Econometrics III
Assignment Part I

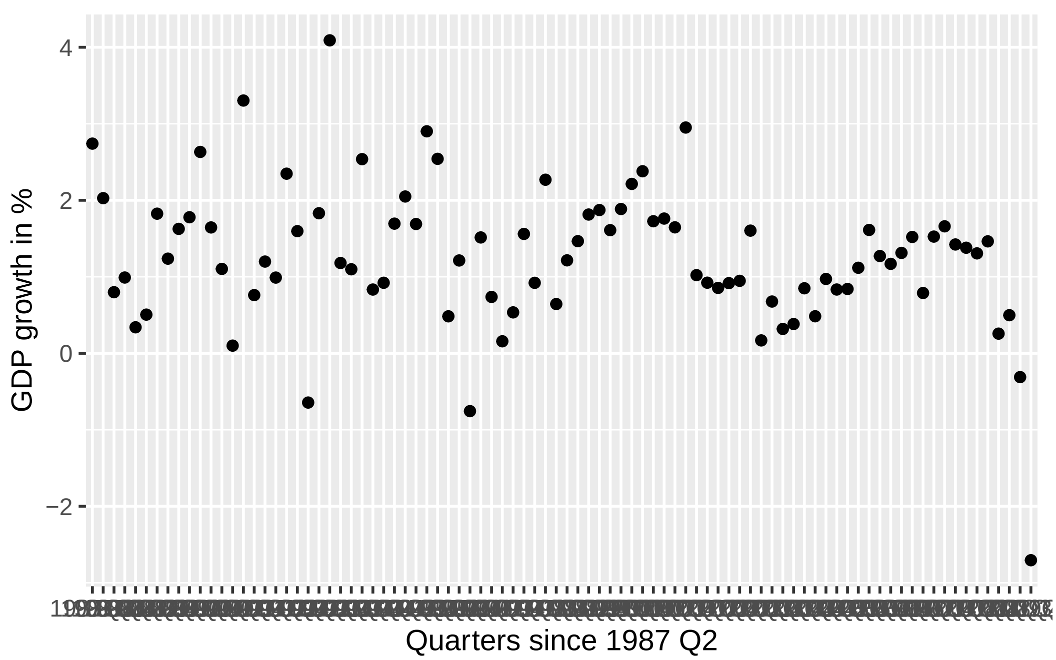
Thao Le (523716)
David Gyaraki (582340)

Contents

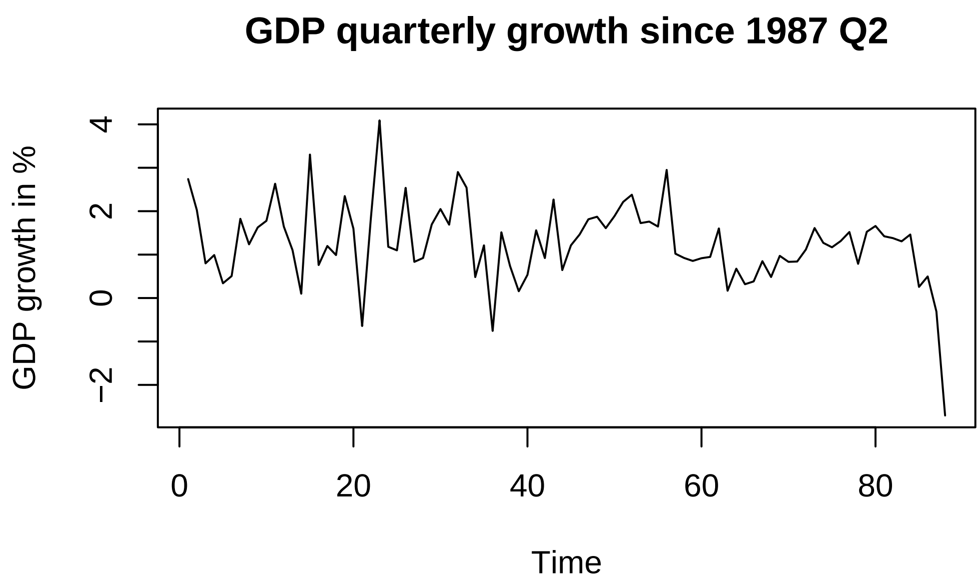
```
# load packages
if(!require(pacman)){install.packages("pacman")}

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

dfAssign_p1 <- as.data.frame(read_csv("data/data_assign_p1.csv"))

# Encode quarters
dfAssign_p1 <-
 ↪  cbind(dfAssign_p1,c(seq(1,nrow(dfAssign_p1),length.out=nrow(dfAssign_p1))))
colnames(dfAssign_p1) <- c("obs", "GDP_QGR", "ind")
```

