STAT 1331 - Assignment 4

Gordon Lu

11/9/2020

Problem 1:

Consider the monthly US unemployment rate from January 1949 to November 2011 in the file $m-unrate-4811.txt$. The data are seasonally adjusted and obtained from the Federal Reserve Bank at St. Louis.

1a) Does the monthly unemployment rate have a unit root? Why?

Solution:

```r
library("fUnitRoots")
```

```
## Warning: package 'fUnitRoots' was built under R version 3.6.3

## Loading required package: timeDate

## Loading required package: timeSeries

## Warning: package 'timeSeries' was built under R version 3.6.3

## Loading required package: fBasics

## Warning: package 'fBasics' was built under R version 3.6.3
```

```r
library("quantmod")
```

```
## Warning: package 'quantmod' was built under R version 3.6.3

## Loading required package: xts

## Warning: package 'xts' was built under R version 3.6.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.6.3
```

```
##
## Attaching package: 'zoo'

## The following object is masked from 'package:timeSeries':
##
##     time<-

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: TTR

## Warning: package 'TTR' was built under R version 3.6.3

##
## Attaching package: 'TTR'

## The following object is masked from 'package:fBasics':
##
##     volatility

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Version 0.4-0 included new data defaults. See ?getSymbols.
```

```r
df <- read.table("C:/Users/gordo/Desktop/Fall 2020 Classes/STAT 1331 - Financial Econometrics/Code/Assi
unemprate<-ts(df,start =c(1948,1), frequency=12, end = c(2011,11))
m1 <- ar(diff(unemprate),method='ols')
m1$order
```

```
## [1] 27
```

```r
#l_unrate <- log(df[,1]) #ignore date column
#m1 <- ar(diff(l_unrate), method = 'ols')
#m1$order
```

```r
#now perform augmented dickey-fuller test
adfTest(unemprate, lags = 12, type = c("c"))
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 12
##   STATISTIC:
```

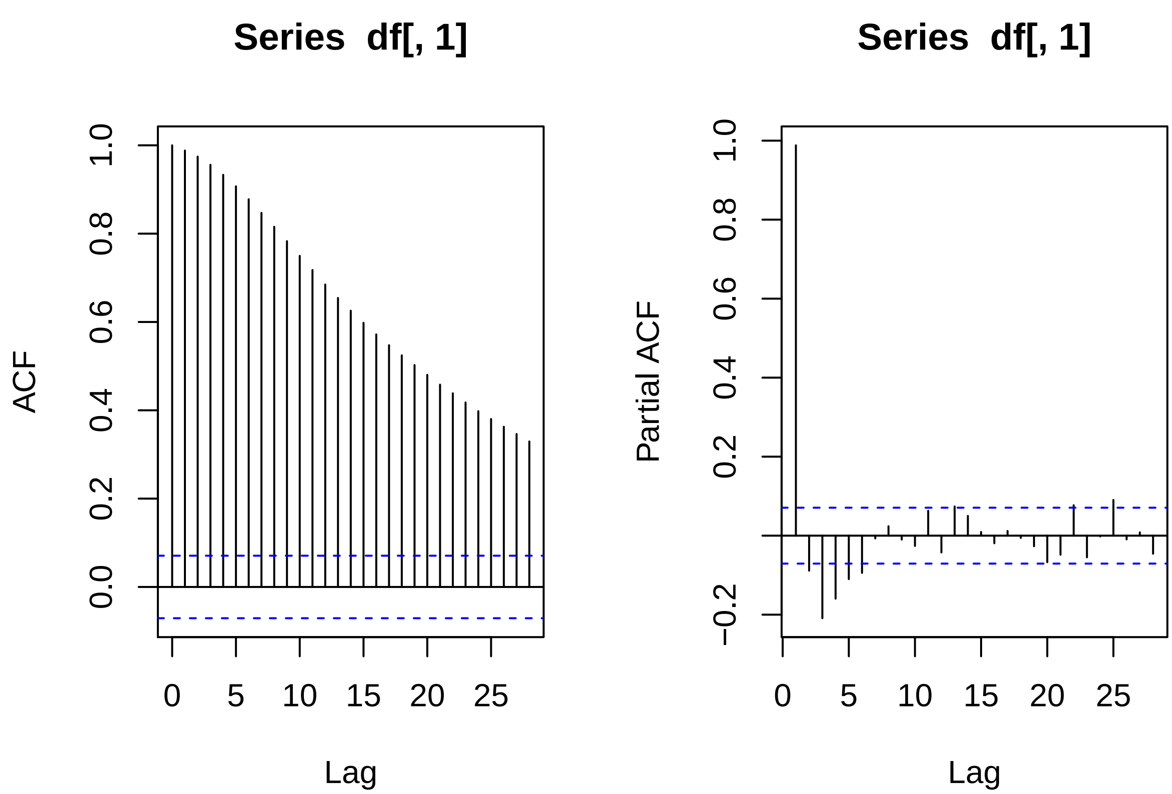
```
##     Dickey-Fuller: -2.72
##   P VALUE:
##     0.07501
##
## Description:
##  Tue Nov 10 22:17:28 2020 by user: gordo
```

