CSC 425: Time Series Analysis - Homework 3

Akhil Kumar Ramasagaram

Friday, October 09, 2015

Problem 0

Suppose that the time series follows a MA(1) model: $X_t = a_t + 0.2a_{t-1}$, with $\sigma_a = 0.025$. Assume that $a_1 00 = 0.01$.

- Compute the 1-step ahead and the 2-step ahead forecasts of the return at the forecast origin $t = 100$.
  - The mean here is 0. So a 1-step ahead will be $\hat{X_n}(1) = \mu - \theta a_n = 0 - (-0.02) * 0.01 = 0.002$ and the mean of 2-step forcast is 0 since the MA(1) model converges to 0.
- Also compute the lag-1 and lag-2 autocorrelations of the MA(1) time series.
  - The formula for lag 1 is

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2} = \frac{0.2}{1.4} = 0.19$$

Problem 1

Describe the different behaviors of the ACF functions for an MA(q) model and an AR(p) model, and highlights the main differences

- An AR(p) model includes lagged terms of the time series itself and an MA(q) model includes lagged terms on the noise or residuals. If the ACF of the differenced series displays a sharp cutoff and /or the lag 1 AC is negative then consider MA. The ACF pattern for an AR(1) model will be, the ACF declines in geometric progression from its highest value at lag-1 & the pattern for MA(1) is that it cuts off abruptly after lag 1. ACF function decays exponentially to zero for AR model. ACF function is not zero for lags less than or equal to the order q of the MA(q) model.

Problem 2

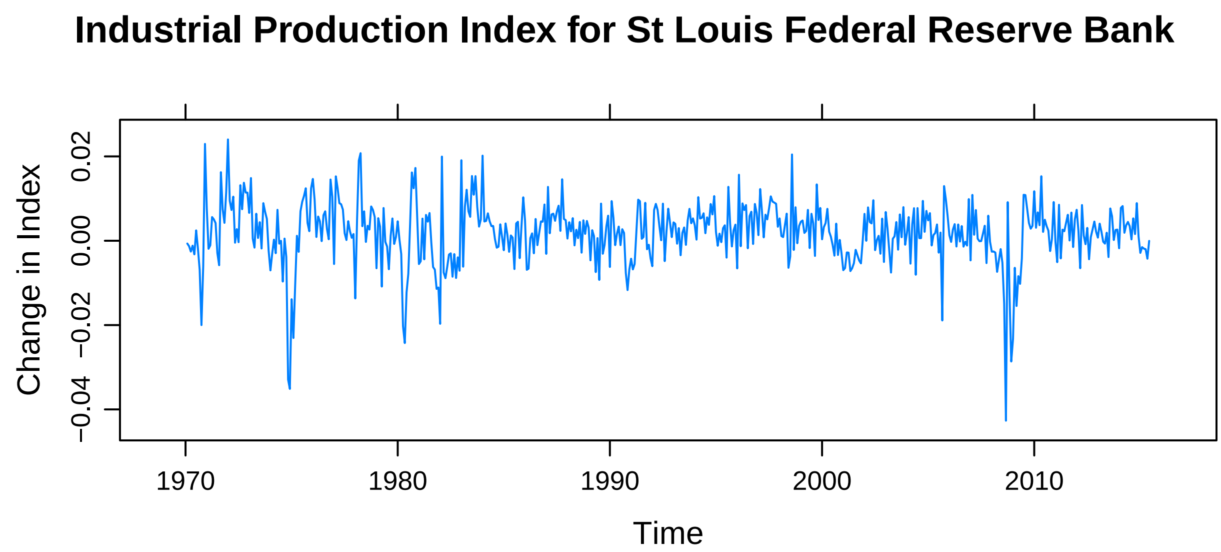
The Industrial Production Index (INDPRO) is an economic indicator that measures real output for all facilities located in the United States manufacturing, mining, and electric, and gas utilities. Since 1997, the Industrial Production Index has been determined from 312 individual series. The index is 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 values of the INDPRO index from February 1970 to June 2015 were obtained from the St Louis Federal Reserve Bank. The dataset contains two variables: date, growth. The following problem focuses on building a TS model for the index growth series.