1 Question 1

```
ggplot(dfAssign_p1, aes(x = obs, y = GDP_QGR)) + geom_point() +
 ↪  ylab("GDP growth in %") + xlab("Quarters since 1987 Q2")
```



```r
plot.ts(dfAssign_p1$GDP_QGR, main="GDP quarterly growth since 1987 Q2",
  ylab="GDP growth in %")
```

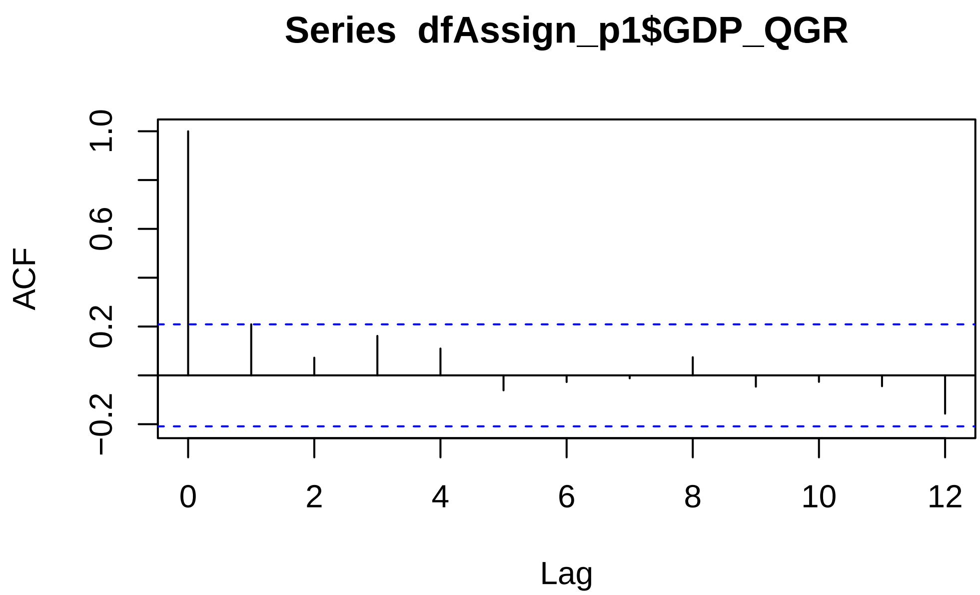


The Dutch quarterly growth rates seems to be fluctuating around 1 % every quarter until it experiences a sharp drop at Quarter 1 of 2009. This growth rate reflect the influence of the economic recession.

```r
acf(dfAssign_p1$GDP_QGR,12,pl=F)
```

```
Autocorrelations of series 'dfAssign_p1$GDP_QGR', by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.209  0.072  0.162  0.110 -0.061 -0.027 -0.012  0.075 -0.046 -0.027
    11     12
-0.045 -0.157
```

