CSC 425 Time Series Analysis: Homework 4

Akhil Kumar Ramasagaram

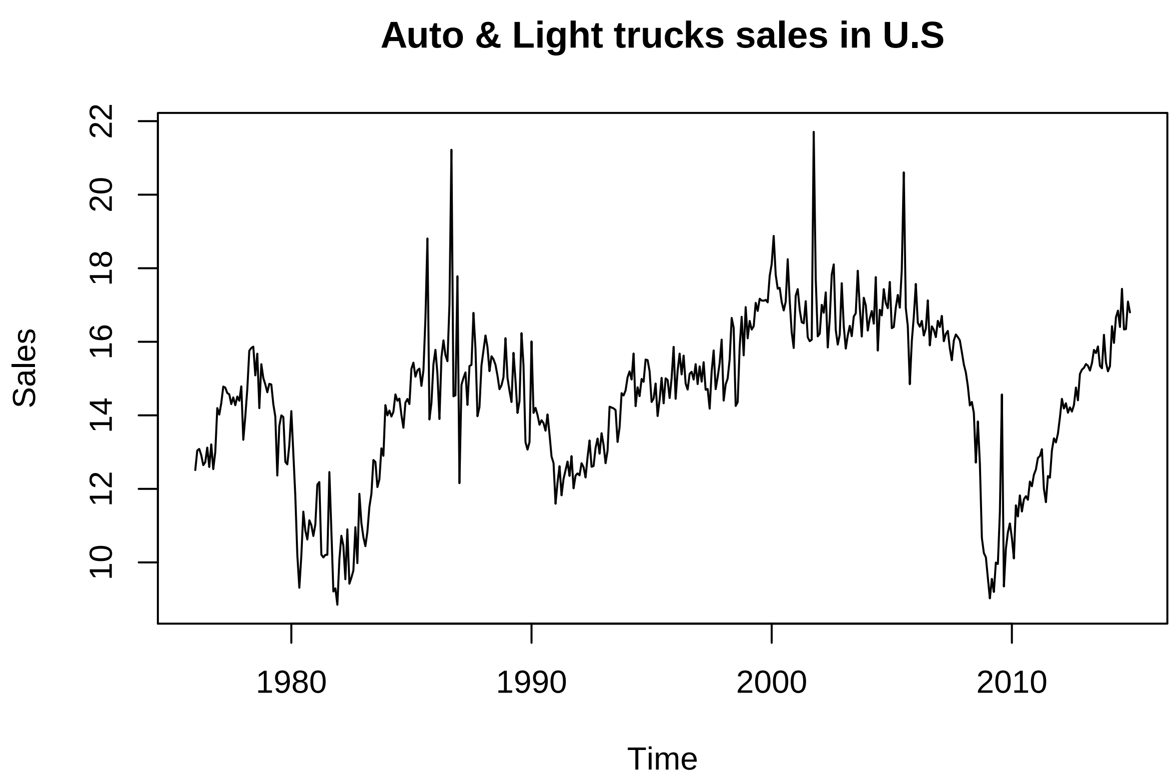
Wednesday, October 14, 2015

Problem 1

The dataset ALSALESRAW.csv contains actual monthly sales values X_t for autos and light trucks in the US from 2/1/1976 to 12/1/2014.

