Solution for Problem Set 3

Mahrad Sharifvaghefi

1. (a) We have $(1 - 0.5L + 0.04L^2)y_t = u_t$. By setting $1 - 0.5L + 0.04L^2 = 0$, we get,

$$L^* = \frac{0.5 \pm \sqrt{0.25 - 0.16}}{0.08} = \frac{0.5 \pm 0.3}{0.08}$$

So $L_1^* = 10$ and $L_2^* = 2.5$. Since all the roots are outside of the unit circle, we can conclude that the process is weakly stationary.

(b) $\mathbb{E}(y_t) = 0.5\mathbb{E}(y_{t-1}) - 0.04\mathbb{E}(y_{t-2}) = (0.5 - 0.04)\mathbb{E}(y_t) \Rightarrow \mathbb{E}(y_t) = 0.46\mathbb{E}(y_t) \Rightarrow \mathbb{E}(y_t) = 0.$

$$\begin{aligned} \text{var}(y_t) =& \text{cov}(y_t, y_t) = 0.5\text{cov}(y_{t-1}, y_t) - 0.04\text{cov}(y_{t-2}, y_t) + \text{cov}(\varepsilon_t, y_t) \\ =& 0.5\text{cov}(y_{t-1}, y_t) - 0.04\text{cov}(y_{t-2}, y_t) + \text{var}(\varepsilon_t) \Rightarrow \end{aligned}$$

or

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + 1. \tag{1}$$

We also have

$$\begin{aligned} \text{cov}(y_t, y_{t-k}) =& \text{cov}(0.5y_{t-1} - 0.04y_{t-2} + \varepsilon_t, y_{t-k}) \\ =& 0.5\text{cov}(y_{t-1}, y_{t-k}) - 0.04\text{cov}(y_{t-2}, y_{t-k}) \end{aligned}$$

for all $k = 1, 2, \cdots$. or

$$\gamma(k) = 0.5\gamma(k-1) - 0.04\gamma(k-2) \text{ for all } k = 1, 2, \cdots.$$

By substituting, $k = 1$ and $k = 2$, we get,

$$\gamma(1) = 0.5\gamma(0) - 0.04\gamma(1), \tag{2}$$

and

$$\gamma(2) = 0.5\gamma(1) - 0.04\gamma(0). \tag{3}$$

From equations (1), (2) and (3), we can find the values of $\gamma(0)$, $\gamma(1)$, and $\gamma(2)$.

(c) Given the value of $\gamma(0)$, $\gamma(1)$, and $\gamma(2)$, we can compute $\gamma(k)$ for $k = 3, 4, \cdots$ by

$$\gamma(k) = 0.5\gamma(k-1) - 0.04\gamma(k-2). \tag{4}$$

(d) We have $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ for all $k = 0, 1, 2, \cdots$. So, $\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$. By dividing both side of equation (2) by $\gamma(0)$, we have

$$\rho(1) = 0.5\rho(0) - 0.04\rho(1) \Rightarrow \rho(1) = \frac{0.5}{1.04} \approx 0.48. \tag{5}$$

By dividing both side of equation (3) by $\gamma(0)$, we have

$$\rho(2) = 0.5\rho(1) - 0.04\rho(0) \approx 0.2 \tag{6}$$

Finally, by dividing both side of equation (4) by $\gamma(0)$, we have

$$\rho(k) = 0.5\rho(k-1) - 0.04\rho(k-2), \tag{7}$$

for $k = 3, 4, \cdots$. So given the values of $\rho(1)$, and $\rho(2)$, we can compute $\rho(k)$ for $k = 3, 4, \cdots$ from (7).

(e) The partial auto-correlation of order one is equal the auto-correlation of order one, $\alpha(1) = \rho(1) \approx 0.48$. Since the process is AR(2), the partial auto-correlation of order two is equal to the slope coefficient of y_{t-2} in the data generating process for y_t, $\alpha(2) = -0.04$. Again, since the process is AR(2), all the higher order partial auto-correlations are equal to zero, $\alpha(k) = 0$ for all $k = 3, 4, \cdots$.

2. (a) $\mathbb{E}(y_t) = 1 + \mathbb{E}(u_t) + 2\mathbb{E}(u_{t-3}) = 1$.

   (b) $\text{var}(y_t) = \text{var}(u_t) + 4\text{var}(u_{t-3}) = 5$.