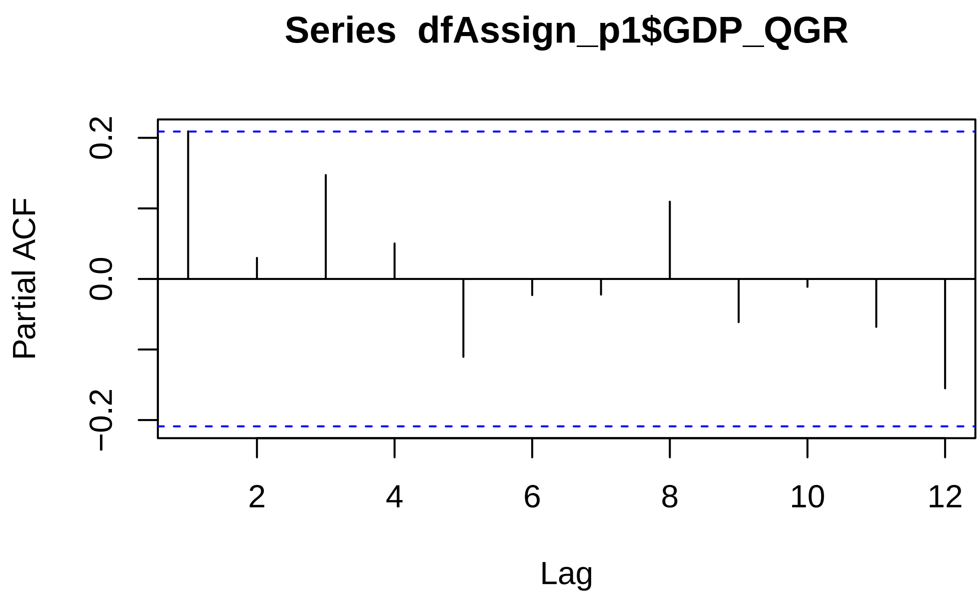
```r
acf(dfAssign_p1$GDP_QGR,12,pl=T)
```




```
pacf(dfAssign_p1$GDP_QGR,12,pl=F)
```

```
Partial autocorrelations of series 'dfAssign_p1$GDP_QGR', by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.209  0.030  0.147  0.050 -0.110 -0.023 -0.022  0.110 -0.061 -0.011 -0.068
    12
-0.155
```

```
pacf(dfAssign_p1$GDP_QGR,12,pl=T)
```



From the sample ACF and partial ACF plots, we can observe the autocorrelation of the data with each defined lag up to 12. From the sample ACF plot (former), we can observe that there is quite a bit of positive autocorrelation with the first, second, third and fourth lags (although only the first reaches the white-noise threshold). After these, there is little correlations, with changing signs, only the twelfth shows a relatively larger negative autocorrelation. From the partial plot, we can observe more or less the same picture, except that the first and third legs are quite enhanced while the second and fourth are less pronounced now. Furthermore we can see that the fifth lag is more pronounced negative now along with the twelfth, and interestingly the eighth shows quite a large positive correlation too.

Using the sample ACF, the dynamics seems to be that GDP quarterly growth rate seems to be correlated to the first 4 lags.

2 Question 2

```
ar4 <- arima(dfAssign_p1$GDP_QGR, order=c(4,0,0))
#coeftest(ar4)

ar3 <- arima(dfAssign_p1$GDP_QGR, order=c(3,0,0))
#coeftest(ar3)
```

```r
ar2 <- arima(dfAssign_p1$GDP_QGR, order=c(2,0,0))
#coeftest(ar2)

ar1 <- arima(dfAssign_p1$GDP_QGR, order=c(1,0,0))
#coeftest(ar1)
```

We start by estimating an AR(4) model and we attempt to investigate the coefficients in the model. Table 1 shows us the estimated model parameters in the AR(4) model and the subsequent restricted models. We find that the fourth lag coefficient (along with second and third) is not significant and therefore we start an elimination of lags from the model one-by-one. Column (2) contains the AR(3) model's estimated parameters then, where we can see that by removing the fourth lag, the third one became more enhanced, but this is not enough for our criteria of 5% significance level. This implies that the third lag might be important to us later, but for now we will remove it from the model. Then the second lag should also be removed due to insignificance, and then we are left with only one, the AR(1) model with first-order autocorrelation, which is significant in predicting this quarter's GDP growth. Hence we can say that the first lag of GDP growth seems to be a significant predictor of the GDP growth in any quarter shown by (4) column in Table 1, in a positive relationship, i.e. if the GDP growth was positive last period, we predict 0.272 times that growth in this period (plus the intercept, which is significant positive). This also implies that GDP tends to exhibit similar patterns in one lag with some moderated coefficient, so if GDP growth was positive in a period, this is likely to repeat in the next period with some moderated effect.

```r
stargazer::stargazer(ar4,ar3,ar2,ar1, title="Estimating the AR(4) to
↪  AR(1) models on GDP data",  align=TRUE, label = "tab_ar4",
↪  table.placement="H", out = "tab_ar4.tex")
```

