STAT 443: Assignment 3

Flora Zhang 52135365

19 Mar, 2022

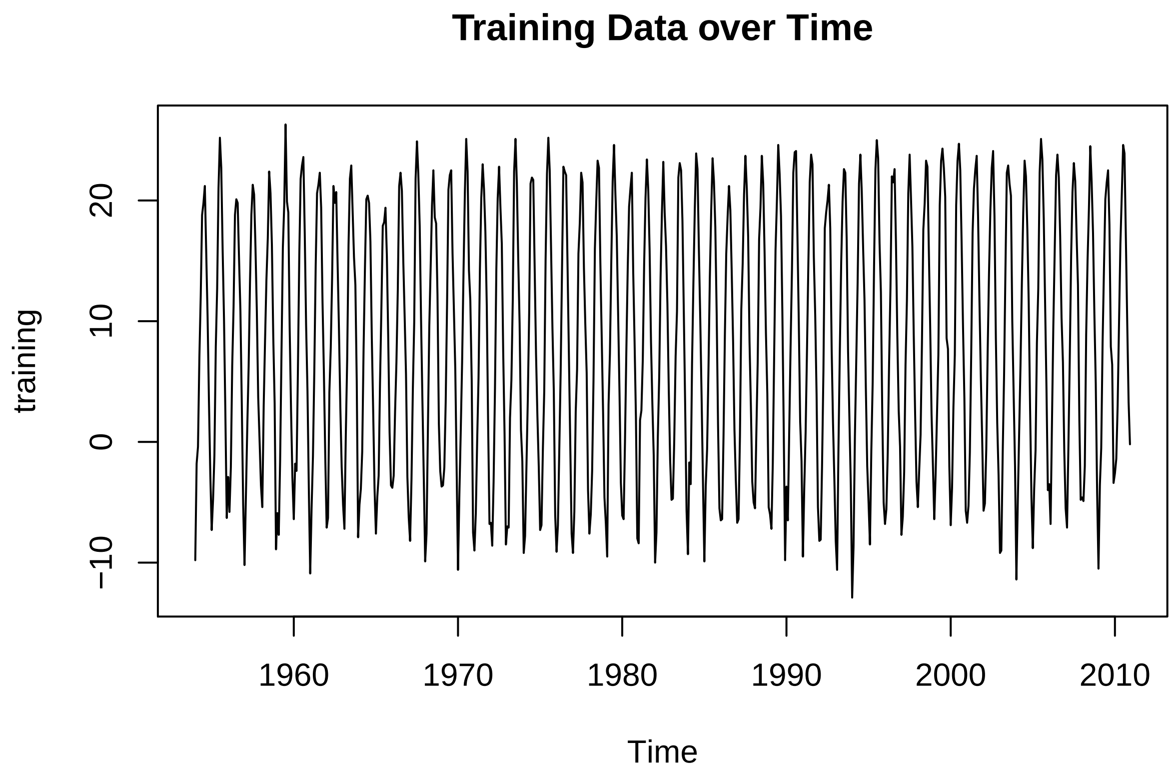
Question 1:

1a)

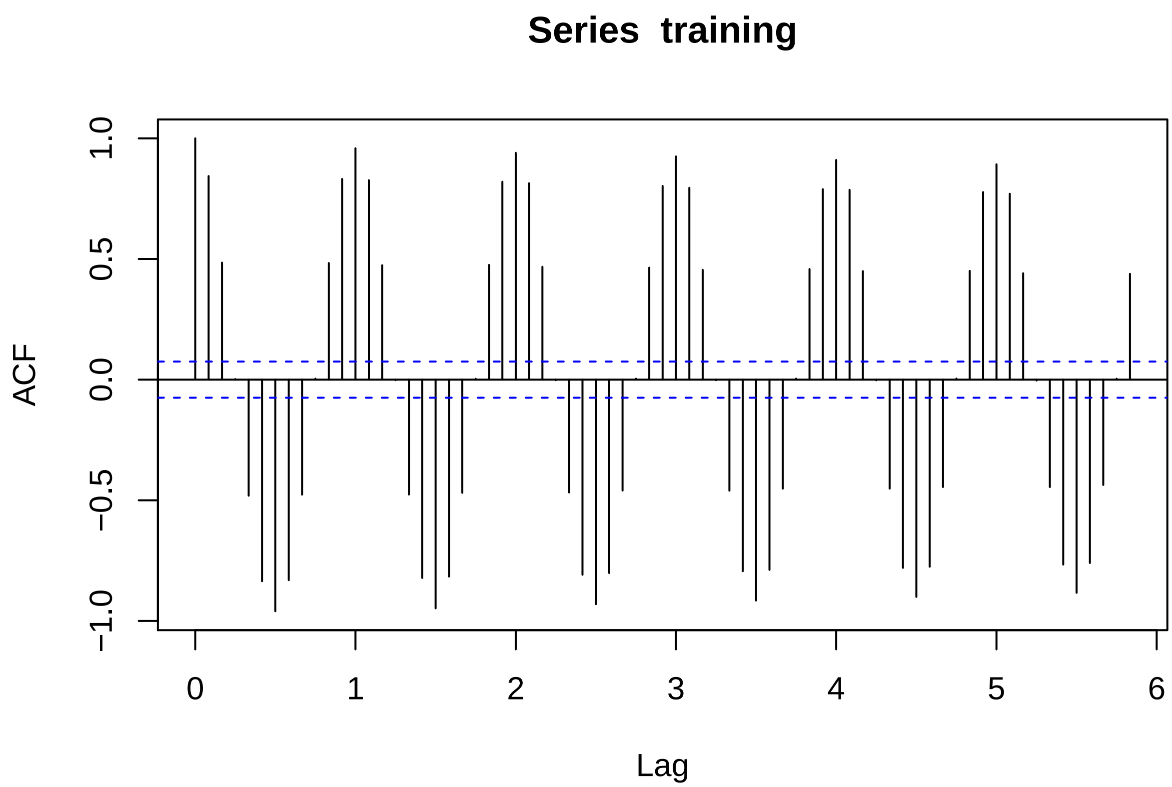
```
rimouski = read_csv("rimouski.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]

##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_double(),
##   Station.Name = col_character(),
##   Climate.ID = col_double(),
##   Date.Time = col_character(),
##   Year = col_double(),
##   Mean.Max.Temp = col_double(),
##   Mean.Min.Temp = col_double()
## )
```

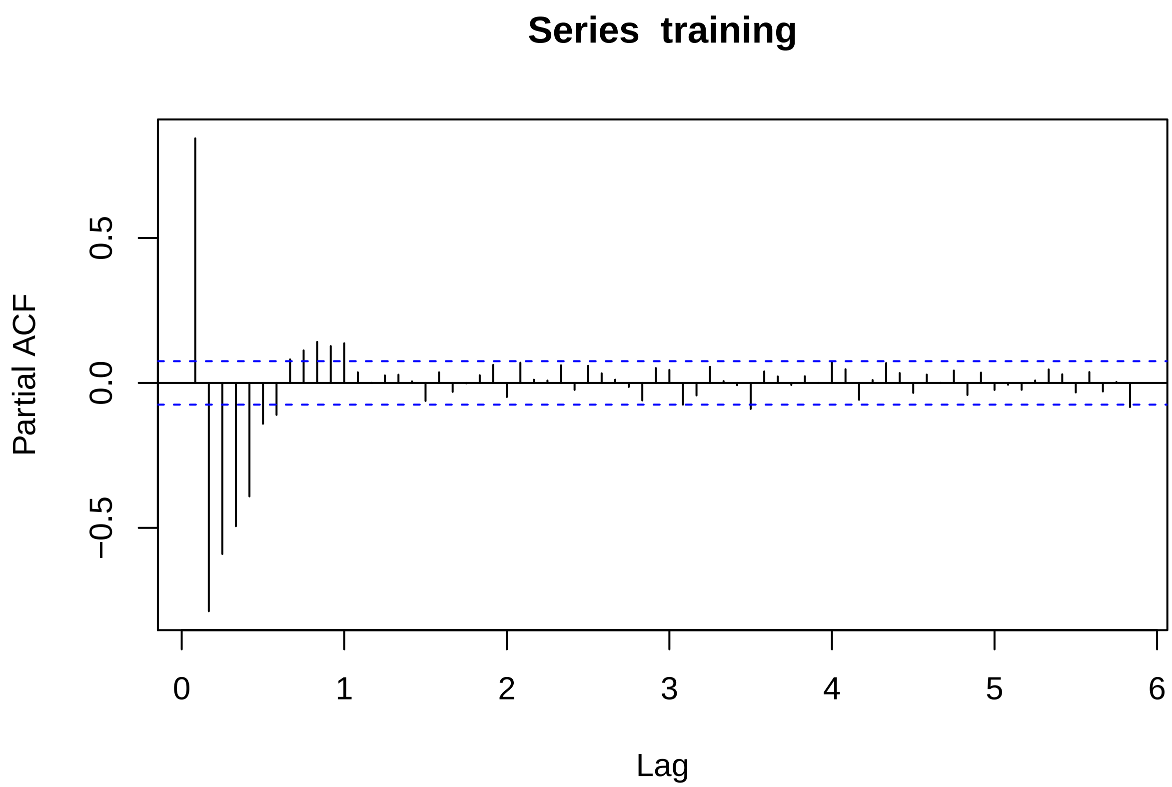
```
rimouski_ts = ts(data = rimouski$Mean.Max.Temp, start = c(1954,1), end = c(2017,8), frequency = 12) #cr
training = window(x=rimouski_ts, start = c(1954,1), end=c(2010,12))   #training
testing = window(x= rimouski_ts, start = c(2011,1), end=c(2016,12)) #testing

#plot training data, acf and pacf
plot(training, main="Training Data over Time")
```




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




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



From the plot of the training data, we can see that there appears to be seasonality but no obvious trend. The ACF oscillates like a damped sine wave, and oscillates at a frequency of 12 (lag 1) due to the seasonal fluctuations. The PACF appears significant, at earlier lags, then cuts off at around lag 1 (frequency of 12).

