Sebastian Zdarowski
Time Series DSC 425
November 9, 2019

Assignment #2

Problem 1 [20 points]

Consider the following AR(2) time series process: $r = 0.10 + 0.4r + a$, where {a_t} is a $_{t\ t-2t}$

Gaussian white noise series with mean zero and constant variance σ^2=0.02. Note that the AR(2) process has coefficient φ_1=0 for r_{t-1}

A) What is the mean of the time series r_t?

Ans: Mean = phi0/(1-phi1 – phi2 – phi3 …… phiN)

B) Determine if the AR(2) model is stationary. Explain.

Ans: Based on the three different conditions the model is stationary:

For an AR(2) model:

- **•−1<ϕ2<1**
- **•ϕ1+ϕ2<1**
- **•ϕ2−ϕ1<1**

C) Assume that r_{100}= - 0.01 and r_{99} = 0.02. Compute the 1-step and 2-step ahead forecasts of the AR(2) series at the forecast origin t=100.

rt = .10 + .40(t-2) + Et

r101 = r100(1) = .10 + .40(r99) = .108

r012 = r100(2) = .10 +.40(r100) = .096

D) Compute the lag-1 and lag-2 autocorrelations of r_t (this is just a simple application of

formulas – see page 73 in TSAR or week 3 slides)

Ans:

$$\rho_k = \frac{\text{cov}(X_t, X_{t-k})}{\text{var}(X_t)}$$

P0 = 1

P1 = phi1 / 1 – phi2 = 0 / 1 - .4 = 0

P2 = phi1 * P1 + phi2 Po = 0*0 + 0.4*(1) = 0.4

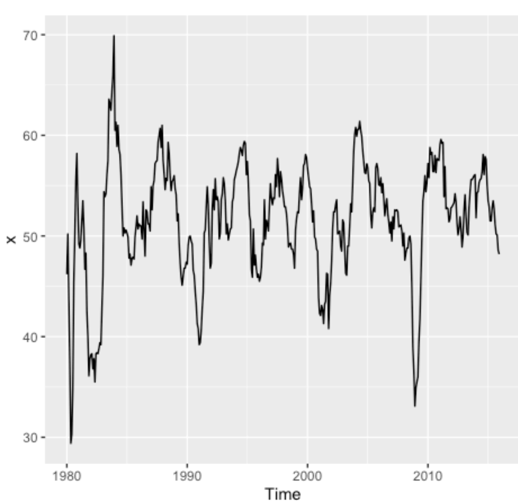
Problem 2 [40 points]

The Institute for Supply Management (ISM) (http://www.ism.ws/) is responsible for maintaining the Purchasing Managers Index (PMI). The index is derived from monthly surveys of private sector companies. Monthly data from 01/01/1980 to 12/01/2015 are saved in the NAPM.csv file. This problem asks you to apply an AR(p) autoregressive model to describe the dynamic behavior of the index.

A) Import the data in R. In R you need to create a time series object for index using the ts() function where the starting date is first month of 1980, and frequency is set equal to 12. What is the dataset size?

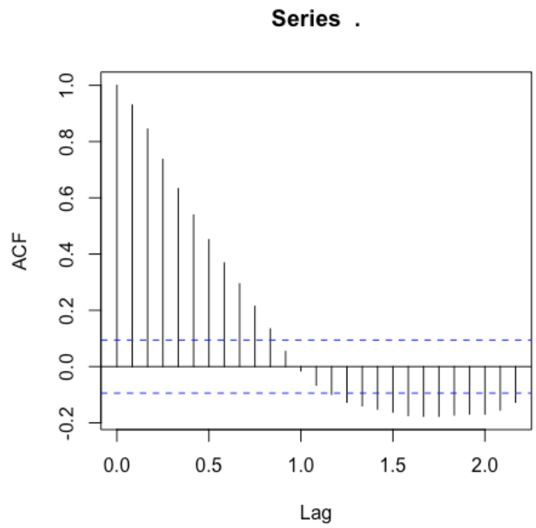
Dataset size is = to 432 Rows x 2 Columns Which contain Date column and Index column

