Solution_HW2.R

I will hold the office hours on this Saturday from: 10:00 - 15:00 and on this Monday from: 9:00 - 15:00

wasin_siwasarit

Fri Mar 1 09:31:44 2019

```
#EE435 Wasin Siwasarit Assignment2
setwd("/Users/wasin_siwasarit/Desktop/EE435")
cat(rep("\n",50))  #clear R Console
```

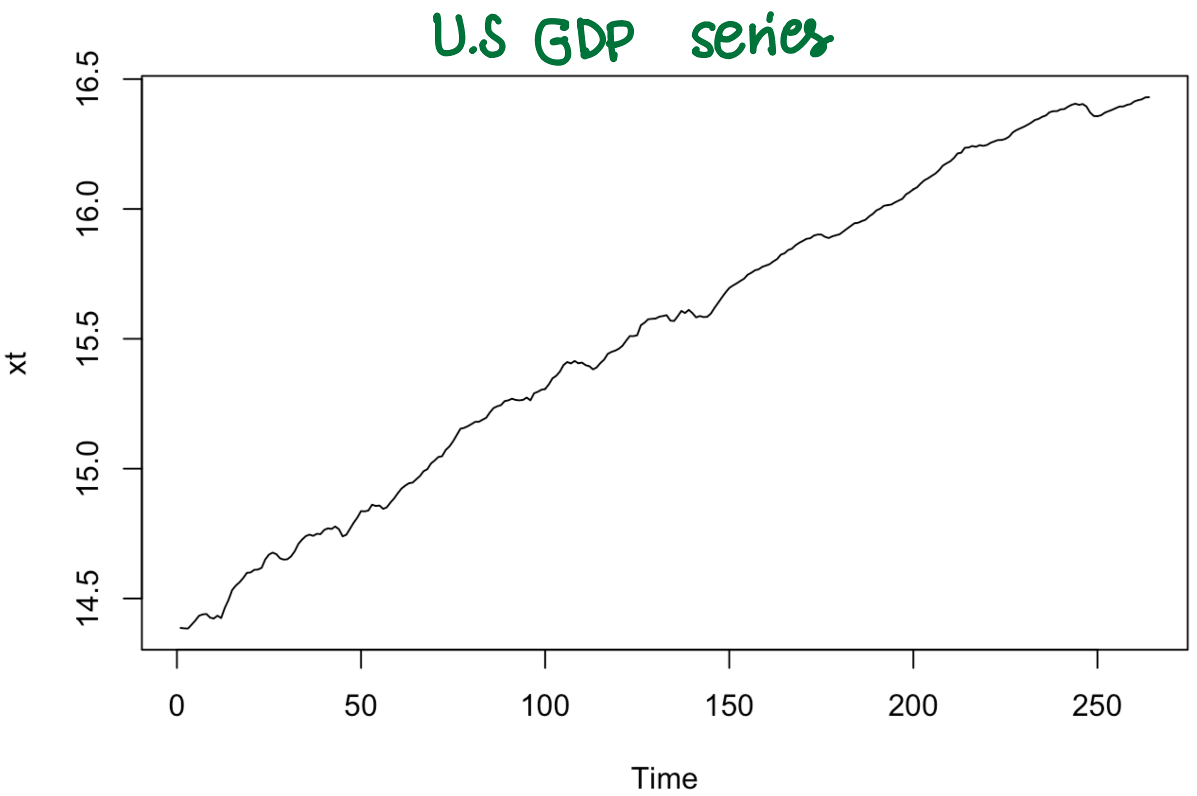
```
da=read.table("q-gdpmc1.txt",header=T)
head(da)
```

```
##   year month day     gdp
## 1 1947     1   1 1770691
## 2 1947     4   1 1767976
## 3 1947     7   1 1766523
## 4 1947    10   1 1793310
## 5 1948     1   1 1821809
## 6 1948     4   1 1855345
```