(b)

i. A SARIMA Model is expressed as: $\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)Z_t$

We have a SARIMA $(0,0,0) \times (0,1,0)_{12}$ model. Here, $\phi(B)\Phi(B^s) = 1$, and $\theta(B)\Theta(B^s)$=1 And $W_t = \nabla^d \nabla_{12}^D X_t = \nabla_{12} X_t = X_t - X_{t-12}$

So, we have \$X_t - X_{t-12} = Z_t \$ \$X_t = X_{t-12} + Z_t \$

We have an ARMA(12,0) process, or an AR(12) process

ii.

```
#fit to training data
arima_function_b = arima(x=training, order =c(0,0,0),  seasonal=list(order=c(0,1,0), period=12), includ
arima_function_b
```

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

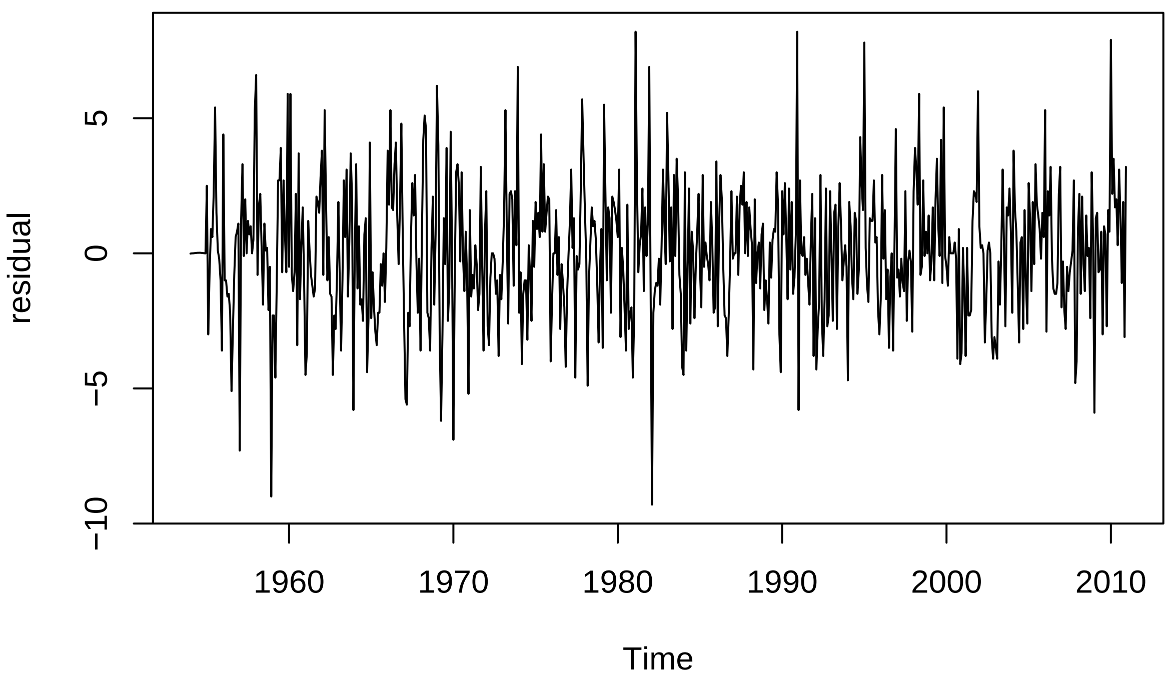
An ARMA(12,0) process should have the following skeleton:

$$X_t = \alpha_1 X_{t-1} + ... + \alpha_{12} X_{t-12} + Z_t$$

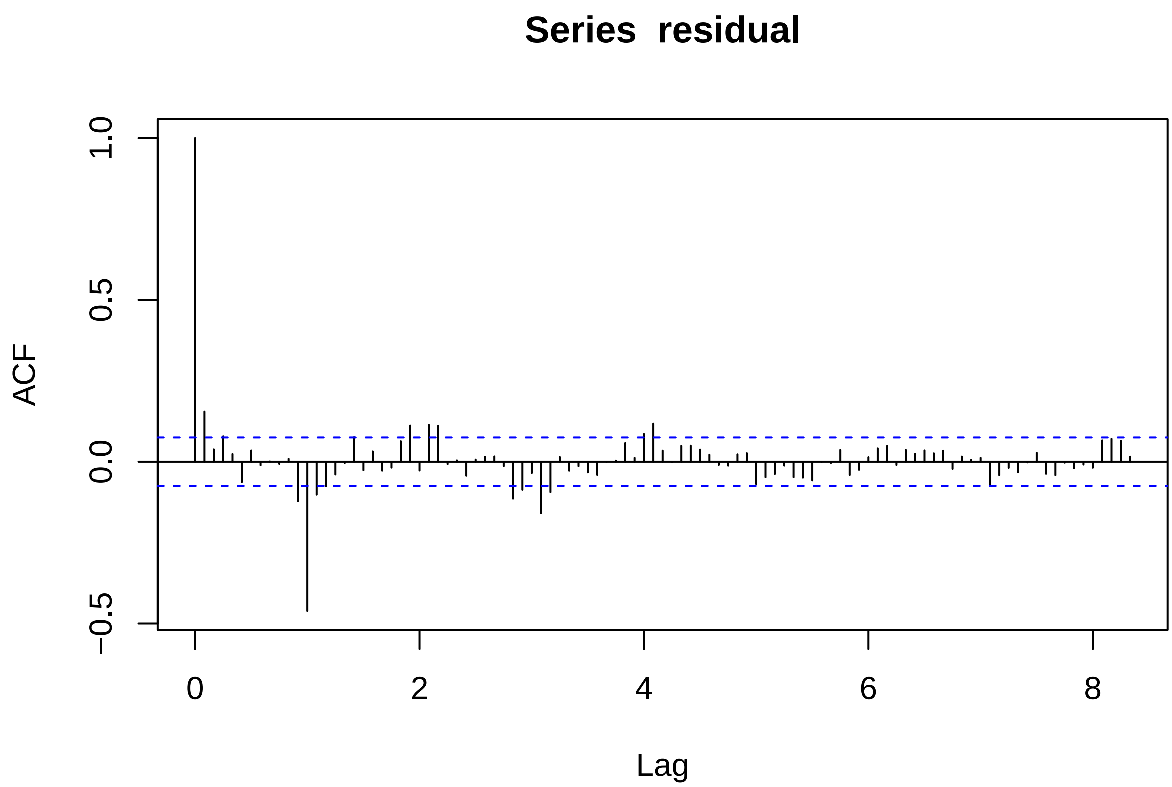
, where $Z_t \sim$ WN(0, 6.122).

iii.

