Sebastian Zdarowski

November 11, 2019

DSC425 – Time Series Analysis and Forecasting

Assignment 3

PROBLEMS

Problem 1 [30 pts]

Consider the following MA(3) time series process: $X_t = 5 + a_t - 0.5a_{t-1} + 0.25a_{t-2} - 0.1a_{t-3}$, where $\{a_t\}$ is a Gaussian white noise series with mean zero and constant variance $\sigma^2=0.025$.

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

$$x_t = c + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$$

Mean $E(X_t)=c$, so the mean would be 5

b) Discuss if the MA (3) model is stationary. Explain.

MA models are always stationary, if not they are not moving average models. Mean and variance are constant and auto-correlation functions decay to zero.

MA(q) models that are driven by Gaussian noise are always strictly stationary, based on the condition that the sum of MA coefficient should be finite)

For an MA(1) model: $-1<\theta_1<1$

For an MA(2) model: $-1<\theta_2<1,\ \theta_2+\theta_1>-1,\ \theta_1-\theta_2<1$**. This will hold true for Ma(3) as well**

c) Assume that $a_{100}= -0.01$ and $a_{99} = 0.02$ and $a_{98}=-0.04$. Compute the 1-step and 2-step ahead forecasts of the MA(3) series at the forecast origin t=100.

Step 1

A101 = 5 +.025 – 0.5(-0.01) + .25(.02-0.1) – 0.1(-0.04) = 5.014

A102 = 5 + .25(-0.01) + .1(0.02) = 4.9955

d) Identify the number of autocorrelation functions for the MA(3) model that are not zero.

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

$$\rho_k = 0 \ for\ k > 2$$

P1 = .5 - .5(.25) / (1 + -.5^2 + .25^2) = .285
P2 = -(.25) / 1 -.5^2 + 0.25^2 =-.3076
P3 = 0

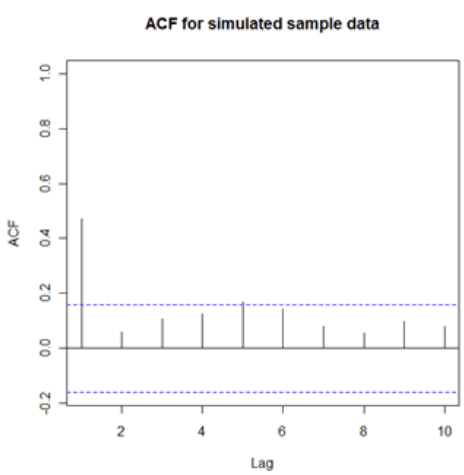
Problem 2 [10 pts]

a) Explain why the MA(q) models are regarded as finite memory models.
Has a finite memory, in the sense that observations spaced more than q time units apart are uncorrelated. Autocovariance function of ma(q) cuts off beyond lag q.

b) Discuss the ACF behavior of an ARMA(p, q) process

ACF will theoretical decrease quickly towards zero, but it never attains zero, it tails off.