Table 1: Estimating the AR(4) to AR(1) models on GDP data

	Dependent variable:			
	(1)	(2)	(3)	(4)
ar1	0.242^{**}	0.256^{**}	0.263^{**}	0.272^{**}
	(0.118)	(0.117)	(0.119)	(0.118)
ar2	0.030	0.030	0.058	
	(0.120)	(0.120)	(0.121)	
ar3	0.189	0.200^{*}		
	(0.119)	(0.119)		
ar4	0.086			
	(0.120)			
intercept	1.214^{***}	1.228^{***}	1.249^{***}	1.253^{***}
	(0.204)	(0.180)	(0.140)	(0.130)
Observations	88	88	88	88
Log Likelihood	-113.211	-113.468	-114.849	-114.963
σ^2	0.765	0.770	0.796	0.798
Akaike Inf. Crit.	238.421	236.936	237.698	235.925
Note:				*p<0.1; **p<0.05; ***p<0.01

3 Question 3

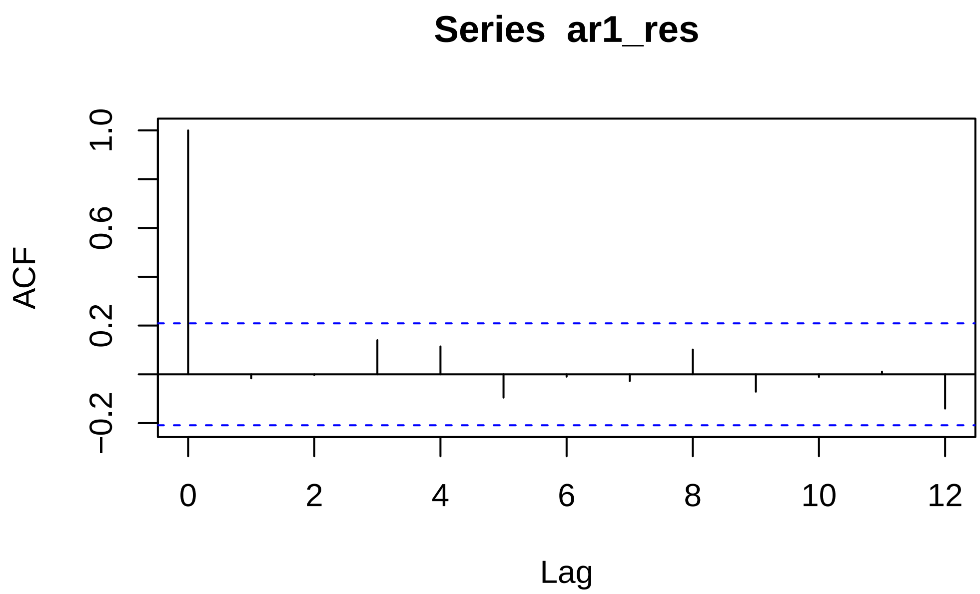
The final model we chose is the AR(1) model.

```
ar1_res <- as.numeric(ar1[["residuals"]])  #estimated residual ACF
 ↪  function
acf(ar1_res,12,pl=F)
```

```
Autocorrelations of series 'ar1_res', by lag

     0      1      2      3      4      5      6      7      8      9     10
```

```
1.000 -0.017 -0.003  0.140  0.114 -0.096 -0.010 -0.028  0.102 -0.071 -0.011
   11     12
0.011 -0.140
```

```r
acf(ar1_res,12,pl=T)
```




```r
resar4 <- arima(ar1_res, order=c(4,0,0))
coeftest(resar4)
```

```
z test of coefficients:

           Estimate Std. Error z value Pr(>|z|)
ar1       -0.036058   0.116826 -0.3086  0.75759
ar2        0.025976   0.117600  0.2209  0.82518
ar3        0.205153   0.118412  1.7325  0.08318 .
ar4        0.180039   0.118957  1.5135  0.13016
intercept -0.037084   0.147513 -0.2514  0.80151
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this case, we can investigate the autocorrelation of the residuals in the AR(1) model. The plot above investigates the autocorrelation in the residuals up to 12 lags, and immediately we can see that the previously large first order autocorrelation has been removed thanks to our AR(1) model estimation. However, there are quite some correlations appearing in the third and fourth lags, along with some levels of correlation in the fifth, eighth and twelfth as previously seen. This would imply that the autoregressive model is not defined as best, since we still leave some autocorrelation behind. However, if we estimate an AR(4) model on the residuals, we still find that the third lag is only significant with 10% level so this could provide us with some justification even if our model is not so well specified.

4 Question 4

