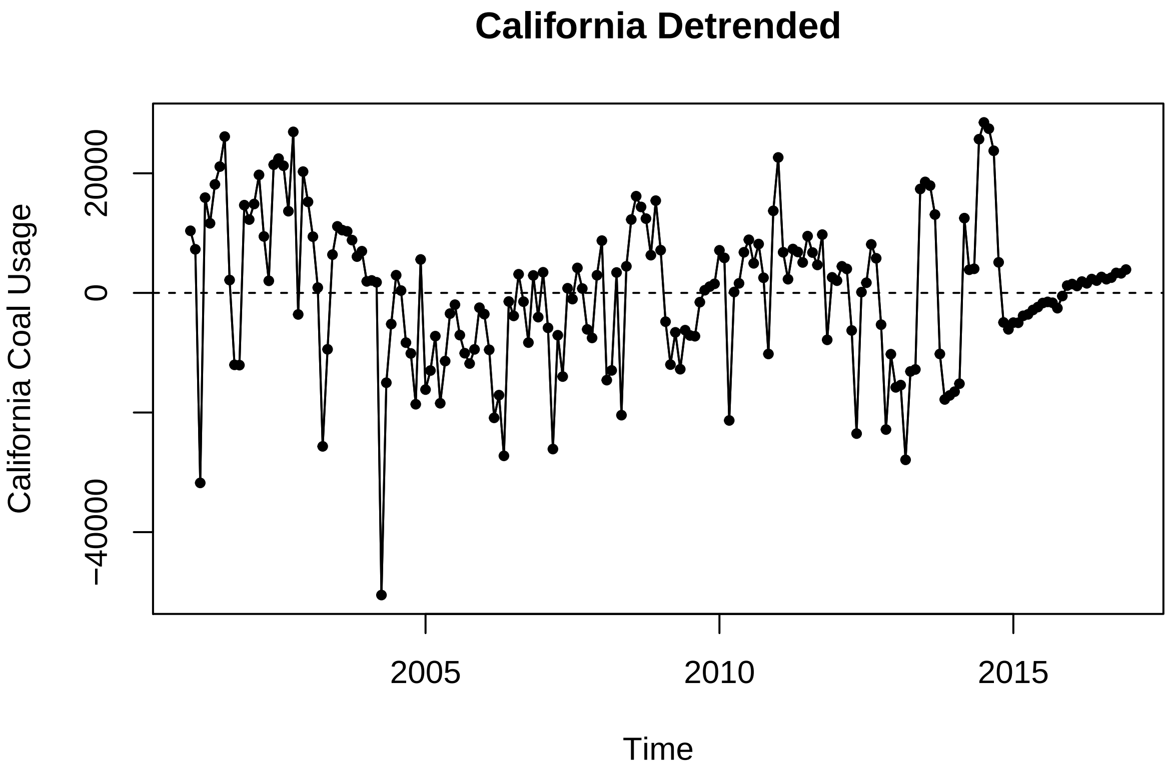
MATH 545 - Final Project

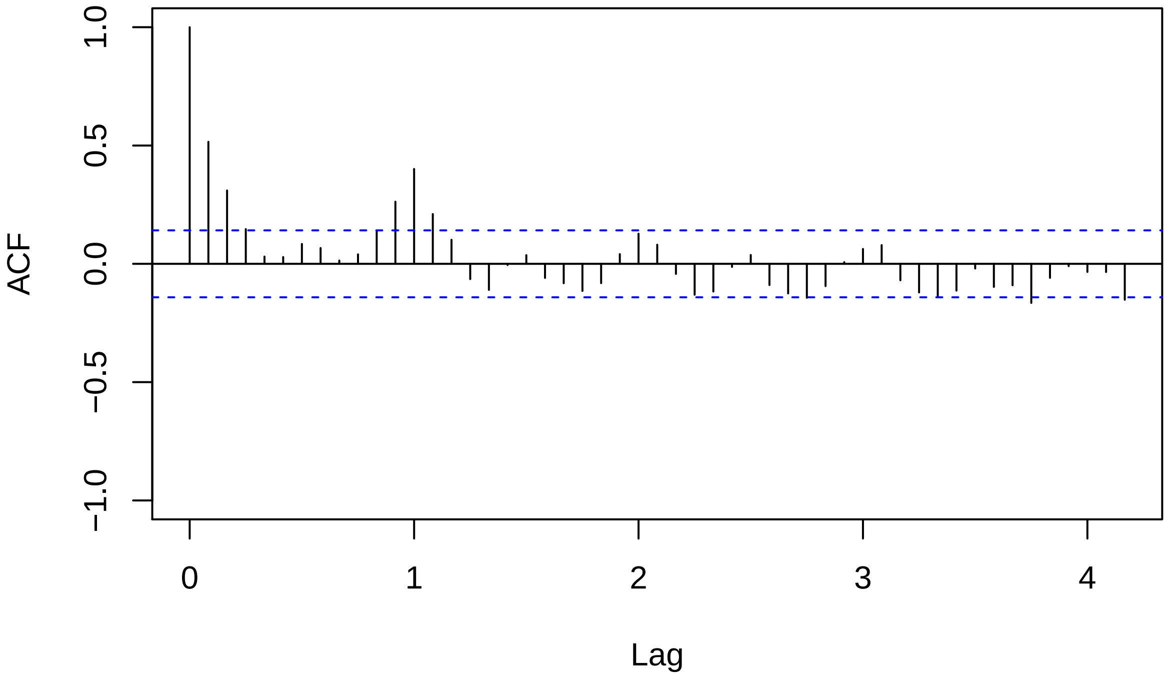
Lea Collin, 260618407

December 20, 2017

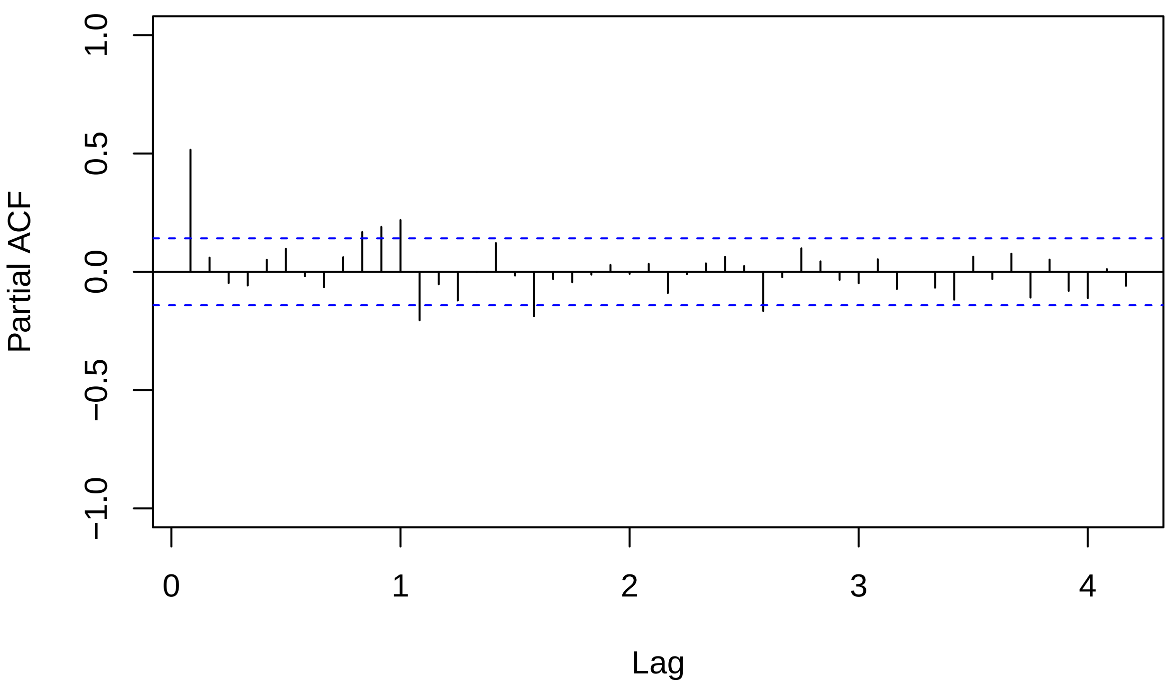
We begin our analysis with California (included in the appendix) and start by plotting the data. There is quite an interesting change in the mean for California; it's kind of flat at first, then drops off quite significantly, and then almost seems to flat line by the beginning of 2015. This mean is obviously not constant and we will actually try to fit a sigmoidal function to imitate this changing mean. The detrended data is plotted below.



This data is still not perfectly zero mean nor does it have an obviously constant variance, but it's not bad.



This acf actually seems fairly stationary. There is a somewhat large spike around lag 12 which we may handle by fitting a SARIMA model and setting Q equal to 1. Now we will look at the pacf.

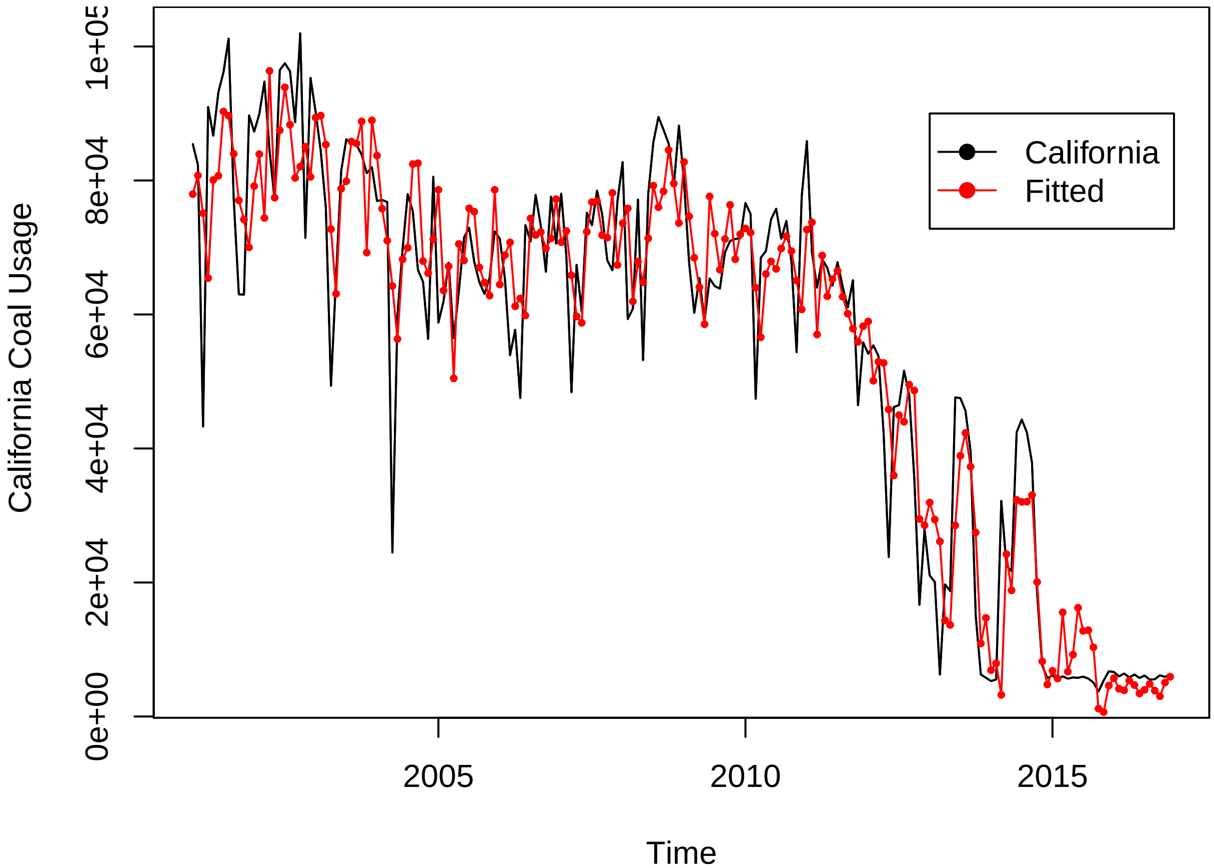


According to the pacf, P is probably 0 or 1 and the p values are quite small. We will now fit several SARIMA models with d = 0, D = 0, P = 0 or 1, Q = 1 and period 12 and then further inspect the mean squared error of a few models based on their AIC and BIC values. This code is provided in the appendix.

Because we tried out several combinations of P and Q values, we have several models to look at in more detail. The models for which we calculate the mean squared error are the following: $SARIMA(3,0,2)X(0,0,1)_{12}$, $SARIMA(1,0,0)X(0,0,1)_{12}$, $SARIMA(3,0,2)X(1,0,1)_{12}$, and $SARIMA(1,0,0)X(1,0,1)_{12}$. This code is provided in the appendix.

Based on the mean squared error and AIC/BIC values, the best model we can fit to this data is a

$SARIMA(3,0,2)X(1,0,1)_{12}$ model. A plot of this fitted model and the original data is shown below.

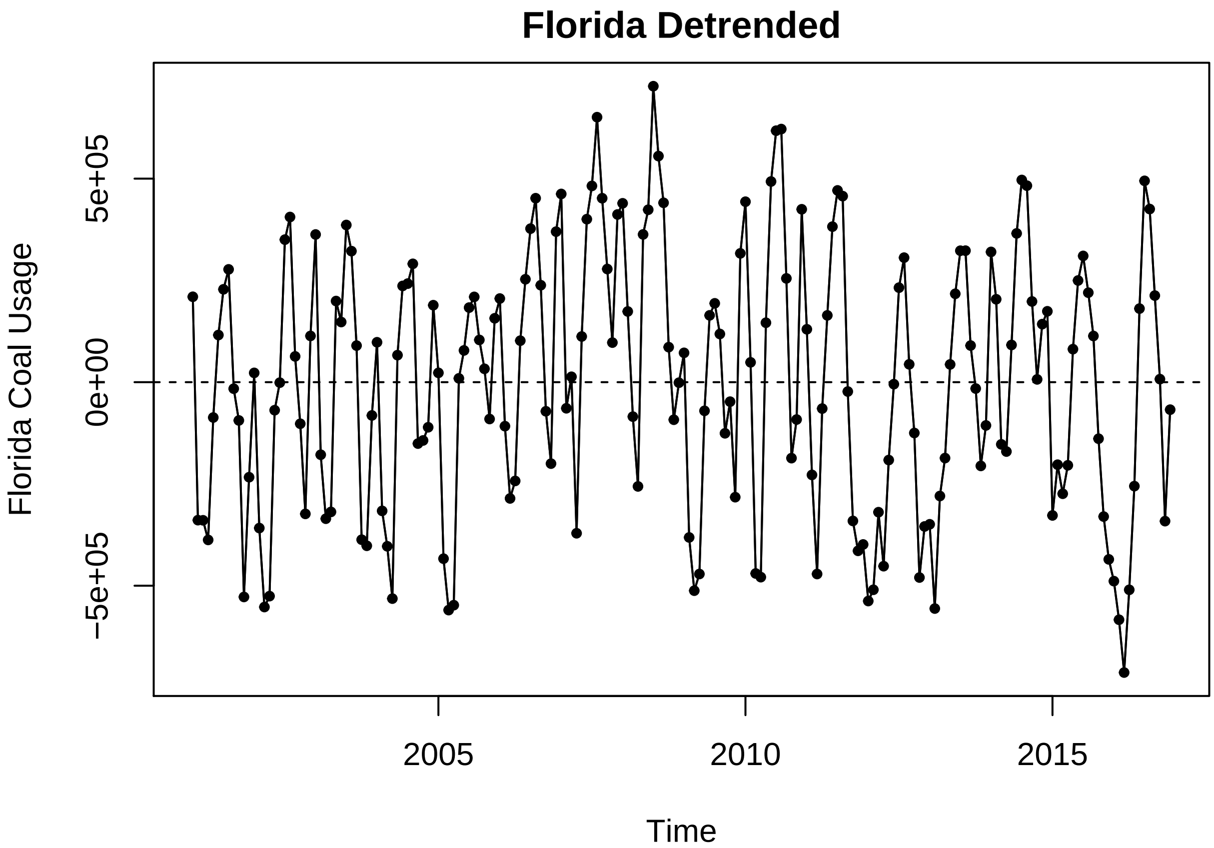


Though the fitted model does not do a perfect job of fitting the data, it does do a decent job of catching the rise and fall of the data for each month and over the years. The fitted values also seem to fluctuate around a constant value after 2015.

