Econometrics III
Assignment Part 1 & 2
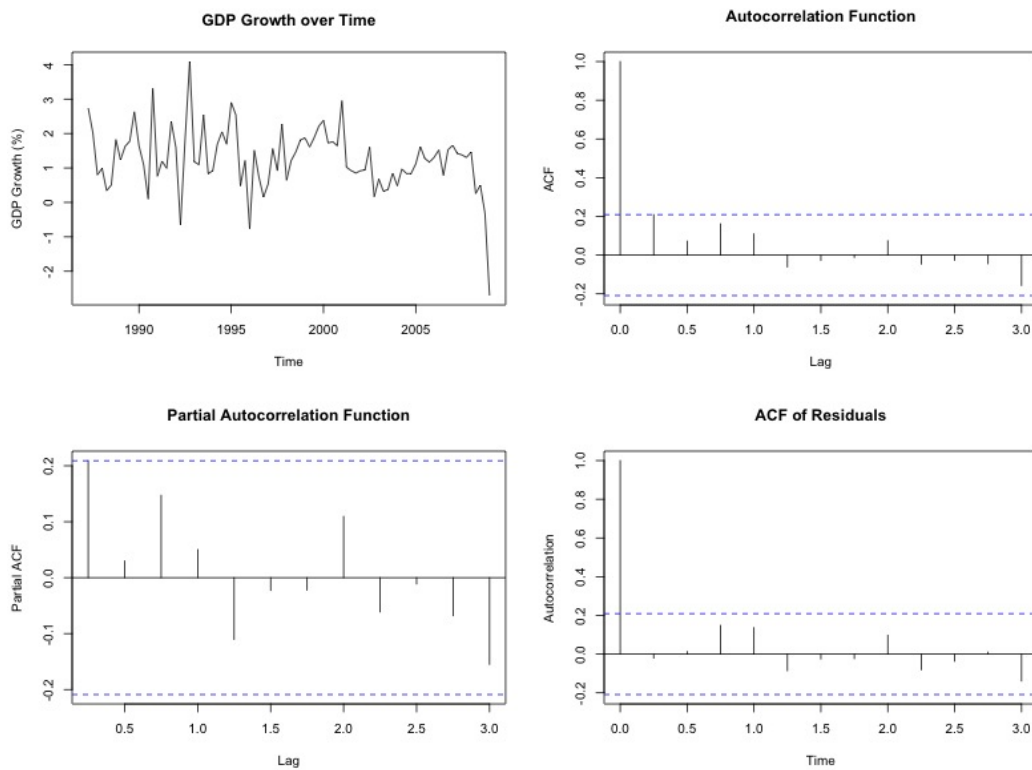# Tinbergen Insitute

Stanislav Avdeev
590050sa
stnavdeev@gmail.com

Bas Machielsen
590049bm
590049bm@eur.nl

March 25, 2021

Question 1

Part 1


```r
# Part 1: Plot the Dutch GDP, ACF, and PACF
df <- readr::read_csv("./data/data_assign_p1.csv")

df <- ts(df$GDP_QGR,
         frequency = 4,
         start = c(1987, 2))
```

In the above figure, upper left, we observe the growth rates of Dutch GDP from 1987 to 2009.

```
autoc <- acf(df,
    lag.max = 12, plot = F)

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

```
##
##  Box-Ljung test
##
## data:  df
## X-squared = 12.106, df = 12, p-value = 0.4372
```

```
pautoc <- pacf(df,
    lag.max = 12, plot = F)
```

In the above figure, we also plot the ACF and PACF, which is the ACF controlled for the other lagged correlations. The ACF tells us that the correlation of GDP growth with its lags is very low - hinting at very little time-dependence in this time-series. More precisely, the estimated correlation coefficients are not higher than 0.2 (for the lag of 1 period). If the GDP is indeed generated by an $AR(p)$ process, the estimates show that the process has low ϕ's (in absolute value), indicating a low time dependence.

Part 2

