Cointegration: The Engle and Granger approach

Introduction

Generally one would find most of the economic variables to be non-stationary – I(1) variables. Hence, any equilibrium theories that involve these variables require the existence of a combination of the variables to be stationary. Otherwise, any deviation from equilibrium will not be temporary.

Consider the equilibrium demand for money equation (all variables except the interest rate are in logs):

$$m_t - p_t = \beta_1 + \beta_2 y_t + \beta_3 r_t + \varepsilon_t$$

m – money demand; p – price level; y – real income; r – interest rate.

Economic theory requires $\beta_2>0$ and $\beta_3<0$.

We want the deviation ε_t to be stationary. This will ensure that the variables do not wander off from the equilibrium path.

So we need **rules** concerning linear combination of integrated series.

a) $x_t \sim I(0) \Rightarrow a + b\, x_t \sim I(0)$

$x_t \sim I(1) \Rightarrow a + b\, x_t \sim I(1)$

b) $x_t \sim I(0)$ and $y_t \sim I(0) \Rightarrow a\, x_t + b\, y_t \sim I(0)$

c) $x_t \sim I(0)$ and $y_t \sim I(1) \Rightarrow a\, x_t + b\, y_t \sim I(1)$ i.e. I(1) is a dominant property.

d) Generally if $x_t \sim I(1)$ and $y_t \sim I(1)$ then $ax_t + by_t \sim I(1)$.

But under certain conditions the linear combination may be I(0).

We then say that x and y are **cointegrated**. I.e. we have a stationary equilibrium relationship.

i.e. $z_t = a\, x_t + by_t \sim I(0)$;

e) Adding or subtracting a constant from a cointegrating equation does not alter its properties.

Definition

The components of a (k×1) vector, y_t, are said to be cointegrated of order d, b, denoted, $y_t \sim CI(d, b)$, if (i) all the components of the vector y_t are I(d), that is, they need d differences to induce stationarity, and (ii) there exists a vector $\beta\ (\neq 0)$ so that $z_t = \beta' y_t \sim I(d-b)$. The vector β is called the cointegrating vector. Usually we consider the case with d=b=1.

This is an important result as any arbitrary linear combination of I(1) series will be I(1) (unless the series are cointegrated).

Cointegrating combinations are "equilibria". So it is important to be able to discover and model these relationships.

An alternative approach to the analysis of "long-run" (equilibrium) relationship would be to analyse the relationships between the differences of the series, i.e. among I(0) series. However, this approach is only concerned with short-run movements, while it throws useful long-run information.

Spurious versus cointegrating relationships

The spurious regression problem

Completely unrelated time series may appear to be related using conventional testing procedures.

Suppose $y_t=\rho y_{t-1}+v_t$ where v_t iid $N(0,\sigma^2_v)$ (1)

and

$x_t=\rho x_{t-1}+\varepsilon_t$ where ε_t iid $N(0,\sigma^2_\varepsilon)$ (2)

with v_t and ε_t independent, i.e. $E(v_t\, \varepsilon_s)=0$ for all t and s.

When $\rho=1$, y_t and x_t are random walks:

$$y_t=y_{t-1}+v_t \quad (3)$$

$$x_t=x_{t-1}+\varepsilon_t$$

If we run a regression between y_t and x_t :

$$y_t=\alpha+\beta x_t+u_t \quad (4)$$

despite lack of causal relationship, we are likely to find a significant t-ratio for the null H_0:$\beta=0$. A simulation study would show that Pr(reject H_0 using 5% level test) >75%.

The problem is that the t-test of $\beta=0$ is not N(0,1) even asymptotically. The standard asymptotic distribution theory does not apply when variables have unit root.

So, it is important to discriminate between two situations:

1. Spurious regressions. Apparently significant relationship between unrelated series.
2. Genuine relationships which arise when the time series are cointegrated.

Cointegrating regressions and Granger representation theorem

Here we are primarily concerned with testing for cointegration in a system of k=2 , I(1) variables, in which case there is at most r=1 cointegrating relationship.

