

# Econometrics III HW - part 4

A. Schmidt and P. Assunção

April, 2020

## Loading packages and helper functions

See the source code if interested in all functions (chunks were omitted unless relevant for the assignment).  
Click [here](#) to access the code.

## Assignment 4

### Introduction

*A large Dutch retailer of consumer goods is interested in predicting the effects of a potential increase in VAT and other consumption taxes over its sales. In particular, this retailer would like you to explore the relation between the total consumption of non-durable goods and the fluctuations in the total disposable income of families in The Netherlands. Luckily, they have turned to you for technical support on this matter!*

### Importing and checking data

```
urlRemote <- "https://raw.githubusercontent.com/aishameriane"
pathGithub <- "/Mphil/master/EconIII/data_assign_p4.csv"
token <- "?token=AAVGJTREMSHGOMB4JUOXYGC6TLPEY"

url <- paste0(urlRemote, pathGithub, token)
dfData01 <- read.csv2(url, sep = ",", dec = ".", header = TRUE)
```

Check if everything is ok with the dataset: header and tail and summary statistics to check for missing data/outliers. The head and tail of the table below show us the Data set indeed starts at the first quarter of 1988 and ends at the first quarter of 2012.

obs	CONS	INC	TIME	obs	CONS	INC	TIME
1988Q1	53803.3	69522.0	725736	2010Q4	88001.2	120510.3	734045
1988Q2	54124.1	70033.3	725827	2011Q1	87213.2	121219.3	734137
1988Q3	54383.8	70634.5	725918	2011Q2	87176.8	121229.9	734227
1988Q4	54783.6	71321.8	726010	2011Q3	86841.6	120748.0	734318
1989Q1	54980.6	72377.2	726102	2011Q4	86579.1	119971.0	734410
1989Q2	55462.3	73251.1	726192	2012Q1	86767.4	120271.0	734502

The column with the dates (TIME) is in numeric format and not in a nice position with respect to the order of the variables, so we will change it to y-m-d and also switch places with the CONS column.

```
dfData01$TIME <- as.Date(dfData01$TIME, origin="0001-01-01")
dfData01      <- dfData01[,c(1,4,2,3)]
```

Next, we present the descriptive statistics for the quarterly aggregate consumption and household income. We observe that indeed there is no missing information and all values are numeric (there are no problems of formatting).

	Aggregate Consumption (cons)	Aggregate household income (inc)
Observations	97.00	97.00
Minimum	53803.30	69522.00
1st quartile	61419.10	81879.90
Mean	72556.16	98885.57
Median	74959.30	103286.80
3rd quartile	82378.80	113800.80
Maximum	88001.20	122816.00
Desv. Pad.	11215.00	16805.94

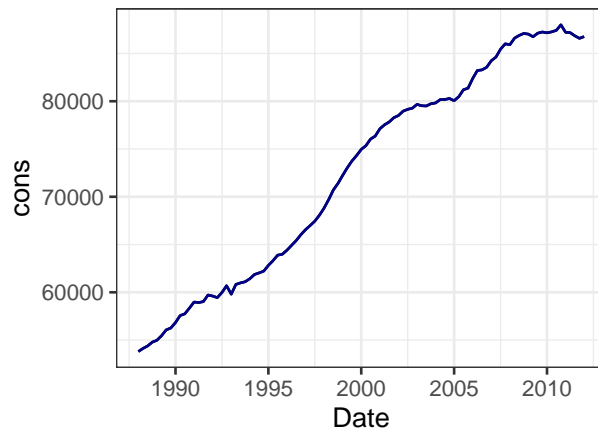
## Question 1

*Plot both the aggregate consumption and aggregate income time-series. Compute and report 12-period ACF and PACF functions for each series and comment on their shape.*

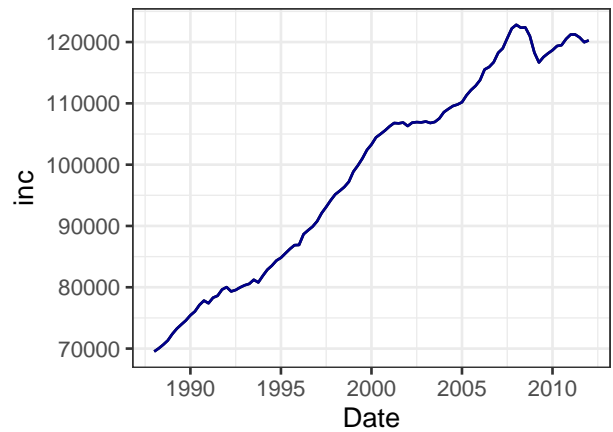
Based on the visual inspection (prior to formally testing), both series seem to not be stationary, since the unconditional average does not seem to be constant over time. This can be seen from the fact that the series does not move around some level: both aggregate income and consumption shows a linear positive increasing behaviour over time. Note that, around the financial crisis of 2008, we observe a brief decrease over aggregate income, but soon after the series is slowly back to its increasing trend. As for consumption, we observe a smoothening of consumption growth during the crisis (which is better seen when looking at the series alone, given the difference in scale), and a somewhat “flat” behaviour after-crisis, with aggregate consumption neither increasing or decreasing. When comparing both series together, we observe that the gap between income and consumption increased between 1987 and 2012.

Looking at the ACFs, we observe a slow decay of the autocorrelation function and several positive significant lags at a 5% significance level, which offers further support for the fact that the series are not stationary - this is based also on the ACF for 60 lags, which is not reported here. It is also important to notice that this result would be sustained even for other confidence levels such as 90% or lower. The “long period ACF” shows a very slow decay in the ACF without an abrupt cut, which suggests an AR model, with no MA component. As for the PACFs, it tells us that only the first lag autocorrelations is significant. Overall, both ACF and PACF, for both aggregate consumption and income, provide evidence of autocorrelation in the series. Such autocorrelation has a persistent behaviour, even considering larger lags.

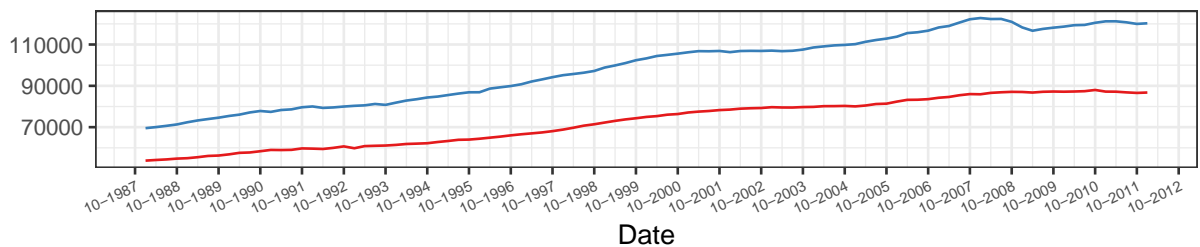
Quarterly aggregate consumption  
1988Q1 to 2012Q1



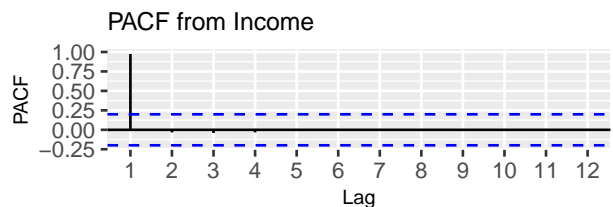
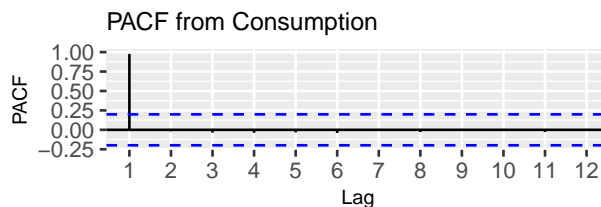
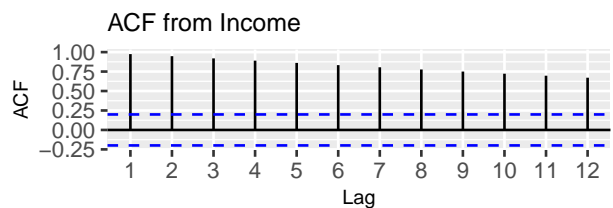
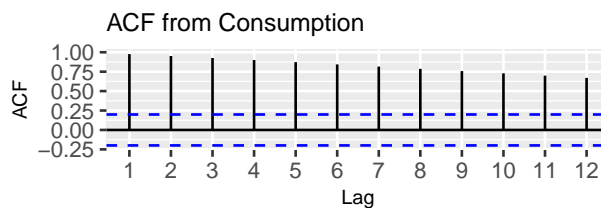
Quarterly aggregate income  
1988Q1 to 2012Q1



Quarterly aggregate consumption and income (in Euros)  
1988Q1 to 2012Q1



Variable — GDP Growth — Unemployment



## Question 2

Perform an ADF unit-root test on each series using the general-to-specific approach and report the values of the test statistics. Is the unit-root hypothesis rejected in any of them?