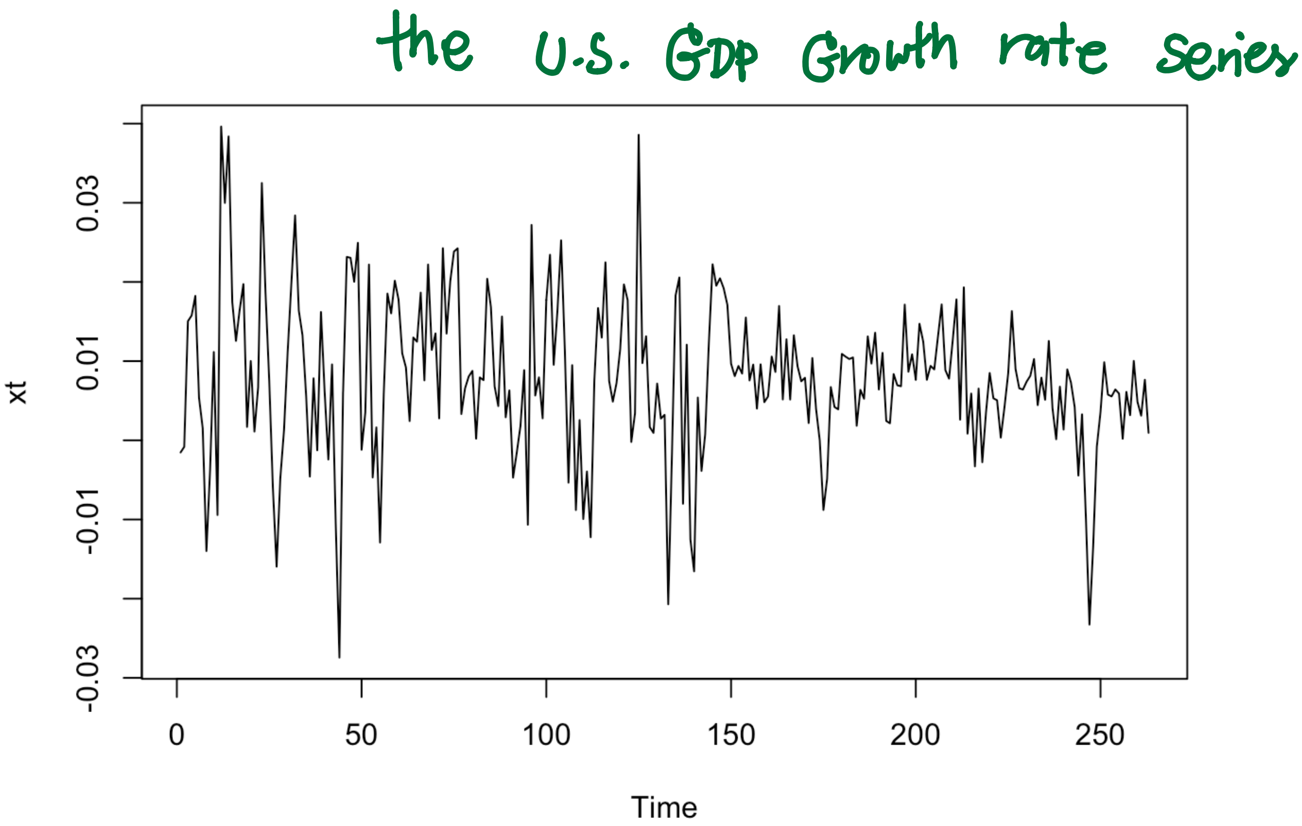
1.(30 points) Consider the growth rates of the U.S. real gross domestic product (GDP) from 1947.I to 2012.IV. The original data, from Federal Reserve Bank of St Louis, are in the file q-gdpmc1.txt (year, month, day, gnp), and the GDP are in millions of 2005 chained dollars. The growth rate is the first differenced series of the log(GDP).

```
gdp=log(da$gdp)

gdprate=diff(gdp)

ts.plot(gdp,ylab="xt")
```

We calculate the growth rate first.




```
ts.plot(gdprate,ylab="xt")
```



```
m1=ar(gdprate,method="mle")
m1
```

(a) Build an AR model for the growth rate series. [Use the command ar with method mle to find the order.] Perform model checking to validate the fitted model. Write down the model.

(b) Does the model confirm the existence of business cycles? Why?