a) Create a time plot for monthly sales and analyze trends

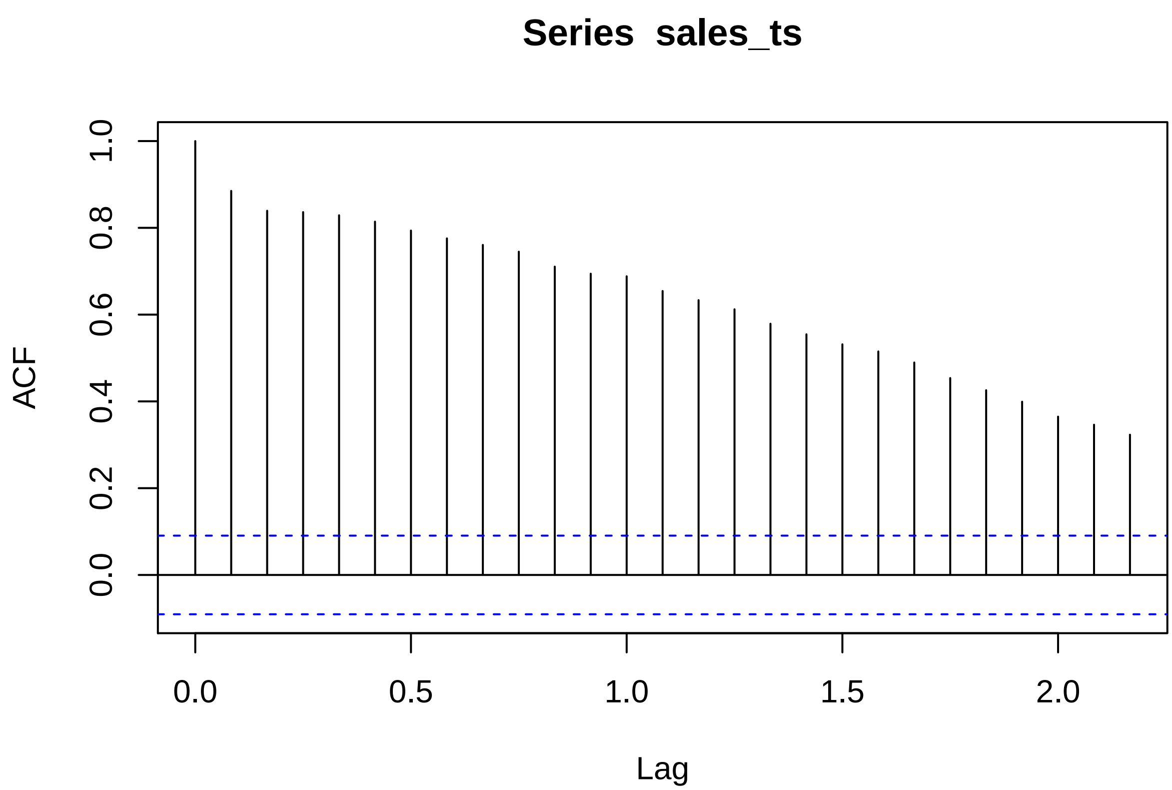
```
library(tseries)
library(fBasics)
library(forecast)
library(fUnitRoots)
library(lmtest)
source("backtest.R")
sales <- read.csv("ALTSALESRAW.csv")
sales_ts <- ts(sales$sales, start=c(1976,1), freq=12)
plot(sales_ts, ylab = "Sales", main = "Auto & Light trucks sales in U.S")
```



There are two prominent depressions, around 1981 and 2009. The highest sales have been during 2002. The sales after 2008 has seen gradually increasing trend.

b) Analyze if the series is stationary using both the ACF function and the Dickey Fuller test to check if TS is unit-root non-stationary.

```
acf(sales_ts)
```




```
adfTest(sales_ts, lags=5, type=c("ct"))
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 5
##   STATISTIC:
##     Dickey-Fuller: -2.1304
##   P VALUE:
##     0.5228
##
## Description:
##  Sat Oct 17 21:45:28 2015 by user: Akhilkumar
```

Since the acf plots gradually decays to zero, we can say that the given time series in not stationary.The Dickey Fuller test tells the given time series is unit root non stationary and we cannot reject the H_0

c) Use the BIC order selection method to identify the order of the *best* ARIMA(p,1,q) model.

```
auto.arima(sales$sales, ic = c("bic"), max.p = 5, max.q = 5, trace = T, allowdrift = F)
```

```
##
##  ARIMA(2,1,2)                    : 1261.981
##  ARIMA(0,1,0)                    : 1351.274
##  ARIMA(1,1,0)                    : 1312.223
##  ARIMA(0,1,1)                    : 1258.444
##  ARIMA(1,1,1)                    : 1258.074
##  ARIMA(1,1,2)                    : 1258.85
##  ARIMA(2,1,1)                    : 1256.937
##  ARIMA(2,1,0)                    : 1272.848
##  ARIMA(3,1,2)                    : 1266.064
##  ARIMA(3,1,1)                    : 1263.559
##
##  Best model: ARIMA(2,1,1)
```

```
## Series: sales$sales
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1      ar2      ma1
##       0.0715  -0.1717  -0.5490
## s.e.  0.0843   0.0579   0.0767
##
## sigma^2 estimated as 0.8163:  log likelihood=-615.46
## AIC=1238.92   AICc=1239.01   BIC=1255.5
```