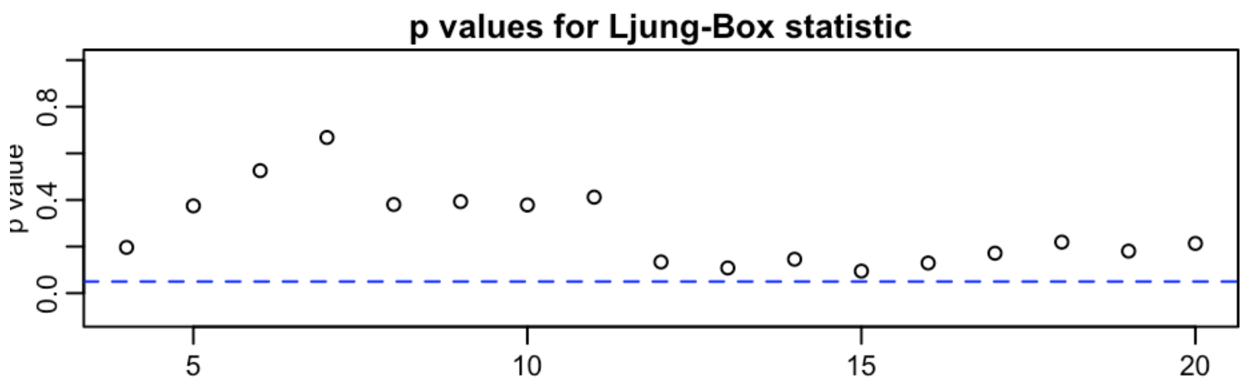
B) Create the time plot of the index and analyze the time trend displayed by the plot. What can you tell about the data based on the plot alone?



There seems to be seasonality within the model.

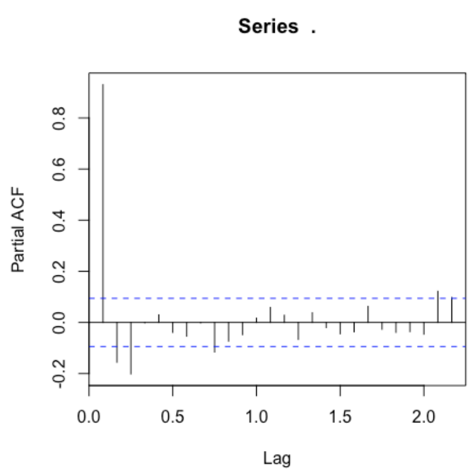
C) Analyze if the time series is serially correlated using the ACF plot and the Ljung Box test.





D) Analyze the PACF plot and identify the order “p” of the AR(p) model.

Order of P = 3 based on the PACF plot we see 3 beginning points go over the threshold



E) Fit an adequate AR model:

1. Examine the significance of the model coefficients, and discuss which coefficients are significantly different from zero.

   Fit AR model:

   m = Arima(x,order =c(3,0,0))

   coeftest(m)

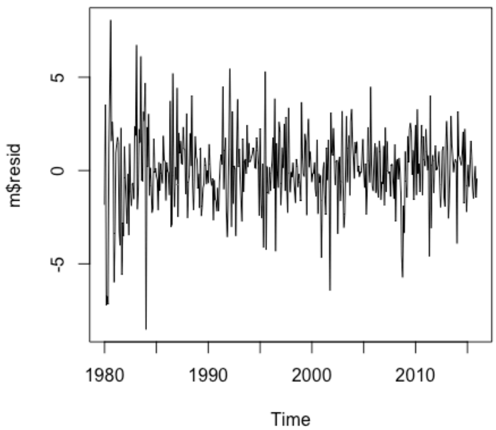
   z test of coefficients:

   ```
             Estimate Std. Error z value  Pr(>|z|)
   ar1       1.053415   0.047479 22.1868 < 2.2e-16 ***
   ar2       0.041382   0.070667  0.5856    0.5582
   ar3      -0.190335   0.048032 -3.9627 7.411e-05 ***
   intercept 51.414235   1.025118 50.1545 < 2.2e-16 ***
   ```

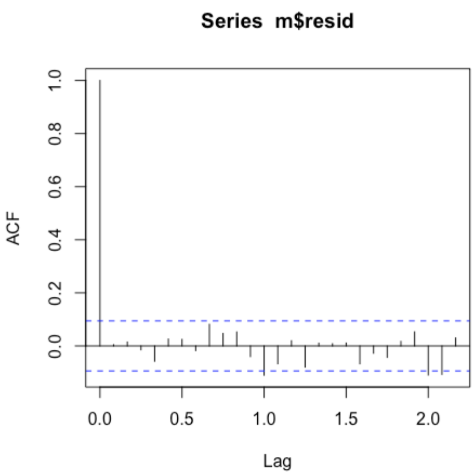
   Based on the coefficient test, the best model to use would be of using p1 and p3 since they have significant p values. Ar2 wouldn't be used since its not a significant parameter within the model.

2. Perform a residual analysis and discuss if the selected model is adequate i. compute ACF functions of residuals,

   plot(m$resid, type='l') #residual plot



Within this residual plot, we are seeing the correct results as intended when looking at the residual. We are seeing a zero mean and constant variance. Looking at white noise within the plot currently



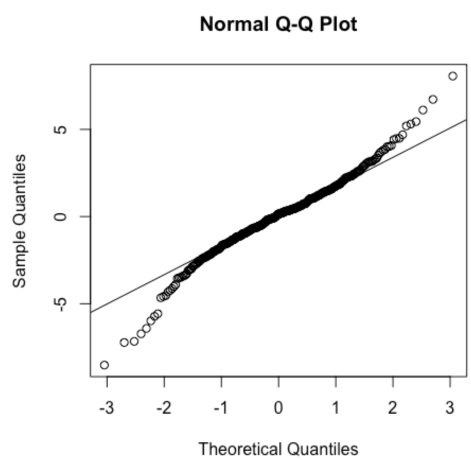
Looking at that residual ACF plot, we are not seeing any instances of serial correlation

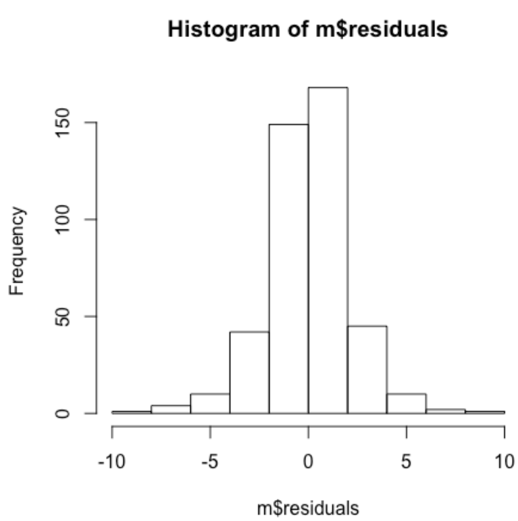
ii. test if residuals are White Noise (try several different lags),

```
        Box-Ljung test

data:  m$residuals
X-squared = 1.9638, df = 5, p-value = 0.8541
```

```
        Box-Ljung test

data:  m$residuals
X-squared = 7.5026, df = 10, p-value = 0.6773
```

iii. plot histogram and normal quantile plots of residuals.





The histogram shows the tell-tale bell-curve shape.

F) Discuss the results of your residual analysis, and draw conclusions on whether the selected model is appropriate to describe the time behavior of the PMI index.

Based on the residual analysis, we are seeing white noise within the plot, which is a good indicator. Since it showing to be independent and identically distributed.

