

STAT 443 Assignment 3

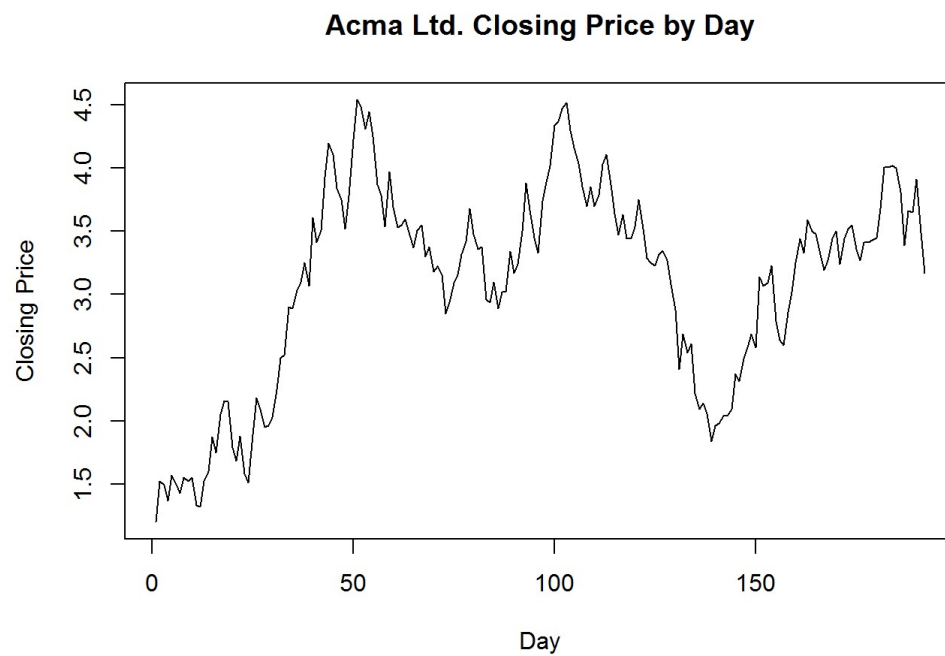
David Yin (13922159)

16 March, 2017

Question 1

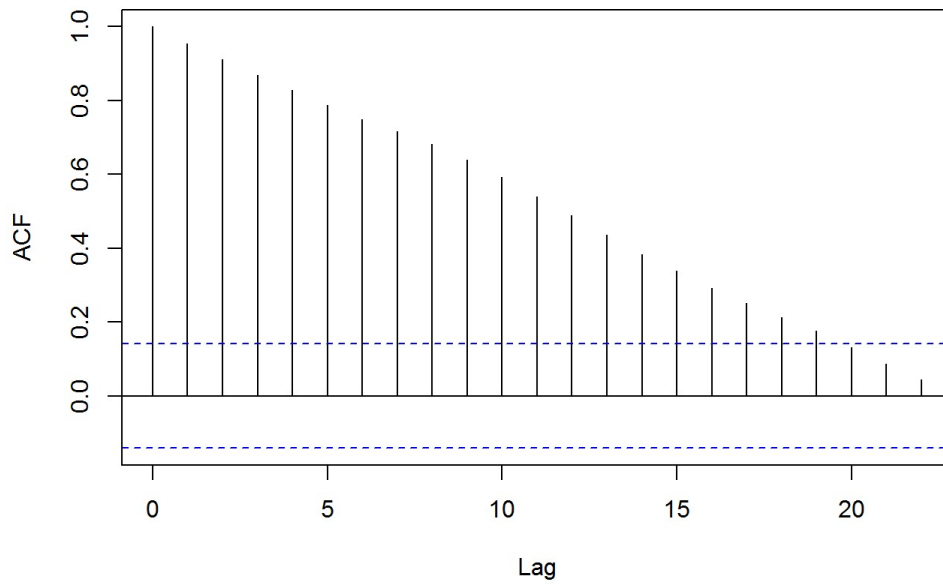
a)

```
acma = read.table("Acma.TXT")
acma <- ts(acma)
plot(acma, xlab = "Day", ylab = "Closing Price", main = "Acma Ltd. Closing Price by Day")
```