① Find the optimal lag by using the AIC

```
##
## Call:
## ar(x = gdprate, method = "mle")
##
## Coefficients:
##       1        2        3
##  0.3475   0.1283  -0.1121
##
## Order selected 3  sigma^2 estimated as  8.17e-05
```

② the optimal lag = 3

③ estimate the AR(3)

```
m2=arima(gdprate,order=c(3,0,0)) #  <== AR(3) model is selected
m2
```

we estimate series X_t

```
##
## Call:
## arima(x = gdprate, order = c(3, 0, 0))
##
## Coefficients:
##          ar1     ar2      ar3  intercept
##       0.3476  0.1282  -0.1122     0.0077
## s.e.  0.0612  0.0644   0.0613     0.0009
##
## sigma^2 estimated as 8.17e-05:  log likelihood = 864.45,  aic = -1718.9
```

```
tsdiag(m2,gof=20)
```

Ans 1(b)

Since the U.S. GDP growth can be estimated by AR(3) with the complex roots in the fitted AR polynomial, it confirms the existence of business cycles!

ans: 1(a)

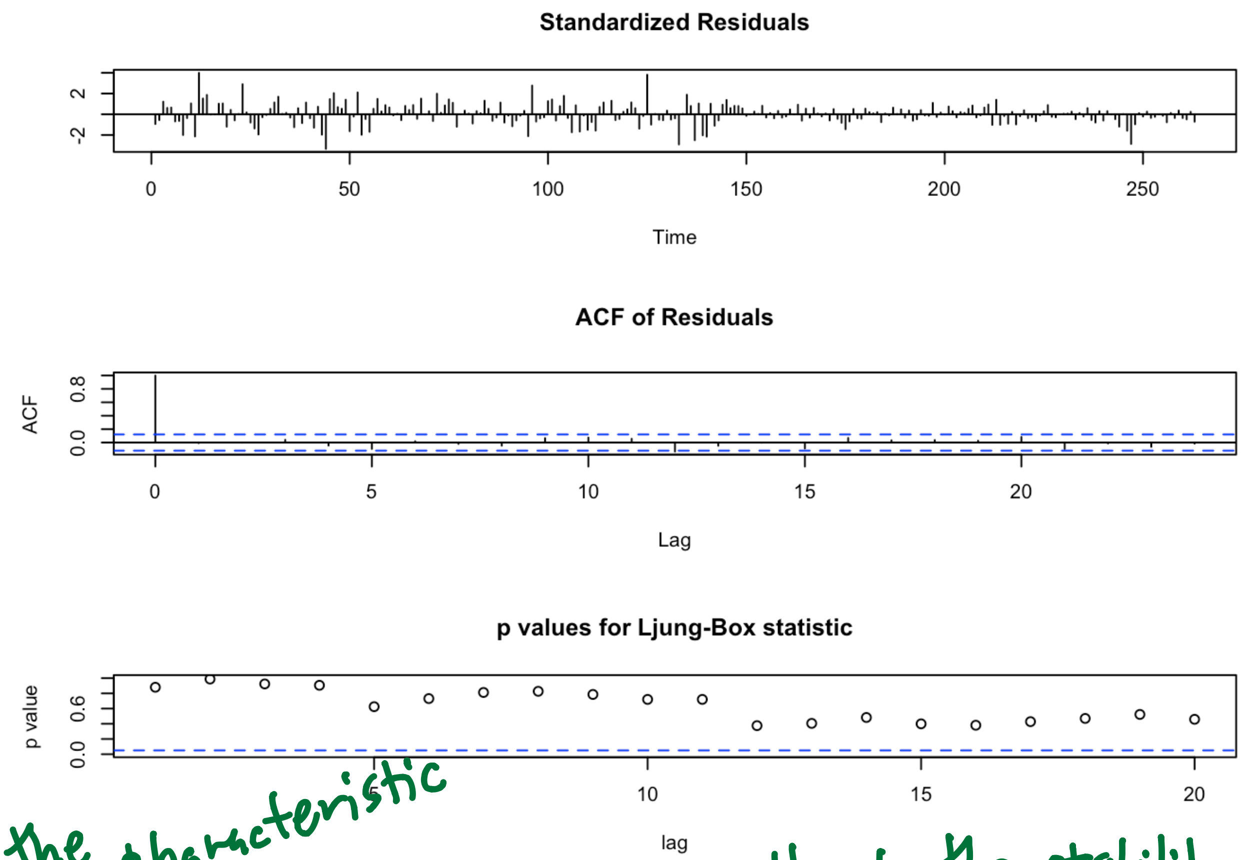
Population equation:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + a_t$$

