Vector Error Correction Model (VECM), Cointegrated VAR

Ercan Karadas

New York University
Department of Economics

Spring, 2018

Content

- **Motivation: plausible economic relations**
- **Model with I(1) variables: spurious regression, bivariate cointegration**
- **Cointegration**
- **Examples: unstable VAR(1), cointegrated VAR(1)**
- **VECM, vector error correction model**
- **Cointegrated VAR models, model structure, estimation, testing, forecasting**
- **Bivariate cointegration**

Resources

- Tsay, Chapter 8.5-6
- Greene, Chapter 21
- Phillips and Ouliaris, Asymptotic properties of residual based tests for cointegration, Econometrica 58, 165-193, 1990.

Motivation

Paths of Dow JC and DAX: 10/2009 - 10/2010

We observe a parallel development. Remarkably this pattern can be observed for single years at least since 1998, though both are assumed to be geometric random walks. They are non stationary, the log-series are I(1).

If a linear combination of I(1) series is stationary, i.e. I(0), the series are called **cointegrated**.
If 2 processes x_t and y_t are both I(1) and

$$y_t - \alpha x_t = \epsilon_t$$

with ϵ_t trend-stationary or simply I(0), then x_t and y_t are called cointegrated.

Cointegration in economics

This concept origins in macroeconomics where series often seen as I(1) are regressed onto, like private consumption, C, and disposable income, Y^d. Despite I(1), Y^d and C cannot diverge too much in either direction:

$$C > Y^d \quad \text{or} \quad C \ll Y^d$$

Or, according to the theory of competitive markets the profit rate of firms *(profits/invested capital)* (both I(1)) should converge to the market average over time. This means that profits should be proportional to the invested capital in the long run.

Common stochastic trend

The idea of cointegration is that there is a common stochastic trend, an I(1) process Z, underlying two (or more) processes X and Y. E.g.

$$X_t = \gamma_0 + \gamma_1 Z_t + \epsilon_t$$

$$Y_t = \delta_0 + \delta_1 Z_t + \eta_t$$

ϵ_t and η_t are stationary, I(0), with mean 0. They may be serially correlated.

Though X_t and Y_t are both I(1), there exists a linear combination of them which is stationary:

$$\delta_1 X_t - \gamma_1 Y_t \sim I(0)$$

Models with I(1) variables

Spurious regression

The spurious regression problem arises if *arbitrarily*

- trending or
- nonstationary

series are regressed on each other.

- In case of (e.g. deterministic) *trending* the spuriously found relationship is due to the trend (growing over time) governing both series instead to economic reasons.
  t-statistic and R^2 are implausibly large.
- In case of *nonstationarity* (of I(1) type) the series - even without drifts - tend to show local trends, which tend to comove along for relative long periods.

Spurious regression: independent I(1)'s

We simulate paths of 2 RWs without drift with independently generated standard normal white noises, ϵ_t, η_t.

$$X_t = X_{t-1} + \epsilon_t, \quad Y_t = Y_{t-1} + \eta_t, \quad t = 1, 2, 3, \ldots, T$$

Then we estimate by LS the model

$$Y_t = \alpha + \beta X_t + \zeta_t$$

In the population $\alpha = 0$ and $\beta = 0$, since X_t and Y_t are independent.
Replications for increasing sample sizes shows that

- the DW-statistics are close to 0. R^2 is too large.
- ζ_t is I(1), nonstationary.
- the estimates are inconsistent.
- the t_β-statistic *diverges* with rate $\sqrt{T}$.

Spurious regression: independence

As both X and Y are independent I(1)s, the relation can be checked consistently using first differences.

$$\Delta Y_t = \beta \, \Delta X_t + \xi_t$$

Here we find that

- $\hat{\beta}$ has the usual distribution around zero,
- the t_β-values are t-distributed,
- the error ξ_t is WN.

