Group Assignment

TI - Econometrics III

Block 4, 2021

Siem Jan Koopman (coordinator and teacher)

Francisco Blasques (course developer)

Quint Wiersma (teacher)

Yasmine van der Straten (teaching assistant)

Quick notes and instructions:

1. This assignment is mandatory.
2. The assignment is to be made in groups of two students. No more, no less.
3. The assignment has five parts. Each part must be delivered within the following deadlines:

   Parts 1 and 2: (week 4) Friday, March 26, at 23:59

   Parts 3, 4, and 5: (week 6) Friday, April 9, at 23:59
4. The assignment must be uploaded in pdf format to the Canvas before the deadline mentioned above.
5. The first page of the assignment must state the names and student numbers of all the group members.
6. The answers should be correct, clear and complete. Deliver a report with appropriate justifications and insightful comments and remarks. Please consider all the theory that you know in analyzing the empirical results. Think about your report: Too much information is unreadable. Too little information is ill advised.

General Information

This assignment is mandatory. The assignment is to be made in groups of four students. No more, no less. Each part of the assignment must be uploaded to the blackboard, in pdf format, within the deadlines mentioned in page 1. There will be no tolerance period for late deliveries as the upload link will be automatically closed at 23:59. Deliveries after the assigned deadline imply that you obtain zero points for that part of the assignment.

The assignment requires you to program using any sort of programming language or the use of some software package. Those who are familiar with a programming language (e.g. Python, R, MATLAB, OCTAVE, OxMetrics, Gauss, C++, Julia, Fortran, etc.) are encouraged to use it in the assignment. Those who are not familiar with a programing language are advised to use one of the following software packages: EViews (proprietary), gretl (free open-source equivalent to EViews (`http://gretl.sourceforge.net`)), or STATA (proprietary but made available through the VU[1]). All parts of the assignment make use of the time series data contained in the csv files which will be made available through *Canvas*.

The assignment is composed of four parts. The data required for solving the assignments is available in four separate csv files called `data_assign_X.csv`, where `X` = `p1`,`p2`,`p3`,`p4`,`p5`.

Not all questions can be solved by the computer. On the contrary, some questions must be solved 'with pen and paper'.

When you find a question unclear or impossible to answer, then explain your thoughts. Explain why you find the question unclear and how it should be formulated instead.

Finally, please note that you may have to make some decisions. Which confidence level should I adopt in a certain question? Which statistical test or criteria should I take into consideration to give my final answer? If two tests disagree, then what should I conclude? These questions are part of the daily life of econometricians. Sometimes there is no right or wrong answer. The only thing that really matters, is that you explain carefully and clearly the problems you face, and you justify your decisions convincingly.

[1]For Windows and Linux users STATA can be dowloaded through `https://download.vu.nl`. Mac users can download STATA via the VU Software Center (`https://munki.mac.vu.nl/installers/`).

PART I: An autogregressive Model for Dutch GDP

Let us go back in time to the first quarter of 2009. The world economy has just been hit by a major financial crisis. In just one year, the Dutch quarterly GDP growth rate has fallen from 1.4%, in the first quarter of 2008, to -2.7%, in the first quarter of 2009.

In the first quarter of 2009, at the peak of the economic recession, suppose that government officials ask you to describe the dynamics of the Dutch GDP quarterly growth rate and deliver a forecast for the two years ahead. The available sample of observed GDP growth rates spans from the second quarter of 1987 to the first quarter of 2009. You can find this data set in the csv file labeled `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 tells you about the dynamic properties of GDP quarterly growth rates?

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?

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?

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.

5. Suppose that the innovations in your AR model are iid Gaussian. Produce 95% confidence intervals for your 2-year forecast.

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

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.

Time	08q3	08q4	09q1	09q2	09q3	09q4	10q1	10q2	10q3	10q4	11q1
GDP qgr	0.49	-0.31	-2.70	-1.63	0.28	0.33	0.66	1.59	0.51	0.71	0.81

Part II: An ADL Model for Okun's Law

By the beginning of 2014, in the face of fast rising unemployment, the Dutch Ministry of Social Affairs is concerned that the provisions for future government expenditure with social pensions may be severely under-estimated. This is particularly true if the economy is hit again by a large negative shock. Suppose that you have been asked to analyze alternative unemployment scenarios.