We divided this question in two parts, one for the model fitting and another for the unit root test.

## Fitting an AR(p) model

Given the behavior of the ACF and PACF from the previous items, there is no clear indicator that we are dealing with an ARMA(p,q) model with  $q \neq 0$ . So, using the same parcimonious argument from the Homework 3 (which helps to avoid parameter redundancy, as discussed in Chapter 7 of Box et al. (2015)), we are going to fit directly an AR(q) model. In here we are also using our function from HW3 to run models with different lags up until lag 4 and decide using the BIC which models fits the best, before performing the ADL test.

About the choice of  $p = 4$  as maximum lag, we tried to use more lags, but this incurred in numerical errors when using the `Arima()` function. More specifically, the method is based on ML estimation and requires to invert a Hessian matrix at some point, but depending on the specification of the model, this matrix will not be (numerically) non-singular, causing the algorithm to crash. Upon further investigation, we found out that the function does not return a model with inverse roots outside the unit circle (source: <https://otexts.com/fpp2/arima-r.html>). As we are going to find out in this homework, both series are  $I(1)$  and hence this might be the (unit) root of the problem with the algorithm. Nevertheless, we considered that starting from an AR(4) is a fair start because this represents an entire year in terms of our series.

The results from the best models for each series are summarized below. As pointed out in the feedback for parts 1, 2 and 3, the `Arima()` package does not compute the intercept when the AR component is present in the model. We therefore adjusted our code to make notation and computation consistent (more specifically, our tables now exhibits “mean” instead “intercept”, as suggested in this post by Prof. Stoffer).

We notice that for the consumption model, the statistics for the mean of the process are treated as NaN (not a number) - more specifically, the problem starts with the standard error, which is then carried on to the t-statistics and subsequently to the p-value. This also supports the hypothesis that there is something suspicious with the series and it requires further investigation to assume a model, in particular to assume stationarity (since this holds only when the first two unconditional moments are well defined). As for the income, we see that it has some (very high mean) and a finite standard error which is also very high.

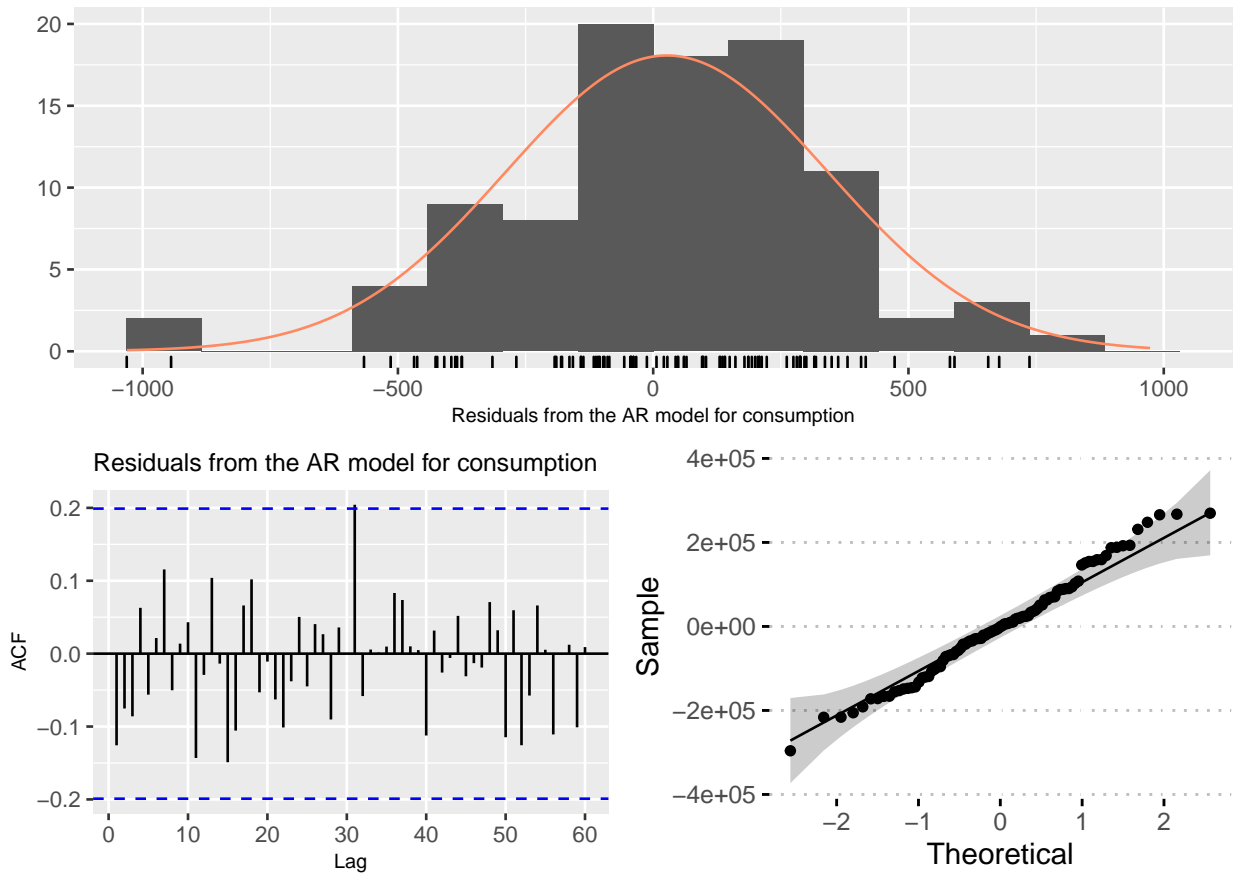
## [1] "Best AR model for Consumption"

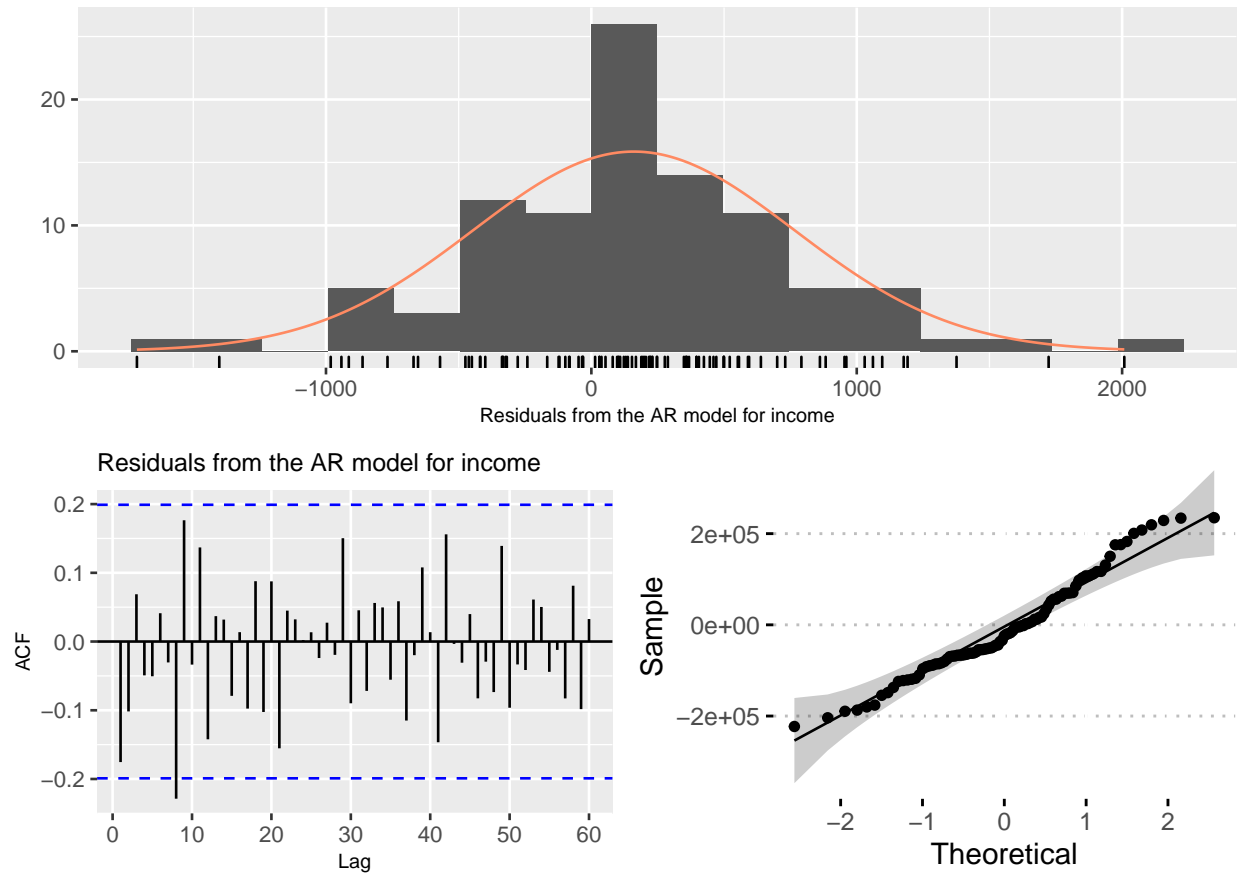
Variable	Estimate	Std. Error	t-statistic	P-Value
Lag 1	1.11	0.07	15.53	0.00
Lag 2	0.15	0.07	2.14	0.03
Lag 3	0.25	0.08	3.22	0.00
Lag 4	-0.50	0.07	-6.81	0.00
Mean	92612.28	NaN	NaN	NaN