a) Import the data in R and create a time object for the growth variable using the ts() function.

```
INDPRO <- read.csv("INDPRO.csv")
I_growth <- INDPRO$growth
I_ts <- ts(I_growth, start=c(1970,2), freq=12)
```

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

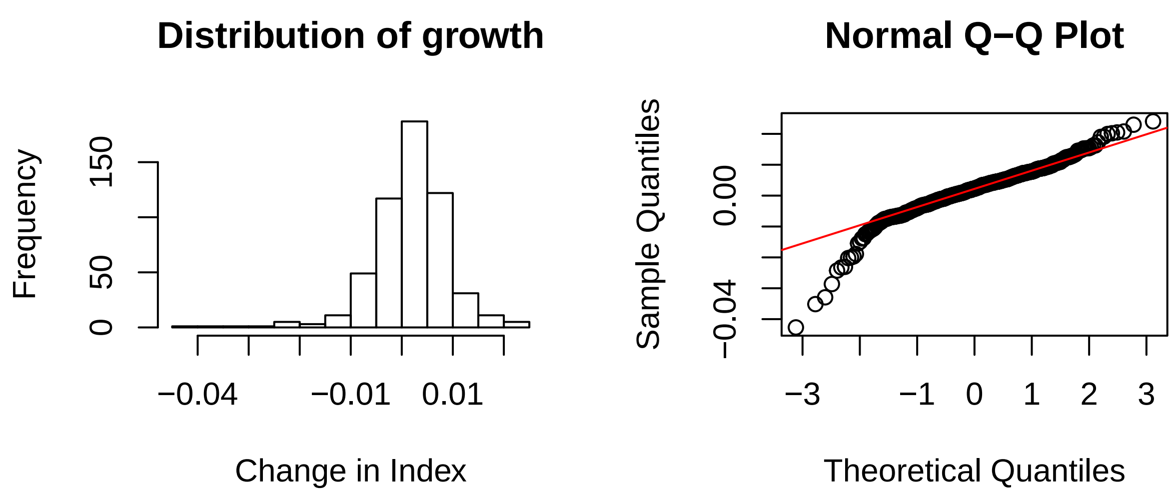
```r
xyplot(I_ts, ylab = "Change in Index",
       main = "Industrial Production Index for St Louis Federal Reserve Bank")
```



There are two predominant peaks here, both refer the economic crisis during 1973-75 and the 2008 economic meltdown which happened in the United States.

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

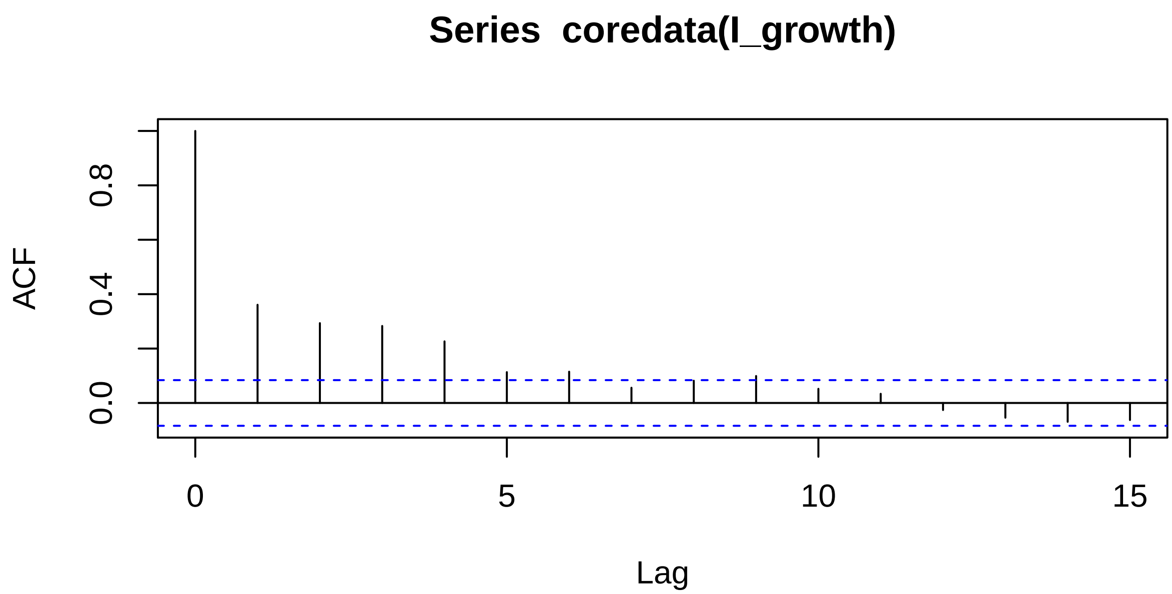
```r
par(mfrow = c(1,2))
hist(I_growth, xlab = "Change in Index", main = "Distribution of growth")
qqnorm(I_growth)
qqline(I_growth, col = 2)
```