The best Arima model is *ARIMA(2,1,1)* where $p = 2, q = 1$

d) Fit the selected ARIMA model, and analyze good ness of fit. Check if coefficients are significant and conduct residual analysis.Discuss results and explain if you are satisfied with the model chosen by the BIC criterion. If the model is not adequate, find a better model

```
m1 <- Arima(sales_ts, order = c(2,1,1), method = 'ML', include.mean = T)
coeftest(m1)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  0.071565   0.084318  0.8487  0.396022
## ar2 -0.171700   0.057916 -2.9646  0.003031 **
## ma1 -0.549024   0.076664 -7.1614 7.986e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow = c(2,2))
plot(m1$residuals)
qqnorm(m1$residuals)
qqline(m1$residuals)
Box.test(m1$residuals,lag=5,type='Ljung-Box',fitdf=3)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 0.8582, df = 2, p-value = 0.6511
```

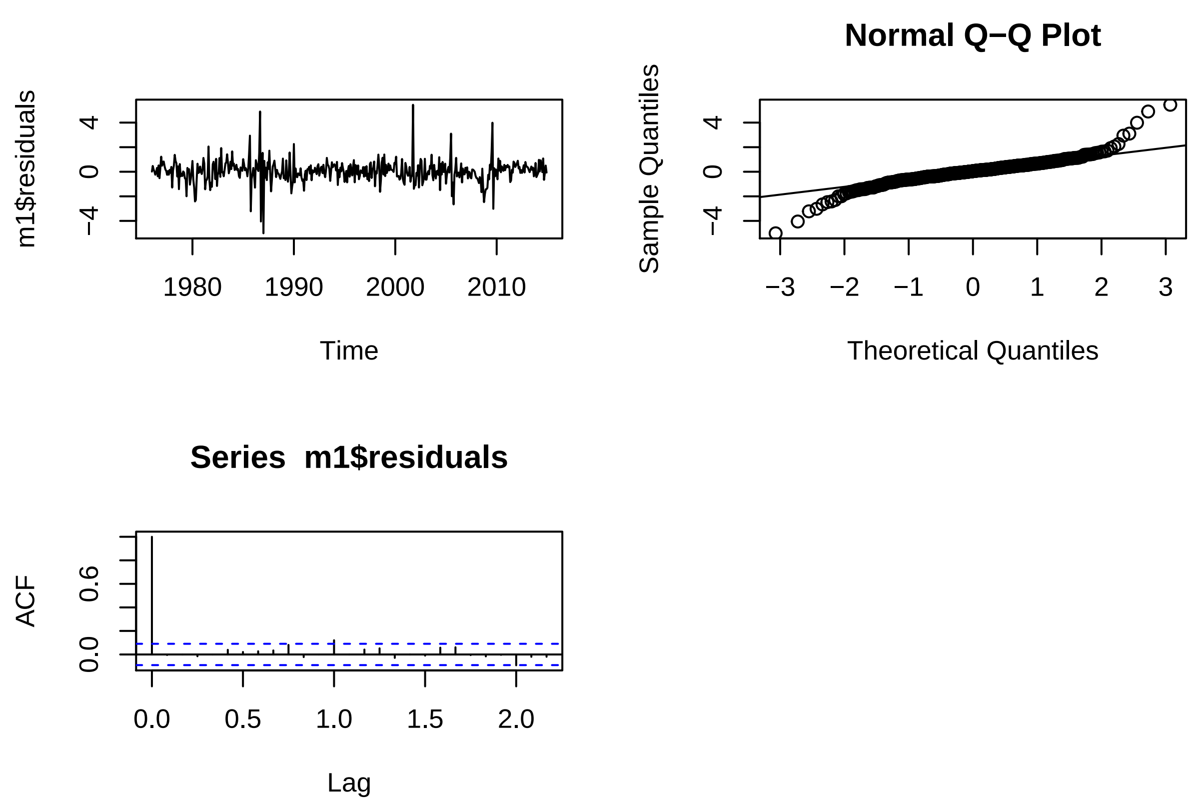
```r
Box.test(m1$residuals,lag=7,type='Ljung-Box',fitdf=3)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 1.4241, df = 4, p-value = 0.84
```

```r
Box.test(m1$residuals,lag=9,type='Ljung-Box',fitdf=3)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 5.1209, df = 6, p-value = 0.5284
```

```r
acf(m1$residuals)
```



