```
> rm(list=ls())
> graphics.off()
> library(astsa)
> #data <- read.csv("/Volumes/Macintosh HD/users/michaelmcphail/documents/tamu/stat626/project/fi
naldataset.csv",header=TRUE)
>
> PlotFile="C:\\Users\\Madhavi\\Documents\\MS\\626\\Proj\\Plot.pdf"
> pdf(file=PlotFile)
>
> data <- read.csv("C:\\Users\\Madhavi\\Documents\\MS\\626\\Proj\\finaldataset.csv",header=TRUE)
>
>
> #Plot the raw lumber data and independent variables
> dates <- seq(as.Date("12/01/1999", format = "%m/%d/%Y"),
+              by = "months", length = nrow(data))
>
> plot(dates, data$Lumber, xaxt="n", type="l", xlab="Month", ylab = "Lumber PPI", main = "Plot of
 Lumber PPI")
> axis.Date(side = 1, dates, format = "%m/%d/%Y")
>
> plot(dates, data$S.P_Price, xaxt = "n", type="l", xlab = "Month", ylab = "S&P Index Price", mai
n="Plot of S&P Index Price")
> axis.Date(side = 1, dates, format = "%m/%d/%Y")
>
> plot(dates, data$NewHousingStarts, xaxt = "n", type="l", xlab = "Month", ylab = "Number of New
Housing Starts", main="Plot of New Housing Starts")
> axis.Date(side = 1, dates, format = "%m/%d/%Y")
>
> #Take the first difference to create a stationary dataset
> Lumber_diff <- diff(data$Lumber)
> NHS_diff = diff(data$NewHousingStarts)
> SP_diff = diff(data$S.P_Price)
>
> #Plot the differenced lumber data and independent variables
> dates_diff <- seq(as.Date("01/01/2000", format = "%m/%d/%Y"),
+                   by = "months", length = nrow(data)-1)
>
> plot(dates_diff, Lumber_diff, xaxt="n", type="l", xlab="Month", ylab = "First Difference of Lum
ber PPI", main = "Plot of First Difference of Lumber PPI")
> axis.Date(side = 1, dates_diff, format = "%m/%d/%Y")
>
> plot(dates_diff, NHS_diff, xaxt="n", type="l", xlab="Month", ylab = "First Difference of Number
 of New Housing Starts", main = "Plot of First Difference of Number of New Housing Starts")
> axis.Date(side = 1, dates_diff, format = "%m/%d/%Y")
>
> plot(dates_diff, SP_diff, xaxt="n", type="l", xlab="Month", ylab = "First Difference of S&P Ind
ex Price", main = "Plot of First Difference of S&P Index Price")
> axis.Date(side = 1, dates_diff, format = "%m/%d/%Y")
>
> #ACF and PACF of differenced lumber data beginning 12 points after beginning of data (due to NH
S_lag11) and ending
> #12 points before end of data (so only model on training dataset)
> acf(dataLag$Lumber_diff[1:(nrow(dataLag)-12)], main="ACF of Lumber_Diff Training Data") #sugges
ts MA(1)
Error in as.ts(x) : object 'dataLag' not found
> pacf(dataLag$Lumber_diff[1:(nrow(dataLag)-12)], main="PACF of Lumber_Diff Training Data") #sugg
ests AR(1)
Error in pacf(dataLag$Lumber_diff[1:(nrow(dataLag) - 12)], main = "PACF of Lumber_Diff Training D
ata") :
  object 'dataLag' not found
>
> #Plot CCF between differenced lumber data and independent variables
> ccf(NHS_diff[1:(nrow(dataLag)-12)], Lumber_diff[1:(nrow(dataLag)-12)], main = "CCF Between Lumb
er_Diff and NHS_Diff", ylab="CCF") #Lag 11 looks predictive
Error in nrow(dataLag) : object 'dataLag' not found
> ccf(SP_diff[1:(nrow(dataLag)-12)], Lumber_diff[1:(nrow(dataLag)-12)], main = "CCF Between Lumbe
r_Diff and SP_Diff", ylab="CCF") #Lag 1 looks predictive
Error in nrow(dataLag) : object 'dataLag' not found
>
> #Create Lagged Dataset
```

```
> dataLag <- data.frame(Lumber_diff, NHS_diff, SP_diff)
> NHS_diff_lag11 <- rep(0, nrow(dataLag))
> SP_diff_lag1 <- rep(0, nrow(dataLag))
> dataLag <- data.frame(Lumber_diff, NHS_diff, SP_diff, NHS_diff_lag11, SP_diff_lag1)
>
> for (i in 12:(nrow(dataLag))) {
+   dataLag$NHS_diff_lag11[i] = dataLag$NHS_diff[(i-11)]
+ }
>
> for (i in 2:(nrow(dataLag))) {
+   dataLag$SP_diff_lag1[i] = dataLag$SP_diff[(i-1)]
+ }
>
>
> #Fit ARIMA(1,0,1) to differenced data
> fit <- arima(dataLag$Lumber_diff[1:(nrow(dataLag)-12)], order=c(1,0,1))
> fit #Fit shows that both AR terms and MA terms are NOT significant; AIC = 1023.04

Call:
arima(x = dataLag$Lumber_diff[1:(nrow(dataLag) - 12)], order = c(1, 0, 1))

Coefficients:
         ar1     ma1  intercept
      0.0210  0.3265     0.0815
s.e.  0.2189  0.2082     0.3570

sigma^2 estimated as 13.64:  log likelihood = -534.28,  aic = 1076.56
>
> #Fit ARIMA(0,0,1) to differenced data
> fit2 <- arima(dataLag$Lumber_diff[1:(nrow(dataLag)-12)], order = c(0,0,1))
> fit2 #MA term is significant; AIC = 1021.05

Call:
arima(x = dataLag$Lumber_diff[1:(nrow(dataLag) - 12)], order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.3453     0.0809