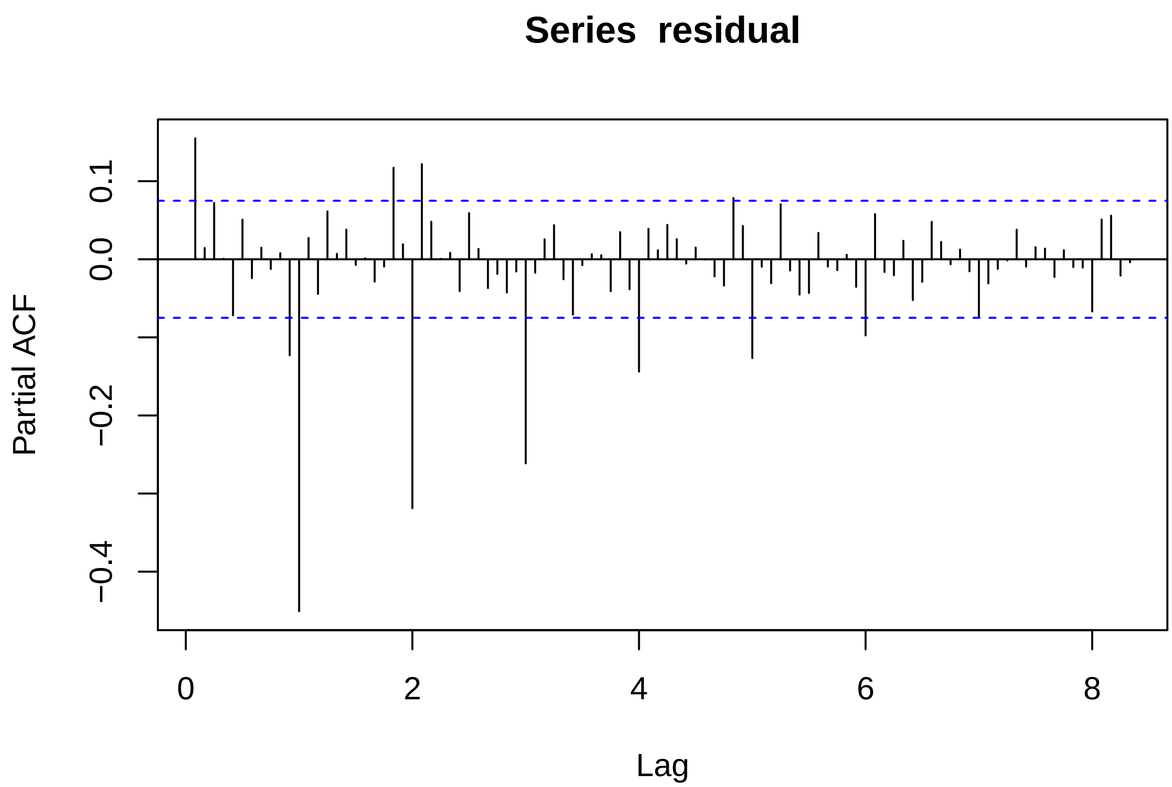
```
residual = residuals(arima_function_b)
plot(residual)
```




```
acf(residual, lag.max = 100)
```




```
pacf(residual, lag.max = 100)
```



The plot of the residuals appears to have no apparent trend. The ACF plot appears to look like a damped sine wave, and displays significant values at lag 1 (12 months), and around lag 3 (3 years). The PACF decays slowly and displays significant values at the seasonal periods, such as lags 1 (12 months) and 2 (24 months), and so on. This suggests that we might want to add a seasonal component to the data.

(c) We propose adding a seasonal MA(Q = 1) component.

i. From the residual ACF, there is a negative spike around lag 1 (12 months). The PACF tapers in multiples of seasonality, meaning it has significant lags at 1 year, 2 years, ... This is similar to a seasonal MA(1) component, hence why we add it in.

ii. A SARIMA Model is expressed as: $\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)Z_t$

We have a SARIMA $(0, 0, 0) \times (0, 1, 1)_{12}$ model. Here, $\phi(B)\Phi(B^s) = 1$, and $\theta(B) = 1$ and $\Theta(B^s) = 1+\beta B^{12}$.

$W_t = \nabla^d \nabla_{12}^D X_t = \nabla_{12} X_t = X_t - X_{t-12}$

So, we have $X_t - X_{t-12} = (1 + \beta B^{12})Z_t$ $X_t - X_{t-12} = Z_t + \beta Z_{t-12}$ $X_t = X_{t-12} + Z_t + \beta_{12} Z_{t-12}$

We have an ARMA(12,12) process.

iii. Fit the model using the arima() function, and specifically identify the model parameters and their estimates (do not just print the function results).

- you only need beta and sigma squared terms

```r
arima_function_c = arima(x=training, order =c(0,0,0),  seasonal=list(order=c(0,1,1), period=12), includ
arima_function_c
```

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

An ARMA(12,12) process should have the following skeleton: $X_t = \alpha_1 X_{t-1} + ... + \alpha_{12} X_{t-12} + Z_t + \beta_1 Z_{t-1} + ... + \beta_1 Z_{t-1} + ... + \beta_{12} Z_{t-12}$

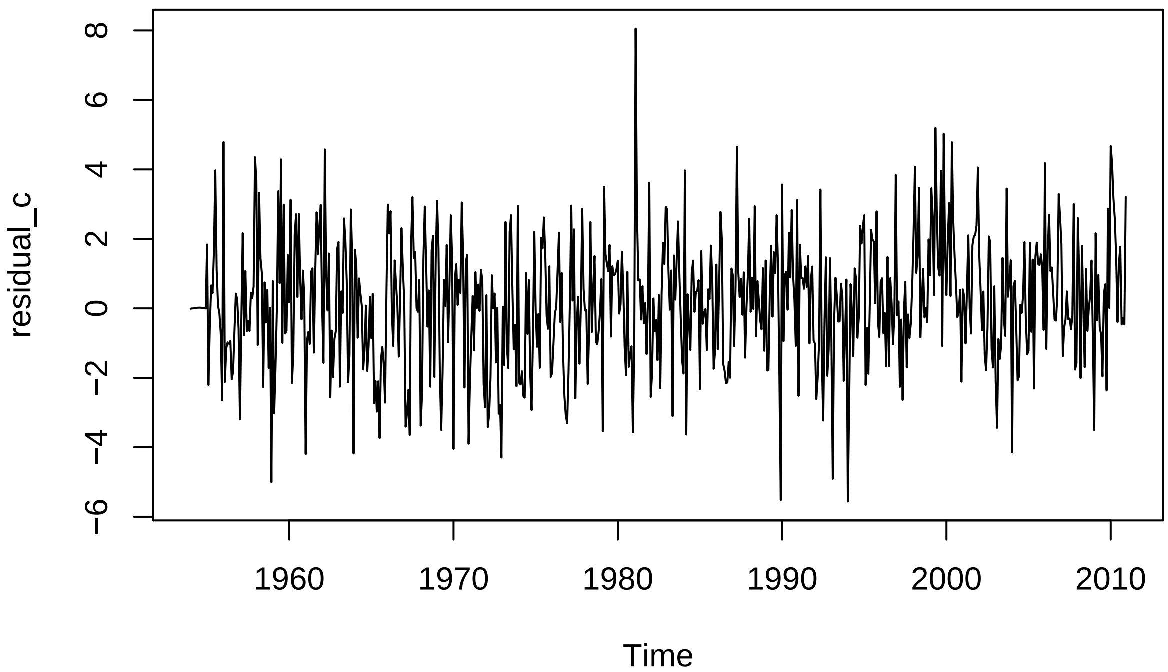
Here,$\beta_1 = ... = \beta_{11} = 0$, $\beta_{12} = -0.921$, and Z_t~WN(0, 3.657)

iv. The AIC of this model is 2755.96, and the AIC of the previous model is 3126.62. I would pick the SARIMA (0,0,0)x(0,1,1)s model over the $(0, 0, 0)\times(0, 1, 0)$s model, because it has a smaller AIC.

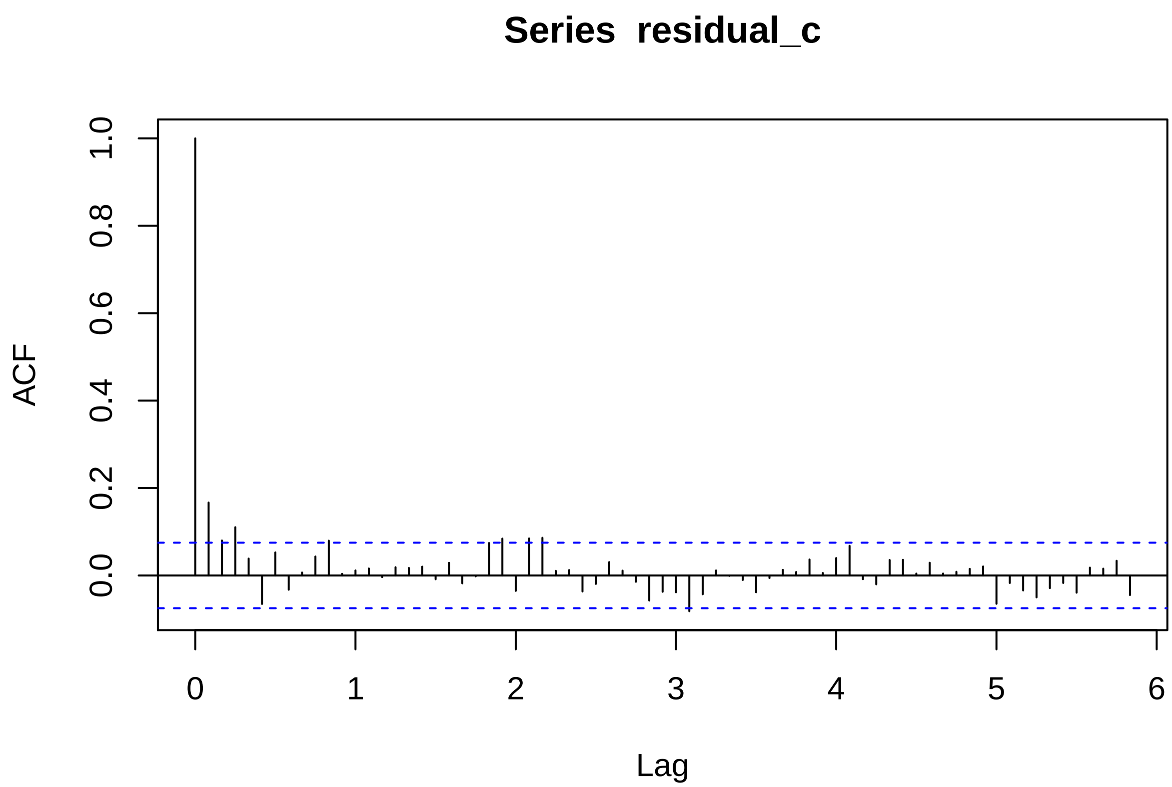
(d) We now add an AR(p = 1) component.

i.