It looks like a normal distribution, but its actually skewed towards left. If we look at the normal plot, we can see the skweness towards left.

d) Analyze the first 15 lags of ACF for growth and discuss if time series can be considered stationary.

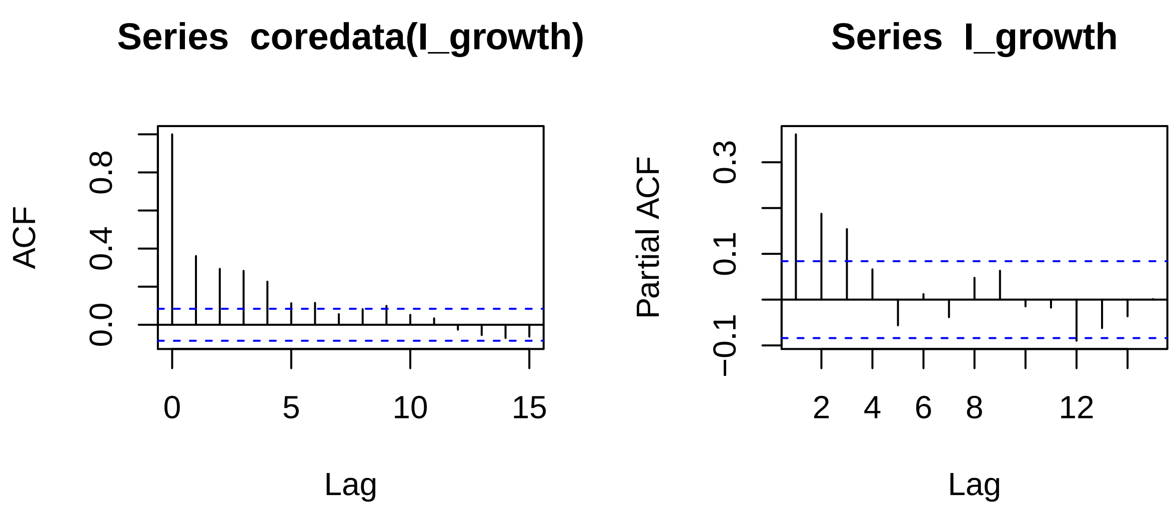
```r
acf(coredata(I_growth), plot=T, lag=15)
```



The ACF value doesn't decays, so we can assume this plot is stationary.

e) Analyze the ACF and the PACF functions and determine which model seems more appropriate to describe the time series. Does the process show clear AR behavior or MA behavior, or neither? Explain your answer.

```r
par(mfrow = c(1,2))
acf(coredata(I_growth), plot=T, lag=15)
pacf(I_growth, lag = 15)
```



For simplicity, i have condifered lag=15 for both ACF & PACF. From the ACF plot it looks like the first 6 lags looks significant, but if you look at the PCAF plot, only the first two are significant. I think here PCAF is more appropriate to describe this time series and the behavior is clearly Auto regressive, because if it was a MA model, then the ACF would have cut off abruptly after lag 1.

