E 4101/5101

Lecture 11: Cointegration, estimation and testing: Part 1

Ragnar Nymoen

Department of Economics, University of Oslo

6 april 2011
Introduction

For the vector $y_t$ consisting of $n \times 1$ variables, we have the Gaussian VAR($p$):

$$y_t = \Phi(L)y_{t-1} + \epsilon_t$$  \hspace{1cm} (1)

We assume that if there are unit-roots in the associated characteristic equation, they are located at the zero frequency. By using the transformed equation

$$\Delta y_t = \Phi^*(L)\Delta y_{t-1} + \Pi y_{t-1} + \epsilon$$  \hspace{1cm} (2)

the “long-run” roots can be found as the roots of the characteristic equation associated with the matrix of levels coefficients $\Pi$. We write $\Pi$ as the product of two matrices $\alpha_{n \times r}$ and $\beta'_{r \times n}$ where $r \equiv rank(\Pi)$:

$$\Pi = \alpha \beta'$$  \hspace{1cm} (3)
Introduction II

We have: The stationary case:

\[ \text{rank}(\Pi) = n \implies y_t \sim I(0) \]

The cointegrating case (reduced rank)

\[ 0 < \text{rank}(\Pi) < n \implies y_t \sim I(1) \]

and no cointegration (danger of spurious regression)

\[ \text{rank}(\Pi) = 0 \implies y_t \sim I(1) \]
Introduction III

- We follow Davidson and MacKinnon, Ch 14, and discuss estimation of known cointegrating vectors first, and after that, the testing of the hypothesis about absence of cointegration (more generally: reduced rank).

- This is a little “back to front” compared to Hamilton’s Ch 19, but the material covered is the same, and this organization has the advantage of bringing out the thin but consequential line between the *spurious regression* case and the *cointegrating regression* case.

The organization for this and the next lecture is:
Introduction IV

1. \( \text{rank}(\Pi) \) is 1
   Estimating a unique cointegrating vector by means of:
   The cointegration regression
   The ECM estimator

2. \( \text{rank}(\Pi) \) is 0 or 1
   Test \( \text{rank}(\Pi) = 0 \) against \( \text{rank}(\Pi) = 1 \), by
   Engle-Granger tests
   ECM test

3. Test and ML estimation based on VAR
   Johansen test for \( \text{rank}(\Pi) \) (other than 0 or 1)
   ML estimation of \( \beta \) for the case of \( \text{rank}(\Pi) \geq 2 \)
   No assumptions about weak exogeneity of variables with respect to \( \beta \).
The cointegrating regression 1

When \( \text{rank}(\Pi) = 1 \), the cointegration vector is unique (subject only to normalization).

Without loss of generality we set \( n = 1 \) and write \( y_t = (y_t, x_t) \) as in a usual regression.

The cointegration parameter \( \beta \) can be estimated by OLS on

\[
y_t = \beta x_t + u_t
\]

where \( u_t \sim I(0) \) by assumption.

\[
(\hat{\beta} - \beta) = \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}.
\]

Since \( x_t \sim I(1) \) we are in the same situation as with the first order AR case with autoregressive parameter equal to one.
The cointegrating regression II

In direct analogy we need to multiply \((\hat{\beta} - \beta)\) by \(T\) in order to obtain a non-degenerate asymptotic distribution:

\[
T(\hat{\beta} - \beta) = \frac{1}{T^2} \sum_{t=1}^{T} x_t u_t - \frac{1}{T^2} \sum_{t=1}^{T} x_t^2,
\]

\(\Rightarrow (\hat{\beta} - \beta)\) converges to zero at rate \(T\), instead of \(\sqrt{T}\) as in the stationary case.

▶ This result is called the Engle-Granger super-consistency theorem.

▶ Davidson and MacKinnon also call \(\hat{\beta}\) from OLS on (4) the levels estimator.
The cointegrating regression III

- In the seminar, we saw that $\hat{\beta}$ is consistent also in the case of a simultaneous equations system.
- Remember that we assume $r = 1$ so the cointegration vector is unique if it exists.
- More generally there is an identification issue, cf Lecture 10.
The distribution of the Engle-Granger (levels) estimator I

The E-G estimator $\hat{\beta}$ has a non-standard distribution which will depend on the details of the DGP.

For the DGP

$$y_t = \beta x_t + u_t, \ u_t \sim N(0, \sigma_u^2)$$

$$\Delta x_t = \varepsilon_{xt}, \ \varepsilon_{xt} \sim N(0, \sigma_\varepsilon^2)$$

$$E[u_t \varepsilon_{xt}] = \sigma_{u\varepsilon}$$

it can be shown, see Banerjee et al. (1983), ch 6.2.2
The distribution of the Engle-Granger (levels) estimator II

\[ T(\hat{\beta} - \beta) \overset{L}{\to} \frac{\sigma_{ue}}{2} \left( W_{\epsilon}(1)^2 + 1 \right) + \sigma_{\epsilon} \cdot h \cdot N(0, \int_0^1 [W_{\epsilon}(r)]^2 dr) \]

where

\[ h = \sqrt{\sigma_u^2 - \sigma_{ue}^2 / \sigma_{\epsilon}^2} \]

- If \( \sigma_{ue} = 0 \) the non-centrality disappears from the numerator, but the bias is still a function of Brownian motions.
  - Not normally distributed.
  - The same applies to the \( t \)-value based on \( \hat{\beta} \): It is not a normal
The distribution of the Engle-Granger (levels) estimator III

Inference “in” the cointegration regression is generally impractical (because standard inference is not valid).