## [1] "Best AR model for Income"

Variable	Estimate	Std. Error	t-statistic	P-Value
Lag 1	1.69	0.07	23.25	0
Lag 2	-0.69	0.07	-9.46	0
Mean	95525.58	22423.43	4.26	0

*Comments:* Based on the ACF for the residuals of both AR models using a significance level of 5%, it looks like that we have a white noise (there is one significant lag for income around  $h = 8$  and one for consumption around  $h = 31$ , however this could be due to a type I error). However, when we look at the QQplots, it seems that we have some problems on the income residuals (several values fall outside the 95% confidence interval) when assuming a normal distribution. Since the question does not asks for the gaussian specification, we are not going to pursue this for now.





### Unit root tests

We are using the same function used for homework 3 and the reported results are shown below.

```
## [1] "Results from the ADF test"
```

Variable	p of AR(p)	Test-statistic	p-value
CONS	4	-1.2140	0.9006
INC	2	-1.4257	0.8140

*Comments:* The ADF test for the Consumption AR(4) model has a test statistic of -1.214 (p-value = 0.9006) while the ADF test for the AR(2) fitted for income had a test statistic of -1.4257 (p-value = 0.814), using the null hypothesis that the series is not stationary. This means that at any reasonable confidence level (either 5% or even 10%), we do not reject the null hypothesis of having unit roots and both series are considered non-stationary.

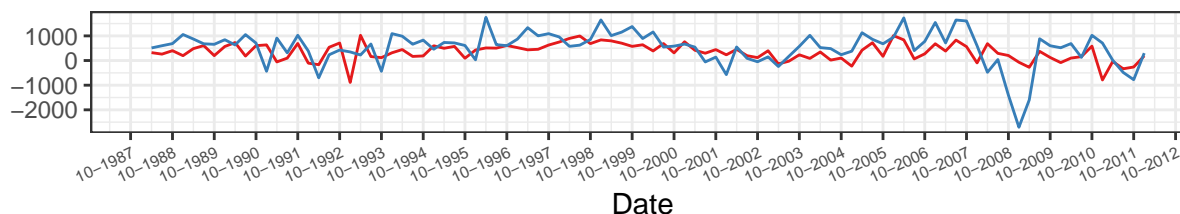
### Question 3

*Perform an ADF unit-root test on the first difference of each series using the general-to-specific approach and report the values of the test statistics. Is the unit-root hypothesis rejected in any of them? What do you conclude about the order of integration of these time series?*

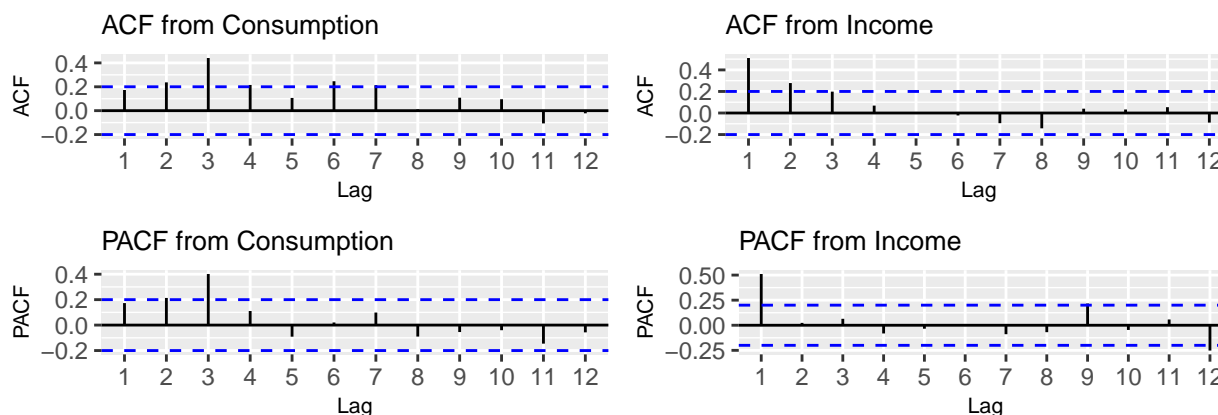
Since we did the fit in the previous question up until lag 4, we will also use this as maximum lag here. As for the differencing factor, as we discussed in class, several series in economics are  $I(1)$  and very few variables exhibit a  $I(2)$  behavior. To help us decide, we have below the plots of the two series after taking the first difference, as well as the ACFs and PACFs.

## Quarterly consumption and income (1st difference)

1988Q1 to 2012Q1



Variable — Consumption — Income



*Comments* We can see from the plots that the series seems to be stationary (only based on the visual inspection, but testing is needed to verify this hypothesis) when we take the first difference, because they move randomly around some mean and do not appear to exhibit clusters of volatility in general, with the exception maybe for the beginning of the 90s (since this is not a GARCH exercise, we will not focus on this now).

There is a prominent outlier for the Income in the aftermath of the financial crisis, while the same did not happen (at least with the same magnitude) to the consumption. This might be explained by the hypothesis that households smooth consumption and in this case could be that they used another source as income, for example, savings or loans. Overall, it seems that, by taking the first difference, we were able to remove the positive linear trend from before.

From the ACF/PACF, we see that the income seems to have a more short dependency of its lags (in the PACF only the first lag remains significative), while for consumption it seems that the third still plays a role. So, without estimating, we would expect an AR(p) for consumption with a higher  $p$  than for income. This is also supported by the graphs of the series: the income differentiated series is less smooth, with more prominent jumps at every period than the consumption.