Only ar2 & ma1 coeff is significant in the ARIMA model.

```r
m2 <- Arima(sales_ts, order = c(2,1,1), method = 'ML', include.mean = T, fixed = c(0, NA, NA))
coeftest(m2)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error  z value  Pr(>|z|)
## ar2 -0.196229   0.049046  -4.0009  6.31e-05 ***
## ma1 -0.491335   0.044354 -11.0775 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow = c(2,2))
plot(m2$residuals)
qqnorm(m2$residuals)
qqline(m2$residuals)
Box.test(m2$residuals,lag=5,type='Ljung-Box',fitdf=3)
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 1.6691, df = 2, p-value = 0.4341
```

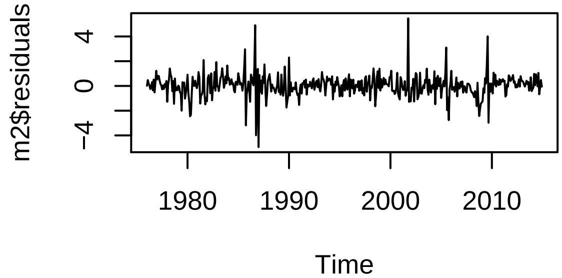
```r
Box.test(m2$residuals,lag=7,type='Ljung-Box',fitdf=3)
```

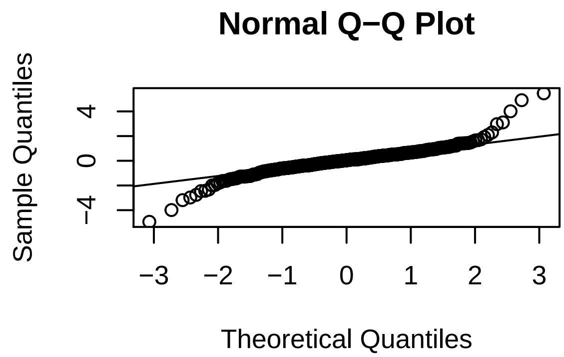
```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 2.1992, df = 4, p-value = 0.6992
```

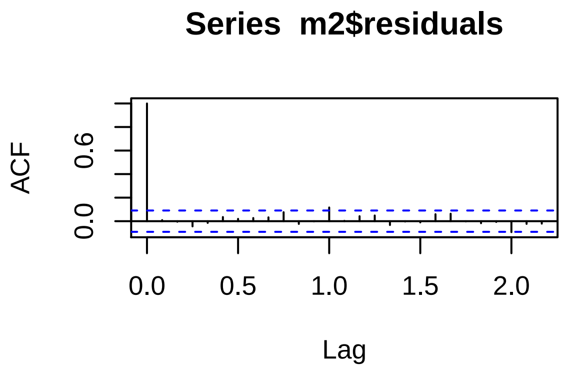
```r
Box.test(m2$residuals,lag=9,type='Ljung-Box',fitdf=3)
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 5.4809, df = 6, p-value = 0.4838
```

```r
acf(m2$residuals)
```



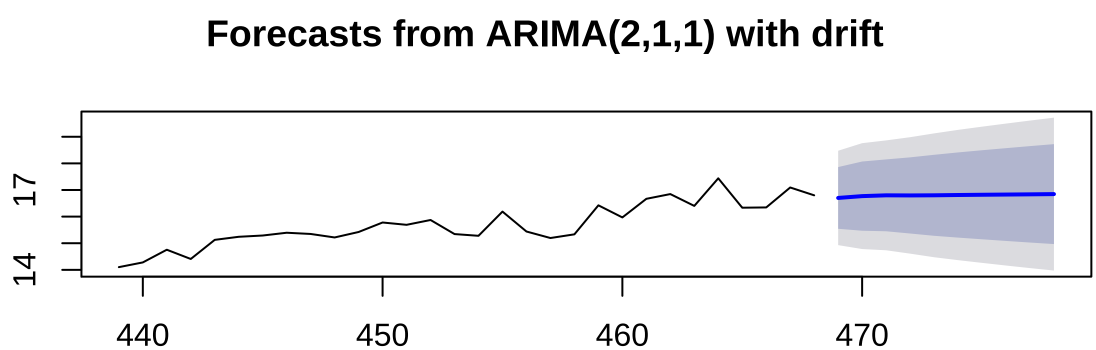




Again here only ar2 & ma1 coeff is significant in the ARIMA model.