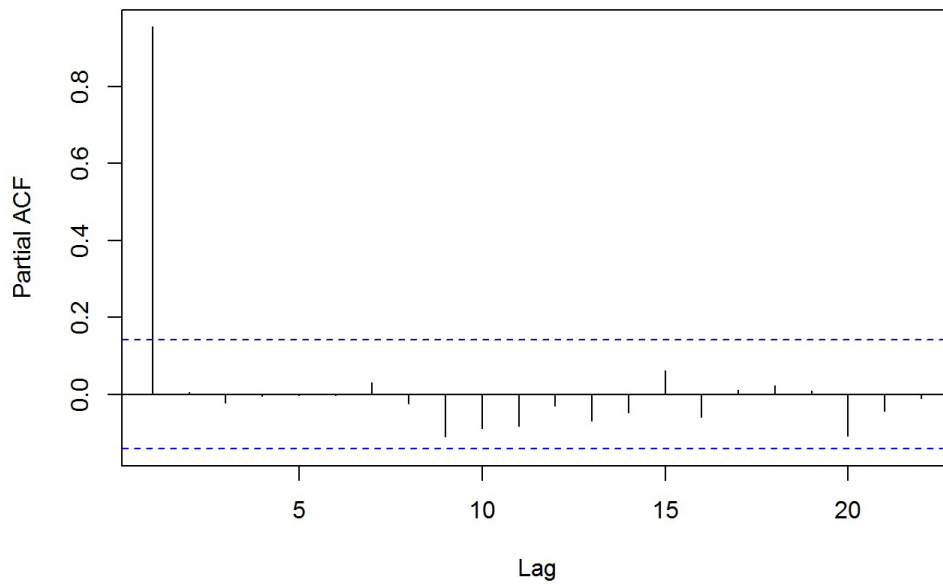
```
acf(acma, main = "Autocorrelation")
```

Autocorrelation



```
pacf(acma, main = "Partial Autocorrelation")
```

Partial Autocorrelation

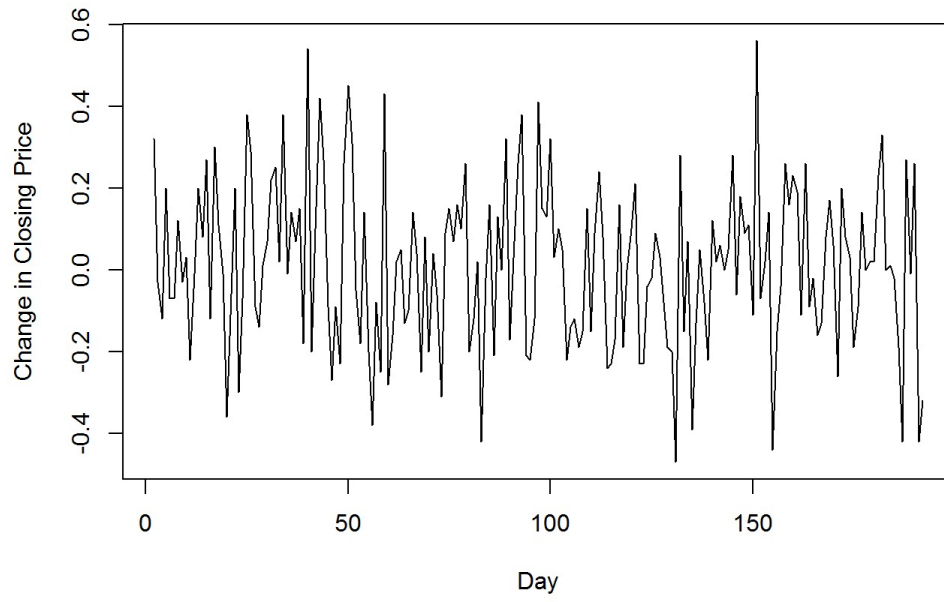


The acf shows a slow decay, reflective of a typical AR model. The pacf shows a sharp cut off at lag 1, suggesting an AR(1) model.

