

# Econometrics III - Assignment 1

## The Explosive Invertibles

25 Mar 2022

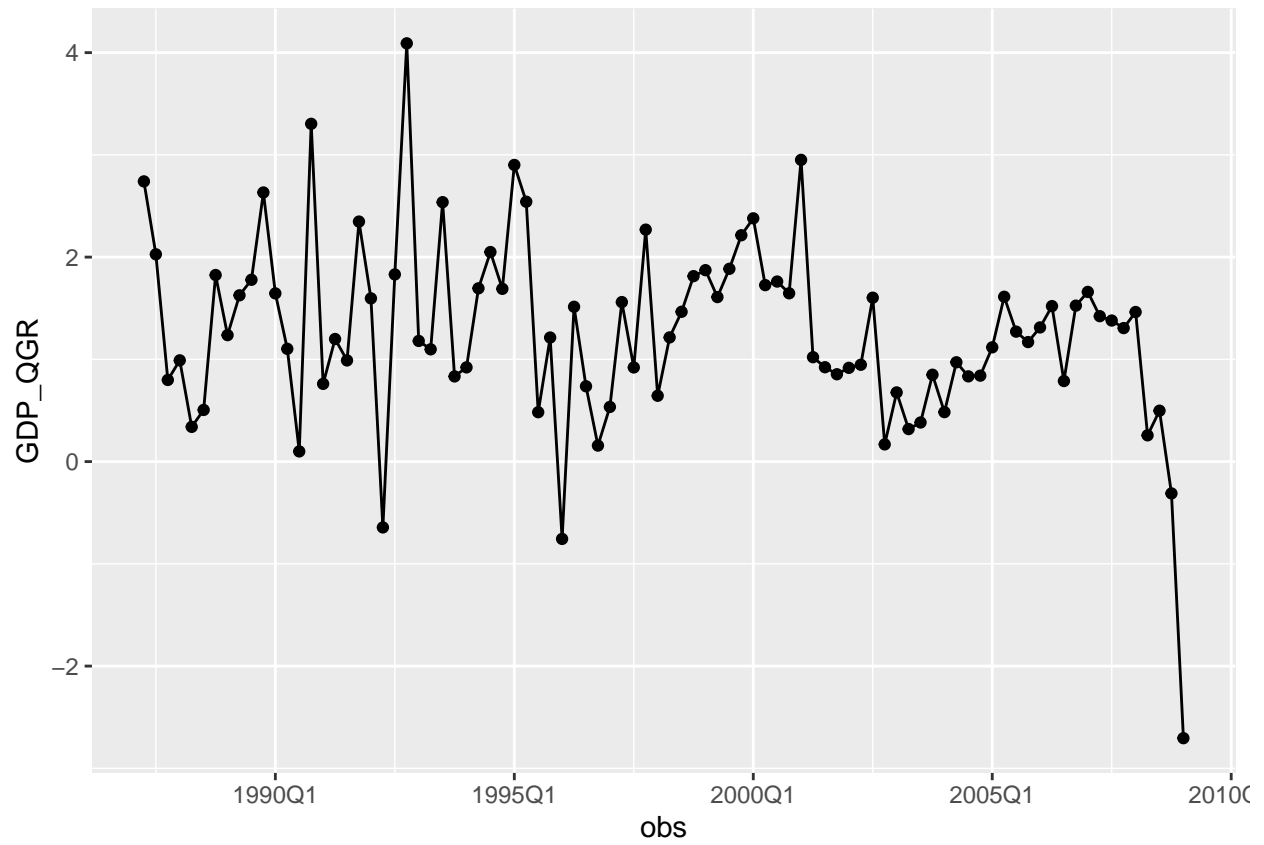
### Question 1

```
# Load data
data <- read.csv("data_assign_p1.csv")
```

1.) Plot the sample of Dutch GDP quarterly growth rates that you have at your disposal. Report the 12-period sample ACF and PACF functions and comment on their shape. What does the sample ACF tell you about the dynamic properties of GDP quarterly growth rates?

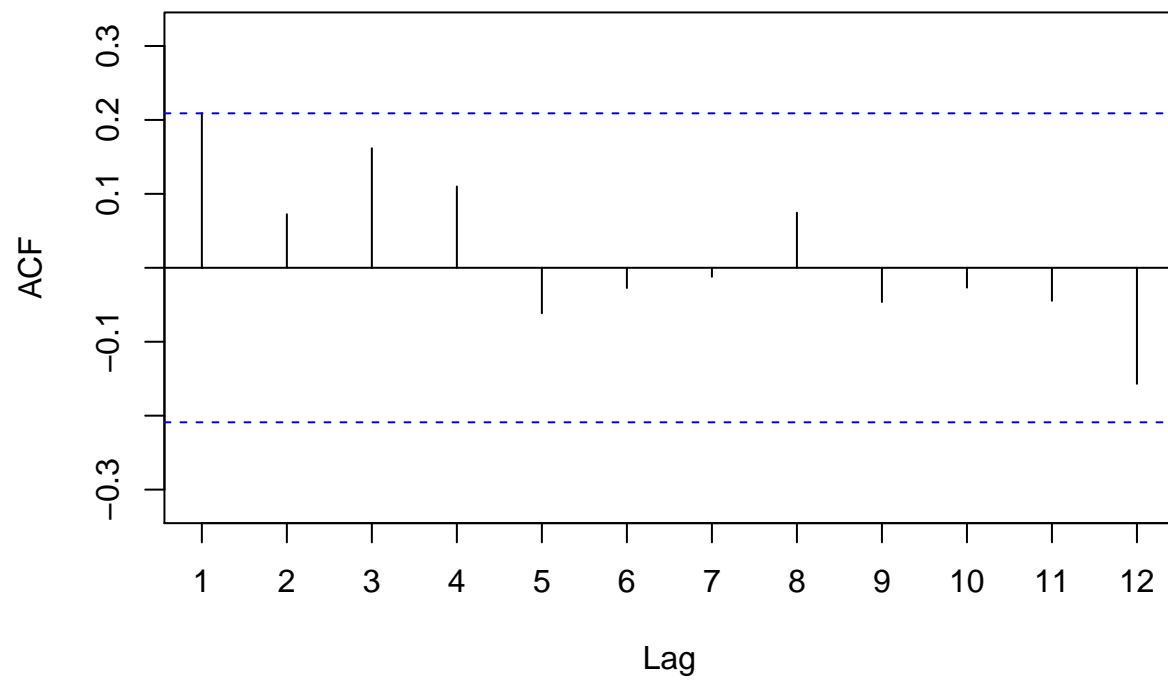
```
fmt <- "%YQ%q"
data$obs <- as.yearqtr(data$obs, format = fmt)
```

```
ggplot(data, aes(obs, GDP_QGR)) +
  geom_point() +
  geom_line() +
  scale_x_yearqtr(format = fmt)
```

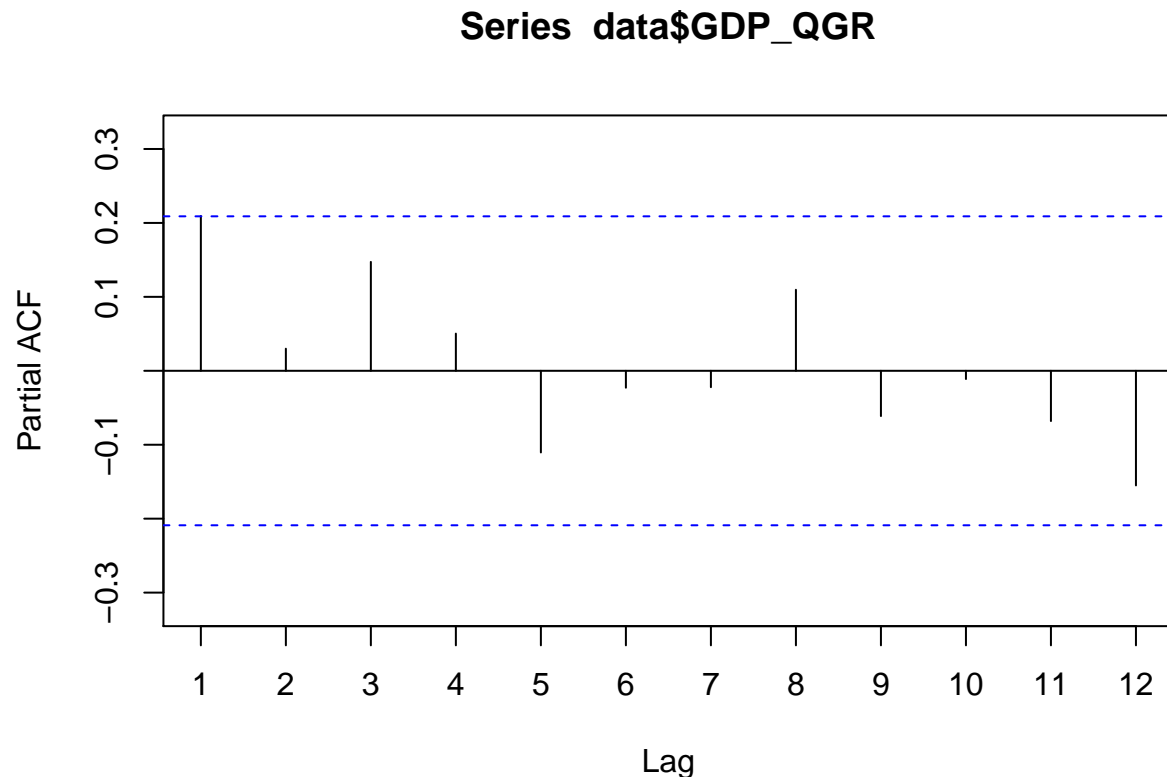


```
Acf(  
  data$GDP_QGR,  
  lag.max = 12  
)
```

### Series data\$GDP\_QGR



```
Pacf(  
  data$GDP_QGR,  
  lag.max = 12  
)
```



The ACF is the autocorrelation function and measures the autocorrelation between  $X_t$  and  $X_{t-h} \forall h$ , in our case for  $h = \{1, 2, \dots, 12\}$ . The autocorrelation is an indicator for memory: the higher the correlation, the higher the time dependence. Hence, the ACF also is a good first indicator for the selection of the lags of our time series model (corresponding to the Box-Jenkins approach). The partial autocorrelation function (PACF) goes one step further, controlling for other lags. Looking at our ACF and PACF graphs, we can see that the first lag seem to have a significant effect on our GDP while higher lags have a rather small autocorrelation with  $X_t$ .

