ECON 4160, Spring term 2015. Lecture 10+11

Co-integration

Ragnar Nymoen

Department of Economics

26 October 2014

Introduction I

So far we have considered:

- Stationary VAR, ("no unit roots")
  - Standard inference
- Non-stationary VAR ("all unit-roots")
  - Danger of spurious relationships
  - Need Dickey-Fuller distribution to test the null hypothesis of unit-root for a single time series

We next consider *cointegration*, the case of "some, but not only unit-roots" in the VAR.

- In such systems, there exist one or more linear combinations of $I(1)$ variables that are $I(0)$—they are called *cointegration relationships*.

Introduction II

- We may see already that cointegration is the "flip of the coin" of spurious regression: If we have two **dependent** $I(1)$ variables, they are cointegrated.
- We can also guess the correct distrubtion to use for testing the null hypothesis of no cointegration is going to be of the Dickey-Fuller type:

  Why: If we reject $Y_t \sim I(1)$ against $Y_t \sim I(0)$ using critical values from DF-distributions, we have shown that Y_t is "cointegrated with itself!"

Introduction III

- In these two lectures we sketch the theory of cointegration more fully:
  - The cointegrated VAR: VARs with some, but not all unit-roots
  - Testing the null-hypothesis of no cointegration
    - The cointegrating regression
    - The conditional ECM
    - VAR methods, testing hypotheses about multiple cointegrating relationships
  - Estimating the cointegrated VAR.

Introduction IV

- Ci- relationships correspond to equilibrium relationships from economic theory.
  - Finding no evidence for cointegration should lead us to question whether equilibrium is rightfully such a central concept in macroeconomics.
  - Finding too many (spurious) ci-relationships may lead us to being too optimistic about the economy's ability to regulate itself.
- **References**:
  - HN: Ch. 17.
  - DM: Ch. 14.
  - BN(2104): Kap. 11.
  - The posted paper by Ericsson and MacKinnon.

The VAR with one unstable and one stable root I

Consider the bi-variate first order VAR

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1}$$

where $\mathbf{y}_t = (Y_t, X_t)$, $\mathbf{\Phi}$ is a 2×2 matrix with coefficients and $\boldsymbol{\varepsilon}_t$ is a vector with Gaussian disturbances.
The characteristic equation for $\mathbf{\Phi}$:

$$|\mathbf{\Phi} - z\mathbf{I}| = 0,$$

Our interest is the case with one unit-root and one stationary root:

$$z_1 = 1, \text{and } z_2 = \lambda, |\lambda| < 1. \tag{2}$$

implying that both X_t and Y_t are $I(1)$. Why?

The VAR with one unstable and one stable root II

$\boldsymbol{\Phi}$ has full rank, equal to 2. It can be diagonalized in terms of its eigenvalues and the corresponding eigenvectors:

$$\boldsymbol{\Phi} = \mathbf{P} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \mathbf{Q} \tag{3}$$

$\mathbf{P}$ has the eigenvectors as columns:

$$\mathbf{P} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \tag{4}$$

Cointegrated VAR—ECM implication I

Using the above assumptions and diagonalization (1) can be written as:

$$\begin{bmatrix} W_t \\ -EC_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} W_{t-1} \\ -EC_{t-1} \end{bmatrix} + \boldsymbol{\eta}_t, \tag{5}$$

$\boldsymbol{\eta}_t$ contains linear combinations of the original VAR disturbances. EC_t and W_t are given by:

$$W_t = \delta Y_t - \beta X_t \tag{6}$$

$$EC_t = -\gamma Y_t + \alpha X_t. \tag{7}$$

- $W_t \sim I(1)$,is a stochastic trend (Random-Walk)
- $EC_t \sim I(0)$,a stationary variable

Cointegrated VAR—ECM implication II

- We say that there is **cointegration** between X_t and Y_t, since EC_t is a stationary variable, and it is a linear combination of X_t and Y_t.
- $-\gamma$ and α are the **cointegrating parameters** in this example.