f) Use the BIC criterion to identify orders p and q of an initial model in the ARMA(p,q) family. What is the selected model?

```r
auto.arima(I_growth, ic = "bic")
```

```
## Series: I_growth
## ARIMA(1,0,1) with zero mean
##
## Coefficients:
##          ar1      ma1
##       0.8529  -0.5774
## s.e.  0.0392   0.0607
##
## sigma^2 estimated as 4.49e-05:  log likelihood=1954.53
## AIC=-3903.05   AICc=-3903.01   BIC=-3890.15
```

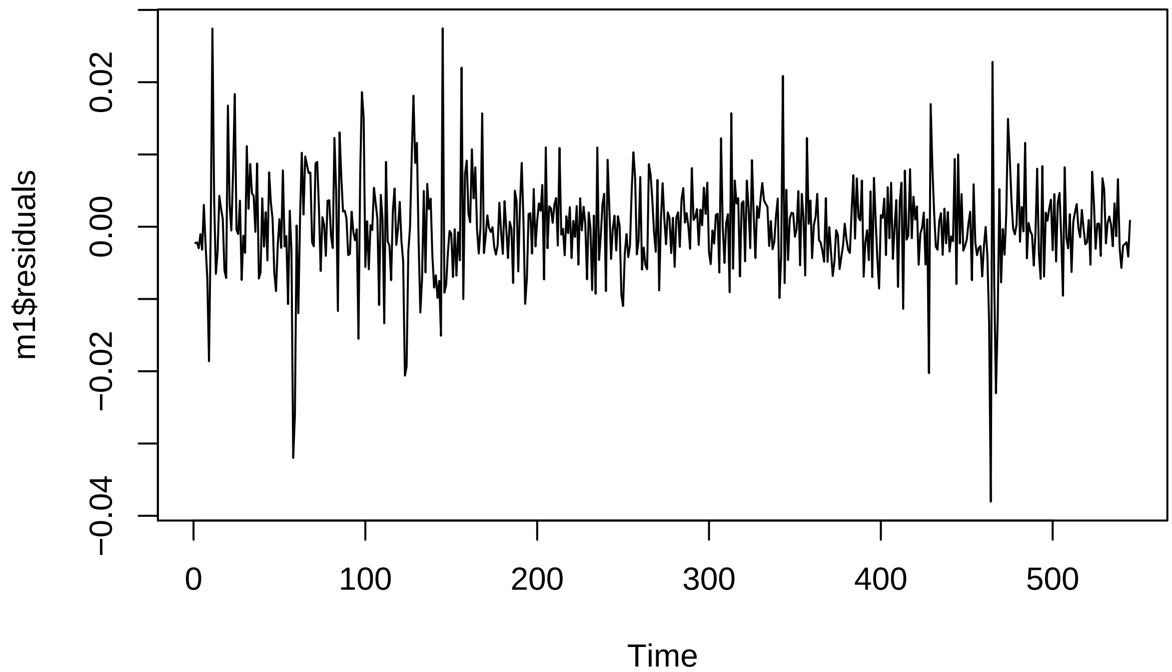
P = 1 & Q = 1

g) MODEL M1: Fit the model selected by the BIC criterion and apply the diagnostic methods to evaluate goodness of fit of the model: examine if model coefficients are significant, and perform residual analysis. Discuss if this model provides a good fit for the data.

```r
m1 <- Arima(I_growth, order = c(1,0,1),method='ML', include.mean=T)
coeftest(m1)
```

```
##
## z test of coefficients:
##
##              Estimate  Std. Error z value  Pr(>|z|)
## ar1        0.81695278  0.04680431 17.4546 < 2.2e-16 ***
## ma1       -0.54851790  0.06616913 -8.2896 < 2.2e-16 ***
## intercept  0.00183301  0.00070149  2.6130  0.008975 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(m1$residuals, type = "l")
```




```r
Box.test(m1$residuals, lag = 5, type = "Ljung", fitdf = 1)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 5.6329, df = 4, p-value = 0.2283
```

```r
Box.test(m1$residuals, lag = 10, type = "Ljung", fitdf = 1)
```

