Erasmus School of Economics

MOOC Econometrics

Training Exercise 6.1

Notes:

- This exercise uses the datafile `TrainExer61` and requires a computer.
- The dataset `TrainExer61` is available on the website.

Questions

The datafile TrainExer61 contains values of four series of length 250. Two of these series are uncorrelated white noise series denoted by ε_{xt} and ε_{yt}, where both variables are NID(0, 1), that is, normally and independently distributed standard normal random variables. The other two series are so-called random walks constructed from these two white noise series by $x_t = x_{t-1} + \varepsilon_{xt}$ and $y_t = y_{t-1} + \varepsilon_{yt}$, with starting values $x_1 = 0$ and $y_1 = 0$.

As ε_{xt} and ε_{ys} are independent for all values of t and s, the same holds true for all values of x_t and y_s. The purpose of this exercise is to experience that, nonetheless, the regression of y on x indicates a highly significant relation between y and x if evaluated by standard regression tools. This kind of result is called 'spurious regression' and is caused by the trending nature of the variables x and y. The lesson we learn is that standard regression tools are not applicable if the variables contain trends similar to those of the random walks considered here.

(a) Use dataset TrainExer61 to make the following graphs: the time series plot of x_t against time t, the time series plot of y_t against time t, and the scatter plot of y_t against x_t. What conclusion could you draw from these three graphs?

(b) To check that the series ε_{xt} and ε_{yt} are uncorrelated, regress ε_{yt} on a constant and ε_{xt}. Report the t-value and p-value of the slope coefficient.

(c) Extend the analysis of part (b) by regressing ε_{yt} on a constant, ε_{xt}, and three lagged values of ε_{yt} and of ε_{xt}. Perform the F-test for the joint insignificance of the seven parameters of ε_{xt} and the three lags of ε_{xt} and ε_{yt}. Report the degrees of freedom of the F-test and the numerical outcome of this test, and draw your conclusion. Note: The relevant 5% critical value is 2.0.

(d) Regress y on a constant and x. Report the t-value and p-value of the slope coefficient. What conclusion would you be tempted to draw if you did not know how the data were generated?

(e) Let e_t be the residuals of the regression of part (d). Regress e_t on a constant and the one-period lagged residual e_{t-1}. What standard assumption of regression is clearly violated for the regression in part (d)?

Erasmus