   (c) We have $y_t = 1+u_t+2u_{t-3}$ and $y_{t-3} = 1+u_{t-3}+2u_{t-6}$. Hence, $\text{cov}(y_t, y_{t-k}) = 2\text{cov}(u_{t-3}, u_{t-3}) = 2\text{var}(u_{t-3}) = 2$. For any other choice of $k > 0$, where exists no common shocks among y_t and y_{t-k} and hence $\gamma(k) = 0$.

   (d) We can write y_t in lag operator format as

$$y_t = 1 + (1 + 2L)u_t.$$

   By setting $1 + 2L = 0$, we get $L = -\frac{1}{2}$ which is in absolute value less than one. So the MA process is not invertible.

   (e) $\rho_y(0) = 1$ and $\rho_y(3) = \frac{2}{5}$. $\rho_y(k) = 0$ for all the other values of $k > 0$ since $\gamma(k) = 0$.

   (f) We have $\rho_x(0) = 1$, $\rho_x(3) = \frac{\theta}{1+\theta^2}$ and $\rho_x(k) = 0$ for all the other values of $k > 0$. So $\rho_x(k) = \rho_y(k)$ for all $k \geq 0$, if and only if $\rho_x(3) = \rho_y(3) = \frac{2}{5}$. So we have

$$\frac{\theta}{1+\theta^2} = 2/5 \text{ or } 2\theta^2 - 5\theta + 2 = 0$$

   By solving for *theta*, we get

$$\theta_1 = \frac{5 + \sqrt{25 - 16}}{4} = 2 \text{ and } \theta_2 = \frac{5 - \sqrt{25 - 16}}{4} = 0.5.$$

   But we have shown in part (d) that if $\theta = \theta_1$ then the MA process in not invertible. Here we show that if $\theta = \theta_2 = 0.5$ then the MA process is invertible. By writing x_t in the lag operator format we get

$$x_t = \mu + (1 + \theta L)v_t.$$

   Now if we set $1 + \theta L = 0$, then $L = -\frac{1}{\theta}$. But since $\theta = \theta_2 = 0.5$, the root L is in absolute value greater that one and hence the MA process is invertible.

3.

The following lines call the required packages and source the required local functions for this problem set

```
source("r_functions/model_selection_function.R") # function for model selection
source("r_functions/ols_function.R") # function for OLS estimation
source("r_functions/t_test_function.R") # fucntion for t test
source("r_functions/expanding_window_forecast_function.R") # function for forecasting
#using expanding windows
source("r_functions/rolling_window_forecast_function.R") # function for forecasting
# using rolling windows
library("quantmod") # add quantmod to the list of Packages
library("fBasics") # add fBasics to the list of Packages
```

Now, we start with answering the questions.

(a) We can use the getSymbols command to fetch the data from Fred as

```
getSymbols(Symbols ="GNP",src = "FRED",warnings = FALSE) # Download Quarterly data for GNP
```

```
## [1] "GNP"
```

