seasonal period of 12 months. We can add a seasonal MA(1) component to address this.

(1c.ii)

If  $X_t$  is a SARIMA(0,0,0)x(0,1,1)<sub>12</sub> process, then the model equation is

$$W_t = \beta Z_{t-12} + Z_t$$

where  $W_t = X_t - X_{t-12}$ . Thus we have

$$X_t - X_{t-12} = \beta Z_{t-12} + Z_t$$

or

$$X_t = X_{t-12} + \beta Z_{t-12} + Z_t.$$

This is an ARMA(12,12) process.

(1c.iii)

```
rimouski_fit2 <- arima(rimouski_train,
                      order = c(0,0,0),
                      seasonal = list(order = c(0, 1, 1), period = 12))
c(rimouski_fit2$coef, sigma2 = rimouski_fit2$sigma2)

##          sma1      sigma2
## -0.9205594   3.3998211
```

The estimates of the parameters are

$$\hat{\beta} = -0.921, \hat{\sigma}^2 = 3.998, \alpha_{12} = 1, \alpha_i = 0$$

for  $i = 1, 2, \dots, 11$ .

(1c.iv)

```
AIC(rimouski_fit2)
```

```
## [1] 2755.956
```

```
AIC(rimouski_fit)
```

```
## [1] 3126.616
```

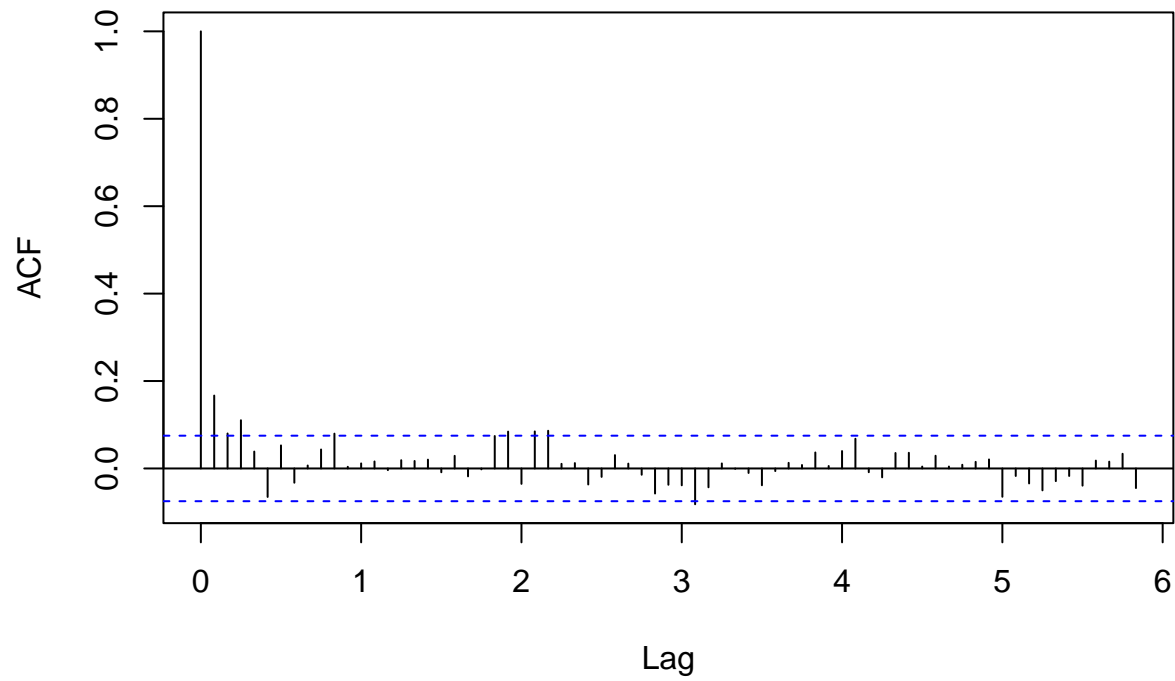
Since the AIC for the SARIMA(0,0,0)x(0,1,1)<sub>12</sub> model is less, I would choose it.

(1d)

(1d.i)

```
acf(rimouski_fit2$residuals, lag.max = 70, main = "ACF of Residuals")
```