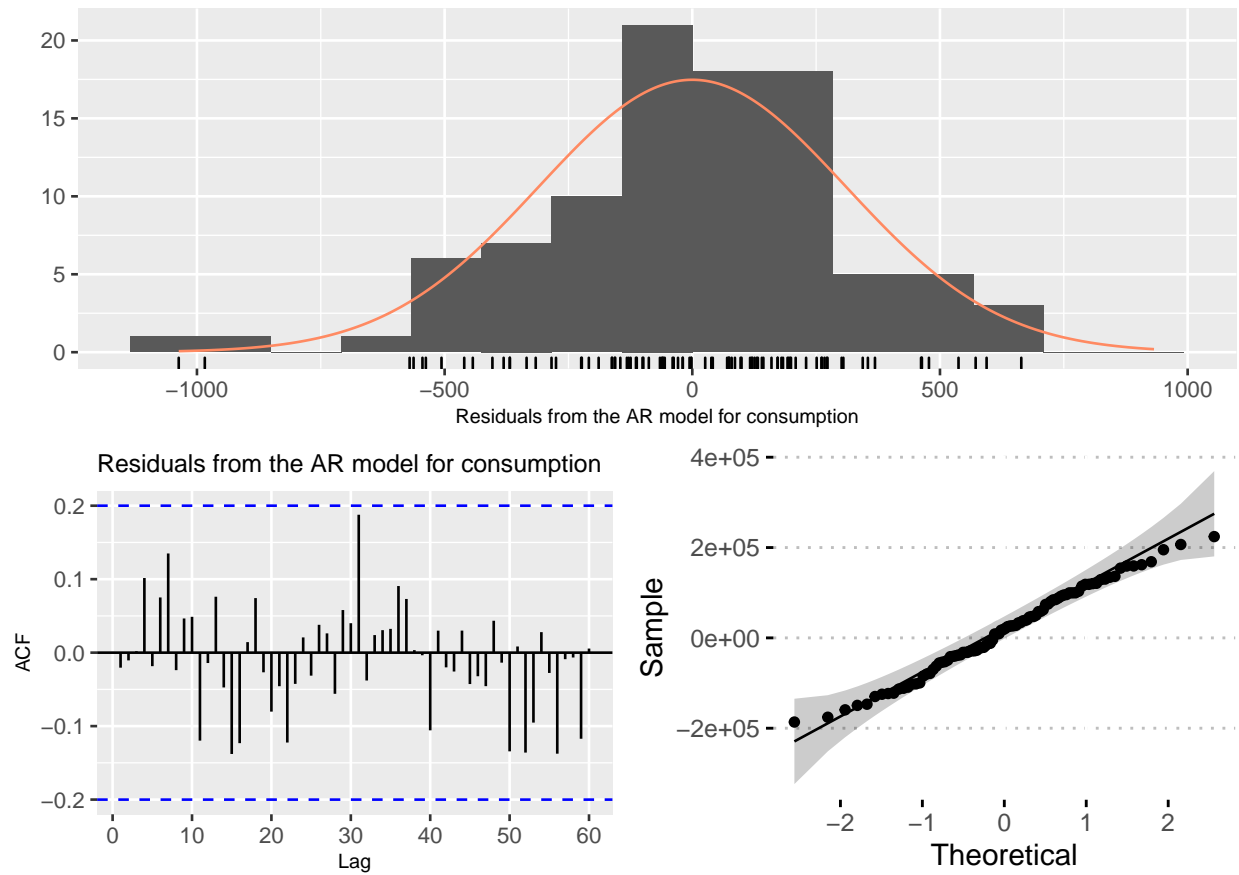
As in the previous exercises, we will use the `Arima()` function from the forecast package, the change is that we are doing the combinations for the differenced series.

```
## [1] "Best AR model for Consumption"
```

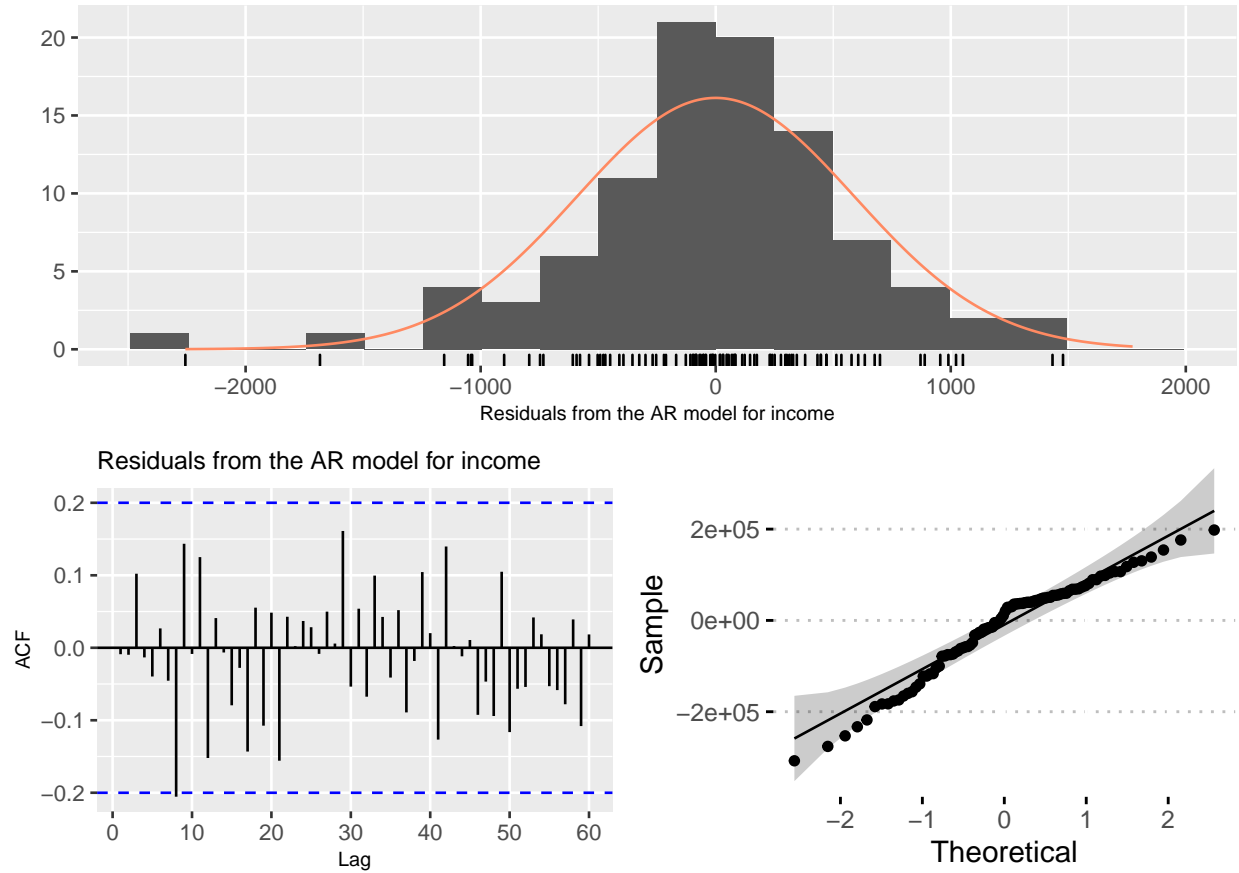
Variable	Estimate	Std. Error	t-statistic	P-Value
Lag 2	0.18	0.09	2.02	0.04
Lag 3	0.43	0.09	4.79	0.00
Mean	323.38	79.08	4.09	0.00

## [1] "Best AR model for Income"

Variable	Estimate	Std. Error	t-statistic	P-Value
Lag 1	0.51	0.09	5.81	0
Mean	526.11	120.18	4.38	0







**Comments:** Different from question 2, the residuals of the income series now fall better within the confidence bands of the qqplot (assuming a gaussian distribution, except for some points at the right tail of the residuals from income). Also, we see from the ACF and PACF that we don't have significant lags, except for the 8th lag of the income residuals. Given that these diagnostics are not with 100% confidence, it might be only due to a type I error.

As for the ADF test, now we are able to reject the null hypothesis that the series of differentiated income is explosive, using a confidence level of 95% (but not at 99% - however the confidence level should rely also on the problem at hand, and in this case, we considered adequate to use 95% given that the theory supports the idea that consumption and income indeed have a non-spurious relation). This was not the case for the first difference of consumption. We did run again the models (which includes running the models manually instead using the function we created to test different lag specifications), but with no changes. We then tried to use the `auto.arima()` function, which fits automatically an ARIMA(p,d,q) model to a series. To our surprise, the consumption series in level has  $d = 2$  and consequently the first difference has a better adjustment when  $d = 1$ . This suggests that the consumption series is  $I(2)$  and not  $I(1)$  as we previously thought. However, since the next questions ask us to assume that both are  $I(1)$ , we considered that this was not going to prevent us to solve the remaining problems.

```
## [1] "Results from the ADF test"
```

Variable	p of AR(p)	Test-statistic	p-value
Consumption	3	-2.2080	0.4905
Income	1	-4.7048	0.0100

## Question 4

Assuming both series are  $I(1)$ , test for cointegration between consumption and income by re-gressing consumption on income and performing a unit-root test on the residuals. Report the estimated regression coefficients. Plot the regression residuals. Use the Schwartz Information Criterion (SIC) to determine the number of ADF lags in your unit-root residual test. Report the cointegration test statistic. Do you reject cointegration?