The Common Trends representation I

The *Common Trends* representation for Y_t and X_t is:

$$Y_t = \alpha W_t - \beta EC_t \tag{8}$$

$$X_t = \gamma W_t - \delta EC_t. \tag{9}$$

- X_t and Y have a common stochastic trend, namely W_t.

The Common Trends representation II

Two consequences for forecasts

1. Forecasts for $X_{T+h|T}$ and $Y_{T+h|T}$ become dominated by the common stochastic trend
2. Cointegration is maintained in the forecasts, so $EC_{T+h|T} = -\gamma X_{T+h|T} + \alpha Y_{T+h|T} = 0$ for large h.

The ECM representation of the cointegrated VAR I

As before, can re-parameterize the VAR (1) as

$$\Delta \mathbf{y}_t = \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{10}$$

with

$$\boldsymbol{\Pi} = (\boldsymbol{\Phi} - \mathbf{I}) \tag{11}$$

Next, define two (2×1) parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in such a way that the product $\boldsymbol{\alpha}\boldsymbol{\beta}'$ gives $\boldsymbol{\Pi}$:

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{12}$$

In our example, it can be shown (compare BN 2014 Kap 11)

$$\boldsymbol{\Pi} = \underbrace{\begin{bmatrix} (1-\lambda)\beta \\ (1-\lambda)\delta \end{bmatrix}}_{\boldsymbol{\alpha}} \underbrace{\begin{bmatrix} \gamma & -\alpha \end{bmatrix}}_{\boldsymbol{\beta}'}$$

The ECM representation of the cointegrated VAR II

and then (10) can be expressed as:

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \boldsymbol{\alpha}\boldsymbol{\beta}' \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t, \tag{13}$$

- $\boldsymbol{\alpha}$ is known as the (matrix) of **equilibrium correction coefficients** (aka adjustment coefficients, or loadings),

$$\boldsymbol{\alpha} = \begin{bmatrix} (1-\lambda)\beta \\ (1-\lambda)\delta \end{bmatrix} \tag{14}$$

- $\boldsymbol{\beta}$ is the matrix of long-run cointegration coefficients

The ECM representation of the cointegrated VAR III

$$\boldsymbol{\beta} = \begin{bmatrix} \gamma \\ -\alpha \end{bmatrix} \quad (15)$$

In this formulation we see that

- $rank(\mathbf{\Pi}) = 0$, reduced rank and no cointegration. Both eigenvalues are zero.
- $rank(\mathbf{\Pi}) = 1$, reduced rank and cointegration. One eigenvalue is different from zero.
- $rank(\mathbf{\Pi}) = 2$, full rank, both eigenvalues are different from zero and the VAR (1) is stationary.

Cointegration and Granger causality

Since $\lambda < 1$ is equivalent with cointegration, we see from (14) that cointegration also implies Granger-causality in at least one direction: $(1-\lambda)\beta \neq 0$ and/or $(1-\lambda)\beta \neq 0$.

The ECM representation of the cointegrated VAR IV

Cointegration and weak exogeneity

- Assume $\delta = 0$, from (14). This implies

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = (1-\lambda) \begin{bmatrix} \beta \\ 0 \end{bmatrix} [\gamma Y_{t-1} - \alpha X_{t-1}] + \boldsymbol{\varepsilon}_t$$

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \begin{bmatrix} (1-\lambda)\beta[\gamma Y_{t-1} - \alpha X_{t-1}] + \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{bmatrix}$$

- The marginal model contains no information about the cointegration parameters $(\gamma, -\alpha)'$. Y_t is Weakly Exogenous (WE) for the cointegration parameters $\boldsymbol{\beta}' = (\gamma, -\alpha)'$.
- So how can be test for WE of X_t with respect to $\boldsymbol{\beta}$?