b)

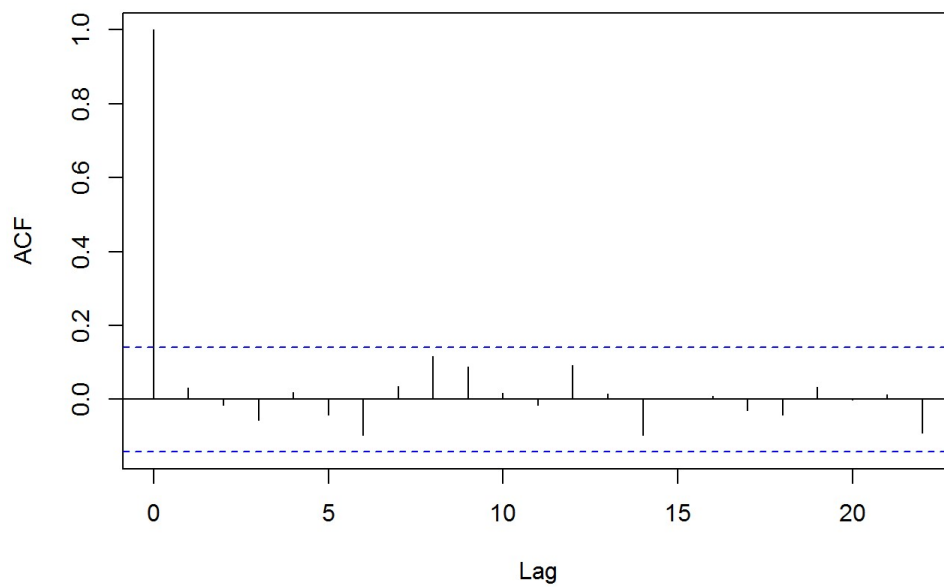
```
acma_d = diff(acma)
plot(acma_d, xlab = "Day", ylab = "Change in Closing Price", main = "First-order differencing of ACMA")
```

First-order differencing of ACMA



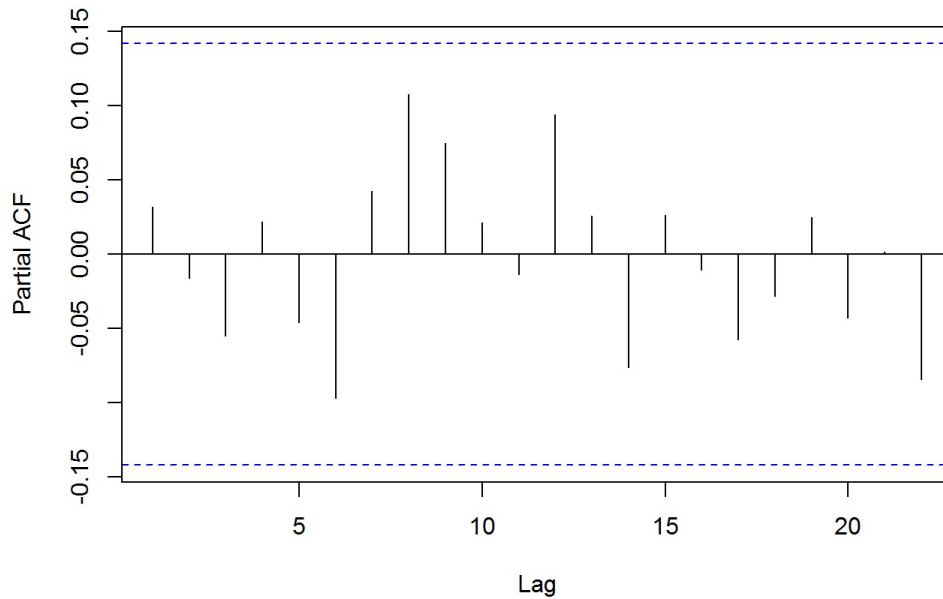
```
acf(acma_d, main = "Autocorrelation")
```