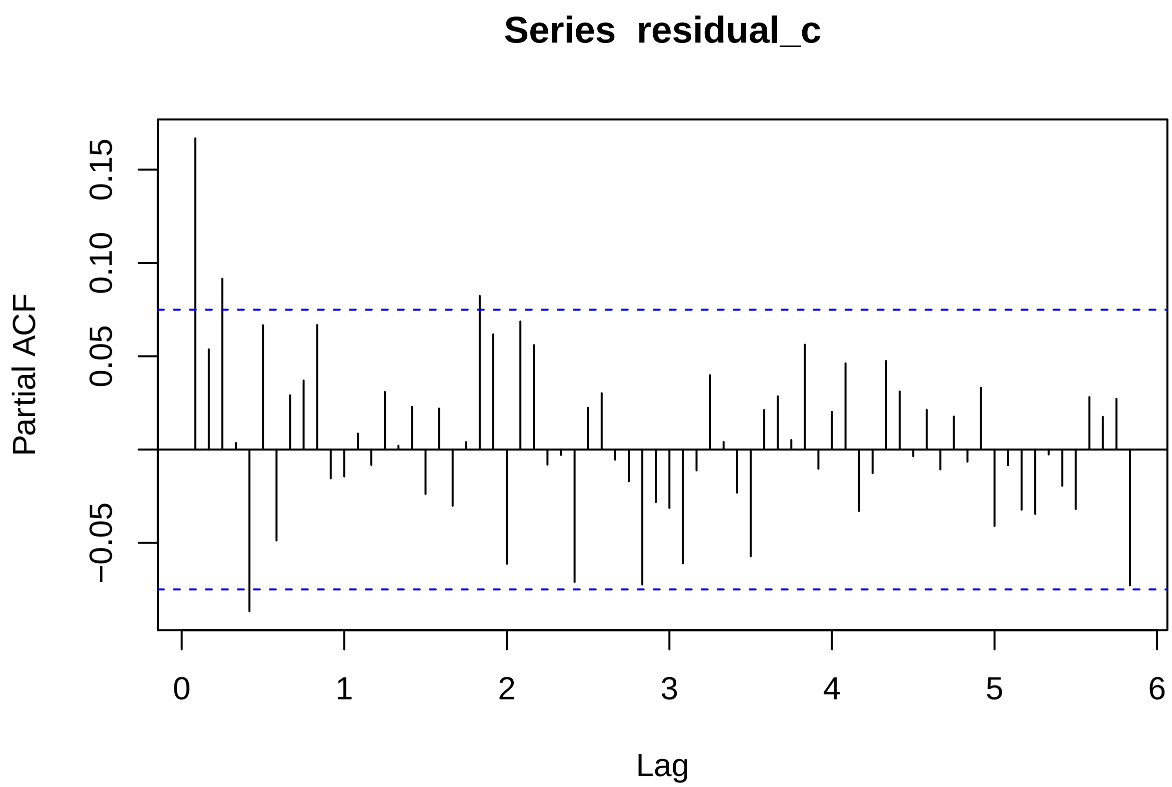
```r
residual_c = residuals(arima_function_c)
plot(residual_c)
```




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




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



The residual ACF appears to decay exponentially. The PACF shows clear spikes in the earlier lags. This is accompanied by a tapering pattern in the early lags of the ACF. Therefore, a non-seasonal AR(1) component should be added.

ii. Write down the SARIMA (1, 0, 0) × (0, 1, 1)s model equations and express Xt as an ARMA(p, q) process.

A SARIMA Model is expressed as: $\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)Z_t$

We have a SARIMA $(1,0,0)Ö(0,1,1)_{12}$ model. Here, $\phi(B) \ = 1 - \alpha B \ Phi(B^s) = 1$, and $\theta(B) = 1$ $\Theta(B^s) = 1 + \beta B^{12}$ And $W_t = \nabla^d \nabla_{12}^D X_t = \nabla_{12} X_t = X_t - X_{t-12}$

So, we have $(1 - \alpha B)(X_t - X_{t-12}) = (1 + \beta B^{12})Z_t \ X_t - X_{t-12} - \alpha X_{t-1} + \alpha X_{t-13} = Z_t + \beta Z_{t-12} \ X_t = \alpha_1 X_{t-1} + X_{t-12} - \alpha_{13} X_{t-13} + Z_t + \beta_{12} Z_{t-12}$

We have an ARMA(13,12) process.

iii. Fit the model to the data using arima() and specify each parameter's estimate.

```
arima_function_d = arima(x=training, order =c(1,0,0),  seasonal=list(order=c(0,1,1), period=12), includ
arima_function_d
```

```
##
## Call:
## arima(x = training, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1),
##     period = 12), include.mean = T)
##
```

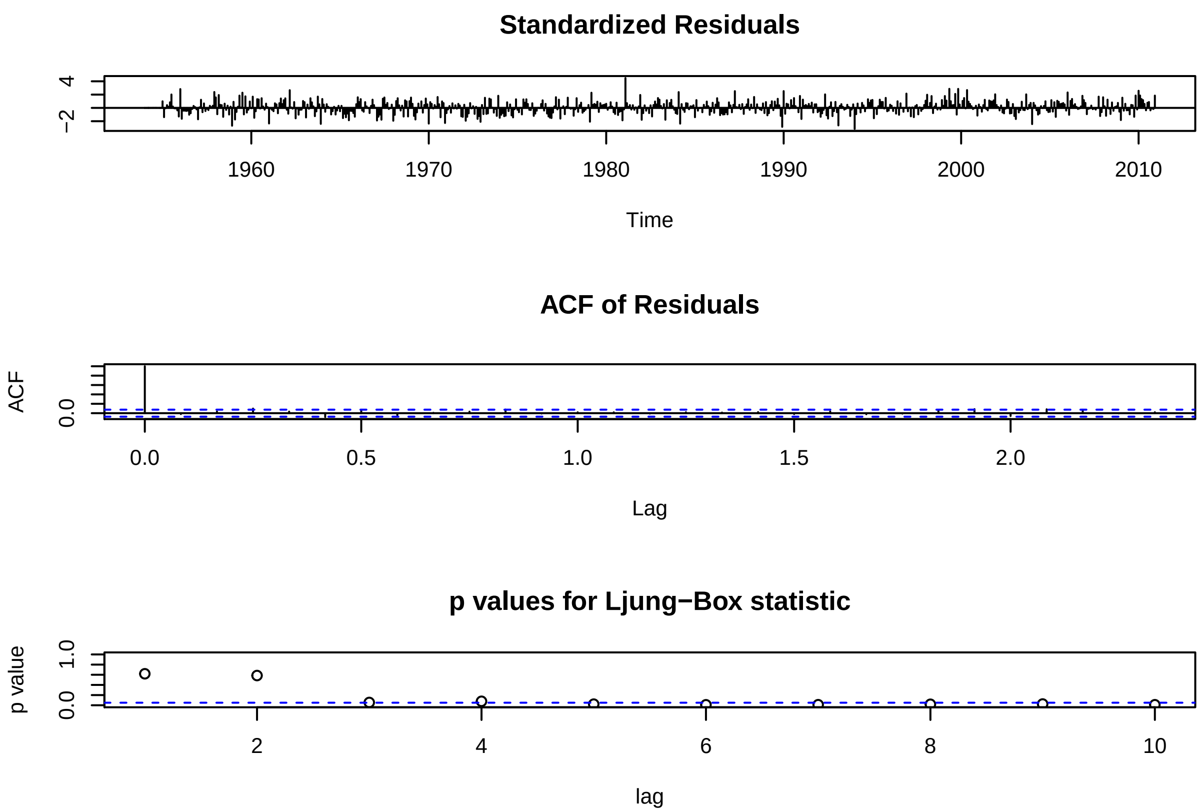
```
## Coefficients:
##          ar1     sma1
##       0.1808  -0.9354
## s.e.  0.0387   0.0190
##
## sigma^2 estimated as 3.281:  log likelihood = -1365.21,  aic = 2736.43
```