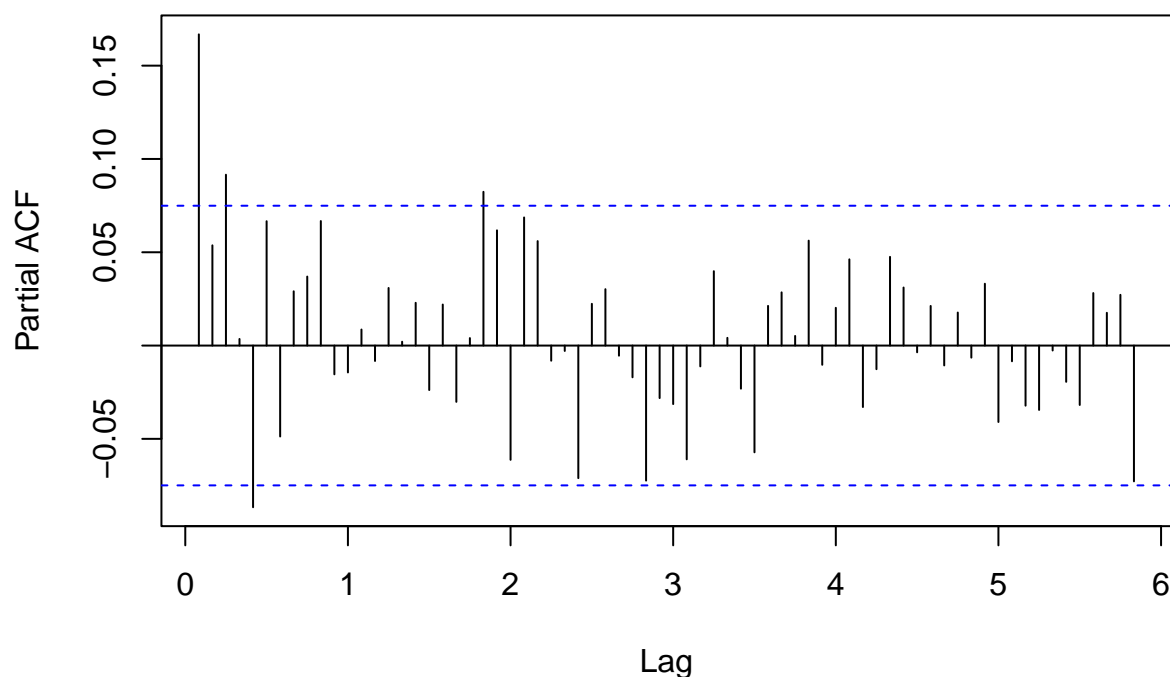
## ACF of Residuals



```
pacf(rimouski_fit2$residuals, lag.max = 70, main = "PACF of Residuals")
```



## PACF of Residuals



The acf and pacf of the residuals are both significant at lag = 1 month which is a sign that we may have a consistent issue with our predictions at 1 lag apart. By adding an AR(1) component we can update how our model incorporates a correlation at 1 lag apart to address this.

(1d.ii)

If  $X_t$  is a SARIMA(1,0,0)x(0,1,1)<sub>12</sub> process then the model equation is

$$(1 - \alpha B)(W_t) = (1 + \beta B^{12})Z_t$$

with  $W_t = X_t - X_{t-12}$ . Then we have

$$X_t - X_{t-12} - \alpha X_{t-1} + \alpha X_{t-13} = \beta Z_{t-12} + Z_t$$

or

$$X_t = \alpha X_{t-1} + X_{t-12} - \alpha X_{t-13} + \beta Z_{t-12} + Z_t.$$

(1d.iii)

```
rimouski_fit3 <- arima(rimouski_train,
                      order = c(1, 0, 0),
                      seasonal = list(order = c(0, 1, 1), period = 12))
c(rimouski_fit3$coef, sigma2 = rimouski_fit2$sigma2)
```

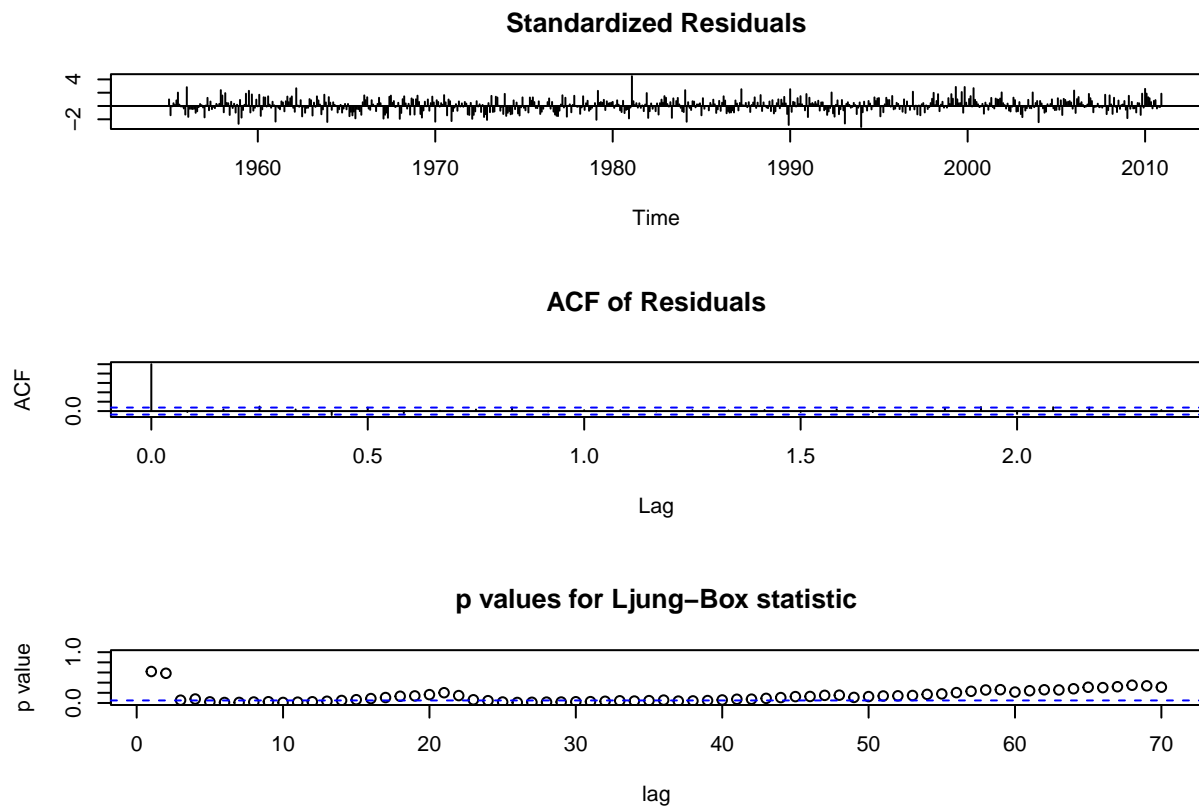
```
##          ar1          sma1          sigma2
## 0.1808350 -0.9354127  3.3998211
```

The estimates of the parameters are

$$\hat{\alpha} = 0.181, \hat{\beta} = -0.935, \hat{\sigma}^2 = 3.400$$

(1d.iv)

```
tsdiag(rimouski_fit3, gof.lag = 70)
```



From the plot of the standardized residuals and the acf of the residuals, we might conclude that the residuals are small and random, which is a good sign for the model fit. But the p-values for the Ljung-Box statistic are low which is a sign that the model fits poorly. Considering this, we might say that the model is not perfect but might be good enough for our purposes.