G) Write down the expression of the estimated AR(p) model, and check if the AR model represents a stationary process. Explain your results. (NOTE: tests on the model parameters will show that one coefficient is not significantly different from zero. You can still use the model for forecasting as long as the residuals are white noise/independent)

Before when model had Ar2 included: #xt = -1.05xt-1 + .04xt-2 - 0.19xt-3

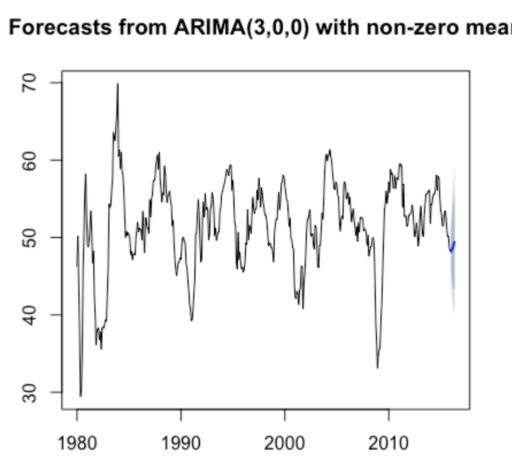
Removed AR from the model

Intercept = 51.413

Calculate intercept = 51.413(1-1.07-.16) = -11.19

Xt = 1.07(xt-1) – 0.16(xt-3) + -11.89 + Et

H) Compute up to 5-step ahead forecasts with origin at the end of the data, i.e. 12/01/2015. Write down the forecasts and their 95% prediction intervals.




```
$lower
              80%      95%
Jan 2016 45.53111 44.12661
Feb 2016 44.52758 42.46652
Mar 2016 43.79069 41.16915
Apr 2016 43.38341 40.36186
May 2016 43.16710 39.85418

$upper
              80%      95%
Jan 2016 50.83745 52.24196
Feb 2016 52.31445 54.37551
Mar 2016 53.69510 56.31664
Apr 2016 54.79910 57.82065
May 2016 55.68361 58.99653
```

I) Plot the 10-step ahead forecasts and discuss whether the forecasts exhibit a trend that is consistent with the observed dynamic behavior of the process

```
$lower
               80%      95%
Jan 2016 45.50679 44.10121
Feb 2016 44.53430 42.49273
Mar 2016 43.78580 41.18087
Apr 2016 43.37937 40.37439
May 2016 43.15998 39.86024
Jun 2016 43.07677 39.56898
Jul 2016 43.07355 39.41909
Aug 2016 43.11793 39.36146
Sep 2016 43.18663 39.35954
Oct 2016 43.26542 39.38974

$upper
               80%      95%
Jan 2016 50.81720 52.22278
Feb 2016 52.24753 54.28910
Mar 2016 53.62745 56.23238
Apr 2016 54.73244 57.73742
May 2016 55.62669 58.92642
Jun 2016 56.32952 59.83731
Jul 2016 56.88043 60.53490
Aug 2016 57.31020 61.06667
Sep 2016 57.64570 61.47278
Oct 2016 57.90807 61.78374
```

Trend seems to be occurring where values are tending to the 95% gap is wideing during the later months in the prediction. Within the earlier months we are able to get our high and low predictions to be closer margin, but as we predict further, that gap seems to spread.

J) A PMI reading above 50 percent indicates that the manufacturing economy is generally expanding; below 50 percent that it is generally declining. Do the model forecasts predict that manufacturing economy is generally expanding or contracting?

