Problem Set 2

Jianqiu Bei, Xinru Huang, Jingyi Liu, Zhirui Wang

June 25, 2016

Problem 1

```
suppressMessages(library(lmtest))
suppressMessages(library(car))
ts <- read.csv("D:/Dropbox/16summer/Macroeconometrics/Problem Set2/1.csv")
```

i

Estimate the dynamic version of Okun's law

```
GDP <- ts[,3]
L.GDP <- c(NA,GDP[-239])
L2.GDP <- c(NA,L.GDP[-239])
L3.GDP <- c(NA,L2.GDP[-239])
arima(ts[,4],order = c(2,0,0),xreg =cbind(GDP,L.GDP,L2.GDP))
```

```
##
## Call:
## arima(x = ts[, 4], order = c(2, 0, 0), xreg = cbind(GDP, L.GDP, L2.GDP))
##
## Coefficients:
##          ar1      ar2  intercept      GDP    L.GDP   L2.GDP
##       0.2787  -0.1252     1.4694  -0.2036  -0.1393  -0.0686
## s.e.  0.0655   0.0668     0.1083   0.0158   0.0154   0.0155
##
## sigma^2 estimated as 0.7897:  log likelihood = -308.36,  aic = 630.72
```

Although the magnitude of the coefficients are not the same as the paper, the sign of the coefficients are consistent with the paper.

ii

```
arima(ts[-(2:3),4],order = c(2,0,0),xreg =cbind(GDP,L.GDP,L2.GDP)[-c(2:3),])$aic
```

```
## [1] 628.5374
```

```
arima(ts[-(2:3),4],order = c(3,0,0),xreg =cbind(GDP,L.GDP,L2.GDP,L3.GDP)[-c(2:3),])$aic
```

```
## [1] 629.604
```

```r
arima(ts[-(2:3),4],order = c(3,0,0),xreg =cbind(GDP,L.GDP,L2.GDP)[-c(2:3),])$aic
```

```
## [1] 630.0607
```

```r
arima(ts[-(2:3),4],order = c(2,0,0),xreg =cbind(GDP,L.GDP,L2.GDP,L3.GDP)[-c(2:3),])$aic
```

```
## [1] 628.6941
```

```r
arima(ts[-(2:3),4],order = c(1,0,0),xreg =cbind(GDP,L.GDP)[-c(2:3),])$aic
```

```
## [1] 645.908
```

```r
arima(ts[-(2:3),4],order = c(2,0,0),xreg =cbind(GDP,L.GDP)[-c(2:3),])$aic
```

```
## [1] 644.9237
```

```r
arima(ts[-(2:3),4],order = c(1,0,0),xreg =cbind(GDP,L.GDP,L2.GDP)[-c(2:3),])$aic
```

```
## [1] 629.4827
```

The aic of the original model is the smallest. So the author chose the right model.

iii

```r
SSRp <- sum(arima(ts[,4],order = c(2,0,0),xreg =cbind(GDP,L.GDP,L2.GDP))$residuals[-(1:2)]^2)
SSR1 <- sum(arima(ts[1:143,4],order = c(2,0,0),
                  xreg =cbind(GDP,L.GDP,L2.GDP)[1:143,])$residuals[-(1:2)]^2)
SSR2 <- sum(arima(ts[144:239,4],order = c(2,0,0),xreg =cbind(GDP,L.GDP,L2.GDP)[144:239,])$residuals^2)
chow <- ((SSRp-SSR1-SSR2)/(SSR1+SSR2))*((239-2*6)/6)
print(chow)
```

```
## [1] 2.961727
```

```r
qf(0.95,6,(239-6*2))
```

```
## [1] 2.138668
```

The chow test shows that the F-statistic is significant, so the structure is unstable over the pre-1984 and post-1984.