2.) Estimate an AR(p) model for the same time-series. Please use the general-to-specific modeling approach by starting with a total  $p = 4$  lags and removing insignificant lags sequentially. Report the final estimated AR(p) model, working at a 5% significance level. Comment on the estimated coefficients. What do these coefficients tell you about the dynamic properties of the GDP quarterly growth rate?

```
ar4 <- arima(data$GDP_QGR, order = c(4,0,0))
coeftest(ar4)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1      0.241960  0.117749  2.0549  0.03989 *
## ar2      0.029955  0.119538  0.2506  0.80213
## ar3      0.189023  0.119463  1.5823  0.11359
## ar4      0.086413  0.120151  0.7192  0.47201
## intercept 1.214421  0.203636  5.9637 2.466e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ar3 <- arima(data$GDP_QGR, order = c(3,0,0))
coeftest(ar3)
```

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value Pr(>|z|)
## ar1         0.255536   0.116737  2.1890  0.02860 *
## ar2         0.029868   0.120377  0.2481  0.80404
## ar3         0.200007   0.119049  1.6800  0.09295 .
## intercept 1.228330   0.179893  6.8281 8.605e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ar2 <- arima(data$GDP_QGR, order = c(2,0,0))
coeftest(ar2)
```

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value Pr(>|z|)
## ar1         0.263192   0.118876  2.2140  0.02683 *
## ar2         0.057902   0.121458  0.4767  0.63356
## intercept 1.249173   0.139580  8.9495 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ar1 <- arima(data$GDP_QGR, order = c(1,0,0))
coeftest(ar1)
```

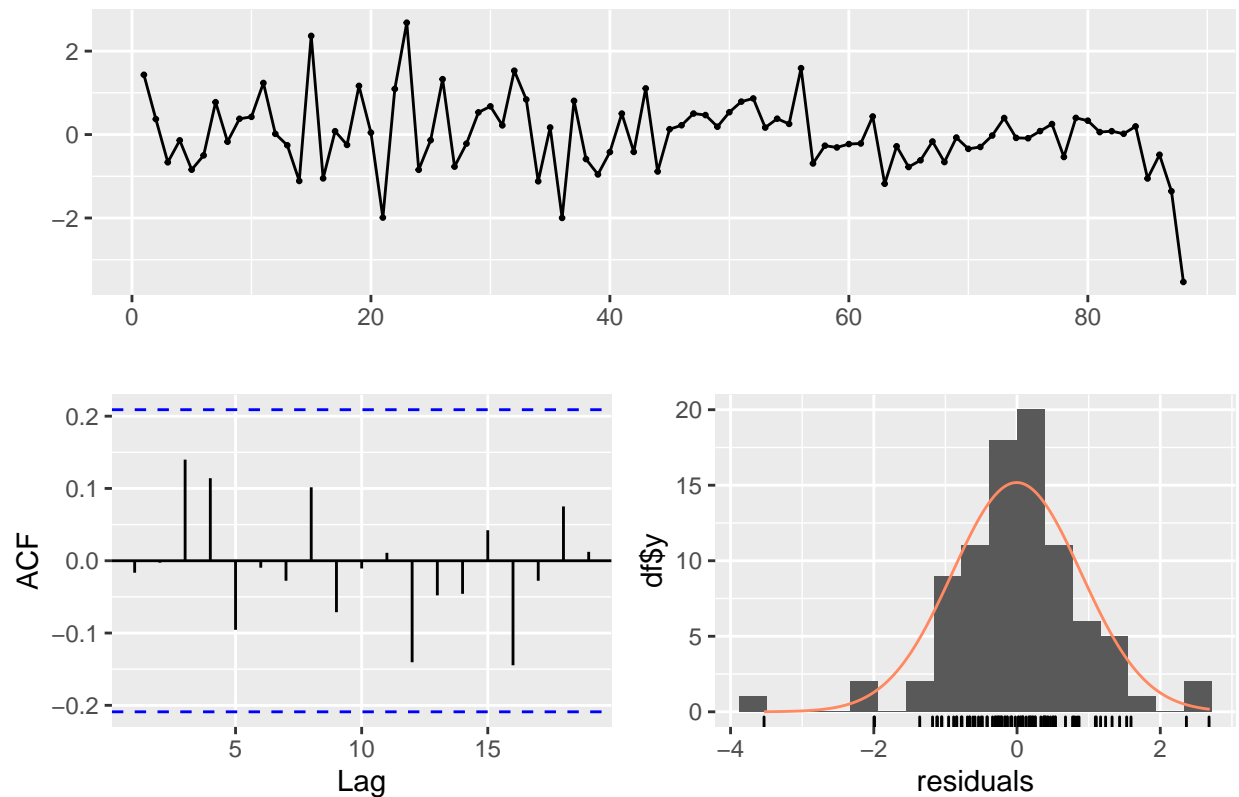
```
##
## z test of coefficients:
##
##           Estimate Std. Error z value Pr(>|z|)
## ar1         0.27191    0.11775  2.3092  0.02093 *
## intercept 1.25269    0.13036  9.6091 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As only the estimator of the first lag has a significant effect (which can be observed in all model specification), our final model specification is an AR(1) model. The estimator is around 0.27191. Hence, an increase in the previous period's GDP growth by 1%, increases the next period's GDP growth by approximately 0.22%.

3.) Check the regression residuals of the estimated AR(p) model for autocorrelation by plotting the estimated residual ACF function. Does the model seem well specified?

```
checkresiduals(ar1)
```

### Residuals from ARIMA(1,0,0) with non-zero mean



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 5.5739, df = 8, p-value = 0.6948
##
## Model df: 2.   Total lags used: 10
```

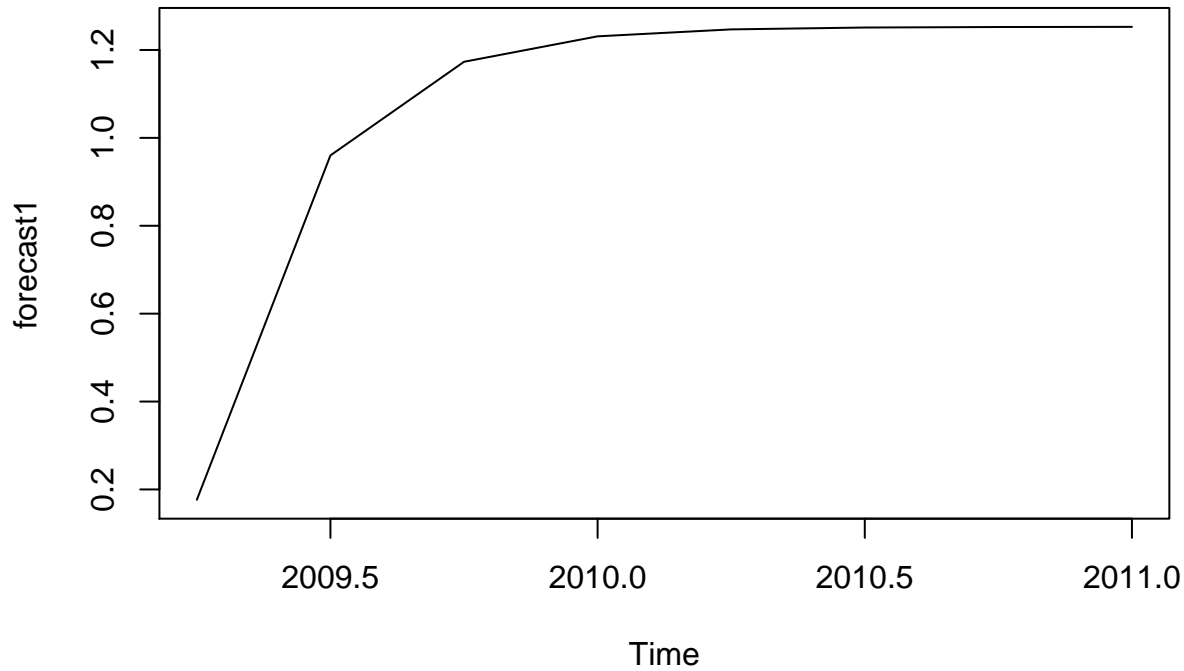
Looking at the plot of the estimated residual ACF, there does not seem to be any autocorrelation between the regression residuals of the estimated AR(1) model. Therefore, the model seems to be well specified.

4.) Make use of your estimated AR model to produce a 2-year (8 quarters) forecast for the Dutch GDP quarterly growth rate that spans until the first quarter of 2011. Report the values you obtained and explain how you derived them.

```
seriesdata <- ts(data$GDP_QGR, start= c(1987, 2), end = c(2009, 1), frequency = 4)
forecast1 <- ts(predict(ar1, n.ahead=8)$pred, start = c(2009,2), frequency = 4)
forecast1
```

```
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2009           0.1765306 0.9600700 1.1731221
## 2010 1.2310530 1.2468050 1.2510881 1.2522527
## 2011 1.2525694
```

```
ts.plot(forecast1, type="l")
```



Our forecast is an eight step ahead forecast:

$$\begin{aligned}\hat{X}_{T+8} &= E(X_{T+8}|X_1, \dots, X_T) \\ &= E(\phi^8 X_T + \phi^7 \epsilon_{T+1} + \dots + \epsilon_{T+8} | X_1, \dots, X_T) = \phi^8 X_T\end{aligned}$$

In order to get an efficient forecast,  $X_t$  (so the GDP growth) needs to follow a stable AR(1) process which means that  $|\phi| < 1$ . This is the case as our estimation gives us a  $\phi \approx 0.27191$ . Further, we need  $\epsilon_t \sim IID(0, \sigma_\epsilon^2)$ .

For our forecast, we used the function *predict()* to forecast the next 8 quarters based on the AR(1) model. It gives us a time series of predictions, using finite-history prediction via the Kalman filter. Note it is only statistically significant if the MA part is stationary (as an AR model does not have a MA part, this is not an issue for us).

