

E 4101/5101  
Lecture 11/12: Co-integration: tests and  
estimation (1 of 2)  
Ragnar Nymoen

Department of Economics, University of Oslo

23 April 2014

## Introduction I

For the vector  $\mathbf{y}_t$  consisting of  $n \times 1$  variables, we have the Gaussian  $VAR(p)$ :

$$\mathbf{y}_t = \Phi(L)\mathbf{y}_{t-1} + \varepsilon_t \quad (1)$$

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

$$\Delta\mathbf{y}_t = \Phi^*(L)\Delta\mathbf{y}_{t-1} + \Pi\mathbf{y}_{t-1} + \varepsilon \quad (2)$$

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  $r \equiv \text{rank}(\Pi)$  :

$$\Pi = \alpha\beta' \quad (3)$$

## Introduction II

We have: The stationary case:

$$\text{rank}(\mathbf{\Pi}) = n \implies \mathbf{y}_t \sim \mathbf{I}(0)$$

The cointegrating case (reduced rank)

$$0 < \text{rank}(\mathbf{\Pi}) < n \implies \mathbf{y}_t \sim \mathbf{I}(1)$$

and no cointegration (danger of spurious regression)

$$\text{rank}(\mathbf{\Pi}) = 0 \implies \mathbf{y}_t \sim \mathbf{I}(1)$$

- ▶ 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).

## Introduction III

- ▶ 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:

1.  $rank(\mathbf{\Pi})$  is 1

Estimating a unique cointegrating vector by means of:

The cointegration regression

The ECM estimator

## Introduction IV

- $rank(\mathbf{\Pi})$  is 0 or 1  
Test  $rank(\mathbf{\Pi}) = 0$  against  $rank(\mathbf{\Pi}) = 1$ , by  
Engle-Granger tests  
ECM test
- Test and ML estimation based on VAR  
Johansen test for  $rank(\mathbf{\Pi})$  (other than 0 or 1)  
ML estimation of  $\beta$  for the case of  $rank(\mathbf{\Pi}) \geq 2$  No assumptions about weak exogeneity of variables with respect to  $\beta$ .

## The cointegrating regression I

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

Without loss of generality we set  $n = 1$  and write  $\mathbf{y}_t = (y_t, x_t)$  as in a usual regression.

The cointegration parameter  $\beta$  can be estimated by OLS on

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

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

$$(\hat{\beta} - \beta) = \frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}. \quad (5)$$

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:

$$T(\hat{\beta} - \beta) = \frac{\frac{1}{T} \sum_{t=1}^T x_t u_t}{\frac{1}{T^2} \sum_{t=1}^T x_t^2}, \quad (6)$$

$\implies (\hat{\beta} - \beta)$  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

- ▶ 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 slide set.
- ▶ In seminar 6 we investigate  $\hat{\beta}$  when  $y_t$  and  $x_t$  are generated by simultaneous equations system



## 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

$$\begin{aligned}y_t &= \beta x_t + u_t, \quad u_t \sim N(0, \sigma_u^2) \\ \Delta x_t &= \varepsilon_{xt}, \quad \varepsilon_{xt} \sim N(0, \sigma_\varepsilon^2) \\ E[u_t \varepsilon_{xt}] &= \sigma_{u\varepsilon}\end{aligned}$$

it can be shown, cf. Banerjee et al. (1983), ch 6.2.2

## The distribution of the Engle-Granger (levels) estimator II

$$T(\hat{\beta} - \beta) \xrightarrow{L} \frac{\frac{\sigma_{ue}}{2}(W_\varepsilon(1)^2 + 1) + \sigma_\varepsilon \cdot h \cdot N(0, \int_0^1 [W_\varepsilon(r)]^2 dr)}{\sigma_\varepsilon^2 \int_0^1 [W_\varepsilon(r)]^2 dr}, \quad (7)$$

where  $h = \sqrt{\sigma_u^2 - \sigma_{u\varepsilon}^2 / \sigma_\varepsilon^2}$

- ▶ If  $\sigma_{u\varepsilon} = 0$  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
- ⇒ Inference “in” the cointegration regression is generally impractical (because standard inference is not valid)

## The distribution of the Engle-Granger (levels) estimator III

- ▶ 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.
- ▶  $\sigma_{u\varepsilon} \neq 0$ , 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.
- ▶ 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.

## The distribution of the Engle-Granger (levels) estimator IV

Note:

Hamilton contains more general results for the cointegrating regression: Proposition 19.2, and Ch 19.3, p 601-608.