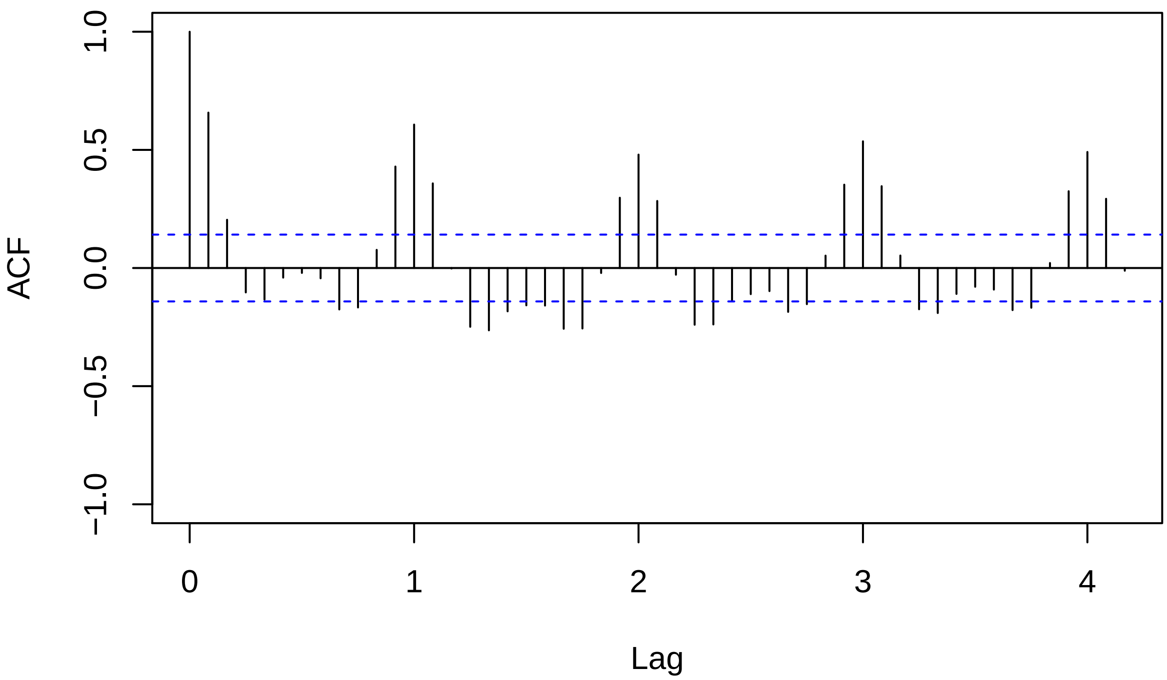
```
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2017 5137.306 3805.814 5531.396 4367.612 4120.164 4244.220 4488.245
##           Aug      Sep      Oct      Nov      Dec
## 2017 4635.244 4361.830 3575.078 2711.734 2933.921
```

Above are the predictions for California's coal usage in 2017. They seem to be relatively constant, with a slight decreasing trend. The predictions are 1000 to 2000 tons less than the coal usage in 2016 which may not be perfectly reasonable as after 2015, California's coal usage seemed to fluctuate between 5000 and 6000 tons. A probable explanation for why California's coal usage dropped so drastically at the start of 2015 is that new legislation was passed that set limits on the amount of coal to be used.

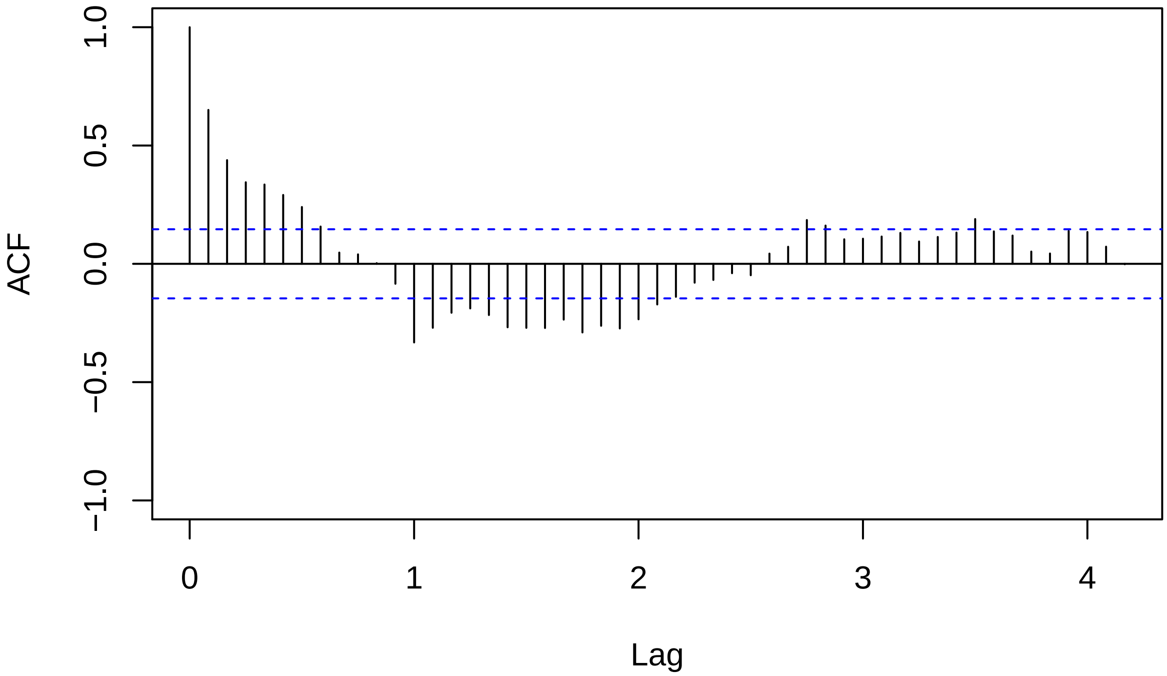
Let's take a look at Florida now. From plots of just the original data (included in the appendix), we see that there is a more straightforward, linear downward trend in the mean than in the California coal usage data. Because there seems to be a fairly deterministic, decreasing trend in the mean, we can fit a linear regression model and subtract this mean from the data. The detrended data are plotted below.



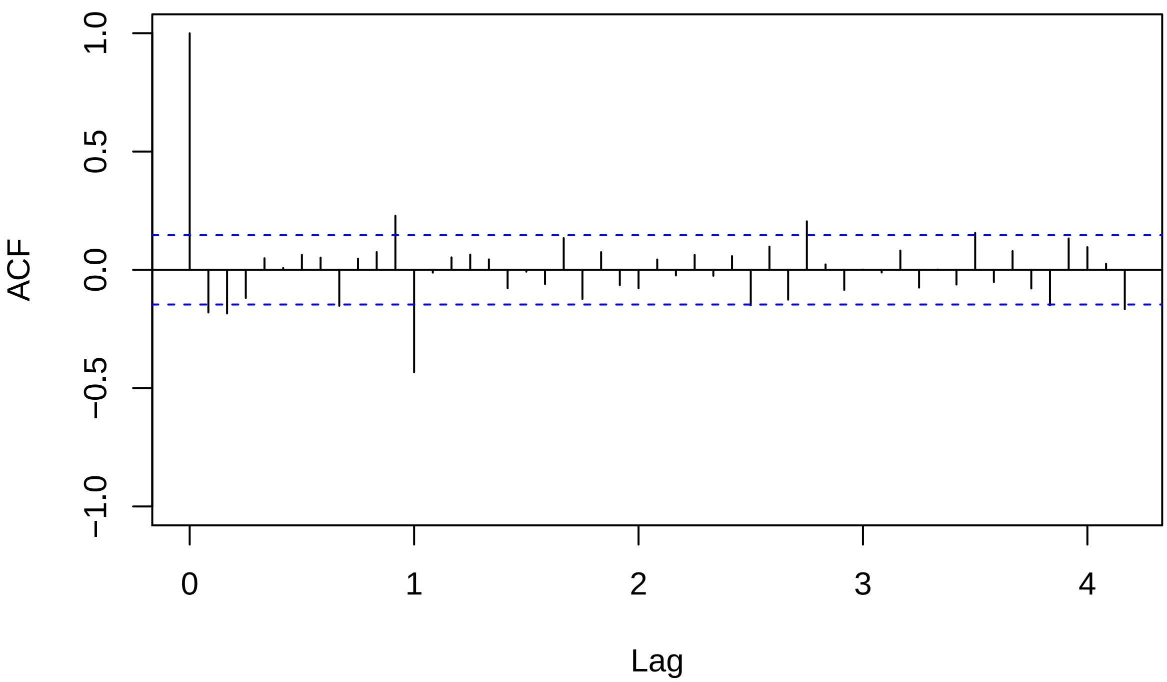
These data seem to be zero mean and we now look at the acf.

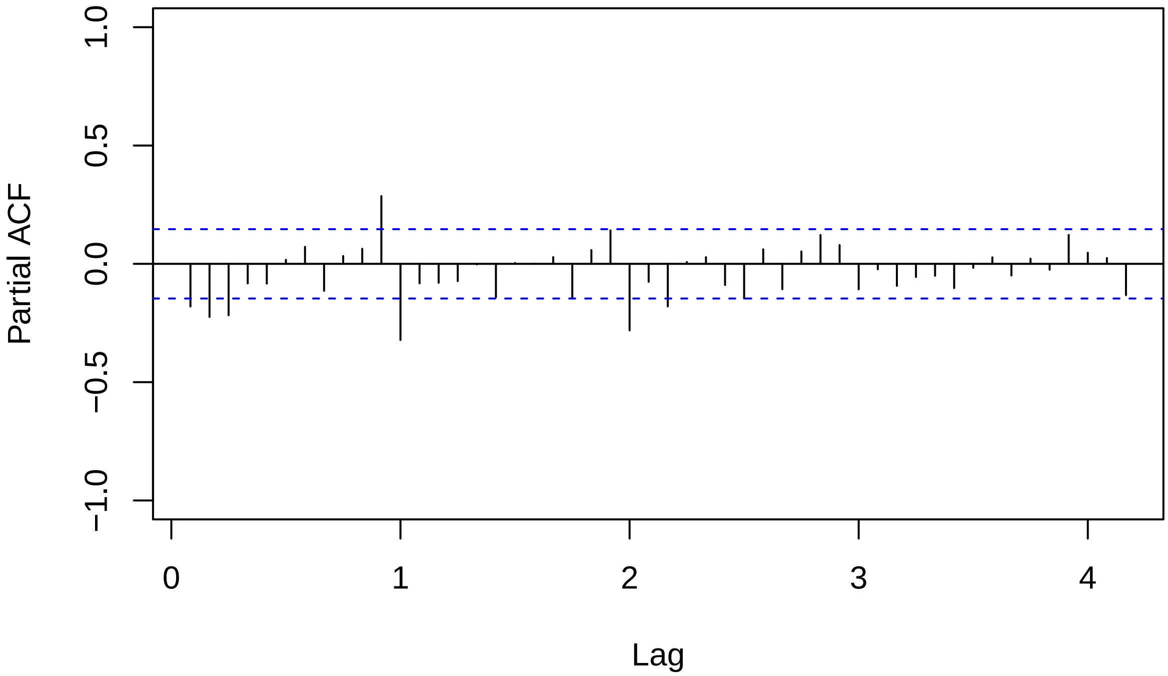


From this acf, we see that there is some seasonal behaviour with seasonality of 12 months. We can take care of this seasonality with seasonal differencing at lag 12 and once again we inspect the acf.



Though we have gotten rid of the seasonal behaviour in this data, the acf still shows signs of a non-stationary process. We difference again, this time simply at lag 1, to attempt to remove this non-stationary behaviour, and again insepct the acf and also the pacf.

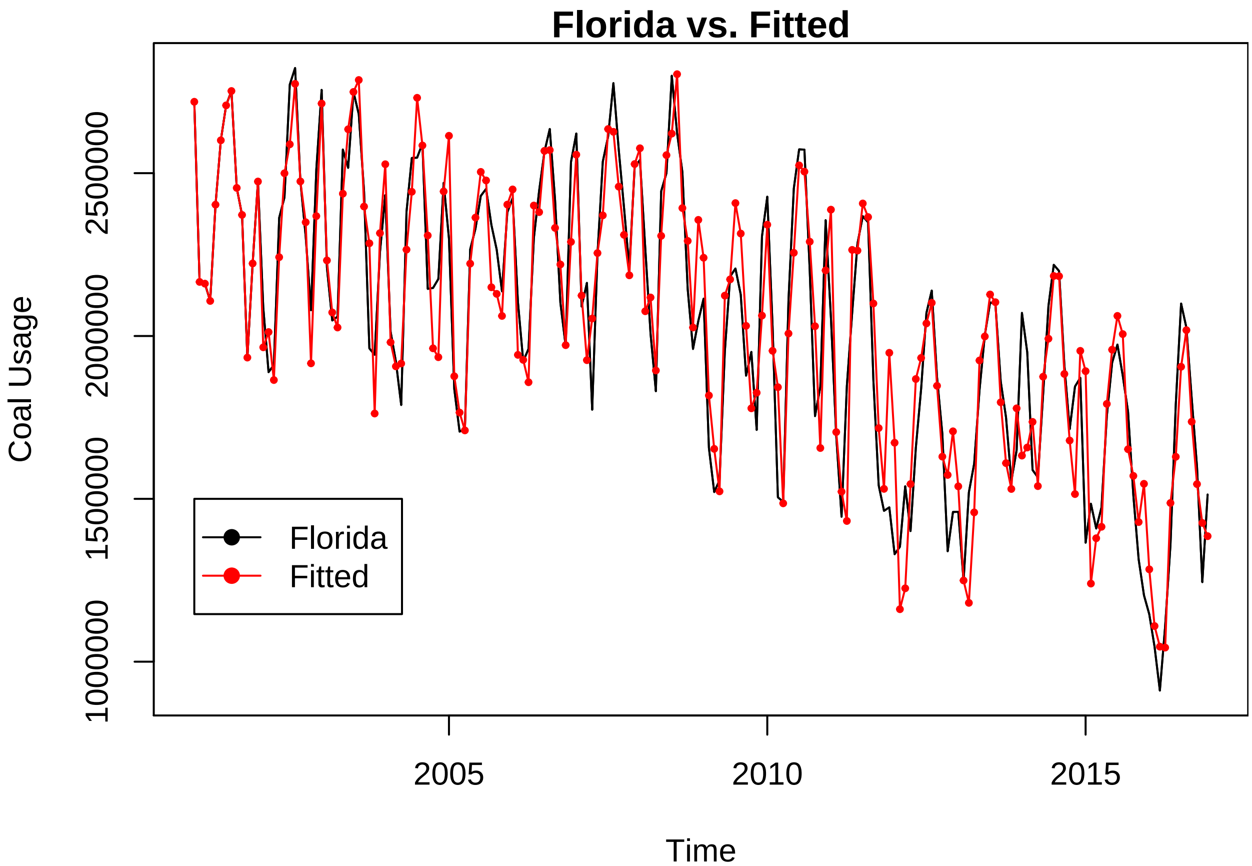




This acf now seems to belong to a stationary process. Upon close inspection, the spike at lag 12 suggests that Q is 0 or 1 and according to the pacf, P is probably also either 0 or 1. q seems to be quite small, probably just 1 whereas we might have to try several different p values. We will fit several seasonal arima models and then compare a few based on the AIC and BIC values. Because we are not too sure if Q is 0 or 1 or P is 0 or 1, we will fit models for all and compare the mean squared error of the different fits to help us decide, similar to what we did for California.

From this we get minimum AIC and BIC values for a $SARIMA(1,1,1)X(1,1,1)_{12}$, $SARIMA(2,1,1)X(1,1,1)_{12}$, and $SARIMA(1,1,1)X(1,1,0)_{12}$ models. We can fit these models to the detrended data, retrieve the fitted values and add back the linearly decreasing mean that we fit during pre-processing. We then compute the mean-squared error for each of these models to compare fit.

We achieve the lowest mean-squared error and AIC/BIC values with the $SARIMA(1,1,1)X(0,1,1)_{12}$ model. We plot this model's fitted values along with the original data below.

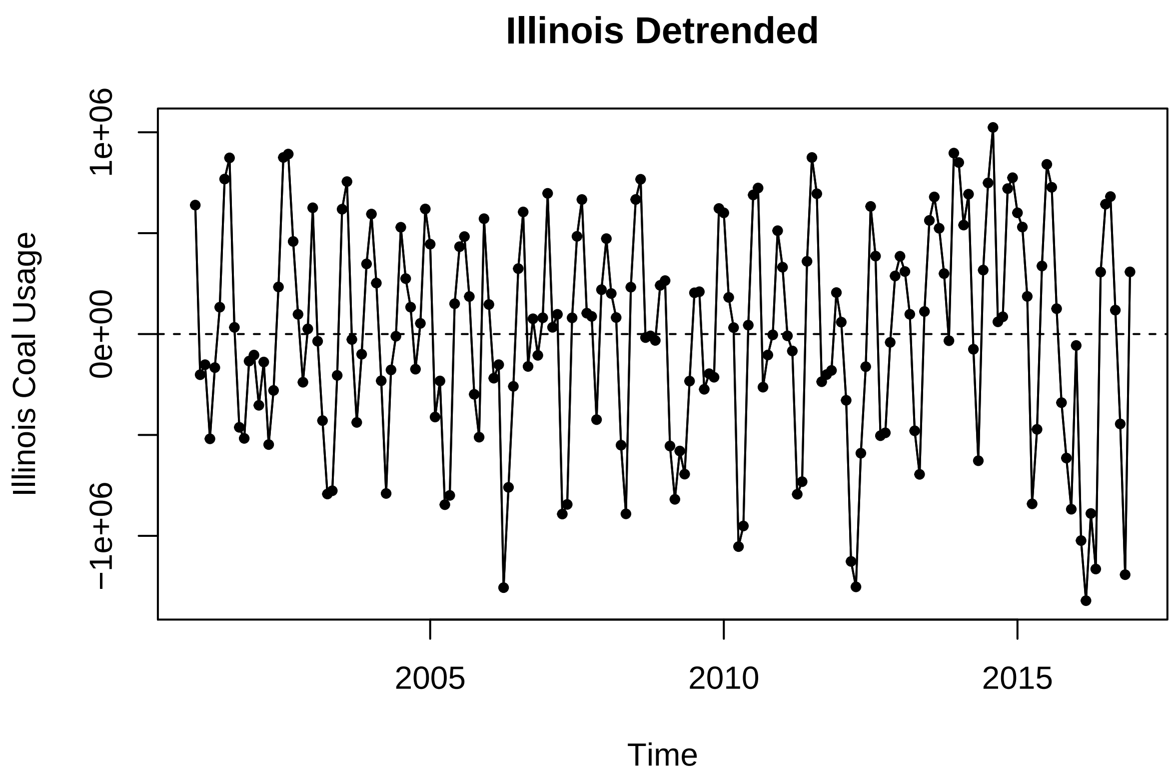


Overall this model fits the data fairly well.