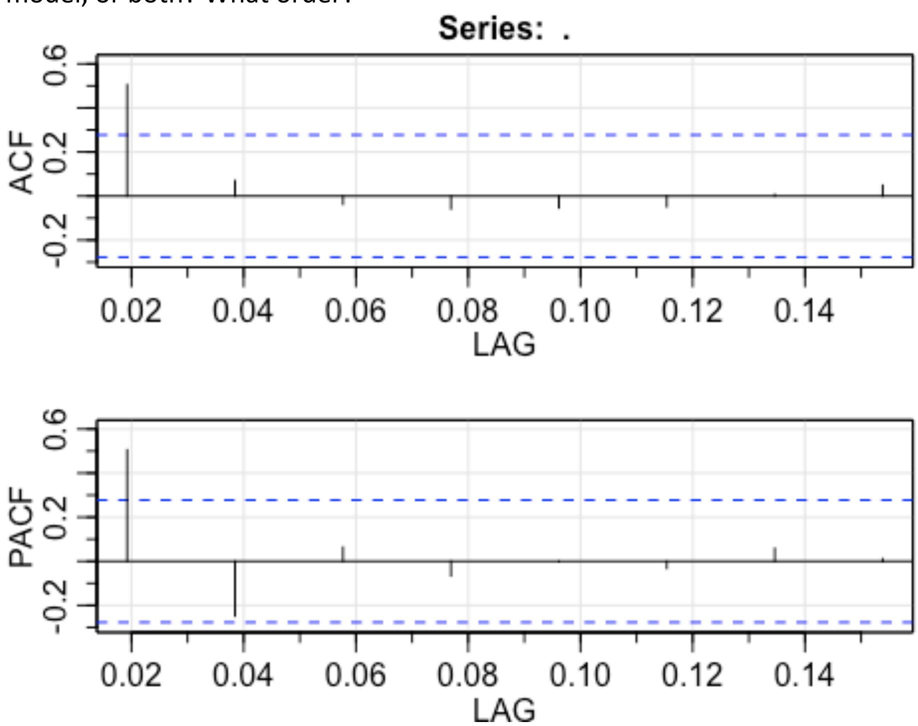
On the higher end of the 95% confidence interval we are seeing scores that are consistently above 50 point mark. Lower end we are seeing scores around 40. So if we would average all the amounts we should come out to be around the 50 point mark. Seems that they are expanding based on these numbers

K) What do the model forecasts converge to?
Model will converge to the mean after long term forecasting.

Problem 3 [40 pts]

Consider the groceries data set you analyzed in homework 1. The dataset “groceries.csv” contains weekly units sold for three grocery items: ToothPaste (100ml container of toothpaste), PeanutButter (340g. jat of crunchy peanut butter), and Biscuits (200g., 10 finger package of shortbread cookies). The variable Date is defined as the first day of the week for the sales period. The following problem focuses on building an AR model for the **toothpaste** series.

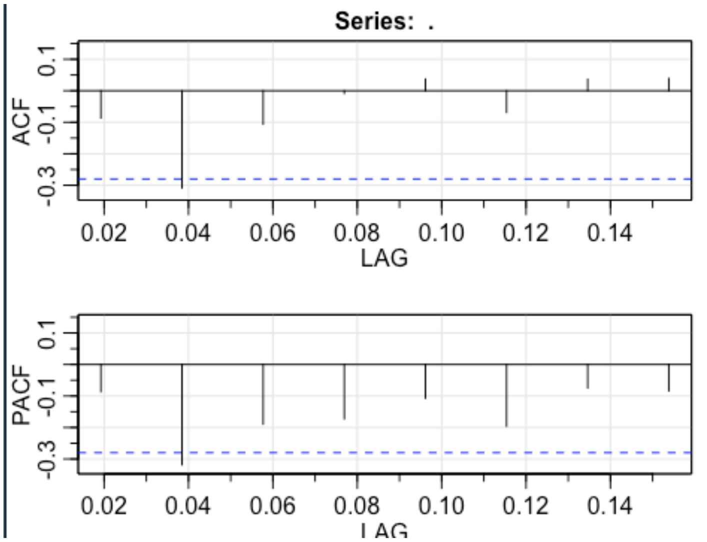
1. a) Analyze the ACF and the PACF plots for the toothpaste sales data and discuss which time series model is suggested by the plots. Do the plots suggest either an AR(p) or a MA(q) model, or both? What order?