(1d.v)

```
AIC(rimouski_fit3)
```

```
## [1] 2736.425
```

```
AIC(rimouski_fit2)
```

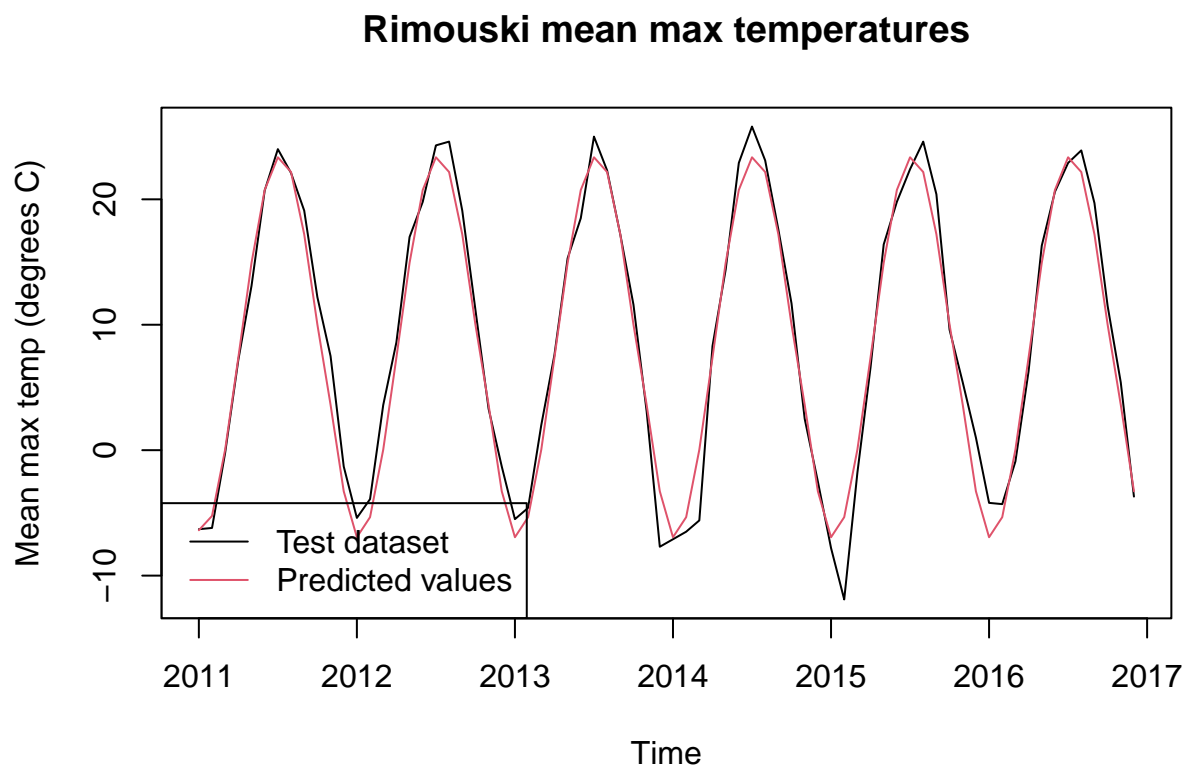
```
## [1] 2755.956
```

The AIC is lower than the models without the AR(1) component, which tells us that the model has improved.

(1e)

```
plot(rimouski_test,
     main = "Rimouski mean max temperatures",
     ylab = "Mean max temp (degrees C)")

rimouski_sarima_prediction <- predict(rimouski_fit3, n.ahead = 72)
lines(rimouski_sarima_prediction$pred, type = "l", col = 2)
legend("bottomleft",
     legend = c("Test dataset", "Predicted values"),
     col = c(1, 2),
     lwd = 1)
```