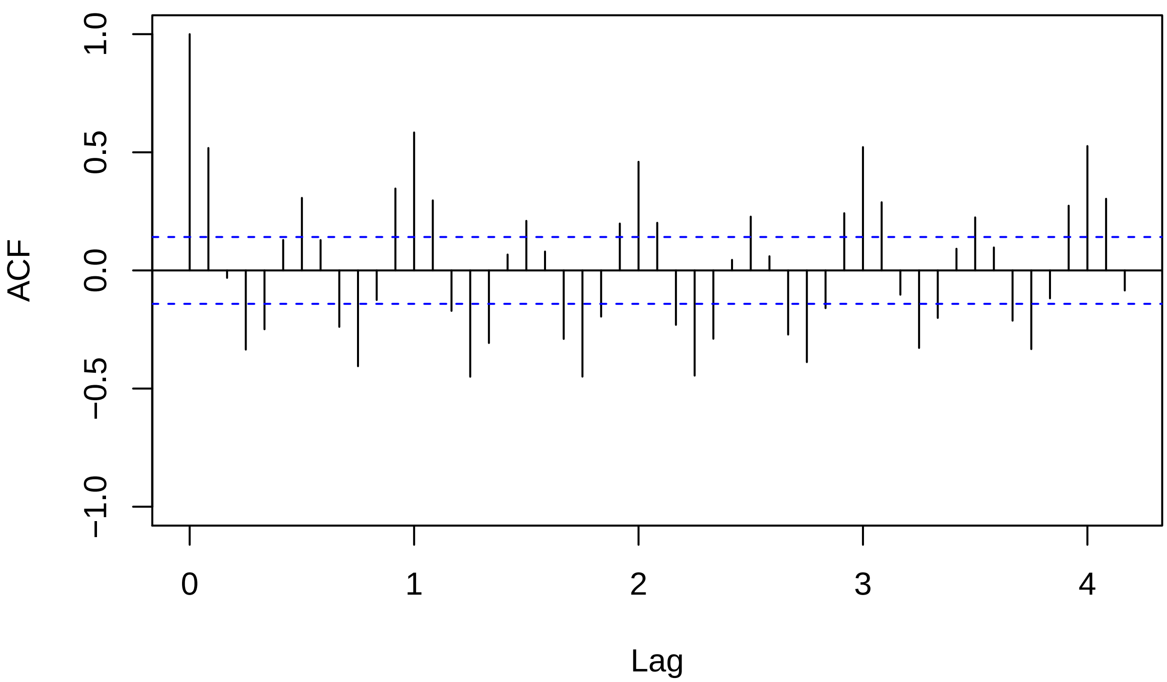
```
##          Jan     Feb     Mar     Apr     May     Jun     Jul     Aug
## 2017 1433212 1265829 1143949 1198669 1502801 1749044 1906125 1870929
##          Sep     Oct     Nov     Dec
## 2017 1629377 1415598 1239901 1391154
```

The predictions for 2017 are shown above. These values follow the trend of the rest of the data which showed higher coal usage in the summer months and lower coal usage in the cooler months (possibly due to increased usage of air conditioning during the extremely warm months) while also still incorporating the decreasing trend overall.

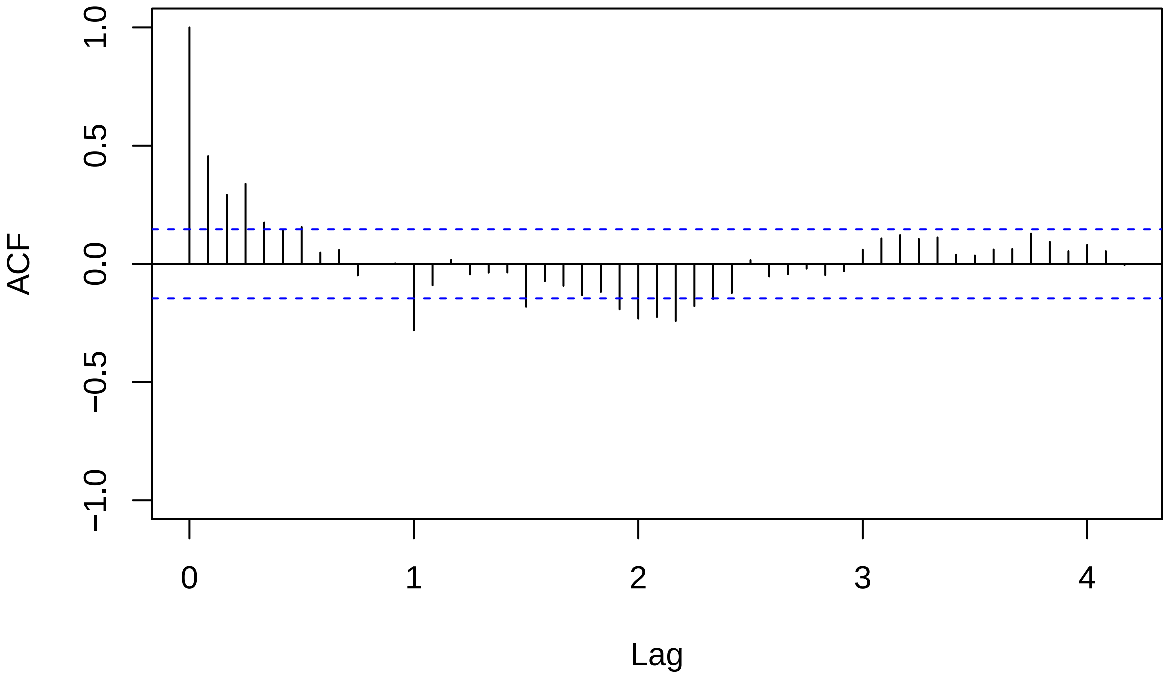
We now look to Illinois. These data (graph included in the appendix) seem to have more of a quadratic looking mean compared to the states we have already looked at, California and Florida. We can do something similar to what we did for Florida which is fit a linear regression model, this time with a squared term to see if we can eliminate this nonconstant mean. After performing this multiple linear regression, we have a process that seems to have a fairly constant, zero mean. The detrended data are plotted below.



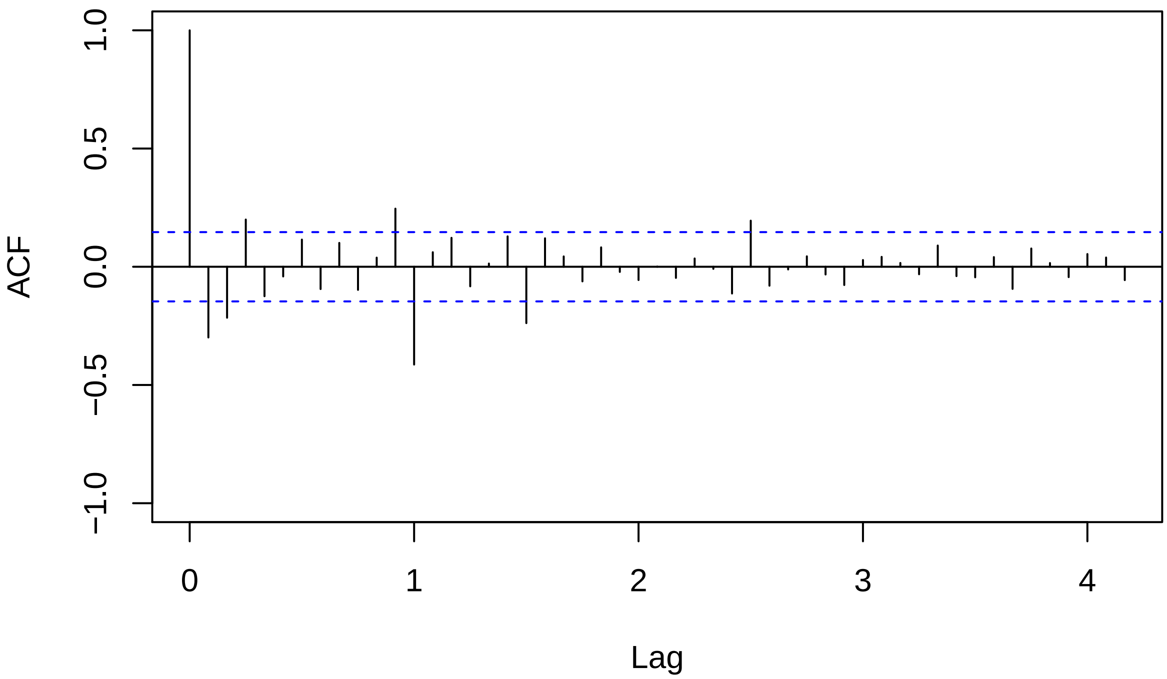
We can now take a look at the acf to determine if the process is stationary or if there is any sort of seasonality or possibly a unit root.

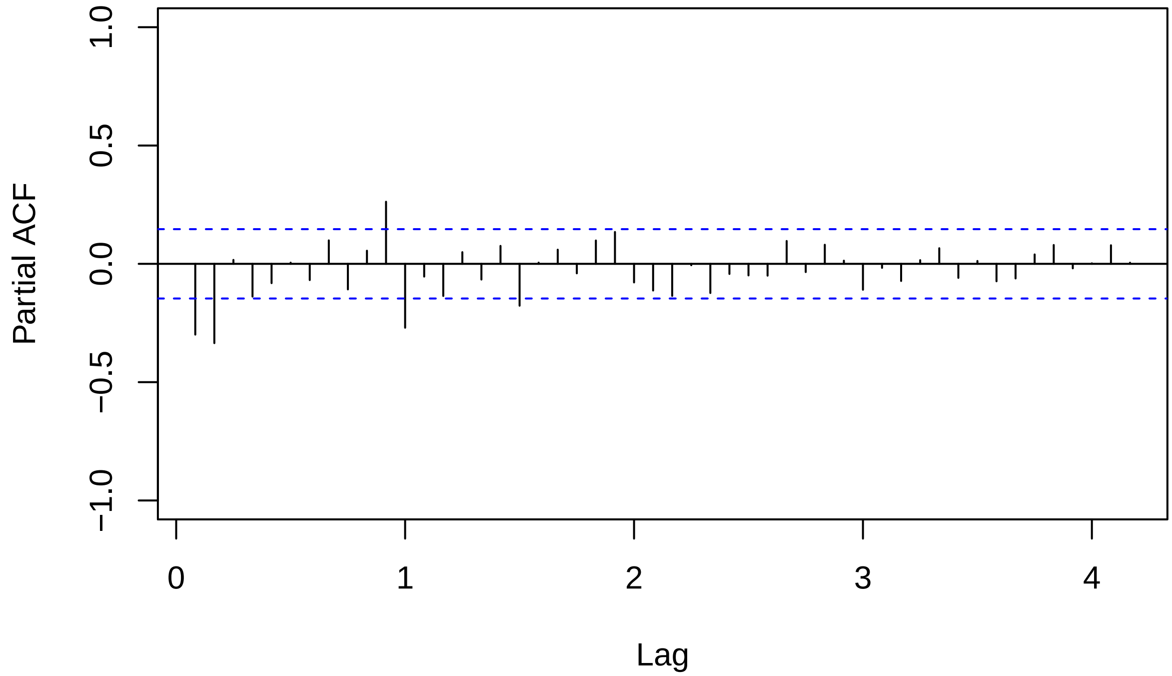


Similar to what we saw with the Florida data, there seems to be a seasonality component with 12 month periodicity. We can difference at lag 12 and once again plot the acf to see if we have a stationary process.



Again similar to what we saw in Florida, this seasonal differencing still does not fully get us to a process that seems stationary so we once again difference at lag one and plot the acf and pacf.



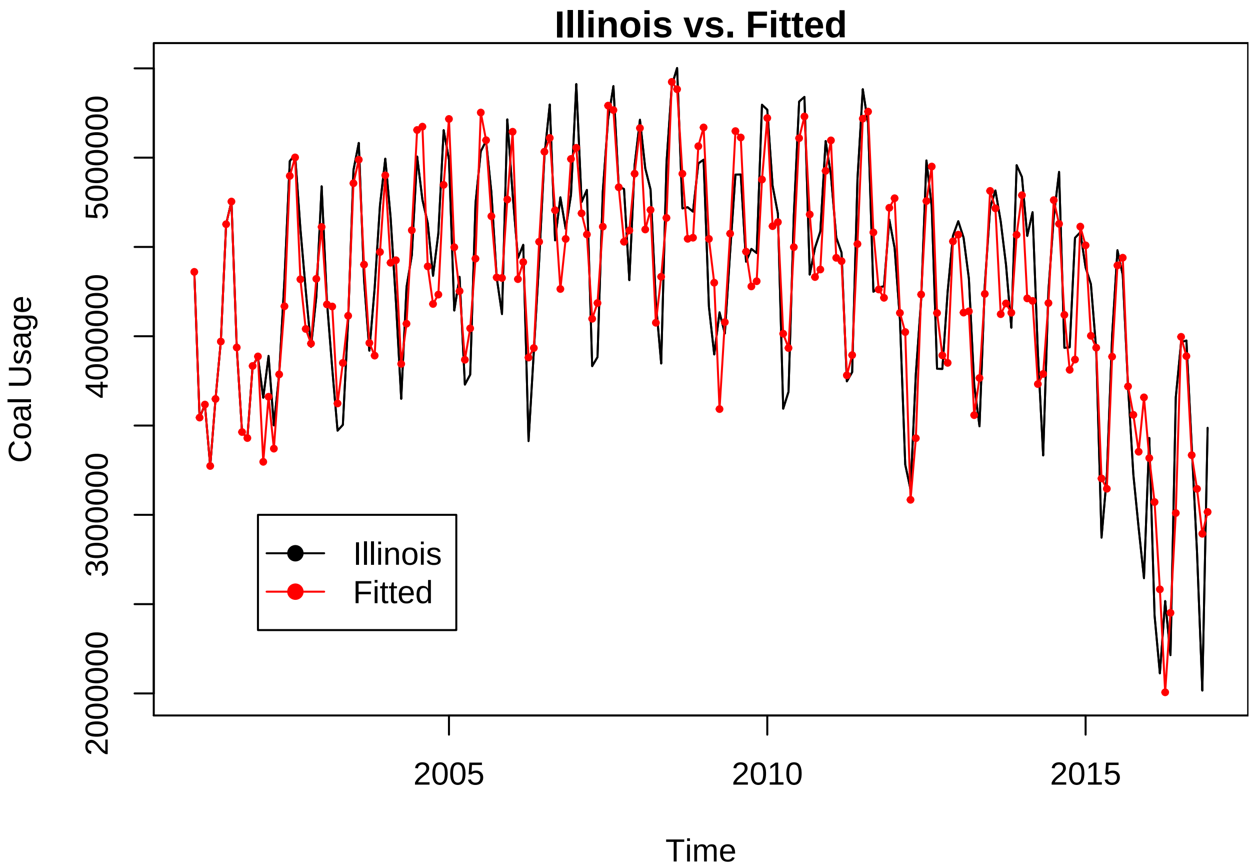


Based on the acf, we have once again that Q and P are both either 0 or 1. Similar again to what we did with California and Florida, we can fit several of these SARIMA models, select a few based on the AIC/BIC values and then select a model based on the mean square error. Based on the acf and pacf, both p and q seem relatively small, probably no more than 2 or 3.

From the code which is included in the appendix, we get minimum AIC/BIC values for $SARIMA(1,1,1)X(1,1,1)_{12}$, $SARIMA(3,1,1)X(1,1,0)_{12}$, $SARIMA(3,1,1)X(0,1,1)_{12}$, and $SARIMA(1,1,1)X(0,1,1)_{12}$. We will now calculate the mean squared error and select the model with the lowest one.

The model with the lowest mean square error was $SARIMA(3,1,1)X(0,1,1)_{12}$. We can plot the fitted values against the original data to visually assess fit.