iv

Estimate the difference version of Okun's law

```r
lm(ts[,4]~GDP)
```

```
##
## Call:
## lm(formula = ts[, 4] ~ GDP)
##
## Coefficients:
## (Intercept)          GDP
##      0.9968      -0.2757
```

Test stability

```r
D <- c(rep(0,143),rep(1,(239-143)))
reg <- lm(ts[,4]~D+GDP*D)
coeftest(reg)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  1.200087   0.116497  10.3015  < 2e-16 ***
## D           -0.608033   0.239712  -2.5365  0.01184 *
## GDP         -0.284748   0.019110 -14.9004  < 2e-16 ***
## D:GDP        0.055751   0.057582   0.9682  0.33394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
linearHypothesis(reg,c("D=0","D:GDP=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## D = 0
## D:GDP = 0
##
## Model 1: restricted model
## Model 2: ts[, 4] ~ D + GDP * D
##
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    237 299.90
## 2    235 288.44  2    11.458 4.6675 0.01028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reject the null hypothesis. It is unstable over pre-1984 and post-1984, so we have to calculate the GDP growth rate separately.

```r
reg1 <- lm(ts[1:143,4]~GDP[1:143])
reg2 <- lm(ts[144:239,4]~GDP[144:239])
```

pre-1984 GDP growth rate