Bivariate cointegration

However, if we observe two I(1) processes X and Y, so that the linear combination

$$Y_t = \alpha + \beta X_t + \zeta_t$$

is stationary, i.e. ζ_t is stationary, then

- X_t and Y_t are cointegrated.

When we estimate this model with LS,

- the estimator $\hat{\beta}$ is not only consistent, but **superconsistent.** It converges with the rate T,
  instead of $\sqrt{T}$.
- However, the t_β-statistic is asy normal only if ζ_t is not serially correlated.

Bivariate cointegration: discussion

- The Johansen procedure (which allows for correction for serial correlation easily) (see below) is to be preferred to single equation procedures.
- If the model is extended to 3 or more variables, more than one relation with stationary errors may exist. Then when estimating only a multiple regression, it is not clear what we get.

Cointegration

Definition: Cointegration

Definition: Given a set of I(1) variables $\{x_{1t}, \ldots, x_{kt}\}$. If there exists a linear combination consisting of all vars with a vector $\boldsymbol{\beta}$ so that

$$\beta_1 x_{1t} + \ldots + \beta_k x_{kt} = \boldsymbol{\beta}' \boldsymbol{x}_t \quad \ldots \quad \text{trend-stationary}$$

$\beta_j \neq 0$, $j = 1, \ldots, k$. Then the x's are cointegrated of order CI(1,1).

- $\boldsymbol{\beta}' \boldsymbol{x}_t$ is a (trend-)stationary variable.
- The definition is symmetric in the vars. There is no interpretation of endogenous or exogenous vars. A simultaneous relationship is described.

Definition: Trend-stationarity means that after subtracting a deterministic trend the process is I(0).

Definition: Cointegration (cont)

- $\boldsymbol{\beta}$ is defined only up to a scale.
  If $\boldsymbol{\beta}'\boldsymbol{x}_t$ is trend-stationary, then also $c(\boldsymbol{\beta}'\boldsymbol{x}_t)$ with $c \neq 0$.
  Moreover, any linear combination of cointegrating relationships (stationary variables) is stationary.
- More generally we could consider $\boldsymbol{x} \sim I(d)$ and $\beta'\boldsymbol{x} \sim I(d-b)$ with $b > 0$.
  Then the x's are $\mathrm{CI}(d, b)$.
- We will deal only with the standard case of CI(1,1).

An unstable VAR(1), an example

An unstable VAR(1): $\boldsymbol{x}_t = \Phi_1 \boldsymbol{x}_{t-1} + \epsilon_t$

We analyze in the following the properties of

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & -1. \\ -.25 & 0.5 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

ϵ_t are weakly stationary and serially uncorrelated.

We know a VAR(1) is stable, if the eigenvalues of Φ_1 are less 1 in modulus.

- The eigenvalues of Φ_1 are $\lambda_{1,2} = 0, 1$.
- The roots of the characteristic function $|\boldsymbol{I} - \Phi_1 z| = 0$ should be outside the unit circle for stationarity.
  Actually, the roots are $z = (1/\lambda)$ with $\lambda \neq 0$. $z = 1$.

Φ_1 has a root on the unit circle. So process $\boldsymbol{x}_t$ is not stable.

Remark: Φ_1 is singular; its rank is 1.

Common trend

For all Φ_1 there exists an invertible (i.g. full) matrix $\boldsymbol{L}$ so that

$$\boldsymbol{L}\Phi_1\boldsymbol{L}^{-1} = \Lambda$$

Λ is (for simplicity) diagonal containing the eigenvalues of Φ_1.