e) Do you believe that the time series has a linear trend? If you include the drift (or constant term) in the ARIMA model, is the drift significant? Discuss if your findings suggest that the time series follows a linear time trend.

```
m3=Arima(sales$sales, order = c(2, 1, 1), method = 'ML', include.drift = T, fixed = c(0, NA, NA, NA))
f1=forecast.Arima(m3, h=10)
plot(f1, include=30)
```



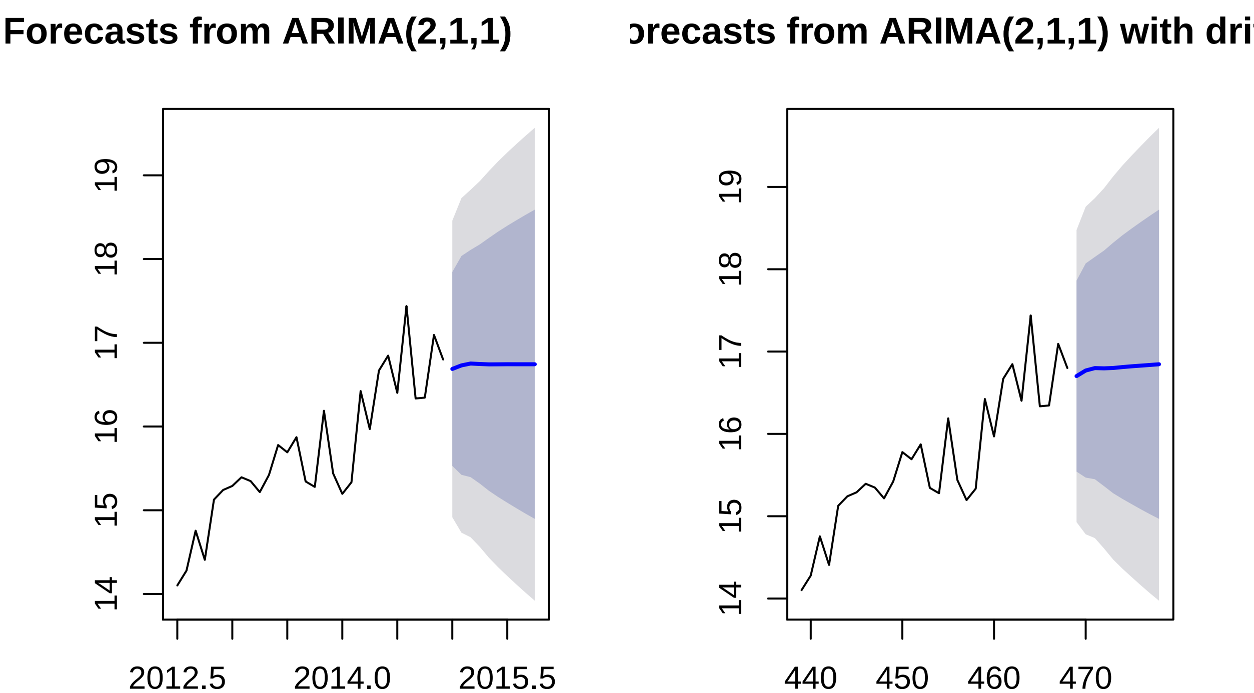
The time series follows a linear trend but the forecast are converged to the mean.

f) Write down the model expression

$$X_t = -0.1962X_{t-1} - 0.4913X_{t-2} + a_t$$

g) Plot the model forecasts and discuss if the forecasted trend is consistent with the past process behavior.

```
par(mfrow = c(1,2))
f2=forecast.Arima(m1, h=10)
plot(f2, include=30)
f3=forecast.Arima(m3, h=10)
plot(f3, include=30)
```



There is no change in forecast even after adding drift.

h) Use the backtesting procedures to compute the RMSE and the MAPE for the model. Interpret the result of MAPE.

```
backtest(m1, sales_ts, 398, 1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.7988595
```

```
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.5126948
## [1] "Mean Absolute Percentage error"
## [1] 0.03051754
## [1] "Symmetric Mean Absolute Percentage error"
## [1] 0.04018938
```

```
  backtest(m2, sales_ts, 398, 1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.7988595
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.5126948
## [1] "Mean Absolute Percentage error"
## [1] 0.03051754
## [1] "Symmetric Mean Absolute Percentage error"
## [1] 0.04018938
```

```
  backtest(m3, sales_ts, 398, 1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.7988595
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.5126948
## [1] "Mean Absolute Percentage error"
## [1] 0.03051754
## [1] "Symmetric Mean Absolute Percentage error"
## [1] 0.04018938
```

Looks like adding drift hasn't changed anything.

