CSC 425 Homework 2
Sarah Cummings

1) Given AR(1) time series process: $r_t = 0.10 + 0.4r_{t-1} + a_t$, where {a_t} is a Gaussian white noise series with mean zero and constant variance σ^2=0.02.

a. Mean of the time series rt:

$x_t = \phi_0 + \phi_1 x_{t-1} + a_t$ AR(1)= $0.10 + 0.4r_{t-1} + a_t$

Mean= $\Phi_0/(1- \Phi_1)$ 0.10/(1-0.4) = **0.166666**

b. Is AR(1) model stationary? Explain.

Yes. A necessary and sufficient condition for stationarity of a AR(1) is $|\Phi_1| < 1$. Our Φ_1 = 0.4 so we can confirm the AR(1) model is stationary.

c. With r_{100}= - 0.01 and r_{99} = 0.02 computation of 1-step and 2-step ahead forecasts

of the AR(1) series at the forecast origin t=100.

$\hat{Y}_t(1) = \mu + \phi(Y_t - \mu)$

1 step ahead=0.1666 + 0.4(0.02-0.1666) = 0.10796

2 step ahead= 0.1666 + 0.4(-0.01-0.1666) = 0.0959936

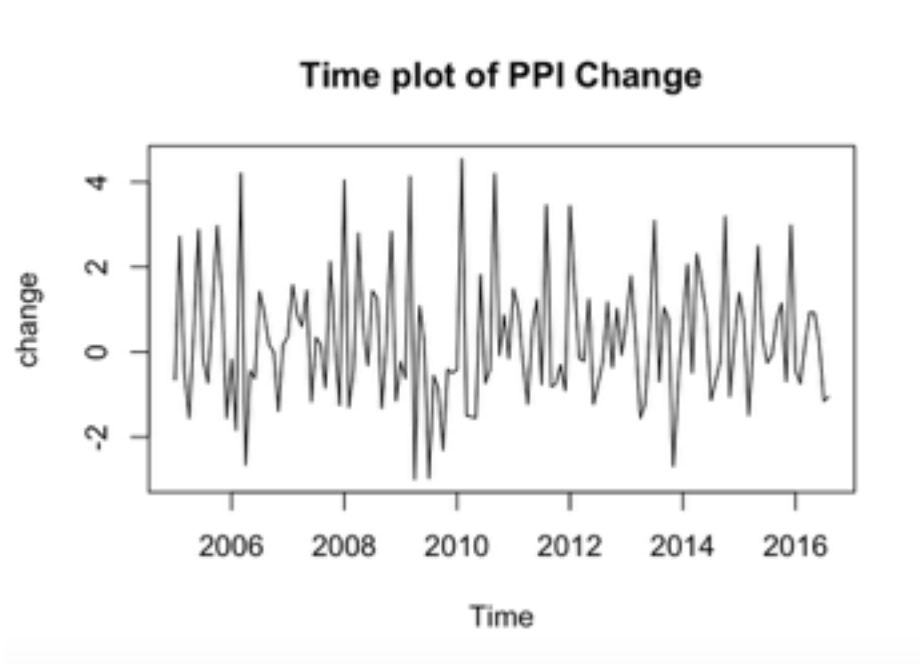
d. Lag-1 and lag-2 autocorrelations of r_t

$\rho_1 = \Phi_1$, $\rho_2 = \Phi_1^2$

lag 1= 0.4 lag 2= 0.16

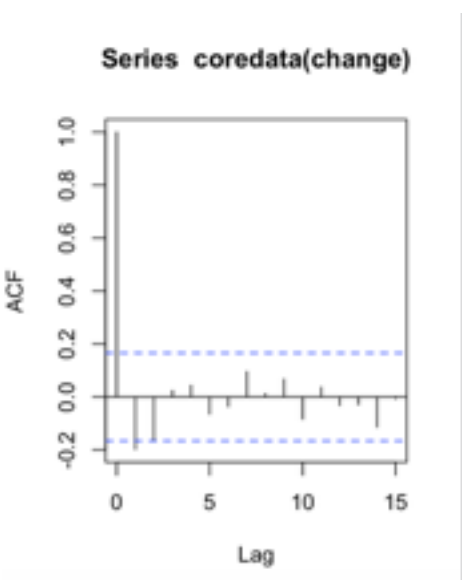
2) Producer Price Index Problem

a. Time plot of the index and analysis of the time trend displayed by the plot



It appears as though January of each year starts on a peak, with a valley in February. The series could definitely have some seasonality to it. It also looks like in more recent years, the peaks and valleys have become less drastic- signifying less change in the producer price index.