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

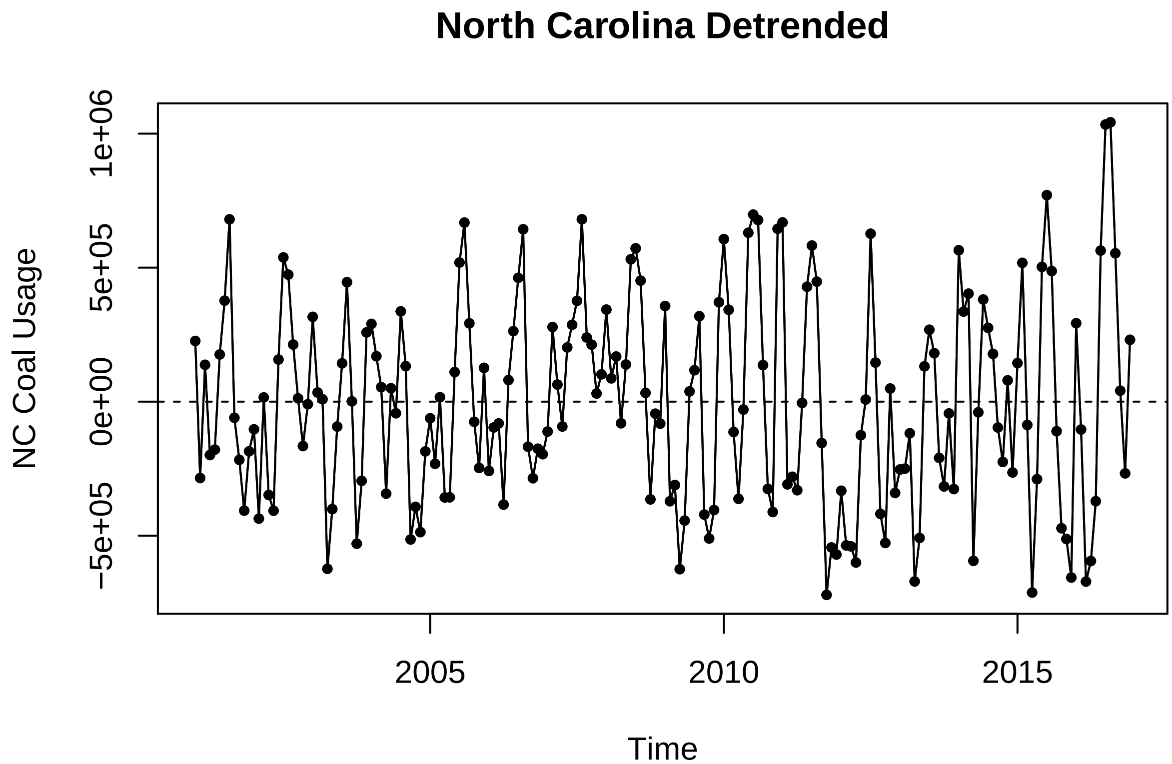


This model seems to fit the data similarly well to the model that we fit for the Florida data and better than the model we fit for California.

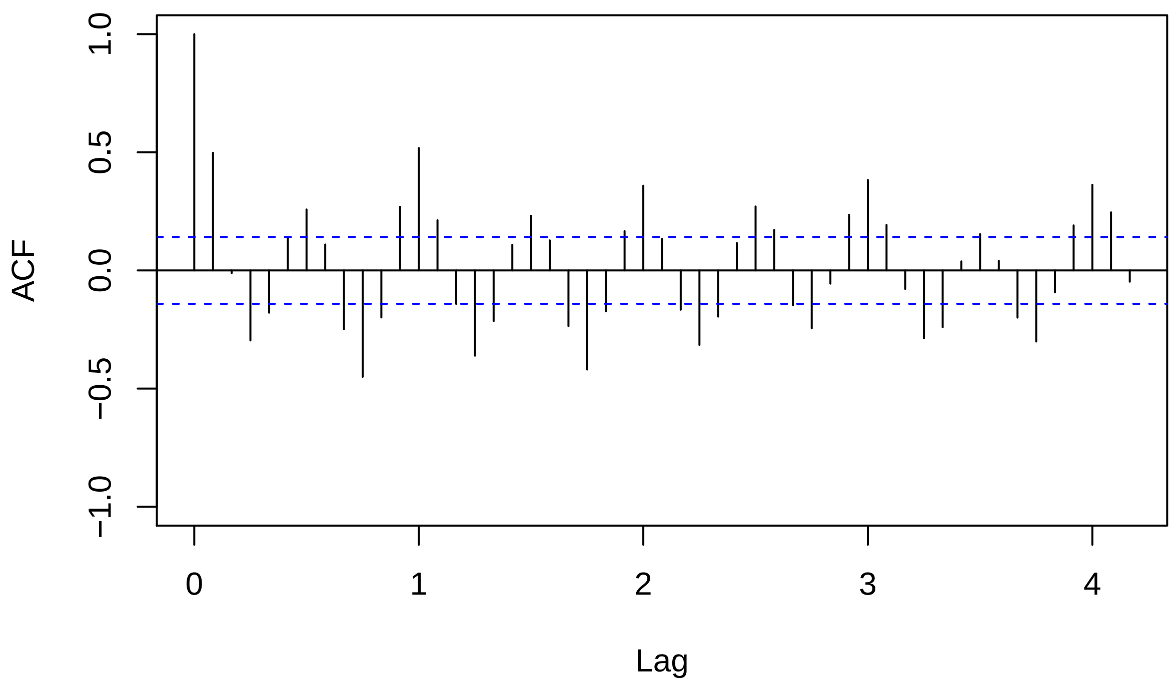
```
##          Jan     Feb     Mar     Apr     May     Jun     Jul     Aug
## 2017 3451859 2827791 2757482 2211234 2274871 3015869 3525609 3534438
##          Sep     Oct     Nov     Dec
## 2017 2873323 2673571 2585684 3068940
```

Above are the predictions for Illinois' coal usage for 2017. These predictions also follow the pattern of the previous data for Illinois where coal usage is highest in December, January, July and August (possibly due to increase use of heat in the coldest winter months and use of air conditioning in the hottest summer months) and these predictions follow the steadily decreasing trend in the data.

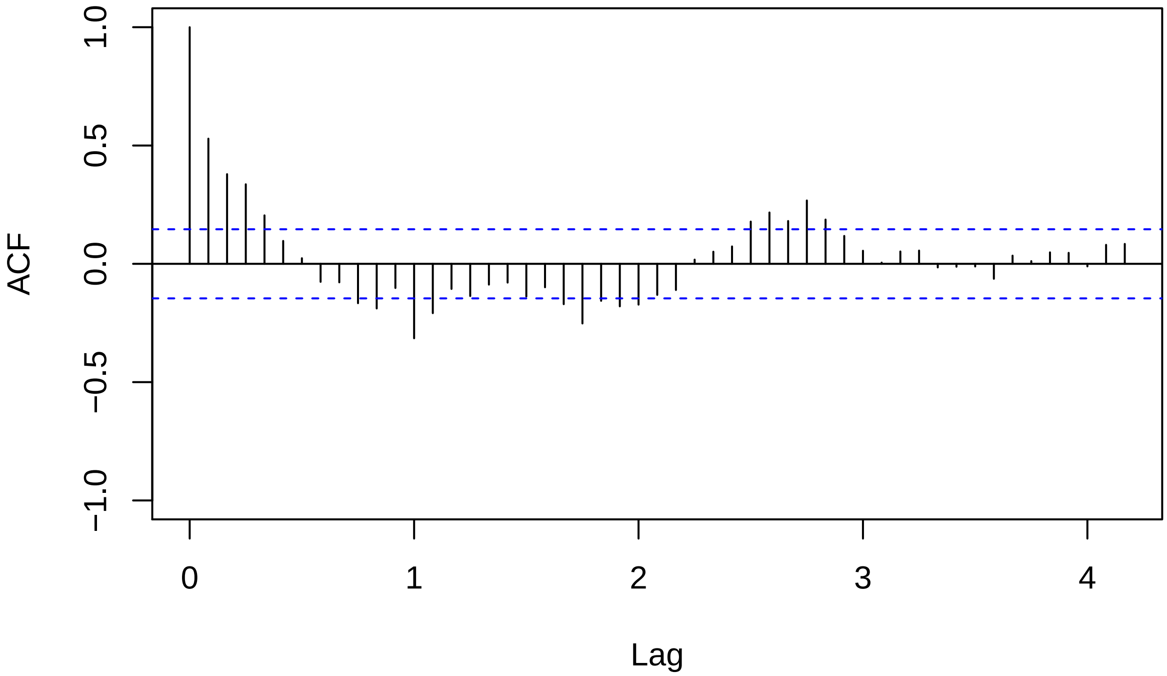
We now move on to North Carolina. Based on the plot of the original North Carolina data (included in the appendix), we see a similar pattern to what we just saw for Illinois which is a rising and then once again falling trend in the mean. We handle this in the same way that we handled the trend for Illinois: by fitting a multiple linear regression model with respect to time with a time squared term. The plot of the detrended data is included below.



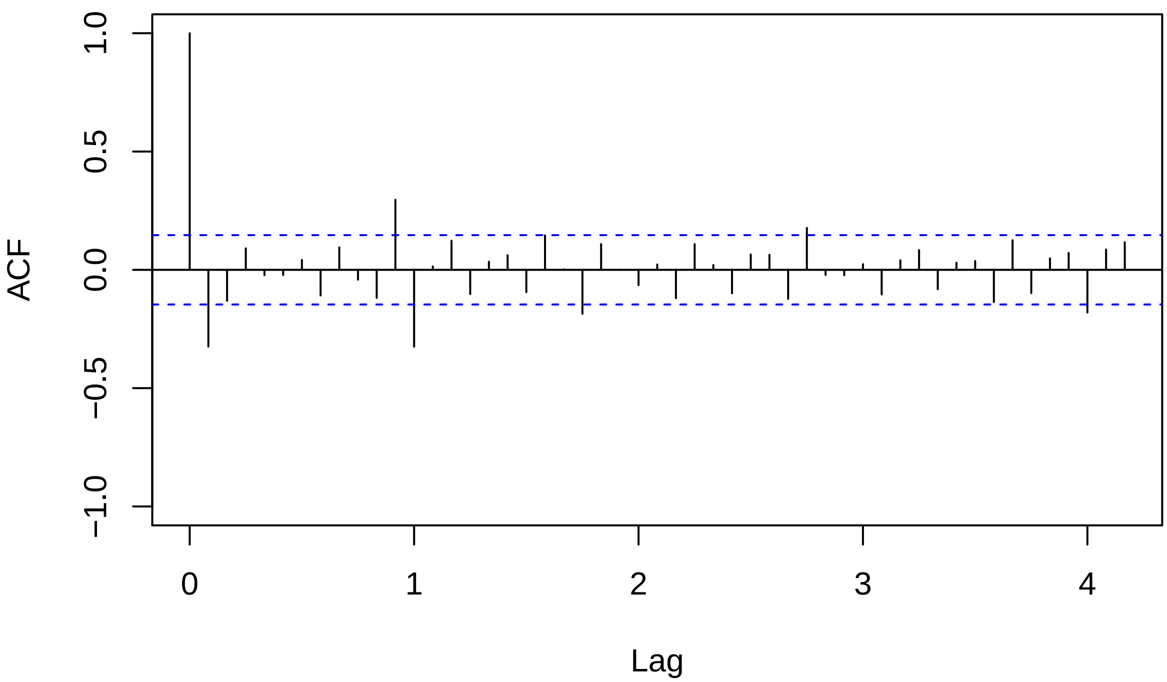
The data now look approximately zero mean, save for maybe a slight increase in variance towards the end of the dataset. We now look at the acf to see if we have a stationary process.

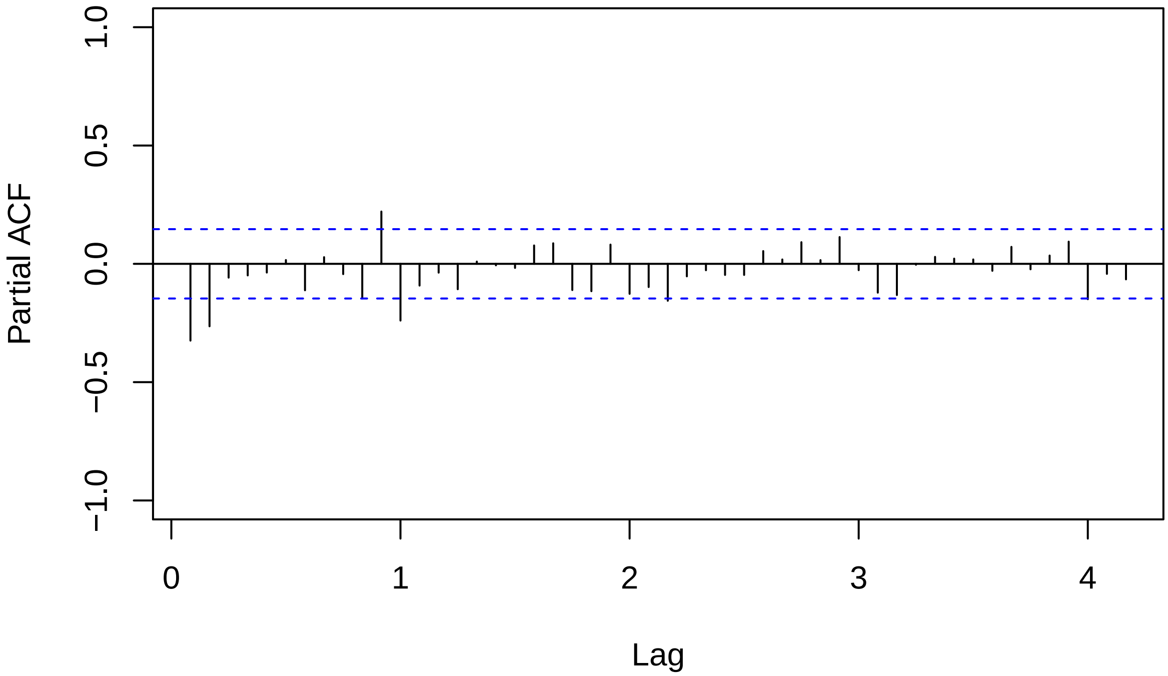


Similar to what we have seen in Florida and Illinois already, there is clear seasonality at lags of 12. We again handle this in the same way that we did for Florida and Illinois: differencing at lag 12.



Once again, very similar to what we have seen in Florida and Illinois, there is still some non-stationary behaviour which we will eliminate by differencing at lag one. We do this and then once again inspect the acf and pacf plots.

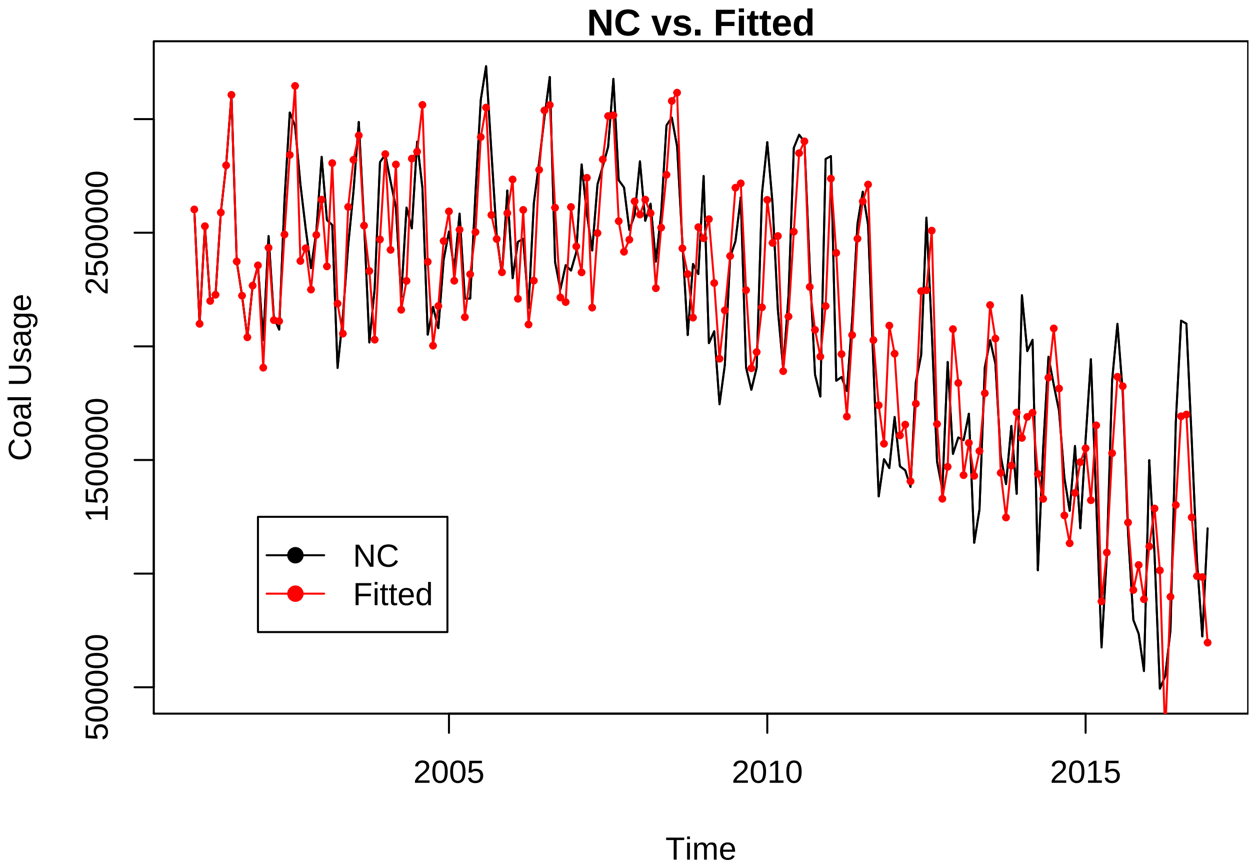




Based on these plots, p and q both seem to be relatively small. It may even be that Q and P are both 0 as even the spikes around lag 12 are quite small. We will set P equal to 0 and try out different SARIMA models with Q equal to 0 or 1. We will again compare AIC and BIC values between each of these models, select the few that have the lowest and then actually fit one to the data based on the mean-square error.