```
forecast8 <- forecast(ar1, 8, level = c(95))
predlbl <- c("2009Q2", "2009Q3", "2009Q4", "2010Q1", "2010Q2", "2010Q3",
↪   "2010Q4", "2011Q1")
forecast_q4 <- cbind(predlbl, as.data.frame(forecast8))

pred <- predict(ar1, 8)

kable(forecast_q4[,1:2], caption="The predicted quarterly GDP growth
↪   rates using ARIMA(1,0,0)", col.names = c("Quarter", "Predicted
↪   growth %"), digits = 3, row.names = F)
```

Table 2: The predicted quarterly GDP growth rates using ARIMA(1,0,0)

Quarter	Predicted growth %
2009Q2	0.177
2009Q3	0.960
2009Q4	1.173
2010Q1	1.231
2010Q2	1.247
2010Q3	1.251
2010Q4	1.252
2011Q1	1.253

```
mean(dfAssign_p1$GDP_QGR)
```

```
[1] 1.263169
```

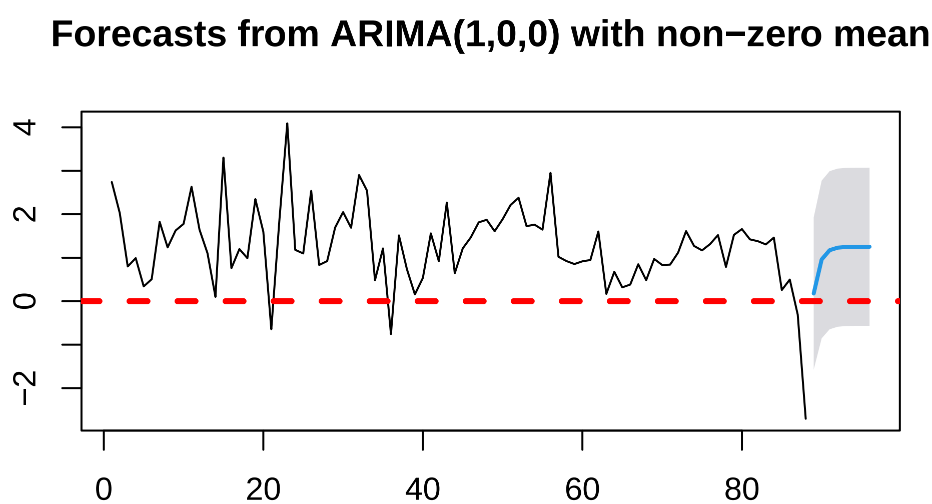
Since we estimated the AR(1) model with nonzero mean, our prediction for the first period follows the calculation:

$$\hat{Y}_{T+1} = \delta + \phi(Y_T - \mu), \tag{1}$$

where δ is the estimated intercept, ϕ is the coefficient of the AR(1) lag term and $\mu = 1.26317$ is the mean of the GDP growth in the estimation period. Therefore, we can calculate the value for the Q2 of 2009 by $(-2.7051 - 1.26317) * 0.272 + 1.253 = 0.174$, which closely corresponds to the function estimates of 0.177. For all further values in time period t, we simply use the predicted values for period $t-1$. We can then manually confirm that indeed the values predicted by the function are correct.

5 Question 5