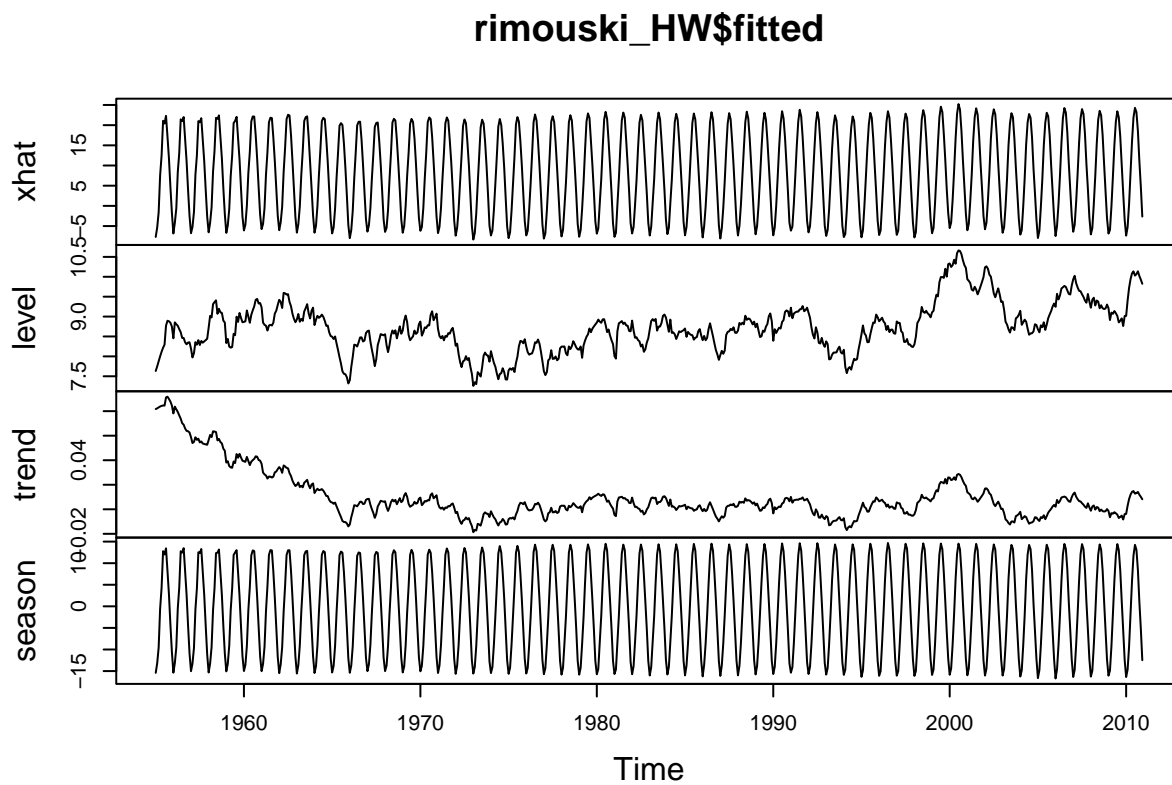


From this plot we see that the predictions are close to the observed values. We see the largest residuals in the winter months and the summer months, when monthly mean max temperatures reach either lows or peaks.

(1f)

(1f.i)

```
rimouski_HW <- HoltWinters(rimouski_train)
plot(rimouski_HW$fitted)
```



```
rimouski_HW$beta
```

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

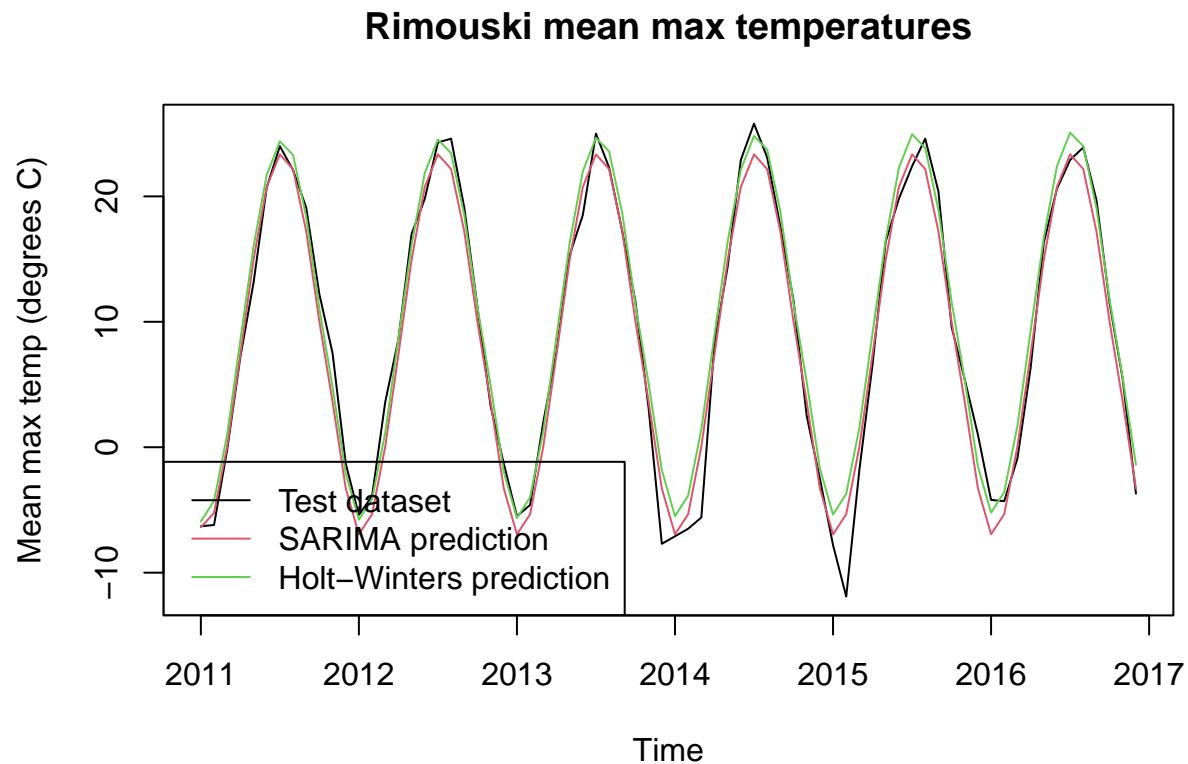
There is not enough evidence to suggest a trend, as the  $\beta$  parameter is very close to 0.

