Erasmus School of Economics

MOOC Econometrics

Training Exercise 6.5

Notes:

- This exercise uses the datafile `TrainExer65` and requires a computer.
- The dataset `TrainExer65` is available on the website.
- The exercise can be split in two parts. Parts (a-d) consider the nature of trends and Granger causality, and parts (e-g) consider forecast models and out-of-sample forecast evaluation. If you wish, you can choose to make either parts (a-d) or parts (e-g). Of course, if you have the time, it is of interest to make all parts.

Questions

The datafile TrainExer65 contains yearly data on industrial production (IP) in the USA and on the Composite Leading Index (CLI), both in logarithms. The example in Lecture 6.5 considered monthly data, and now we consider yearly data, both for the two log-series and for their first differences that correspond to the yearly growth rates. The main question of interest is whether past values of the growth rate in CLI have predictive power for future growth rates of IP. We denote the yearly growth rates by GIP $= \Delta\log(\text{IP})$ and GCLI $= \Delta\log(\text{CLI})$.

Note: In all questions, use 1960-2002 as estimation and test sample, and use 2003-2007 as hold-out forecast evaluation sample. That is, the data for 2003-2007 are nowhere used to estimate models or to perform tests, and these data are only used for forecast comparison in parts (e-g).

(a) Make time series plots of log(IP) and log(CLI), and also of the yearly growth rates GIP and GCLI. What conclusions do you draw from these plots?

(b) (i) Perform the Augmented Dickey-Fuller (ADF) test for log(IP). In the ADF test equation, include (among others) a constant (α), a deterministic trend term (βt), and two lags of GIP $= \Delta\log(\text{IP})$. Report the coefficient of $\log(\text{IP}_{t-1})$ and its standard error and t-value, and draw your conclusion.

(ii) Perform a similar ADF test for log(CLI).

Note that the 5% critical value differs from the usual one, see lecture 6.3.

(c) Perform the two-step Engle-Granger test for cointegration of the time series log(IP) and log(CLI). The second-step regression is of the type $e_t = \alpha + \rho e_{t-1} + \beta_1 \Delta e_{t-1} + \beta_2 \Delta e_{t-2} + \omega_t$, where e_t are the residuals of step 1. What is your conclusion?
Note that the 5% critical value differs from the usual one, see lecture 6.3.

(d) Perform two F-tests, one for the Granger causality of GIP for GCLI and the other for the Granger causality of GCLI for GIP. Include a constant and two lags of both variables in the test equations. Report the degrees of freedom and the numerical values of the two F-tests, and draw your conclusion. The relevant 5% critical value is 3.3.

(e) Show that the coefficients of both lags in an AR(2) model for GIP are insignificant. Show also that even the slope coefficient in the AR(1) model $\text{GIP}_t = \alpha + \beta \text{GIP}_{t-1} + \varepsilon_t$ is insignificant. Make two forecasts for GIP for the five years from 2003-2007, one from the AR(1) model and another from the simple model $\text{GIP}_t = \alpha + \varepsilon_t$.

(f) Estimate the ADL(2,2) model $\text{GIP}_t = \alpha + \beta_1 \text{GIP}_{t-1} + \beta_2 \text{GIP}_{t-2} + \gamma_1 \text{GCLI}_{t-1} + \gamma_2 \text{GCLI}_{t-2} + \varepsilon_t$, and show by means of an F-test that the null hypothesis that $\beta_1 = \beta_2 = \gamma_2 = 0$ is not rejected. Then estimate the ADL(0,1) model $\text{GIP}_t = \alpha + \gamma \text{GCLI}_{t-1} + \varepsilon_t$ and use this model to forecast GIP for the five years from 2003-2007.

(g) Compare the three series of forecasts of parts (e) and (f) by computing their values of the root mean squared error (RMSE), mean absolute error (MAE), and the sum of the forecast errors (SUM). Check that it seems quite difficult to forecast the IP growth rates for 2003-2007 from models estimated from 1960-2002. Can you think of possible reasons why this is the case?