```
ar4 <- dynlm(df ~ L(df, 1) + L(df, 2) + L(df, 3) + L(df, 4))
ar3 <- dynlm(df ~ L(df, 1) + L(df, 3) + L(df, 4))
ar2 <- dynlm(df ~ L(df, 1) + L(df, 3))
ar1 <- dynlm(df ~ L(df, 1))

ar1_m <- forecast::Arima(df, c(1,0,0), method = "CSS")

stargazer(ar4, ar3, ar2, ar1,
        type = "latex", header = FALSE,
        font.size = "small", column.sep.width = "0pt",
        omit.stat = c("adj.rsq", "ser", "f"))
```

The table below shows us the coefficients of the $AR(p)$ models with 1 to 4 lags included. Using a general-to-specific approach, we eliminate insignificant lags ($\alpha = 0.05$ level) at each step. Column (1) shows that only the first lag is significant at the 0.1 level. The second lag has the lowest ψ coefficient and the highest s.e., thus, we eliminate it first. Doing this iteratively, we are left with the first lag in the model which is significant at the 0.05 level. Investigating the coefficient, we observe there is a 1-period persistence of the time series on its lagged value. The *degree* of persistence, however, is not large, as the magnitude of the coefficient is fairly small (0.26).

Part 3

In the fourth plot above, we show the autocorrelation of the residuals. We observe that none of the coefficients attain significance. Hence, we think that the model is well-specified. Formally, we compute the Durbin-Watson test statistic, of which the output is shown below. The null hypothesis of no autocorrelation is not rejected. In an unreported plot, we also investigate the *partial* autocorrelation of residuals, in which there is also no sign of autocorrelation.

```
durbinWatsonTest(ar1)
```

Table 1:

	Dependent variable:			
	df			
	(1)	(2)	(3)	(4)
L(df, 1)	0.232*	0.240*	0.257**	0.267**
	(0.124)	(0.122)	(0.119)	(0.117)
L(df, 2)	0.055			
	(0.126)			
L(df, 3)	0.203	0.210*	0.210*	
	(0.126)	(0.124)	(0.120)	
L(df, 4)	0.094	0.092		
	(0.125)	(0.124)		
Constant	0.479	0.533*	0.631***	0.896***
	(0.299)	(0.271)	(0.238)	(0.181)
Observations	84	84	85	87
R^2	0.099	0.097	0.089	0.058
Note:			*p<0.1; **p<0.05; ***p<0.01	

```
##  lag Autocorrelation D-W Statistic p-value
##    1      -0.02081405        1.85759    0.476
##  Alternative hypothesis: rho != 0
```

```
autocred <- acf(ar1_m$resid, 12,
                plot = FALSE)
```

Part 4

```
# Part 4: Forecast AR model for 2 years
df_pred <- predict(ar1_m, n.ahead = 8)$pred
```

We derived the forecasts by making use of the conditional expectation as the forecast minimizing the forecast error under quadratic loss. The conditional expectation of the forecast for $T + 1$ is: $\mathbb{E}[X_{T+1}|X_1, \ldots, X_T] = \mathbb{E}[\phi X_T + \alpha + \epsilon_{T+1}|X_1, \ldots, X_T] = \phi X_T + \alpha$. Similarly, the forecast for $T + 2$ equals $\phi \mathbb{E}[X_{T+1} + \alpha] = \phi(\alpha + \phi X_T) + \alpha$. Generalizing this pattern, we then end up with the following recursive forecasts (for $j > 1$):

$$\mathbb{E}[X_{T+j}] = \alpha + \alpha \cdot \left(\sum_{i=1}^{j-1} \phi^i \right) + \phi^j X_T$$

Part 5

We construct the forecasts using a 95% symmetric confidence interval around the forecast.

```
# Part 5: Produce CI
phi <- ar1_m$coef[1]
sigma <- sqrt(ar1_m$sigma2)

df_ciu <- predict(ar1_m, n.ahead = 8)$pred + predict(ar1_m, n.ahead = 8)$se*1.96
df_cil <- predict(ar1_m, n.ahead = 8)$pred - predict(ar1_m, n.ahead = 8)$se*1.96
```

We can reproduce the forecasts manually by using a standard confidence interval around the aforementioned predicted value, $CI(\mathbb{E}[X_{T+j}]) = \alpha + \alpha \cdot \left(\sum_{i=1}^{j-1} \phi^i\right) + \phi^j X_T \pm 1.96 \cdot \sigma$, where $\sigma_j^2 = \sigma_\epsilon^2 \cdot (1 + \phi^2 + \cdots + \phi^{2(j-1)})$.

Part 6

We check for normality of the residuals using the Jarque-Bera test:

```
jb.norm.test(ar1_m$resid)
```

```
##
##  Jarque-Bera test for normality
##
## data:  ar1_m$resid
## JB = 31.722, p-value = 0.0015
```

According to the test, we have to reject H0, indicating that normality of the innovations is violated. Hence, empirically, the assumption of normal residuals does not seem to be appropriate. The error variance might therefore not be estimated consistently, which (possibly) negatively affects the certitude of our forecasts. On the other hand, the zero autocorrelation in the residuals shows that our model is well-specified and there are no lagged GDP-terms in the error-term.

Part 7

In the figure below, we compare the forecasts and actual values by plotting the forecasts (green, CIs in red) next to the actual data (black).

