Econometrics III
Assignment Part II

Thao Le (523716)
David Gyaraki (582340)

Contents

1 Question 1 **2**

2 Question 2 **9**

3 Question 3 **10**

4 Question 4 **10**

5 Question 5 **10**

6 Question 6 **10**

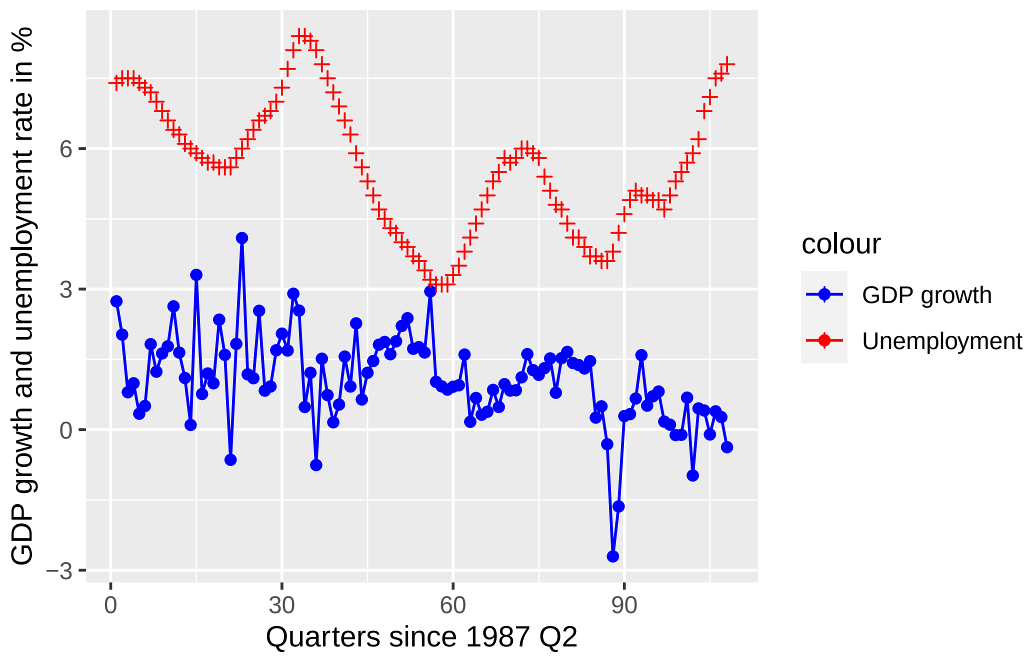
6.1 Question 6a 10

6.2 Question 6b 10

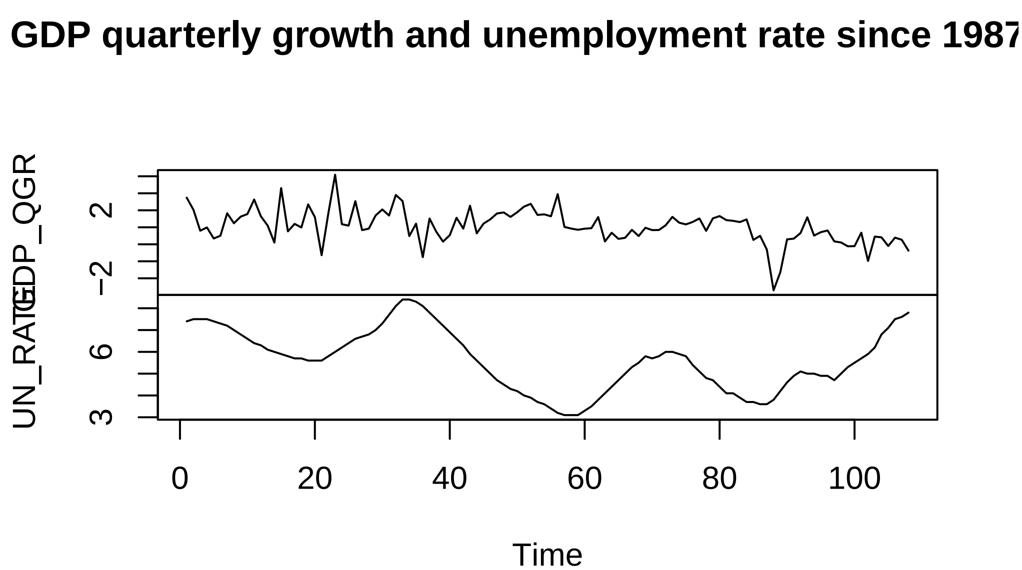
```r
# load packages
if(!require(pacman)){install.packages("pacman")}

p_load(devtools,tidyverse,dplyr,ggplot2,latex2exp,stargazer, fixest,
↪  modelsummary, knitr, readr, tseries, lmtest, forecast, dynlm)

dfAssign_p2 <- as.data.frame(read_csv("data/data_assign_p2.csv"))

# Encode quarters
dfAssign_p2 <-
↪  cbind(dfAssign_p2,c(seq(1,nrow(dfAssign_p2),length.out=nrow(dfAssign_p2))))
colnames(dfAssign_p2) <- c("obs", "GDP_QGR", "UN_RATE", "ind")
```