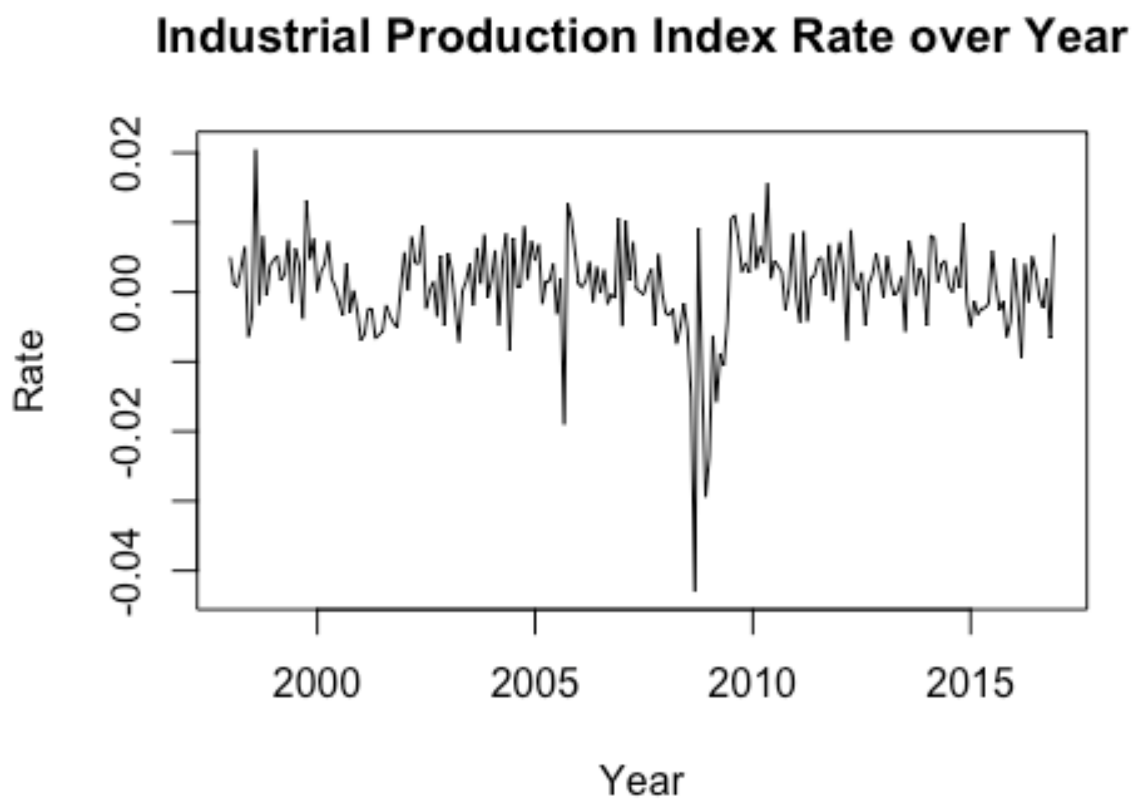
Moving AVG ACF below



Problem 3 [60 pts]

The Industrial Production Index (INDPRO) is an economic indicator that measures real output for all facilities located in the United States manufacturing, mining, and electric, and gas utilities. Since 1997, the Industrial Production Index has been determined from 312 individual series. The index is compiled on a monthly basis to bring attention to short- term changes in industrial production. Growth in the production index from month to month is an indicator of growth in the industry. Monthly percentage changes of the INDPRO index from January 1998 to December 2016 were obtained from the St Louis Federal Reserve Bank and are defined as $X_t = (p_t-p_{t-1})/p_{t-1}$ where p_t is the monthly Industrial Production Index. The dataset contains two variables: date, rate, where rate is the INDPRO index growth rate series. The following problem focuses on building a TS model for the INDPRO index growth rate series X_t.

a) Import the data either in R. In R you need to create a time object for index using the ts() function where the starting date is the first month of 1998, and frequency is set equal to 12.

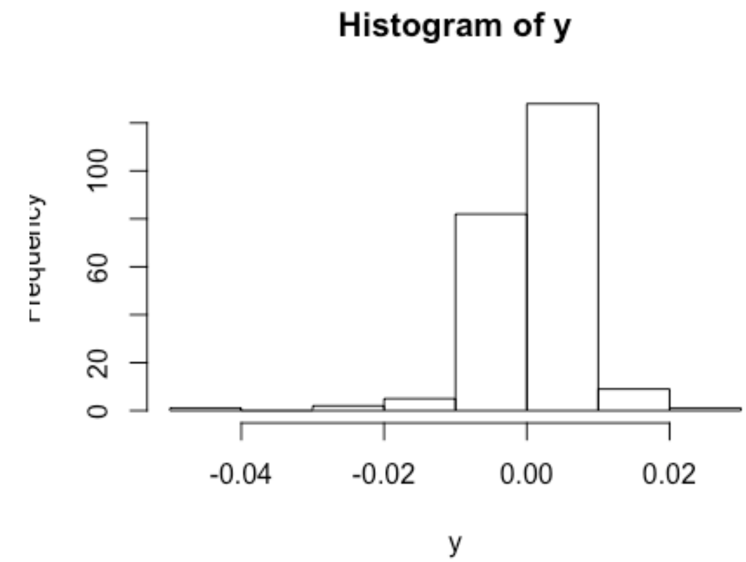


y = ts(x$rate, start=c(1998,1),frequency = 12)