An ARMA(12,12) process should have the following skeleton: $X_t = \alpha_1 X_{t-1} + ... + \alpha_{12} X_{t-12} + Z_t + \beta_1 Z_{t-1} + ... + \beta_1 Z_{t-1} + ... + \beta_{12} Z_{t-12}$

Here, $\beta_1 = ... = \beta_{12} = 0$, $\beta_{13} = - - 0.9354$, $\alpha_{13} = 0.1808$ and Z_t~WN(0, 3.281).

iv. Using the diagnostic plots from tsdiag(), say whether this model seems like a good fit.

```
tsdiag(arima_function_d)
```



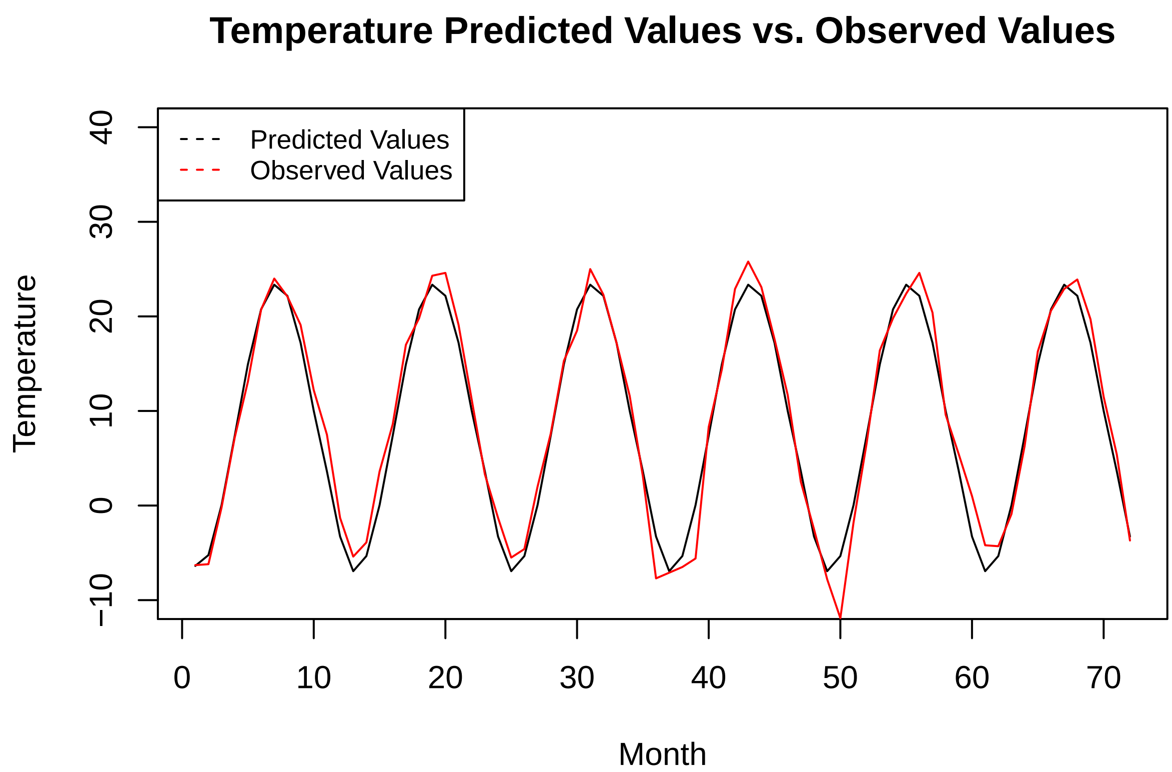
From the tsdiag, we see that only the odd residual lies outside +/- 2. The ACF looks like there is no obvious serial correlation. The p-values for the Ljung Box statistic are non-significant up to around lag 5. These suggest an acceptable fit for the model.

v. Using AIC, have we improved our model by adding the AR(1) component?

Based on AIC, the model has improved by adding the AR(1) component, as AIC decreased from 2755.96 to 2736.43.

(e) Use the predict(sarima.object, n.ahead=p) function to predict the monthly mean temperatures for 2011-2016 using the model fit from part (d). In a single plot, show both the predictions and the test set data, using a legend to distinguish between the predicted and observed values. Comment on whether the predictions are reasonable and how they differ from the test data.

```
predictions = predict(arima_function_d, n.ahead = 72)

plot(1:72, predictions$pred, col ="black", type="l", xlab = "Month", ylim = c(-10, 40), ylab = "Tempera
lines(1:72, testing, col ="red")
legend("topleft", legend = c("Predicted Values", "Observed Values"), col=c("black", "red"), lty=2, cex=
```



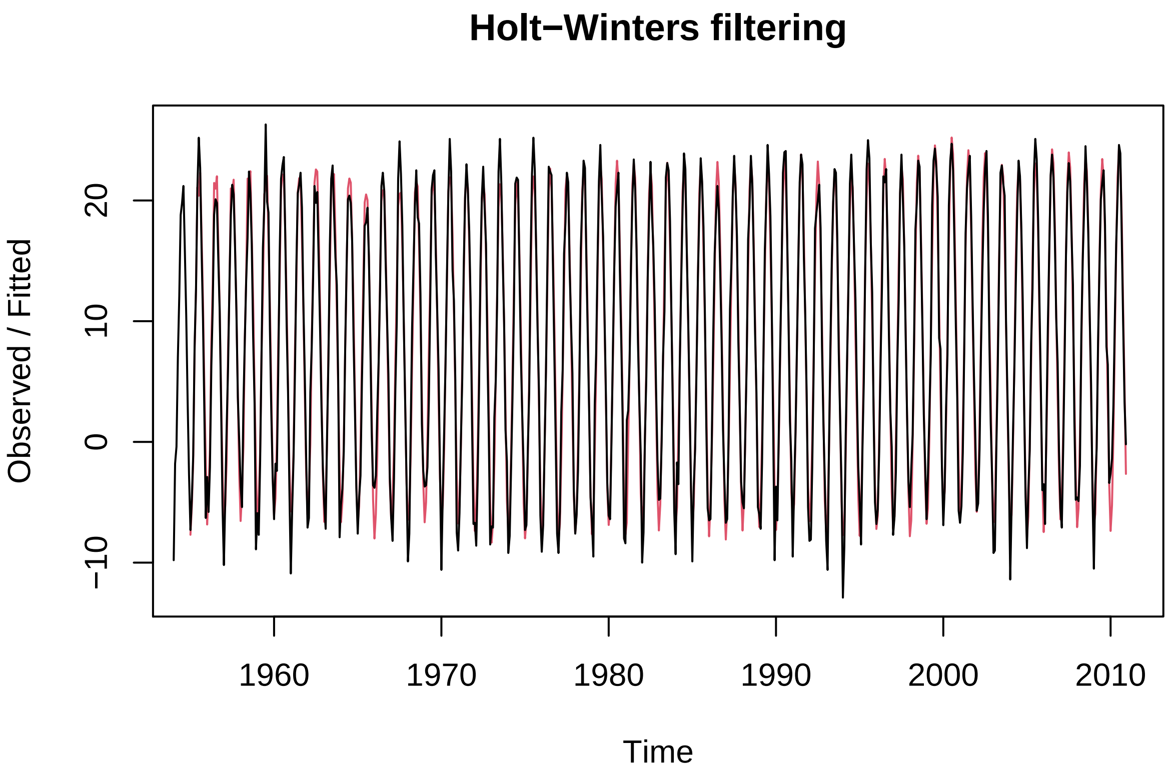
Looking at the plot of predicted vs observed values, the predictions appear to be reasonable as they look very similar to the observed values. The predictions for earlier months are better than those for later months. For example, we start to see larger deviations at Month = 50.

(f) Using the HoltWinters() function, fit a Holt-Winters model to the data.

```
holtwinters_result = HoltWinters(training)
```

i. Do the results of the Holt-Winters model indicate a trend?