## 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)

$$\Delta y_t = b\Delta x_t + \underbrace{\phi}_{\alpha_{11}\beta_{11}} y_{t-1} + \underbrace{\gamma}_{\alpha_{11}\beta_{12}} x_{t-1} + \epsilon_t \quad (8)$$

$$\Delta x_t = \epsilon_{xt} \quad (9)$$

## ECM estimator II

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

$$\begin{aligned}\Delta y_t &= b\Delta x + \phi\left(y_{t-1} + \frac{\gamma}{\phi}x_{t-1}\right) + \epsilon_t \\ &= b\Delta x + \phi\left(y_{t-1} + \frac{\beta_{12}}{\beta_{11}}x_{t-1}\right) + \epsilon_t\end{aligned}$$

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

$$\Delta y_t = b\Delta x + \phi(y_{t-1} - \beta x_{t-1}) + \epsilon_t$$

## ECM estimator III

The ECM estimator  $\hat{\beta}^{ECM}$ , is obtained from OLS on (8)

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

$\hat{\beta}^{ECM}$  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 also plays a central role in the Johansen method (Lecture 13).



## ECM estimator IV

Therefore, by direct reasoning:

$$\hat{\gamma} \xrightarrow{T \rightarrow \infty} \gamma, \hat{\phi} \xrightarrow{T \rightarrow \infty} \phi \text{ and } \hat{\beta}^{ECM} \xrightarrow{T \rightarrow \infty} \beta \quad (11)$$

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  $T$ .
  - ▶ Their variances are stochastic variables rather than parameters.

## ECM estimator V

- ▶ 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

$$\left\{ \begin{array}{c} \hat{\gamma} \\ \hat{\phi} \end{array} - \beta \right\} / \sqrt{\text{Var}(\hat{\beta}^{ECM})} \xrightarrow[T \rightarrow \infty]{} N(0, 1) \quad (12)$$

where,  $\text{Var}(\hat{\beta}^{ECM})$  can be found by the delta method (In PcGive slo by choosing Dynamic Analysis-Static Solution when the model is in ADL form).

- ▶ The generalization to  $n - 1$  explanatory variables, intercept and dummies is also unproblematic.
- ▶ 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 test the 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 critical values are shifted to the left as deterministic terms, and/or more  $I(1)$  variables in the regression are added.
- ▶ See Figure e.g. 14.4 in Davidson and MacKinnons book.

## 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 (2004)
- ▶ or Ericsson and MacKinnon (2002) for a self-contained exposition with critical values, and comparison with E-G test and Johansen's reduced rank test.

## Why use ECM test instead of the Engle-Granger test? I

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

$$t_{\phi}^{ECM} \cong \frac{\sigma_e}{\sigma_{\epsilon}} t_{\tau}^{EG}, \quad (13)$$

where  $t^{EG}$  is the E-G test using

$$\Delta \hat{u}_t = \tau \hat{u}_{t-1} + e_t \quad (14)$$

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:

$$\frac{(b + (\gamma - b)L)}{(1 - (1 + \phi)L)} = \beta \quad (15)$$

so that

$$b = \beta$$

$$(\gamma - b) = -\beta(1 + \phi)$$

## Why use ECM test instead of the Engle-Granger test? III

$$y_t = \beta x_t + (1 + \phi)y_{t-1} - \beta(1 + \phi)x_{t-1} + \epsilon_t \quad (16)$$

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

If we replace  $\beta$  by  $\hat{\beta}$ , we have

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

$$\underbrace{\Delta y_t - \hat{\beta} \Delta x_t}_{\Delta \hat{u}_t} = \phi \underbrace{(y_{t-1} - \hat{\beta} x_{t-1})}_{\hat{u}_{t-1}} + \epsilon_t \quad (17)$$

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.
  - ▶ For example if the restriction in (15) holds,  $y$  in (16) has the same impact and long-run multiplier with respect to  $x$ .



## Why use ECM test instead of the Engle-Granger test? V

- ▶ 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

- Banerjee, A., J. Dolado, J. W. Galbraith and D. F. Hendry (1993): *Co-integration, Error-Correction and the Econometric Analysis of Non-Stationary Data*, Oxford University Press
- Ericsson, N.E. and J.G. MacKinnon (2002): Distributions of error-correction tests for cointegration, *Econometrics Journal*, 5, 285–318