(b)

```r
Y <- as.matrix(GNP[,1])
DY <- diff(Y)/Y[1:(dim(Y)[1]-1),]
head(DY)
```

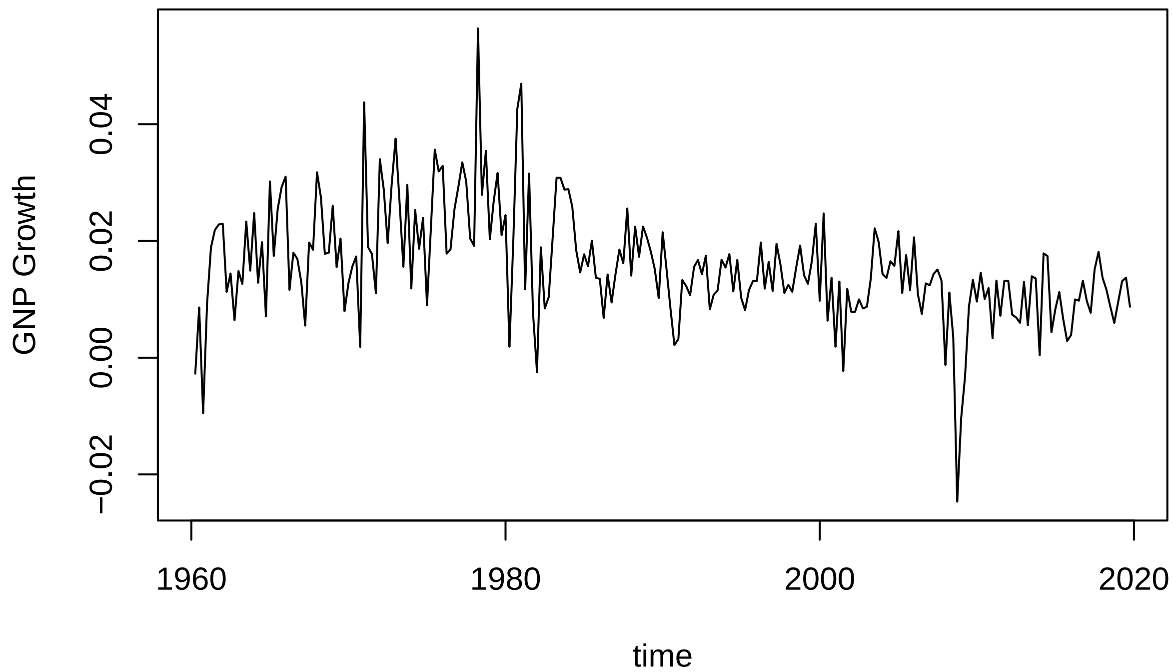
```
##                   GNP
## 1947-04-01 0.01196435
## 1947-07-01 0.01478570
## 1947-10-01 0.04094274
## 1948-01-01 0.02357260
## 1948-04-01 0.02587849
## 1948-07-01 0.02420397
```

```r
tail(DY)
```

```
##                    GNP
## 2022-04-01 0.022580326
## 2022-07-01 0.017074777
## 2022-10-01 0.015584918
## 2023-01-01 0.014233663
## 2023-04-01 0.009771869
## 2023-07-01 0.019756463
```

```r
keep_data <- seq(from = as.Date("1960-04-01"), to = as.Date("2019-10-1"), by = "quarter")
DY_new = as.matrix(DY[as.Date(rownames(DY)) %in% keep_data,])
colnames(DY_new) = "GNP Growth"
n_obs = dim(DY_new)[1]
DY_new_date = as.Date(row.names(DY_new))

plot(x = DY_new_date, y = DY_new,xlab='time',ylab='GNP Growth',type='l',col="black")
```

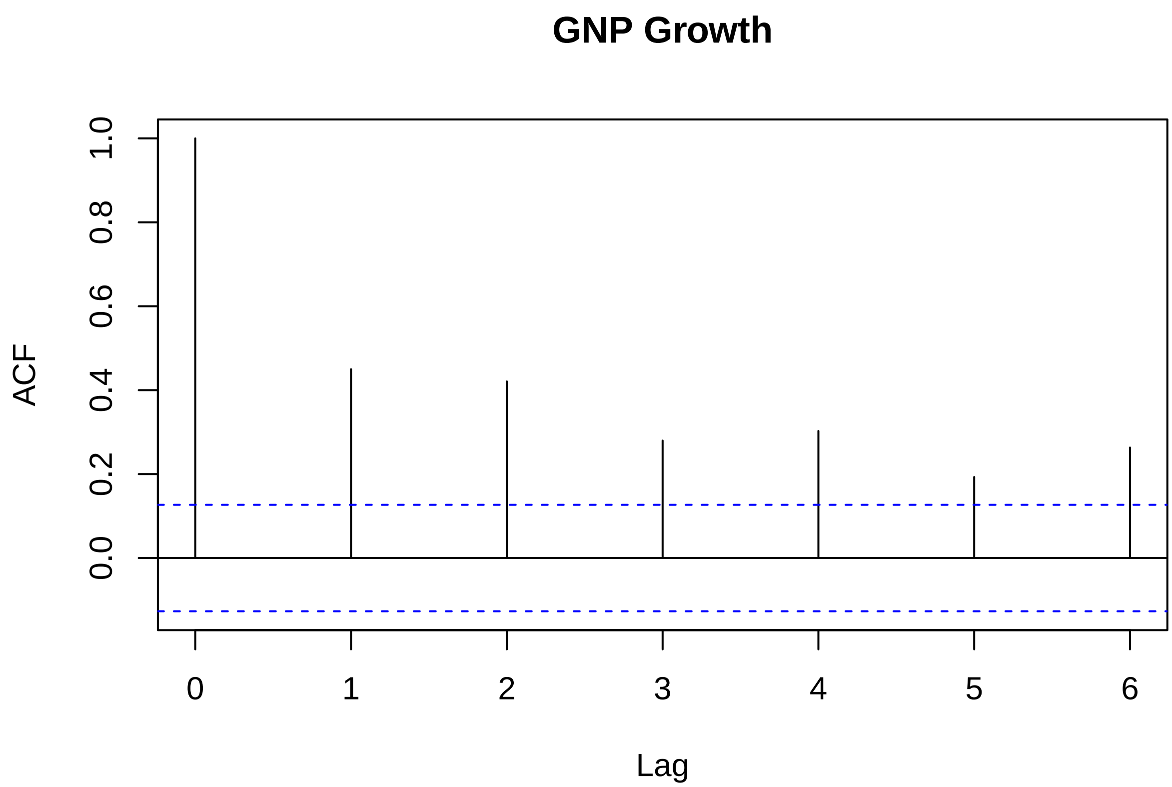



```r
basicStats(DY_new)
```

```
##              GNP.Growth
## nobs         239.000000
## NAs            0.000000
## Minimum       -0.024656
## Maximum        0.056421
## 1. Quartile    0.010217
## 3. Quartile    0.019952
## Mean           0.015665
## Median         0.014327
## Sum            3.744017
## SE Mean        0.000618
## LCL Mean       0.014448
## UCL Mean       0.016883
## Variance       0.000091
## Stdev          0.009556
## Skewness       0.378849
## Kurtosis       2.611224
```