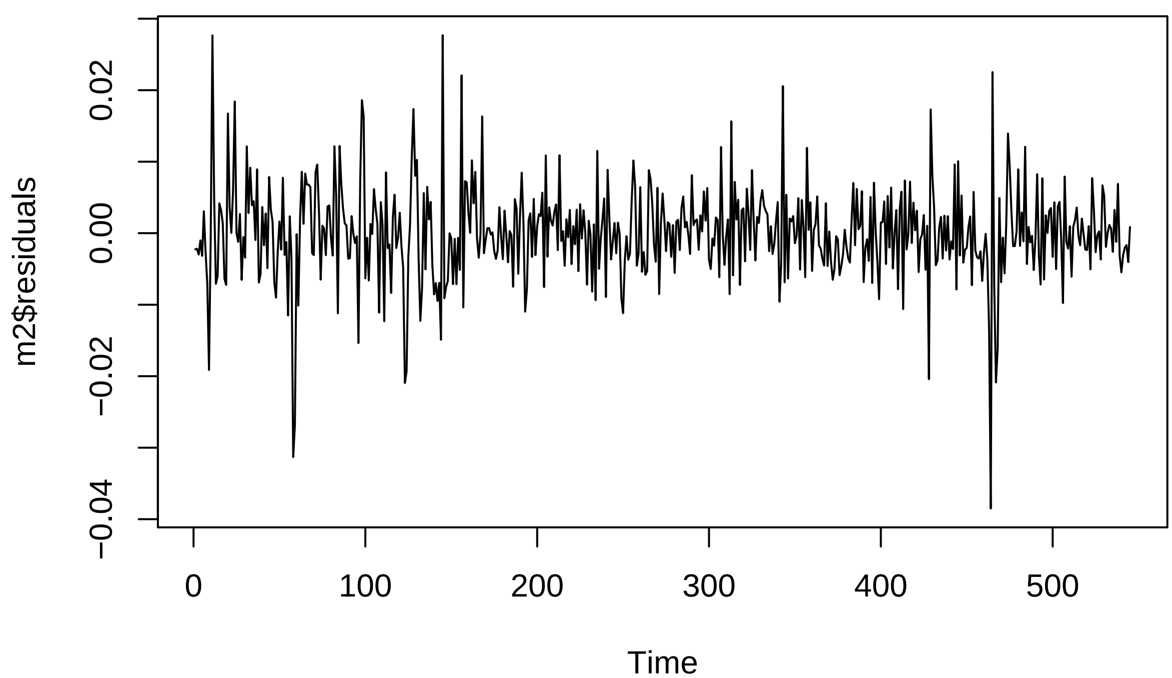
```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 10.1682, df = 9, p-value = 0.337
```

h) MODEL M2: Identify a possible order p for an AR(p) model and fit an AR(p) model for index growth (growth) time series. Examine the significance of the model coefficients and analyze the residuals to check adequacy of the model. Finding an appropriate model is an iterative process. If you are not satisfied with your initial model, modify it and find a more adequate model.

```r
m2 <- Arima(I_growth, order = c(3,0,0))
coeftest(m2)
```

```
##
## z test of coefficients:
##
##              Estimate Std. Error z value  Pr(>|z|)
## ar1       0.26374134 0.04225040  6.2423 4.311e-10 ***
## ar2       0.14233981 0.04331347  3.2863 0.0010152 **
## ar3       0.15411497 0.04223764  3.6488 0.0002635 ***
## intercept 0.00183420 0.00064702  2.8349 0.0045847 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(m2$residuals, type = "l")
```




```
Box.test(m2$residuals, lag = 5, type = "Ljung", fitdf = 3)
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 4.7376, df = 2, p-value = 0.09359
```

```
Box.test(m2$residuals, lag = 10, type = "Ljung", fitdf = 3)
```

```
##
```