For this question, we will follow the steps described in Lecture 7, slide 22:

1. Run a static regression of Consumption ( $Y_t$ ) on Income ( $X_t$ )

$$Cons_t = \delta + \lambda Inc_t + Z_t$$

The option for the income as the exogenous variable is because the households are constrained by their available income (i.e., income will determine how much you can purchase in goods and services) and not the otherway around - it is not reasonable to expect that our wage will be higher just because we want to buy a new laptop;

2. Obtain the estimates for  $\hat{\delta}$  and  $\hat{\lambda}$ ;
3. Obtain the estimated residuals

$$\hat{Z}_t = Cons_t - \hat{\delta} - \hat{\lambda} Inc_t;$$

4. Test if the series of the residuals has a unit root, i.e., test whether  $\hat{Z}_t \sim I(0)$  using the appropriate critical values.

Given that our function to test combinations of lags in a AR(p) model does not work with differenced series and we would have some time constraints to make the proper adaptations, to perform step 4 we tested AR(p) until  $p = 10$  (without the intermediate combinations of lags) and computed the BIC of those to determine how many lags should be included in the ADF test.

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	6783.37	554.99	12.22	0
Income	0.67	0.01	120.19	0

**Comments:** We started with  $\Delta \hat{Z}_t = \beta \hat{Z}_{t-1} + \phi_1^* \Delta \hat{Z}_{t-1} + \dots + \phi_{10}^* \Delta \hat{Z}_{t-10}$  which gives us a BIC of 1239.2479 and sequentially removed  $\phi_i^*$  for  $i$  ranging from 10 to 1 and we ended up having the lowest BIC with the largest model. We used the updated version of MacKinnon (2010) (available at: [http://qed.econ.queensu.ca/working\\_papers/papers/qed\\_wp\\_1227.pdf](http://qed.econ.queensu.ca/working_papers/papers/qed_wp_1227.pdf)) for the critical values (in particular, we used the second line of table 2 for the computation of the critical value). It is worth noticing that we followed the convention from the slides and did not include a constant terms in the regressions, although MacKinnon (2010) advises against this. The critical value computed was considering a sample size of 87 and a significance level of 5%. The critical value is equal to -1.9445, while the test statistics is equal to -4.6865. Since TRUE, we reject the null hypothesis that  $\hat{Z}_t \sim I(1)$ , which is equivalent to reject the hypothesis of no-cointegration between consumption and income.

## Question 5

Estimate an error correction model for consumption using the estimated residuals from the cointegration regression above. Use a general-to-specific modeling approach for the short-run dynamics. Report the estimated model. Report and interpret the short-run and long-run multipliers. Report and interpret the error correction coefficient.

A motivation for using the error correction model for the relationship between consumption and income is that, as we found before, the series for consumption and income are cointegrated. Hence, according to

Granger's Representation Theorem, they admit an ECM representation of the form

$$\Delta Y_t = \alpha + \gamma[Y_{t-1} - \delta - \lambda X_{t-1}] + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \sum_{i=0}^q \beta_i \Delta X_{t-i} + \varepsilon_t,$$

where we have not only a regression in first differences and white noise error, but also an error correction term that deals with the long term relationship between consumption and income. In our case,  $Y_t$  represents aggregate consumption and  $X_t$  represents aggregate income.

The model estimation is inline with the Engle and Granger 2-step procedure, where we first estimate a static cointegration regression of consumption on income, retrieve the residuals and then we estimate an ECM that has the residuals from the first regression as one of the regressors:

$$\begin{aligned} \Delta CONS_t &= \alpha + \gamma[CONS_{t-1} - \delta - \lambda INC_{t-1}] + \sum_{j=1}^p \phi_j \Delta CONS_{t-j} + \sum_{i=0}^q \beta_i \Delta INC_{t-i} + \varepsilon_t \\ &= \alpha + \gamma Res_{t-1} + \sum_{j=1}^p \phi_j \Delta CONS_{t-j} + \sum_{i=0}^q \beta_i \Delta INC_{t-i} + \varepsilon_t \end{aligned}$$

The first differences gives us the short run dynamics of the relationship between consumption and income. The long-run component is given by the residual (error correction term).

First, we regress consumption on income and retrieve the residuals, in the same way we did in the previous item. Next, we find the appropriate specification for the error correction model. To do so, we use the general-to-specific approach for the short-run components. We start with an error correction model with four lags for the short run components, and then proceed to sequentially removing insignificant lags, using as threshold a 5% significance level and removing first the lags who present a higher p-value.

*ECM with 4 lags for both  $\Delta INC$  and  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	124.65	73.67	1.69	0.09
Res(-1)	0.02	0.06	0.36	0.72
D_CONS(-1)	-0.07	0.11	-0.64	0.52
D_CONS(-2)	0.14	0.12	1.23	0.22
D_CONS(-3)	0.40	0.11	3.45	0.00
D_CONS(-4)	0.15	0.11	1.35	0.18
D_INC	0.10	0.06	1.56	0.12
D_INC(-1)	0.12	0.06	1.85	0.07
D_INC(-2)	-0.15	0.06	-2.40	0.02
D_INC(-3)	0.05	0.07	0.75	0.46
D_INC(-4)	-0.08	0.06	-1.25	0.21

*ECM with 4 lags for  $\Delta INC$  and lags 2, 3 and 4 for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	115.51	72.02	1.60	0.11
Res(-1)	0.02	0.06	0.36	0.72
D_CONS(-2)	0.15	0.12	1.25	0.21
D_CONS(-3)	0.39	0.11	3.40	0.00
D_CONS(-4)	0.13	0.11	1.21	0.23
D_INC	0.10	0.06	1.56	0.12
D_INC(-1)	0.11	0.06	1.78	0.08
D_INC(-2)	-0.16	0.06	-2.59	0.01
D_INC(-3)	0.06	0.06	0.93	0.35
D_INC(-4)	-0.08	0.06	-1.31	0.20

*ECM with lags 0, 1, 2 and 4 for  $\Delta INC$  and lags 2, 3 and 4 for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	97.96	69.47	1.41	0.16
Res(-1)	0.04	0.06	0.71	0.48
D_CONS(-2)	0.18	0.11	1.67	0.10
D_CONS(-3)	0.42	0.11	4.01	0.00
D_CONS(-4)	0.14	0.11	1.32	0.19
D_INC	0.11	0.06	1.72	0.09
D_INC(-1)	0.12	0.06	1.86	0.07
D_INC(-2)	-0.15	0.06	-2.46	0.02
D_INC(-4)	-0.07	0.06	-1.14	0.26

*ECM with lags 0, 1, 2 for  $\Delta INC$  and lags 2, 3 and 4 for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	114.49	68.05	1.68	0.10
Res(-1)	0.01	0.05	0.22	0.82
D_CONS(-2)	0.17	0.11	1.53	0.13
D_CONS(-3)	0.37	0.10	3.90	0.00
D_CONS(-4)	0.08	0.09	0.86	0.39
D_INC	0.10	0.06	1.64	0.10
D_INC(-1)	0.11	0.06	1.71	0.09
D_INC(-2)	-0.15	0.06	-2.42	0.02

*ECM with lags 0, 1, 2 for  $\Delta INC$  and lags 2 and 3 for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	120.17	61.99	1.94	0.06
Res(-1)	0.01	0.05	0.23	0.82
D_CONS(-2)	0.18	0.11	1.73	0.09
D_CONS(-3)	0.37	0.09	3.95	0.00
D_INC	0.10	0.06	1.69	0.10
D_INC(-1)	0.12	0.06	1.96	0.05
D_INC(-2)	-0.14	0.06	-2.34	0.02

*ECM with lags 1 and 2 for  $\Delta INC$  and lags 2 and 3 for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	157.72	58.46	2.70	0.01
Res(-1)	-0.03	0.04	-0.58	0.56
D_CONS(-2)	0.17	0.11	1.55	0.13
D_CONS(-3)	0.39	0.10	4.11	0.00
D_INC(-1)	0.14	0.06	2.35	0.02
D_INC(-2)	-0.13	0.06	-2.11	0.04