No, since the Augmented Dickey-Fuller test had a test statistic > -3, we can say that the process is non-stationary.

1b) Build a time series model for the monthly unemployment rates. Check the fitted model for adequacy. Then use the model to forecast the unemployment rate for December 2011 and the first three months of 2012. (Note that there are more than one model that fits the data well. You only need an adequate model.)

Solution:

```r
par(mfrow=c(1,2))
acf(df[,1]) #display ACF of log return series
pacf(df[,1])
```



```r
info<-arima(unemprate,order = c(1,0,0))
info
```

```
##
## Call:
## arima(x = unemprate, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.9933     6.0109
## s.e.  0.0043     1.0119
##
## sigma^2 estimated as 0.046:  log likelihood = 90.34,  aic = -174.69
```

So, the model is:

$$x_t = 1.7466 + 0.9927x_{t-1} + \varepsilon_t$$

```r
predict(info, 4) #predict unemployment rate for December 2011, and first 3 months of 2012
```

```
## $pred
##           Jan      Feb      Mar Apr May Jun Jul Aug Sep Oct Nov      Dec
## 2011                                                            8.682115
## 2012 8.664350 8.646702 8.629172
##
## $se
##            Jan       Feb       Mar Apr May Jun Jul Aug Sep Oct Nov       Dec
## 2011                                                               0.2144783
## 2012 0.3023112 0.3690277 0.4247081
```

Problem 6:

Consider the two bond yield series of the previous exercise. What is the relationship between the two series? To answer this question, take the log transformation of the data to build a time series model for the Aaa yields using Baa yields as an explanatory variable. Write down the fitted model, including model checking.