Based on lowest AIC and BIC values of the models tested, we have the choice of the following models: $SARIMA(1,1,1)X(0,1,1)_{12}$, $SARIMA(0,0,3)X(0,1,0)_{12}$ and $SARIMA(0,0,1)X(0,1,0)_{12}$. We compare their mean-squared errors now to select one of these three.

Based on this analysis and the models selected, the best model we have found for this data is a $SARIMA(1,1,1)X(0,1,1)_{12}$. Below is a plot of the fitted values and the original data.

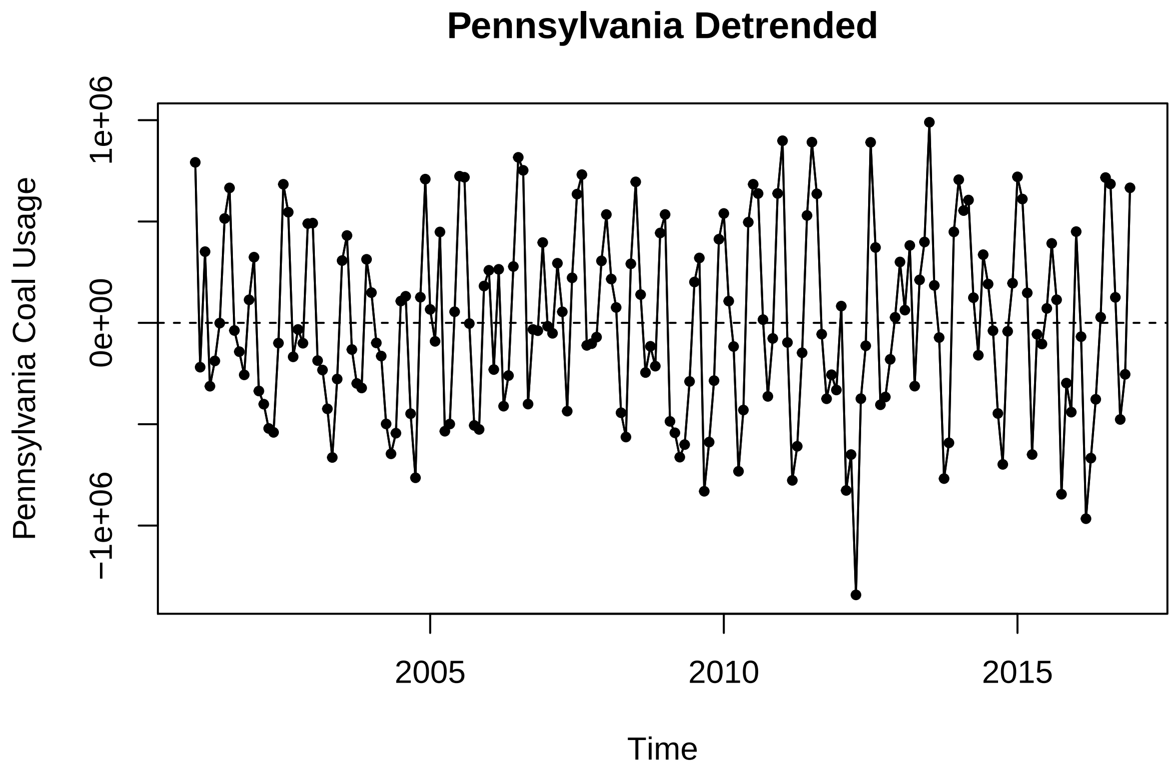


Though this model definitely fits the data better than the model we fit for California, it does not seem to do as good of a job at fitting the data as the models we fit for Florida and Illinois.

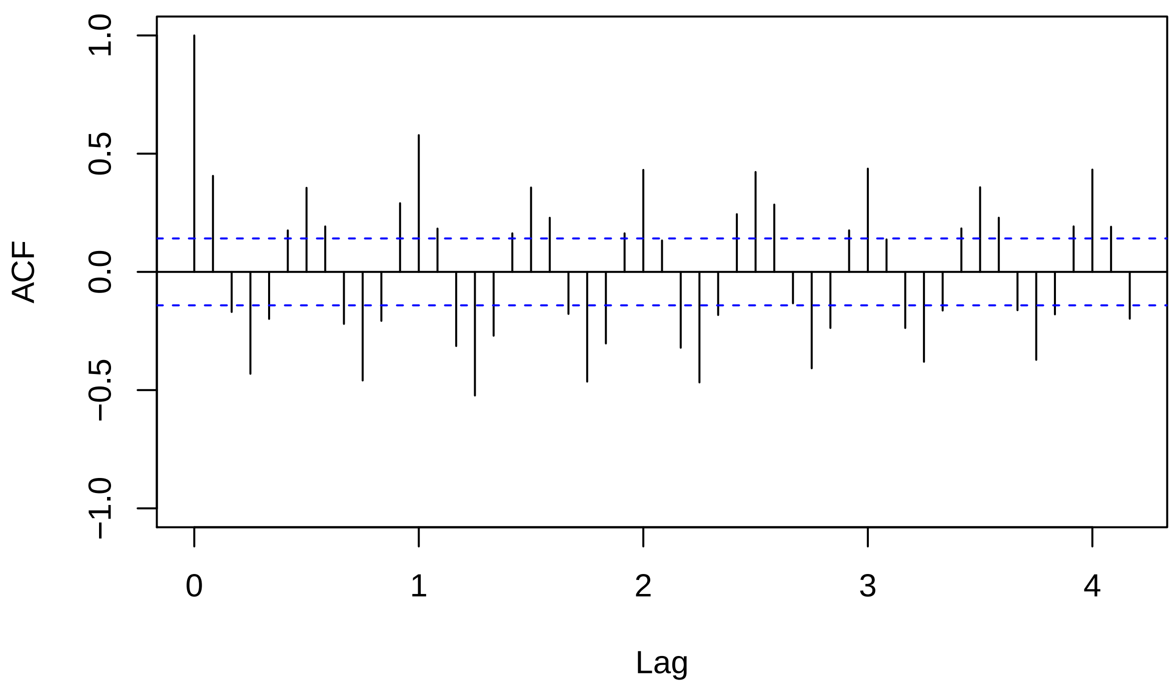
```
##             Jan       Feb       Mar       Apr       May       Jun       Jul
## 2017 1403843.5 1076027.8  784564.8  373286.3  659522.3 1212472.8 1426768.8
##             Aug       Sep       Oct       Nov       Dec
## 2017 1306356.8  783097.6  457215.1  490631.8  547781.5
```

Above are the predictions for North Carolina's coal usage for 2017. The North Carolina data had larger fluctuations between months than the data for Florida and Illinois and the predictions for 2017 do a fairly good job of reflecting that. Once again the months that had higher coal usage in the previous years also have higher coal usage in the predictions, and continue to show a steadily decreasing trend.

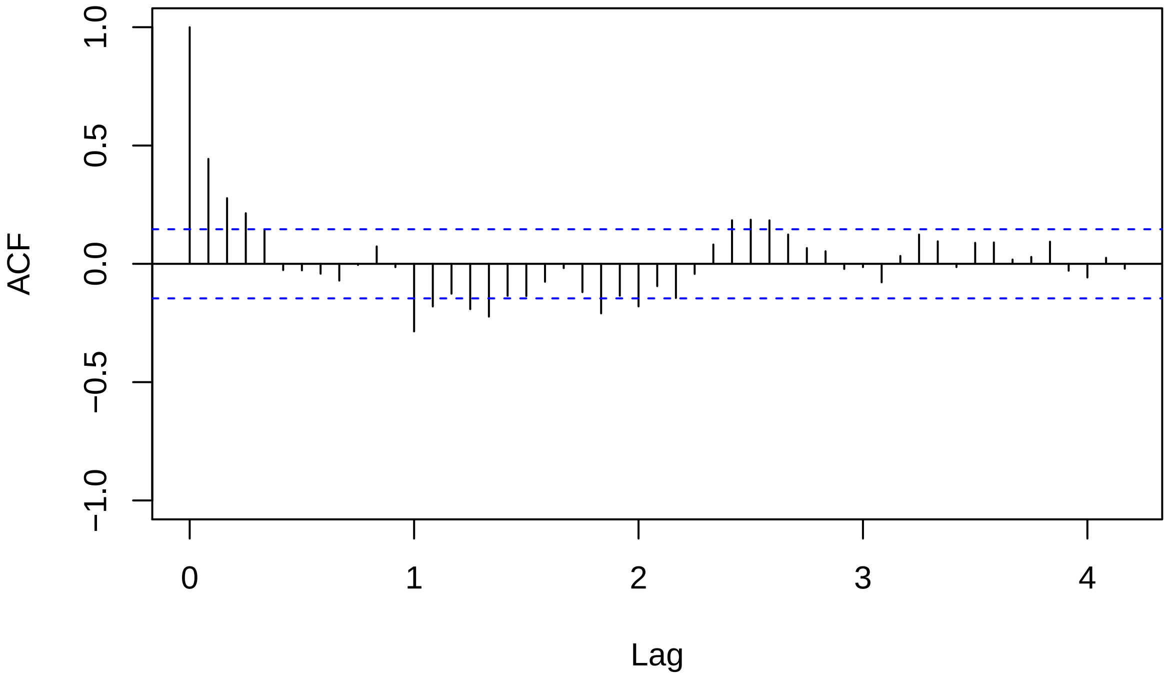
We end our analysis with Pennsylvania. Again very similar to what we saw for Illinois and North Carolina (graph included in the appendix), there is not a straightforward linear trend in the mean so we will fit a multiple linear regression model with respect to time for the mean which includes a time squared term. Below is a plot of the detrended data.



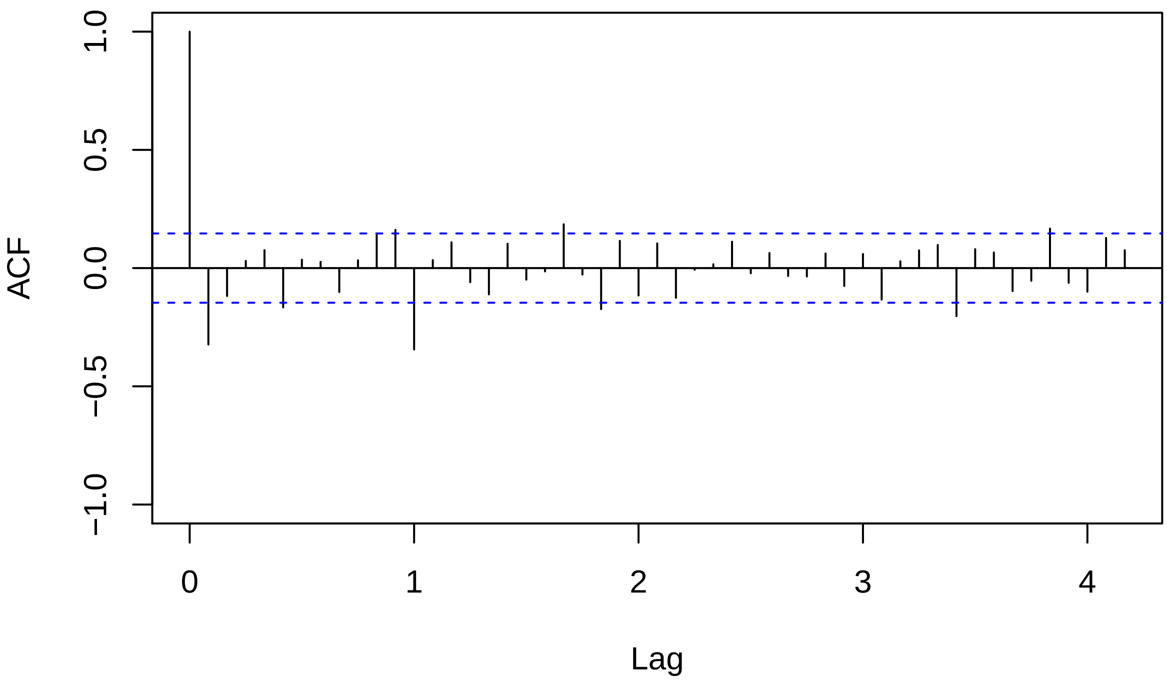
The data appear to have approximately zero mean now. As we have done for all four other states, we now take a look at the acf to see if we have a stationary process but considering that three of the four previous states had a very clear 12 month seasonal period, we can assume that this is probably the pattern we will see for Pennyslvania as well.

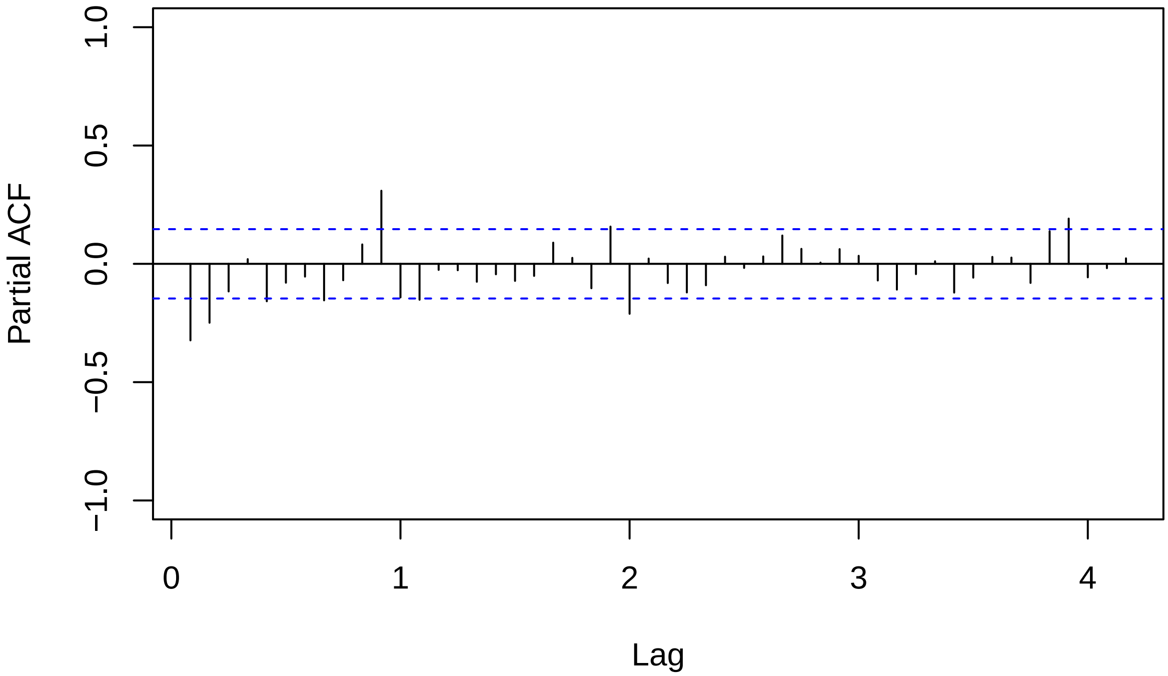


As guessed, there is a seasonality of period 12 in this data. We can also guess that we will likely have to difference the data at lag 1 to get an acf that resembles that of a stationary process.



Based on this acf, it does seem like we need to difference once at lag 1. We look at the acf and pacf of this differenced data below.

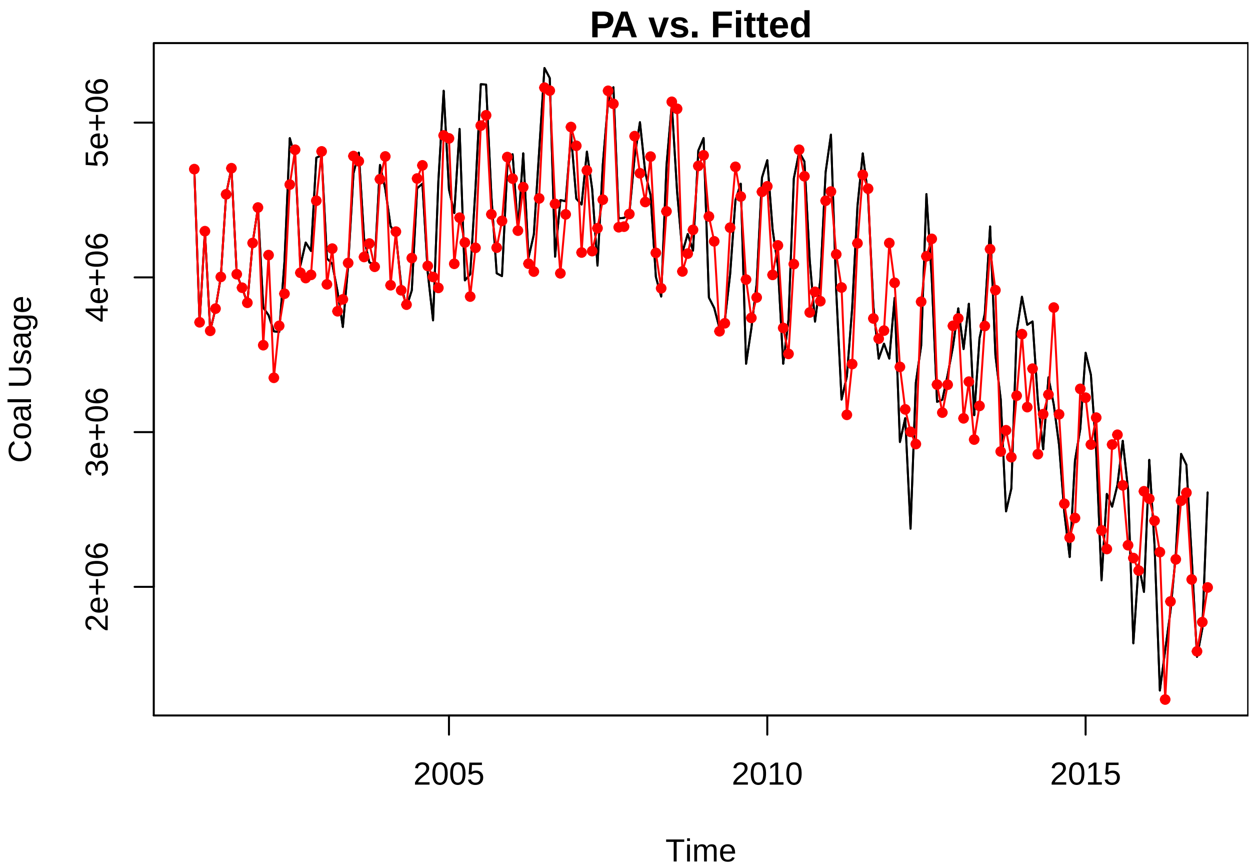




Very similar to what we saw for North Carolina, p and q seem quite small and we can once again probably say that P and Q are probably equal to 0 or 1 so we will try to fit different SARIMA models with these four different combinations of P and Q values, consider a few based on their AIC and BIC values and then fit and plot one based on the lowest mean squared error.