We *define* new variables $\quad \boldsymbol{y}_t = \boldsymbol{L}\boldsymbol{x}_t \quad$ and $\quad \eta_t = \boldsymbol{L}\epsilon_t$.
Left multiplication of the VAR(1) with $\boldsymbol{L}$ gives

$$\boldsymbol{L}\boldsymbol{x}_t = \boldsymbol{L}\Phi_1\boldsymbol{x}_{t-1} + \boldsymbol{L}\epsilon_t$$

$$(\boldsymbol{L}\boldsymbol{x}_t) = \boldsymbol{L}\Phi_1\boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{x}_{t-1}) + (\boldsymbol{L}\epsilon_t)$$

$$\boldsymbol{y}_t = \Lambda\boldsymbol{y}_{t-1} + \eta_t$$

Common trend: x's are I(1)

In our case $\boldsymbol{L}$ and $\boldsymbol{\Lambda}$ are

$$\boldsymbol{L} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Then

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}$$

- $\boldsymbol{\eta}_t = \boldsymbol{L}\epsilon_t$: $\quad \eta_{1t}$ and η_{2t} are linear combinations of stationary processes. So they are stationary.
- So also y_{2t} is stationary.
- y_{1t} is obviously integrated of order 1, I(1).

Common trend y_{1t}, x's as function of y_{1t}

$\boldsymbol{y}_t = \boldsymbol{L}\boldsymbol{x}_t$ with $\boldsymbol{L}$ invertible, so we can express $\boldsymbol{x}_t$ in $\boldsymbol{y}_t$.
Left multiplication by $\boldsymbol{L}^{-1}$ gives

$$\boldsymbol{L}^{-1}\boldsymbol{y}_t = \boldsymbol{L}^{-1}\boldsymbol{\Lambda}\boldsymbol{y}_{t-1} + \boldsymbol{L}^{-1}\boldsymbol{\eta}_t$$

$$\boldsymbol{x}_t = (\boldsymbol{L}^{-1}\boldsymbol{\Lambda})\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t$$

$\boldsymbol{L}^{-1} = \ldots$

$$\begin{aligned} x_{1t} &= (1/2)y_{1,t-1} + \epsilon_{1t} \\ x_{2t} &= -(1/4)y_{1,t-1} + \epsilon_{2t} \end{aligned}$$

- Both x_{1t} and x_{2t} are I(1), since y_{1t} is I(1).
- y_{1t} is called the **common trend** of x_{1t} and x_{2t}. It is the common nonstationary component in both x_{1t} and x_{2t}.

Cointegrating relation

Now we eliminate $y_{1,t-1}$ in the system above by multiplying the 2nd equation by 2 and adding to the first.

$$x_{1t} + 2x_{2t} = (\epsilon_{1,t} + 2\epsilon_{2,t})$$

This gives a stationary process, which is called the **cointegrating relation.** This is the only linear combination (apart from a factor) of both nonstationary processes, which is stationary.

A cointegrated VAR(1) example

A cointegrated VAR(1)

We go back to the system and proceed directly.

$$\boldsymbol{x}_t = \boldsymbol{\Phi}_1 \boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t$$

and subtract $\boldsymbol{x}_{t-1}$ on both sides (cp. the Dickey-Fuller statistic).

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} -.5 & -1. \\ -.25 & -.5 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

The coefficient matrix $\boldsymbol{\Pi}$, $\boldsymbol{\Pi} = -(\boldsymbol{I} - \boldsymbol{\Phi}_1)$, in

$$\Delta \boldsymbol{x}_t = \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t$$

has only rank 1. It is singular.

Then $\boldsymbol{\Pi}$ can be factorized as

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

$$(2 \times 2) = (2 \times 1)(1 \times 2)$$

A cointegrated VAR(1)

k the number of endogenous variables, here $k = 2$.
$m = \mathrm{Rank}(\mathbf{\Pi}) = 1$, is the number of cointegrating relations.

