Emprical Exercises - 4
Unit Roots, Cointegration and Error Correction

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due May 3)

Assignment: These empirical exercises deal with unit roots, cointegration and error correction. Exercise 1 asks you to carry out unit root test for the US time series; Exercise 2 takes you through the steps of an Engle-Granger analysis for the US aggregate consumption, while Exercise 3 guides you through a cointegration analysis based on a dynamic regression model.

Data: Download the dataset `timeSeriesUSA.csv` from the course webpage on NYU Home. The dataset contains the following quarterly time series for the US economy between 1990:1 - 2017:3:

EXNOM	Effective (nominal) exchange rate (1980=100).
CANOM	Current account (current prices).
INVBUSS	Business machinery investments (constant prices).
GDPREAL	Total GDP (constant prices).
BOND	Bond yield (pro anno yield in ratio).
HOUSE	House price index (2000=1).
CONNP	Consumption deflator (2000=1).
URATE	Unemployment rate (in ratio).
TT	Terms of trade, export price/import price (2000=1).

Exercise 1 (Unit Root)

[1] Choose one of the time series in the file `timeSeriesUSA.csv`. Take logs if you think it is appropriate and make a time series graph.

[2] Based on the graph and your knowledge of economic theory, what would you suggest as a deterministic specification for the Dickey-Fuller test? Is the relevant alternative stationarity or trend-stationarity?

[3] Set up an autoregressive model in R and find the preferred lag length. Is the model well specified?

[4] Perform the augmented Dickey-Fuller t-test. What is the conclusion regarding a unit root? How would you relate the finding to economic theory?

[5] Perform the LR test for the joint hypothesis involving also the relevant deterministic variable. What is the conclusion regarding a unit root?

Exercise 2 (Engle-Granger Analysis for the US Aggregate Consumption)
In this exercise we want to construct a single equation cointegration model for the US private consumption. We define the vector of variables

$$\mathbf{Z}_t = \begin{pmatrix} C_t \\ Y_t \\ W_t \\ \text{ELOSS}_t \end{pmatrix} = \begin{pmatrix} \text{CPRIV}_t \\ \text{DPRIV}_t \\ \text{REALW}_t \\ \text{ELOSS}_t \end{pmatrix}$$

where

CPRIV	Private sector aggregate consumption, constant prices.
CPRIVP	Deflator for private consumption, 1995=100.
DPRIV	Private disposable income, constant prices.
REALW	Private wealth including owner occupied housing, constant prices.
ELOSS	Expected income loss from changes in unemployment.

The variables are located in `consumptionUSA.csv`.

[1] Test for unit roots in each of the four variables in $\mathbf{Z}_t$. Make sure that you use a sensible deterministic specification.

What do your conclusions imply for the estimation of a cointegration relation for the variables in $\mathbf{Z}_t$?

[2] Use OLS to estimate the static long-run relation

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 W_t + u_t \tag{1}$$

and interpret the signs and the magnitudes of the coefficients. Do you think it is reasonable to use the output to test statistical hypothesis on the coefficients?

[3] Perform an Engle-Granger residual based test for whether (1) is a cointegrating relation. Explain the test and the outcome.

[4] Irrespective of the conclusions in question [3], define the error correction term, $ecm_t = \hat{u}_t$, corresponding to the relation in (1).

Construct a graph of ecm_t and try to explain why there is not strong support for cointegration.

[5] Construct single equation error correction models for $\Delta C_t, \Delta Y_t$, and ΔW_t, where you include also ELOSS_t as an explanatory variable, e.g.

$$\begin{aligned} \Delta C_t = {} & \alpha_0 + \alpha_1 \Delta C_{t-1} + \alpha_2 \Delta Y_t + \alpha_3 \Delta Y_{t-1} \\ & + \alpha_4 \Delta W_t + \alpha_5 \Delta W_{t-1} + \alpha_6 \text{ELOSS}_t + \alpha_7 ecm_{t-1} + \epsilon_t \end{aligned} \tag{2}$$

Does any of the variables seem to error correct? Is that in line with your expectations based on economic theory?

[6] Concentrate on the model for ΔC_t. Delete insignificant variables and reestimate the model (2). Perform the usual misspecification tests. Does the model appear well specified?

[7] Does any of the residuals are very large compared to the rest of the residuals and compared to the expected range for observations drawn from a standard normal distribution?