Granger, 1983, Co-integrated Variables and Error-Correcting Models, Unpublished discussion paper, 83-13, University of California, San Diego.

Engle and Granger, 1987, Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55 251-276.

If a set of variables are cointegrated, then there exists a valid error correction representation of the data, and viceversa.

If y and x are both I(1) and have a long run relationship, there must be some force which pulls the equilibrium error back to zero.

Engle and Granger (1987, Econometrica) recommend a two-step procedure for cointegration analysis.

(i) Estimate the long-run (equilibrium) equation: $y_t = \delta_0 + \delta_1 x_t + u_t$ (5)

The OLS residuals from (5) are a measure of disequilibrium: $\hat{u}_t = y_t - \hat{\delta}_0 - \hat{\delta}_1 x_t$

A test of cointegration is a test of whether $\hat{u}_t$ is stationary. This is determined by ADF tests on the residuals, with the MacKinnon (1991) critical values adjusted for the number of variables (which MacKinnon denotes as n).

If cointegration holds, the OLS estimator of (5) is said to be super-consistent. Implications: as $T \to \infty$ (i) there is no need to include I(0) variables in the cointegrating equation.

NOTE: The t-ratios from equation (5) are not interpretable, as it is a long-run equation, and therefore will have serial correlation (due to misspecified dynamics) as well as omitted variable problems, and as such the distribution of the t-ratio is not known.

The traditional diagnostic tests from (5) are unimportant as the only important question is the stationarity or otherwise of the residuals.

(ii) Second step: estimate the Error Correction Model

$$\Delta y_t = \phi_0 + \sum_{j=1} \phi_j \Delta y_{t-j} + \sum_{h=0} \theta_h \Delta x_{t-h} + \alpha \hat{u}_{t-1} + \varepsilon_t$$

by OLS as this equation has only I(0) variables, standard hypothesis testing using t-ratios and diagnostic testing of the error term is appropriate. The adjustment coefficient α **must be negative**.

Special case:

$$\Delta y_t = \phi_0 + \phi_1 \Delta y_{t-1} + \theta_1 \Delta x_{t-1} + \alpha (y_{t-1} - \hat{\delta}_0 - \hat{\delta}_1 x_{t-1}) + \varepsilon_t$$

ECM describes how y and x behave in the short run consistent with a long run cointegrating relationship.

Dynamic approach to ECM and cointegration

The estimates from OLS in the static equation (equation 5), although consistent, can be substantially biased in small samples, partly due to serial correlation in the residuals. The bias can be reduced by allowing for some dynamics.

In stage (i) we can estimate, with OLS, an ADL model :

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \gamma y_{t-1} + \varepsilon_t \qquad (6)$$

and solve for the long run equation

$$y_t = \frac{\alpha}{1-\gamma} + \left(\frac{\beta_0 + \beta_1}{1-\gamma} \right) x_t + u_t$$

The residuals from (6)

$$\hat{u}_t = y_t - \frac{\alpha}{1-\gamma} - \left(\frac{\beta_0 + \beta_1}{1-\gamma} \right) x_t$$

are a measure of disequilibrium and a test of cointegration is a test of whether $\hat{u}_t$ is stationary.

As an alternative to the two-step Engle and Granger procedure, the ECM model can be estimated using the residuals from (6). If cointegration holds, the OLS estimator of (6) are super-consistent.

Balanced regressions

Consider again, as an example, the equilibrium demand for money equation (all variables except the interest rate are in logs):

$$m_t = \beta_1 + \beta_2 y_t + \beta_3 r_t + \varepsilon_t$$

where this time m – real money demand; y – real income; r – interest rate.

If m is I(1), what are the necessary conditions for a balanced regression?

(i) **either or both y** and **r** are **I(1)**;

(ii) **if just y** or **r** is **I(1)** then **the other must be I(0)**;

(iii) a balance could also be achieved, in principle, **if y** and **r** were of the **same** but **higher order of integration**, for example **I(2),** but they **cointegrated** on a pairwise basis **to be I(1)** (Patterson, p. 441).