The Kalman Filter is basically an iterative process that entails two phases. In the first phase, the filter produces estimations of the current state variables. In the second phase, the estimates are weighted with weighted averages. The higher the certainty of the estimation, the higher the weight. The filter also is often used for DSGE models.

5.) Suppose that the innovations in your AR model are iid Gaussian. Produce 95% confidence intervals for your 2-year forecast. Furthermore, comment on the following statement issued by government officials: “Given the available GDP data, we believe that there is a low probability that the Dutch GDP growth rate will remain negative in the second quarter of 2009.”

Under the assumption that the innovations in the AR(1) model are Gaussian, confidence intervals for the h-step ahead forecast  $\hat{X}_{T+h}$  can be produced based on the estimated variance of the corresponding h-step

ahead forecast error  $\mathbb{V}\hat{\mathcal{D}}\backslash(e_{T+h})$ . For an AR(1)-process it holds that:

$$\mathbb{V}\hat{\mathcal{D}}\backslash(e_{T+h}) = \sigma^2(1 + \phi^2 + \phi^4 + \dots + \phi^{2(h-1)})$$

Since the true variance of the innovations is unknown, it is estimated based on the obtained residuals to the fitted values:

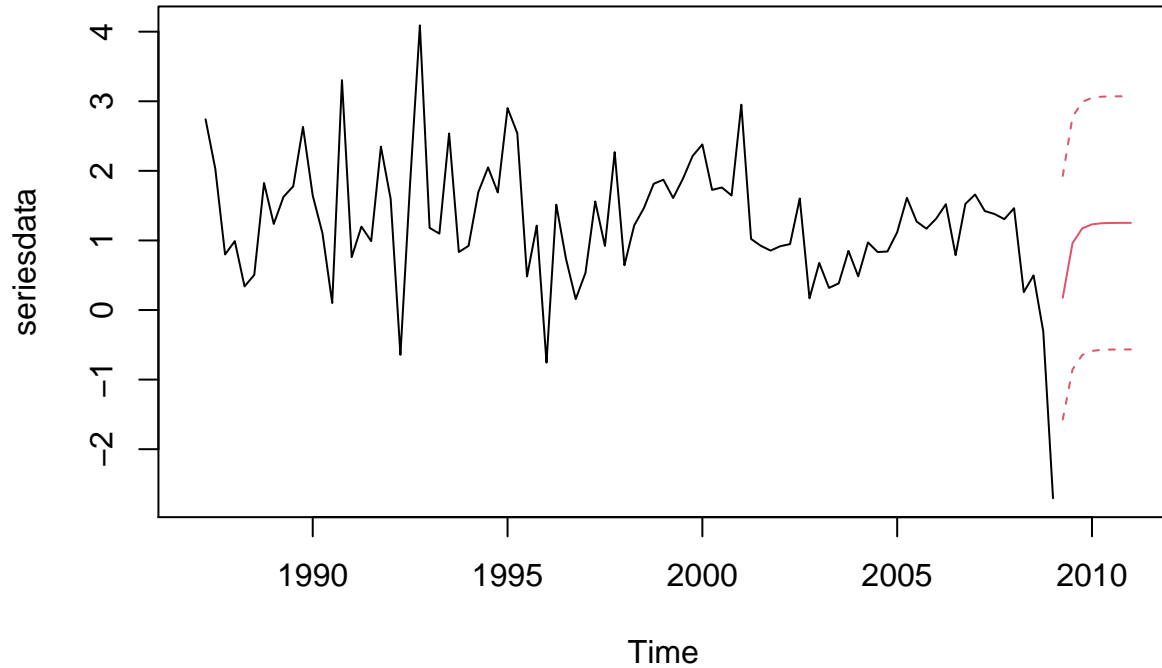
$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=2}^T (x_t - \hat{\alpha} - \hat{\phi}x_{t-1})^2$$

Under the Gaussianity assumption, the 95%-confidence intervals for  $\hat{X}_{T+h}$  then formally read as follows:

$$\hat{X}_{T+h} \pm 1.96 \cdot \sqrt{\mathbb{V}\hat{\mathcal{D}}\backslash(e_{T+h})}$$

The code below produces and plots the confidence intervals together with the h-step ahead forecasts along the observed time series. The estimated standard deviations of the forecast errors which reflect only innovation uncertainty are obtained from the function `predict()` described above.

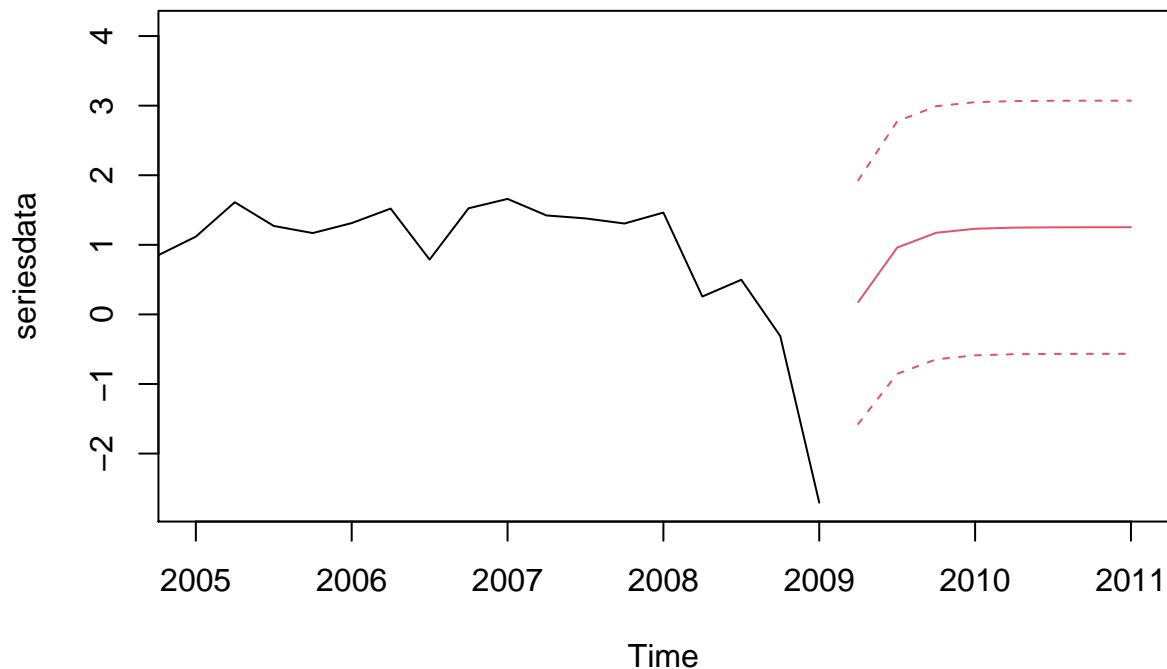
```
ts.plot(seriesdata, xlim = c(1987, 2011))
se1 <- ts(predict(ar1, n.ahead=8)$se, start = c(2009,2), frequency = 4)
points(forecast1, type="l", col=2)
points(forecast1 - 1.96*se1, type="l", col=2, lty=2)
points(forecast1 + 1.96*se1, type="l", col=2, lty=2)
```



```
ts.plot(seriesdata, xlim = c(2005, 2011))
se1 <- ts(predict(ar1, n.ahead=8)$se, start = c(2009,2), frequency = 4)
points(forecast1, type="l", col=2)
```



```
points(forecast1 - 1.96*se1, type="l", col=2, lty=2)
points(forecast1 + 1.96*se1, type="l", col=2, lty=2)
```



To comment on the statement we can also leverage the Gaussianity assumption which implies that under the assumption of an AR(1)-process future realizations of Dutch GDP growth follow the following estimated distribution:

$$X_{T+h} \sim N\left(\hat{X}_{T+h}, \sqrt{\mathbb{V}\hat{\mathcal{D}} \setminus (e_{T+h})}\right) \Rightarrow X_{T+1} \sim N\left(\hat{\alpha} + \hat{\phi}x_T, \hat{\sigma}\right)$$

This allows us to estimate the probability that the realization of Dutch GDP growth in the next period, i.e. the second quarter of 2009, will fall below 0%:

```
pnorm(0, mean=forecast1[1], sd = se1[1])
```

```
## [1] 0.4216612
```

Given an estimated probability of 43.16% of a negative realization of Dutch GDP growth, our findings do not support the statement.

6.) Do you find the assumption of iid Gaussian innovations reasonable? How does this affect your answer to the previous question?

To test whether the assumption of iid Gaussian innovations is reasonable we can use a Jarque-Bera test based on the distribution moments of the residuals obtained above.

```
jarque.bera.test(ar1$residuals)
```

```
##
```

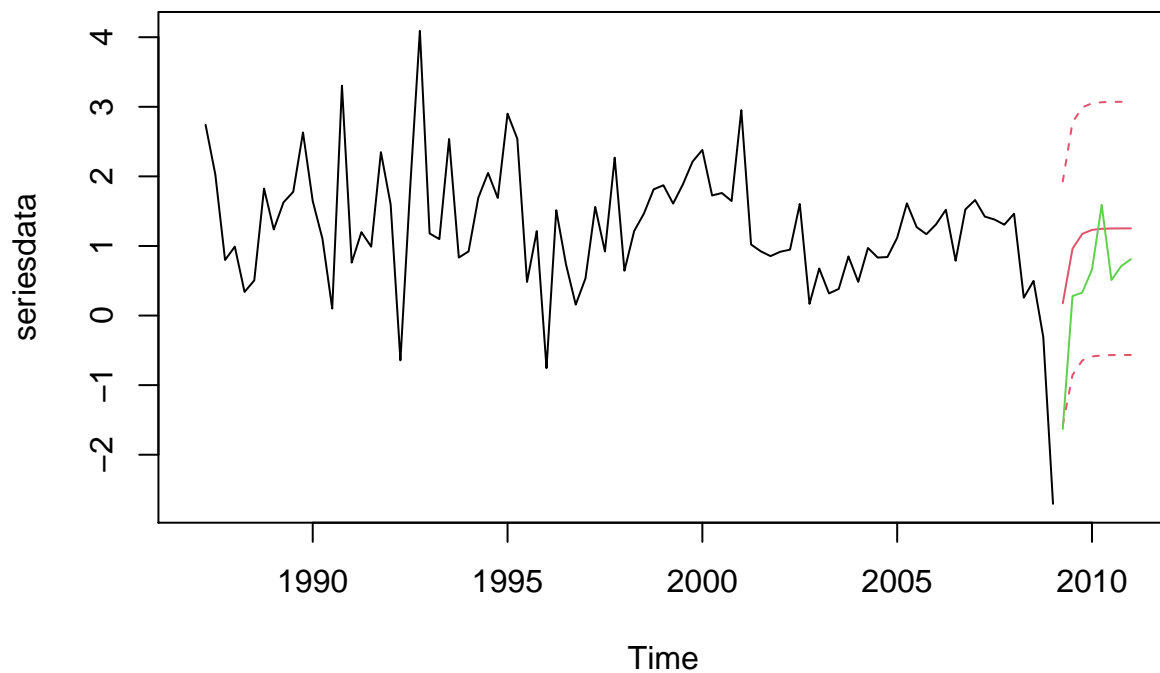
```
## Jarque Bera Test
##
## data: ar1$residuals
## X-squared = 26.298, df = 2, p-value = 1.947e-06
```

Since the test rejects the null hypothesis of normality the assumption of iid Gaussian innovations is under serious jeopardy and weakens the conclusions drawn based upon this normality assumption above.