```r
true <- ts(c(0.49, -0.31, -2.7, -1.63, 0.28, 0.33, 0.66, 1.59, 0.51, 0.71, 0.81),
        frequency = 4,
        start = c(2009, 2))

data <- data.frame(date = as.Date(time(true)),
        actual_values = as.matrix(true),
        prediction = c(rep(NA,3), as.matrix(df_pred)),
        upper = c(rep(NA,3), as.matrix(df_ciu)),
        lower = c(rep(NA,3),as.matrix(df_cil)))

data %>%
  ggplot(aes(x = date)) + geom_line(aes(y = actual_values)) +
  geom_line(aes(y = upper), color = "red", lty = "dashed") +
  geom_line(aes(y = lower), color = "red", lty = "dashed") +
  geom_line(aes(y = prediction), color = "green", lty = "dashed") +
  ylab("Values") + xlab("Date") +
  ggtitle("Actual and Predicted GDP series") +
  theme_minimal()
```

**Actual and Predicted GDP series**
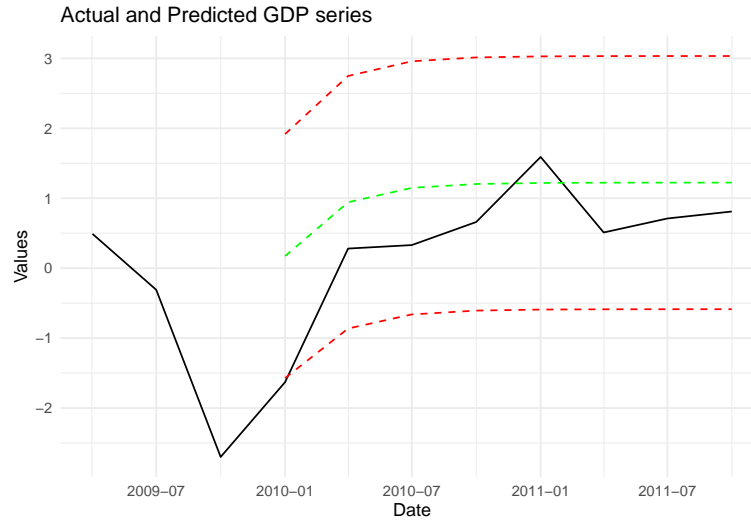
The forecasts show a tendency to revert to the mean of the AR process, which happens quite quickly after the drop. The drop has thus only a small effect on the forecasts. On the other hand, the limited T makes that the confidence bounds are quite large: on the one hand, all realized values fall within the 95% confidence bounds, which is a sign of accuracy, but on the other hand, the standard errors of the estimates are quite large, making the estimates less informative and perhaps less useful to policy makers.

Question 2

Part 1

First, we plot the unemployment and GDP series.

```
plot(df2)
```


df2
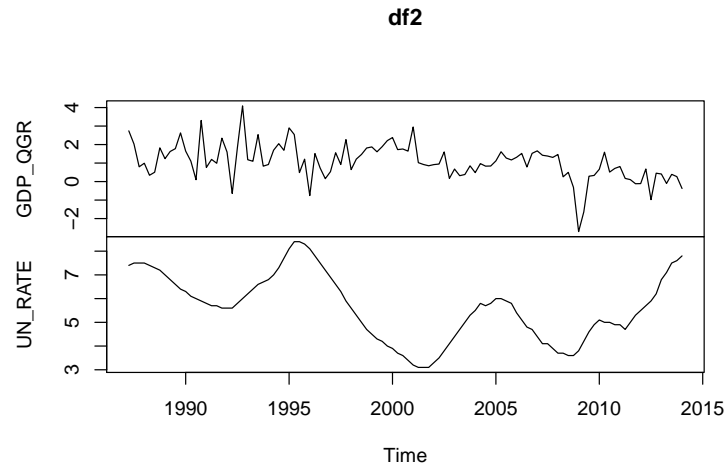
Secondly, we implement the general-to-specific approach for the ADL in R:

```
recursive_reg <- function(dataset, equation){

  initial_reg <- dynlm(formula = as.formula(equation), data = dataset)

  p_values <- initial_reg %>%
    summary() %>%
    .$coefficients %>%
    .[,4]

  while(max(p_values) > 0.05){

    variable_drop <- which.max(p_values) %>%
      labels()

    equation <- equation %>%
      str_c("-", fixed(variable_drop))

    # reestimate the model and print the results
    initial_reg <- dynlm(formula = as.formula(equation), data = dataset)

    p_values <- initial_reg %>%
    summary() %>%
    .$coefficients %>%
    .[,4]

    print(summary(initial_reg))
  }
```

```
   initial_reg

}
```

Thirdly, we estimate the AR-models:

```
ar <- recursive_reg(dataset = df2, equation = "GDP_QGR ~ L(GDP_QGR, 1) +
           L(GDP_QGR, 2) + L(GDP_QGR, 3) + L(GDP_QGR, 4)")
```

Fourthly, we estimate the ADL-models:

```
adl <- recursive_reg(dataset = df2, equation = 'UN_RATE ~ L(UN_RATE, 1) +
           L(UN_RATE, 2) + L(UN_RATE, 3) + L(UN_RATE, 4) +
           L(GDP_QGR, 0) + L(GDP_QGR, 1) + L(GDP_QGR, 2) +
           L(GDP_QGR, 3) + L(GDP_QGR, 4)')
```

```
stargazer(ar, adl, type = "latex", header = FALSE,
          font.size = "small", column.sep.width = "0pt",
          omit.stat = c("adj.rsq", "ser", "f"),
          column.labels = c("AR for GDP", "ADL for Unemp"))
```

Table 2:

	Dependent variable:	
	equation	
	AR for GDP	ADL for Unemp
	(1)	(2)
L(UN_RATE, 1)		1.400***
		(0.029)
L(UN_RATE, 3)		−0.432***
		(0.029)
L(GDP_QGR, 1)	0.353***	−0.027**
	(0.092)	(0.013)
L(GDP_QGR, 3)	0.251***	
	(0.092)	
Constant	0.400***	0.208***
	(0.144)	(0.052)
Observations	104	104
R^2	0.235	0.992
Note:	*p<0.1; **p<0.05; ***p<0.01	

We observe that the GDP AR-process is an AR(3)-process, with the 2-period lag component being insignificant. This means that there is a one-period effect, but also some kind of seasonality. If the economy 2 periods ago was doing better, today's economy tends to do better as well (as evidenced by the positive sign of the coefficient). In general, it is expected that lags in an AR-process for GDP should be positive, indicating a kind of path dependency. The ADL-process for Unemployment is an ADL(3,1) process: focusing on the autoregressive component, the coefficient for the one-period lag is high in magnitude and positive, indicating a persistence effect in unemployment. The 3-period lag, however, is negative, indicating a kind of recovery:

if unemployment decreased two-periods ago, it tends to recover today. Focusing on the distributed-lag components, we note that there is no contemporaneous effect of GDP on unemployment, indicating that a shock in GDP growth is only visible in the unemployment-series in the next period. The coefficient also has the expected sign: an increase in GDP growth is associated with a decrease in unemployment the next period.

Part 2

We estimate short- and long-run multipliers, and the 2-step ahead multiplier. Define the ADL-process as:

$$Unemp_t = \alpha + \phi_1 \cdot Unemp_{t-1} + \phi_3 \cdot Unemp_{t-3} + \theta_1 \cdot GDP_{t-1} + \epsilon_t$$

We then implement the multipliers:

```
osm <- coef(adl)[4]
tsm <- coef(adl)[2] * coef(adl)[4]

lrm <- coef(adl)[4]/(1 - coef(adl)[2] - coef(adl)[3])
```

First, since there is no contemporaneous X_t, the short-run multiplier is equal to zero. Secondly, the one-step ahead multiplier can be easily obtained by observing that Y_t is still at the same value, and the shock enters the process only at $t + 1$: hence, the 1-step multiplier is equal to $\theta_1 = $ -0.0270649. The two-step multiplier has to take into account that now Y_{t+1}^* is different from Y_{t+1}, but there is no influence from the AR-process at X_{t+1}, so that the two-step multiplier is $ _1 _1 $ = -0.0378819.

The long-run multiplier is derived by setting $Y_t = \bar{Y}$ and $X_t = \bar{X}$ for all t. Solving this and looking at the coefficient of $\Delta X = 0$ leads to the long-run multiplier: -0.848268. The long-run relation is given by:

$$\bar{Y} = \frac{\alpha}{1 - \phi_1 - \phi_3} + \frac{\theta_1}{1 - \phi_1 - \phi_3}\bar{X}$$

Substituting in the coefficients from our model gives:

```
constant <- coef(adl)[1]/(1 - coef(adl)[2] - coef(adl)[3])
```

$\bar{Y} = 6.5191658 + $ -0.848268 $\cdot \bar{X}$.

Part 3

Assuming strict exogeneity, we can interpret the coefficient estimates as being causal, in which case, taking $\delta X = 1$, we see that the long-run multiplier $\bar{Y} = \frac{\theta_1}{1 - \phi_1 - \phi_2} < 0$, which means that indeed an increase in the GDP growth rate causes a decrease in the unemployment rate of about 4 percentage points. However, it also seems plausible there is an simultaneous relationship between GDP and Unemployment, or perhaps both of them are shaped by more fundamental parameters. In this case, the coefficients do not warrant a causal explanation, and there is only a negative *correlation* between GDP growth and unemployment.

Part 4

We know that the distribution of $Y_t|Y_{t-1} \sim N(\alpha + \phi_1 Y_{t-1} + \phi_3 Y_{t-3} + \gamma X_{t-1}, \sigma^2)$. Shifting everything by one period, we obtain that the distribution of Y_{t+1} conditional on the realized values of Y_t is:

$$Y_{t+1}|Y_t \sim N(\alpha + \phi_1 Y_t + \phi_3 Y_{t-2} + \gamma_1 X_t, \sigma^2)$$

Denote the mean of this distribution by μ. Then, we are looking for:

$$\mathbb{P}(Y_{t+1} > 0.078|Y_t) = \mathbb{P}\left(\frac{Y_{t+1} - \mu}{\sigma} > \frac{0.078 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{0.078 - \mu}{\sigma}\right)$$

Substituting μ and σ gives:

```
mu <- adl$coefficients[1] + adl$coefficients[2]*7.8 +
  adl$coefficients[3]*7.5 + adl$coefficients[4]*-0.37

sigma <- sqrt(var(adl$residuals))
```

So that the probability of unemployment being over 7.8% is: 0.7904221, and the probability of it being below 7.8 is 0.2095779. On the one hand, the IID gaussian innovations-assumption is very *ad hoc*, and made mainly because of convenience, but on the other hand, a Gaussian often does a good job when errors come from many (uncorrelated) sources, making it a more plausible description of reality. Crises however, show that 'tail events' occur with too high a probability to be described by even a Gaussian with a high variance (which would go at the cost of overall likelihood), so we conclude that the assumption is useful, but not particularly realistic.

Part 5

First, we produce an 8-quarter forecast for the AR-process underlying GDP growth:

```
#input dataframe with one column
forecast_gdp <- function(no_periods, model, df){

  coefficients <- model$coefficients

  for(i in 1:no_periods){
    prediction <- coefficients[1] +
    coefficients[2] * df[length(df)] +
    coefficients[3] * df[length(df) - 2]

    df <- c(df, unname(prediction))

  }

  df

}

plot(ts(start = c(1987, 2),
        frequency = 4,
        forecast_gdp(8, ar, df2[,1])), ylab = "Realized and Predicted GDP Growth")
```
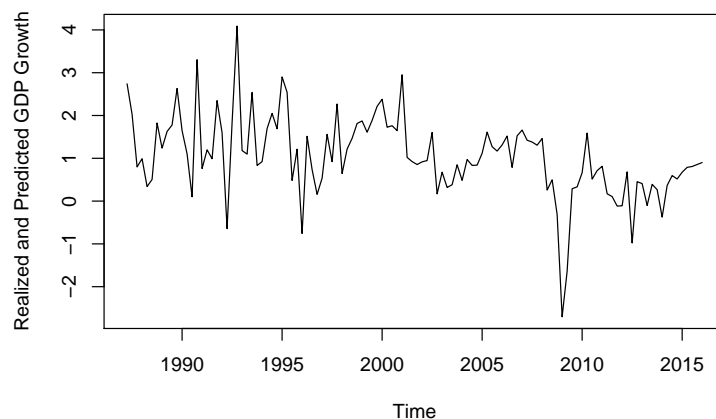
```
# Predicted Values:
ts(start = c(1987, 2),
        frequency = 4,
        forecast_gdp(8, ar, df2[,1])
   ) %>%
  .[(length(.)-8):length(.)]
```