This drawback is even more severe in DGPs with higher order dynamics, because the disturbance of the cointegrating equation is autocorrelated also in the case of cointegration.

\[ \sigma_{ue} \neq 0, \] creates a “second order bias”: Correlation between regressor and disturbance in the cointegration regression does not damage the super-consistency property, but it creates a finite sample bias.

Autocorrelation (due to any omitted higher order dynamics) adds to the second-order bias.
The distribution of the Engle-Granger (levels) estimator IV

All the above (except the detailed expression for $T(\hat{\beta} - \beta)$ of course) applies to more realistic cointegration regressions that contain an intercept and other constant terms.

Note:
Modified Engle-Granger estimator I

There are important developments of the levels estimator that aim at “fixing” the second-order bias problem

- Phillips and Hansen fully modified estimator:
  Subtract an estimate of the bias from \( \hat{\beta} \) (i.e. keep the cointegration regression simple).
  The modified estimator has an asymptotic normal distribution, which allows inference on \( \beta \).

- Saikonnen’s estimator,
  Is based on

\[
y_t = \beta x_t + \gamma_1 \Delta x_{t+1} + \gamma_2 \Delta x_{t-1} + u_t
\]

or higher order lead/lags that “make” \( u_t \) white-noise, see Davidson and MacKinnon p 630.
ECM estimator I

The ECM represents a way of avoiding second order bias due to dynamic mis-specification. This is because, under cointegration, the ECM is implied (the representation theorem)

With \( n = 2, \ p = 1 \) and weak exogeneity of \( x_t (= y_{2t}) \) with respect to the cointegration parameter we have seen that the cointegrated VAR can be re-written as a conditional model and a marginal model (Lecture 10, slides)

\[
\Delta y_t = b \Delta x_t + \phi y_{t-1} + \gamma x_{t-1} + \epsilon_t \tag{8}
\]

\[
\Delta x_t = \epsilon_{xt} \tag{9}
\]
ECM estimator II

where \( b \) is the regression coefficient, and \( \epsilon_t \) and \( \epsilon_{xt} \) are uncorrelated normal variables (by regression).

\[
\Delta y_t = b \Delta x + \phi (y_{t-1} + \frac{\gamma}{\phi} x_{t-1}) + \epsilon_t \\
= b \Delta x + \phi (y_{t-1} + \frac{\beta_{12}}{\beta_{11}} x_{t-1}) + \epsilon_t 
\]

Normalization on \( y_{t-1} \) by setting \( \beta_{11} = -1 \), and defining \( \beta_{12} = \beta \), for comparison with E-G estimator, gives

\[
\Delta y_t = b \Delta x + \phi (y_{t-1} - \beta x_{t-1}) + \epsilon_t 
\]
The ECM estimator $\hat{\beta}^{ECM}$, is obtained from OLS on (8)

$$\hat{\beta}^{ECM} = -\frac{\hat{\gamma}}{\hat{\phi}}$$  \hfill (10)

$\hat{\beta}^{ECM}$ is consistent if both $\hat{\gamma}$ and $\hat{\phi}$ are consistent.

OLS (by construction) chooses the $\hat{\gamma}$ and $\hat{\phi}$ that give the best predictor $y_{t-1} - \hat{\beta}^{ECM}x_{t-1}$ for $\Delta y_t$.

As $T$ grows towards infinity, the true parameters $\gamma$, $\phi$ and $\beta$ will therefore be found.

This is an example of the principle of canonical correlation, which plays a central role in the Johansen method.
ECM estimator IV

Therefore, by direct reasoning:

\[
\hat{\gamma} \xrightarrow{T \to \infty} \gamma, \hat{\phi} \xrightarrow{T \to \infty} \phi \text{ and } \hat{\beta}^{ECM} \xrightarrow{T \to \infty} \beta \quad (11)
\]

In fact:

- \(\hat{\beta}^{ECM}\) is super-consistent
- \(\hat{\beta}^{ECM}\) has better small sample properties than the E-G levels estimator, since it is based on a well specified econometric model (avoids the second-order bias problem).

Inference:

- The distributions of \(\hat{\gamma}\) and \(\hat{\phi}\) (under cointegration) can be shown to be so called “mixed normal” for large \(T\).
ECM estimator V

- Their variances are stochastic variables rather than parameters.
- However, the OLS based t-values of $\hat{\gamma}$ and $\hat{\phi}$ are asymptotically $N(0, 1)$.
- $\hat{\beta}^{ECM}$ is also “mixed normal”, but

$$\left\{ \frac{\hat{\gamma}}{\phi} - \beta \right\} / \sqrt{\text{Var}(\hat{\beta}^{ECM})} \xrightarrow{T \to \infty} N(0, 1) \quad (12)$$

where, despite the change in notation, it is clear that $\text{Var}(\hat{\beta}^{ECM})$ can be found by the formula as in Lecture 7. (Directly available from PcGive when the model is written in ADL form).