Problem 2

The file sugarprice.csv contains monthly sugar prices ($) in US cents per pound from September 2000 to August 2015.

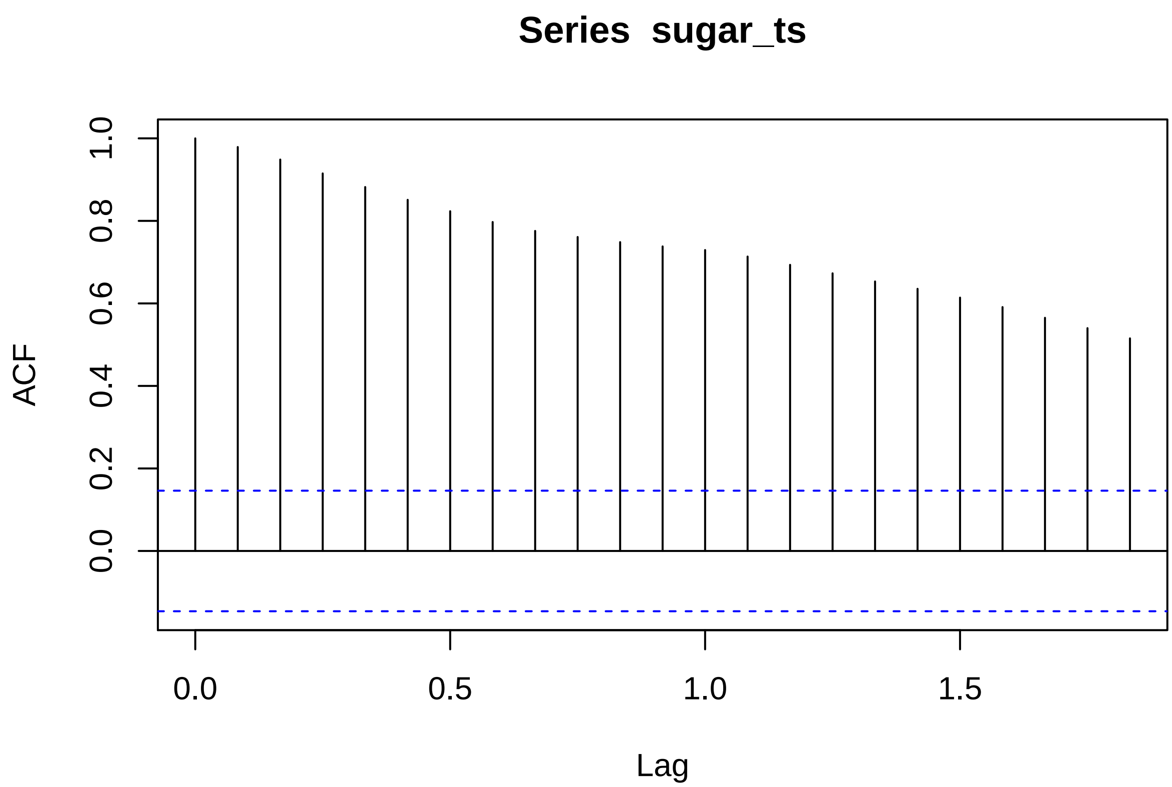
a) Plot the observed time series and its ACFs (20 lags). Analyze trends and patterns shown by the data.

```
sugar <- read.csv("sugar.csv")
sugar_ts <- ts(sugar[,2], start=c(2000,9), freq=12)
```

Since the acf plot gradually decays to zero, the given time series is stationary

b) Analyze if the series is stationary using both the ACF function and the Dickey Fuller test to check if TS is unit-root non-stationary.

```
acf(sugar_ts)
```




```r
adfTest(sugar_ts, lags=3, type=c("ct"))
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 3
##   STATISTIC:
##     Dickey-Fuller: -1.8561
##   P VALUE:
##     0.6364
##
## Description:
##  Sat Oct 17 21:45:33 2015 by user: Akhilkumar
```

```r
adfTest(sugar_ts, lags=5, type=c("ct"))
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
```

```
##   PARAMETER:
##     Lag Order: 5
##   STATISTIC:
##     Dickey-Fuller: -1.5037
##   P VALUE:
##     0.7836
##
## Description:
##  Sat Oct 17 21:45:33 2015 by user: Akhilkumar
```