VAR(p) —> ECM general case I

If $\mathbf{y}_t$ is $n \times 1$ with $I(1)$ variables. The VAR is:

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

where $\boldsymbol{\varepsilon}_t$ is multivariate Gaussian and

$$\mathbf{\Phi}(L) = \sum_{i=0}^{p} \mathbf{\Phi}_{i+1} L^i \tag{16}$$

In analogy to the scaler case, the matrix lag-polynomial can be written as

$$\mathbf{\Phi}(L) = \mathbf{\Phi}(1) + \Delta\mathbf{\Phi}^*(L)$$

where the $\mathbf{\Phi}_i^*$ matrices

$$\mathbf{\Phi}^*(L) = \mathbf{\Phi}_1^* + \mathbf{\Phi}_2^* L + \ldots + \mathbf{\Phi}_{p-1}^* L^{p-1}$$

VAR(p) —> ECM general case II

are linear transformations of $\mathbf{\Phi}_i$ $(i = 1, \ldots, p)$. Substitution yields

$$\begin{aligned} \mathbf{y}_t &= \mathbf{\Phi}^*(L)\Delta\mathbf{y}_{t-1} + \mathbf{\Phi}(1)\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \\ \Delta\mathbf{y}_t &= \mathbf{\Phi}^*(L)\Delta\mathbf{y}_{t-1} + \mathbf{\Pi}(1)\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \end{aligned} \tag{17}$$

where $\mathbf{\Pi}(1) \equiv \mathbf{\Phi}(1) - \mathbf{I}_N = \mathbf{0}$ in the case of no cointegration but

$$\mathbf{\Pi}(1) = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{18}$$

in the case of r cointegrating-vectors.

- $\boldsymbol{\beta}_{n\times r}$ contains the CI-vectors as columns, while $\boldsymbol{\alpha}_{n\times r}$ shows the strength of equilibrium correction in each of the equations for $\Delta Y_{1t}, \Delta Y_{2t}, \ldots, \Delta Y_{nt}$. In general $\text{rank}(\boldsymbol{\beta}) = r$ and $\text{rank}(\mathbf{\Pi}) = r < n$.

VAR(p) —> ECM general case III

- If $\boldsymbol{\beta}$ is known, the system

$$\Delta \mathbf{y}_t = \boldsymbol{\Phi}^*(L)\Delta \mathbf{y}_{t-1} + \boldsymbol{\alpha}[\boldsymbol{\beta}'\mathbf{y}]_{t-1} + \boldsymbol{\varepsilon}_t \tag{19}$$

  contains only $I(0)$ variables and conventional asymptotic inference applies.
- Moreover: If $\boldsymbol{\beta}$ is regarded as known, *after first estimating* $\boldsymbol{\beta}$, conventional asymptotic inference also applies.
- (19) is then a stationary VAR, called the VAR-ECM or the cointegrated VAR.
- This system can be identified and modelled with the concepts that we have developed for the stationary case.

Restricted and unrestricted constant term I

- Usually we include separate *Constants* in each row of the VAR.
- We call them unrestricted constant terms. In the unit-root the implication is that each Y_{jt} contains a deterministic trend (think of a Random Walk with drift)
- However if the constants are *restricted* to be in the EC_{t-1} variables, there are no drifts and therefore no trend in the levels variables. We don't give the precise argument here.
- We mention it here because it reminds us that, in the same way as with DF-test, the role of deterministic terms is important when there are unit-roots.
- It also matters for the construction of the tests we use (again, the DF test is a parallel).

Conditional cointegrated ECM I

Assume that $\alpha_{21} = 0$, i.e. Y_{2t} is weakly exogenous for $\boldsymbol{\beta}$.
With Gaussian disturbances $\boldsymbol{\varepsilon}_t = N(0, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ has elements ω_{ij},we can derive the conditional model for ΔY_{1t}:

$$\Delta Y_{1t} = \underbrace{\omega_{21}\omega_{22}^{-1}}_{b}\Delta Y_{2t} + \alpha_{11}\boldsymbol{\beta}' \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \underbrace{\varepsilon_{1t} - \omega_{21}\omega_{22}^{-1}\varepsilon_{2t}}_{u_t} \quad (20)$$

the single equation ECM we have discussed before.
(20) is an example of an open system, since x_{t-1} is determined outside the model.
If we write it as

$$\Delta Y_{1t} = b\Delta Y_{2t} + \alpha_{11}\beta_{11} Y_{1t-1} + \alpha_{11}\beta_{12} Y_{2t-1} + u_t$$

we see that $\boldsymbol{\Pi}$=$\alpha_{11}\beta_{11} \neq 0$, i.e., the $\boldsymbol{\Pi}$ "matrix" has full rank.