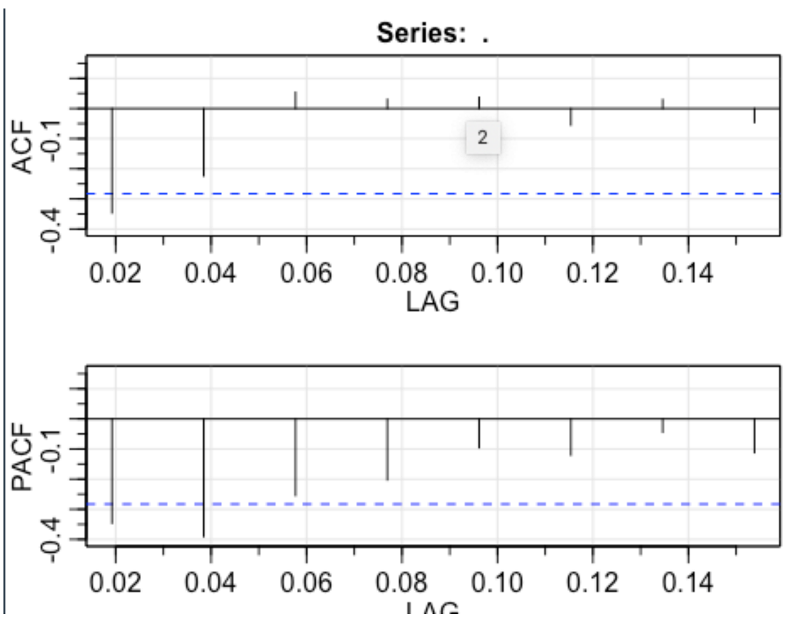
Model doesn’t seem to be any of these the models generally. But it seems to have lag 1 order from the charts.

Tried differencing the TS model for toothpaste and after taking the difference the model seems to have the order of 1, and seems to be an AR model based on the charts below.

Diff 1



Diff 2



2. b) Fit an AR model of the toothpaste sales data:
   - Examine the significance of the model coefficients, and discuss if all coefficients are significantly different from zero.

m = Arima(tpasteTS,order =c(2,0,0)) ##This model worked overall better

m = Arima(tpasteTS,order =c(2,1,0)) #### One seems to be working better with d = 1 (update(difference didn't help as orginal thought with model calculation.

coeftest(m)

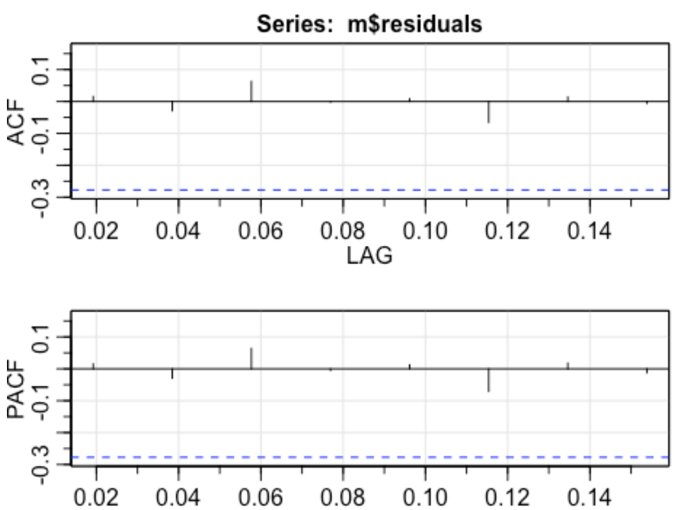
plot(m$residuals, type='l') #residual plot

m = sarima(tpasteTS,2,1,0) #use for charts

Took the difference of the model by having d =1 (taking difference once within toothpaste sales)