- The generalization to $n - 1$ explanatory variables, intercept and dummies is also unproblematic.
ECM estimator VI

- Remember: The efficiency of the ECM estimator depends on the assumed weak exogeneity of $x_t$. 
Engle-Granger test

- The easiest approach is to use an ADF regression to the test null-hypothesis of a unit-root in the residuals $\hat{u}_t$ from the cointegrating regression (4).
- The motivation for the $\Delta \hat{u}_{t-j}$ is as before: to whiten the residuals of the ADF regression
- The DF critical values are shifted to the left as deterministic terms, and/or more $I(1)$ variables in the regression are added.
- See Figure 14.4 in Davidson and Mackinnon
Introduction

Estimating a single cointegrating vector

Testing r=0 against r=1

Figure 14.4 Asymptotic densities of Engle-Granger $\tau_c$ tests
The ECM test

- As we have seen, \( r = 0 \) corresponds to \( \phi = 0 \) in the ECM model in (8):

\[
\Delta y_t = b\Delta x_t + \phi y_{t-1} + \gamma x_{t-1} + \epsilon_t
\]

- It also comes as no surprise that the t-value \( t_\phi \) have typical DF-like distributions under \( H_0 : \phi = 0 \).
- See Figure 14.5 in Davidson and Mackinnon,
Figure 14.5 Asymptotic densities of ECM $\kappa_c$ tests
Why use ECM test instead of the Engle-Granger test? 1

The size of the test (the probability of type 1 error) is more or less the same for the two tests. However, the power of the ECM test is generally larger than for the E-G test.

If \( t_{ECM} \) is the ECM test based on (8), it can be shown that

\[
t_{\phi}^{ECM} \asymp \frac{\sigma_e}{\sigma_\epsilon} t_{EG},
\]

(13)

where \( t_{EG} \) is the E-G test using

\[
\Delta \hat{u}_t = \tau \hat{u}_{t-1} + e_t
\]

(14)

The “t-values”, and therefore the power, will be equal when

\[
\sigma_e = \sigma_\epsilon.
\]
Why use ECM test instead of the Engle-Granger test? II

We can say something about when this will happen: Start with the ECM and bring it on ADL form:

\[ y_t = bx_t + (1 + \phi)y_{t-1} + (\gamma - b)x_{t-1} + \epsilon_t \]

\[ (1 - (1 + \phi)L)y_t = (b + (\gamma - b)L)x_t + \epsilon_t \]

Assume next that the following *Common factor* restriction holds:

\[ \frac{(b + (\gamma - b)L)}{(1 - (1 + \phi)L)} = \beta \]

so that

\[ b = \beta \]

\[ (\gamma - b) = -\beta(1 + \phi) \]
Why use ECM test instead of the Engle-Granger test? III

\[ y_t = \beta x_t + (1 + \phi)y_{t-1} - \beta(1 + \phi)x_{t-1} + \epsilon_t \]  
\[ \Delta y_t - \beta \Delta x_t = \phi(y_{t-1} - \beta x_{t-1}) + \epsilon_t \]  

If we replace \( \beta \) by \( \hat{\beta} \), we have

The ECM model (8) implies the Dickey-Fuller regression

\[ \underbrace{\Delta y_t - \hat{\beta} \Delta x_t}_{\Delta \hat{u}_t} = \phi(y_{t-1} - \hat{\beta} x_{t-1}) + \epsilon_t \]  
\[ \hat{u}_{t-1} \]  

when the Common factor restriction in (15) is true.

- If the Common factor restriction is invalid, the E-G test is based on a mis-specified model.
- As a consequence \( \sigma_e > \sigma_{\epsilon} \), and there is a loss of power relative to ECM test.
Why use ECM test instead of the Engle-Granger test? IV

- Common factor restrictions are easy to test in PcGive, also for higher order multivariate ADL
- You can read more about Common factor restrictions in Davidson and MacKinnon p 294-298
- There the focus is that if we have an ADL\((p, p)\) model with white-noise errors and \(m\) Common factor restrictions hold, we can obtain a more parsimonious ADL\((p - m, p - m)\).
- The disturbance will then be autocorrelated but this autocorrelation is a “convenient simplification, not a nuisance”.
- Note that Common factors restrict the dynamic multipliers in a quite marked way.
Why use ECM test instead of the Engle-Granger test? V

- For example if the restriction in (15) holds, $y$ in (16) has the same impact and long-run multiplier with respect to $x$.
- We can see these implications in our case:

$$y_t = \frac{(b + (\gamma - b)L)}{(1 - (1 + \phi)L)} x_t + \frac{1}{1 - (1 + \phi)L} \varepsilon_t$$

which gives

$$y_t = \beta x_t + u_t,$$

$$u_t = (1 + \phi)u_{t-1} + \varepsilon_t$$

Same impact and long-run multiplier, and AR(1) disturbance.
References