(1f.ii)

```
plot(rimouski_test,
     main = "Rimouski mean max temperatures",
     ylab = "Mean max temp (degrees C)")

rimouski_HW_prediction <- predict(rimouski_HW, n.ahead = 72)
```

```
lines(rimouski_sarima_prediction$pred, type = "l", col = 2)
lines(rimouski_HW_prediction, type = "l", col = 3)
legend("bottomleft", legend =
      c("Test dataset", "SARIMA prediction", "Holt-Winters prediction"),
      col = c(1, 2, 3),
      lwd = 1)
```



We see that both prediction methods produce similar results. For the summer months, the Holt-Winters predictions appear to be more accurate, while for the winter months, the Box-Jenkins method appears to be more accurate.

(1f.iii)

```
mspe_sarima <- sum((rimouski_test - rimouski_sarima_prediction$pred)^2) /
  length(rimouski_test)
mspe_hw <- sum((rimouski_test - rimouski_HW_prediction)^2) /
  length(rimouski_test)

c(MSPE_BoxJenkins = mspe_sarima, MSPE_HoltWinters = mspe_hw)
```

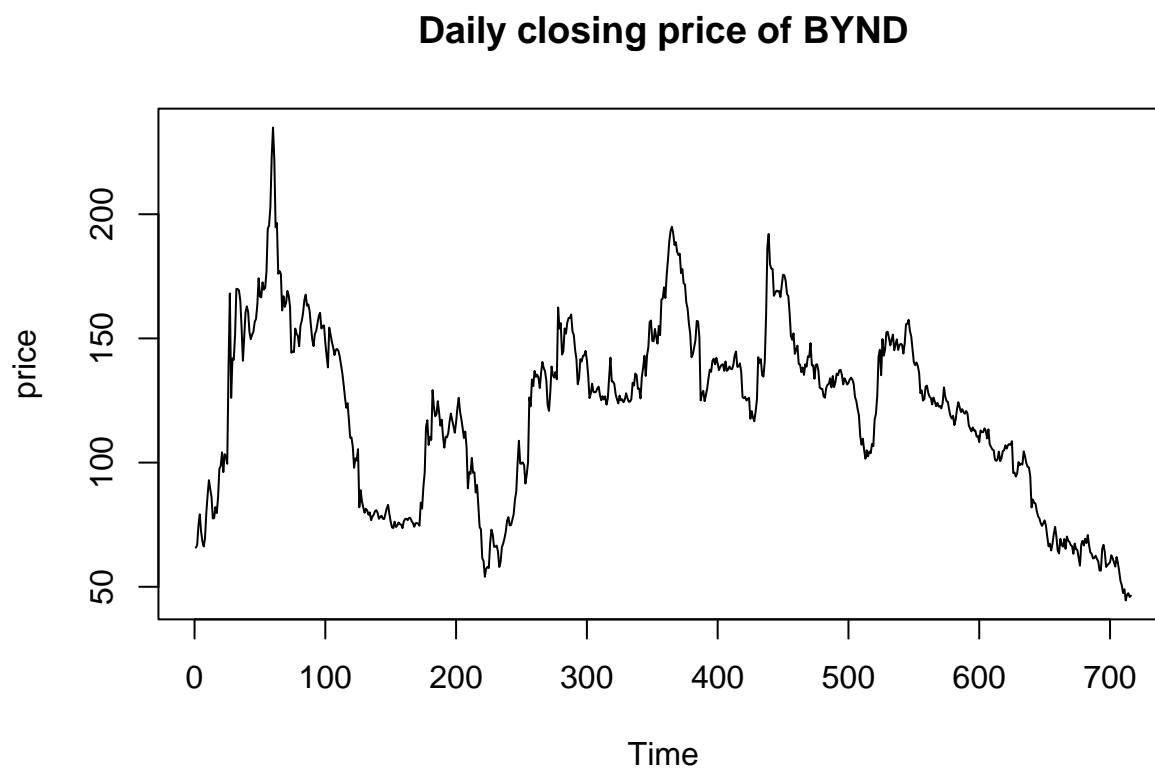
```
## MSPE_BoxJenkins MSPE_HoltWinters
##          3.832878          4.496152
```

The Box-Jenkins model has a lower MSPE, so it performs better.