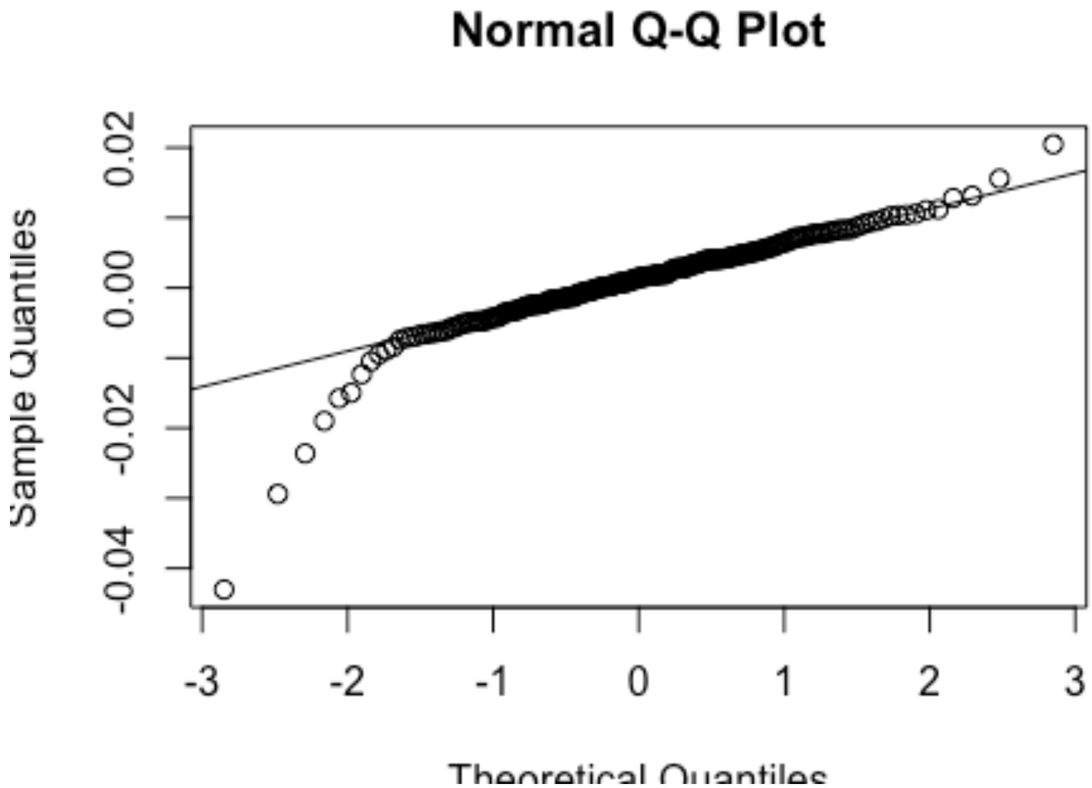
b) Create the time plot of the index growth rate Xt and analyze trends displayed by the plot?

Looking at the chart above, we see a DIP during the 2008-2009 period, which to me indicates the recession of 2008. Also within the chart, we are seeing the chart vary around .00 to .02 besides the dip, which is showing seasonality trends within the market.

c) Analyze the distribution of Xt. Can you assume that Xt is normally distributed?

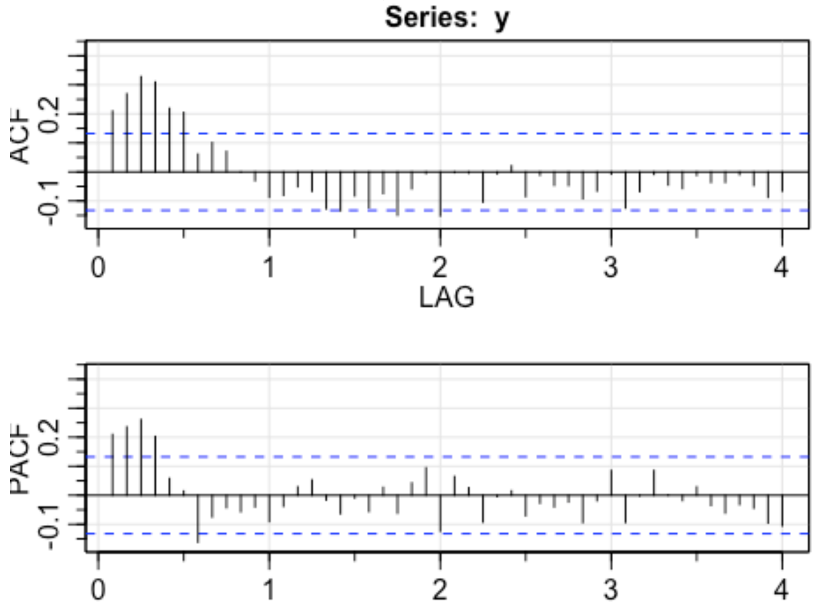


Histogram is showing skewness to the right side of the graph, it seems that its not normally distributed.