Conditional cointegrated ECM II

- Open system are often relevant. Ideally after first testing $\alpha_{21} = 0$. But not always that the research purpose requires this: Can be interested in modelling the interaction between for example wages and prices conditional on productivity.
- The common $I(1)$-trend is now the non-modelled but observable variable Y_{2t-1}.
- Care must be taken: The relevant distribution for testing $\text{rank}(\mathbf{\Pi}) = 0$ is (as we shall see) different from the distribution that applies for the closed system.
- *Generalization:* If the open system contain n_1 endogenous $I(1)$ variables and n_2 non-modelled $I(1)$ variables. Cointegration is consistent with:

$$0 < rank(\mathbf{\Pi}) \leq n_1$$

Identification I

- As we have seen, if $n = 2$, cointegration implies $rank(\mathbf{\Pi}) = 1$
  - There is one cointegration vector

$$(\beta_{11}, \beta_{12})'$$

    which is uniquely identified after normalization. For example with $\beta_{11} = -1$ the ECM variable becomes

$$ECM_{1t} = -Y_{1t} + \beta_{12} Y_{2t} \sim I(0)$$

- When $n > 2$, we can have $rank(\mathbf{\Pi}) > 1$, and in these cases the cointegrating vectors are not identified.

Identification II

- Assume that ˝ is known (in practice, consistently estimated), and $\boldsymbol{\beta}$ is a $n \times r$ cointegrating vector:

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

However for a $r \times r$ non-singular matrix $\boldsymbol{\Theta}$:

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\Theta}\boldsymbol{\Theta}^{-1}\boldsymbol{\beta}' = \boldsymbol{\alpha}_{\Theta}\boldsymbol{\beta}'_{\Theta}$$

showing that $\boldsymbol{\beta}'_{\Theta}$ is also a cointegrating vector.

This problem is equivalent to the identification problem in simultaneous equation models!

Identification III

- Assume $rank(\mathbf{\Pi}) = 2$ for a $n = 3$ VAR

$$-Y_{1t} + \beta_{12}Y_{2t} + \beta_{13}Y_{3t} = ECM_{1t}$$
$$\beta_{21}Y_{1t} - Y_{2t} + \beta_{13}Y_{3t} = ECM_{2t}$$

- By simply viewing these as a pair of simultaneous equations, we see that they are not identified on the order-condition.
- Exact identification requires for example 1 linear restrictions on each of the equations.
  - For example $\beta_{13} = 0$ and $\beta_{21} + \beta_{13} = 0$ will result in exact identification
  - Identification = theory !!!
- Restrictions of the loading matrix can also help identification (then we impose hypotheses about causation)

Identification IV

- A very useful estimator of $\mathbf{\Pi}$ is the Maximum-Likelihood estimator (OLS on each equation in the VAR). A natural test-statistic for any overidentifying restrictions is the LR test.
- The identification issue applies equally for open systems. Again, in direct analogy to the simultaneous equation model.

The cointegrating regression I

When $rank(\mathbf{\Pi}) = 1$, the cointegration vector is unique (subject only to normalization).
Without loss of generality we set $n = 1$ and write $\mathbf{y}_t = (Y_t, X)$ as in a usual regression.
The cointegration parameter β can be estimated by OLS on

$$Y_t = \beta X_t + u_t \tag{21}$$

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

$$(\hat{\beta} - \beta) = \frac{\sum_{t=1}^{T} X_t u_t}{\sum_{t=1}^{T} X_t^2}. \tag{22}$$

The cointegrating regression II

Since $X_t \sim I(1)$, we are in a the same situation as with the first order AR case with autoregressive parameter equal to one (Lecture 9)

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

$$T(\hat{\beta} - \beta) = \frac{\frac{1}{T}\sum_{t=1}^{T} X_t u_t}{\frac{1}{T^2}\sum_{t=1}^{T} X_t^2}, \tag{23}$$

$\Longrightarrow (\hat{\beta} - \beta)$ 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*.