## Question 2

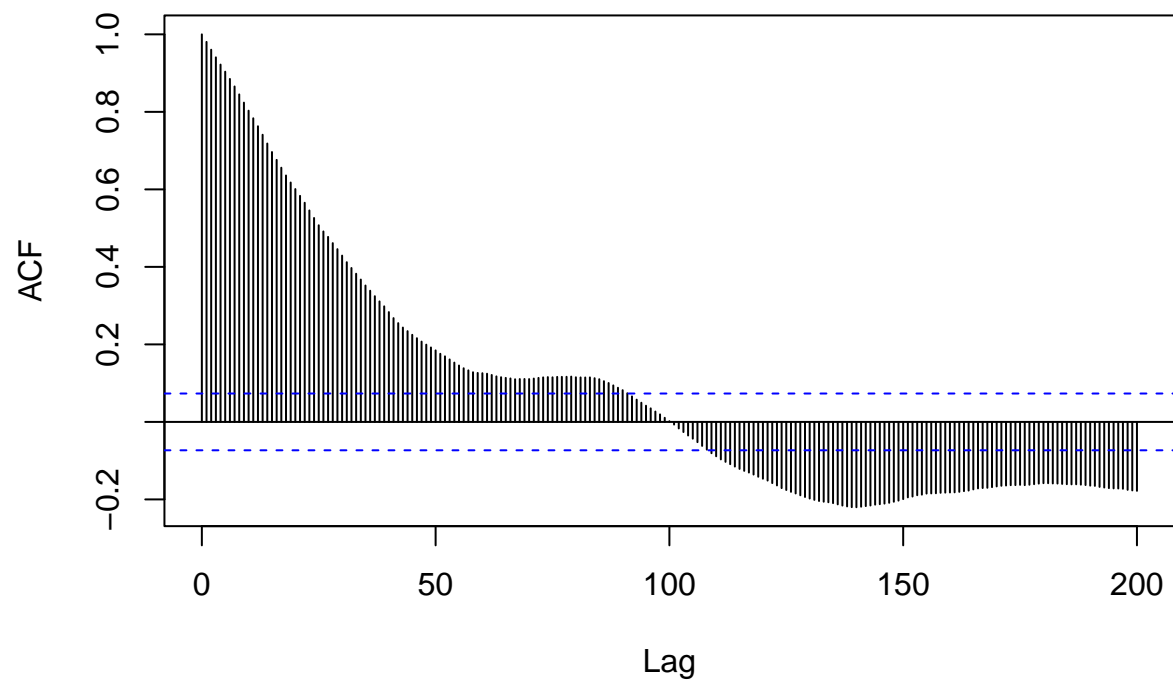
(2a)

```
bynd <- read.table("bynd.txt", header = FALSE, sep = " ")  
price <- ts(bynd$V2, start = 1, end = 716)  
plot(price, main = "Daily closing price of BYND")
```