Because we tried out four different combinations of P and Q values, we have several models that achieved low AIC and BIC values. We will now calculate the mean squared error for the following models: $SARIMA(1,1,1)X(0,1,1)_{12}$, $SARIMA(1,1,1)X(0,1,0)_{12}$, $SARIMA(3,1,3)X(0,1,0)_{12}$, $SARIMA(1,1,1)X(1,1,1)_{12}$, $SARIMA(3,1,4)X(0,1,1)_{12}$ and $SARIMA(1,1,1)X(1,1,0)_{12}$.

With the combination of low AIC/BIC values and low mean squared error compared to the other models tested, we will fit a $SARIMA(1,1,1)X(1,1,1)_{12}$. A plot of the fitted values along with the original data is included below.



This model seems similar in quality of fit to the models fitted for Florida and Illinois.

```
##          Jan     Feb     Mar     Apr     May     Jun     Jul     Aug
## 2017 2632643 1962621 1574789 1252793 1432915 1789164 2250330 2093053
##          Sep     Oct     Nov     Dec
## 2017 1462765 1046407 1245595 1773042
```

Above are the predictions for Pennsylvania's coal usage in 2017. Similar to the other states, these predictions stay true to the rest of the trends in the state's previous data, where there is higher usage in January, July and August and also shows a steadily decreasing trend.

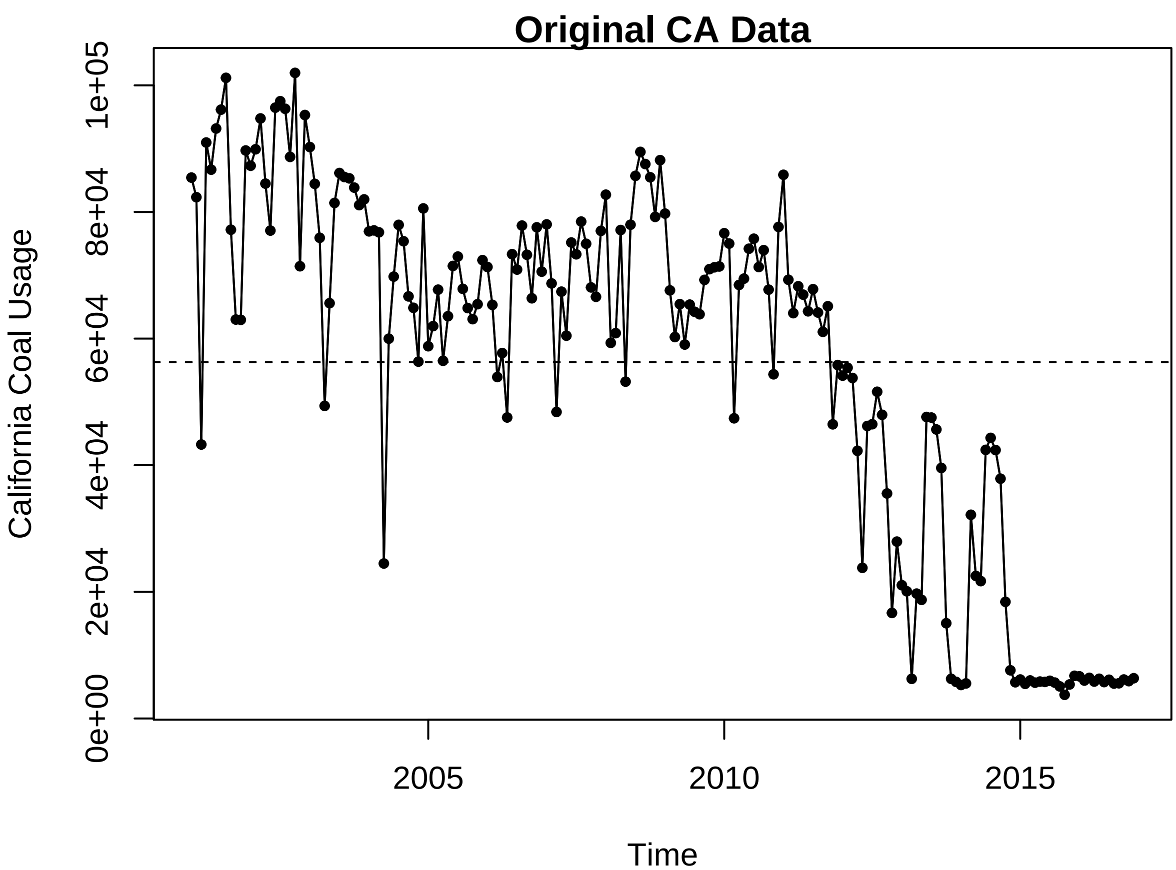
With the exception of California, all the states needed seasonal differencing and lag one differencing to get a stationary process. However, all states were fit with a seasonal ARIMA model rather than a simpler ARMA model. Also save for California, the states required preprocessing in the form of either simple or multiple linear regression. California was the one state that did not exhibit clear seasonal behaviour and whose mean did not really have a clear underlying function that could be modeled using linear regression, hence the use of a sigmoidal function. Because California had quite some jumps in the data, especially the drop in coal usage at the beginning of 2015, it was especially difficult to find a model for this data that could be rationalized and that didn't overfit the data either. Each state did show an overall decrease in coal usage since 2001, most likely due to stricter regulations and a move towards more renewable energy sources.

APPENDIX

```r
library(ggplot2)
library(reshape2)
library(tseries)
library(forecast)
setwd("~/Dropbox/U4/MATH545")

USCoal<-read.csv('USCoal-2016.csv',header=TRUE)
Y.ca<-ts(USCoal[,3],start=c(2001,1),frequency=12)
Y.fl<-ts(USCoal[,4],start=c(2001,1),frequency=12)
Y.il<-ts(USCoal[,5],start=c(2001,1),frequency=12)
Y.nc<-ts(USCoal[,6],start=c(2001,1),frequency=12)
Y.pa<-ts(USCoal[,7],start=c(2001,1),frequency=12)
```

California Analysis

```r
par(mar = c(4,4,1,2))
plot(Y.ca, ylab = "California Coal Usage", xlab = "Time", main = "Original CA Data"); lines(Y.ca); poin
```




```r
n <- length(Y.ca)
tvec <- c(1:n)
fit1 <- nls(Y.ca ~ SSlogis(tvec, Asym, xmid, scal))
```

```r
coef.sig <- coef(summary(fit1))[,1]
mt<-coef.sig[1]/(1+exp((coef.sig[2]-tvec)/coef.sig[3]))
mt <- ts(mt, start = c(2001,1), frequency = 12)

Yt <- Y.ca - mt

plot(Yt, ylab = "California Coal Usage", xlab = "Time", main = "California Detrended")
lines(Yt); points(Yt, pch = 19, cex = 0.6); abline(h = 0, lty = 2)
```

```r
acf(Yt, ylim = range(-1,1), lag.max = 50, main = "")
```

```r
pacf(Yt, ylim = range(-1,1), lag.max = 50, main = "")
```

```r
Q <- 1; D <- 0; d <- 0
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,4,4)
for(p in 0:3){
  for(q in 0:3){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
rownames(P1AIC.mat)<-paste("p=",c(0:3),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:3),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:3),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:3),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0    q=1    q=2    q=3
## p=0 37.5237 6.9743 3.3184 4.5986
## p=1  0.0000 1.9489 3.9463 5.8755
## p=2  1.9483 3.9485 5.9406 7.1808
## p=3  3.9469 5.9405 4.9831 8.8908
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3
## p=0 34.2662  6.9743  6.5759 11.1136
## p=1  0.0000  5.2064 10.4613 15.6480
## p=2  5.2058 10.4635 15.7131 20.2108
## p=3 10.4619 15.7130 18.0131 25.1783
```

```r
Q <- 0; D <- 0; d <- 0
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,4,4)
for(p in 0:3){
  for(q in 0:3){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
```

```r
}
rownames(P1AIC.mat)<-paste("p=",c(0:3),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:3),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:3),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:3),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0    q=1    q=2    q=3
## p=0 38.4783 4.4647 1.2444 3.2289
## p=1  0.0000 1.6260 3.1889 5.2354
## p=2  1.5864 3.5813 4.8063 6.5153
## p=3  3.5675 5.1324 3.5060 8.4824
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3
## p=0 35.2208  4.4647  4.5019  9.7439
## p=1  0.0000  4.8835  9.7039 15.0079
## p=2  4.8439 10.0963 14.5787 19.5453
## p=3 10.0825 14.9049 16.5360 24.7699
```

```r
fit1 <- arima(Yt, order = c(3,0,2), seasonal = list(order=c(0,0,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit1.resids <- residuals(fit1)
fit1.fitted <- Yt - fit1.resids
fit1.fitted <- fit1.fitted + mt

sum((fit1.fitted - Y.ca)^2)
```

```
## [1] 2.928436e+15
```

```r
fit2 <- arima(Yt, order = c(1,0,0), seasonal = list(order=c(0,0,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit2.resids <- residuals(fit2)
fit2.fitted <- Yt - fit2.resids
fit2.fitted <- fit2.fitted + mt

sum((fit2.fitted - Y.ca)^2)
```

```
## [1] 2.925829e+15
```

```r
fit3 <- arima(Yt, order = c(1,0,0), seasonal = list(order=c(1,0,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit3.resids <- residuals(fit3)
fit3.fitted <- Yt - fit3.resids
fit3.fitted <- fit3.fitted + mt

sum((fit3.fitted - Y.ca)^2)
```

```
## [1] 2.934654e+15
```

```r
fit4 <- arima(Yt, order = c(3,0,2), seasonal = list(order=c(1,0,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit4.resids <- residuals(fit4)
fit4.fitted <- Yt - fit4.resids
fit4.fitted <- fit4.fitted + mt

sum((fit4.fitted - Y.ca)^2)
```

```
## [1] 2.936185e+15
```

```r
fit5 <- arima(Yt, order = c(3,0,2), seasonal = list(order=c(1,0,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit5.resids <- residuals(fit5)
fit5.fitted <- Yt - fit5.resids
fit5.fitted <- fit5.fitted + mt

sum((fit5.fitted - Y.ca)^2)
```

```
## [1] 2.935966e+15
```

```r
fit6 <- arima(Yt, order = c(1,0,0), seasonal = list(order=c(1,0,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit6.resids <- residuals(fit6)
fit6.fitted <- Yt - fit6.resids
fit6.fitted <- fit6.fitted + mt

sum((fit6.fitted - Y.ca)^2)
```

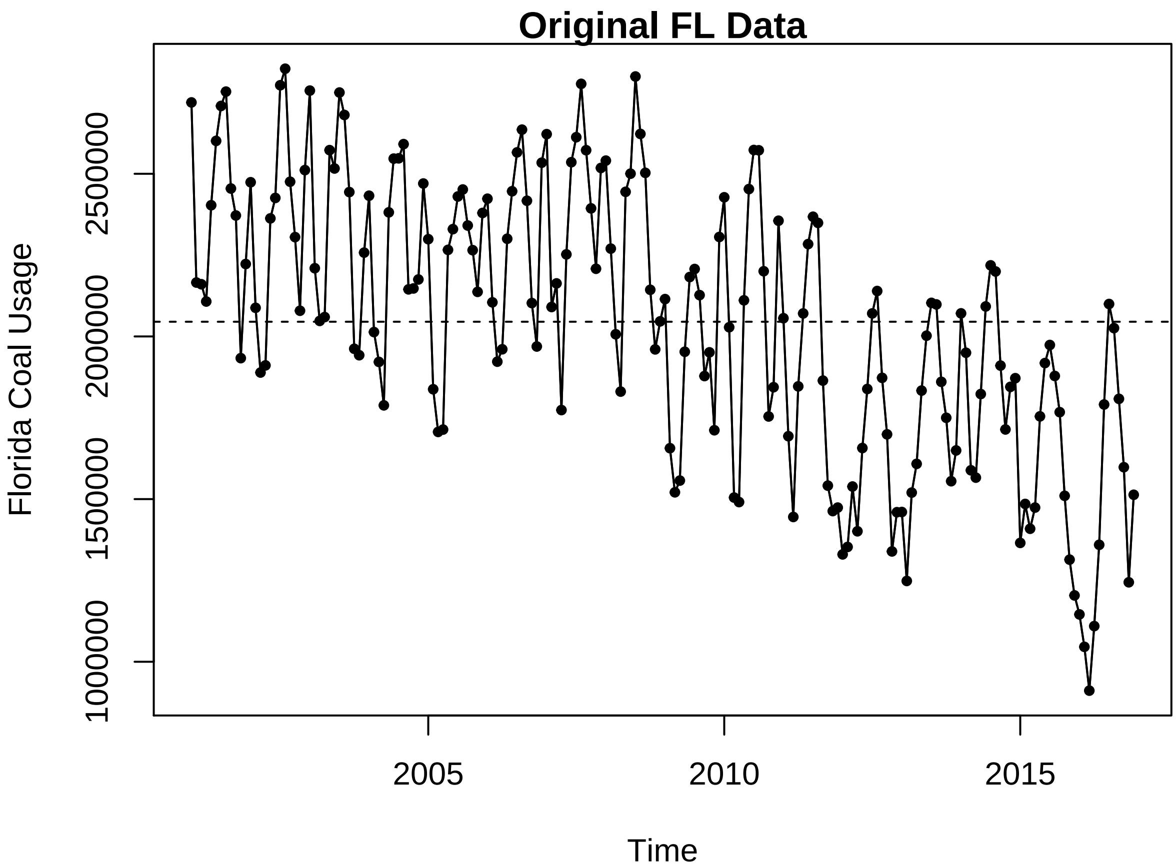
```
## [1] 2.936887e+15
```

```r
fit <- arima(Yt, order = c(3,0,2), seasonal = list(order=c(1,0,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit.resids <- residuals(fit)
fit.fitted <- Yt - fit.resids
fit.fitted <- fit.fitted + mt
par(mar = c(4,4,0,1))
plot(Y.ca,type='l', xlab = "Time", ylab = " California Coal Usage");
lines(Y.ca,pch=19,cex=0.5)
points(fit.fitted, pch = 19, cex = 0.4, col = 'red');lines(fit.fitted,col='red')
legend(2013, 90000, c("California", "Fitted"), col = c("black", "red"), pch = 19, lty = 1)
```

```r
pred1<- predict(fit1, n.ahead = 12)$pred
t <- c(193:204)
mt<-coef.sig[1]/(1+exp((coef.sig[2]-t)/coef.sig[3]))
pred1 <- pred1 + mt
pred1
```