(c)

```r
acf(DY_new,lag=round(n_obs^(1/3))) # command to obtain sample ACF of the data
```



Yes, since the computed sample autocorrelation of order 1 to 6 are significantly different from zero.

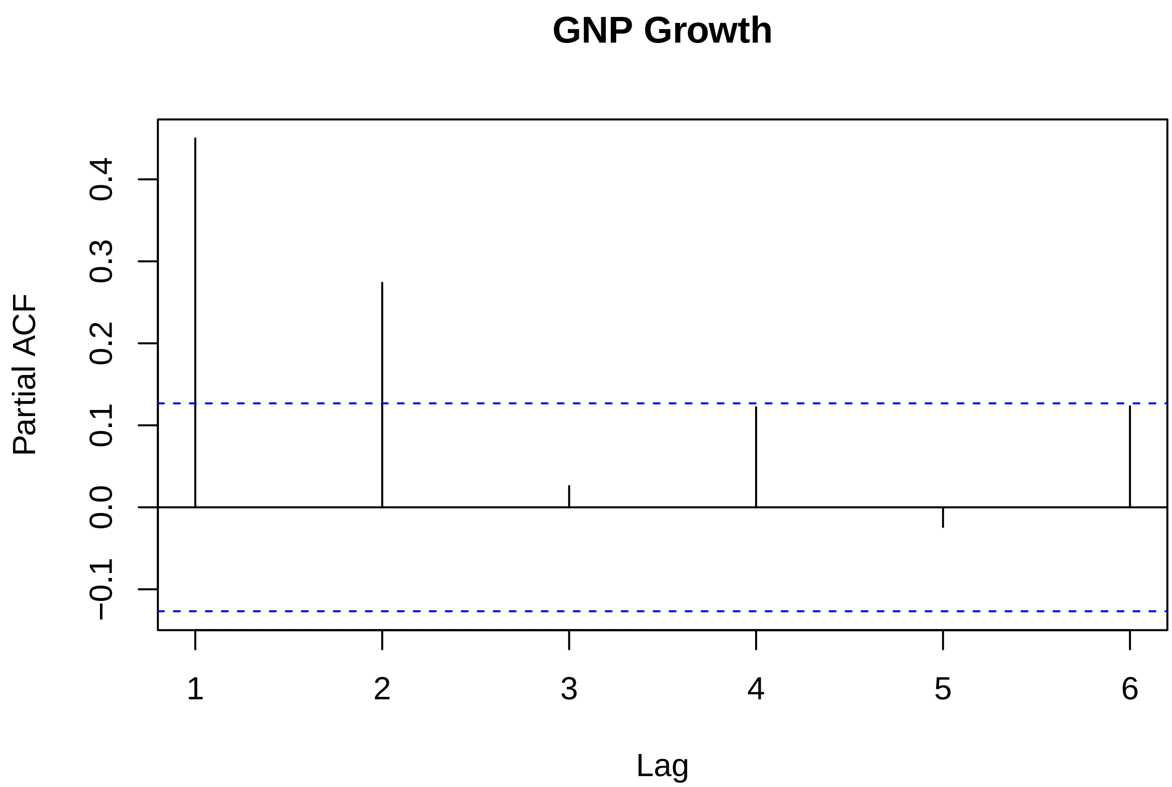
```
# applying Ljung and Box (1978) joint test of auto correlations
Box.test(DY_new, lag = round(n_obs^(1/3)), type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  DY_new
## X-squared = 160, df = 6, p-value < 2.2e-16
```

As expected the Ljung and Box (1978) test reject the null that there exists no serial correlation.