Autocorrelation



```
pacf(acma_d, main = "Partial Autocorrelation")
```

Partial Autocorrelation



It appears stationary. No obvious trend, and acf shows cut off at lag 0 and no pattern in pacf. It resembles white noise.

c)

It is an AR(1) model, and a random walk specifically. We expected the first order difference to be white noise, so $X_t - X_{t-1} = Z_t$ where Z_t is white noise. So this matched our expectation of a AR(1) model from part a).

d)

```
mod = arima(acma, order = c(1,0,0))
mod$sigma2
```

```
## [1] 0.04065574
```

From running our model on the dataset, we obtain that the variance of Z_t is 0.04065574.

e)

$$X_{n+l} = X_{n+l-1} + Z_{n+l} = X_{n+l-2} + Z_{n+l-1} + Z_{n+l} = \dots = X_n + Z_{n+1} + \dots + Z_{n+l}$$

Because the expected values of the Z_{n+i} are 0, our best pick will be X_n , the closest actual observation we have to X_{n+l} . So it would be $X_{192} = 3.17$.

f)

$$e_n(l) = X_{n+l} - \hat{X}_n(l)$$

$$e_n(l) = Z_{n+l} + Z_{n+l-1} + \dots + Z_{n+1}$$

The variance of the error is the sum of the variances of the white noises (covariances of independent white noises are 0), so

$$\text{Var}(e_n(l)) = l * \sigma^2$$

Therefore, the 90% prediction interval is:

$$X_n \pm \Phi^{-1}(0.95)\sigma\sqrt{l}$$

Since our last observed value is:

```
acma[length(acma)]
```

```
## [1] 3.17
```

And our $\Phi^{-1}(0.95)$ is:

```
qnorm(0.95)
```

```
## [1] 1.644854
```

And our σ^2 is 0.04066, we have our 90% prediction interval to be:

$$3.17 \pm 0.33\sqrt{l}$$

g)

As explained earlier, we know that our forecast for closing price at lead time l is always the last observed value of 3.17, so there would be no difference in the expected share price as time moves forward. So there is no advantage in selling or holding the shares at any point in time in the future. However, if he were to lock the shares for a period of time and then selling it, the longer he waits to sell the shares, the more risky it is, as we have shown that the prediction interval depends on l , and the prediction error increases as l increases. Thus, if he is not so much of a risk taker, it might be wise to not hold his shares for too long before selling them.