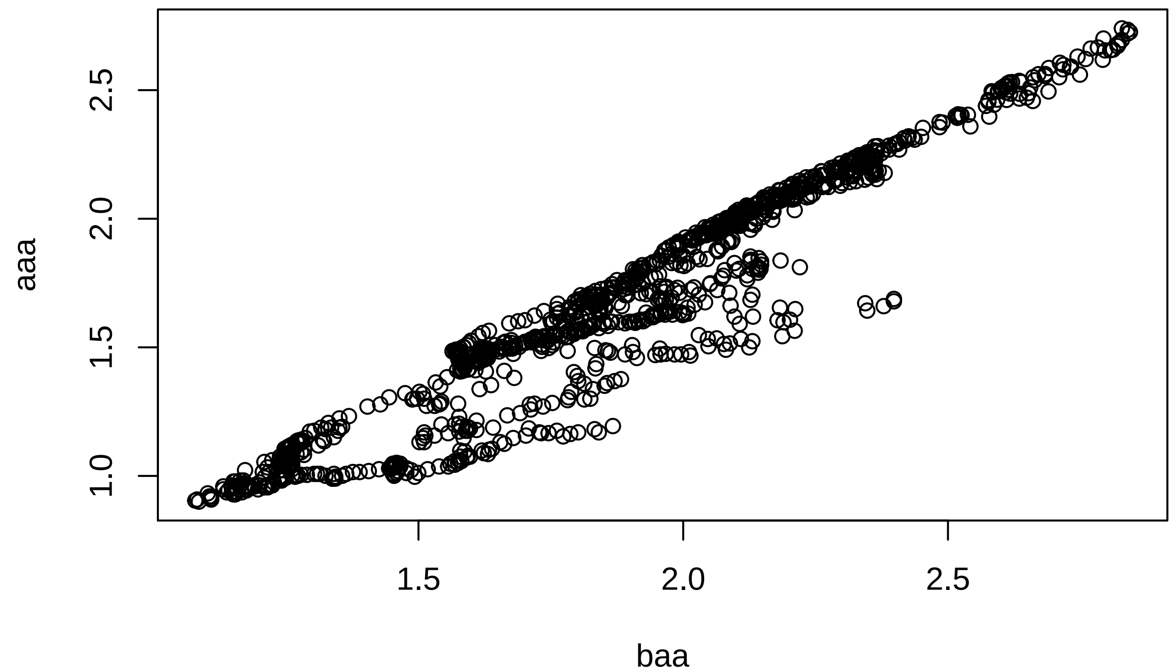
Solution:

First, set up a linear regression model of ln(Aaa) and ln(Baa):

```r
da <- read.table("C:/Users/gordo/Desktop/Fall 2020 Classes/STAT 1331 - Financial Econometrics/Code/Assi
da1 <- read.table("C:/Users/gordo/Desktop/Fall 2020 Classes/STAT 1331 - Financial Econometrics/Code/Ass

aaa <- log(da$yield)
baa <- log(da1$yield)

plot(baa, aaa)
```




```r
m1 <- lm(aaa ~ baa) #linear model of log(aaa) and log(baa)
summary(m1)
```

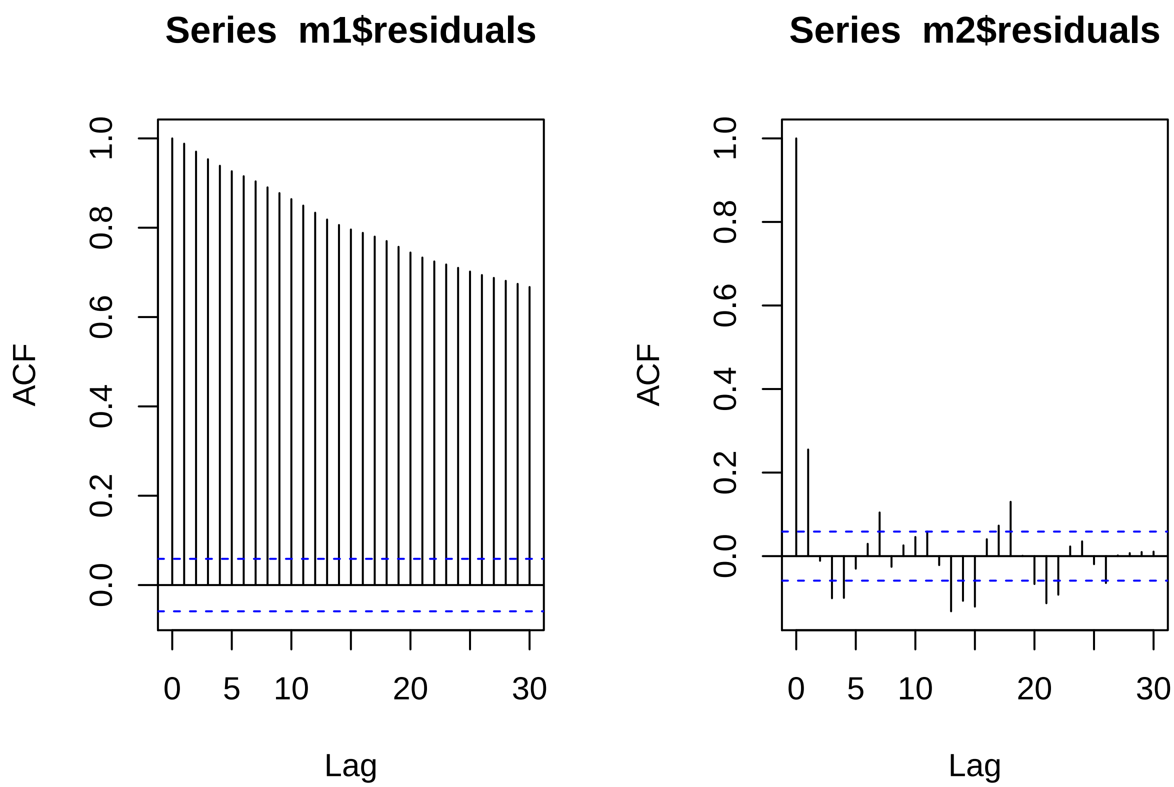
```
##
## Call:
## lm(formula = aaa ~ baa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55330 -0.02714  0.04803  0.08299  0.16292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.358993   0.018628  -19.27   <2e-16 ***
## baa          1.080637   0.009694  111.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1285 on 1113 degrees of freedom
## Multiple R-squared:  0.9178, Adjusted R-squared:  0.9177
## F-statistic: 1.243e+04 on 1 and 1113 DF,  p-value: < 2.2e-16
```