b. Analysis of serial correlation using the ACF plot and the Ljung Box test



Box-Ljung test data: change

X-squared = 9.7571, df = 3, p-value = 0.02075

Box-Ljung test data: change

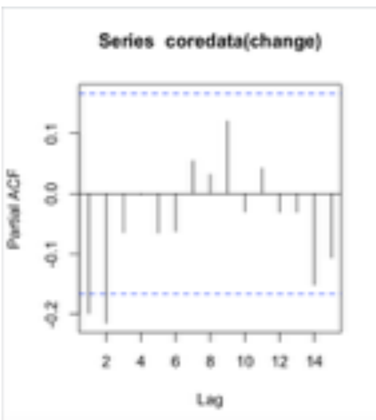
X-squared = 10.864, df = 6, p-value = 0.09267

The series shows evidence of serial correlation, with quick decay to zero visible in the ACF plot. Non-zero sample autocorrelations (significantly different than 0) occur up to lag 2, but the ACF values decay to zero as lags increase.

With our first Box Ljung test (df=3), p= 0.02 so we reject the null hypothesis of independence and conclude the change rates are serially correlated. With our second box test, however, we cannot reject the null hypothesis at the same level of certainty.

There is evidence change rates are serially correlated, but the evidence is not as strong as it was for the datasets on our previous homework assignments.

c.. Analysis of the PACF plot and identify the order “p” of the AR(p) model.



Looking at our PACF plot, we see that lags >2 are not significantly different from zero. Thus the PACF plot suggests we use an AR(2) model, or P=2.

d. Fit an adequate AR model:

Series: change
ARIMA(2,0,0) with non-zero mean

Coefficients:

	ar1	ar2	intercept
	-0.2431	**-0.2197**	**0.3036**
s.e.	**0.0827**	**0.0832**	**0.0847**