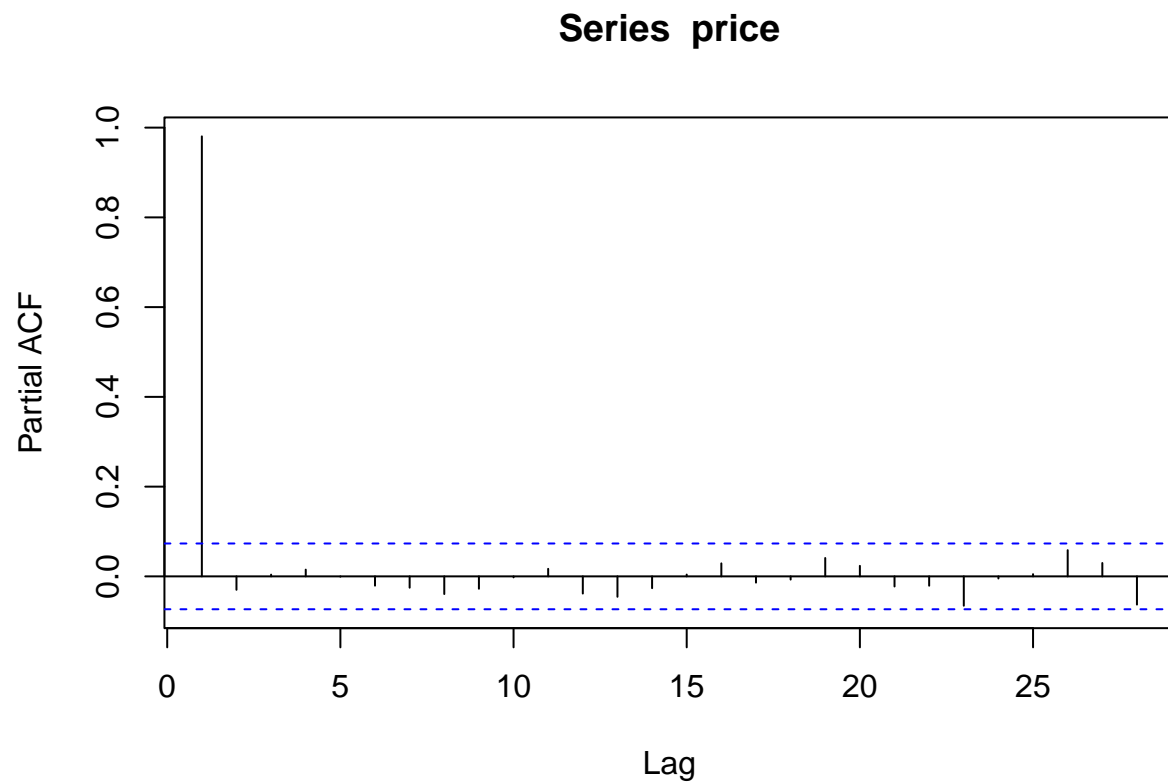


```
acf(price, lag.max = 200)
```

### Series price



```
pacf(price)
```

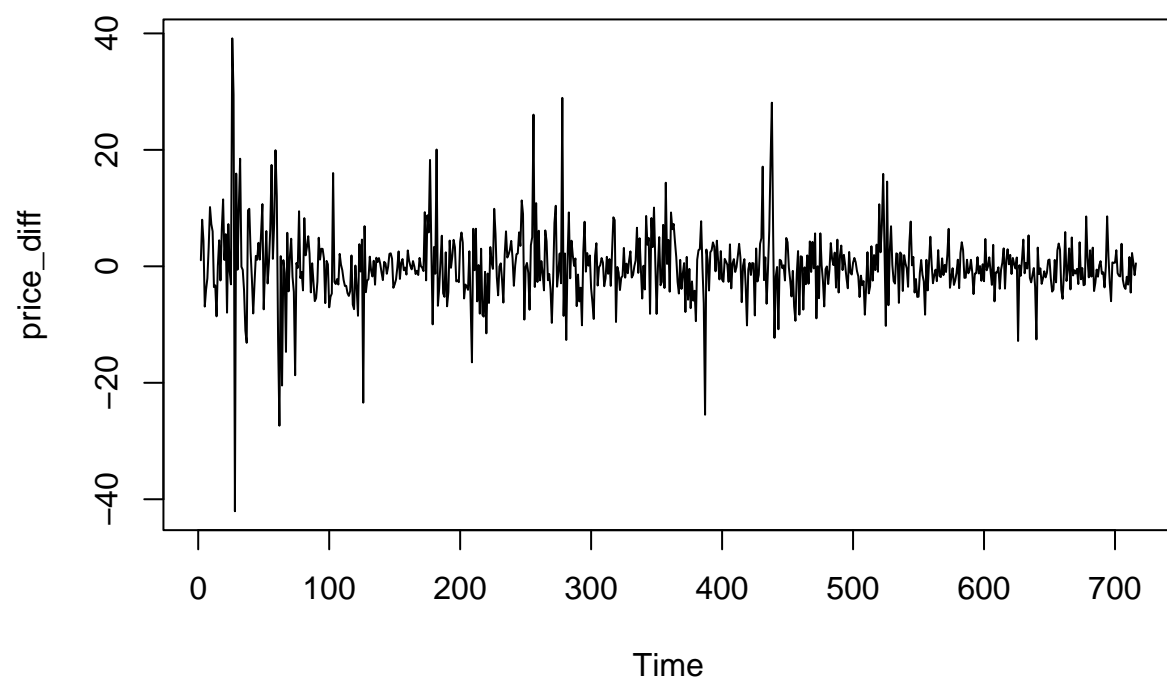


The acf decays slowly and the pacf at lag 1 is very high, and then cuts off.

(2b)

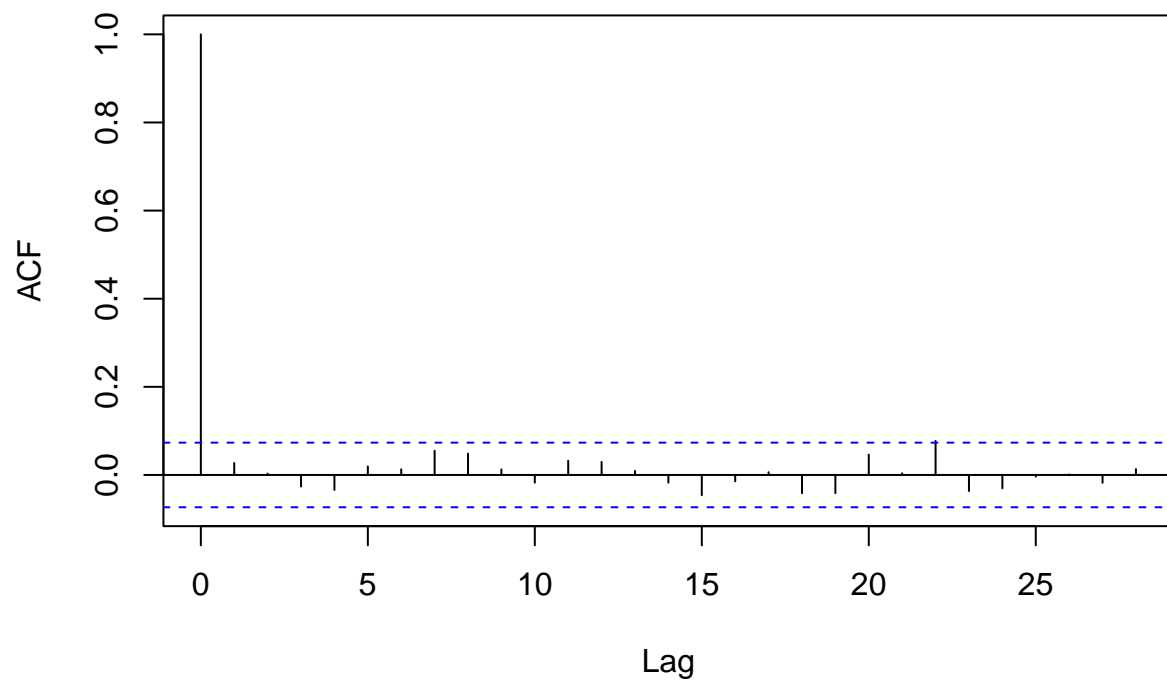
```
price_diff <- diff(price)
plot(price_diff)
```