The estimated result:

$$\therefore \hat{X}_t = \underset{(0.0009)}{0.0077}[1-0.3476-0.1282+0.1122] + \underset{(0.0612)}{0.3476} X_{t-1} + \underset{(0.0644)}{0.1282} X_{t-2} - \underset{(0.0613)}{0.1122} X_{t-3}$$

$$\hat{\sigma}_a^2 = 8.17 \times 10^{-5}$$



use the characteristic eq.

check the stability of AR(3)

```
p1=c(1,-m2$coef[1:3])
p1
```

```
##                  ar1        ar2        ar3
##  1.0000000 -0.3476309 -0.1282098  0.1121601
```

```
roots=polyroot(p1)
Mod(roots)
```

modulus

From $x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + a_t$

```
## [1] 2.025805 2.172531 2.025805
```

> 1 ∴ weakly stationarity

```
p2=predict(m2,8)

names(p2)
```

$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \phi_3 x_{t-3} = \phi_0 + a_t$

```
## [1] "pred" "se"
```

$$\underbrace{(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)}_{\text{Characteristic equation}} x_t = \phi_0 + a_t$$

```
p2$pred-1.96*p2$se # <== lower bound of 95% intervals
```

If we use the characteristic eq. to check the stability, it is weakly stationarity if

⇒ the reverse characteristic eq.

$$\lambda^3 - \phi_1 \lambda^2 - \phi_2 \lambda - \phi_3 = 0$$

every modulus has the value greater than 1!

If we use the reverse characteristic eq. to check the stability of AR(p), it is weakly stationarity when every modulus has the value less than 1.

```
## Time Series:
## Start = 264
## End = 271
## Frequency = 1
## [1] -0.01183604 -0.01252798 -0.01153506 -0.01152362 -0.01136471 -0.01148177
## [7] -0.01150285 -0.01154238
```

```
p2$pred+1.96*p2$se # <== upper bound of 95% intervals
```

```
## Time Series:
## Start = 264
## End = 271
## Frequency = 1
## [1] 0.02359671 0.02498469 0.02700164 0.02701896 0.02717787 0.02707152
## [7] 0.02705246 0.02701374
```

(c) Obtain 1-step to 8-step ahead point and 95 % interval forecasts for the U.S. quarterly GDP growth rate at the forecast origin October 1, 2012 (the last data point).

Conduct the 95% CI for $\hat{x}_h(l)$.

```
#### Problem 3 ####
m3=arima(gdprate,order=c(1,0,0))
m3
```

```
##
## Call:
## arima(x = gdprate, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.3705     0.0077
## s.e.  0.0572     0.0009
##
## sigma^2 estimated as 8.344e-05:  log likelihood = 861.71,  aic = -1717.43
```

2.(40 points) Consider, again, the quarterly U.S. real GDP growth rates from 1947 to 2012 in Problem 1.

(a) Fit a simple AR(1) model to the series. Write down the model.

2.a

Ans:

$$\hat{X}_t = \underset{(0.0009)}{0.0077}(1-0.3705) + \underset{(0.0572)}{0.3705}\, X_{t-1}$$

$$\hat{\sigma}_a^2 = 8.344 \times 10^{-5}$$

Ans: we check the model adequacy

```
tsdiag(m3,gof=20)
dim(da)
```

(b) Is the model adequate? Why?

```
## [1] 264   4
```

by using the Ljung-Box test (see figures on next page)

```
source("/Users/wasin_siwasarit/Desktop/backtest.R")

## 263-20   <== find the first forecast origin
backtest(m2,gdprate,243,1)
```

∴ model is adequate!

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.008160635
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.005301588
```

(c) Compare the AR(1) model with the AR model built in Problem 1. In terms of in-sample fitting, which model is preferred? Why?

Ans: Based on the in-sample fitting, the AR(3) model is selected because its has the smaller

AIC (−1718.9) and smaller residual variance ($\hat{\sigma}_a^2$) = 8.17×10^{-5}.