Question 2

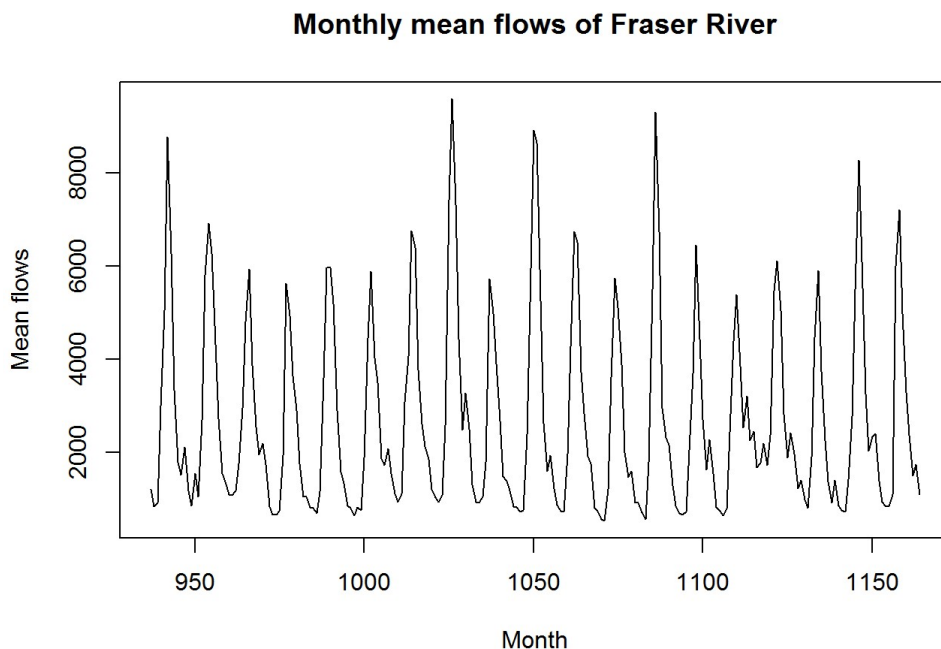
Reorganizing the data into a single time-series column:

```
dfr = read.csv("dataFraserRiver.csv")
dfr <- dfr[-1]
dat = c()
for (i in 1:99) {
  dat <- c(dat,dfr[i,])
}
dat <- as.numeric(dat)
dat <- data.frame(dat)
dat <- ts(dat)
```

Making the training and test data:

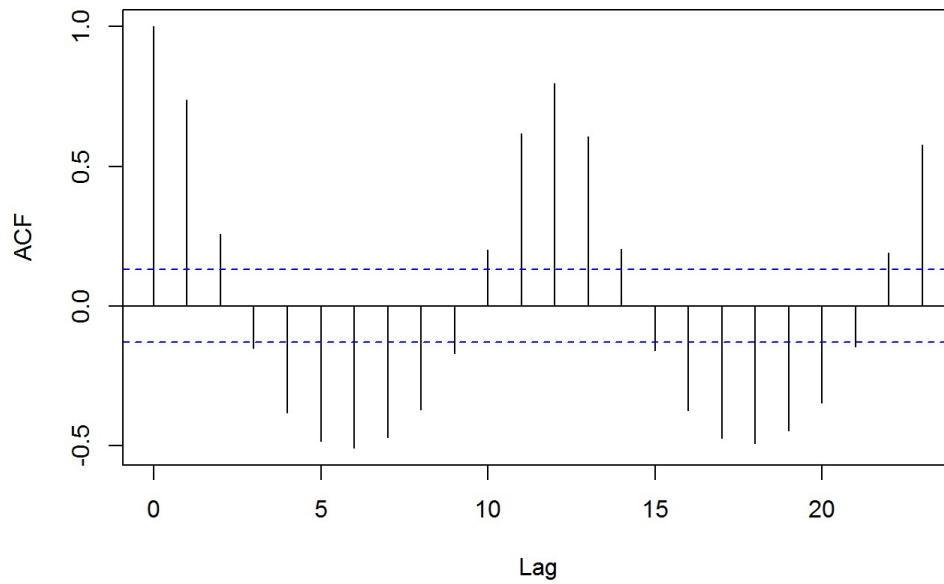
```
flow = window(dat, start = 78*12+1, end = 97*12)
flow.test = window(dat, start = 97*12+1)
```

```
plot(flow, xlab = "Month", ylab = "Mean flows", main = "Monthly mean flows of Fraser River")
```