Florida Analysis

```r
par(mar = c(4,4,1,2))
plot(Y.fl, ylab = "Florida Coal Usage", xlab = "Time", main = "Original FL Data"); lines(Y.fl); points(Y
```




```r
n <- length(Y.fl)
tvec <- c(1:n)
Y.fl.fit <- lm(Y.fl~tvec)
mt <- Y.fl.fit$coef[1] + Y.fl.fit$coef[2]*tvec
mt <- ts(mt, start = c(2001,1), frequency = 12)

Yt <- Y.fl - mt

par(mar= c(4,4,2,1))
plot(Yt, ylab = "Florida Coal Usage", xlab = "Time", main = "Florida Detrended"); lines(Yt);
points(Yt, pch = 19, cex = 0.6); abline(h = 0, lty = 2)
```

```r
acf(Yt, ylim = range(-1,1), lag.max = 50, main = "")
```

```r
Y <- diff(Yt, 12)
acf(Y, lag.max = 50, ylim = range(-1,1), main = "")
```

```r
Y <- diff(Y)
acf(Y, lag.max = 50, ylim = range(-1,1), main = "")
pacf(Y, lag.max = 50, ylim = range(-1,1), main = "")
```

```
Q <- 1; D <- 1; d <- 1
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,6,3)
for(p in 0:5){
  for(q in 0:2){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:2),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:2),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1    q=2
## p=0 45.8004 14.0574 8.5668
## p=1 27.7297  0.0000 0.7417
## p=2 19.3423  0.9690 3.2165
## p=3 18.3402  2.3349 4.3262
## p=4 20.3401  4.2806 5.3863
## p=5 18.4429  2.1290 4.1233
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2
## p=0 39.4256 10.8701  8.5668
## p=1 24.5424  0.0000  3.9291
## p=2 19.3423  4.1564  9.5913
## p=3 21.5276  8.7097 13.8883
## p=4 26.7149 13.8428 18.1359
## p=5 28.0050 14.8785 20.0602
```

```
Q <- 0; D <- 1; d <- 1
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,6,3)
for(p in 0:5){
  for(q in 0:2){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:2),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:2),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2
## p=0 42.4334 14.3672 13.1460
## p=1 22.0729  1.1284  0.2310
## p=2 18.2784  0.0000  1.9642
## p=3 19.6274  1.9100  3.8002
## p=4 21.5892  2.5470  5.9100
## p=5 17.5566  0.8629  2.7423
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2
## p=0 34.9302 10.0515 12.0176
## p=1 17.7572  0.0000  2.2900
## p=2 17.1501  2.0590  7.2106
## p=3 21.6865  7.1565 12.2340
## p=4 26.8356 10.9808 17.5312
## p=5 25.9904 12.4841 17.5509
```

```r
Q <- 1; D <- 1; d <- 1
P <- 0
P1AIC.mat <- P1BIC.mat <- matrix(0,6,3)
for(p in 0:5){
  for(q in 0:2){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:2),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:2),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1    q=2
## p=0 44.7789 16.3553 9.5097
## p=1 27.3952  0.0000 0.9427
## p=2 21.1075  1.0160 2.8507
## p=3 20.9463  2.8057 4.7718
## p=4 22.9231  4.4707 5.7668
## p=5 19.9696  2.1111 4.1109
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2
## p=0 38.4041 13.1680  9.5097
## p=1 24.2078  0.0000  4.1301
## p=2 21.1075  4.2034  9.2254
## p=3 24.1337  9.1805 14.3340
## p=4 29.2979 14.0329 18.5163
## p=5 29.5318 14.8606 20.0478
```

```r
fit1 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit1.resids <- residuals(fit1)
fit1.fitted <- Yt - fit1.resids
fit1.fitted <- fit1.fitted + mt

sum((fit1.fitted - Y.fl)^2)
```

```
## [1] 6.995852e+14
```

```r
fit2 <- arima(Yt, order = c(2,1,1), seasonal = list(order=c(1,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
fit2.resids <- residuals(fit2)
fit2.fitted <- Yt - fit2.resids
fit2.fitted <- fit2.fitted + mt

sum((fit2.fitted - Y.fl)^2)
```

```
## [1] 7.004011e+14
```

```
fit3 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit3.resids <- residuals(fit3)
fit3.fitted <- Yt - fit3.resids
fit3.fitted <- fit3.fitted + mt

sum((fit3.fitted - Y.fl)^2)
```

```
## [1] 7.060875e+14
```

```
fit4 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit4.resids <- residuals(fit4)
fit4.fitted <- Yt - fit4.resids
fit4.fitted <- fit4.fitted + mt

sum((fit4.fitted - Y.fl)^2)
```

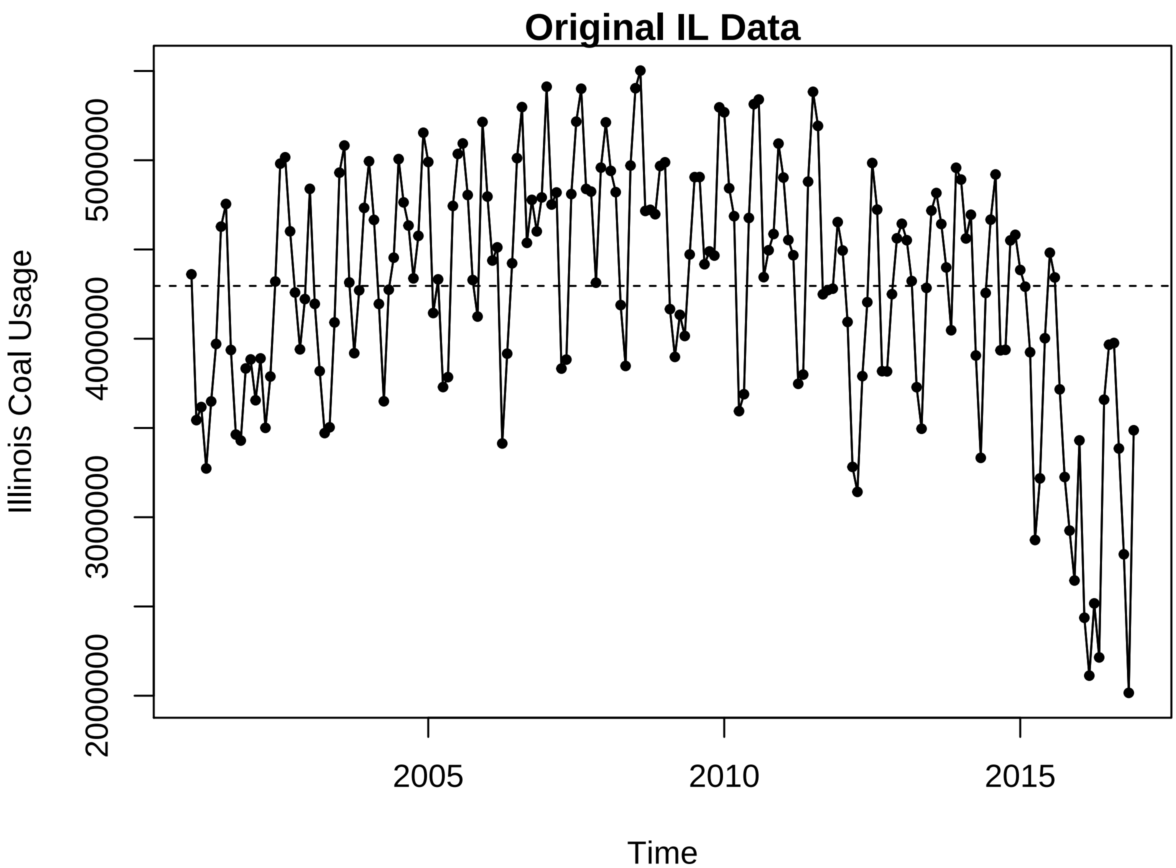
```
## [1] 6.992642e+14
```

```
fit1 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit1.resids <- residuals(fit1)
fit1.fitted <- Yt - fit1.resids
fit1.fitted <- fit1.fitted + mt
par(mar=c(4,4,1,0));plot(Y.fl,type='l', main = "Florida vs. Fitted", xlab = "Time", ylab = "Coal Usage")
lines(Y.fl,pch=19,cex=0.5)
points(fit1.fitted, pch = 19, cex = 0.4, col = 'red');lines(fit1.fitted,col='red')
legend(2001, 1500000, c("Florida", "Fitted"), col = c("black", "red"), pch = 19, lty = 1)
```

```
pred1<- predict(fit1, n.ahead = 12)$pred
t <- c(193:204)
m <- Y.fl.fit$coef[1] + Y.fl.fit$coef[2]*t
pred1 <- pred1 + m
pred1
```

Illinois Analysis

```
par(mar = c(4,4,1,2))
plot(Y.il, ylab = "Illinois Coal Usage", xlab = "Time", main = "Original IL Data"); lines(Y.il); points
```




```
n <- length(Y.il)
tvec <- c(1:n)
Y.il.fit <- lm(Y.il~tvec+I(tvec^2))
mt <- Y.il.fit$coef[1] + Y.il.fit$coef[2]*tvec + Y.il.fit$coef[3]*(tvec^2)
mt <- ts(mt, start = c(2001,1), frequency = 12)
Yt <- Y.il - mt

plot(Yt, ylab = "Illinois Coal Usage", xlab = "Time", main = "Illinois Detrended"); lines(Yt); points(Y
```

```
acf(Yt, ylim = range(-1,1), lag.max = 50, main = "")
```

```
Y <- diff(Yt, 12)
acf(Y, ylim = range(-1,1), lag.max = 50, main = "")
```

Again similar to what we saw in Florida, this seasonal differencing still does not fully get us to a process that seems stationary so we once again difference at lag one and plot the acf and pacf.

```
Y <- diff(Y)
acf(Y, ylim = range(-1,1), lag.max = 50, main = "")
pacf(Y, ylim = range(-1,1), lag.max = 50, main = "")
```

```
Q <- 1; D <- 1; d <- 1
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,2,2)
```

```r
for(p in 0:1){
  for(q in 0:1){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
rownames(P1AIC.mat)<-paste("p=",c(0:1),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:1),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:1),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:1),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1
## p=0 45.8004 14.0574
## p=1 27.7297  0.0000
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1
## p=0 39.4256 10.8701
## p=1 24.5424  0.0000
```

```r
Q <- 0; D <- 1; d <- 1
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,6,6)
for(p in 0:5){
  for(q in 0:5){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:5),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:5),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
```

```
## p=0 50.5021 22.4360 21.2148 16.3120 11.0814  7.4774
## p=1 30.1417  9.1971  8.2998  9.7234  9.3359  9.2657
## p=2 26.3472  8.0688 10.0329 12.1771  3.5121 11.1253
## p=3 27.6962  9.9788 11.8690 11.2861  0.0000  7.0153
## p=4 29.6580 10.6158 13.9788  1.5126 11.3503  5.6559
## p=5 25.6254  8.9317 10.8111 12.8122  3.4821  4.9544
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 34.9302 10.0515 12.0176 10.3023  8.2591  7.8425
## p=1 17.7572  0.0000  2.2900  6.9011  9.7010 12.8181
## p=2 17.1501  2.0590  7.2106 12.5421  7.0645 17.8651
## p=3 21.6865  7.1565 12.2340 14.8386  6.7398 16.9425
## p=4 26.8356 10.9808 17.5312  8.2524 21.2774 18.7704
## p=5 25.9904 12.4841 17.5509 22.7394 16.5966 21.2563
```

```r
Q <- 1; D <- 1; d <- 1
P <- 0
P1AIC.mat <- P1BIC.mat <- matrix(0,6,6)
for(p in 0:5){
  for(q in 0:5){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:5),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:5),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2    q=3     q=4    q=5
## p=0 48.7113 20.2878 13.4421 8.9708  8.3179 4.9686
## p=1 31.3276  3.9325  4.8752 6.8713  7.7346 6.9412
## p=2 25.0399  4.9485  6.7831 8.3766 10.2715 8.9392
## p=3 24.8788  6.7382  8.7043 5.0173  6.1809 0.7521
## p=4 26.8555  8.4031  9.6992 5.3474  7.0165 2.3368
## p=5 23.9021  6.0435  8.0433 9.9802  2.5401 0.0000
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 38.4041 13.1680  9.5097  8.2258 10.7602 10.5983
## p=1 24.2078  0.0000  4.1301  9.3136 13.3643 15.7583
## p=2 21.1075  4.2034  9.2254 14.0063 19.0885 20.9437
## p=3 24.1337  9.1805 14.3340 13.8344 18.1854 15.9439
## p=4 29.2979 14.0329 18.5163 17.3518 22.2084 20.7161
## p=5 29.5318 14.8606 20.0478 25.1721 20.9193 21.5666
```

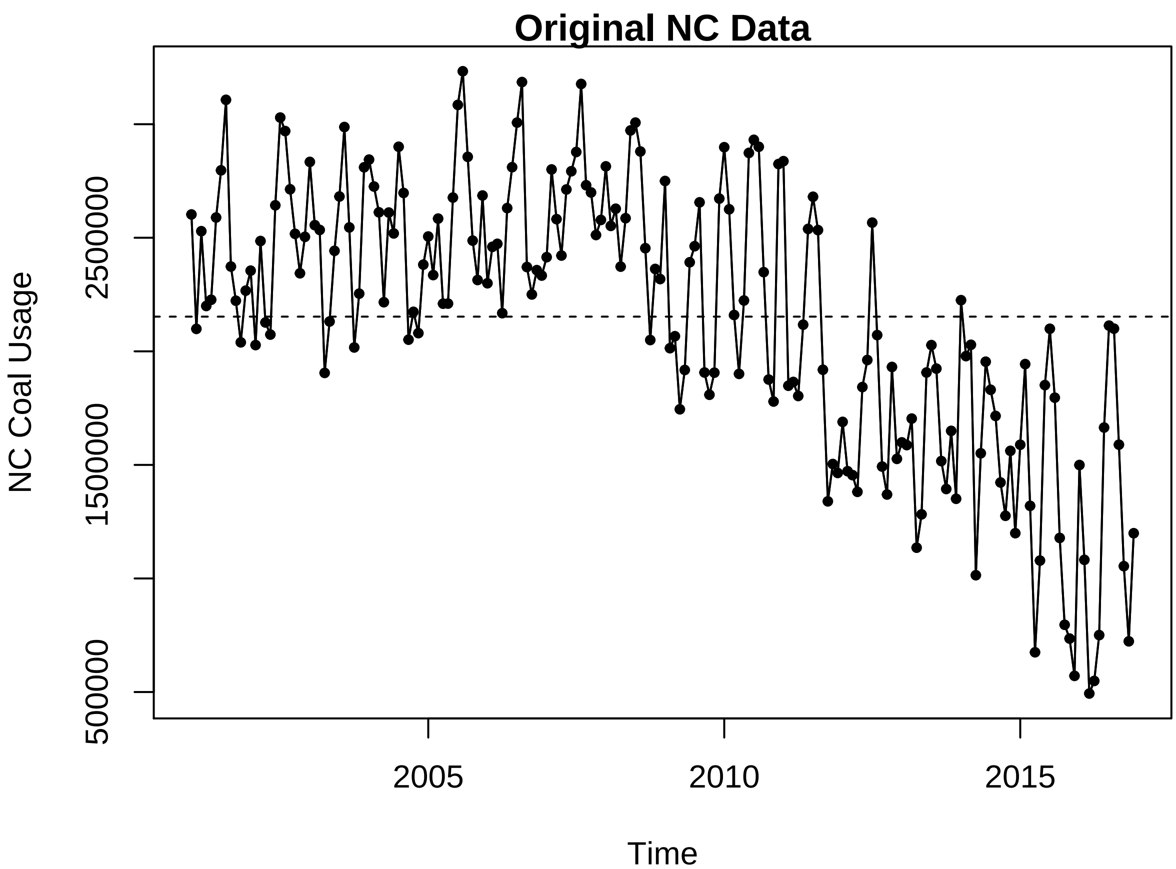
```
fit1 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit1.resids <- residuals(fit1)
fit1.fitted <- Yt - fit1.resids
fit1.fitted <- fit1.fitted + mt

sum((fit1.fitted - Y.il)^2)

fit2 <- arima(Yt, order = c(3,1,1), seasonal = list(order=c(1,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit2.resids <- residuals(fit2)
fit2.fitted <- Yt - fit2.resids
fit2.fitted <- fit2.fitted + mt

sum((fit2.fitted - Y.il)^2)

fit3 <- arima(Yt, order = c(3,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit3.resids <- residuals(fit3)
fit3.fitted <- Yt - fit3.resids
fit3.fitted <- fit3.fitted + mt

sum((fit3.fitted - Y.il)^2)

fit4 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit4.resids <- residuals(fit4)
fit4.fitted <- Yt - fit4.resids
fit4.fitted <- fit4.fitted + mt

sum((fit4.fitted - Y.il)^2)
```

```
fit <- arima(Yt, order = c(3,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit.resids <- residuals(fit)
fit.fitted<- Yt - fit.resids
fit.fitted <- fit.fitted + mt
par(mar=c(4,4,1,0));plot(Y.il,type='l', main = "Illinois vs. Fitted", xlab = "Time", ylab = "Coal Usage
lines(Y.il,pch=19,cex=0.5)
points(fit.fitted, pch = 19, cex = 0.4, col = 'red');lines(fit.fitted,col='red')
legend(2002, 3000000, c("Illinois", "Fitted"), col = c("black", "red"), pch = 19, lty = 1)
```

```
pred1<- predict(fit, n.ahead = 12)$pred
t <- c(193:204)
m <- Y.il.fit$coef[1] + Y.il.fit$coef[2]*t + Y.il.fit$coef[3]*(t^2)
pred1 <- pred1 + m
pred1
```