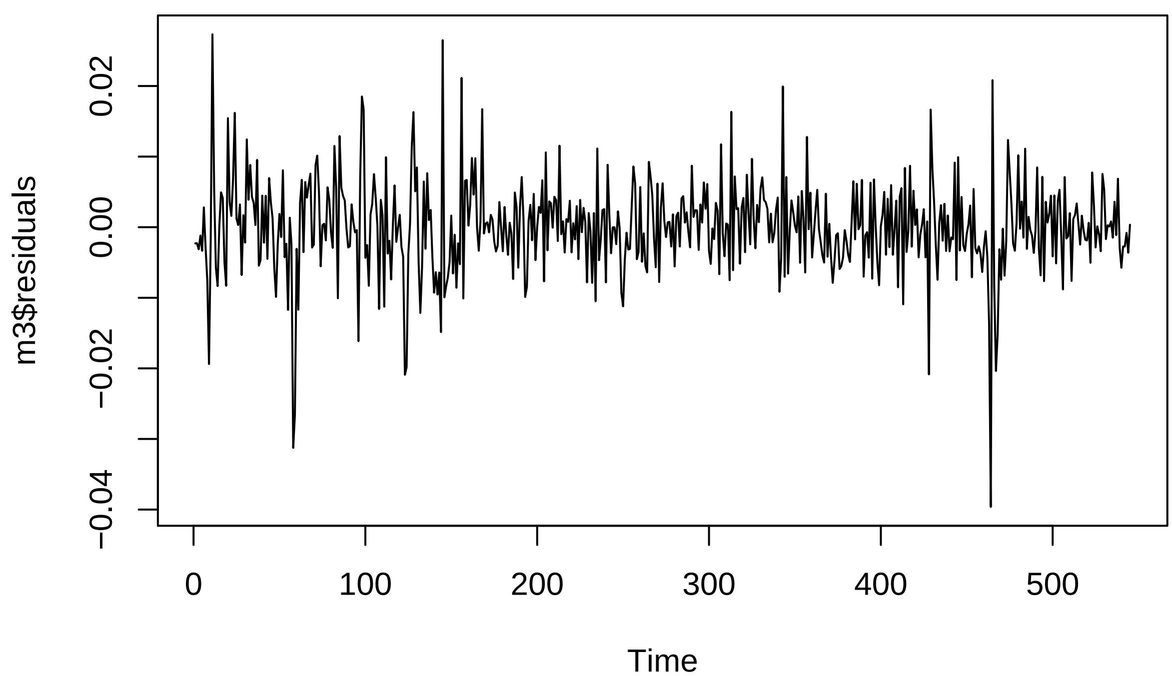
```
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 10.0724, df = 7, p-value = 0.1845
```

i) MODEL M3: Fit an MA(4) model for the index growth (growth) time series. Examine the significance of the model coefficients and analyze the residuals to check adequacy of the model. Discuss if this model provides a good fit for the data.

```
m3 <- Arima(I_growth, order = c(0,0,4), method='ML', include.mean=T)
coeftest(m3)
```

```
##
## z test of coefficients:
##
##             Estimate Std. Error z value  Pr(>|z|)
## ma1       0.25154570 0.04227609  5.9501 2.680e-09 ***
## ma2       0.18193410 0.04134291  4.4006 1.079e-05 ***
## ma3       0.22739295 0.04019237  5.6576 1.535e-08 ***
## ma4       0.20467424 0.04352368  4.7026 2.569e-06 ***
## intercept 0.00185604 0.00053171  3.4907 0.0004817 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(m3$residuals, type = "l")
```



```r
Box.test(m3$residuals, lag = 5, type = "Ljung", fitdf = 4)
```

```
##
##  Box-Ljung test
##
## data:  m3$residuals
## X-squared = 3.1869, df = 1, p-value = 0.07423
```