*ECM with lags 1 and 2 for  $\Delta INC$  and third lag for  $\Delta CONS$*

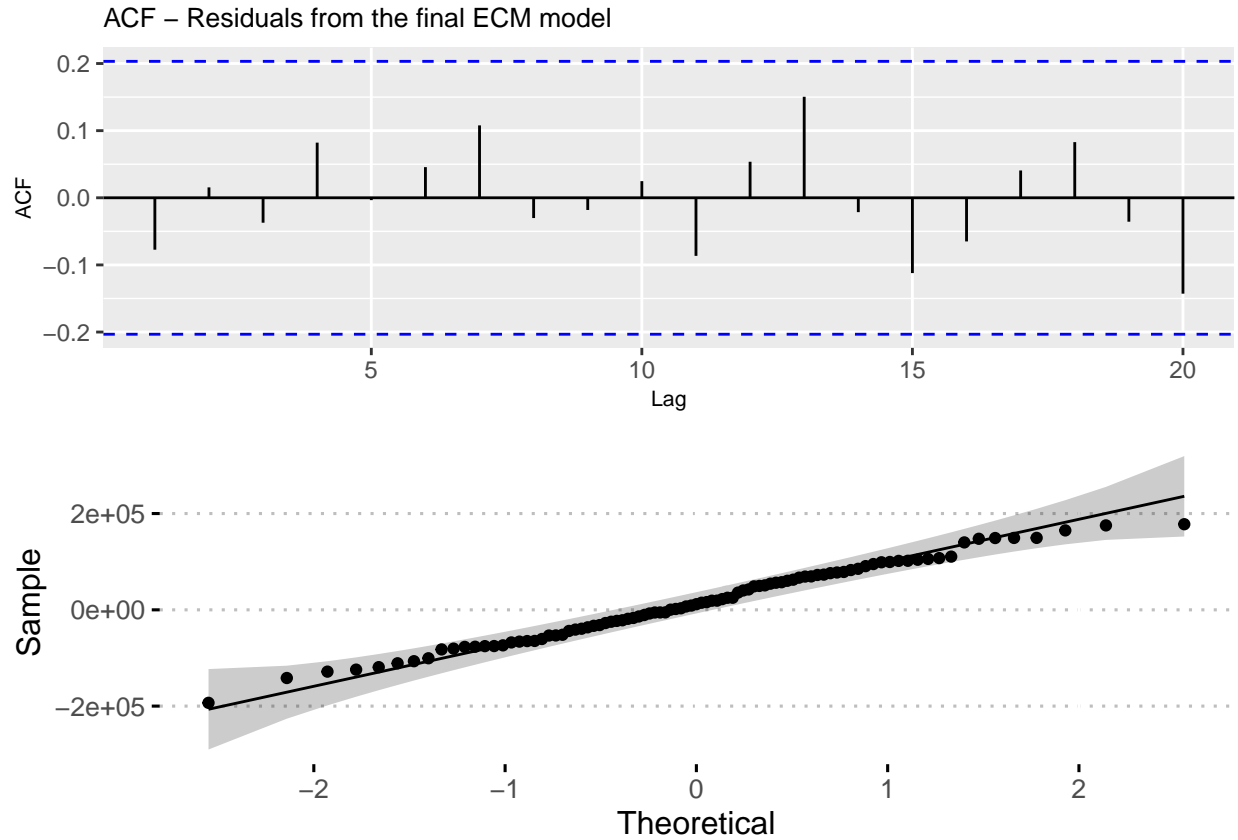
Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	205.19	50.16	4.09	0.00
Res(-1)	-0.06	0.04	-1.44	0.15
D_CONS(-3)	0.37	0.10	3.93	0.00
D_INC(-1)	0.13	0.06	2.12	0.04
D_INC(-2)	-0.08	0.05	-1.56	0.12

*ECM with first lag for  $\Delta INC$  and third lag for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	198.71	50.38	3.94	0.00
Res(-1)	-0.07	0.04	-1.83	0.07
D_CONS(-3)	0.34	0.09	3.64	0.00
D_INC(-1)	0.08	0.05	1.52	0.13

*ECM with third lag for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	235.03	44.70	5.26	0.00
Res(-1)	-0.10	0.04	-2.59	0.01
D_CONS(-3)	0.36	0.09	3.82	0.00



The final model is

$$\Delta CONS_t = 235.03 - 0.10Res_{t-1} + 0.36\Delta CONS_{t-3} + \varepsilon_t$$

The residual analysis show that there is no significant residual autocorrelation at the 5% significance level. Also, the QQ-plot show that, except for the some points at the right part, the residuals fit a normal distribution around zero. Overall, we believe the final model has a good fit to the data and the residuals can be taken as being a result from a white noise process. Now, we proceed with the discussion about the multipliers and coefficients.

Since we do not have the current first difference for aggregate income in the final model (and hence no current income affecting current consumption), the short-run multiplier is zero. That is, current changes in income do not affect changes in current consumption, according to our model. This is supported by the idea that households smooth consumption and try to protect themselves from temporary income fluctuations (unemployment, recessions, end of the year bonuses, etc) - so in the short run, we do not observe an effect of income playing a role in the consumption path. This is a result of households transferring surpluses in income from “good” periods to the “bad” periods, resulting in overall consumption not changing, diminishing uncertainty and providing more predictability. Both Ando and Modigliani (1963) and Friedman (1957) theories supports this smoothing in consumption behavior.

To find the long-run multiplier, we need the equilibrium coefficients  $\delta$  and  $\lambda$ . From the initial regression of consumption on income, we have that  $\hat{\delta} = 6783.67$  and  $\hat{\lambda} = 0.67$ . The long-run multiplier will be given by  $\hat{\delta}$ . To see this, let us find the long-run multiplier as follows:

$$\begin{aligned}\Delta \overline{\text{CONS}} &= 235.03 - 0.10[\overline{\text{CONS}} - 6783.37 - 0.67 \overline{\text{INC}}] + 0.36\Delta \overline{\text{CONS}} \\ 0 &= 235.03 - 0.10[\overline{\text{CONS}} - 6783.37 - 0.67 \overline{\text{INC}}] + 0 \\ \overline{\text{CONS}} &= \frac{235.03}{0.10} + 6783.37 + 0.67 \overline{\text{INC}}\end{aligned}$$

Therefore, the long-run multiplier is 0.67, same as the value for  $\hat{\delta}$ . This value gives us the impact over aggregate consumption resulting from a permanente change in aggregate income.

The error correction coefficient is  $-0.10$  (p-value = 0.01), i.e. the coefficient of the residuals in of the regression of consumption on income. It tells us the adjustment speed of consumption with respect to the long-run equilibrium. If this coefficient is negative and close to zero (as it is our case, using a 5% significance level), consumption shows a partial adjustment towards the long-run equilibrium. Note that, the fact that the error correction coefficient is negative provide us evidence that there indeed are “corrections” for deviations from the long run equilibrium: if the residual was positive in the previous period (aggregate consumption yesterday is larger than  $\delta + \lambda \text{INC}_{t-1}$ , then we expect that consumption today falls a little bit in order to adjust itself in view of this “overconsumption” in the previous period. How much it falls (adjusts) is given by  $\hat{\gamma} = -0.10$ .

## Question 6

*How strong is the correction to equilibrium? Is there over-shooting? Do you find evidence of Granger causality? Justify your answer.*

According to the estimate obtained for  $\gamma$  above, we have that the correction to equilibrium is not very large. As discusse above, the estimate being negative shows us that there is indeed a correction to equilibrium. However, since the estimative is between 0 and -1, we have a *partial adjustment*, not an overshooting, to deviations from long run equilibrium.

As discussed in the lectures, the over-shooting effect is related to the coefficient of the ECM model that appears in front of the lagged residuals from the level equation. If this coefficient is negative and higher than 1 in absolute value, then the adjustment is higher than the the initial change (over-shooting of consumption). However, as pointed out in the previous question, this is not the case for the estimated model and we observe small adjusts, or partial adjustments, in consumption over time in response to deviations from the long run equilibrium values.

As for Granger causality in this ECM model, since we do not have a significant (considering a 95% confidence level) coefficient of the type  $\Delta INC_{t-j}$ , for  $j \geq 1$  in the model, we cannot infer that changes (the  $\Delta$ 's, i.e., the differenced terms) of income in the past will Granger cause changes in consumption today.