North Carolina Analysis

```
par(mar = c(4,4,1,2))
plot(Y.nc, ylab = "NC Coal Usage", xlab = "Time", main = "Original NC Data"); lines(Y.nc); points(Y.nc,
```




```
n <- length(Y.nc)
tvec <- c(1:n)
Y.nc.fit <- lm(Y.nc~tvec+I(tvec^2))
mt <- Y.nc.fit$coef[1] + Y.nc.fit$coef[2]*tvec + Y.nc.fit$coef[3]*(tvec^2)
mt <- ts(mt, start = c(2001,1), frequency = 12)

Yt <- Y.nc - mt

plot(Yt, ylab = "NC Coal Usage", xlab = "Time", main = "North Carolina Detrended"); lines(Yt); points(Y
```

```
acf(Yt, ylim = range(-1,1), lag.max = 50, main = "")
```

```r
Y <- diff(Yt, 12)
acf(Y, ylim = range(-1,1), lag.max = 50, main = "")
```

```r
Y <- diff(Y)
acf(Y, ylim = range(-1,1), lag.max = 50, main = "")
pacf(Y, ylim = range(-1,1), lag.max = 50, main = "")
```

```r
Q <- 1; D <- 1; d <- 1;
P <- 0
P1AIC.mat <- P1BIC.mat <- matrix(0,5,5)
for(p in 0:4){
  for(q in 0:4){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
rownames(P1AIC.mat)<-paste("p=",c(0:4),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:4),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:4),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:4),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1    q=2    q=3    q=4
## p=0 44.7789 16.3553 9.5097 5.0384 4.3854
## p=1 27.3952  0.0000 0.9427 2.9388 3.8021
## p=2 21.1075  1.0160 2.8507 4.4442 6.3390
## p=3 20.9463  2.8057 4.7718 1.0848 2.2485
## p=4 22.9231  4.4707 5.7668 1.4149 3.0840
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4
## p=0 38.4041 13.1680  9.5097  8.2258 10.7602
## p=1 24.2078  0.0000  4.1301  9.3136 13.3643
## p=2 21.1075  4.2034  9.2254 14.0063 19.0885
## p=3 24.1337  9.1805 14.3340 13.8344 18.1854
## p=4 29.2979 14.0329 18.5163 17.3518 22.2084
```

```r
Q <- 0; D <- 1; d <- 1;
P <- 0
P1AIC.mat <- P1BIC.mat <- matrix(0,5,5)
for(p in 0:4){
  for(q in 0:4){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
```

```
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
rownames(P1AIC.mat)<-paste("p=",c(0:4),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:4),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:4),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:4),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4
## p=0 50.3154 17.5157 15.2323 12.6458 11.1448
## p=1 31.7722  6.1873  5.7270 18.7674  8.3674
## p=2 21.9573  6.4127  7.6615  9.6813 11.6073
## p=3 21.0726  6.8699  8.8686  0.0000  0.6873
## p=4 23.0239  8.8647  9.8259  0.1432  2.3133
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4
## p=0 37.7533  8.1410  9.0450  9.6459 11.3322
## p=1 22.3975  0.0000  2.7271 18.9548 11.7422
## p=2 15.7700  3.4127  7.8490 13.0562 18.1696
## p=3 18.0727  7.0573 12.2434  6.5622 10.4369
## p=4 23.2114 12.2395 16.3882  9.8928 15.2503
```

```
fit1 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit1.resids <- residuals(fit1)
fit1.fitted <- Yt - fit1.resids
fit1.fitted <- fit1.fitted + mt

sum((fit1.fitted - Y.nc)^2)

fit2 <- arima(Yt, order = c(0,0,3), seasonal = list(order=c(0,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit2.resids <- residuals(fit2)
fit2.fitted <- Yt - fit2.resids
fit2.fitted <- fit2.fitted + mt

sum((fit2.fitted - Y.nc)^2)

fit3 <- arima(Yt, order = c(0,0,1), seasonal = list(order=c(0,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit3.resids <- residuals(fit3)
fit3.fitted <- Yt - fit3.resids
fit3.fitted <- fit3.fitted + mt
```

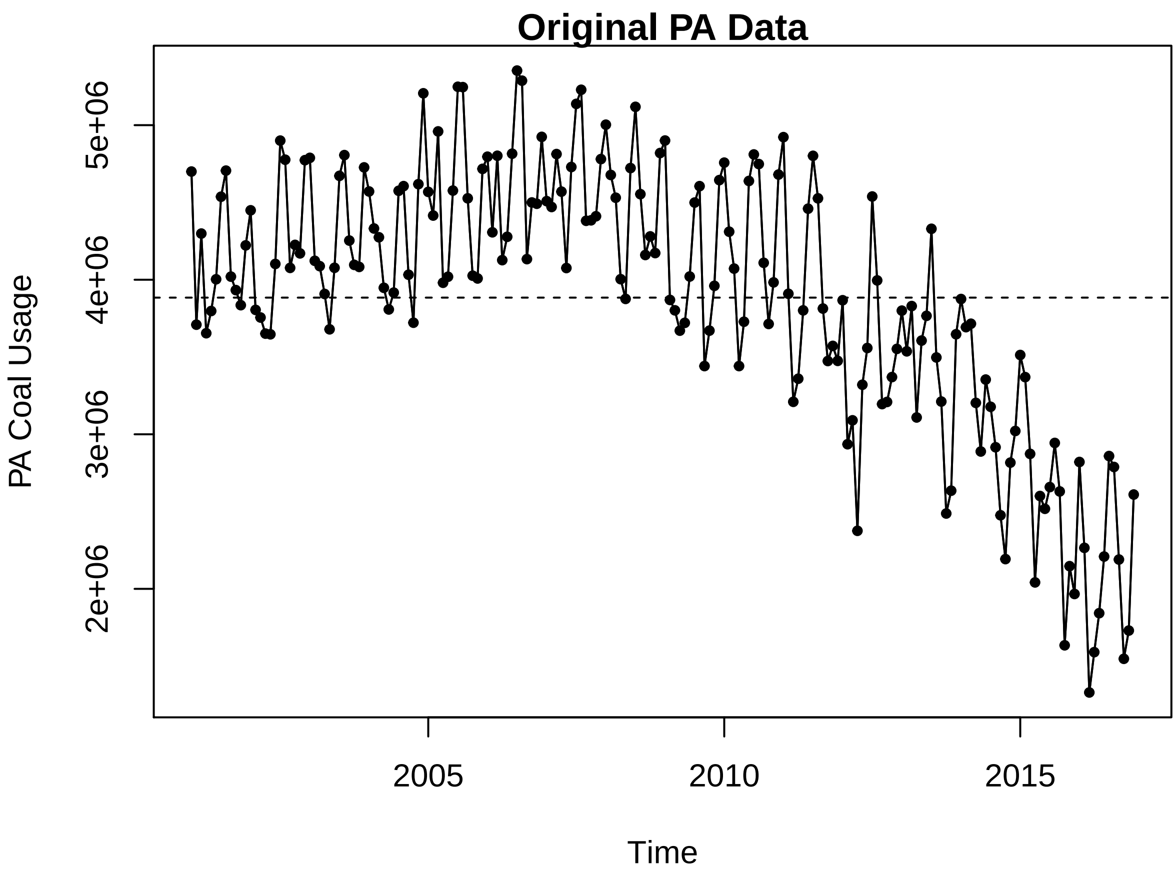
```r
sum((fit3.fitted - Y.nc)^2)

fit4 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit4.resids <- residuals(fit4)
fit4.fitted <- Yt - fit4.resids
fit4.fitted <- fit4.fitted + mt

sum((fit4.fitted - Y.nc)^2)
```

```r
fit <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit.resids <- residuals(fit)
fit.fitted <- Yt - fit.resids
fit.fitted <- fit.fitted + mt
par(mar=c(4,4,1,0));plot(Y.nc,type='l', main = "NC vs. Fitted", xlab = "Time", ylab = "Coal Usage");
lines(Y.nc,pch=19,cex=0.5)
points(fit.fitted, pch = 19, cex = 0.4, col = 'red');lines(fit.fitted,col='red')
legend(2002, 1250000, c("NC", "Fitted"), col = c("black", "red"), pch = 19, lty = 1)
```

```r
pred1<- predict(fit, n.ahead = 12)$pred
t <- c(193:204)
m <- Y.nc.fit$coef[1] + Y.nc.fit$coef[2]*t + Y.nc.fit$coef[3]*(t^2)
pred1 <- pred1 + m
pred1
```

Pennsylvania Analysis

```r
par(mar = c(4,4,1,2))
plot(Y.pa, ylab = "PA Coal Usage", xlab = "Time", main = "Original PA Data"); lines(Y.pa); points(Y.pa,
```




```
n <- length(Y.pa)
tvec <- c(1:n)
Y.pa.fit <- lm(Y.pa~tvec+I(tvec^2))
mt <- Y.pa.fit$coef[1] + Y.pa.fit$coef[2]*tvec + Y.pa.fit$coef[3]*(tvec^2)
mt <- ts(mt, start = c(2001,1), frequency = 12)

Yt <- Y.pa - mt

plot(Yt, ylab = "Pennsylvania Coal Usage", xlab = "Time", main = "Pennsylvania Detrended"); lines(Yt); p
```

```
acf(Yt, ylim = range(-1,1), lag.max = 50, main = "")
```

```
Y <- diff(Yt, 12)
acf(Y, ylim = range(-1,1), lag.max = 50, main = "")
```

```
Y <- diff(Y)
acf(Y, ylim = range(-1,1), lag.max = 50, main = "")
pacf(Y, ylim = range(-1,1), lag.max = 50, main = "")
```

```
Q <- 1; D <- 1; d <- 0;
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,6,6)
for(p in 0:5){
  for(q in 0:5){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 6), method = 'ML', opti
```

```
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:5),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:5),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 33.2256 13.9651 12.4643 12.6317  9.1969 11.1897
## p=1  8.6370  7.3722  8.9927 10.6645 10.3447 12.9075
## p=2  8.9510  9.0823 10.8961 12.9924 10.4962 11.9007
## p=3  8.8094 10.5211 12.2574  3.0554 11.8296  5.6904
## p=4 10.1880  8.8661  9.5851 11.5925  0.0000  4.2148
## p=5 11.0092 10.4107 11.5846 13.2527  4.2624 12.5424
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 21.3628  5.3280  7.0530 10.4462 10.2371 15.4557
## p=1  0.0000  1.9609  6.8071 11.7047 14.6107 20.3992
## p=2  3.5398  6.8967 11.9363 17.2584 17.9879 22.6181
## p=3  6.6239 11.5613 16.5234 10.5471 22.5471 19.6336
## p=4 11.2282 13.1320 17.0768 22.3100 13.9432 21.3838
## p=5 15.2751 17.9024 22.3021 27.1959 21.4314 32.9371
```

```
Q <- 0; D <- 1; d <- 1;
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,6,6)
for(p in 0:5){
  for(q in 0:5){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', opt
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:5),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:5),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 50.5021 22.4360 21.2148 16.3120 11.0814  7.4774
## p=1 30.1417  9.1971  8.2998  9.7234  9.3359  9.2657
## p=2 26.3472  8.0688 10.0329 12.1771  3.5121 11.1253
## p=3 27.6962  9.9788 11.8690 11.2861  0.0000  7.0153
## p=4 29.6580 10.6158 13.9788  1.5126 11.3503  5.6559
## p=5 25.6254  8.9317 10.8111 12.8122  3.4821  4.9544
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 34.9302 10.0515 12.0176 10.3023  8.2591  7.8425
## p=1 17.7572  0.0000  2.2900  6.9011  9.7010 12.8181
## p=2 17.1501  2.0590  7.2106 12.5421  7.0645 17.8651
## p=3 21.6865  7.1565 12.2340 14.8386  6.7398 16.9425
## p=4 26.8356 10.9808 17.5312  8.2524 21.2774 18.7704
## p=5 25.9904 12.4841 17.5509 22.7394 16.5966 21.2563
```

```
Q <- 1; D <- 1; d <- 1;
P <- 1
P1AIC.mat <- P1BIC.mat <- matrix(0,2,6)
for(p in 0:1){
  for(q in 0:5){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
rownames(P1AIC.mat)<-paste("p=",c(0:1),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:5),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:1),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:5),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1    q=2    q=3    q=4    q=5
## p=0 45.8004 14.0574 8.5668 4.9928 4.5684 0.5467
## p=1 27.7297  0.0000 0.7417 2.7111 3.6106 2.5264
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1    q=2    q=3     q=4     q=5
## p=0 39.4256 10.8701 8.5668 8.1802 10.9431 10.1089
## p=1 24.5424  0.0000 3.9291 9.0858 13.1727 15.2760
```

```
Q <- 1; D <- 1; d <- 1;
P <- 0
P1AIC.mat <- P1BIC.mat <- matrix(0,6,6)
for(p in 0:5){
  for(q in 0:5){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
rownames(P1AIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:5),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:5),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:5),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2    q=3     q=4    q=5
## p=0 48.7113 20.2878 13.4421 8.9708  8.3179 4.9686
## p=1 31.3276  3.9325  4.8752 6.8713  7.7346 6.9412
## p=2 25.0399  4.9485  6.7831 8.3766 10.2715 8.9392
## p=3 24.8788  6.7382  8.7043 5.0173  6.1809 0.7521
## p=4 26.8555  8.4031  9.6992 5.3474  7.0165 2.3368
## p=5 23.9021  6.0435  8.0433 9.9802  2.5401 0.0000
```

```
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4     q=5
## p=0 38.4041 13.1680  9.5097  8.2258 10.7602 10.5983
## p=1 24.2078  0.0000  4.1301  9.3136 13.3643 15.7583
## p=2 21.1075  4.2034  9.2254 14.0063 19.0885 20.9437
## p=3 24.1337  9.1805 14.3340 13.8344 18.1854 15.9439
## p=4 29.2979 14.0329 18.5163 17.3518 22.2084 20.7161
## p=5 29.5318 14.8606 20.0478 25.1721 20.9193 21.5666
```

```
Q <- 0; D <- 1; d <- 1;
P <- 0
P1AIC.mat <- P1BIC.mat <- matrix(0,5,5)
for(p in 0:4){
  for(q in 0:4){
    fit.pq <- arima(Yt, order = c(p,d,q), seasonal=list(order=c(P,D,Q), period = 12), method = 'ML', op
    P1AIC.mat[1+p, 1+q] <- fit.pq$aic
```

```r
    P1BIC.mat[1+p, 1+q] <- BIC(fit.pq)
  }
}
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```r
rownames(P1AIC.mat)<-paste("p=",c(0:4),sep="");colnames(P1AIC.mat)<-paste("q=",c(0:4),sep="")
rownames(P1BIC.mat)<-paste("p=",c(0:4),sep="");colnames(P1BIC.mat)<-paste("q=",c(0:4),sep="")

round(P1AIC.mat-min(P1AIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4
## p=0 50.3154 17.5157 15.2323 12.6458 11.1448
## p=1 31.7722  6.1873  5.7270 18.7674  8.3674
## p=2 21.9573  6.4127  7.6615  9.6813 11.6073
## p=3 21.0726  6.8699  8.8686  0.0000  0.6873
## p=4 23.0239  8.8647  9.8259  0.1432  2.3133
```

```r
round(P1BIC.mat-min(P1BIC.mat),4)
```

```
##         q=0     q=1     q=2     q=3     q=4
## p=0 37.7533  8.1410  9.0450  9.6459 11.3322
## p=1 22.3975  0.0000  2.7271 18.9548 11.7422
## p=2 15.7700  3.4127  7.8490 13.0562 18.1696
## p=3 18.0727  7.0573 12.2434  6.5622 10.4369
## p=4 23.2114 12.2395 16.3882  9.8928 15.2503
```

```r
fit1 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit1.resids <- residuals(fit1)
fit1.fitted <- Yt - fit1.resids
fit1.fitted <- fit1.fitted + mt

sum((fit1.fitted - Y.pa)^2)

fit2 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(0,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit2.resids <- residuals(fit2)
fit2.fitted <- Yt - fit2.resids
fit2.fitted <- fit2.fitted + mt

sum((fit2.fitted - Y.pa)^2)

fit3 <- arima(Yt, order = c(3,1,3), seasonal = list(order=c(0,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit3.resids <- residuals(fit3)
fit3.fitted <- Yt - fit3.resids
fit3.fitted <- fit3.fitted + mt
```

```r
sum((fit3.fitted - Y.pa)^2)

fit4 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit4.resids <- residuals(fit4)
fit4.fitted <- Yt - fit4.resids
fit4.fitted <- fit4.fitted + mt

sum((fit4.fitted - Y.pa)^2)

fit5 <- arima(Yt, order = c(3,1,4), seasonal = list(order=c(1,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit5.resids <- residuals(fit5)
fit5.fitted <- Yt - fit5.resids
fit5.fitted <- fit5.fitted + mt

sum((fit5.fitted - Y.pa)^2)

fit6 <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,0), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit6.resids <- residuals(fit6)
fit6.fitted <- Yt - fit6.resids
fit6.fitted <- fit6.fitted + mt

sum((fit6.fitted - Y.pa)^2)
```

```r
fit <- arima(Yt, order = c(1,1,1), seasonal = list(order=c(1,1,1), period = 12),
                include.mean = T, method = 'ML', optim.control = list(maxit = 1000))
fit.resids <- residuals(fit)
fit.fitted <- Yt - fit.resids
fit.fitted <- fit.fitted + mt
par(mar=c(4,4,1,0));plot(Y.pa,type='l', main = "PA vs. Fitted", xlab = "Time", ylab = "Coal Usage");
lines(Y.pa,pch=19,cex=0.5)
points(fit.fitted, pch = 19, cex = 0.6, col = 'red');lines(fit.fitted,col='red')
```

```r
pred1<- predict(fit, n.ahead = 12)$pred
t <- c(193:204)
m <- Y.pa.fit$coef[1] + Y.pa.fit$coef[2]*t + Y.pa.fit$coef[3]*(t^2)
pred1 <- pred1 + m
pred1
```