1 Question 1

```r
ggplot(dfAssign_p2) +
  geom_point(aes(x = ind, y = GDP_QGR, color="GDP growth")) +
  geom_line(aes(x = ind, y = GDP_QGR, color="GDP growth")) +
  geom_point(data=dfAssign_p2, aes(x=ind, y=UN_RATE,
  ↪  color="Unemployment"), shape=3) +
  ylab("GDP growth and unemployment rate in %") +
  xlab("Quarters since 1987 Q2") +
  scale_color_manual(values = c(
    "Unemployment" = "red",
    "GDP growth" = "blue"))
```




```r
plot.ts(dfAssign_p2[,2:3], plot.type = c("multiple"), main="GDP
↪  quarterly growth and unemployment rate since 1987 Q2")
```



The Dutch quarterly growth rates seems to be fluctuating around 1 $\%$ every quarter until it experiences a sharp drop to almost -3$\%$ at Quarter 1 of 2009. This growth rate reflect the influence of the economic recession. Afterwards, it rises again and fluctuate around 0$\%$, indicating that the economy have recovered, but not back to the full swing before the recession.

Regarding the unemployment rate, it resembles a wave pattern. As of now, it is unclear how to relate this pattern to the GDP growth rate or the eco nomic recession. Thus, we proceed to model it in the following steps.

```r
# Creating the autoregression of GDP growth
ar4_gdp <- arima(dfAssign_p2$GDP_QGR, order=c(4,0,0))
coeftest(ar4_gdp) #Remove the least significant lag: the second lag
```

```
z test of coefficients:

          Estimate Std. Error z value  Pr(>|z|)
ar1       0.332074   0.097217  3.4158  0.000636 ***
ar2       0.056504   0.100260  0.5636  0.573047
ar3       0.207863   0.100008  2.0785  0.037667 *
ar4       0.045652   0.097498  0.4682  0.639614
intercept 1.064601   0.220502  4.8281 1.379e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ar1_3_4_gdp <- arima(dfAssign_p2$GDP_QGR, order=c(4,0,0), fixed =
  ↪  c(NA,0,NA,NA,NA))
coeftest(ar1_3_4_gdp) #Remove the least significant lag: the forth lag
```

```
z test of coefficients:

          Estimate Std. Error z value  Pr(>|z|)
ar1       0.349605   0.092313  3.7872 0.0001524 ***
ar3       0.224838   0.095511  2.3541 0.0185694 *
ar4       0.049150   0.097448  0.5044 0.6139976
intercept 1.063916   0.210443  5.0556  4.29e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ar1_3_gdp <- arima(dfAssign_p2$GDP_QGR, order=c(3,0,0), fixed =
 ↪  c(NA,0,NA,NA))
coeftest(ar1_3_gdp) #Stop here, 1st and 3rd lag coefficients are
 ↪  significant
```

```
z test of coefficients:

          Estimate Std. Error z value  Pr(>|z|)
ar1       0.361536   0.089338  4.0468 5.192e-05 ***
ar3       0.241553   0.089670  2.6938  0.007064 **
intercept 1.065869   0.200668  5.3116 1.087e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The final AR model is one with the first and third lag, we call it the AR(1-3) model.

$$X_t = 1.066 + 0.362 * X_{t-1} + 0.242 * X_{t-3} + \epsilon_t \quad (1)$$

Looking at the estimated coefficients, we can say that the quarterly GDP growth rate is influenced by the growth of the previous quarter and the third previous quarters. (MAYBE SOME MORE HYPOTHESIS WHY?).