So, we have the linear regression model:

$$ln(Aaa_t) = -0.359 + 1.081 ln(Baa_t) + \varepsilon_t$$

Then, based on the ACF of the residuals of the linear regression model and the ACF of the differenced series of the residuals of the linear regression model:

```
par(mfrow = c(1,2))
acf(m1$residuals)
m2 <- lm(diff(baa) ~ -1+diff(aaa))
acf(m2$residuals)
```



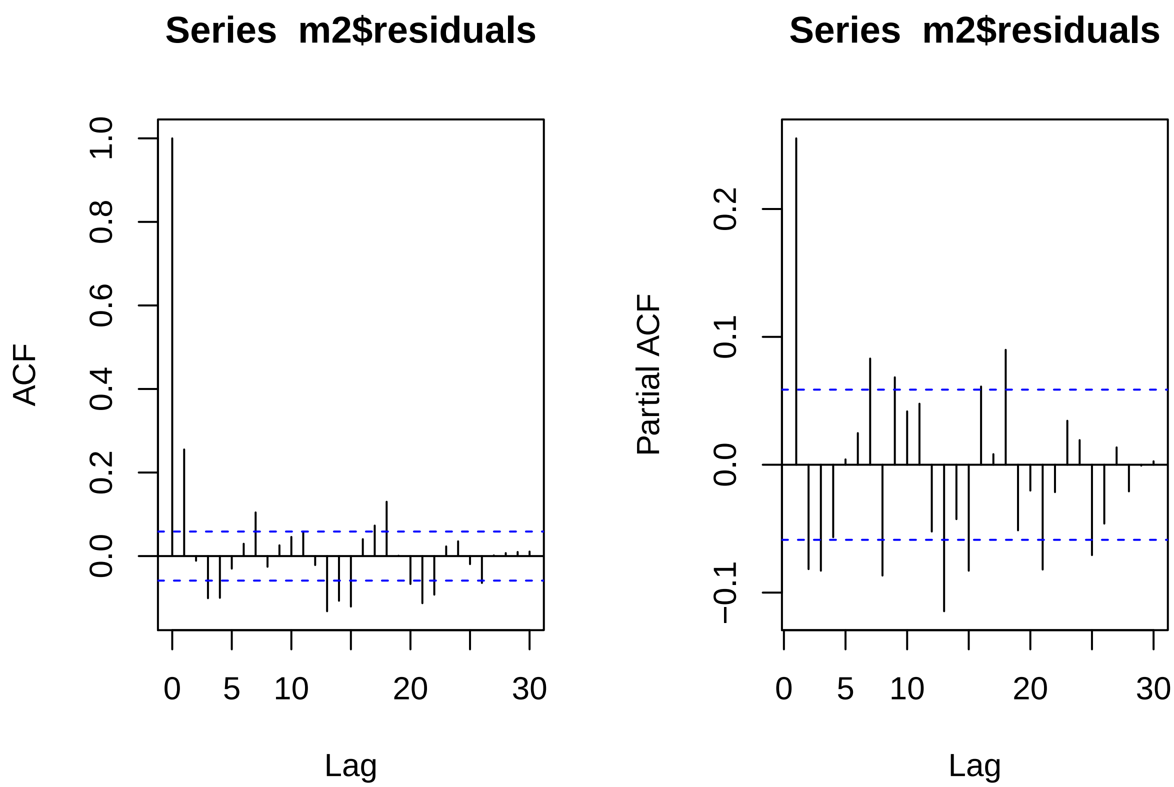
So, we see that a unit root does indeed exist.