Suppose that the standardized residuals for 1975:4, 1977:3 and 1977:4 are very large. These observations might be related to some transitory policy measures. Discuss why it might be preferable to try to remove the effects of these observations. To remove the effects of the observations, we can construct dummy variables, i.e. variables of the form 0, ..., 0, 1, 0, ..., 0 with the value 1 located at the particular point in time. Construct dummy variables for the three special episodes, and include them in the model. What happened to the residuals of the observations with dummy variables, e.g. $\epsilon_{1975:4}$? How do the parameter estimates change? How do the misspecification tests change?

Now we introduce the concept of recursive estimation, which is very useful in analyzing the structural stability of an estimated model. Recursive estimation is done by estimating a model like (2) for increasing samples $t = 1973:2 - T_0$, where T_0 takes the values

$$1973:2 + N, \qquad 1973:2 + N + 1, \qquad 1973:2 + N + 2, \qquad \ldots 2016:4$$

That is, we first estimate the model with only the first N observations; and then we successively add a new observation and reestimate the model. For each sample we do an OLS estimation and obtain all the usual statistics. Afterwards we can consider the sample paths of the different statistics calculated for each sample. For example we can consider the estimated coefficients for the expanding samples

$$\alpha_i(T_0), \qquad \text{for} \quad T_0 = 1973:2 + N, \ldots, 2016:4$$

The model (2) is estimated under the assumption of constant coefficients, which implies that a graph for $\alpha_i(T_0)$, $i = 0, \ldots, 7$, should not fluctuate too much.

[8] Do a recursive estimation of your error correction model and look at the recursive parameter estimates. Does the model look stable?

Exercise 3 (ADL Analysis of the US Aggregate Consumption)
This exercise introduces cointegration analysis based on an unrestricted ADL or ECM model. That is an alternative to the Engle-Granger procedure discussed above.

[1] Explain why the estimates of a cointegration vector based on an unre- stricted dynamic model (ADL or ECM) are expected to be superior to the estimates derived from the Engle-Granger two-step approach.

[2] Estimate an unrestricted ADL model

$$\begin{aligned}\Delta C_t = \beta_0 &+ \beta_1 C_{t-1} + \beta_2 C_{t-2} + \beta_3 Y_t + \beta_4 Y_{t-1} + \beta_5 Y_{t-2} \\ &+ \beta_6 W_t + \beta_7 W_{t-1} + \beta_8 W_{t-2} + \beta_9 \text{ELOSS}_t \\ &+ \beta_{10} \text{Dum1}_t + \beta_{11} \text{Dum2}_t + \text{etc.} + \epsilon_t\end{aligned}$$

and delete the insignificant terms.

[3] Test the model for misspecifications.

[4] Derive the long-run solution and interpret the results.

[5] Perform the test for no-cointegration in R. What do you conclude concerning the cointegration properties?

[6] Derive and interpret the dynamic multipliers.

[7] Finally, estimate the unrestricted error correction model:

$$\begin{aligned}\Delta C_t = \alpha_0 &+ \alpha_1 \Delta C_{t-1} + \alpha_2 \Delta Y_t + \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta W_t + \alpha_5 \Delta W_{t-1} \\ &+ \alpha_6 C_{t-1} + \alpha_7 Y_{t-1} + \alpha_8 W_{t-1} + \alpha_9 \text{ELOSS}_t \\ &+ \alpha_{10} \text{Dum1}_t + \alpha_{11} \text{Dum2}_t + \text{etc.} + \epsilon_t\end{aligned}$$

Compare the estimates with the estimates from the ADL model.

Derive (manually) the long-run solution and compare with the ADL model. manually perform the R test for no-cointegration.

Formalities:

[1] You must hand in a report that (i) presents your graphical analysis, (ii) describes the econometric model, (iii) outlines the modeling progress (e.g., the approach you have taken, the alternative models you have tried, etc.), (iv) presents your preferred model including interpretation and statements on economic and statistical significance, and (v) discusses the potential weaknesses of the model.

[2] The report must be a maximum of 10 pages of pdf produced from an R markdown file. You must hand in your R markdown file and pdf files together. You can use the R markdown template that I have posted on course webpage. Please name your files as follows: `empEx4-surname(s).Rmd`

[3] If you prefer, you are allowed to work in groups of up to three persons (not necessarily in the same exercise class as yours). The requirements and assessment criteria are the same for assignments written by one, two, or three persons.