```
backtest(m3,gdprate,243,1)
```

```
## [1] "RMSE of out-of-sample forecasts"
## [1] 0.008126095
## [1] "Mean absolute error of out-of-sample forecasts"
## [1] 0.005362249
```

(d) Use **backtest** to compare the two AR models. You may use the data in 2008-2012 (inclusive) as the forecasting period. Which model is preferred? Why?

```
###Question3##############################
require(quantmod)
```

<u>Ans</u>: Based on the backtest, the selection is mixed. The RMSE selects the AR(1) model, but the MAE prefers the AR(3) model.

```
## Loading required package: quantmod
```

```
## Warning: package 'quantmod' was built under R version 3.4.4
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 3.4.4
```

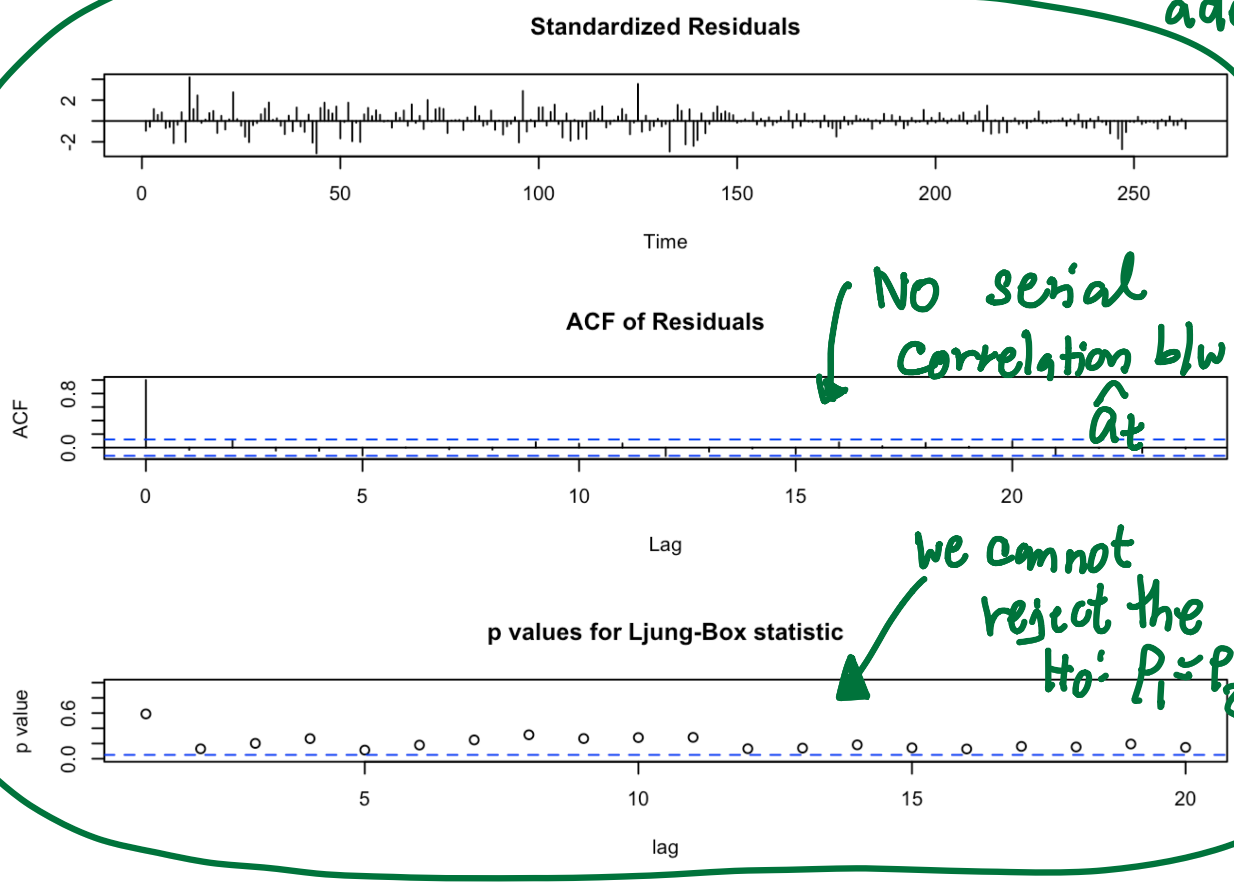
```
##
## Attaching package: 'zoo'
```

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