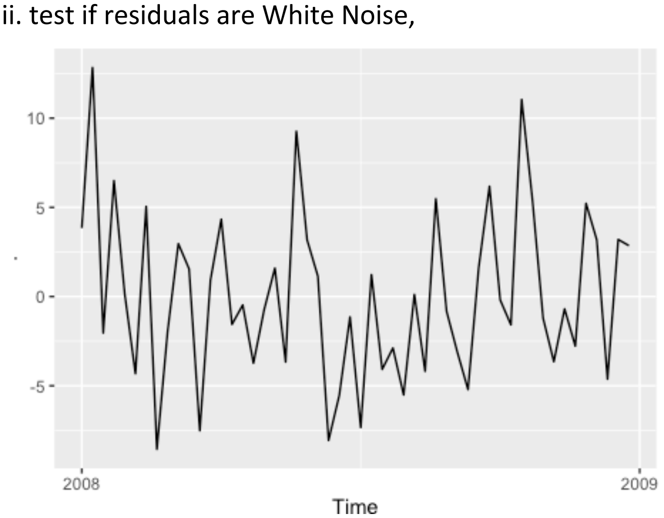
AIC score was better on model setup of m = sarima(tpasteTS,2,0,0) #use for charts AIC Score of 6.09 vs 6.32 score.

- Perform a residual analysis and discuss if the selected model is adequate i. compute ACF functions of residuals,



Not seeing any serial correlation within the charts.

ii. test if residuals are White Noise,



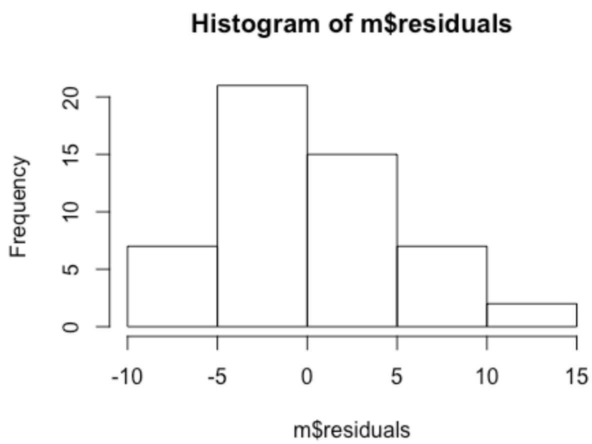
JB Test:

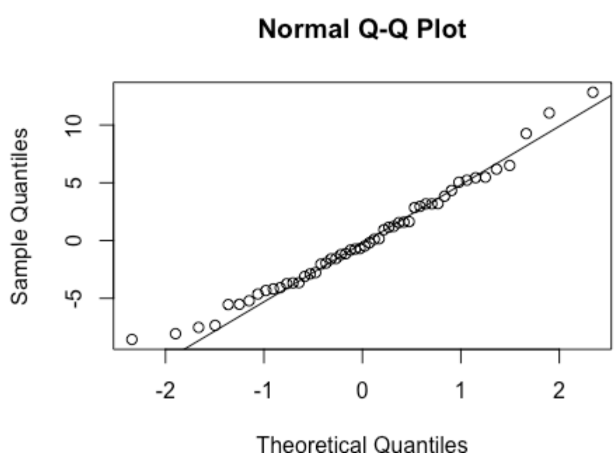
```
        Jarque Bera Test

data:  m$residuals
X-squared = 1.8748, df = 2, p-value = 0.3917

>
```

Box-Ljung test

```
        Box-Ljung test

data:  m$residuals
X-squared = 0.29154, df = 5, p-value = 0.9978
```

plot histogram and normal quantile plots of residuals.





- Discuss the results of your residual analysis, and draw conclusions on whether the selected model is appropriate to describe toothpaste sales process.

Based on the residual analysis, I don't think that this is an proper model for the toothpaste sales process. The histogram and didn't line up like the TS that we had present in problem 2, no bell shape curve on histogram. All the points on the QQ plot do not aline on the plot.

3. c) Write down the expression of the estimated AR(p) model, and test if the AR model represents a stationary process. Explain your results.