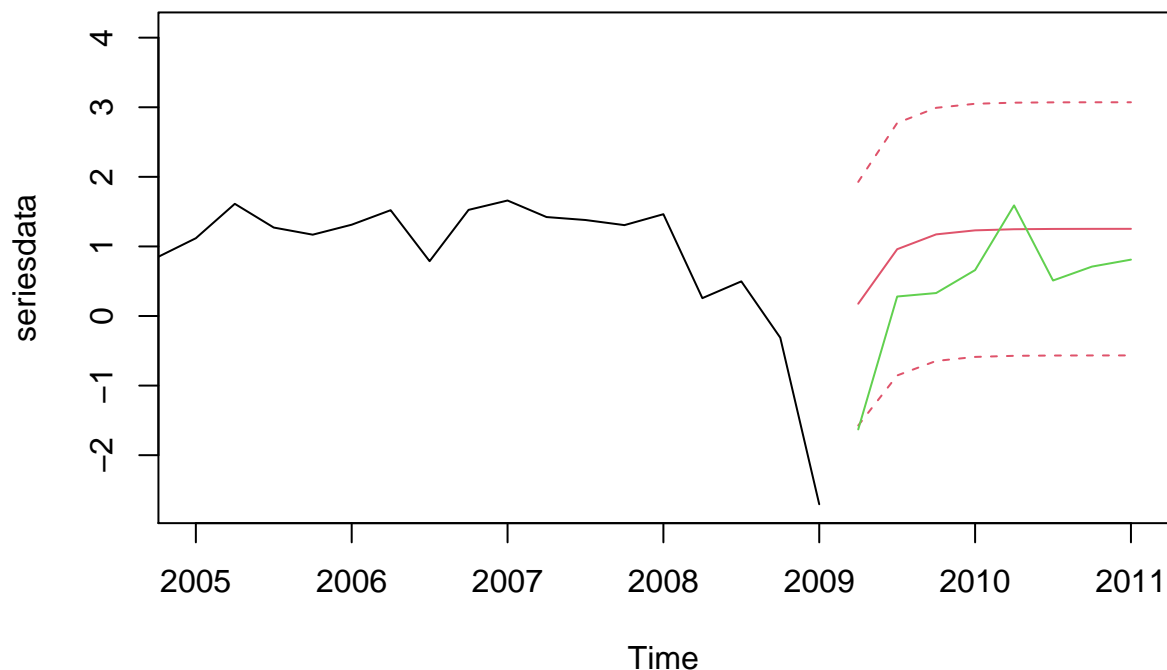
7.) Suppose that 2 years have passed since you delivered your forecasts to the government, in the first quarter of 2009. Compare your point forecasts and confidence bounds with the following actual observed values for the 12 quarters from 2009q2 to 2011q1. Please comment on the accuracy of your forecasts.

The given realizations are plotted alongside the point forecasts and confidence bounds below.

```
ts.plot(seriesdata, xlim = c(1987, 2011))
re <- ts(c(-1.63, 0.28, 0.33, 0.66, 1.59, 0.51, 0.71, 0.81), start = c(2009, 2), frequency = 4)
points(forecast1, type="l", col=2)
points(forecast1 - 1.96*se1, type="l", col=2, lty=2)
points(forecast1 + 1.96*se1, type="l", col=2, lty=2)
points(re, type="l", col=3)
```



```
ts.plot(seriesdata, xlim = c(2005, 2011))
re <- ts(c(-1.63, 0.28, 0.33, 0.66, 1.59, 0.51, 0.71, 0.81), start = c(2009, 2), frequency = 4)
points(forecast1, type="l", col=2)
points(forecast1 - 1.96*se1, type="l", col=2, lty=2)
points(forecast1 + 1.96*se1, type="l", col=2, lty=2)
points(re, type="l", col=3)
```



At a glance, the AR(1)-model estimated above produces forecasts of reasonable accuracy, though without benchmarking such a statement seems fairly unfounded. At least, the true realization of Dutch GDP growth only lays outside the provided confidence intervals for the one-step ahead forecast, i.e. the second quarter of 2009. Generally, the model yields predictions which lay above the true realizations in the validation sample considered here.

8.) Repeat question 2 above, but this time using a 10% significance level for the general to-specific modeling approach. Use the newly estimated model to produce a 2-year (8 quarters) point forecast of the Dutch GDP quarterly growth rate. Comment on the accuracy of the forecast generated by the newly estimated AR model. Is it better than the model you estimated before?

From the discussion in 2. it directly follows that the G2S model selection procedure selects a AR(3)-model if a 10% significance model is adopted. Its estimation is reported again below.

```
ar3

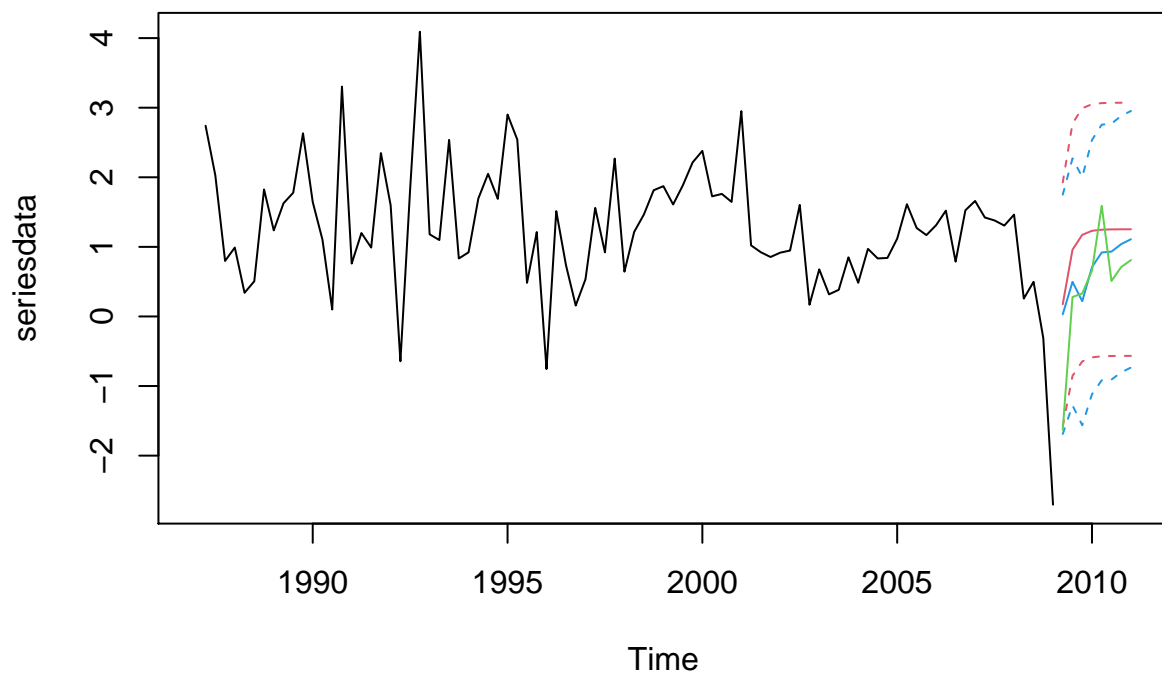
##
## Call:
## arima(x = data$GDP_QGR, order = c(3, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3  intercept
##          0.2555  0.0299  0.200      1.2283
## s.e.      0.1167  0.1204  0.119      0.1799
##
## sigma^2 estimated as 0.7698:  log likelihood = -113.47,  aic = 236.94
```