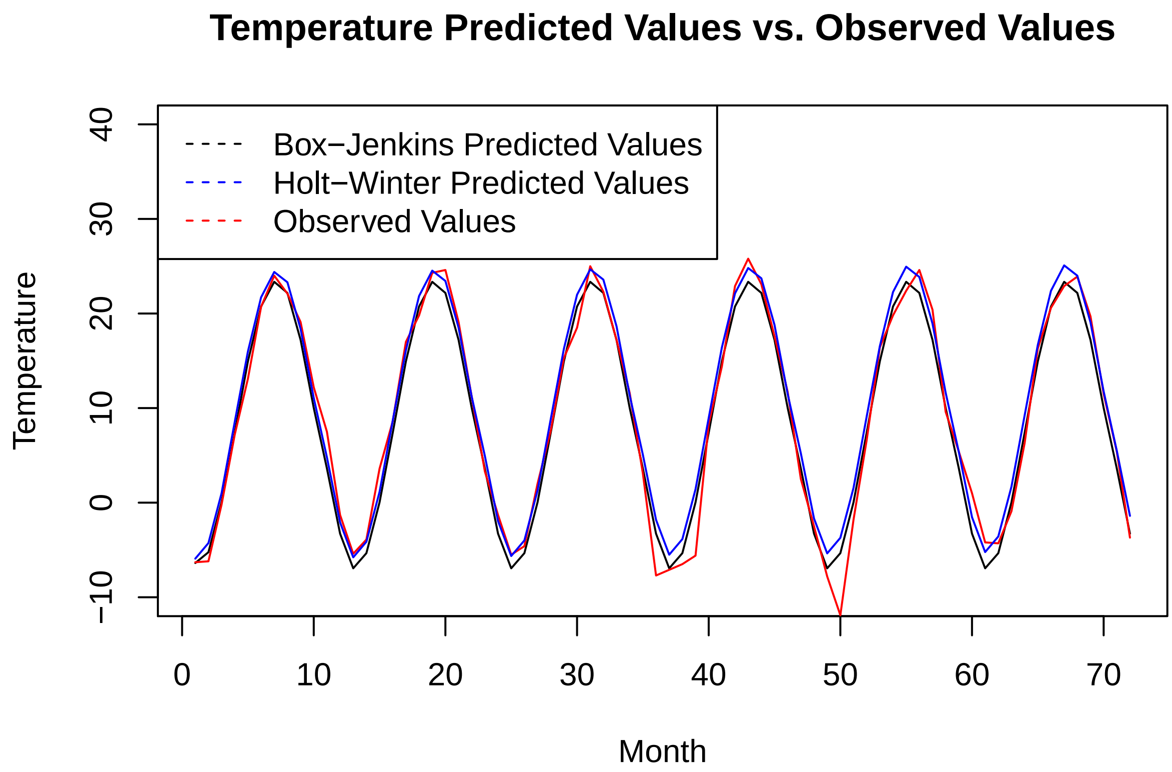
```
plot(holtwinters_result)
```



From the plot of the Holt-Winters model, we do not see a trend.

ii. Use predict() to predict the monthly mean max temperatures for 2011-2016. Plot the Holt-Winter predictions, the Box-Jenkins predictions from part (e), and the testing data in the same plot. Comment on the comparison between the two forecasting methods.

```
predictions_holt_winters = predict(holtwinters_result, n.ahead = 72)

plot(1:72, predictions$pred, col ="black", type="l", xlab = "Month", ylim = c(-10, 40), ylab = "Tempera
lines(1:72, testing, col ="red")
lines(1:72, predictions_holt_winters, col="blue")

legend("topleft", legend = c("Box-Jenkins Predicted Values", "Holt-Winter Predicted Values ", "Observed
```



Comparing the two forecasting methods, we see that Holt-Winters appears to fit the model better than the Box-Jenkins. During the earlier months, these two methods appear to fit the model quite similarly, but the Holt-Winters continues to fit the model smoothly, compared to Box-Jenkings for later months. For example, while Box-Jenkings has a large deviation at Month = 50, the Holt-Winters values are still quite accurate.

iii. Calculate the mean squared prediction error (MSPE) for the Box-Jenkins model and the Holt-Winters model. Which method performs better?

```
MSPE_Box_Jenkins = mean((testing - predictions$pred)^2)
MSPE_Holt_Winters = mean((testing - predictions_holt_winters)^2)

MSPE_Box_Jenkins
```

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

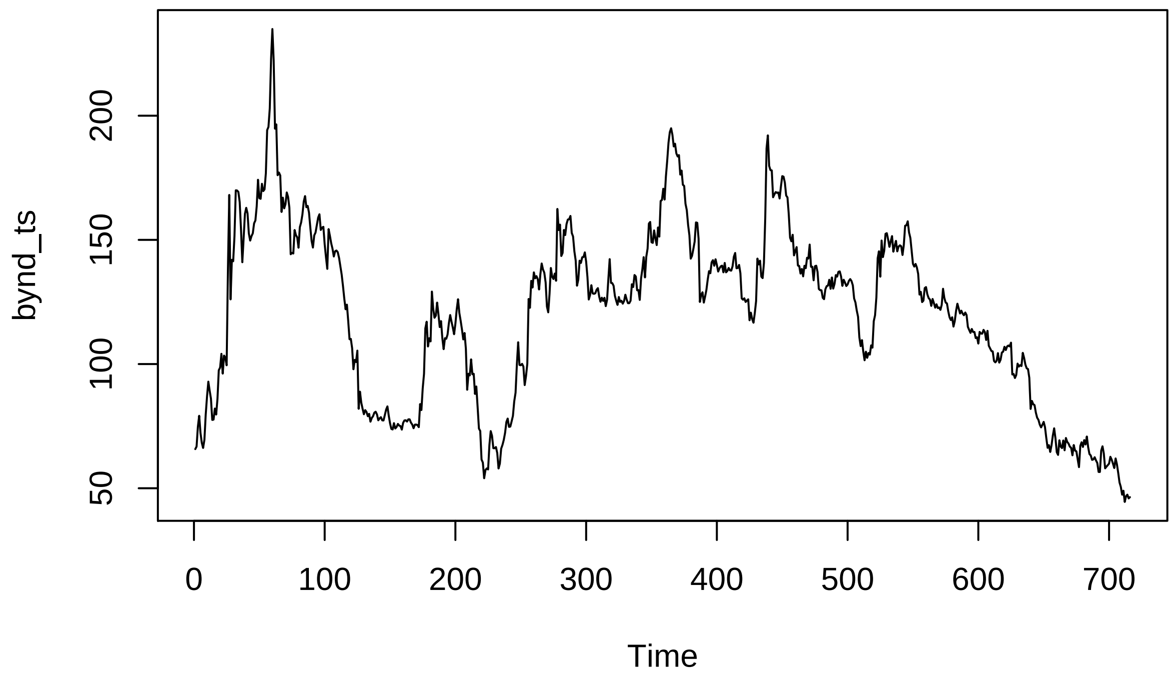
```
MSPE_Holt_Winters
```

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

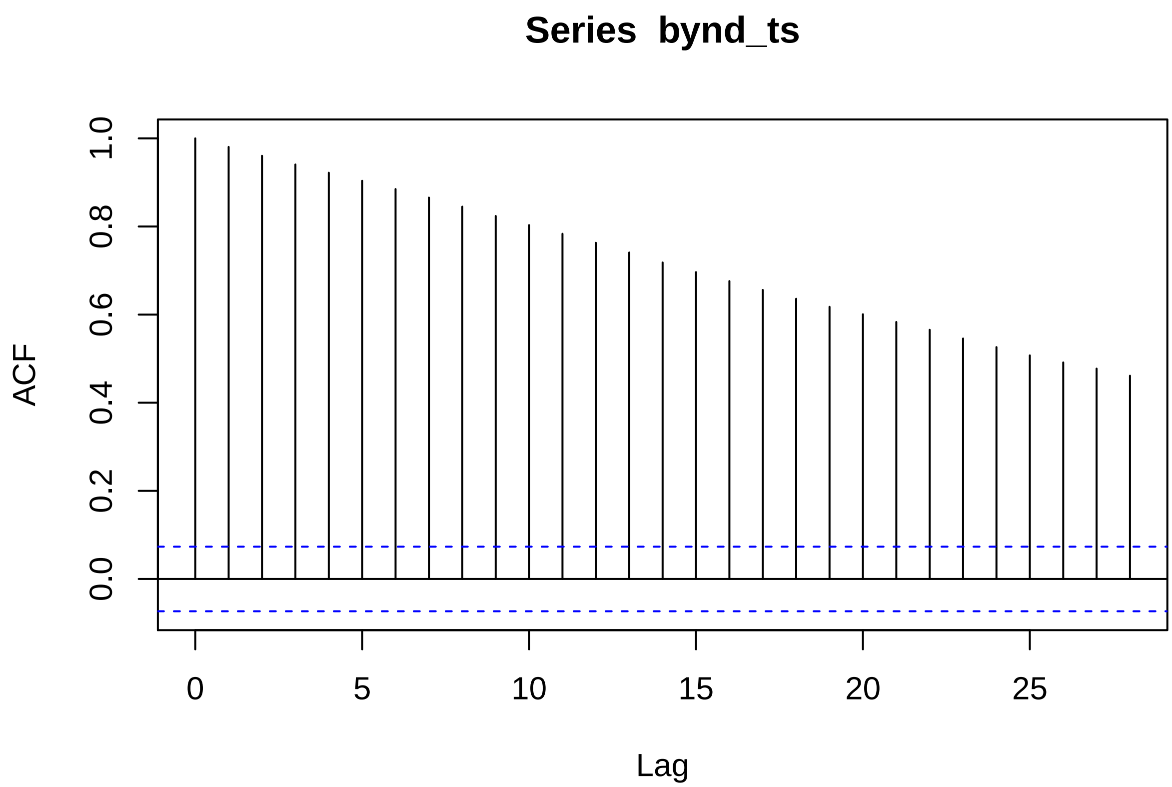
The Box Jenkins method performs better because it has a lower MSPE compared to the Holt Winters prediction.