So, now we set up the linear model as:

$$z_t = 0.6414523y_t + \varepsilon_t$$

Then, we plot the ACF and PACF of the residuals of the residuals of the coefficients of the regression model:

```
par(mfrow = c(1,2))
acf(m2$residuals)
pacf(m2$residuals)
```



Based on the PACF plot, the peak appears at lag 2. Thus, we employ an AR(2) model.

```
m3 <- arima(diff(aaa), order = c(2,0,0))
m3
```

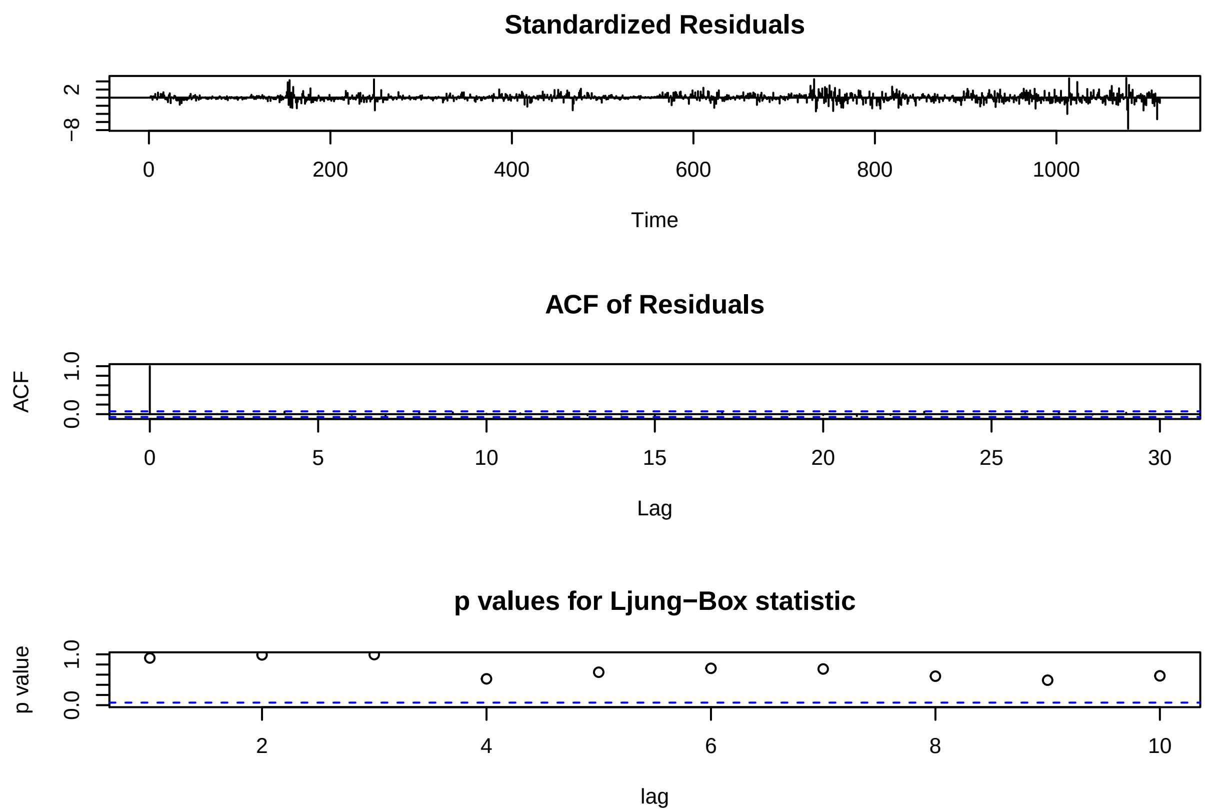
```
##
## Call:
## arima(x = diff(aaa), order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       0.3733  -0.1596     -3e-04
## s.e.  0.0296   0.0296      8e-04
##
## sigma^2 estimated as 0.0004644:  log likelihood = 2694.07,  aic = -5380.13
```