The cointegrating regression III

- Remember: This is based on $r = 1$ so the cointegration vector is unique if it exists.

The distribution of the Engle-Granger (levels) estimator I

- Even with simple DGPs the E-G estimator is not normally distributed.
- The same applies to the $t-$value based on $\hat{\beta}$: It does *not* have a normal distribution
  $\Longrightarrow$ Inference "in" the cointegration regression is generally impractical (because standard inference in 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.

Modified Engle-Granger estimator I

- Phillips and Hansen fully modified estimator:
  Subtract an estimate of the finite sample bias from $\hat{\beta}$ (i.e. keep the cointegration regression simple).
  The modified estimator has an asymptotic normal distribution, which allows inference on β.
- 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 DM p 630.

ECM estimator I

The ECM represents a way of avoiding second order bias due to dynamic mis-specification.
This is because, under the assumption of 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

$$\Delta Y_t = b\Delta X_t + \underbrace{\phi}_{\alpha_{11}\beta_{11}} Y_{t-1} + \underbrace{\gamma}_{\alpha_{11}\beta_{12}} X_{t-1} + \epsilon_t \tag{24}$$

$$\Delta X_t = \varepsilon_{xt} \tag{25}$$

ECM estimator II

where b is the regression coefficient, and ϵ_t and ε_{xt} are uncorrelated normal variables (by regression).

$$\Delta Y_t = b\Delta X_t + \phi(Y_{t-1} + \frac{\gamma}{\phi}X_{t-1}) + \epsilon_t$$
$$= b\Delta X + \phi(Y_{t-1} + \frac{\beta_{12}}{\beta_{11}}X_{t-1}) + \epsilon_t$$

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

$$\Delta Y_t = b\Delta X_t + \phi(Y_{t-1} - \beta X_{t-1}) + \epsilon_t$$

ECM estimator III

The ECM estimator $\hat{\beta}^{ECM}$, is obtained from OLS on (24)

$$\hat{\beta}^{ECM} = -\frac{\hat{\gamma}}{\hat{\phi}} \tag{26}$$

$\hat{\beta}^{ECM}$ 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 Δy_t.
As T grows towards infinity, the true parameters γ, ϕ and β will therefore be found.
This is an example of *canonical correlation*, known from multivariate statistics.

ECM estimator IV

Therefore, by direct reasoning:

$$\widehat{\gamma} \underset{T\to\infty}{\longrightarrow} \gamma,\ \widehat{\phi} \underset{T\to\infty}{\longrightarrow} \phi \text{ and } \hat{\beta}^{ECM} \underset{T\to\infty}{\longrightarrow} \beta \tag{27}$$

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 $\widehat{\gamma}$ and $\widehat{\phi}$ (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 $\widehat{\gamma}$ and $\widehat{\phi}$ are asymptotically $N(0,1)$.
- $\hat{\beta}^{ECM}$ is also "mixed normal", but

$$\left\{\frac{\widehat{\gamma}}{\widehat{\phi}} - \beta\right\} / \sqrt{Var(\hat{\beta}^{ECM})} \underset{T\longrightarrow\infty}{\longrightarrow} N(0,1) \qquad (28)$$

  where, despite the change in notation, it is clear that $Var(\hat{\beta}^{ECM})$ can be found by using the delta-method.
- 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 the test the null-hypothesis of a unit-root in the residuals $\hat{u}_t$ from the cointegrating regression (21).
- The motivation for the $\Delta\hat{u}_{t-j}$ terms 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.

The ECM test

- As we have seen, $r = 0$ corresponds to $\phi = 0$ in the *ECM* model in (24):

$$\Delta Y_t = b\Delta X_t + \phi Y_{t-1} + \gamma X_{t-1} + \epsilon_t$$

- It also comes as no surprise that the t-value t_ϕ have typical DF-like distributions under $H_0 : \phi = 0$.
- See DN and/or Ericsson and MacKinnon (2002) for critical values.