```
-reg1$coefficients[1]/reg1$coefficients[2]
```

```
## (Intercept)
##    4.214554
```

post-1984 GDP growth rate

```
-reg2$coefficients[1]/reg2$coefficients[2]
```

```
## (Intercept)
##    2.585415
```

2.(i) $\gamma_0 = E y_t y_t = E(\varepsilon_t - 0.3\varepsilon_{t-1} + 0.17\varepsilon_{t-2})^2$

$= E(\varepsilon_t^2) - 0 + 0 - 0 + 0.09E(\varepsilon_{t-1}^2) - 0 + 0 - 0 + 0.17^2 E(\varepsilon_{t-2}^2)$

$= Var(\varepsilon_t) + 0.09\, Var(\varepsilon_t) + 0.17^2\, Var(\varepsilon_t) = 1.1189\sigma^2$

$\gamma_1 = E y_t y_{t-1} = E(\varepsilon_t - 0.3\varepsilon_{t-1} + 0.17\varepsilon_{t-2})(\varepsilon_{t-1} - 0.3\varepsilon_{t-2} + 0.17\varepsilon_{t-3})$

$= -0.3E(\varepsilon_{t-1}^2) - 0.3\times 0.17\, E(\varepsilon_{t-2}^2)$

$= -0.351\sigma^2$

$\gamma_2 = E y_t y_{t-2} = E(\varepsilon_t - 0.3\varepsilon_{t-1} + 0.17\varepsilon_{t-2})(\varepsilon_{t-2} - 0.3\varepsilon_{t-3} + 0.17\varepsilon_{t-4})$

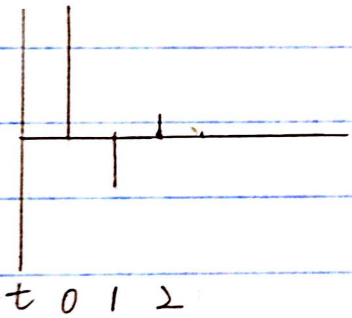
$= 0.17\, E(\varepsilon_{t-2}^2)$

$= 0.17\sigma^2$

$\gamma_3 = E y_t y_{t-3} = 0$

$\gamma_4 = E y_t y_{t-4} = 0$

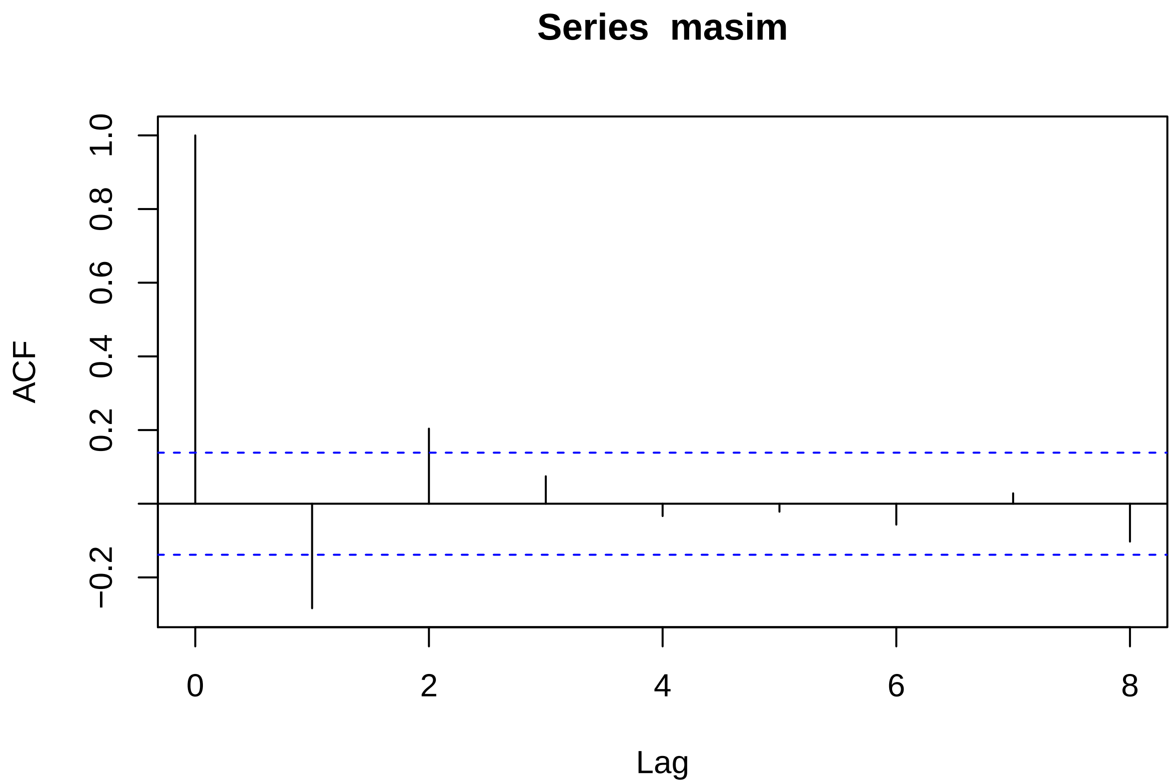
$\rho_0 = 1 \quad \rho_1 = -\frac{0.351}{1.1189} = -0.314 \quad \rho_2 = 0.152 \quad \rho_3 = 0 \quad \rho_4 = 0$



Problem 2

ii

```r
suppressMessages(library(forecast))
masim <- arima.sim(list(ma=c(-0.3,0.17)),n=200)
acf(masim,8)$acf
```




```
## , , 1
##
##             [,1]
##  [1,]  1.00000000
##  [2,] -0.28383716
##  [3,]  0.20384244
##  [4,]  0.07459754
##  [5,] -0.03351591
##  [6,] -0.02172824
##  [7,] -0.05685270
##  [8,]  0.02833621
##  [9,] -0.10289715
```

Yes, it matches the theoretical prediction because the lag 1 and 2 are significant and similar to what we calculated, and all following lags are non-significant.