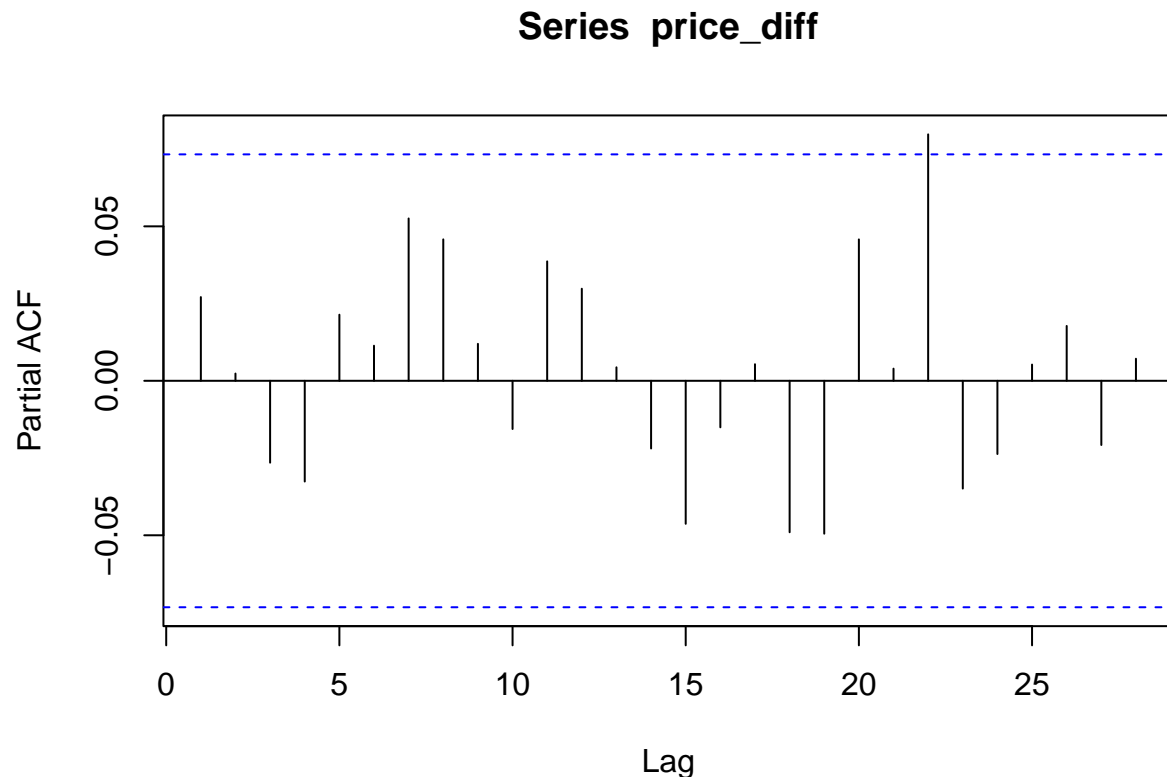


```
acf(price_diff)
```

### Series price\_diff



```
pacf(price_diff)
```



The plot of the differenced series looks like a white noise plot. The acf and pacf are insignificant which also suggests the differenced series is white noise. Since it looks like white noise, it definitely appears stationary.

### (2c)

First we discuss why an AR(1) model fits. From part (a), we see that the acf is slowly decaying and the pacf cuts off after lag = 1, which suggests that an AR(1) model is appropriate.

Next we discuss why it is appropriate to have the  $\alpha$  parameter in the model equal to 1, giving us a random walk process. From part (b), we see that the differenced series can be appropriately modeled by white noise. This means that the change in the closing price of BYND can be expressed as the noise term  $Z_t$ . Then, a way to express the random variable for the price at time  $t$ ,  $X_t$ , is the price at time  $t - 1$  plus the random noise  $Z_t$ , i.e.  $X_t = X_{t-1} + Z_t$ .

### (2d)

As discussed it can be an AR(1) process with  $\alpha = 1$ . It can also be considered an ARIMA(0,1,0) process. Both interpretations lead to the same model equation.

### (2e)

```
bynd_fit <- arima(price, order = c(0, 1, 0))
bynd_fit$sigma2
```

```
## [1] 35.98248
```

```
var(price_diff)
```

```
## [1] 36.03215
```

The estimate for the variance of  $Z_t$  is

$$\hat{\sigma}^2 = 36.$$

(2f)

Assume  $\ell \in \mathbb{Z}, \ell > 0$ . The forecast at lead time  $\ell$  is

$$\begin{aligned} x_t(\ell) &= E(X_{t+\ell} | X_t = x_t, X_{t-1} = x_{t-1}, \dots) \\ &= E(X_{t+\ell-1} + Z_{t+\ell} | X_t = x_t, X_{t-1} = x_{t-1}, \dots) \\ &= E(X_t + Z_t + Z_{t+1} + \dots + Z_{t+\ell} | X_t = x_t, X_{t-1} = x_{t-1}, \dots). \end{aligned}$$

And since  $\forall i, E(Z_i) = 0$ , we have

$$\begin{aligned} x_t(\ell) &= E(X_t | X_t = x_t, X_{t-1} = x_{t-1}, \dots) \\ &= x_t \end{aligned}$$

Thus I would use  $x_t$  for the forecast at lead time  $\ell$ .

(2g)

The 90% prediction interval for the forecast at lead time  $\ell$  is

$$[x_t(\ell) - 1.64\sqrt{\text{Var}(\hat{e}_n(\ell))}, x_t(\ell) + 1.64\sqrt{\text{Var}(\hat{e}_n(\ell))}]$$

where  $\text{Var}(\hat{e}_n(\ell)) = \hat{\sigma}^2(\ell) = 36\ell$ .

Then the prediction for the stock price one week in the future is

$$x_{716}(\hat{\ell}) = x_{716} = 46.41$$

and the 90% confidence interval is

$$\begin{aligned} &[46.41 - 1.64\sqrt{36 \times 5}, 46.41 + 1.64\sqrt{36 \times 5}] \\ &= [24.407, 68.413] \end{aligned}$$

(2h)

Hello Director, we have found that the daily change in closing price of the stock can be modeled well by a white noise process centered around 0. What this means is that mean daily change in price is 0, and you can think of the change that occurs on a given day as being sampled from a normal distribution. If we had many identical stocks to BYND, over the long run we might expect the net change in price of all these stocks to be about 0, but each individual stock's net change within the time period could be quite different from 0. Essentially, it is hard to predict what the price of the stock will be in the long term, but in the short term we can expect it to remain quite similar to the current price. My recommendation is to sell some shares if you are satisfied with the current market value.