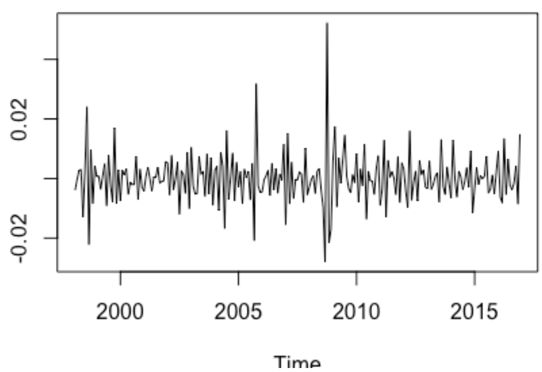
QQ plot is showing us, that most of the values are falling around .02 but we do see outliers on the data which I assume is from that recession dip.

d) Analyze the ACF and the PACF functions and discuss the following questions (a yes or no answer is not sufficient)

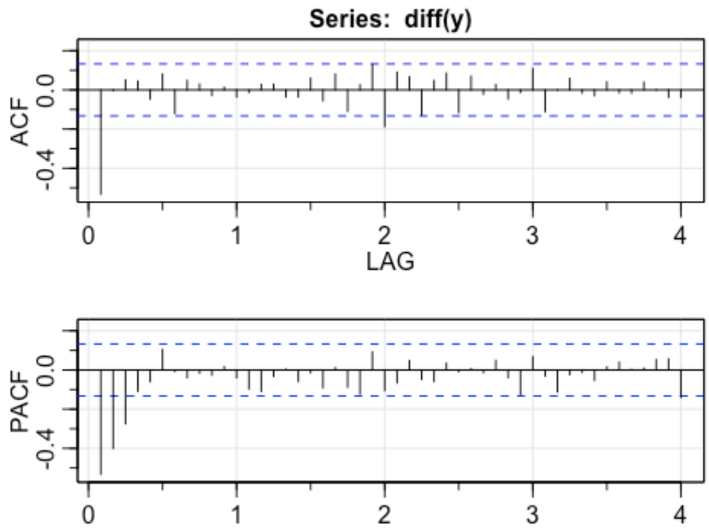


- Does the TS exhibits a stationary behavior?

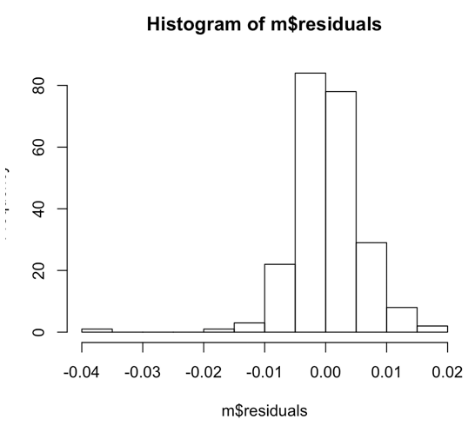
Currently without taking the difference, the time series is not showing stationary behavior, Once we took the difference from the TS, we are seeing stationary, due removing the trends from the model.

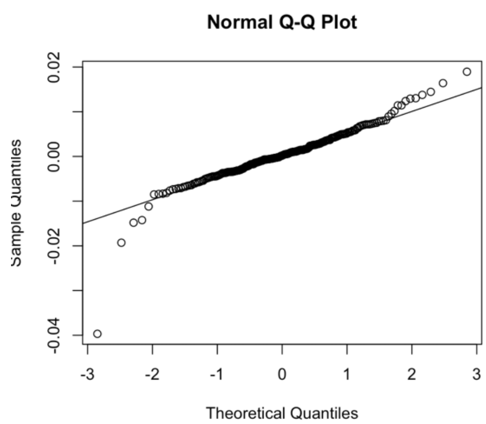


- Does the process show more of an AR behavior or MA behavior?
  Currently the process is showing more of a MA behavior, we notice this because the ACF took a dip right away instead of slowly decaying to 0 as an AR model does. Also noticing that the PACF is slowly decaying as well.