```
## Loading required package: TTR
```

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

checking model adequacy



No serial correlation b/w $\hat{a}_t$

we cannot reject the $H_0: \rho_1 = \rho_2 = \dots \rho_m = 0$

```
getSymbols('GOLDAMGBD228NLBM', src='FRED')
```

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.
```

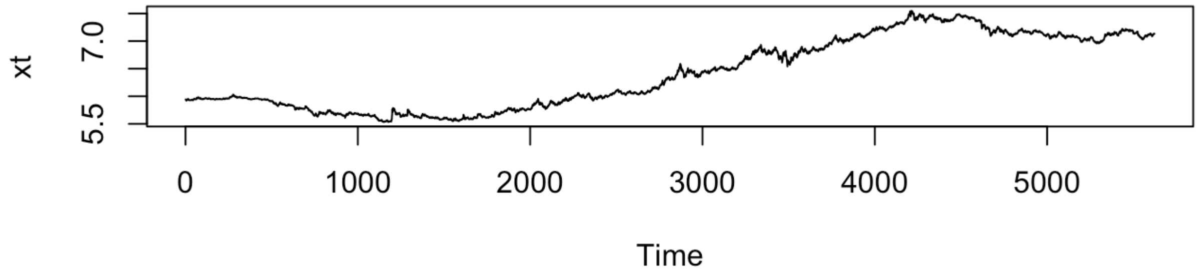
```
## Warning in strptime(xx, f <- "%Y-%m-%d", tz = "GMT"): unknown timezone
## 'zone/tz/2018i.1.0/zoneinfo/Asia/Bangkok'
```

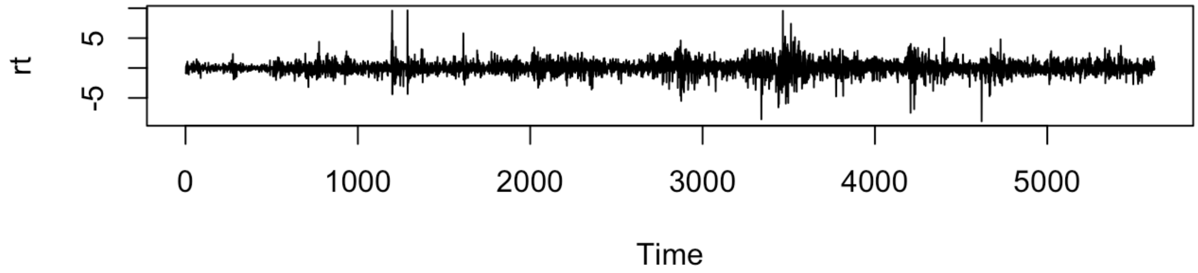
```
## [1] "GOLDAMGBD228NLBM"
```

```
GOLD <- GOLDAMGBD228NLBM[6982:12784]
idx <- c(1:nrow(GOLD))[is.na(GOLD)]
GOLD <- GOLD[-idx]
xt <- log(as.numeric(GOLD))
rt <- 100*diff(xt)


par(mfcol=c(2,1))
ts.plot(xt,ylab="xt")
ts.plot(rt,ylab="rt")
```

3.(50 points) (Commodity price). Consider the daily gold fixing price 10:30 am (London time) in London Bullion Market in U.S. dollars per Troy ounce from January 3, 1995 to March 30, 2017. The data can be obtained from FRED using the quantmod package. Since there are some missing values, we need to remove them before analysis. Let x_t = log(gold price). See instructions below.

(a)Obtain the time plots of xt and rt (in one page, using the command par(mfcol=c(2,1)).