So, our fitted model is:

$$(1 + 0.3733B_t - 0.1560B_t^2)(z_t - 0.6414523y_t) = a_t$$

All of the coefficients here are significant, so we perform model checking.

```
tsdiag(m3)
```



The results show that the residuals follow a pattern of white noise series, thus, the model is adequately specified.

Problem 8:

Consider the US quarterly real gross national product from the first quarter of 1947 to the third quarter of 2011. The data are in the file $q-GNPC96.txt$, seasonally adjusted, and in billions of chained 2005 dollars. Let x_t be the growth rate series of real GDP:

8a) The ar command identifies an AR(4) model for x_t via the AIC criterion. Fit the model. Is the model adequate? Why?

Solution:

```
da <- read.table("C:/Users/gordo/Desktop/Fall 2020 Classes/STAT 1331 - Financial Econometrics/Code/Assi
m1 <- ar(diff(log(da$gnp)), method = 'ols')
m2 <- arima(diff(log(da$gnp)), order = c(4,0,0))
m2
```

```
##
## Call:
```

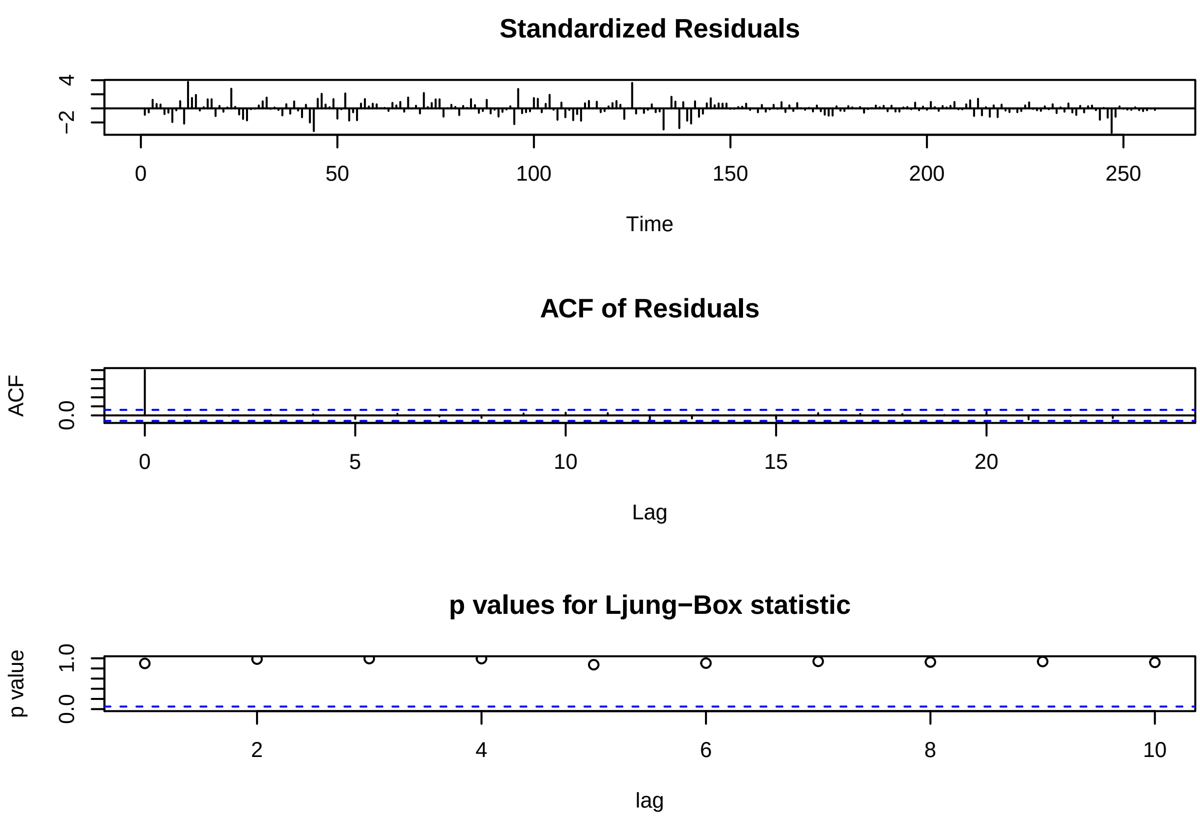
```
## arima(x = diff(log(da$gnp)), order = c(4, 0, 0))
##
## Coefficients:
##          ar1     ar2      ar3      ar4  intercept
##       0.3369  0.1513  -0.1010  -0.0887     0.0078
## s.e.  0.0619  0.0652   0.0651   0.0619     0.0008
##
## sigma^2 estimated as 8.368e-05:  log likelihood = 844.9,  aic = -1677.8
```

The fitted model for the process is:

$$x_t = 0.0078 + 0.3369x_{t-1} + 0.1513x_{t-2} - 0.1010x_{t-3} + 0.0887x_{t-4} + \varepsilon_t$$

Then, perform model checking using the `tsdiag()` function.

```r
tsdiag(m2)
```



So, based on the plots for the model chcking statistics, there does not appear to be any inadequacy for the fitted model. Additionally, from the ACF, we can see that the ACF decays quickly, and the residuals do not seem to have constant variance. So, I would say that the AR(3) model is adequate.

8b) The sample PACF of x_t specifies an AR(3) model. Fit the model. Is it adequate? Why?

Solution:

```
m3 = arima(diff(log(da$gnp)),order = c(3,0,0),season = list(order = c(1,0,1), period=4))
m3
```