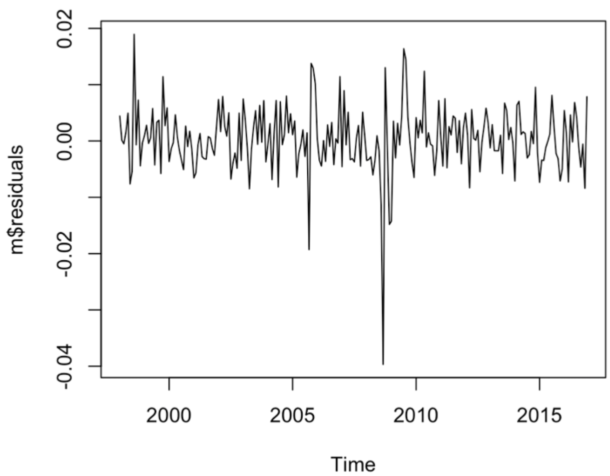


e) Apply the BIC criteria using the auto.arima(xvar, ic=c("bic)) function, and fit the model suggested by the BIC criterion. (M1)
   - Examine the significance of the model coefficients and analyze the residuals to check adequacy of the model.






```
> summary(m$residuals)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.0396827 -0.0031354  0.0001954  0.0003187  0.0035290  0.0189518
```



Mean is close to zero, which is a good sign. Histogram seems to be skewed to the right slightly and qq plot it showing most values along the line with a few outliers.

- Write down the model expression and discuss if this is a good model for the data.

```
Series: y
ARIMA(1,0,3) with zero mean

Coefficients:
         ar1      ma1     ma2     ma3
      0.8221  -0.8154  0.1251  0.1884
s.e.  0.0593   0.0825  0.0802  0.0664

sigma^2 estimated as 3.568e-05:  log likelihood=845.6
AIC=-1681.2   AICc=-1680.93   BIC=-1664.06
```

at = .8221

MA(3) time series process: $X_t = a_t - .81a_{t-1} + 0.125a_{t-2} + 0.18a_{t-3}$

f) Find an alternative model (M2) for the index growth rate time series, using either an MA(q) or an AR(p) model depending on your analysis in d). Make sure the model coefficients are significant and the residuals are white noise. Write down the model expression.

Model expression:

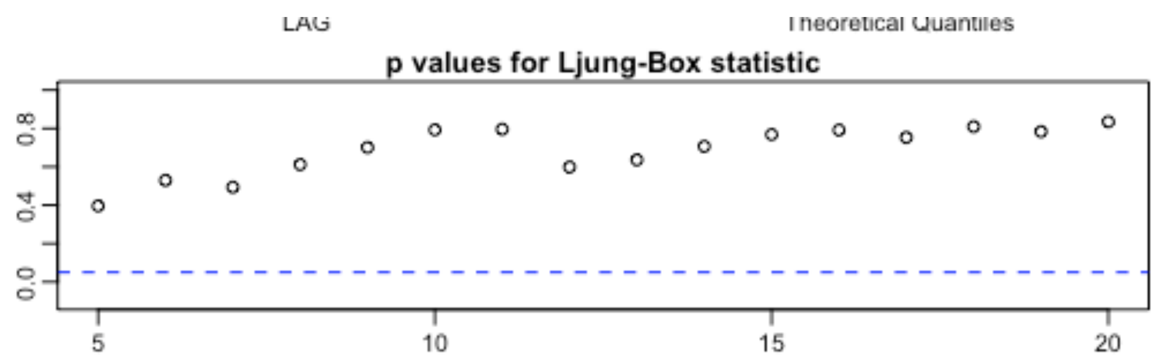
```
           Estimate  Std. Error  z value  Pr(>|z|)
ar1        0.81212363  0.06192488  13.1147 < 2.2e-16 ***
ma1       -0.74954110  0.06908602 -10.8494 < 2.2e-16 ***
ma3        0.25135120  0.05502930   4.5676 4.934e-06 ***
intercept  0.00085836  0.00103479   0.8295    0.4068
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

at = .812

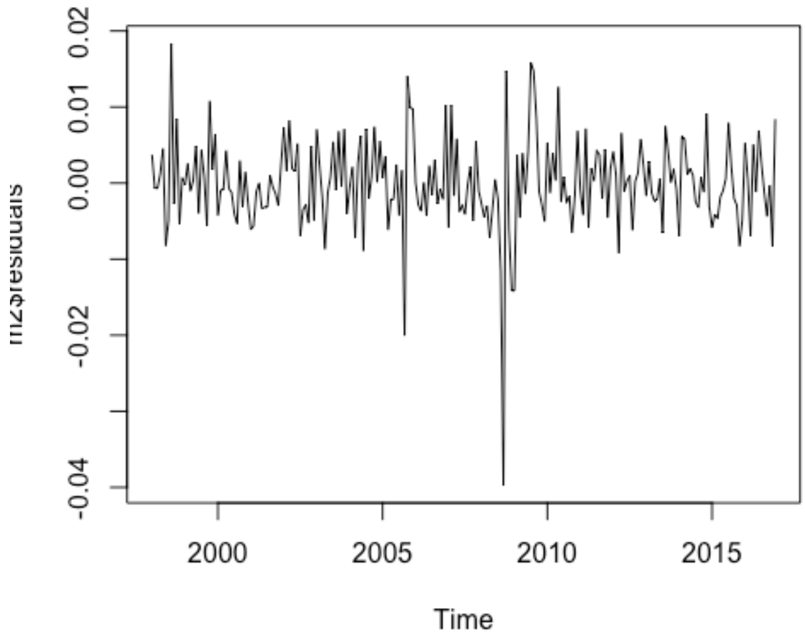
MA(3) time series process: $X_t = .00085836 + a_t - .749a_{t-1} + 0.25a_{t-3}$

m2 = sarima(y,1,0,3)
m2 = Arima(y,order = c(1,0,3), fixed = c(NA,NA,NA,0,NA)) #use this model based on coeffeicant tests that are significant

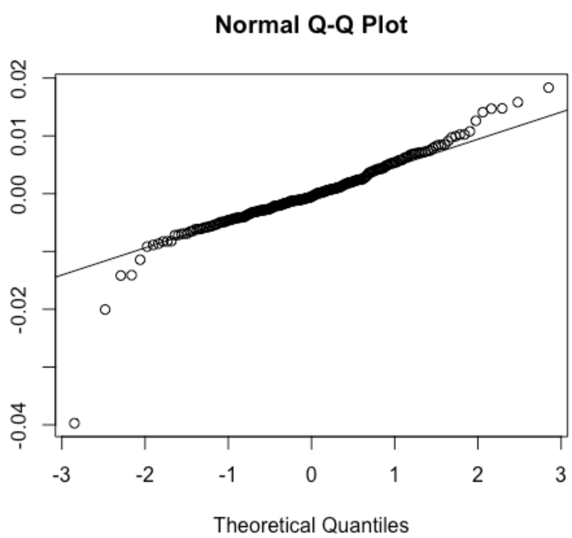
P values are above .05 so no signs of serial correlations



Residual Plot: **Mean is centering around zero which is a good sign besides the dip**

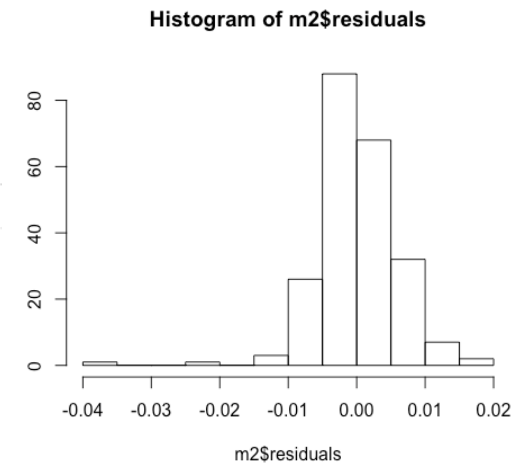


Residual qq plot: **Good amount of lines fallen on qq plot**



Residual Histogram

Slightly skewed due to some outliers, but overall shape is showing bellshape curve

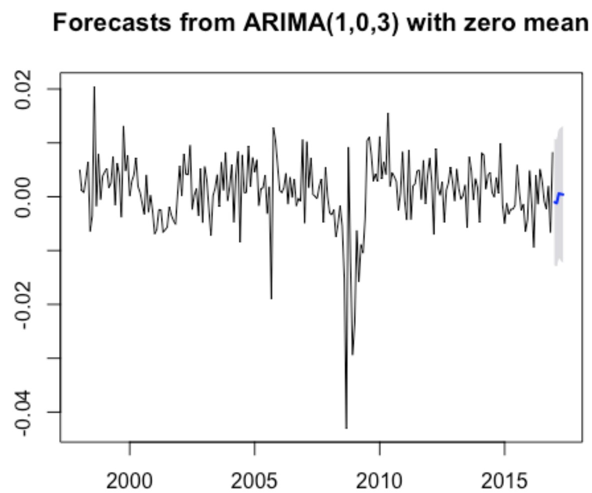


g) Plot the forecasts for the models M1 and M2 and compare their behavior. Are the forecasts similar in trends? Do you notice any major difference?

M1 results are forecasting a lower rate than m2 is, its not a huge difference but it is noticeable when looking at the results below
They are similar in trends, when looking at the forecasting graphs.

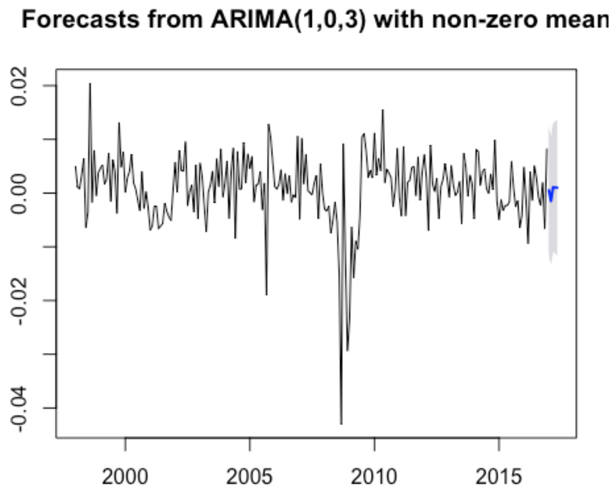
M1 results:

```
$lower
                95%         95%
Jan 2017 -0.01271632 -0.01271632
Feb 2017 -0.01291376 -0.01291376
Mar 2017 -0.01121615 -0.01121615
Apr 2017 -0.01181867 -0.01181867
May 2017 -0.01223015 -0.01223015

$upper
               95%        95%
Jan 2017 0.01069964 0.01069964
Feb 2017 0.01050272 0.01050272
Mar 2017 0.01239910 0.01239910
Apr 2017 0.01279112 0.01279112
May 2017 0.01302956 0.01302956
```



M2 results:

```
$lower
                 95%         95%
Jan 2017 -0.01122977 -0.01122977
Feb 2017 -0.01326903 -0.01326903
Mar 2017 -0.01073021 -0.01073021
Apr 2017 -0.01126122 -0.01126122
May 2017 -0.01160697 -0.01160697

$upper
                95%        95%
Jan 2017 0.01229023 0.01229023
Feb 2017 0.01029699 0.01029699
Mar 2017 0.01286611 0.01286611
Apr 2017 0.01331837 0.01331837
May 2017 0.01360015 0.01360015
```



h) Apply the backtesting procedure using 85% (=194 values) of the data for training and 15% for testing to evaluate the forecasting power of both models. Discuss the results.

```
> backtest(m,y,194,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.004749186
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.003886943
> backtest(m2,y,194,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.004749186
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.003886943
>
```

When running the backtest function, I seemed to get the same results for m1 and m2 models.
Models slightly vary, so I am confused on how the RMSE and MASE are about the same. Could it be that these are both MA models which is causing the issue?

i) At the end of this analysis, which model would you recommend for this analysis? Discuss the trend captured by it.
Both models seem to run the forecasting pretty well. There are slight deviations in the prediction results, but it seems that either model you will use will get similar results. Residuals are slightly better in M2, so I would go ahead with that model instead of m1. M1 was a good building block for m2 as it gave me ideas on how to build a better model.

Submission instructions

1. Submit the homework at the Course Web page d2l.depaul.edu
2. Submit your answers in a word document. Make sure to explain in detail your analyses, and include relevant output and graphs. You should attach code to your submission or include code in your word document.
3. If you have questions about the homework, email me BEFORE the deadline.
4. The assignment will lose **10%** of the points per day, after the due date.
5. Keep a copy of all your submissions!