(b) Let $r_t = 100*(x_t - x_{t-1})$ be the return series of the gold prices, in percentages. Consider the r_t series. Test $H_0 : \rho_1 = ... = \rho_{12} = 0$ versus $H_a : \rho_i \neq 0$ $\exists 1 \leqslant i \leqslant 12$. Draw the conclusion.

```
Box.test(rt,lag=12,type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  rt
## X-squared = 44.989, df = 12, p-value = 1.035e-05
```

```
ar(rt,method="mle",order.max=20)
```

Ans: $H_0: \rho_1 = \rho_2 = \dots \rho_{12} = 0$ (series of return of gold prices)

$H_1: \exists \rho_i \neq 0$

Given $\alpha = 0.05$

we reject the H_0:

that means there is the serial correlation in this series with 95%

CI.

```
##
## Call:
## ar(x = rt, order.max = 20, method = "mle")
##
## Coefficients:
##       1        2        3        4        5        6        7        8
## -0.0437   0.0057   0.0139   0.0019   0.0376  -0.0078  -0.0449   0.0206
##       9       10       11       12       13       14
##  0.0106  -0.0225  -0.0214  -0.0240  -0.0152   0.0226
##
## Order selected 14  sigma^2 estimated as  1.154
```

optimal order = 14

(c)Use the following command to specify the order of an AR model for rt.

```
ar(rt,method= "mle",order.max=20)
```

```
m1 <- arima(rt,order=c(14,0,0))
m1
```

(d) Build an AR model for r_t, including model checking. Refine the model by excluding all estimates with t-ratio less than 1.645. Write down the fitted model.

```
##
## Call:
## arima(x = rt, order = c(14, 0, 0))
##
## Coefficients:
##           ar1     ar2     ar3     ar4     ar5      ar6      ar7     ar8
##       -0.0437  0.0057  0.0139  0.0019  0.0376  -0.0078  -0.0449  0.0206
## s.e.   0.0133  0.0133  0.0133  0.0133  0.0133   0.0133   0.0133  0.0133
##          ar9     ar10     ar11     ar12     ar13    ar14  intercept
##       0.0106  -0.0224  -0.0214  -0.0240  -0.0152  0.0226     0.0211
## s.e.  0.0133   0.0133   0.0133   0.0133   0.0133  0.0133     0.0134
##
## sigma^2 estimated as 1.154:  log likelihood = -8381.14,  aic = 16794.28
```

```
c2 <- c(NA,0,0,0,NA,0,NA,0,0,NA,0,0,0,NA)
m1a <- arima(rt,order=c(14,0,0),fixed=c2,include.mean=F)
```

we exclude the intercept.

```
## Warning in arima(rt, order = c(14, 0, 0), fixed = c2, include.mean = F):
## some AR parameters were fixed: setting transform.pars = FALSE
```