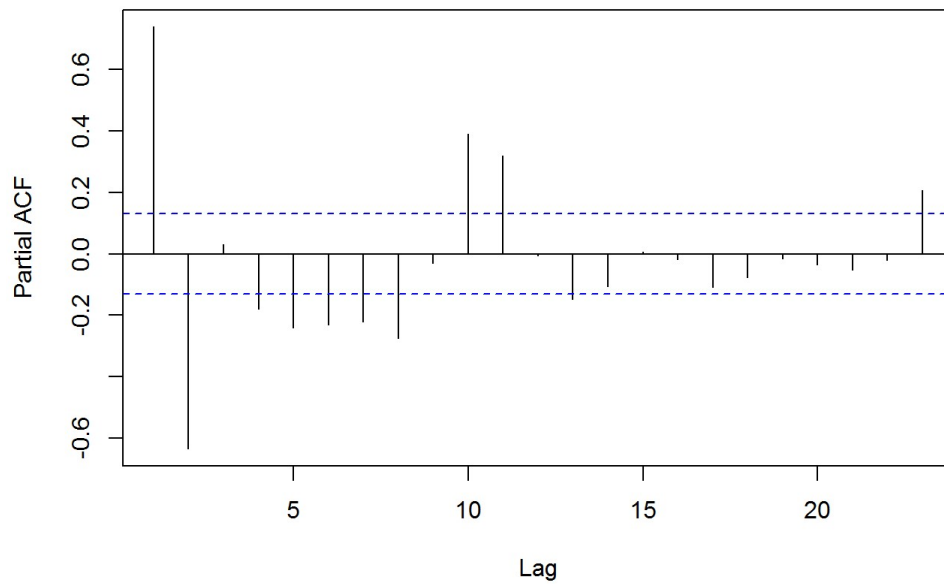
```
acf(flow, main = "Autocorrelation")
```

Autocorrelation



```
pacf(flow, main = "Partial Autocorrelation")
```

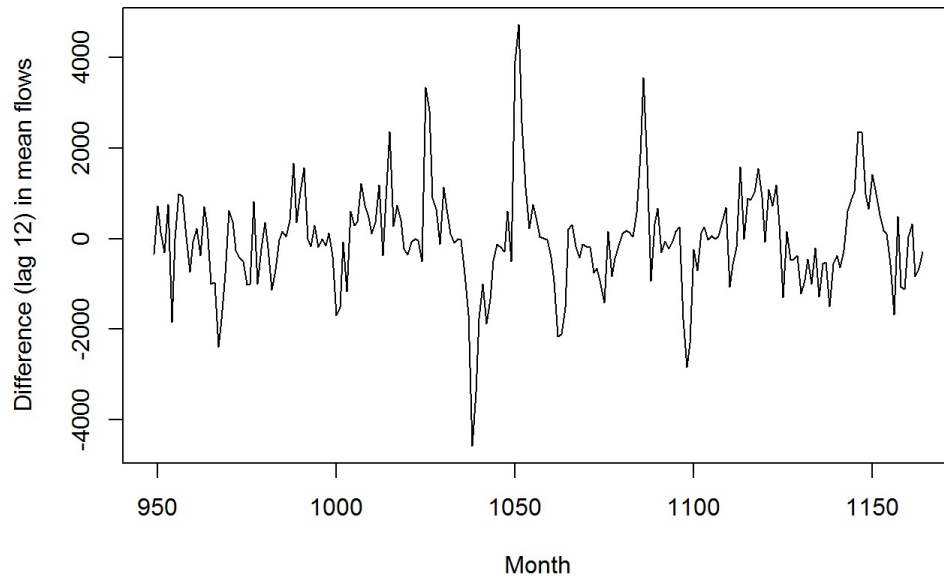
Partial Autocorrelation



From the training data, we can see that the plots show high effects in seasonality with lag 12, which suggests the use of first-order differencing.