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_{ϕ}^{ECM} is the ECM test based on (24), it can be shown that

$$t_{\phi}^{ECM} \cong \frac{\sigma_e}{\sigma_\epsilon} t_{\tau}^{EG}, \tag{29}$$

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

$$\Delta \hat{u}_t = \tau \hat{u}_{t-1} + e_t \tag{30}$$

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 restriction holds:

$$\frac{(b + (\gamma - b)L)}{(1-(1+\phi)L)} = \beta \tag{31}$$

(the is a Common Factor in the lag polynomial) so that

$$b = \beta$$

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

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

$$Y_t = \beta X_t + (1+\phi) Y_{t-1} - \beta(1+\phi) X_{t-1} + \epsilon_t \quad (32)$$

$$\Delta Y_t - \beta \Delta X_t = \phi(Y_{t-1} - \beta X_{t-1}) + \epsilon_t$$

If we replace β by $\hat{\beta}$, we have
The ECM model (24) implies the Dickey-Fuller regression

$$\underbrace{\Delta Y_t - \hat{\beta}\Delta X_t}_{\Delta \hat{u}_t} = \phi \underbrace{(y_{t-1} - \hat{\beta} X_{t-1})}_{\hat{u}_{t-1}} + \epsilon_t \quad (33)$$

when the Common factor restriction in (31) 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.

Testing cointegrating rank I

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

$$\mathbf{y}_t = \mathbf{\Phi}(L)\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{34}$$

and use the re-parameterized equation:

$$\Delta\mathbf{y}_t = \mathbf{\Phi}^*(L)\Delta\mathbf{y}_{t-1} + \mathbf{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{35}$$

We write the levels coefficient matrix $\mathbf{\Pi}$ as the product of two matrices $\boldsymbol{\alpha}_{n\times r}$ and $\boldsymbol{\beta}'_{r\times n}$ where $r \equiv rank(\mathbf{\Pi})$:

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{36}$$

We are interested in both the cointegrating case

$$0 < rank(\mathbf{\Pi}) < n$$

Testing cointegrating rank II

and the case with no cointegration

$$rank(\mathbf{\Pi}) = 0$$

- $rank(\mathbf{\Pi})$ is given by the number of non-zero eigenvalues of $\mathbf{\Pi}$. But can we find the number of eigenvalues that are significantly different from zero?
- Fortunately, this problem has a solution. An eigenvalue of $\mathbf{\Pi}$ is a special kind of squared correlation coefficient known as a *canonical correlation* in multivariate statistics.
- This method has become known as the **Johansen approach**. It is likelihood based, see HN § 17.3.2

Intuition I

- For concreteness, consider $n = 3$ so r can be 0,1 or 2
- $r = 0$ corresponds to $\mathbf{\Pi} = \mathbf{0}$ in the context of cointegration:
- From the representation theorem; with two unit-roots

$$\mathbf{\Pi}=\mathbf{\Phi} - \mathbf{I} = \mathbf{P} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{P}^{-1} - \mathbf{I} = \mathbf{0}.$$

- $r = 1$ corresponds to $\boldsymbol{\alpha}_{3\times 1} \neq \mathbf{0}$ for a single cointegration vector $\boldsymbol{\beta}'_{1\times 3}$.
- For this to make sense, $\boldsymbol{\beta}'_{1\times 3}\mathbf{y}_{t-1}$ must be a $I(0)$ and it must be a significant predictor of $\Delta\mathbf{y}_t$.

Intuition II

- The strength of the relationship can be estimated by the highest squared canonical correlation coefficient, call it $\hat{\rho}_1^2$,between $\Delta \mathbf{y}_t$ and all the possible the linear combinations of the variables in $\mathbf{y}_{t-1}$.
- If $\hat{\rho}_1^2 > 0$ is statistically significant, we reject that $r = 0$.
- $\hat{\rho}_1^2$ is the same as the highest eigenvalue of $\hat{\mathbf{\Pi}}$, and $\hat{\boldsymbol{\beta}}'_{1\times 3}$ is the corresponding eigenvector.
- If $r = 0$ is rejected we can,continue, and test $r = 1$ against $r = 2$.
- If the second largest canonical correlation coefficient $\hat{\rho}_2^2$ is also significantly different from zero, we conclude that the number of cointegrating vectors is two. $\hat{\boldsymbol{\beta}}'_{2\times 3}$ is the corresponding eigenvector

Intuition III

- It can be shown that, for the Gaussian VAR, $\hat{\boldsymbol{\beta}}'_{1\times3}$ and $\hat{\boldsymbol{\beta}}'_{2\times3}$ are ML estimates.