Point forecasts and confidence intervals of the Dutch GDP quarterly growth rate are obtained in the same way as above and plotted alongside the forecast yielded by the estimated AR(1)-model.

```

ts.plot(seriesdata, xlim = c(1987,2011))
forecast3 <- ts(predict(ar3, n.ahead=8)$pred, start = c(2009,2), frequency = 4)
se3 <- ts(predict(ar3, n.ahead=8)$se, start = c(2009,2), frequency = 4)
points(forecast1, type="l", col=2)
points(forecast1 - 1.96*se1, type="l", col=2, lty=2)
points(forecast1 + 1.96*se1, type="l", col=2, lty=2)
points(forecast3, type="l", col=4)
points(forecast3 - 1.96*se3, type="l", col=4, lty=2)
points(forecast3 + 1.96*se3, type="l", col=4, lty=2)
points(re, type="l", col=3)

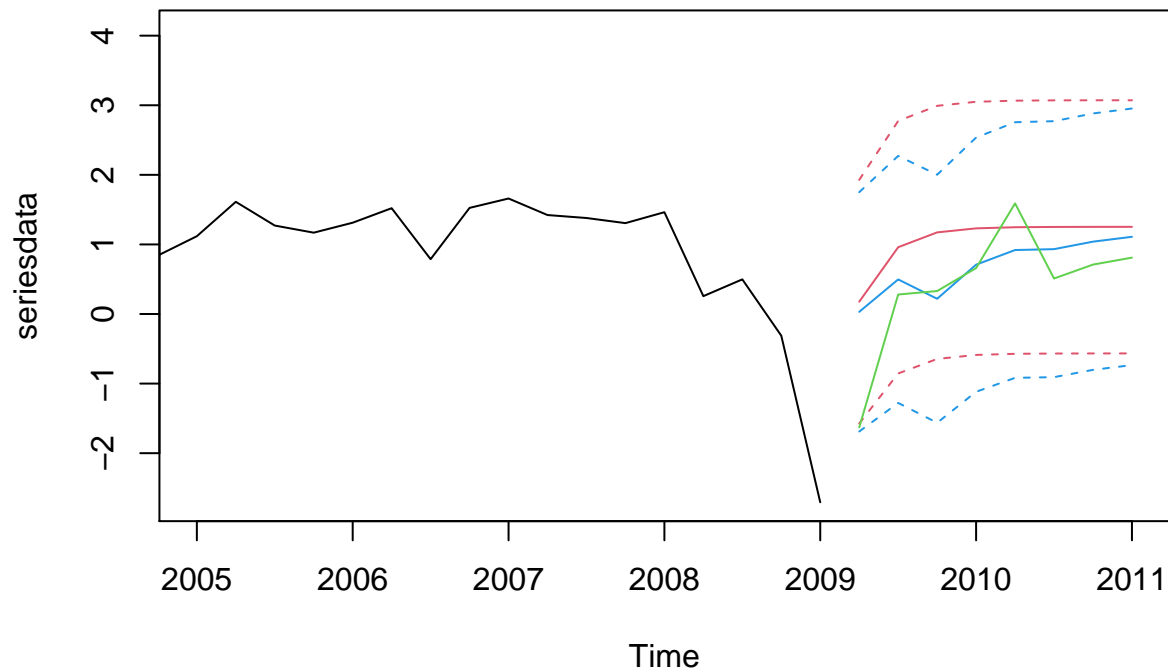
```



```

ts.plot(seriesdata, xlim = c(2005,2011))
forecast3 <- ts(predict(ar3, n.ahead=8)$pred, start = c(2009,2), frequency = 4)
se3 <- ts(predict(ar3, n.ahead=8)$se, start = c(2009,2), frequency = 4)
points(forecast1, type="l", col=2)
points(forecast1 - 1.96*se1, type="l", col=2, lty=2)
points(forecast1 + 1.96*se1, type="l", col=2, lty=2)
points(forecast3, type="l", col=4)
points(forecast3 - 1.96*se3, type="l", col=4, lty=2)
points(forecast3 + 1.96*se3, type="l", col=4, lty=2)
points(re, type="l", col=3)

```



Judging by the naked eye, the AR(3)-model seems to generate more accurate out-of-sample forecasts than the AR(1)-model. This is also confirmed by consulting the forecast mean squared errors for both models:

```
rmse(re, forecast1)
```

```
## [1] 0.8602256
```

```
rmse(re, forecast3)
```

```
## [1] 0.6752132
```

To formally test whether the AR(3)-model produces significantly more accurate forecasts a Diebold-Mariano test for predictive accuracy is calculated (Diebold & Mariano, 1995). Since the test rejects the null hypothesis of equality of prediction mean squared errors at the 5% confidence level, the judgement by naked eye is confirmed.

```
error1 <- re - forecast1
error3 <- re - forecast3
dm.test(error1, error3)
```

```
##
```

```
## Diebold-Mariano Test
```

```
##
```

```
## data: error1error3
```

```
## DM = 2.6014, Forecast horizon = 1, Loss function power = 2, p-value =
```

```
## 0.03535
```

```
## alternative hypothesis: two.sided
```

## Question 2

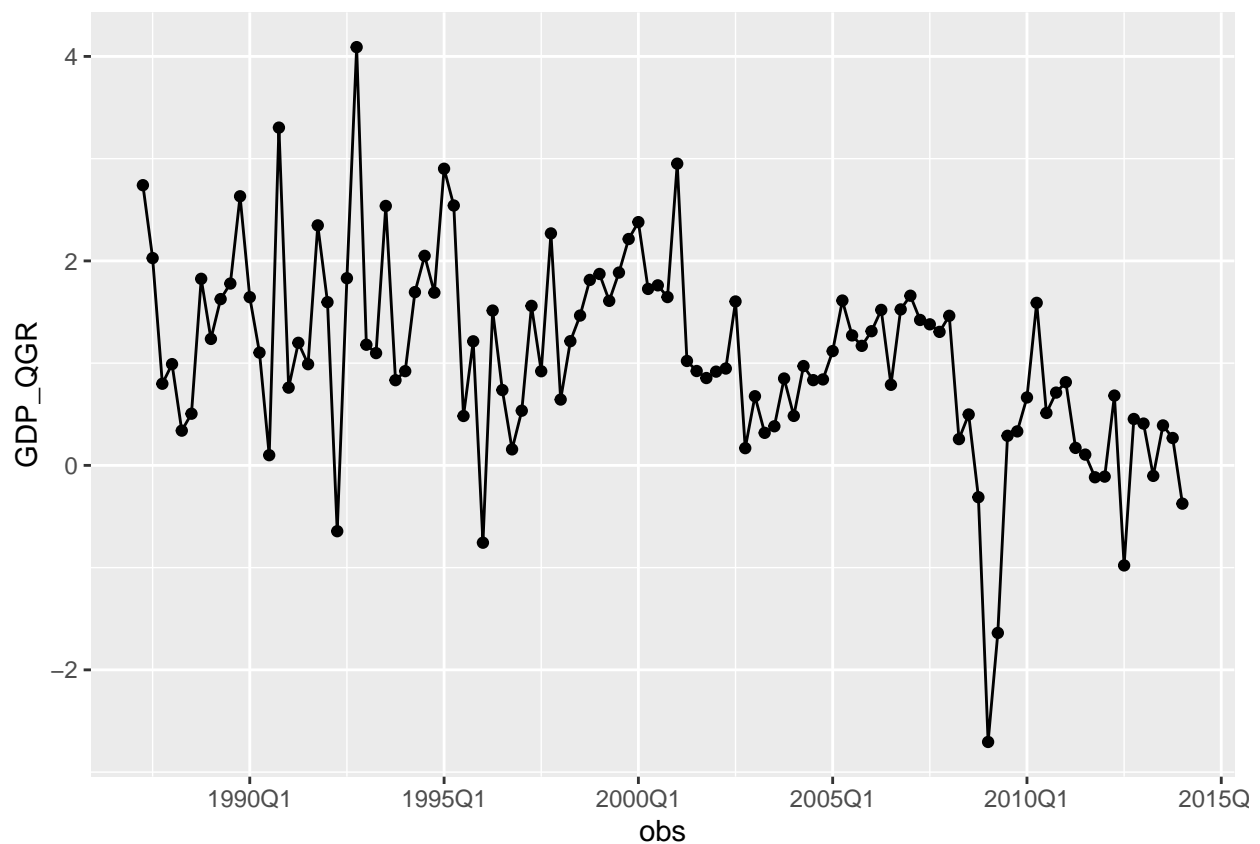
1. Plot the sample of Dutch quarterly unemployment rates and Dutch GDP quarterly growth rates that you have at your disposal. Estimate an AR model for the GDP growth rate and an ADL model for the unemployment rate using the GDP growth rate as an exogenous explanatory variable. Please adopt a general-to-specific methodology for both models by eliminating insignificant lags. For the AR model start with four lags of GDP. For the ADL model start with four lags of each variable. Report the final estimated AR and ADL models, working at a 5% significance level. Comment on the estimated coefficients. What do these coefficients tell you about the dynamic properties of the unemployment rate and the GDP growth rate?

The sample of Dutch quarterly unemployment rates and Dutch GDP quarterly growth rates are plotted below.

```
# Load data
data2 <- read.csv("data_assign_p2.csv")

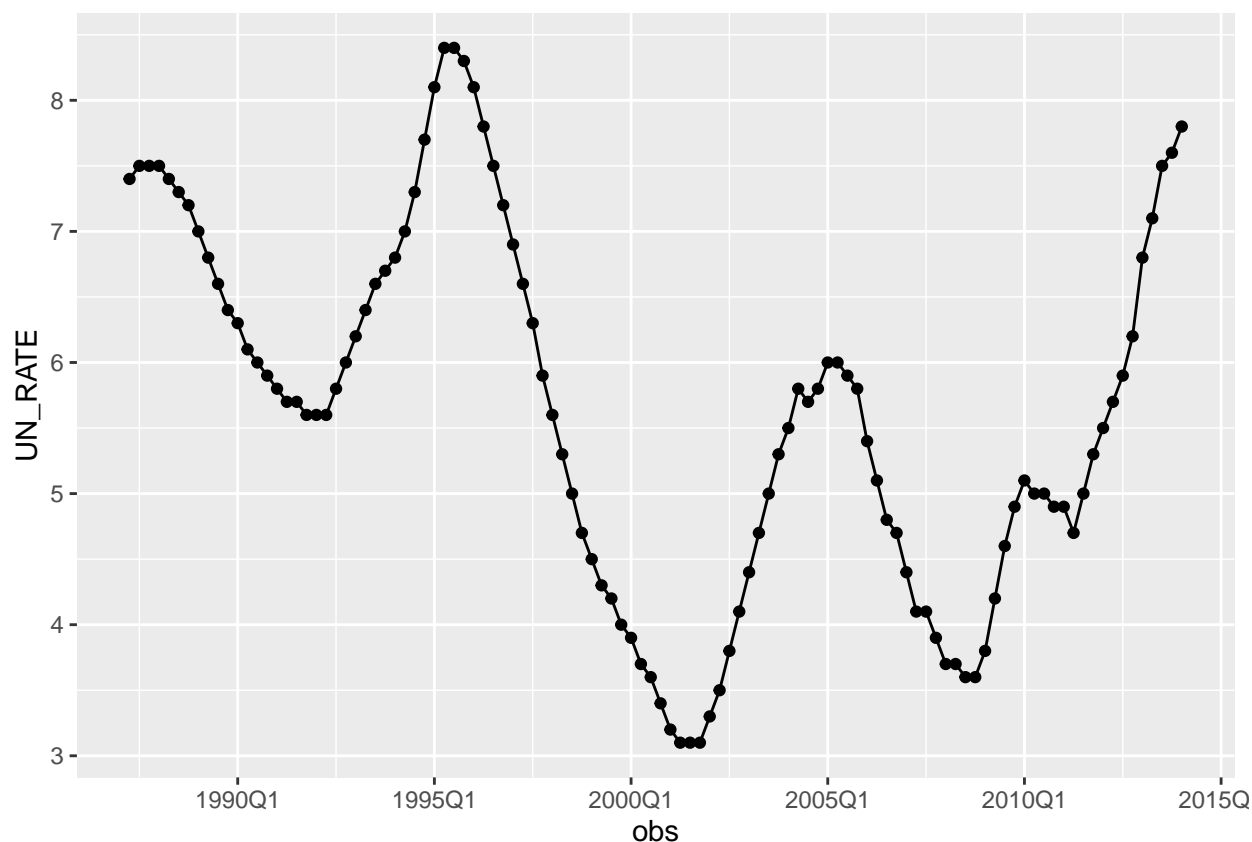
fmt <- "%YQ%q"
data2$obs <- as.yearqtr(data2$obs, format = fmt)

ggplot(data2, aes(obs, GDP_QGR)) +
  geom_point() +
  geom_line() +
  scale_x_yearqtr(format = fmt)
```



```
ggplot(data2, aes(obs, UN_RATE)) +
  geom_point() +
  geom_line() +
```

```
scale_x_yearqtr(format = fmt)
```



