Cointegration

- Problem with differencing is that lose valuable long run information in the data
- One possible solution to this is cointegration methods which get long run solutions from non stationary variables

Define

$$\varepsilon_t \backsim N(0, \sigma^2)$$

as meaning that ε_t is integrated of order 0 or is I(0)

If

$$y_t = y_{t-1} + \varepsilon_t$$
$$\Delta y_t = \varepsilon_t$$

then y is said to be integrated of order 1 i.e. I(1)

So the order of integration is the number of times that a series has to be differenced before it become stationary

When we specify a regression model in time series we have to make sure the variables are integrated to the same degree

$$y_t = \beta x_t + u_t$$
$$u_t \backsim N(0, \sigma^2)$$

- from the assumptions we need for OLS u_t is I(0) and if y and x are also I(0) we can estimate this regression
- but if y and x are I(1) we have problems and cannot estimate the relation
- except in the special case where $y_t - \beta x_t$ is I(0) where $\beta \neq 0$ and

$$u_t = y_t - \beta x_t$$

In this case the variables are cointegrated, the levels are I(1), but the linear combination of the variables is I(0). This means we can estimate an error correction model

$$\Delta y_t = \alpha \Delta x_t + \lambda(y_{t-1} - \beta x_{t-1}) + v_t$$

as all of these transformed variables are I(0)

- This has the advantage over a first difference model that we can get long and short run information..

The long run relation is

$$y_t = \beta x_t$$

and the short run relation is

$$\Delta y_t = \alpha \Delta x_t + \lambda(y_{t-1} - \beta x_{t-1}) + v_t$$

- Note that if we have more than two variable we have to use:
-

$$\Delta y_t = \alpha \Delta x_t + \lambda(y_{t-1} - \widehat{y}_{t-1}) + v_t$$

- Engle and Granger suggest a two stage procedure

First stage

1. To test if two variables are cointegrated estimate

$$y_t = \beta x_t + u_t$$

with AR(1) errors

$$u_t = \rho u_{t-1} + e_t$$

testing $\rho = 1$ then tests the hypothesis that u_t is I(1)

if $\rho \neq 1$ then u_t is I(0) and we can use the ECM

This is called the Dickey Fuller test and tests the null hypothesis that y and x are not cointegrated.

so if y is i(1) and x is I(1) then we want to make sure that u is *not* I(1).

- If it is then they cannot be cointegrated.
- If the two variables are cointegrated an equilibrium relationship exists and the short run disequilibrium relation can always be represented by an error correction model (ECM)

Second stage

- Estimate

$$y_t = \beta x_t + u_t$$

and get $\hat{\beta}$

- Estimate

$$\Delta y_t = \alpha \Delta x_t + \lambda(y_{t-1} - \hat{\beta} x_{t-1}) + v_t$$

or

$$\Delta y_t = \alpha \Delta x_t + \lambda(\hat{u}_{t-1}) + v_t$$

1. on the assumption they are cointegrated.

   This gives the long run and the short information.

So

- So an alternative to general to specific could be to test for cointegration and if it holds estimate an ECM directly
- A general to specific should lead you to an ECM anyway in this situation
- If it doe not hold use general to specific to find another model

But:

- problem of power of tests in small samples (which are relatively large)
- problems when have more than bivariate relations: has led to the development of new 'systems' methods
- these are extremely important developments in modern econometrics