```
plot(forecast8) %>%
abline(h=0, col="red", lwd=3, lty=2)
```



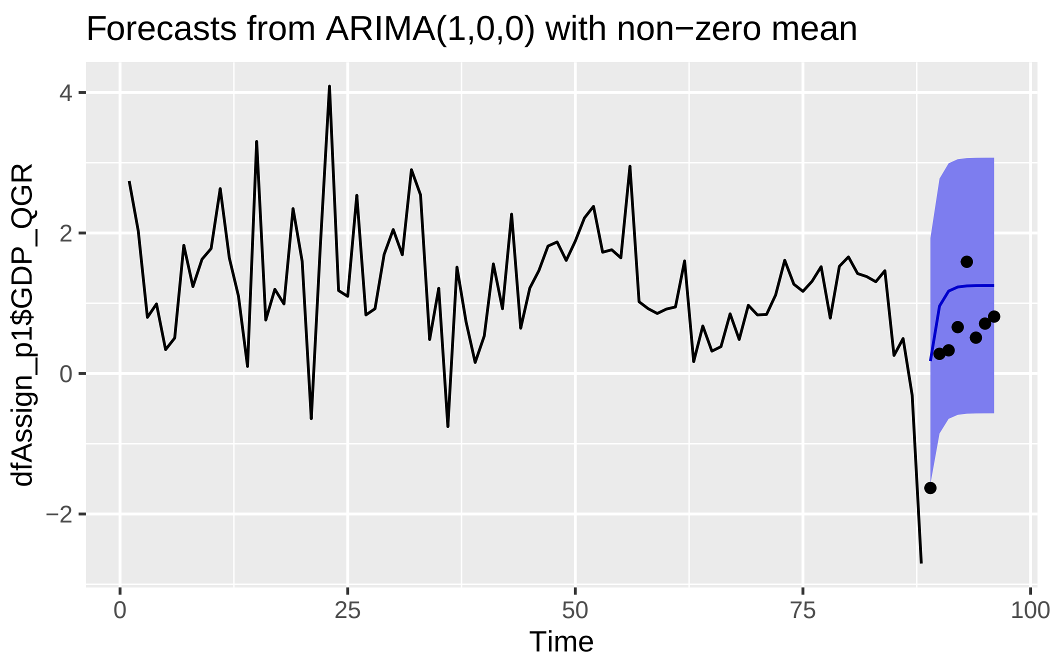
If we assume that the innovations in our AR model are IID Gaussian, we can use the estimated model and its coefficients for forecasting the values and their confidence interval with precision. In this case, we assume that our estimates obtained in Section 4 are not influenced by any remaining correlation in the residuals (potentially from the second, third or fourth lags) and the predictions produced will not be tainted by autocorrelation. From the plot in Section 5 we can see that the prediction interval is quite wide, with 0 continuously within the interval as well as some negative values. Therefore, even if our prediction is positive (expecting growth in GDP in 2009 Q2), there is indeed a really good probability that the growth in Q2 will stay negative, since almost half of the confidence interval lies within the negative territory.

6 Question 6

The assumption that the residuals are iid Gaussian is not reasonable. As we have seen before, even though not at 5% level, but the second, third and fourth lag coefficients came close to being significant, implying that there mighte be some positive autocorrelation up to the fourth lag. Since we have removed them from the model, we will likely introduce autocorrelated errors in the estimated model, which might explain the strong uncertainty in the confidence interval around our predictions. Therefore, as the graph also implies, we may have an overestimation bias by not including the second-to-fourth lags in the model and the CI of the estimates will implode because of the reduced efficiency of the model (although the coefficient estimates will be consistent and unbiased).

7 Question 7

```
dfActuals <- data.frame(x=c(89, 90, 91, 92, 93, 94, 95, 96),
          y=c(-1.63, 0.28, 0.33, 0.66, 1.59, 0.51, 0.71, 0.81))

autoplot(forecast8, showgap = F) + geom_point(data = dfActuals, aes(x=x,
 ↪  y=y))
```



From the plot we can read off the predicted quarterly GDP growth for two years along with the confidence intervals at a 95% level. Then we can compare these predicted values and the confidence interval to the actual observed growth rates per period. While we can claim that most actual values are within our confidence intervals (apart from the first predicted period 2009 Q2, where the actual growth is slightly below the lower bound), this has more to do with the uncertainty of our predictions rather than the accuracy. The CI intervals are quite large, generally spanning 3-3.5% GDP growth, which makes our predictions quite useless (predicting that GDP growth will be between 0 and 3.5% next quarter offers relatively no insight). Also our AR(1) model seems to consistently overestimate the actual values with the predictions, apart from one (2010 Q2), all observations are below the predicted line. Therefore our accuracy is quite weak and it seems that our model consistently overestimates the actual GDP growth in the prediction window.

8 Question 8

Iteratively, we need to estimate the AR(4) model with the lags and at each step, exclude the one with the highest p-value to make sure that we exclude the least significant lags. If we exclude the lags sequentially backwards, we arrive to the conclusion that the third lag coefficient is significant on the 10% level but the second lag is not. Therefore, we need to estimate the model in different steps.