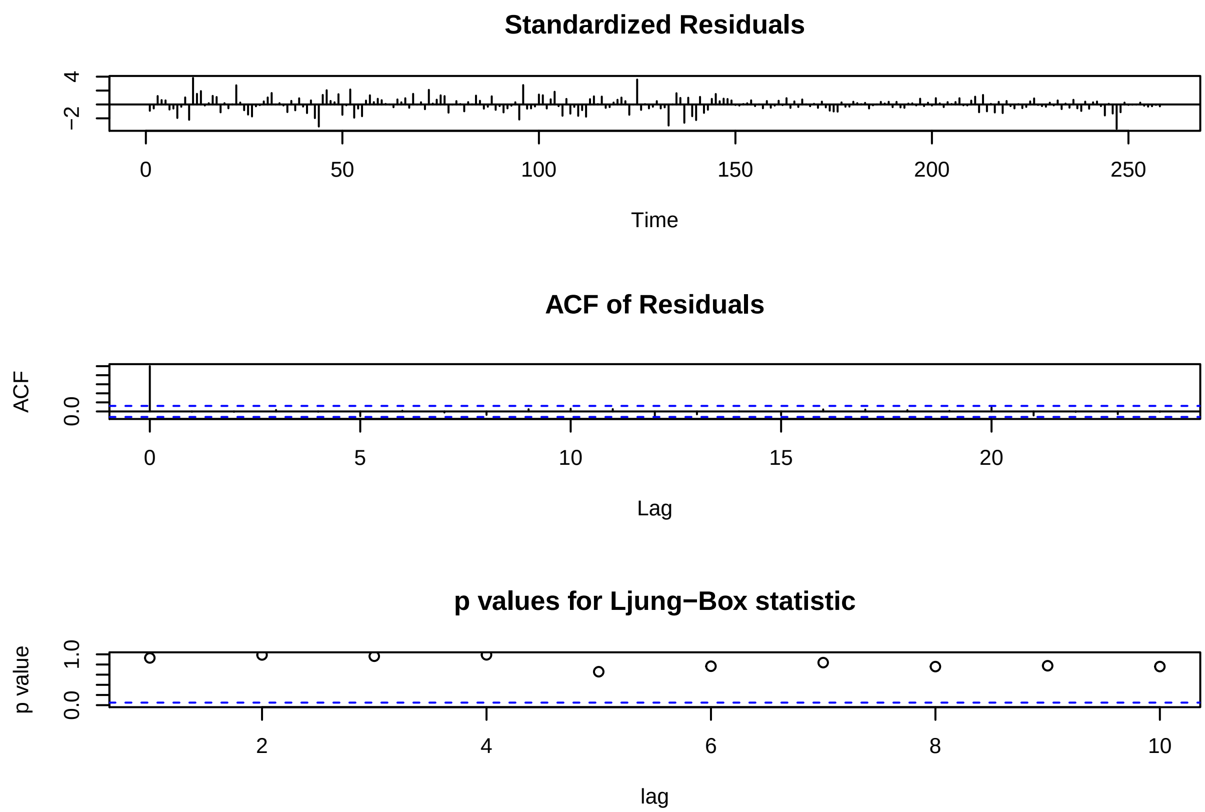
```
##
## Call:
## arima(x = diff(log(da$gnp)), order = c(3, 0, 0), seasonal = list(order = c(1,
##     0, 1), period = 4))
##
## Coefficients:
##          ar1     ar2      ar3     sar1    sma1  intercept
##       0.3385  0.1479  -0.1170  -0.5807  0.5294     0.0078
## s.e.  0.0627  0.0654   0.0638   0.4058  0.4203     0.0009
##
## sigma^2 estimated as 8.407e-05:  log likelihood = 844.32,  aic = -1674.64
```

The fitted model for the process is:

$$x_t = 0.0078 + 0.3484t - 1 + 0.1386x_{t-2} - 0.1317x_{t-3} + \varepsilon_t$$

Then, perform model checking using the `tsdiag()` function.

```
tsdiag(m3)
```



The above plots show that the residuals follow a pattern of a white noise process, thus, the AR(3) model is adequately specified.

8c) What is the model for x_t if one uses in-sample model comparison? Why?

Solution:

We would prefer to use the AR(4) model, since the AIC for AR(4) is smaller than that of the AR(3) model.

8d) Divide the data into estimation and forecasting subsamples using the fourth quarter of 2000 as the initial forecast origin and apply the backtesting procedure with MSFE as the criterion. Select a model for x_t. Justify the choice.

Solution:

```
source("backtest.r") # add functions from the local R file named backtest.R
#Apply backtest on AR(3) model:
m3_backtest = backtest(m3, diff(log(da$gnp)), 215, 1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.007565077
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.005099862
```

```r
#Apply backtest on AR(4) model:
m4_backtest = backtest(m2, diff(log(da$gnp)), 215, 1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.007515004
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.005053905
```

Since the MSFE for AR(4) is less than the MSFE for the AR(3) modl, I would pick the AR(4) model.