(d)

```
pacf(DY_new,lag=round(n_obs^(1/3)),main="GNP Growth") # command to obtain sample PACF of the data
```



Given the partial autocorrelation function, we expect to select 2 lags, as all the higher order number of lags are insignificant. However, We can see 4 or 6 lags can also be selected as their corresponding sample partial autocorrelations are very closed to be significant.

```r
results <- model_selection(round(n_obs^(1/3)),DY_new)
aic_values = results$AIC
bic_values = results$BIC
num_lags_aic = results$op_lag_AIC
num_lags_bic = results$op_lag_BIC
num_lags_aic
```

```
## [1] 2
```

```r
num_lags_bic
```

```
## [1] 2
```

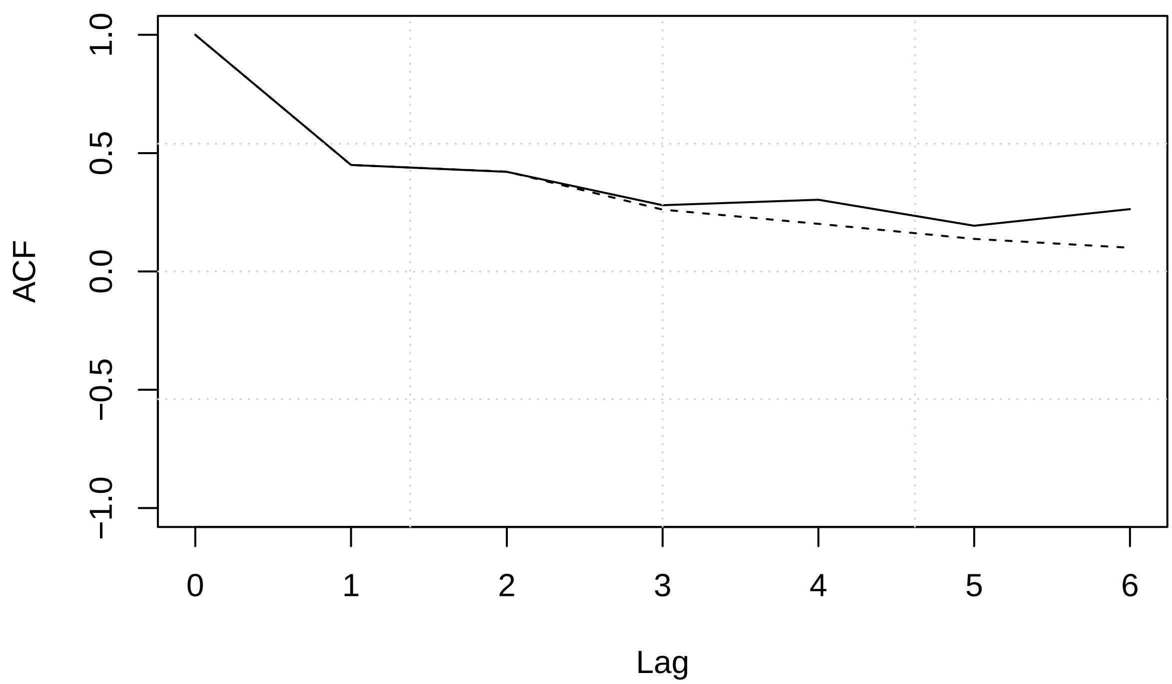
AIC and BIC select 4 and 2 lags, respectively which is expected from PACF analysis.