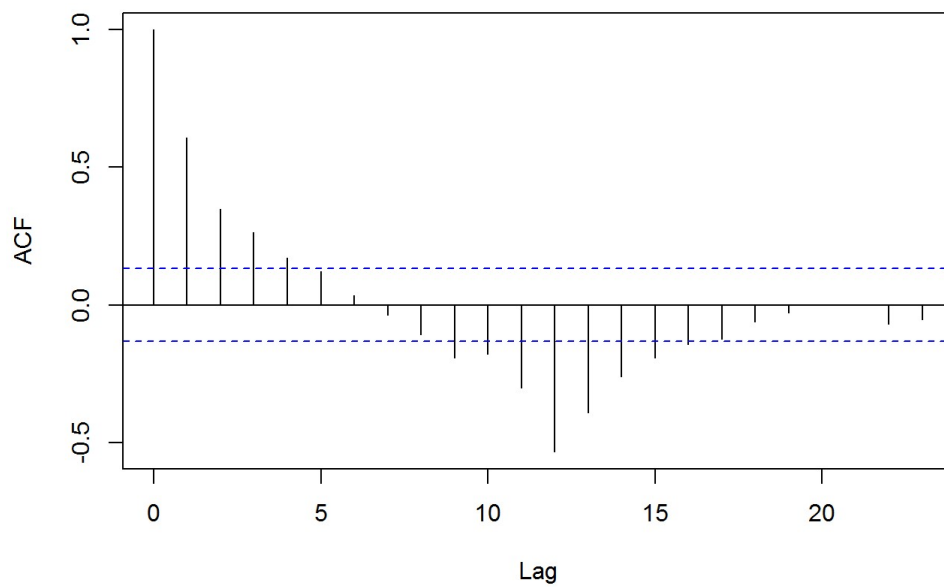
```
flow_d = diff(flow, lag = 12)
plot(flow_d, xlab = "Month", ylab = "Difference (lag 12) in mean flows", main = "Differenced monthly mean flows (without seasonal effect)")
```

Differenced monthly mean flows (without seasonal effect)



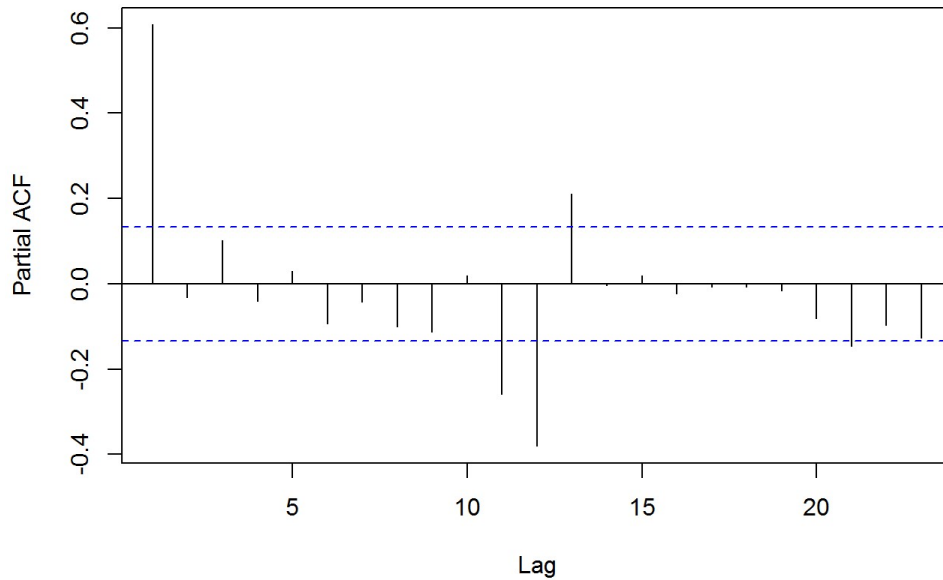
```
acf(flow_d, main = "Autocorrelation")
```

Autocorrelation



```
pacf(flow_d, main = "Partial Autocorrelation")
```

Partial Autocorrelation



From the plots, we can see that the data looks much more stationary, although there are still outliers in the middle of ACF and partial ACF plots. The PACF cut off at lag 1 and the exponential decay in the ACF suggests a P=1 model. Thus, we can fit a model that involve P=1, D=1.

```
mod1 = arima(flow, order = c(0,0,0), seasonal = list(order=c(1,1,0), period = 12))
mod1
```

```
##
## Call:
## arima(x = flow, order = c(0, 0, 0), seasonal = list(order = c(1, 1, 0), period = 12))
##
## Coefficients:
##          sar1
##        -0.5332
## s.e.      0.0558
##
## sigma^2 estimated as 885068:  log likelihood = -1787.39,  aic = 3578.77
```