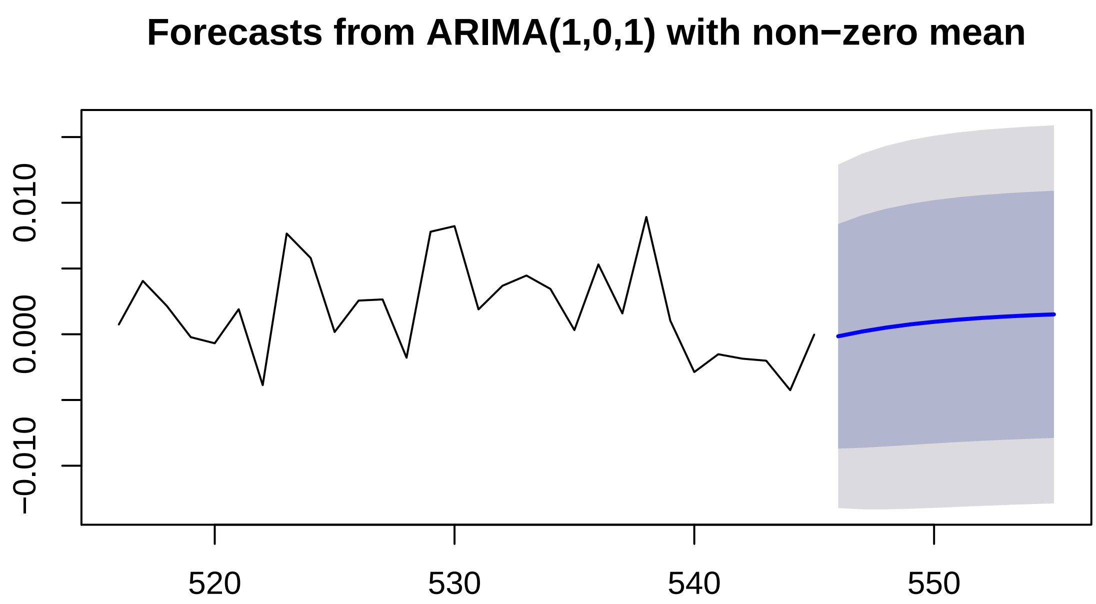
```r
Box.test(m3$residuals, lag = 10, type = "Ljung", fitdf = 4)
```

```
##
##  Box-Ljung test
##
## data:  m3$residuals
## X-squared = 9.3047, df = 6, p-value = 0.1572
```

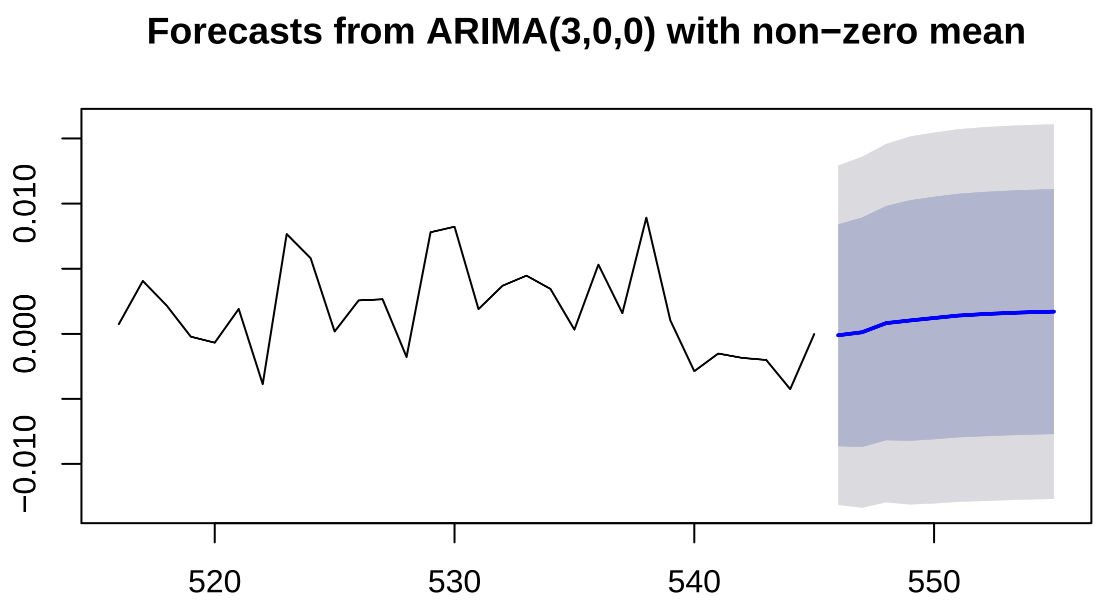
All the coefficients in all the models are significant. The model with the least BIC scores is teh best fit, in this case the MA(4) model is best.