```
## [1] -0.3744039  0.3660043  0.5967171  0.5170298  0.6747751  0.7884422  0.8085984  0.8553275  0.9003790
```

Part 6

First, we record the impulse response function of GDP growth:

```
irf_gdp <- function(model, lags_in_model, e, no_periods, reference_value){

 coefficients <- model$coefficients

 # initial data frame
 irfs <- rep(0, lags_in_model)
 response <- e

 irfs <-  c(irfs, response)

 # all other predictions
  for(i in 2:no_periods){

     response <- coefficients[2] * irfs[length(irfs)] +
     coefficients[3] * irfs[length(irfs) - 2]

     irfs <- c(irfs, unname(response))

  }

 irfs <- irfs + reference_value

 data.frame(time = seq(from = -lags_in_model, to = no_periods -1, by = 1),
            irfs = irfs)

}
```

Now, we calculate the impulse response function of unemployment, using internally the impulse response function of GDP growth:

```
irf_unemp <- function(model, lags_in_model, e, no_periods, reference_value){

  #initialization
  adl_coefficients <- model$coefficients
  ar_coefficients <- ar$coefficients

  #IRF for gdp
  gdp_rates <- irf_gdp(ar, 3, e, no_periods, -0.37)

  unemp_rates <- rep(0, lags_in_model)

  for(i in 1:no_periods){

     response <- adl_coefficients[2] * unemp_rates[length(unemp_rates)] +
```

10

```
      adl_coefficients[3] * unemp_rates[length(unemp_rates)-2] +
      adl_coefficients[4] * gdp_rates$irfs[i+2]

    unemp_rates <- c(unemp_rates, unname(response))

  }

  unemp_rates <- unemp_rates + reference_value

  data.frame(time = seq(from = -lags_in_model, to = no_periods -1, by = 1),
             gdp = gdp_rates$irfs,
             unemp = unemp_rates)

}
```

We observe a familiar pattern: following an exogenous shock in GDP, we see that GDP recovers quickly, owing to the fact that the coefficients from the AR-model are quite low in magnitude, and hence, the process is very stable and shocks do not last long.
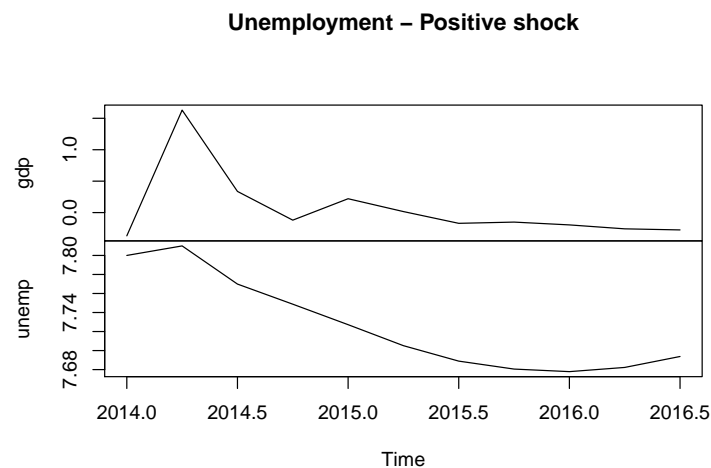
On the contrary, in the impulse responses of the ADL-model, we observe that the coefficient on the past unemployment is very high, indicating a high degree of persistence. This means practically that a shock in unemployment has long-lasting effects, and it takes a long time before unemployment reverts to its stationary value. Note that the effects of the GDP shock make themselves felt only a period after the shock, because the contemporaneous distributed-lag term is not present in the model. The main channel through which the shock permeates is, however, through its own lagged value, rather than through contemporaneous our lagged values of GDP. All of these observations are also in line with commonly accepted macroeconomic theory.

Finally, we note that the impulse responses are somewhat distorted by the fact that the impulses are started out of equilibrium value. Since usually, the equilibrium values are normalized to 0, in this case, we start with different values, causing the IRF's to be a combination of the perturbation and the tendency towards equilibrium.

```
irf_unemp(adl, 3, 2, 10, 7.8) %>%
  .[-c(1:2),2:3] %>%
  ts(frequency = 4,
     start = c(2014, 1)) %>%
  plot(main = "Unemployment - Positive shock")
```

Unemployment – Positive shock

```
irf_unemp(adl, 3, -2, 10, 7.8) %>%
  .[-c(1:2),2:3] %>%
  ts(frequency = 4,
     start = c(2014, 1)) %>%
  plot(main = "Unemployment - Negative shock")
```

Unemployment – Negative shock