```r
ar4_10 <- arima(dfAssign_p1$GDP_QGR, order=c(4,0,0))
coeftest(ar4_10)
```

```
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
```

```r
# Excluding second lag
ar3_10 <- arima(dfAssign_p1$GDP_QGR, order=c(4,0,0),
↪  fixed=c(NA,0,NA,NA,NA))
coeftest(ar3_10)
```

```
z test of coefficients:

          Estimate Std. Error z value  Pr(>|z|)
ar1       0.246717   0.116394  2.1197   0.03403 *
ar3       0.193319   0.118368  1.6332   0.10243
ar4       0.086382   0.120270  0.7182   0.47261
intercept 1.217907   0.194496  6.2619 3.804e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ar2_10 <- arima(dfAssign_p1$GDP_QGR, order=c(4,0,0),
↪  fixed=c(NA,0,NA,0,NA))
coeftest(ar2_10)
```

```
z test of coefficients:

          Estimate Std. Error z value  Pr(>|z|)
ar1        0.26015    0.11542  2.2540   0.02420 *
ar3        0.20417    0.11798  1.7305   0.08354 .
```

```
intercept  1.23099    0.17276  7.1256 1.036e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence our final model contains two lags significant at 10% level, the first and third lags, indicating some sort of quarterly-semiannual mechanism in GDP growth.

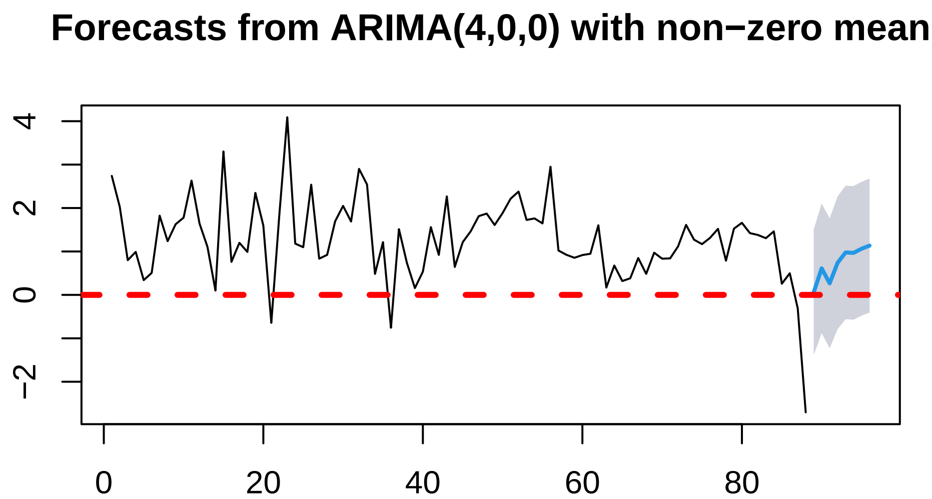
```
forecast_10 <- forecast(ar2_10, 8, level = c(90))
forecast_10db <- cbind(predlbl, as.data.frame(forecast_10))

kable(forecast_10db[,1:2], caption="The predicted quarterly GDP growth
↪ rates using ARIMA(4,0,0) with second and fourth lags removed and 90%
↪ CI", col.names = c("Quarter", "Predicted growth %"), digits = 3,
↪ row.names = F)
```

Table 3: The predicted quarterly GDP growth rates using ARIMA(4,0,0) with second and fourth lags removed and 90% CI