To specify the AR model for the GDP growth rate, the lag order is reduced from  $p = 4$  until the highest-order lag-coefficient is significant at the 5% significance level. This yields an AR(3)-model whose estimation output is printed below. Evidently, a more thorough specification methodology should be based on (multiple) information criteria, but is omitted here. Furthermore, we also do not allow for holes in the autoregressive lag structure.

```
data3 = data2
data3$GDP_QGR <- ts(data3$GDP_QGR, start = c(1987,2), frequency = 4)
data3$UN_RATE <- ts(data3$UN_RATE, start = c(1987,2), frequency = 4)

gdp4 <- arima(data3$GDP_QGR, order = c(4,0,0))
coefest(gdp4)
```

```
##
## 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
```

```
gdp3 <- arima(data3$GDP_QGR, order = c(3,0,0))
coeftest(gdp3)
```

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value Pr(>|z|)
## ar1         0.342181   0.094883  3.6064 0.0003105 ***
## ar2         0.059545   0.100243  0.5940 0.5525106
## ar3         0.222430   0.095130  2.3382 0.0193792 *
## intercept  1.066499   0.211070  5.0528 4.353e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimation output suggests that the unconditional expectation of the GDP growth rate is positive since  $\hat{\alpha}$  is significantly greater than zero. Moreover, the estimated autoregressive lag structure suggests that it is driven mainly by autocorrelation of order one and three, since the coefficient of the second-order autoregressive lag is not significantly different from 0.

We apply the same procedure to specify the ADL model for the unemployment rate, working at a 5% significance level.

```
adl4_4 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, 1:4) + L(data3$GDP_QGR, 0:4))
coeftest(adl4_4)
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.1664281   0.0573297  2.9030 0.004605 **
## L(data3$UN_RATE, 1:4)1  1.6044535   0.1030613 15.5680 < 2.2e-16 ***
## L(data3$UN_RATE, 1:4)2 -0.3044370   0.1908063 -1.5955 0.113951
## L(data3$UN_RATE, 1:4)3 -0.4354095   0.1942187 -2.2419 0.027323 *
## L(data3$UN_RATE, 1:4)4  0.1111869   0.1055289  1.0536 0.294760
## L(data3$GDP_QGR, 0:4)0 -0.0138627   0.0147644 -0.9389 0.350174
## L(data3$GDP_QGR, 0:4)1 -0.0201461   0.0151944 -1.3259 0.188088
## L(data3$GDP_QGR, 0:4)2 -0.0084318   0.0150829 -0.5590 0.577470
## L(data3$GDP_QGR, 0:4)3  0.0205418   0.0153721  1.3363 0.184678
## L(data3$GDP_QGR, 0:4)4 -0.0069244   0.0146330 -0.4732 0.637162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
adl3_3 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, 1:3) + L(data3$GDP_QGR, 0:3))
coeftest(adl3_3)
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.1817707   0.0536139  3.3904 0.001011 **
## L(data3$UN_RATE, 1:3)1  1.5719309   0.0979556 16.0474 < 2.2e-16 ***
## L(data3$UN_RATE, 1:3)2 -0.3330489   0.1832415 -1.8175 0.072221 .
## L(data3$UN_RATE, 1:3)3 -0.2665701   0.0969294 -2.7501 0.007107 **
## L(data3$GDP_QGR, 0:3)0 -0.0107626   0.0142773 -0.7538 0.452777
## L(data3$GDP_QGR, 0:3)1 -0.0216592   0.0147379 -1.4696 0.144900
## L(data3$GDP_QGR, 0:3)2 -0.0097847   0.0148620 -0.6584 0.511861
```



```
## L(data3$GDP_QGR, 0:3)3 0.0172492 0.0143647 1.2008 0.232753
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

adl3_2 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, 1:3) + L(data3$GDP_QGR, 0:2))
coeftest(adl3_2)

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.1896711  0.0533285   3.5567 0.0005805 ***
## L(data3$UN_RATE, 1:3)1  1.5681104  0.0981244 15.9808 < 2.2e-16 ***
## L(data3$UN_RATE, 1:3)2 -0.3393878  0.1835779  -1.8487 0.0675102 .
## L(data3$UN_RATE, 1:3)3 -0.2561725  0.0967593  -2.6475 0.0094497 **
## L(data3$GDP_QGR, 0:2)0 -0.0072895  0.0140127  -0.5202 0.6040933
## L(data3$GDP_QGR, 0:2)1 -0.0210596  0.0147626  -1.4265 0.1568886
## L(data3$GDP_QGR, 0:2)2 -0.0047384  0.0142876  -0.3316 0.7408649
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

adl3_1 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, 1:3) + L(data3$GDP_QGR, 0:1))
coeftest(adl3_1)

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.1869681  0.0524646   3.5637 0.0005646 ***
## L(data3$UN_RATE, 1:3)1  1.5730586  0.0965465 16.2933 < 2.2e-16 ***
## L(data3$UN_RATE, 1:3)2 -0.3452486  0.1819022  -1.8980 0.0606103 .
## L(data3$UN_RATE, 1:3)3 -0.2553056  0.0962883  -2.6515 0.0093334 **
## L(data3$GDP_QGR, 0:1)0 -0.0078777  0.0138374  -0.5693 0.5704360
## L(data3$GDP_QGR, 0:1)1 -0.0224935  0.0140518  -1.6008 0.1126166
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

adl3_0 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, 1:3) + L(data3$GDP_QGR, 0:0))
coeftest(adl3_0)

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.175203  0.052351   3.3467 0.001153 **
## L(data3$UN_RATE, 1:3)1  1.592126  0.096554 16.4894 < 2.2e-16 ***
## L(data3$UN_RATE, 1:3)2 -0.368132  0.182751  -2.0144 0.046653 *
## L(data3$UN_RATE, 1:3)3 -0.252047  0.097016  -2.5980 0.010792 *
## L(data3$GDP_QGR, 0:0) -0.016390  0.012874  -1.2731 0.205940
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since under the ADL(3,0) specification, whose respective estimation output is printed above, the distributed lag coefficient is not significantly different from 0, the G2C methodology would lead us to also conclude an AR(3) model for the unemployment rate. To avoid this, an ADL model with a distributed lag of order one is estimated which points us towards disregarding the autoregressive lag of order 2 which the no longer is significant at the 5% significance level.

```
adl2_0 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, 1:3) + L(data3$GDP_QGR, 1:1))
coeftest(adl2_0)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.184821   0.052152   3.5439 0.0006011 ***
## L(data3$UN_RATE, 1:3)1  1.572280   0.096210  16.3422 < 2.2e-16 ***
## L(data3$UN_RATE, 1:3)2 -0.339772   0.181033  -1.8769 0.0634533 .
## L(data3$UN_RATE, 1:3)3 -0.260515   0.095528  -2.7271 0.0075469 **
## L(data3$GDP_QGR, 1:1) -0.025568   0.012929  -1.9776 0.0507287 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Under final ADL (3,1) model which omits the autoregressive lag of order two and only considers the distributed lag of order one, all estimated coefficients are significant at the 5% significance level.

```
adl3_1 <- dynlm(data3$UN_RATE ~ L(data3$UN_RATE, c(1,3)) + L(data3$GDP_QGR, 1:1))
coeftest(adl3_1)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.207594   0.051350   4.0427 0.0001033 ***
## L(data3$UN_RATE, c(1, 3))1  1.399745   0.028735  48.7114 < 2.2e-16 ***
## L(data3$UN_RATE, c(1, 3))3 -0.431554   0.029004 -14.8789 < 2.2e-16 ***
## L(data3$GDP_QGR, 1:1)      -0.027101   0.013063  -2.0746 0.0405617 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The dynamic properties of the unemployment rate implied by these coefficient estimations seems loosely plausible. First, the fact that the estimated intercept is significantly positive results in a positive unconditional expectation of unemployment which one might interpret as pointing at the fact that the long-run/natural rate of unemployment is positive. Second, the fact that the lagged rather than current GDP growth is negatively related with unemployment rates is in line with frictional notions of the labor market which results in productivity/output shocks to affect unemployment with some delay.

- 2) Use the estimated ADL model to calculate and interpret the short-run multiplier, the 2-step-ahead multiplier, and the long-run multiplier. Finally, report and interpret the long-run relation between the unemployment rate and the GDP growth rate.

The final ADL model specification estimated above reads as follows where  $Y_t$  refers to the unemployment rate in period  $t$  and  $X_t$  refers to the GDP growth rate in period  $t$  (both measured in percentage points).

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_3 Y_{t-3} + \beta X_{t-1} + \varepsilon_t$$

Multiplier estimations are concerned with the effect of a unit shock to  $X_t$  on the value of  $Y_t$  in different periods. Since under the ADL-model considered here  $Y_t$  and  $X_t$  are independent, the short-run multiplier is evidently assumed to be 0.

The 2-step-ahead multiplier is obtained by iteratively applying the model specification above where  $X_t^* = X_{t+1}$  and the multiplier of the  $h$ 'th step equals  $Y_{t+h}^* - Y_{t+h}$  where for  $Y_t^*$  the shock to  $X_t$  reflected by  $X_t^*$  propagates through the system. From above we thus know that  $Y_t^* = Y_t$ . Furthermore:

$$\begin{aligned}
Y_{t+1}^* &= \alpha + \phi_1 Y_t^* + \phi_3 Y_{t-2} + \beta X_t^* + \varepsilon_{t+1} \\
&= \alpha + \phi_1 Y_t + \phi_3 Y_{t-2} + \beta(X_t + 1) + \varepsilon_{t+1} = Y_{t+1} + \beta \\
Y_{t+2}^* &= \alpha + \phi_1 Y_{t+1}^* + \phi_3 Y_{t-1} + \beta X_{t+1} + \varepsilon_{t+2} \\
&= \alpha + \phi_1(Y_{t+1} + \beta) + \phi_3 Y_{t-1} + \beta X_{t+1} + \varepsilon_{t+2} = Y_{t+2} + \phi_1 \beta
\end{aligned}$$

The estimated 2-step-ahead multiplier therefore equals  $\hat{\phi}_1 \hat{\beta}$ :