The p-value is large, which tells us that we cannot reject H_0. So the given time series is unit root non stationary.

c) Use the BIC order selection method to identify the order of the "best" ARIMA(p,1,q) model.

```r
auto.arima(sugar$sugar, max.p = 5, max.q = 5, stationary = F, ic = c("bic"), trace = T, allowdrift = F)
```

```
##
##  ARIMA(2,1,2)                    : 591.6494
##  ARIMA(0,1,0)                    : 583.787
##  ARIMA(1,1,0)                    : 576.0621
##  ARIMA(0,1,1)                    : 576.3285
##  ARIMA(2,1,0)                    : 581.4536
##  ARIMA(1,1,1)                    : 581.2384
##  ARIMA(2,1,1)                    : 586.5583
##
##  Best model: ARIMA(1,1,0)
```

```
## Series: sugar$sugar
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##       0.2709
## s.e.  0.0719
##
## sigma^2 estimated as 1.376:  log likelihood=-282.6
## AIC=569.21   AICc=569.28   BIC=575.58
```

The best model with $ARIMA(p,1,q)$ is $p = 1, q = 0$ d) Fit the selected ARIMA model, and analyze good ness of fit. Check if coefficients are significant and conduct residual analysis.Discuss results and explain if you are satisfied with the model chosen by the BIC criterion. If the model is not adequate, find a better model

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

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 0.270935   0.071922  3.7671 0.0001652 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow = c(2,2))
plot(m1$residuals)
qqnorm(m1$residuals)
qqline(m1$residuals)
Box.test(m1$residuals,lag=5,type='Ljung-Box',fitdf=1)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 1.9706, df = 4, p-value = 0.7412
```

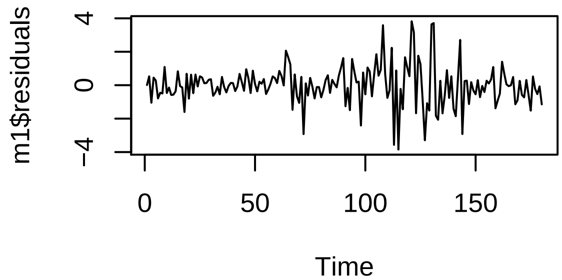
```r
Box.test(m1$residuals,lag=7,type='Ljung-Box',fitdf=1)
```

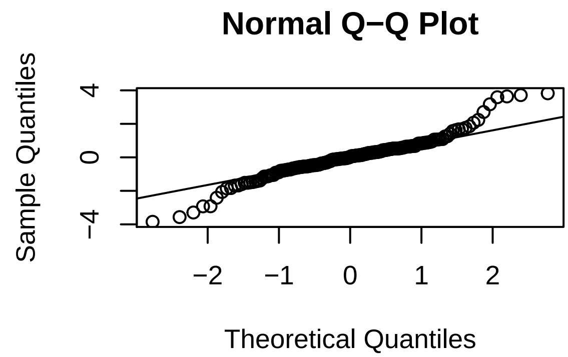
```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 3.7051, df = 6, p-value = 0.7165
```

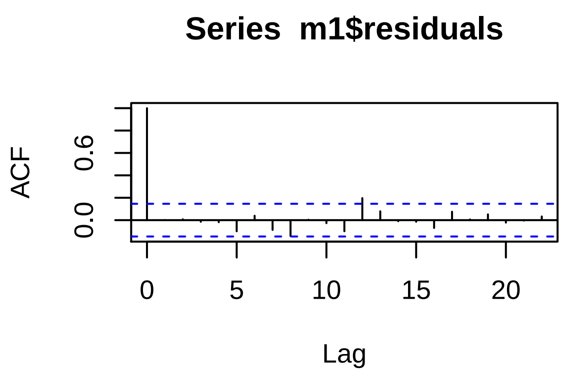
```r
Box.test(m1$residuals,lag=9,type='Ljung-Box',fitdf=1)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 7.5714, df = 8, p-value = 0.4764
```