```r
auto.arima(masim)
```

```
## Series: masim
## ARIMA(2,0,2) with zero mean
##
## Coefficients:
##          ar1      ar2      ma1     ma2
##       0.1180  -0.2585  -0.3910  0.5665
## s.e.  0.2004   0.2178   0.1819  0.1620
##
## sigma^2 estimated as 1.012:  log likelihood=-283.13
## AIC=576.27   AICc=576.58   BIC=592.76
```

```r
Box.test(auto.arima(masim)$residuals[1:8],type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  auto.arima(masim)$residuals[1:8]
## X-squared = 0.45316, df = 1, p-value = 0.5008
```

Based on AIC, the best model is not MA(2), and it has a non-significant Q-statistic, so it is a valid model.It is because the sampling error that we cannot fit a simulated model into its exactly original Data Generating Process.

3. $y_t = 0.28 y_{t-1} + \varepsilon_t$

$(1-0.28L) y_t = \varepsilon_t$

$$y_t = \frac{\varepsilon_t}{1-0.28L} = \left(\sum_0^{\infty} 0.28^i L^i\right)\varepsilon_t = \varepsilon_t + 0.28\,\varepsilon_{t-1} + 0.28^2 \varepsilon_{t-2} + 0.28^3 \varepsilon_{t-3} + \cdots$$

$$Var(y_t) = E(y_t^2) = E(\varepsilon_t^2) + 0.28^2 E(\varepsilon_{t-1}^2) + 0.28^4 E(\varepsilon_{t-2}^2) + \cdots$$

$$= Var(\varepsilon_t) + 0.28^2 Var(\varepsilon_t) + 0.28^4 Var(\varepsilon_t)$$

$$= 1.15 \times \frac{1}{1-0.28^2} = 1.25$$

Problem 4

```r
library(forecast)
library(tseries)
```

a

```r
ts1<- read.csv("D:/Dropbox/16summer/Macroeconometrics/Problem Set2/41.csv")
adf.test(ts1[,2])
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts1[, 2]
## Dickey-Fuller = -3.4015, Lag order = 4, p-value = 0.05802
## alternative hypothesis: stationary
```

Non-stationary. We should allow auto.arima to do difference.

```r
auto.arima(ts1[,2],stationary=FALSE)
```

```
## Series: ts1[, 2]
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1     ar2      ma1
##       0.3372  0.2269  -0.9594
## s.e.  0.1099  0.1069   0.0491
##
## sigma^2 estimated as 5.178:  log likelihood=-229.98
## AIC=467.96   AICc=468.37   BIC=478.5
```

```r
Box.test(auto.arima(ts1[,2])$residuals,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  auto.arima(ts1[, 2])$residuals
## X-squared = 0.0054382, df = 1, p-value = 0.9412
```

The final model is ARIMA(2,1,1).

b

```r
ts2<- read.csv("D:/Dropbox/16summer/Macroeconometrics/Problem Set2/42.csv")
adf.test(ts2[,2])
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts2[, 2]
## Dickey-Fuller = -3.1678, Lag order = 6, p-value = 0.09395
## alternative hypothesis: stationary
```

Non-stationary. We should allow auto.arima to do difference.

```r
auto.arima(ts2[,2],stationary=FALSE)
```

```
## Series: ts2[, 2]
## ARIMA(3,1,0)
##
## Coefficients:
##          ar1      ar2     ar3
##       0.4099  -0.0310  0.2529
## s.e.  0.0624   0.0678  0.0623
##
## sigma^2 estimated as 0.03326:  log likelihood=68.87
## AIC=-129.73   AICc=-129.56   BIC=-115.83
```

```r
Box.test(auto.arima(ts2[,2])$residuals,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  auto.arima(ts2[, 2])$residuals
## X-squared = 0.024162, df = 1, p-value = 0.8765
```

The final model is ARIMA(3,1,0)

c

```r
ts3<- read.csv("D:/Dropbox/16summer/Macroeconometrics/Problem Set2/43.csv")
adf.test(ts3[,2])
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts3[, 2]
## Dickey-Fuller = -1.2572, Lag order = 4, p-value = 0.8846
## alternative hypothesis: stationary
```

Non-stationary. We should allow auto.arima to do difference.