```
as.numeric(adl3_1$coefficients[2]*adl3_1$coefficients[4])
```

```
## [1] -0.03793497
```

The long-run relation between the unemployment rate  $\bar{Y}$  and the GDP growth rate  $\bar{X}$  reads as follows where  $\gamma_1$  corresponds to the long-run multiplier.

$$\bar{Y} = \alpha + \phi_1 \bar{Y} + \phi_3 \bar{Y} + \beta \bar{Y} \Rightarrow \bar{Y} = \frac{\alpha}{1 - \phi_1 - \phi_3} + \frac{\beta}{1 - \phi_1 - \phi_3} \bar{X} \equiv \gamma_1 + \gamma_2 \bar{X}$$

The coefficients of this long-run relationship  $\gamma_1$  and  $\gamma_2$  are estimated as respectively reported below:

```
as.numeric(adl3_1$coefficients[1]/(1-adl3_1$coefficients[2]-adl3_1$coefficients[3]))
```

```
## [1] 6.526205
```

```
as.numeric(adl3_1$coefficients[4]/(1-adl3_1$coefficients[2]-adl3_1$coefficients[3]))
```

```
## [1] -0.8519962
```

This estimated long-run relationship gives an estimation of Okun's law and shows how an ADL-model estimation implies an estimation of Okun's law. The parameter estimation obtained above implies that a one-percentage-point increase in the long-run GDP growth rate is associated with a -0.8519962 percentage point decrease in the long-run unemployment rate. Furthermore, if long-run output were to be constant, i.e.  $\bar{X} = 0$ , the long-run unemployment rate would be 6.526205%.

To verify the convergence of impulse response functions we can obtain an estimate for the unconditional expectation of the GDP growth rate, i.e.  $\bar{X} = \mathbb{E}(X_t)$ , as implied by the AR(3)-model specified above to obtain an estimate for the unconditional expectation of the unemployment rate. More specifically we have that:

$$X_t = \kappa + \psi_1 X_{t-1} + \psi_2 X_{t-2} + \psi_3 X_{t-3} \Rightarrow \mathbb{E}(X_t) = \bar{X} = \frac{\kappa}{1 - \psi_1 - \psi_2 - \psi_3}$$

Which is estimated to equal:

```
as.numeric(gdp3$coef[4]/(1-gdp3$coef[1]-gdp3$coef[2]-gdp3$coef[3]))
```

```
## [1] 2.83761
```

Plugging in the expression for  $\bar{X}$  into the long-run relationship above yields the estimated long-run unemployment rate:

$$\Rightarrow \bar{X} = \frac{\alpha}{1 - \phi_1 - \phi_3} + \frac{\beta}{1 - \phi_1 - \phi_3} \frac{\kappa}{1 - \psi_1 - \psi_2 - \psi_3}$$

Which is estimated to equal:

```
as.numeric(adl3_1$coefficients[1]/(1-adl3_1$coefficients[2]-adl3_1$coefficients[3])) + as.numeric(adl3_1$coefficients[4]/(1-adl3_1$coefficients[2]-adl3_1$coefficients[3]))*as.numeric(gdp3$coef[4]/(1-gdp3$coef[1]-gdp3$coef[2]-gdp3$coef[3]))
```

```
## [1] 4.108573
```

- 3) Please provide a detailed comment on the following statement: “An increase in the GDP growth rate causes a reduction in the unemployment rate.”

As described above, the obtained estimation of Okun’s law indeed suggestst that a permanent increase in the (long-run) GDP growth rate causes a permanent decrease in the (long-run) unemployment rate. This must however be contrasted with a temporary or one-off increase in the GDP growth rate, which leads to a decrease in the unemployment rate in the subsequent quarters (i.e. with a lag) but not to a permanent decrease. Rather, absent further shocks to the GDP growth rate, the unemployment rate will converge (back) to its long-run level estimated above. This notion will also become clear visually when inspecting the impulse response function for the unemployment rate below.

- 4) Suppose that the innovations are iid Gaussian. What is the probability of the unemployment rate rising above 7.8% in the second quarter of 2014? What is the probability that it drops below 7.8%? Do you trust the iid Gaussian assumption?

```
h = 8
forecast_ar <- predict(gdp3, n.ahead=h)$pred
forecast_ar

##           Qtr1      Qtr2      Qtr3      Qtr4
## 2014           0.3755792 0.5665948 0.5338011
## 2015 0.7007721 0.7984413 0.8345098 0.8898068
## 2016 0.9326006

forecast_ar <- as.numeric(forecast_ar)

GDP_QGR <- append(as.numeric(data2$GDP_QGR), forecast_ar)
UN_RATE <- append(as.numeric(data2$UN_RATE), numeric(h))

n <- length(data2$obs)
for (i in 1:h){
  j = i + n
  UN_RATE[j] <- adl3_1$coefficient[1] + adl3_1$coefficient[2] * UN_RATE[j-1] + adl3_1$coefficient[3] * UN_RATE[j-2]
}

fe1 <- summary(adl3_1)$sigma

"The probability that it is below 7.8:"

## [1] "The probability that it is below 7.8:"
pnorm(7.8, mean = UN_RATE[109], sd = fe1)

## [1] 0.2107482
"The probability that it is above 7.8:"

## [1] "The probability that it is above 7.8:"
1 - pnorm(7.8, mean = UN_RATE[109], sd = fe1)

## [1] 0.7892518
jarque.bera.test(adl3_1$residuals)

##
##  Jarque Bera Test
##
## data:  adl3_1$residuals
## X-squared = 6.4896, df = 2, p-value = 0.03898
```

If the iid Gaussian assumption holds, then we have a joint normal distribution. Hence, we conduct a Jarque Bera Test to test for normality (the  $H_0$  is that the residuals are normally distributed). As the resulting p-value is below 5%, we reject the  $H_0$ . This is why we conclude that we cannot trust the iid Gaussian assumption.

- 5) Make use of your estimated AR and ADL models to produce a 2-year (8 quarter) forecast for the Unemployment rate that spans until the first quarter of 2016. Report the obtained values.

```
UN_RATE <- ts(UN_RATE, start = c(1987, 2), frequency = 4)
UN <- ts_split(UN_RATE, sample.out = h)
```

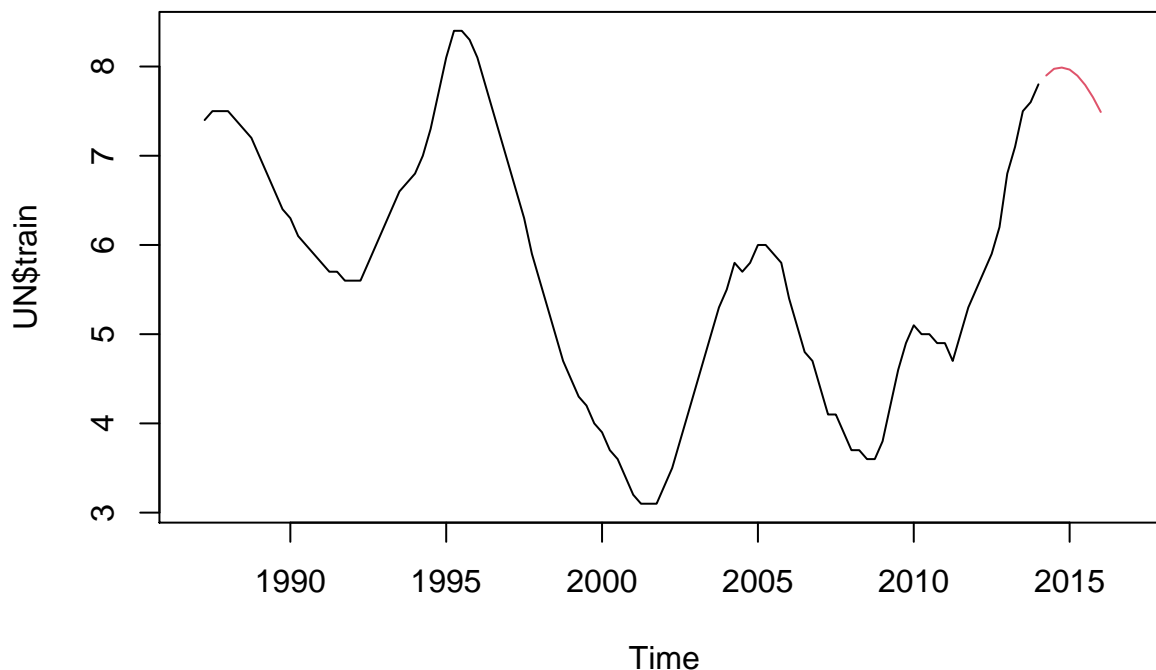
```
"The obtained values are:"
```