```r
dfAssign_p2$GDP_QGR <- ts(dfAssign_p2$GDP_QGR, start = c(1987, 2),
 ↪  frequency = 4)
dfAssign_p2$UN_RATE <- ts(dfAssign_p2$UN_RATE, start = c(1987, 2),
 ↪  frequency = 4)

adl_4_4 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 2) + L(UN_RATE, 3)
 ↪  + L(UN_RATE, 4) + GDP_QGR + L(GDP_QGR, 1) + L(GDP_QGR, 2) +
 ↪  L(GDP_QGR, 3) + L(GDP_QGR, 4), data = dfAssign_p2)
coeftest(adl_4_4)
```

```
t test of coefficients:

                Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.1664281  0.0573297  2.9030  0.004605 **
L(UN_RATE, 1)  1.6044535  0.1030613 15.5680 < 2.2e-16 ***
L(UN_RATE, 2) -0.3044370  0.1908063 -1.5955  0.113951
L(UN_RATE, 3) -0.4354095  0.1942187 -2.2419  0.027323 *
```

```
L(UN_RATE, 4)  0.1111869  0.1055289  1.0536  0.294760
GDP_QGR       -0.0138627  0.0147644 -0.9389  0.350174
L(GDP_QGR, 1) -0.0201461  0.0151944 -1.3259  0.188088
L(GDP_QGR, 2) -0.0084318  0.0150829 -0.5590  0.577470
L(GDP_QGR, 3)  0.0205418  0.0153721  1.3363  0.184678
L(GDP_QGR, 4) -0.0069244  0.0146330 -0.4732  0.637162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
adl_4_3 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 2) + L(UN_RATE, 3)
 ↪  + L(UN_RATE, 4) + GDP_QGR + L(GDP_QGR, 1) + L(GDP_QGR, 2) +
 ↪  L(GDP_QGR, 3), data = dfAssign_p2)
coeftest(adl_4_3)
```

```
t test of coefficients:

                Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.1644335  0.0569405  2.8878  0.004804 **
L(UN_RATE, 1)  1.5984568  0.1018606 15.6926 < 2.2e-16 ***
L(UN_RATE, 2) -0.2929561  0.1884828 -1.5543  0.123442
L(UN_RATE, 3) -0.4361600  0.1934173 -2.2550  0.026428 *
L(UN_RATE, 4)  0.1061712  0.1045656  1.0154  0.312515
GDP_QGR       -0.0139329  0.0147032 -0.9476  0.345733
L(GDP_QGR, 1) -0.0215160  0.0148550 -1.4484  0.150797
L(GDP_QGR, 2) -0.0087843  0.0150028 -0.5855  0.559592
L(GDP_QGR, 3)  0.0185162  0.0147037  1.2593  0.211012
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Remove the 2nd exogenous lag.
adl_4_3 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 2) + L(UN_RATE, 3)
 ↪  + L(UN_RATE, 4) + GDP_QGR + L(GDP_QGR, 1) + L(GDP_QGR, 3), data =
 ↪  dfAssign_p2)
coeftest(adl_4_3)
```

```
t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.160028   0.056248  2.8451  0.005428 **
```

```
L(UN_RATE, 1)  1.607262   0.100399 16.0087 < 2.2e-16 ***
L(UN_RATE, 2) -0.301943   0.187213 -1.6128  0.110063
L(UN_RATE, 3) -0.439869   0.192651 -2.2832  0.024621 *
L(UN_RATE, 4)  0.110183   0.103983  1.0596  0.291973
GDP_QGR       -0.014597   0.014609 -0.9992  0.320223
L(GDP_QGR, 1) -0.023896   0.014239 -1.6782  0.096564 .
L(GDP_QGR, 3)  0.016282   0.014151  1.1506  0.252767
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Remove 2nd exogenous lag, and the exogenous variable with no lag.
adl_4_3 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 2) + L(UN_RATE, 3)
 ↪  + L(UN_RATE, 4) + L(GDP_QGR, 1) + L(GDP_QGR, 3), data = dfAssign_p2)
coeftest(adl_4_3)
```

```
t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.162108   0.056209  2.8840  0.004837 **
L(UN_RATE, 1)  1.599186   0.100072 15.9803 < 2.2e-16 ***
L(UN_RATE, 2) -0.301586   0.187211 -1.6109  0.110441
L(UN_RATE, 3) -0.412020   0.190622 -2.1614  0.033122 *
L(UN_RATE, 4)  0.088545   0.101702  0.8706  0.386104
L(GDP_QGR, 1) -0.028710   0.013399 -2.1427  0.034638 *
L(GDP_QGR, 3)  0.012749   0.013702  0.9304  0.354473
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Looking at the model above, we remove the 4th lag of the unemployment
 ↪  rate (autolag).
adl_3_3 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 2) + L(UN_RATE, 3)
 ↪  + L(GDP_QGR, 1) + L(GDP_QGR, 3), data = dfAssign_p2)
coeftest(adl_3_3)
```

```
t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.176507   0.053005  3.3300  0.001221 **
L(UN_RATE, 1)  1.578350   0.096533 16.3504 < 2.2e-16 ***
```