Question 2: Read the data set into R, and coerce the data into a time series object. Create a plot of the data, and plot the acf and pacf of the series. Comment on what you observe.

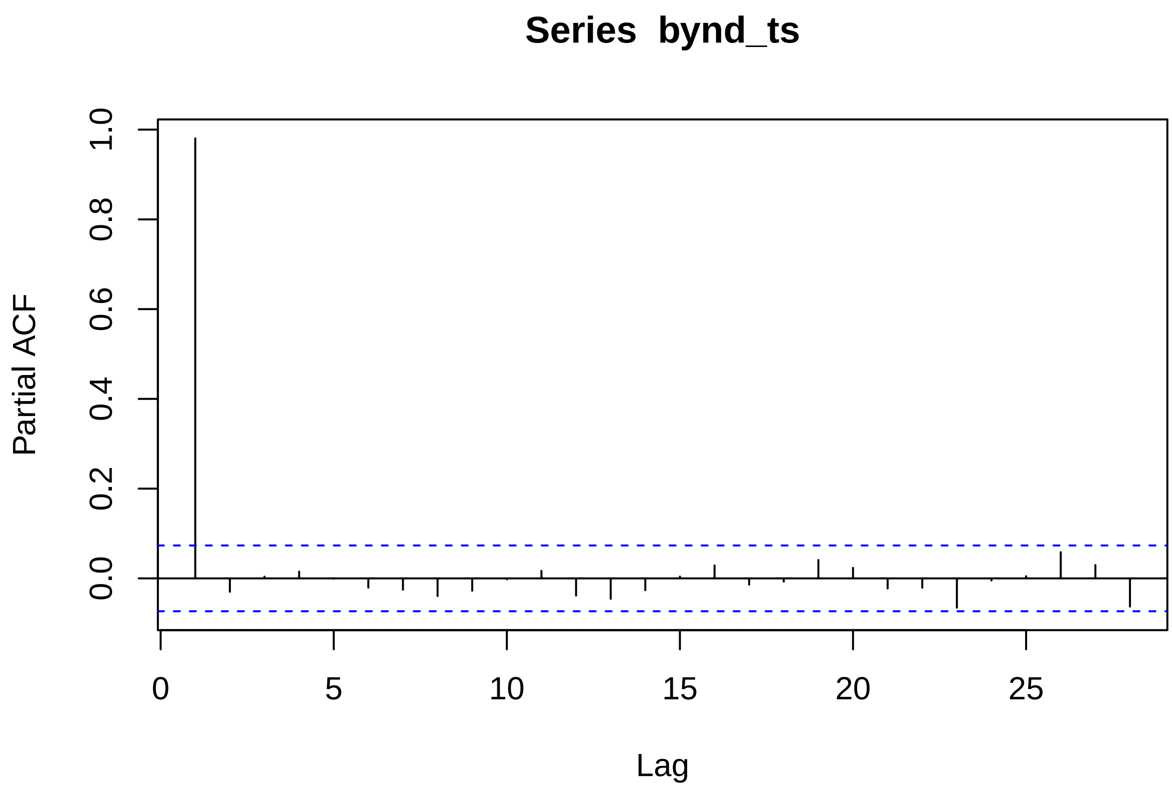
```r
library(zoo)
x = read.table("bynd.txt")
xx = zoo(x$V2, x$V1)
bynd_ts = ts(xx)

plot(bynd_ts)
```




```r
acf(bynd_ts)
```



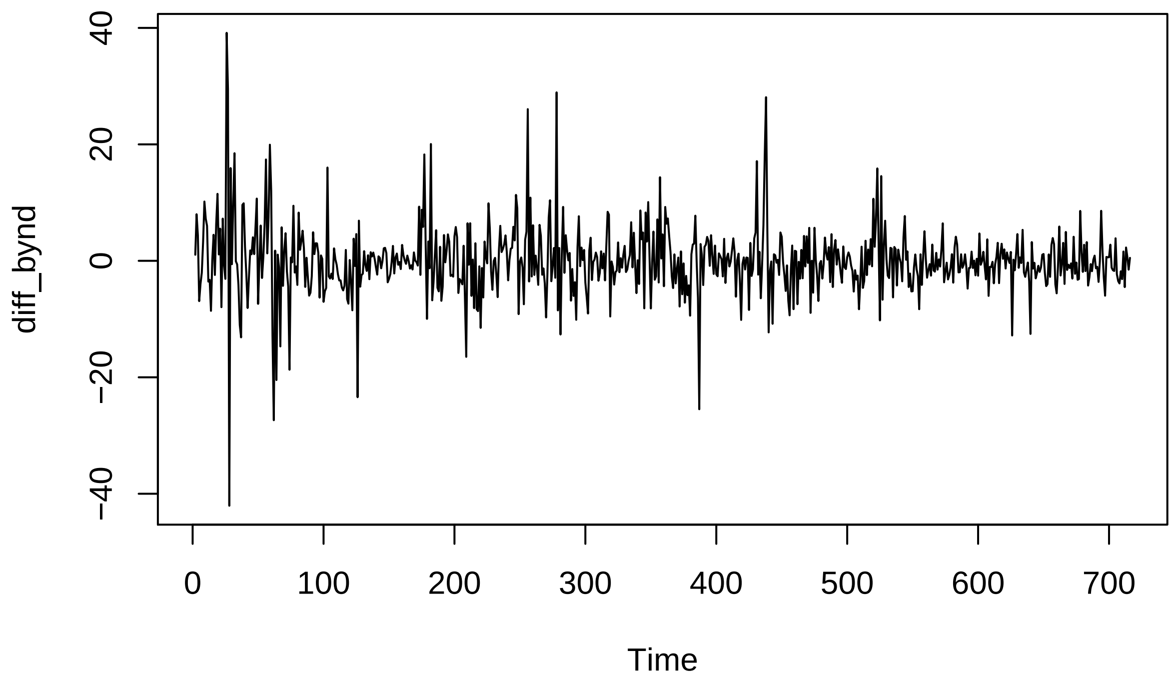

```
pacf(bynd_ts)
```



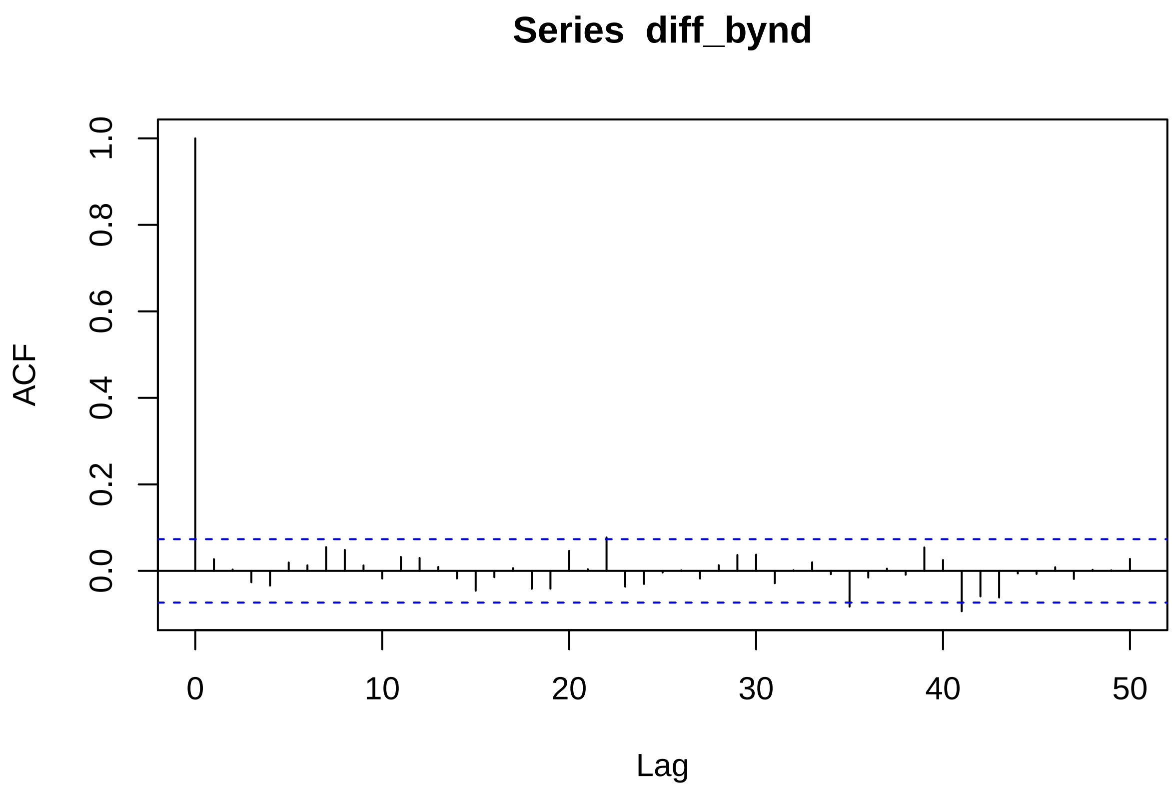
Looking at the plot of the data, we can see that there are large fluctuations in the daily closing price over time. The ACF appears to decay very slowly, which may be indicative of the series having a trend. The PACF cuts off at lag 1.