Quarter	Predicted growth %
2009Q2	0.057
2009Q3	0.611
2009Q4	0.266
2010Q1	0.740
2010Q2	0.977
2010Q3	0.968
2010Q4	1.062
2011Q1	1.135

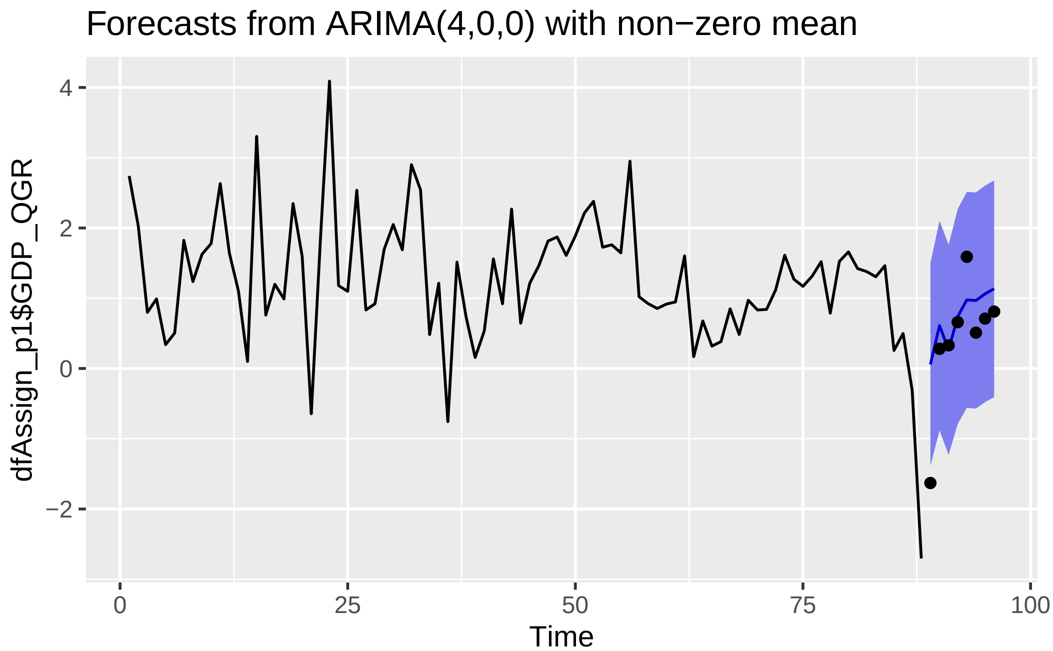
Here we removed lag 4 and stop at lag $p = 3$ because lag 3 is significance at a 10$\%$ significant level.

```
plot(forecast_10) %>%
abline(h=0, col="red", lwd=3, lty=2)
```



We can see that the AR(1-3) model (AR(3) without the second lag) also predicts the growth rate with great uncertainty and the growth rate confidence interval of 90% consistently includes 0, which implies that we are not certain about whether the growth will turn positive again in the coming 2 years, even though our estimates imply that they will.

```
autoplot(forecast_10, showgap = F) + geom_point(data = dfActuals,
↪   aes(x=x, y=y))
```



Now in the case of the 90% confidence interval, we estimate the model using an AR(1-3) since we find that the estimated model's third lag coefficient is significant on a 10% level. Using this AR(1-3) model, we can obtain the predictions as well as the 90% confidence interval similarly as before. We can see that the AR(1-3) model results in a better prediction model, because even the 90% confidence interval estimate around the prediction yields about 3% spread between lower and upper bounds. Furthermore, except for the first prediction in 2009 Q2, our predictions are all within the interval and the 2009 Q3 through 2010 Q1 predictions are quite accurately almost on the prediction line. Just like in the previous case, 2010 Q2 is severely underestimated which implies that the observation is probably an unexpectedly higher GDP growth, almost an outlier of growth in that environment. However the further we progress into time, the more inaccurate our predictions become and there is an appearing trend again that the AR(1-3) model will overestimate the predictions for the second half of 2010. Therefore, I would say that the model is a better predictor than the AR(1), although this might also be expected since the AR(1-3) takes more lags into account.

8.1 AMAZING WORK, I THINK WE SHOULD RENAME THE PLOTS AND THEN WE'RE GOOD.