```r
acf(m1$residuals)
```







Here the model has only single coeff which is significant.

e) Write down the model expression and discuss if your findings suggest that the time series follows a linear time trend

$$X_t = 0.251X_{t-1} + 0.003$$

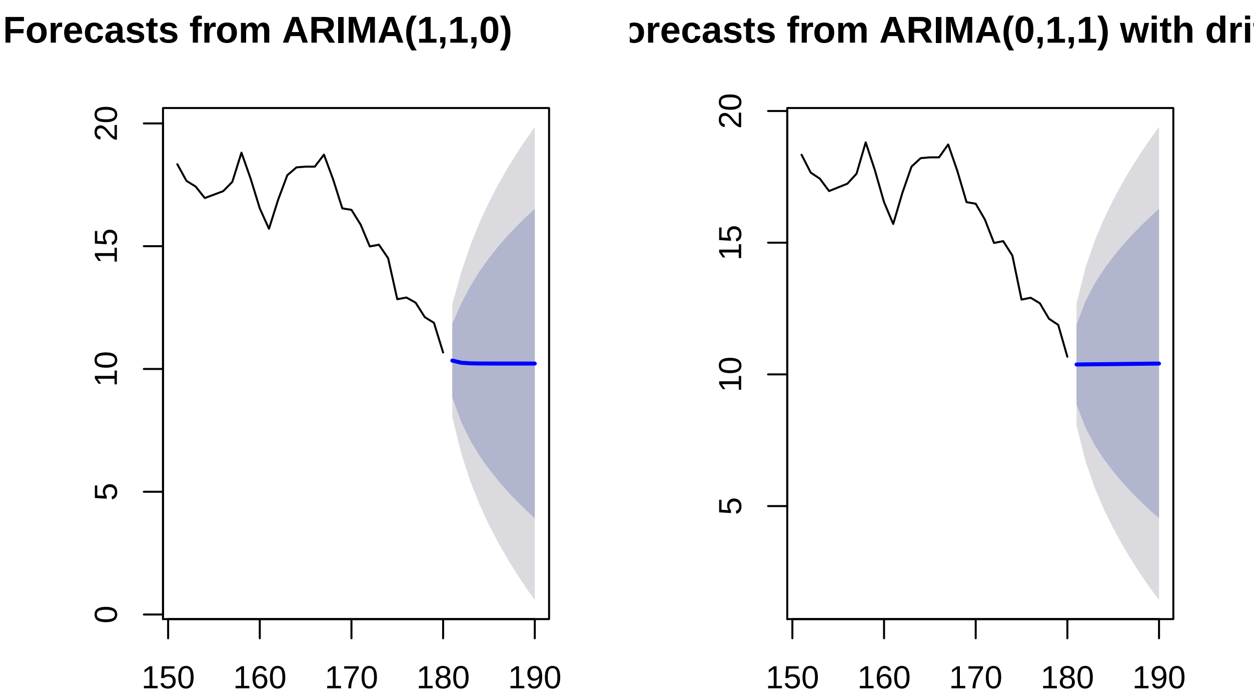
f) Plot the model forecasts and discuss if the forecasted trend is consistent with the past process behavior.

```
m2 <- Arima(sugar$sugar, order = c(0,1,1), method = 'ML', include.drift = T)
coeftest(m2)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error z value  Pr(>|z|)
## ma1   0.2513541  0.0673663  3.7312 0.0001906 ***
## drift 0.0039641  0.1102223  0.0360 0.9713106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow = c(1,2))
f1 <- forecast.Arima(m1, h =10)
plot(f1, include = 30)
```

```r
f2 <- forecast.Arima(m2, h =10)
plot(f2, include = 30)
```



g) Use the backtesting procedures to compute the RMSE and the MAPE for the model. Interpret the result of MAPE.

```r
backtest(m2,sugar_ts,153,1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.7414024
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.6025504
## [1] "Mean Absolute Percentage error"
## [1] 0.05647145
## [1] "Symmetric Mean Absolute Percentage error"
## [1] 0.03865968
```

```r
backtest(m1,sugar_ts,153,1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.7456587
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.6019946
```

```
## [1] "Mean Absolute Percentage error"
## [1] 0.05641936
## [1] "Symmetric Mean Absolute Percentage error"
## [1] 0.03858641
```

Looks like adding drift hasn't changed anything.