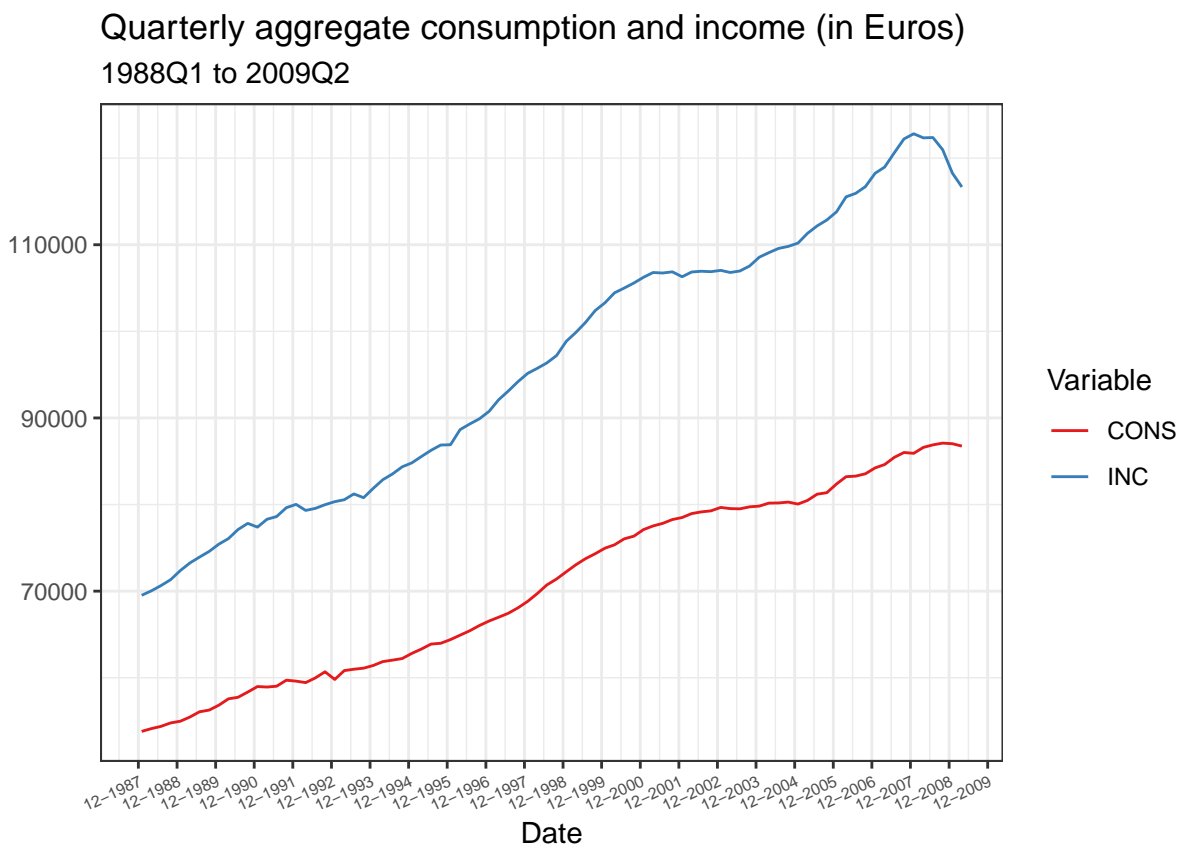
## Question 7

*At the peak of the recession, during the 2nd quarter of 2009, you are asked to forecast the value of consumption for the 3rd quarter of 2009. Do you expect aggregate consumption to raise or fall? Report the predicted change in the value of consumption. How much of this change in consumption is due to the 'correction mechanism' alone? Report your point forecast for consumption for the 3rd quarter of 2009.*

Here, we need to estimate a model and perform a forecast, but only considering the data until 2Q2009. We start by repeating the procedure in question 5 in order to find the appropriate ECM model.

First, we created a new data set with data only until 2Q2009.

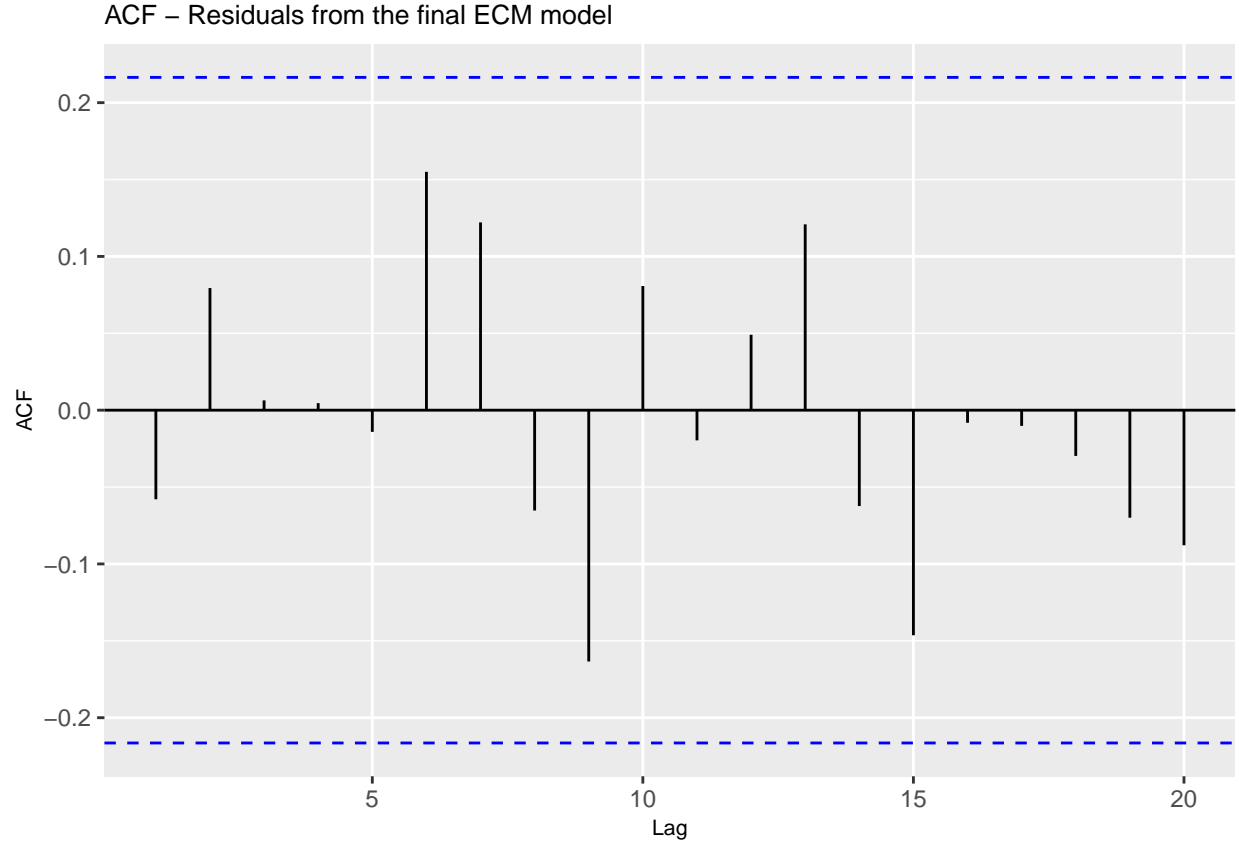
```
dfData03 <- subset(dfData01, TIME <= "2009-04-01")
tsData03 <- xts(dfData03[,3:4], order.by = as.yearqtr(dfData03[,1]), frequency = 4)
```



Next, we make the regression of consumption on income and retrieve the residuals so we can use them in the ECM model estimation. Finally, we proceed with applying the general do specific approach to select the lags for the short-term components. We again start with 4 lags for  $\Delta INC$  and  $\Delta CONS$ , and sequentially eliminate statistically insignificant lags (5% significance level). Below, we present only the estimates for the final model. The intermediate models estimated can be found in the Appendix.