```r
auto.arima(ts3[,2],stationary=FALSE)
```

```
## Series: ts3[, 2]
## ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -0.9323
## s.e.   0.0490
##
## sigma^2 estimated as 0.2821:  log likelihood=-79.92
## AIC=163.83   AICc=163.95   BIC=169.06
```

```
Box.test(auto.arima(ts3[,2])$residuals,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  auto.arima(ts3[, 2])$residuals
## X-squared = 0.13323, df = 1, p-value = 0.7151
```

The final model is ARIMA(0,2,1)

Problem 5

```
library(forecast)
ts<- read.csv("D:/Dropbox/16summer/Macroeconometrics/Problem Set2/ps2prob5_data.csv")
```

First let's find out what lag length is the best based on AIC.

```
aic <- rep(NA,10)
for(i in 1:10){
  aic[i] <- arima(ts[,2],order = c(i,0,0),include.mean = 0)$aic
}
aic
```

```
##  [1] 60.99179 62.57473 54.54597 56.44198 56.29581 58.13532 59.07763
##  [8] 60.41373 62.41164 64.39334
```

Clearly AR(3) is the best based on AIC.

```
arima(ts[,2],c(3,0,0),include.mean = 0)
```

```
##
## Call:
## arima(x = ts[, 2], order = c(3, 0, 0), include.mean = 0)
##
## Coefficients:
##          ar1     ar2      ar3
##       0.2308  0.0072  -0.2607
## s.e.  0.0802  0.0832   0.0808
##
## sigma^2 estimated as 0.08056:  log likelihood = -23.27,  aic = 54.55
```

```r
Box.test(arima(ts[,2],c(3,0,0),include.mean = 0)$residuals,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima(ts[, 2], c(3, 0, 0), include.mean = 0)$residuals
## X-squared = 0.019525, df = 1, p-value = 0.8889
```

AR(3) is valid, but it is actually not the "true" model. In order to get the "true" model AR(8), we have to build our own information criterion.

One way is to correct the parameter of penalty term of lag length in the AIC. We know that when there is no penalty(parameter=0), a higher lag length will always get a higher maximized log-likelihood. When in the AIC the parameter of penalty term is 2, we get the best lag length of 3. So there will definitely be some numbers between 0 and 2 that will make the 8th lag length the best model. So we use numerical computation method to compute the interval of all feasible parameters.

```r
loglik <- rep(NA,10)
for(i in 1:10){
  loglik[i] <- arima(ts[,2],order = c(i,0,0),include.mean = 0)$loglik
}
MyIC <- rep(NA,10)
step <- seq(0,2,1e-5)
parameter <- rep(NA,length(step))
l <- c(1:10)
for (j in 1:length(step)) {
    MyIC<- -2*loglik/145+step[j]*l/145
  if(MyIC[8]==min(MyIC)){
    parameter [j] <- step[j]
  }
}
min(subset(parameter,!is.na(parameter)))
```

```
## [1] 0.0102
```

```r
max(subset(parameter,!is.na(parameter)))
```

```
## [1] 0.62735
```

So here we get our own information criterion as following: IC=-2×loglikelihood/T+parameter×laglength/T where parameter∈[0.0102,0.62735]

```r
arima(ts[,2],c(8,0,0),include.mean = 0)
```

```
##
## Call:
## arima(x = ts[, 2], order = c(8, 0, 0), include.mean = 0)
##
## Coefficients:
##          ar1     ar2      ar3      ar4     ar5     ar6      ar7      ar8
##       0.2196  0.0554  -0.2504  -0.0851  0.1077  0.0583  -0.0751  -0.0702
## s.e.  0.0832  0.0856   0.0855   0.0885  0.0879  0.0869   0.0868   0.0860
##
## sigma^2 estimated as 0.07818:  log likelihood = -21.21,  aic = 60.41
```

```r
Box.test(arima(ts[,2],c(8,0,0),include.mean = 0)$residuals,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima(ts[, 2], c(8, 0, 0), include.mean = 0)$residuals
## X-squared = 0.0031579, df = 1, p-value = 0.9552
```

The “true” model AR(8) is valid.