```
L(UN_RATE, 2) -0.337712   0.181214 -1.8636  0.065340 .
L(UN_RATE, 3) -0.269030   0.096080 -2.8001  0.006144 **
L(GDP_QGR, 1) -0.028172   0.013258 -2.1249  0.036085 *
L(GDP_QGR, 3)  0.012055   0.013346  0.9032  0.368596
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Remove the 3rd exogenous lag.
adl_3_1 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 2) + L(UN_RATE, 3)
 ↪   + L(GDP_QGR, 1), data = dfAssign_p2)
coeftest(adl_3_1)
```

```
t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.184821   0.052152  3.5439 0.0006011 ***
L(UN_RATE, 1)  1.572280   0.096210 16.3422 < 2.2e-16 ***
L(UN_RATE, 2) -0.339772   0.181033 -1.8769 0.0634533 .
L(UN_RATE, 3) -0.260515   0.095528 -2.7271 0.0075469 **
L(GDP_QGR, 1) -0.025568   0.012929 -1.9776 0.0507287 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Remove the second autolag.
adl_3_1 <- dynlm(UN_RATE ~ L(UN_RATE, 1) + L(UN_RATE, 3) + L(GDP_QGR,
 ↪   1), data = dfAssign_p2)
coeftest(adl_3_1)
```

```
t test of coefficients:

               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    0.207594   0.051350   4.0427 0.0001033 ***
L(UN_RATE, 1)  1.399745   0.028735  48.7114 < 2.2e-16 ***
L(UN_RATE, 3) -0.431554   0.029004 -14.8789 < 2.2e-16 ***
L(GDP_QGR, 1) -0.027101   0.013063  -2.0746 0.0405617 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The final ADL model is one with lag 1 and lag 3 of Y_t and lag 1 of the exogenous variable X_t.

$$Y_t = 0.208 + 1.400 * Y_{t-1} - 0.432 * Y_{t-3} - 0.027 * X_{t-1} + \epsilon_t \tag{2}$$

Looking at the coefficients of the final model, we can see that the quarterly unemployment rate is related to the unemployment rates of the previous and the 3rd pervious quarter, moreover, it is related to the GDP growth rate of the previous quarter. (MAYBE SOME MORE HYPOTHESIS WHY?).

2 Question 2

The short-run multiplier is the impact of a unit increase in X_t on Y_t, which is the coefficient β_0 of X_t. Since X_t is not significant in this model, the short run multiplier is $\beta_0 = 0$. Interpretation: with every percent increase in the contemporaneous GDP growth rate, the quarterly unemployment rate of the same period is not affected.

The 2-step-ahead multiplier is the impact of a unit increase in X_t on Y_{t+h}, which is:

$$\phi^2\beta_0 + \phi\beta_1 = 0 + 1.400 * (-0.027) = -0.0378 \tag{3}$$

Interpretation: with every percent increase in the current GDP growth rate, the quarterly unemployment rate of the second following quarter decreases by 0.0378%.

The long-run multiplier is the total impact of a permanent unit increase in X_t on Y_t. Let $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\beta(L) = \beta_0 + \beta_1 L + ... + \beta_q L^q$, then, the long-run multiplier is given by:

$$\begin{aligned} \Delta\bar{Y} &= \frac{\beta(1)}{\phi(1)} = \frac{\beta_1}{1 - \phi_1 - \phi_3} \\ &= \frac{-0.027}{1 - 1.400 + 0.432} = -0.84375 \end{aligned} \tag{4}$$

Interpretation: with every permanent percent increase in the current GDP growth rate, the quarterly un-employment rate decrease by 0.84375% on average.

3 Question 3

Statement: "An increase in the GDP growth rate causes a reduction in the unemployment rate."

Looking at the short-run, 2-step ahead, and the long-run multipliers above, we still cannot confirm this statement. This is because the multipliers above only indicates association, not causation.

The multipliers above measure relationship between the dependent variable and the (lagged values) of the independent variables. It tells us the impact of a change in the independent variable on the dependent variable, holding all other factors constant. However, this change can be caused by 3 main reasons:

1. Causation: the independent variable causes the dependent variable.
2. Reserved causality: the dependent variable causes the independent variable.
3. Spurious association: it can occur when two variables are both influenced by a common cause or when one variable influences a third variable, which in turn influences the second variable.

To establish causation, we need to rely on other methods such as randomized controlled trials or natural experiments, which can help us isolate the effect of the independent variable on the dependent variable while controlling for other factors that may influence the outcome.

Moreover, the ADL model is a time series method that can only suggest a relationship between variables based on observed data. It cannot establish causality on its own. Therefore, we need to exercise caution in interpreting the results of the ADL model and use other methods to establish causation.

4 Question 4

5 Question 5

6 Question 6

6.1 Question 6a

6.2 Question 6b