The available sample of observed unemployment rates and GDP growth rates spans from the second quarter of 1987 to the first quarter of 2014. You can find this data set in the csv file labeled `data_assign_p2.csv`.

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?

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.

3. Please provide a detailed comment on the following statement:

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

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?

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.

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.

PART III: Stock Market Unit-Roots and Spurious Regression

Suppose that you are asked to analyze investment opportunities in the stock market by studying the dynamic behavior of 10 major stock prices. In particular, you are asked to study the stocks of Apple, Intel, Microsoft, Ford, General Electrics, Netflix, Nokia, Exxon Mobil, and Yahoo, as well as the S&P500 stock market index.

You can find the entire sample of daily data at your disposal in the csv file labeled `data_assign_p3.csv`. This data set contains the daily stock prices of all the companies mentioned above and spans from the 14th of February 2007 to the 28th of January of 2013.

1. Provide plots for two stock market time-series at your choice and report 12-period ACF and PACF functions for those two time series. What does the sample ACF tell you about the dynamic properties of these stocks?

2. Perform an ADF unit-root test for all the 10 time series using the general-to-specific approach based on the Schwarz Information Criterion (SIC). Report the values of the ADF test statistics. Is the unit-root hypothesis rejected for any time-series at the 90% confidence level? Did you expect to reject the unit-root hypothesis for some time-series at this confidence level? Justify your answer carefully.

3. Assume that both the stocks of Apple and Microsoft follow a random walk process. Produce a 5-day forecast for the stocks of Apple and Microsoft. Add 95% confidence bounds to your forecasts under the assumption of Gaussian innovations. Is there any investment advice you can give on these stocks? Is their value expected to increase or decrease?

4. Please investigate the following claim:

   > *"Financial analysts have found that changes in the price of Microsoft stocks can be largely explained by fluctuations in the market value of Exxon Mobile. According to these analysts, this shows the extent to which the Microsoft Corporation is currently exposed to the market performance of the oil and gas industry."*

   Do you find a statistically significant contemporaneous relation between Microsoft and Exxon Mobile stock prices? Do you agree that changes in Microsoft stock prices are largely explained by fluctuations in the stock price of Exxon Mobile? Justify your answer.

PART IV: Cointegration in the Consumption Function

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!

The sample of quarterly aggregate consumption in The Netherlands and the sample of aggregate household income that you have at your disposal spans from the first quarter of 1988 to the first quarter of 2012. You can find this data set in the csv file labeled `data_assign_p4.csv`.

1. Plot both the aggregate consumption and aggregate income time-series (these series are called *cons* and *inc* respectively in the csv file). Compute and report 12-period ACF and PACF functions for each series and comment on their shape.

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?

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?

4. Assuming both series are I(1), test for cointegration between *consumption* and *income* by regressing *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?

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.

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

PART V: Forecasting US COVID-19 Cases

The Centers for Disease Control and Prevention (CDC) provides on their webpage a COVID Data Tracker[2] in which some key statistics for the US as a whole, as well as for individual states, are reported. Besides reporting key statistics, the CDC also reports the forecasts for the number of weekly deaths and cases based on a large number of scientific models. However, these forecasts are mostly based on structural models.

In this assignment you are asked to provide forecasts for the number of new infected cases using statistical time series models. The available sample of confirmed number of new infected cases on a daily basis spans from the 22nd of January 2020 until the 25th of February 2021. You can find the dataset in the csv file labeled `data_assign_p5.csv`. Besides the number of newly infected cases the dataset also contains the 7-day moving average of the daily series.

Details of the assignment:

- Produce a 2 week forecast (14 days) for the number of newly infected cases. Note that you could also use the smoothed 7-day moving average data instead of the raw data.
- You are free to manipulate the data in any way you like. For example, the dataset contains a lot of zeros in the beginning, maybe you want to start using the data only when the number of newly infected cases become sufficiently high.
- Use the G2S approach for estimation of your model. In addition you can use subsample validation tecnhiques, as discussed in the additional material on model specification, in combination with MAE and/or RMSE to select your preferred forecasting model.
- Think about potential difficulties of the data, such as non-stationarity, and how to deal with those.
- Write up your results, as well as clarification on how/why you selected your preferred forecasting model, in a small report.

Good Luck!

[2] `https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/index.html`