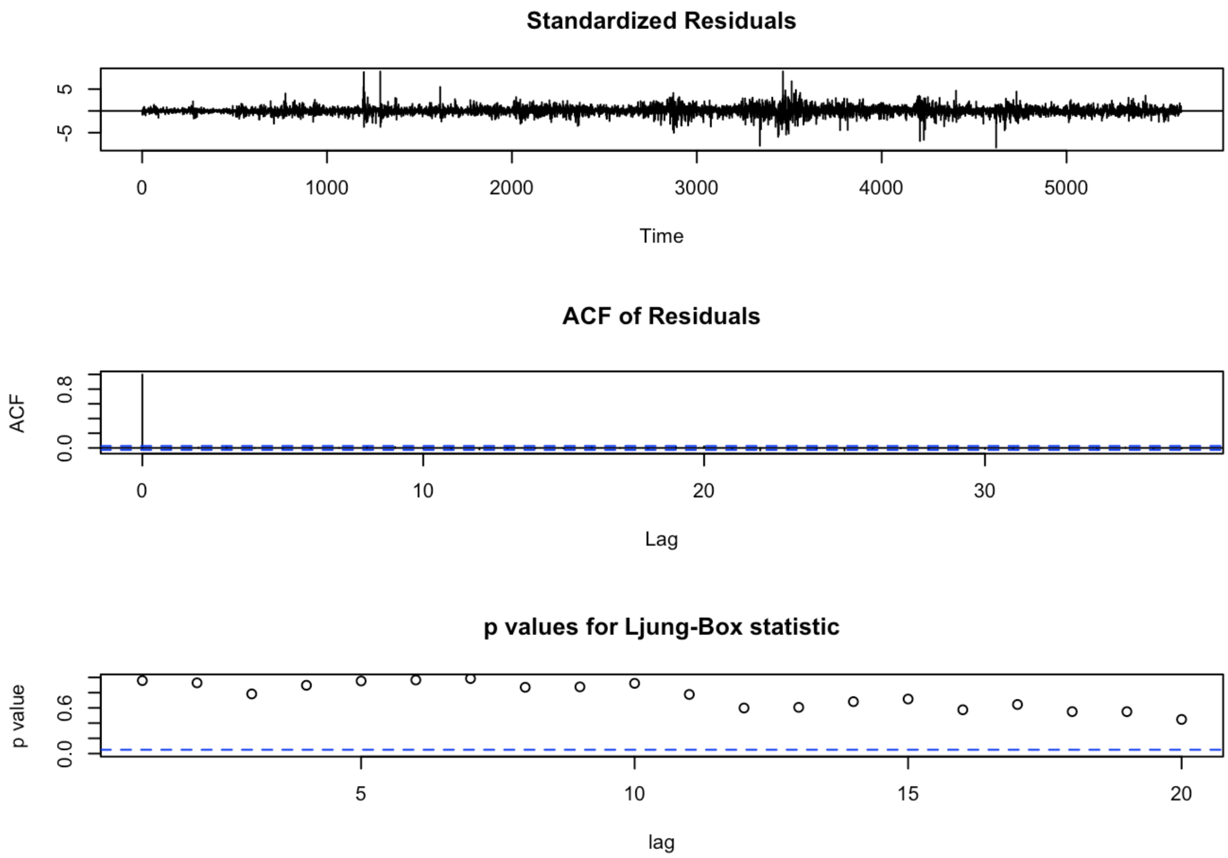
```
m1a
```

(d) After we refine the model, the fitted model is

$$r_t = \underset{(0.0133)}{-0.0432}\, r_{t-1} + \underset{(0.0133)}{0.0396}\, r_{t-5} - \underset{(0.0133)}{0.0456}\, r_{t-7} - \underset{(0.0133)}{0.0224}\, r_{t-10} + \underset{(0.0133)}{0.0288}\, r_{t-14}$$

```
##
## Call:
## arima(x = rt, order = c(14, 0, 0), include.mean = F, fixed = c2)
##
## Coefficients:
##           ar1  ar2  ar3  ar4     ar5  ar6      ar7  ar8  ar9     ar10
##       -0.0432    0    0    0  0.0396    0  -0.0456    0    0  -0.0224
## s.e.   0.0133    0    0    0  0.0133    0   0.0133    0    0   0.0133
##       ar11  ar12  ar13    ar14
##          0     0     0  0.0228
## s.e.     0     0     0  0.0133
##
## sigma^2 estimated as 1.157:  log likelihood = -8387.84,  aic = 16787.68
```

```
tsdiag(m1a,gof=20)
```




```
pm1 <- predict(m1a,4)
pm1
```

```
## $pred
## Time Series:
## Start = 5623
## End = 5626
## Frequency = 1
## [1] -0.041556505  0.020279511  0.002397155 -0.030992681
##
## $se
## Time Series:
## Start = 5623
## End = 5626
## Frequency = 1
## [1] 1.075758 1.076759 1.076761 1.076761
```

(e) Use the fitted AR model to compute 1-step to 4-step ahead forecasts of rt at the forecast origin March 30, 2017. Also, compute the corresponding 95 % interval forecasts.

the 95% CI for r_t : $\hat{r}_h(l) \pm 1.96\, se(e_h(l))$

$\boxed{\text{LB}}$ (Lower bound)	$\boxed{\text{UB}}$ Upper bound
$-0.0416 - 1.96(1.0758)$	$-0.0416 + 1.96(1.0758)$
$0.0203 - 1.96(1.0768)$	$0.0203 + 1.96(1.0768)$
$0.0024 - 1.96(1.0768)$	$0.0024 + 1.96(1.0768)$
$-0.031 - 1.96(1.0768)$	$-0.031 + 1.96(1.0768)$