*ECM with 2 lags for  $\Delta INC$  and third lag for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	223.21	56.82	3.93	0.00
Res(-1)	-0.04	0.05	-0.90	0.37
D_CONS(-3)	0.40	0.10	3.90	0.00
D_INC(-1)	0.16	0.07	2.36	0.02
D_INC(-2)	-0.14	0.05	-2.52	0.01



The final model is

$$\Delta CONS_t = 223.21 - 0.04Res_{t-1} + 0.40\Delta CONS_{t-3} + 0.16\Delta INC_{t-1} - 0.14\Delta INC_{t-2} + \epsilon_t$$

$$\Delta CONS_t = 223.21 - 0.04[CONS_{t-1} - \hat{\delta} - \hat{\lambda}INC_{t-1}] + 0.40\Delta CONS_{t-3} + 0.16\Delta INC_{t-1} - 0.14\Delta INC_{t-2} + \epsilon_t$$

$$\Delta CONS_t = 223.21 - 0.04[CONS_{t-1} - 7486.44 - 0.66INC_{t-1}] + 0.40\Delta CONS_{t-3} + 0.16\Delta INC_{t-1} - 0.14\Delta INC_{t-2} + \epsilon_t$$

The residual analysis of the results show no significant residual autocorrelation (5% significance level), which offers evidence towards white noise innovations in the model.

- *Do you expect aggregate consumption to raise or fall?*

From the error correction coefficient estimate, we expect, in the next quarter, a small decrease on consumption due to adjustments to equilibrium. Note however, that such decrease might be very small and not relevant, since the error correction coefficient is insignificant even for large level of significance since the p-value is above 30%. We believe that any changes in consumption would be more related to past changes in consumption than to adjustments to an equilibrium.



- *Report the predicted change in the value of consumption.*

Now, we proceed with finding a one-step ahead forecast for change in consumption. Since the final model only consider past values of variables, the predicted change in consumption for the third quarter of 2009 can be computed without the need to assign some dynamics for aggregate income. The predicted change in consumption, given all the information until period  $T$  ( $D_T$ ), will be:

$$\begin{aligned}\Delta\hat{CONS}_{T+1} &= E[\Delta CONS_{T+1}|D_T] \\ &= E[\alpha + \gamma[CONS_T - \delta - \lambda INC_T] + \phi_3 \Delta CONS_{T-2} + \beta_1 \Delta_T \\ &\quad + \beta_2 \Delta_{T-1} + \epsilon_{T+1}|D_T] \\ &= \alpha + \gamma Res_T + \phi_3 \Delta CONS_{T-2} + \beta_1 \Delta_T + \beta_2 \Delta_{T-1}\end{aligned}$$

The observed values of the above variables are summarized bellow. Note that  $T + 1$  represents 3Q2009. Thus,  $T$  is 2Q2009,  $T - 1$  is 1Q2009 and  $T - 2$  is 4Q2008.

	Observed value
Res 2Q2009	2627.944
D_CONS 4Q2008	204.200
D_INC 2Q2009	-1600.000
D_INC 1Q2009	-2706.400
CONS 2Q2009	86748.800

The resulting predicted consumption will then be:

$$\begin{aligned}\Delta\hat{CONS}_{T+1} &= 223.21 - 0.04 \cdot 2627.944 + 0.40 \cdot 204.2 + 0.16 \cdot (-1600.0) - 0.14 \cdot (-2706.4) \\ &= 322.67.\end{aligned}$$

- *How much of this change in consumption is due to the ‘correction mechanism’ alone?*

The amount of the above result that is due to the correction mechanism is:

$$-0.04 \cdot 2627.944 = -105.12,$$

which can be seen as a small change relative to the level of aggregate consumption. The relative small correction is due to the fact that the error coefficient is statistically insignificant at 5% or 10% (or even higher) significance levels. Thus, from our final model, we do not expected correction mechanisms to affect changes over consumption in a relevant manner.

- *Report your point forecast for consumption for the 3rd quarter of 2009.*

Since we have the value of consumption in 2Q2009 and the predicted change in 3Q2009, we can retrieve the consumption in 3Q2009. Hence, our forecast for  $CONS_{T+1}$  will be:

$$CONS_T + \Delta\hat{CONS}_{T+1} = CONS_{T+1} \Rightarrow 86748.8 + 322.67 = 87071.47.$$

Overall, with data up until the recession peak, our model predicts a (relatively) small increase over consumption. This result seems to be in line with the somewhat flat behaviour observed after the crisis. As a matter of fact, our forecast for consumption at 3Q2009 seems to be close to the observed consumption (87121.6), at least in level terms. A more complete analysis to understand how close our estimate is would require confidence intervals.

## Appendix to question 7 - intermediate models

*ECM with four lags for  $\Delta INC$  and  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	106.28	97.41	1.09	0.28
Res(-1)	0.04	0.07	0.59	0.56
D_CONS(-1)	-0.07	0.12	-0.62	0.54
D_CONS(-2)	0.13	0.13	1.00	0.32
D_CONS(-3)	0.40	0.13	3.15	0.00
D_CONS(-4)	0.10	0.13	0.83	0.41
D_INC	0.14	0.08	1.74	0.09
D_INC(-1)	0.18	0.08	2.20	0.03
D_INC(-2)	-0.18	0.07	-2.50	0.01
D_INC(-3)	0.02	0.07	0.27	0.79
D_INC(-4)	-0.04	0.06	-0.68	0.50

*ECM with lags 0, 1, 2, and 4 for  $\Delta INC$  and 4 lags for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	99.85	93.72	1.07	0.29
Res(-1)	0.05	0.06	0.71	0.48
D_CONS(-1)	-0.09	0.11	-0.77	0.45
D_CONS(-2)	0.14	0.12	1.15	0.25
D_CONS(-3)	0.41	0.12	3.41	0.00
D_CONS(-4)	0.11	0.12	0.94	0.35
D_INC	0.15	0.08	1.88	0.06
D_INC(-1)	0.18	0.08	2.30	0.02
D_INC(-2)	-0.18	0.07	-2.51	0.01
D_INC(-4)	-0.04	0.06	-0.64	0.52

*ECM with 2 lags for  $\Delta INC$  and 4 lags for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	118.983	88.531	1.344	0.183
Res(-1)	0.026	0.056	0.462	0.645
D_CONS(-1)	-0.083	0.111	-0.742	0.461
D_CONS(-2)	0.130	0.117	1.107	0.272
D_CONS(-3)	0.376	0.109	3.460	0.001
D_CONS(-4)	0.081	0.109	0.739	0.463
D_INC	0.142	0.077	1.848	0.069
D_INC(-1)	0.169	0.076	2.217	0.030
D_INC(-2)	-0.184	0.069	-2.652	0.010

*ECM with 2 lags for  $\Delta INC$  and 3 lags for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	133.40	83.65	1.59	0.12
Res(-1)	0.02	0.06	0.41	0.68
D_CONS(-1)	-0.06	0.10	-0.54	0.59
D_CONS(-2)	0.13	0.12	1.12	0.27
D_CONS(-3)	0.37	0.11	3.44	0.00
D_INC	0.14	0.08	1.88	0.06
D_INC(-1)	0.18	0.07	2.57	0.01
D_INC(-2)	-0.18	0.06	-2.93	0.00

*ECM with 2 lags for  $\Delta INC$  and lags 2 and 3 for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	116.29	76.99	1.51	0.14
Res(-1)	0.03	0.05	0.47	0.64
D_CONS(-2)	0.14	0.11	1.22	0.23
D_CONS(-3)	0.37	0.11	3.51	0.00
D_INC	0.14	0.08	1.85	0.07
D_INC(-1)	0.17	0.07	2.53	0.01
D_INC(-2)	-0.19	0.06	-3.11	0.00

*ECM with 2 lags for  $\Delta INC$  and third lag for  $\Delta CONS$*

Variable	Estimate	Std. Error	t-statistic	P-Value
Intercept	169.17	63.72	2.65	0.01
Res(-1)	0.00	0.05	-0.08	0.93
D_CONS(-3)	0.35	0.10	3.33	0.00
D_INC	0.13	0.08	1.78	0.08
D_INC(-1)	0.16	0.07	2.41	0.02
D_INC(-2)	-0.16	0.05	-2.85	0.01

## References

- Ando, Albert, and Franco Modigliani. 1963. "The" Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests." *The American Economic Review* 53 (1): 55–84.
- Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Friedman, Milton. 1957. "The Permanent Income Hypothesis." In *A Theory of the Consumption Function*, 20–37. Princeton University Press.
- MacKinnon, James G. 2010. "Critical Values for Cointegration Tests." Queen's Economics Department Working Paper.