In the end, I went with Arima(2,0,0) since by taking the difference I was getting mixed results and not getting an intercept for this AR model . Also the AIC value for Armia(2,0,0) was lower than Arimia(2,1,0)

Before when model had Ar2 included: #xt = -1.05xt-1 + .04xt-2 - 0.19xt-3

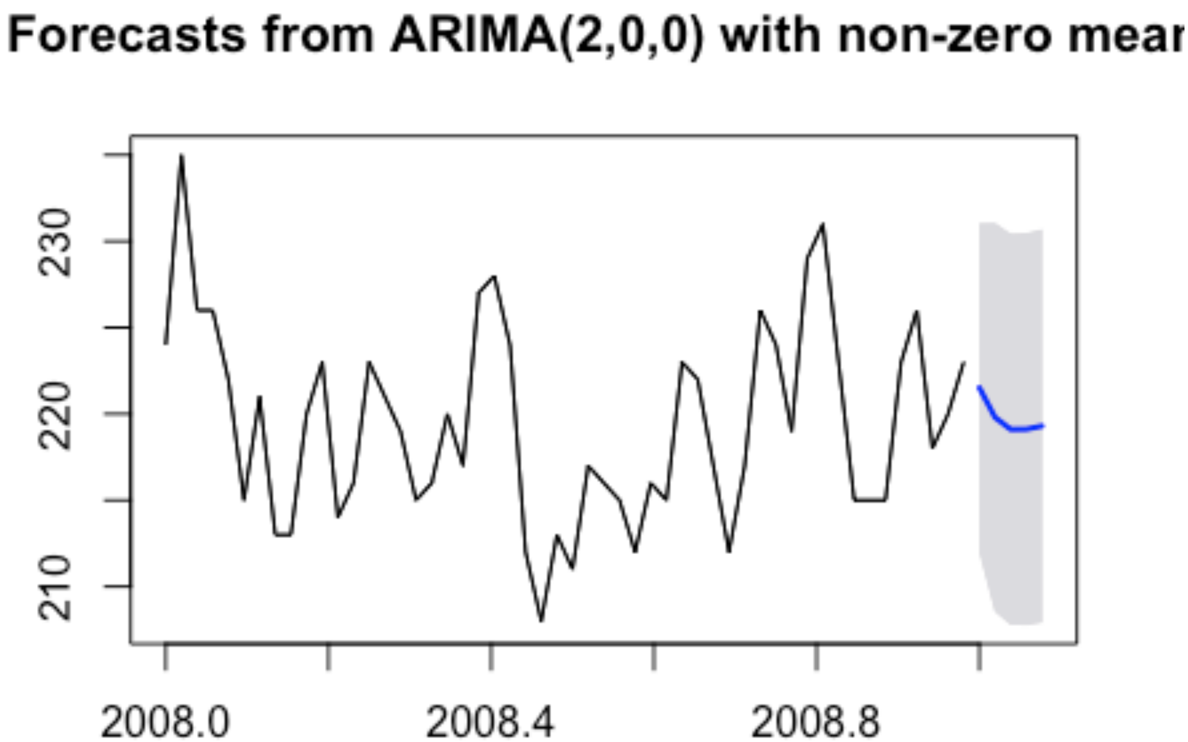
Removed AR from the model

Intercept =219.40

Calculate intercept = 51.413(1-.63-.26) = 24.134

Xt = 1.63(xt-1) – 0.26(xt-2) + 24.134+ Et

4. d) Use the selected AR model to compute up to 5 step-ahead forecasts starting from the last observation in the dataset. Write down the forecasts and their margin of errors or prediction intervals.




```
Time Series:
Start = c(2009, 1)
End = c(2009, 5)
Frequency = 52
                 95%      95%
2009.000 231.0255 231.0255
2009.019 231.0415 231.0415
2009.038 230.4206 230.4206
2009.058 230.4531 230.4531
2009.077 230.6685 230.6685
```