```
## [1] "The obtained values are:"
```

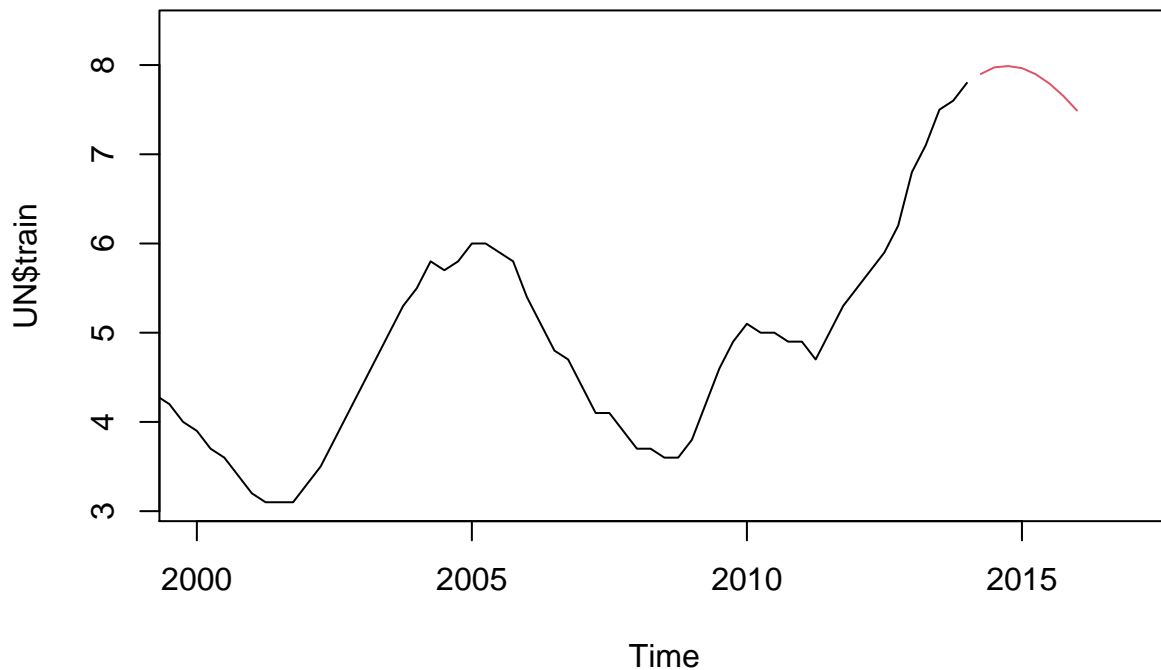
```
UN$test
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 2014          7.899095 7.974321 7.988131
## 2015 7.965586 7.897038 7.792483 7.654884
## 2016 7.490364
```

```
ts.plot(UN$train, xlim = c(1987, 2017))
points(UN$test, type="l", col=2)
```



```
ts.plot(UN$train, xlim = c(2000, 2017))
points(UN$test, type="l", col=2)
```



- 6) Use impulse response functions (IRFs) to analyze two different scenarios for the Dutch unemployment rate:
- (a) In the 'good scenario' the GDP quarterly growth rate is hit by a positive shock of 2%.
  - (b) In the 'bad scenario' the GDP quarterly growth rate suffers a negative shock of 2%. Please use the last observed value of the unemployment rate and gdp growth rate as the origin of your IRFs. In particular, set the origin to -0.37% for GDP and 7.8% for the unemployment rate.

### Good Scenario

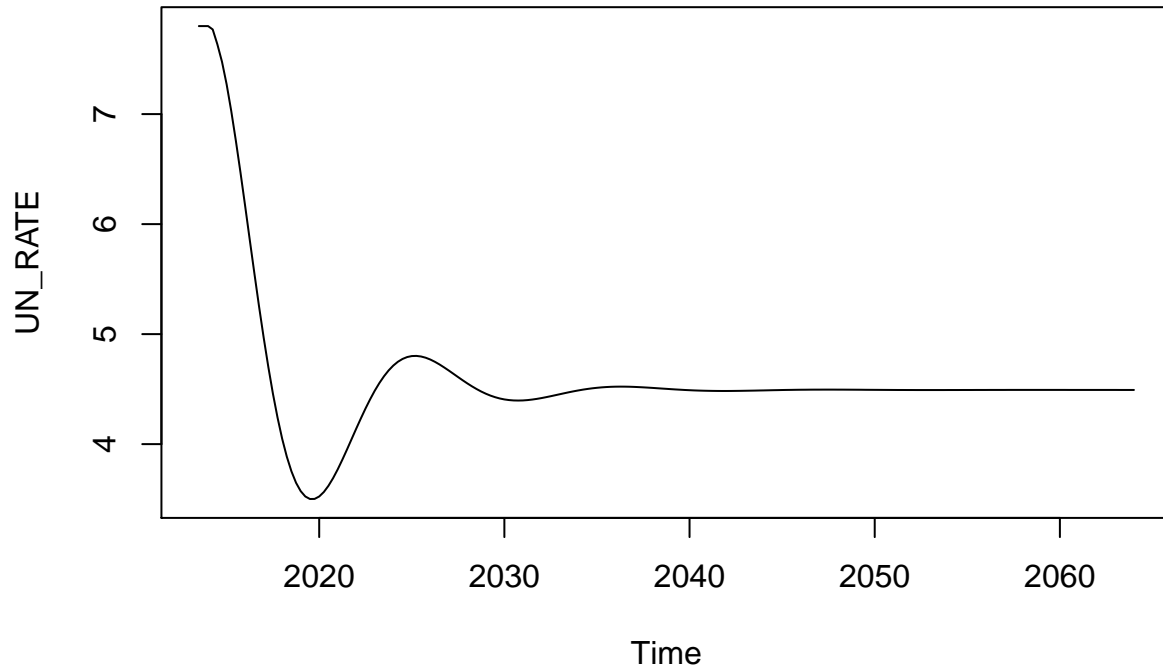
```
h = 200
GDP_QGR <- append(rep(data2$GDP_QGR[n],3), numeric(h))
UN_RATE <- append(rep(data2$UN_RATE[n],3), numeric(h))

GDP_QGR[3+1] = 2
for (i in 1:h){
  j = i + 3
  GDP_QGR[j] <- GDP_QGR[j] + ar3$coef[4] + ar3$coef[1] * GDP_QGR[j-1] + ar3$coef[2] * GDP_QGR[j-2] + ar3$coef[3] * GDP_QGR[j-3]
}

n <- length(data2$obs)
for (i in 1:h){
  j = i + 3
  UN_RATE[j] <- adl3_1$coefficient[1] + adl3_1$coefficient[2] * UN_RATE[j-1] + adl3_1$coefficient[3] * UN_RATE[j-2] + adl3_1$coefficient[4] * UN_RATE[j-3]
}
```

```
UN_RATE <- ts(UN_RATE, start = c(2013, 3), frequency = 4)
```

```
ts.plot(UN_RATE)
```



## Bad Scenario

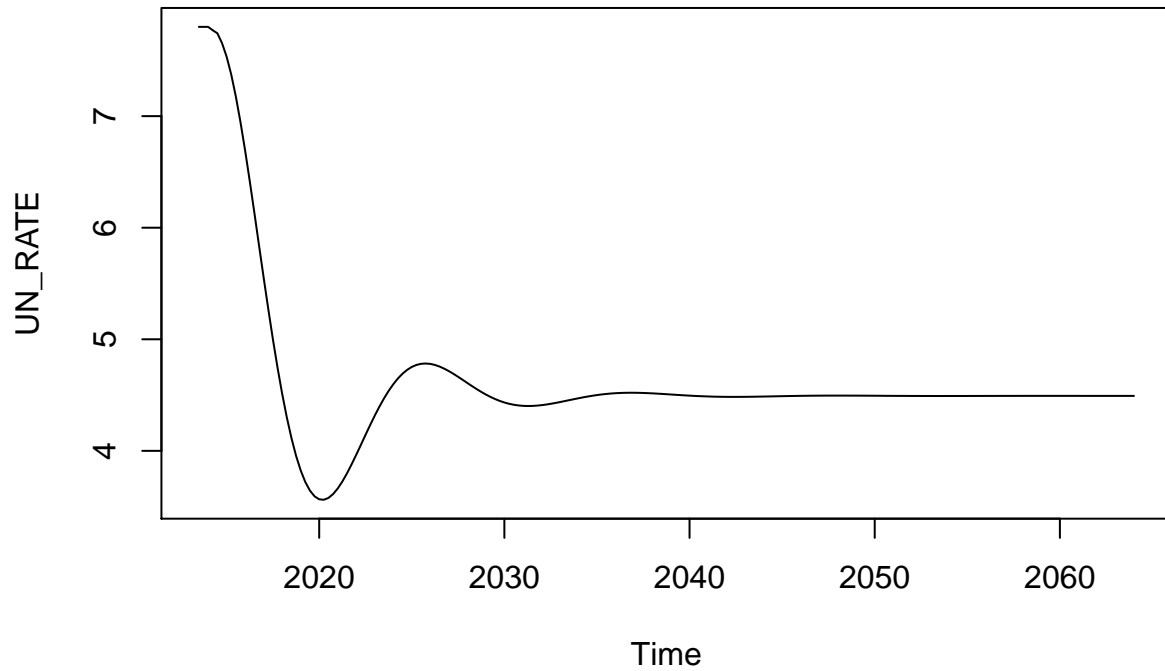
```
h = 200
GDP_QGR <- append(rep(data2$GDP_QGR[n],3), numeric(h))
UN_RATE <- append(rep(data2$UN_RATE[n],3), numeric(h))
```

```
GDP_QGR[3+1] = -2
for (i in 1:h){
  j = i + 3
  GDP_QGR[j] <- GDP_QGR[j] + ar3$coef[4] + ar3$coef[1] * GDP_QGR[j-1] + ar3$coef[2] * GDP_QGR[j-2] + ar3$coef[3] * GDP_QGR[j-3]
```

```
n <- length(data2$obs)
for (i in 1:h){
  j = i + 3
  UN_RATE[j] <- adl3_1$coefficient[1] + adl3_1$coefficient[2] * UN_RATE[j-1] + adl3_1$coefficient[3] * UN_RATE[j-2] + adl3_1$coefficient[4] * UN_RATE[j-3]
```

```
UN_RATE <- ts(UN_RATE, start = c(2013, 3), frequency = 4)
```

```
ts.plot(UN_RATE)
```



As the plots show, although the shocks are different from each other, there is no noticeable difference between the impulse responses. The reason might be that the GDP's effect on the unemployment rate is rather low, especially in comparison to the lags of the unemployment rate. Around the year 2042, both impulse responses converge to around 4.5. However, note that based on our previous calculations, the impulse responses are expected to converge to an value of around 4.1. Unfortunately, we could not determine the reason for the difference.