Trace-test and max-eigenvalue test I

- We order the canonical correlations from largest to smallest and construct the so called trace test:

$$\textit{Trace-test} = -T \sum_{i=r+1}^{3} \ln(1 - \hat{\rho}_i^2), \; r = 0, 1, 2 \qquad (37)$$

- If $\hat{\rho}_1^2$ is close to zero, then clearly *Trace-test* will be close to zero, and we we will not reject the H_0 of $r = 0$ against $r \geq 1$.
- and so on for H_0 of $r = 1$ against $r \geq 2$
- Of course: to make this a formal testing procedure, we need the critical values from the distribution of the *Trace-test* for the sequence of null-hypotheses.

Trace-test and max-eigenvalue test II

- The distributions are non-standard, but at least the main cases are tabulated in PcGive.
- A closely related test is called the *max-eigenvalue* test,(but the trace test is today judged most reliable)
- If there is a single cointegrating vector and there are $n - 1$ weakly-exogenous variables, the Johansen method reduces to the testing and estimation based on a single ECM equation (and OLS estimation as above)

Constant and other deterministic trends I

- It matters a great deal whether the constant is restricted to be in the cointegrating space or not.
- The advise for data with visible drift in levels:
  - include an deterministic trend as *restricted* together with an unrestricted constant.
  - After rank determination, can test significance of the restricted trend with standard inference
- Shift in levels
  - Include restricted step dummy and a free impulse dummy.
- Exogenous I(1) variables, see table and program by MacKinnon, Haug and Michelis (1999).

I(0) variables in the VAR?

- A misunderstanding that sometimes occurs is that "there can be no stationary variables in he cointegrating relationships". Consider for example:

$$-Y_{1t} + \beta_{12} Y_{2t} + \beta_{13} Y_{3t} + \beta_{14} Y_{4t} = ecm_{1t} \quad (38)$$
$$\beta_{21} Y_{1t} - Y_{2t} + \beta_{23} Y_{3t} + \beta_{24} Y_{4t} = ecm_{2t} \quad (39)$$

  If Y_1 is the log of real-wages, Y_2 productivity, Y_3 relative import prices, and Y_4 the rate of unemployment, then the first relationship may be a bargaining based wage and the second a mark-up equation.
- $Y_{4t} \sim I(0)$, most sensibly, but we want to estimate and test the theory $\beta_{14} = 0$.
- Hence: specify the VAR with Y_{4t} included.

From I(1) to I(0)

- When the rank has been determined, we are back in the stationary-case.
- The distribution of the identified cointegration coefficients are "mixed normal" so that conventional asymptotic inference can be performed on this $\hat{\beta}$.
- The determination of rank allows us to move from the $I(1)$ VAR, to the cointegrated VAR that contains only $I(0)$ variables
- Another name for this $I(0)$ model is the *vector equilibrium correction model*, VECM.
- The VECM can be analyzed further, using the tools of the stationary VAR !
- Hence, cointegration analysis is an important step in the analysis, but just one step.

Cointegration: Summary of estimation and testing I

- Depends on how much we know about

$$\mathbf{\Pi}(1) \equiv \mathbf{\Phi}(1) - \mathbf{I}_N$$

  apriori.
- A "typology" is (simplifying notation: $\mathbf{\Pi}(1) = \mathbf{\Pi}$):

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

Cointegration: Summary of estimation and testing II

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

Some important additional references

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press

Juselius, K (2004) *The Cointegrated VAR Model, Methodology and Applications*, Oxford University Press

MacKinnon, J., A. A. Haug and L. Michelis (1999) Numerical Distributions Functions of Likelihood Ratio Tests for Cointegration, with programs.