(e)

```r
num_lags = num_lags_aic
lags_DY_new = matrix(NA,nrow = n_obs, ncol = num_lags)
for (i in 1:num_lags) {
  lags_DY_new[(i+1):n_obs,i] = as.matrix(DY_new[1:(n_obs-i),1])
}
intercept = matrix(1,n_obs)
X = cbind(intercept,lags_DY_new)
y = DY_new
```

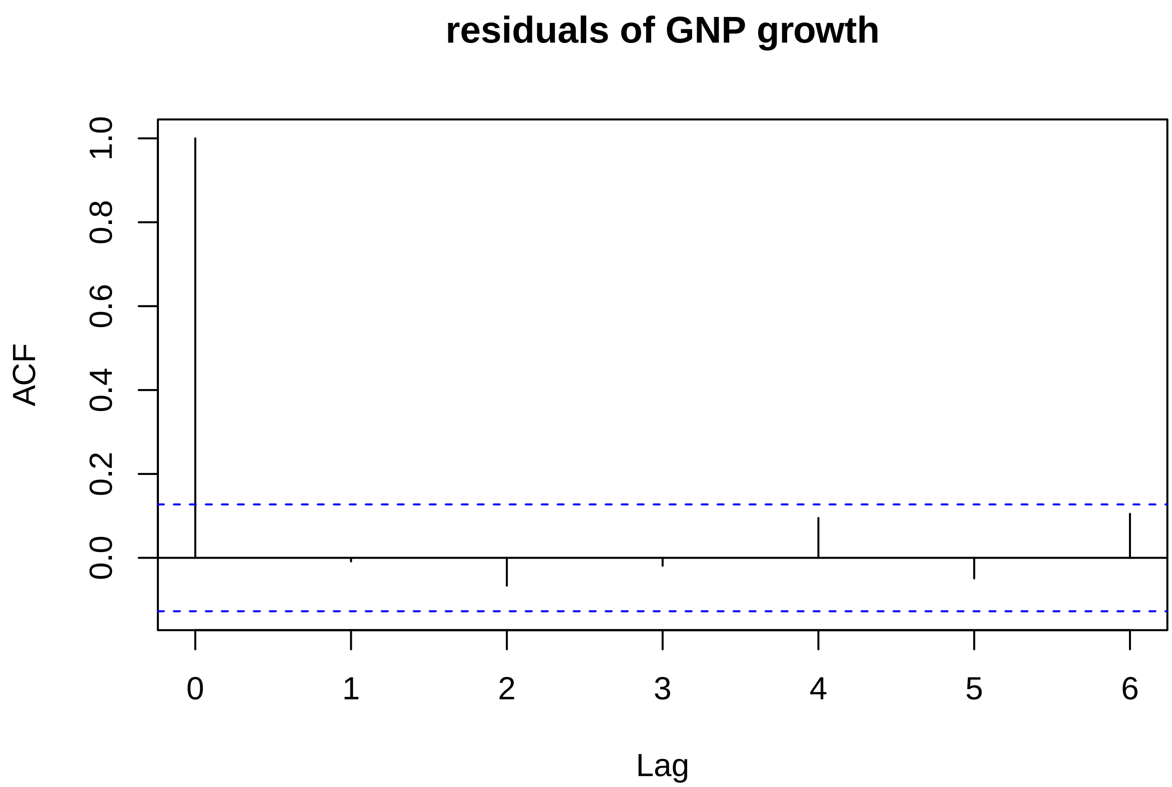
```
reg_result = ols(X[(num_lags+1):n_obs,],as.matrix(y[(num_lags+1):n_obs,1]))
beta_hat = reg_result$beta_hat

ar_coeff <- as.numeric(beta_hat[2:(num_lags+1)])
ma_coeff <- 0
ACF = acf(DY_new,lag=round(n_obs^(1/3)),plot = FALSE) # command to obtain sample ACF of the data
TACF <- ARMAacf(ar_coeff, ma_coeff, lag.max = round(n_obs^(1/3))) # command to obtain theorical ACF
plot(c(0:round(n_obs^(1/3))),ACF$acf,type='l',xlab='Lag',ylab='ACF',ylim=c(-1,1))
lines(0:round(n_obs^(1/3)),TACF,lty=2)
grid(nx = 4, ny = 4)
```



As can be seen in the above figure, the empirical autocorrelation function is close to the AIC selected theoretical autocorrelation function.

```
residuals = reg_result$u_hat # get the AR model residuals
# command to obtain sample ACF of the data
acf(residuals,lag=round(n_obs^(1/3)),main = "residuals of GNP growth")
```



The ACF analysis shows no serial dependence across observations.

```r
# applying Ljung and Box (1978) joint test of auto correlations
Box.test(residuals, lag = round(n_obs^(1/3)), type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  residuals
## X-squared = 6.6587, df = 6, p-value = 0.3536
```

As expected from the ACF analysis, Ljung and Box (1978) joint test of auto correlations cannot reject the null.