s.e.  0.0670     0.3545

sigma^2 estimated as 13.64:  log likelihood = -534.28,  aic = 1074.57
> shapiro.test(fit2$residuals) #Suggests non-normality in residuals

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.98366, p-value = 0.02234

>
> acf(fit2$residuals, main="ACF of MA(1) Model") #Looks good
> pacf(fit2$residuals, main="PACF of MA(1) Model") #Looks good
>
> #Fit ARIMA(1,0,0) to differenced data
> fit3 <- arima(dataLag$Lumber_diff[1:(nrow(dataLag)-12)], order = c(1,0,0))
> fit3 #AR term is significant; AIC = 1023.29

Call:
arima(x = dataLag$Lumber_diff[1:(nrow(dataLag) - 12)], order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.3148     0.0866
s.e.  0.0678     0.3861

sigma^2 estimated as 13.78:  log likelihood = -535.24,  aic = 1076.48
> shapiro.test(fit3$residuals) #Suggests non-normality in residuals

        Shapiro-Wilk normality test

data:  fit3$residuals
```

```
W = 0.98045, p-value = 0.007767

>
> acf(fit3$residuals, main="ACF of AR(1) Model") #Looks good
> pacf(fit3$residuals, main="PACF of AR(1) Model") #Looks good
>
> #Based on the AIC, MA model looks better than AR model.  Let's look at the Sarima
> #diagnostics
>
> sarima(dataLag$Lumber_diff[1:(nrow(dataLag)-12)], 1, 0,0)
initial  value 1.366110
iter   2 value 1.313825
iter   3 value 1.313822
iter   4 value 1.313822
iter   5 value 1.313822
iter   5 value 1.313822
iter   5 value 1.313822
final  value 1.313822
converged
initial  value 1.311873
iter   2 value 1.311870
iter   3 value 1.311869
iter   4 value 1.311869
iter   5 value 1.311869
iter   5 value 1.311869
iter   5 value 1.311869
final  value 1.311869
converged
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), xreg = xmean, include.mean = FALSE, optim.control = list(trace = trc,
    REPORT = 1, reltol = tol))

Coefficients:
         ar1   xmean
      0.3148  0.0866
s.e.  0.0678  0.3861

sigma^2 estimated as 13.78:  log likelihood = -535.24,  aic = 1076.48

$degrees_of_freedom
[1] 194

$ttable
      Estimate     SE t.value p.value
ar1     0.3148 0.0678  4.6406  0.0000
xmean   0.0866 0.3861  0.2242  0.8228

$AIC
[1] 3.643613

$AICc
[1] 3.654454

$BIC
[1] 2.677063

>
> #Now, let's consider additing predictors.  Fit MA(1) with NHS Lag 11 and SP Lag 1 as predictors
> fit.reg <- arima(dataLag[12:(nrow(dataLag)-12), "Lumber_diff"]
+            ,xreg = cbind(dataLag[12:(nrow(dataLag)-12), "NHS_diff_lag11"],
+                          dataLag[12:(nrow(dataLag)-12), "SP_diff_lag1"])
+                          ,order=c(0,0,1))
>
> fit.reg #All terms are significant except for intercept; AIC = 1013.84

Call:
arima(x = dataLag[12:(nrow(dataLag) - 12), "Lumber_diff"], order = c(0, 0, 1),
```

```
    xreg = cbind(dataLag[12:(nrow(dataLag) - 12), "NHS_diff_lag11"], dataLag[12:(nrow(dataLag) -
        12), "SP_diff_lag1"]))

Coefficients:
         ma1  intercept  cbind(dataLag[12:(nrow(dataLag) - 12), "NHS_diff_lag11"], dataLag[12:(nr
ow(dataLag) - 1
      0.2657     0.1313
         0.0531
s.e.  0.0806     0.3396
         0.0232
          12), "SP_diff_lag1"])2
                          0.0139
s.e.                      0.0057

sigma^2 estimated as 13.3:  log likelihood = -501.92,  aic = 1013.84
> shapiro.test(fit.reg$residuals) #Suggests data is normal

        Shapiro-Wilk normality test

data:  fit.reg$residuals
W = 0.98929, p-value = 0.1794

>
> acf(fit.reg$residuals, main="ACF of MA(1) + Predictors Model") #Looks good
> pacf(fit.reg$residuals, main="PACF of MA(1) + Predictors Model") #Looks good
>
> #Create forecasted values for next year and plot them along with raw data and original fitted v
alues
> fore = predict(fit.reg, n.ahead=12, newxreg = cbind(dataLag[(nrow(dataLag)-11):nrow(dataLag), "
NHS_diff_lag11"],
+                dataLag[(nrow(dataLag)-11):nrow(dataLag), "SP_diff_lag1"]))
>
> #par(mfrow=c(1,1))
> plot(dataLag$Lumber_diff[12:(nrow(dataLag))], type="l", xlab = "", xlim=c(0, 207), ylab = "Lumb
er PPI",main="Plot of Actual vs. Fitted/Forecast", xaxt="n")
> lines(dataLag$Lumber_diff[12:(nrow(dataLag)-12)] - fit.reg$residuals, type="l", col="red")
> lines(fore$pred, type="l", col="blue")
> lines(fore$pred + 1.96*fore$se, lty="dashed", col="green")
> lines(fore$pred - 1.96*fore$se, lty="dashed", col="green")
> legend(0,-10, legend=c("Actual", "Fitted", "Forecast", "PI Bounds"),
+        col=c("black", "red", "blue", "green"), lty = c(1,1,1,2),cex=0.5)
>
> dev.off()
null device
          1
>
>
>
```