sigma^2 estimated as 2.182: log likelihood=-251.81
AIC=511.63 AICc=511.92 BIC=523.4

```
z test of coefficients:

          Estimate  Std. Error z value  Pr(>|z|)
ar1       -0.243059  0.082660 -2.9405 0.0032773 **
ar2       -0.219659  0.083231 -2.6391 0.0083117 **
intercept  0.303574  0.084713  3.5836 0.0003389 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

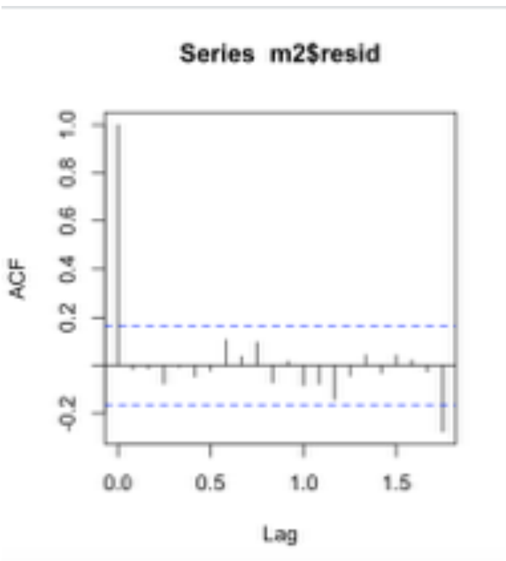
a. Significance of the model coefficients, and discussion of which coefficients are significantly different from zero.

The p value of the ar1 coefficient is 0.032, the p value of the ar2 coefficient is 0.008. Thus the coefficients for this model are all significantly different than zero at the 0.05 level.

b. Residual analysis and discussion if the selected model is adequate

i. ACF functions of residuals can be seen in graph below left.

ii. Results for testing if residuals are white noise can be seen below right.



> Box.test(m2$resid, lag=3, type='Ljung', fitdf=2)

Box-Ljung test data: m2$resid
X-squared = 0.88961, df = 1, p-value = 0.3456

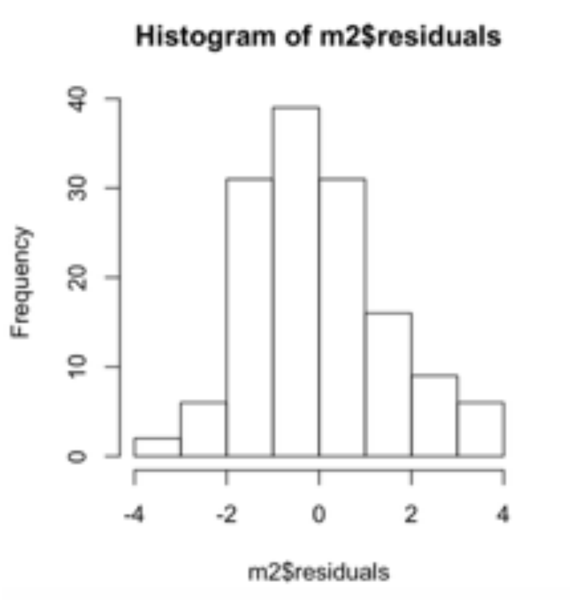
> Box.test(m2$resid, lag=6, type='Ljung', fitdf=2)

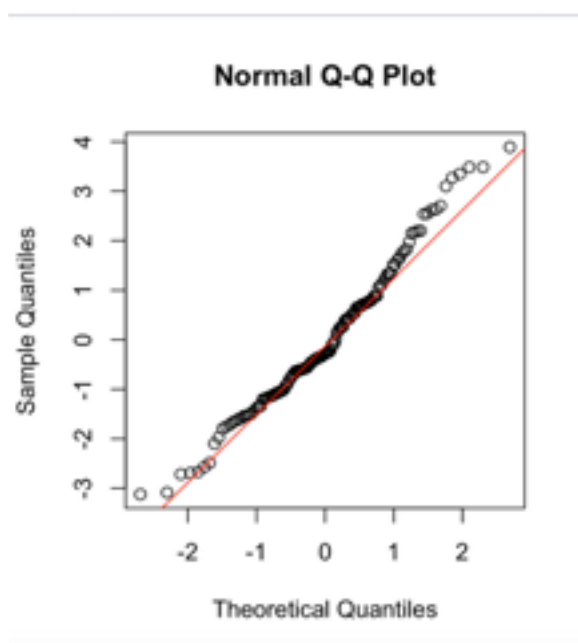
Box-Ljung test data: m2$resid
X-squared = 1.2789, df = 4, p-value = 0.865

> Box.test(m2$resid, lag=9, type='Ljung', fitdf=2)

Box-Ljung test data: m2$resid
X-squared = 4.603, df = 7, p-value = 0.7083

iii. plot histogram and normal quantile plots of residuals.





e. Discussion of results of residual analysis

Since the p values for all of the Box-Ljung tests above are all so large, we `can conclude residuals are white noise (all autocorrelations are zero). Our residual analysis ACF plot also shows no residuals are significantly different from zero. The residuals are normally distributed and centered at zero. This confirms that the model is AR(2) model is appropriate to describe the time behavior of the PPI index

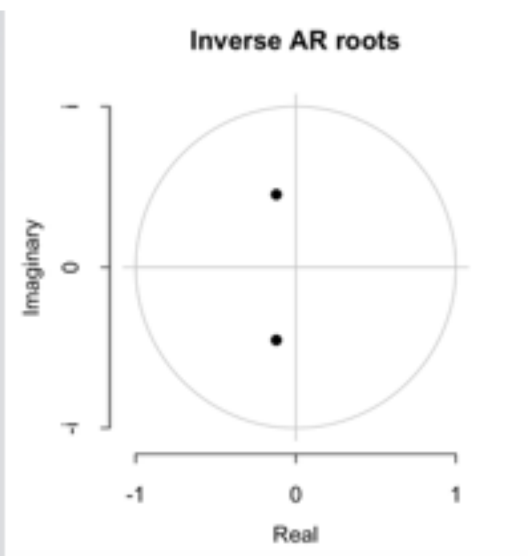
f. Expression of the estimated AR(p) model, and test if the AR model represents a stationary process. Explain.

$(X_t - 0.303574) = -0.2431(X_{t-1} - 0.303574) + -0.2197(X_{t-2} - 0.303574)$ or

Expression of model: $X_t = 0.303574 - 0.2431 X_{t-1} - 0.2197 X_{t-2}$

Thus we need roots of poly: 1+0.2431 b + 0.2197 b^2= 0 to determine stationarity.

We get: **b = -0.553254-2.06048 i** **b = -0.553254+2.06048 i**

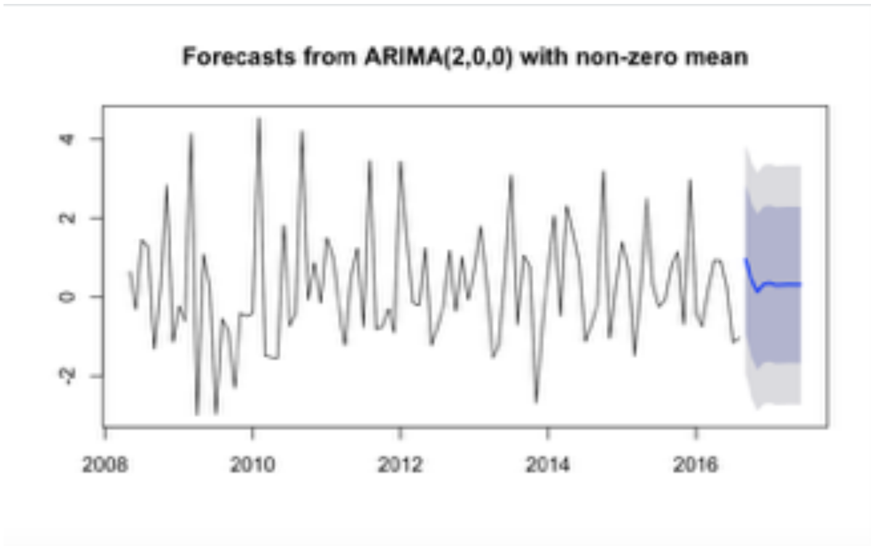


Since one of these values is not less than 1, technically the AR(2) doesn't represent a stationary process. But the problem set told us to keep going so I will :)

g) Compute up to 5-step ahead forecasts. Write down the forecasts and their 95% prediction intervals.

	Point	**Lo 95**	**Hi 95**
Sep 2016	**0.9528072**	**-1.942217**	**3.847831**
Oct 2016	**0.4425344**	**-2.536778**	**3.421847**
Nov 2016	**0.1271890**	**-2.888175**	**3.142553**
Dec 2016	**0.3159225**	**-2.711289**	**3.343134**
Jan 2017	**0.3393175**	**-2.688121**	**3.366756**

h) Plot the 10-step ahead forecasts. Discuss whether the forecasts exhibit a trend that is consistent with the observed dynamic behavior of the process



Yes, the forecasts exhibit a trend that is consistent with the observed dynamic behavior of the process. We saw in 2a that the the change in the producer price index was becoming less variant, and that is what the model is again showing us.

i) What do the model forecasts converge to?
The model forecasts converge to the mean of the process, 0.303574

CODE:

```
library(fBasics)
library(tseries)
library(zoo)

#set working directory
setwd("/Users/sarahcummings/Documents/csc425")
df<-read.csv('ppi_grocery.csv')
head(df)
change<-ts(df$change,start= c(2005,1),end = c(2016,8), frequency = 12)
plot(change, main= "Time plot of PPI Change")
```

```
#Compute Analysis of autocorrelations up to lag 15
myACF<-acf(coredata(change), plot=T, lag=15)
#pring the lags to consol
myACF

#Box test
Box.test(change,lag=3,type='Ljung')
Box.test(change,lag=6,type='Ljung')

myPACF<-pacf(coredata(change),plot = T,lag=15)

library(forecast)
#make a model
m2=Arima(change, c(2,0,0), method="ML")

library(lmtest)
coeftest(m2)
acf(m2$resid)
Box.test(m2$resid, lag=3, type='Ljung', fitdf=2)
Box.test(m2$resid, lag=6, type='Ljung', fitdf=2)
Box.test(m2$resid, lag=9, type='Ljung', fitdf=2)

#hist of residuals
hist(m2$residuals)
#normal prob plot of residuals
qqnorm(m2$residuals)
qqline(m2$residuals, col = 2)

polyroot(c(1,-m2$coef[1:2]))
plot(m2)

predict(m2,n.ahead=5, se.fit=T)
forecast.Arima(m2, h=5)
f=forecast.Arima(m2, h=10)
plot(forecast.Arima(m2, h=10), include=100)
```