I noticed that a similar model gives a smaller AIC:

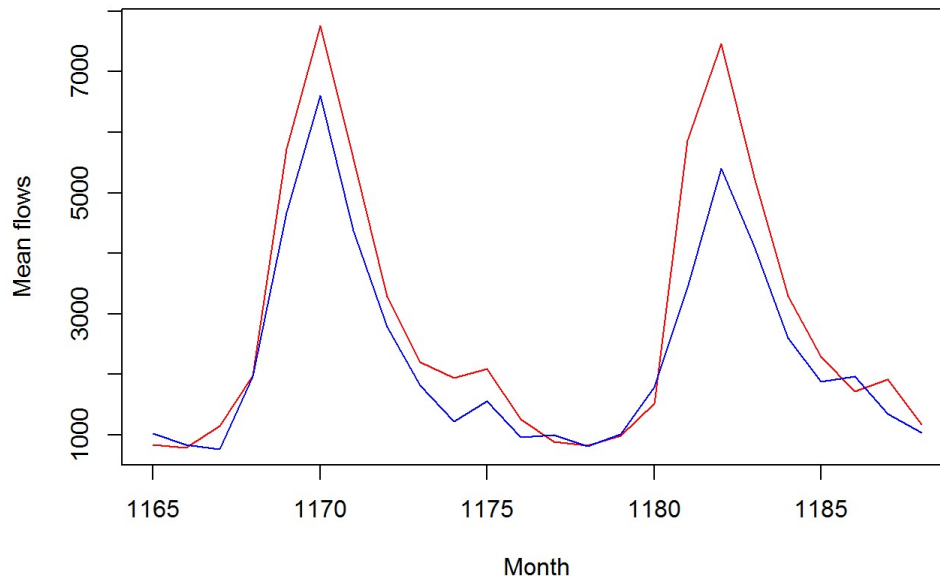
```
mod2 = arima(flow, order = c(1,0,0), seasonal = list(order=c(1,1,0), period = 12))
mod2
```

```
##
## Call:
## arima(x = flow, order = c(1, 0, 0), seasonal = list(order = c(1, 1, 0), period = 12))
##
## Coefficients:
##          ar1          sar1
##         0.5851      -0.5222
## s.e.    0.0548      0.0585
##
## sigma^2 estimated as 579981:  log likelihood = -1741.85,  aic = 3489.7
```

Thus, I will choose my model based on SARIMA(1,0,0)(1,1,0) and do a prediction of the next 24 months using that.


```
mod = mod2
pred = predict(mod, 24)
plot(pred$pred, col = "red", main = "Predicted and actual mean flows for the next 24 months",
      xlab = "Month", ylab = "Mean flows")
lines(flow.test, col = "blue")
```

Predicted and actual mean flows for the next 24 months



The blue line represents the actual observations and the red line represents the predictions. They look very similar in terms of shape, and the predictions seem to be slightly higher than the actual observations.

We can also look at the root mean squared error:

```
sqrt(mean(pred$pred - flow.test)^2)
```

```
## [1] 534.6841
```

The RMSE is 535, and given that the data is between 1000 to 7000, this error is thought to be reasonably small.

We can also look at the 95% prediction intervals and see how many of the actual observations lie in it:

```
sum((flow.test > pred$pred-qnorm(0.975)*pred$se) & (flow.test < pred$pred+qnorm(0.975)*pred$se))
```

```
## [1] 22
```

Since 22 out of 24 of the observed values lie in the 95% prediction interval, the predictions are thought to be appropriate.