(b) Create a new series by differencing consecutive values in the time series. Plot the acf and pacf of the new series. Does the first-differenced series appear stationary?

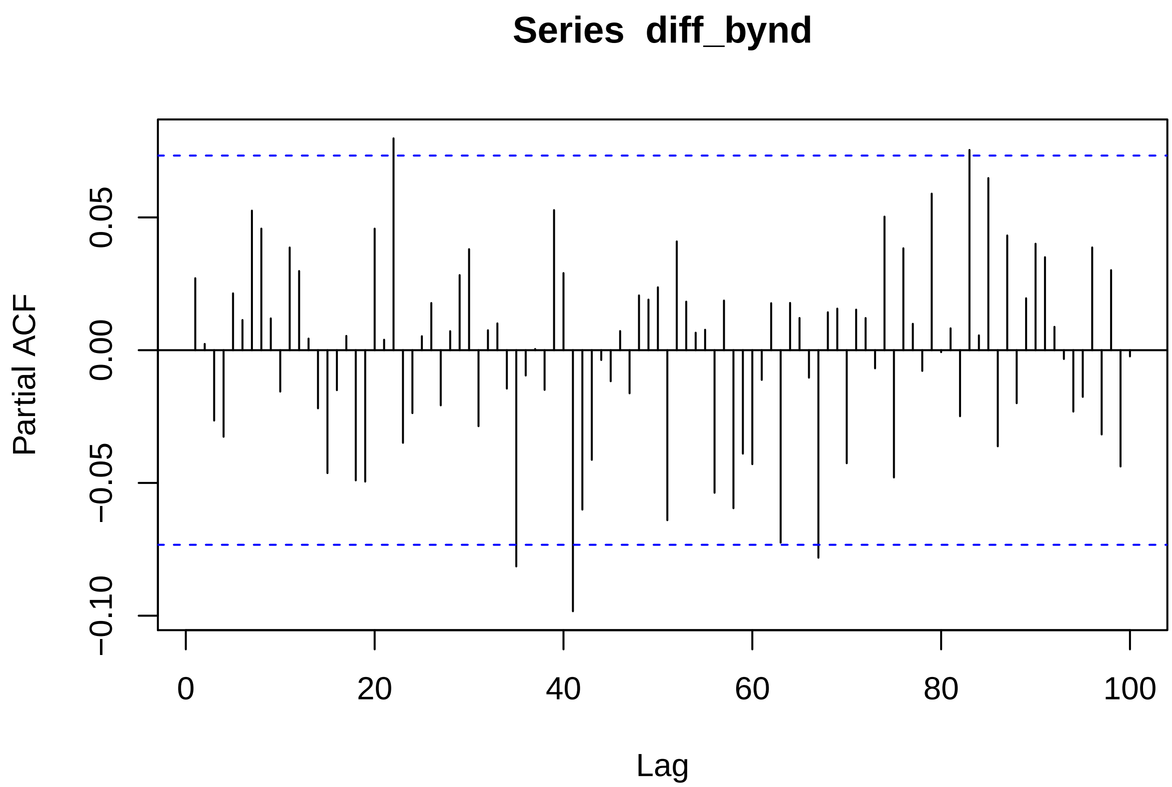
```
diff_bynd = diff(bynd_ts)
plot(diff_bynd)
```




```
acf(diff_bynd, lag.max = 50)
```




```
pacf(diff_bynd, lag.max = 100)
```



The plot of the new series appears to be trendless. The ACF function shows a significant value at lag= 0, and the PACF appears to have some significant values at lags 22, and 41. Yes, the first differenced series appears stationary.

c) A model of the form $X_t = X_{t-1} + Z_t$ would be suitable for share price series, as the daily price seems to depend on the previous day's price.

d) Assuming we have the model, $X_t = X_{t-1} + Z_t$, this would be an ARIMA(1,1,0) model.

e)

```
bynd_arima = arima(x=bynd_ts, order =c(1,1,0), include.mean = T)
bynd_arima
```

```
##
## Call:
## arima(x = bynd_ts, order = c(1, 1, 0), include.mean = T)
##
## Coefficients:
##          ar1
##       0.0271
## s.e.  0.0374
##
## sigma^2 estimated as 35.96:  log likelihood = -2295.21,  aic = 4594.42
```

Here, $\sigma^2 = 35.96$.

f) Assuming the future values of Z_t will be zero, the forecast at time l would be:

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

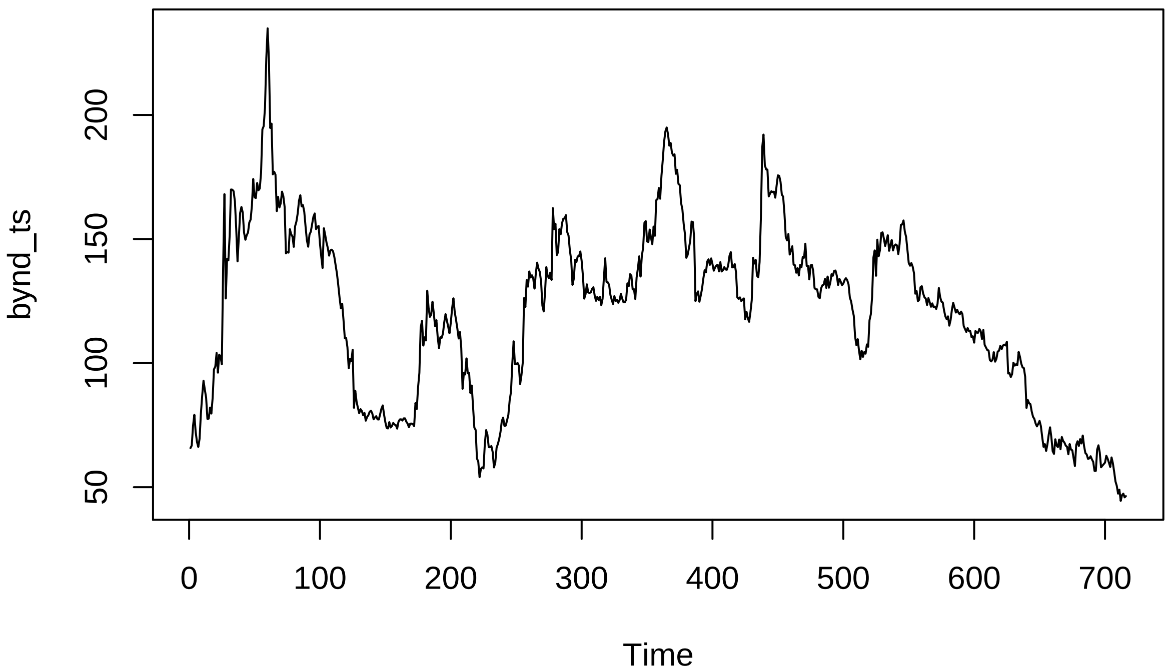
g) Lab - sigma squared variance for the differenced model use function sd apply on the differenced series

```
pred <- predict(bynd_arima, n.ahead = 7, prediction.interval = T, level = 0.90)
pred
```

```
## $pred
## Time Series:
## Start = 717
## End = 723
## Frequency = 1
## [1] 46.42382 46.42419 46.42420 46.42420 46.42420 46.42420 46.42420
##
## $se
## Time Series:
## Start = 717
## End = 723
## Frequency = 1
## [1]  5.996332  8.595732 10.576929 12.241639 13.705622 15.027655 16.242437
```

h)

```
plot(bynd_ts)
```



We consider three options; sell, hold, or do nothing. If we predict prices will increase, we should hold. If we predict prices will decrease, then we should sell now. From the predictions, we see that the numbers just fluctuations, meaning there is no benefit from either selling the shares, or holding them. Therefore, I would advise to just do nothing.