j) Compute up to 5-step ahead forecasts for each of the three models, and compare them. Are they close in value?

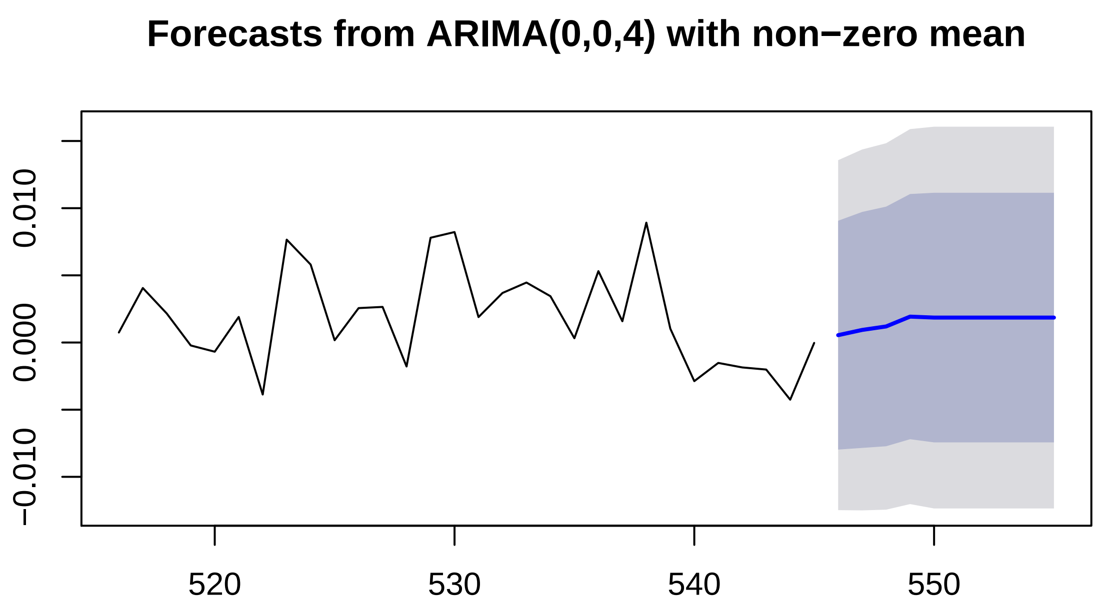
```r
plot(forecast.Arima(m1, h=10), include=30)
```



```r
plot(forecast.Arima(m2, h=10), include=30)
```




```r
plot(forecast.Arima(m3, h=10), include=30)
```



They all are almost close to each other.

k) Now write down the expressions for the three models M1, M2 and M3

the equation for m1 model is

$$(X_t - 0.0019) = 0.2732(X_{t-1} - 0.0019) + a_t$$

$$X_t = 0.0013 + 0.2732X_{t-1} + a_t$$

the equation for m2 model is

$$(X_t - 0.0019) = 0.3604(X_{t-1} - 0.0019) + a_t$$

$$X_t = 0.00121 + 0.3604X_{t-1} + a_t$$

the equation for m3 model is

$$(X_t - 0.0018) = 0.2515(X_{t-1} - 0.0018) + 0.1819(X_{t-2} - 0.0018) + 0.2274 * (X_{t-3} - 0.0018) + 0.2047 * (X_{t-4} - 0.0018) + a_t$$

$$X_t = 0.00016 + 0.2512X_{t-1} + 0.1819X_{t-2} + 0.2274X_{t-3} + 0.2474X_{t-4} + a_t$$