A solution for $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is

$$\begin{bmatrix} -.5 & -1. \\ -.25 & -.5 \end{bmatrix} = \begin{pmatrix} -.5 \\ -.25 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix}' = \begin{pmatrix} -.5 \\ -.25 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix}$$

Substituted in the model

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{pmatrix} -.5 \\ -.25 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

A cointegrated VAR(1)

Multiplying out

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{pmatrix} -.5 \\ -.25 \end{pmatrix} \begin{pmatrix} x_{1,t-1} + 2x_{2,t-1} \end{pmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

The component $(x_{1,t-1} + 2x_{2,t-1})$ appears in both equations.
As the lhs variables and the errors are stationary, this linear combination is stationary.
This component is our **cointegrating relation** from above.

Vector error correction, VEC

VECM, vector error correction model

Given a VAR(p) of I(1) x's (ignoring consts and determ trends)

$$\boldsymbol{x}_t = \boldsymbol{\Phi}_1 \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t$$

There always exists an **error correction** representation of the form (trick $\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \Delta \boldsymbol{x}_t$)

$$\Delta \boldsymbol{x}_t = \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\Pi}$ and the $\boldsymbol{\Phi}^*$ are functions of the $\boldsymbol{\Phi}$'s. Specifically,

$$\boldsymbol{\Phi}_j^* = -\sum_{i=j+1}^{p} \boldsymbol{\Phi}_i, \quad j = 1, \ldots, p-1$$

$$\boldsymbol{\Pi} = -(\boldsymbol{I} - \boldsymbol{\Phi}_1 - \ldots - \boldsymbol{\Phi}_p) = -\boldsymbol{\Phi}(1)$$

The characteristic polynomial is $\quad \boldsymbol{I} - \boldsymbol{\Phi}_1 z - \ldots - \boldsymbol{\Phi}_p z^p = \boldsymbol{\Phi}(z)$.

Interpretation of $\Delta \boldsymbol{x}_t = \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \epsilon_t$

- If $\boldsymbol{\Pi} = \mathbf{0}$, (all $\lambda(\boldsymbol{\Pi}) = 0$) then there is no cointegration. Nonstationarity of I(1) type vanishes by taking differences.
- If $\boldsymbol{\Pi}$ has full rank, k, then the x's cannot be I(1) but are stationary. $(\boldsymbol{\Pi}^{-1} \Delta \boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Pi}^{-1} \epsilon_t)$
- The interesting case is, $\mathrm{Rank}(\boldsymbol{\Pi}) = m, 0 < m < k$, as this is the case of cointegration. We write

$$\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$$
$$(k \times k) = (k \times m)[(k \times m)']$$

where the columns of $\boldsymbol{\beta}$ contain the *m cointegrating* vectors, and the columns of $\boldsymbol{\alpha}$ the *m adjustment* vectors.

$$\mathrm{Rank}(\boldsymbol{\Pi}) = \min[\,\mathrm{Rank}(\boldsymbol{\alpha}),\ \mathrm{Rank}(\boldsymbol{\beta})\,]$$

Long term relationship in $\Delta \boldsymbol{x}_t = \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \epsilon_t$

There is an adjustment to the 'equilibrium' $\boldsymbol{x}^*$ or long term relation described by the cointegrating relation.

- Setting $\Delta \boldsymbol{x} = \mathbf{0}$ we obtain the long run relation, i.e.

$$\boldsymbol{\Pi} \boldsymbol{x}^* = \mathbf{0}$$

This may be wirtten as

$$\boldsymbol{\Pi} \boldsymbol{x}^* = \alpha(\beta' \boldsymbol{x}^*) = \mathbf{0}$$

In the case $0 < \mathrm{Rank}(\boldsymbol{\Pi}) = \mathrm{Rank}(\alpha) = m < k$ the number of equations of this system of linear equations which are different from zero is m.

$$\beta' \boldsymbol{x}^* = \mathbf{0}_{m \times 1}$$

Long term relationship

- The long run relation does not hold perfectly in $(t-1)$. There will be some deviation, an *error*,
$$\beta' \boldsymbol{x}_{t-1} = \xi_{t-1} \neq \boldsymbol{0}$$
- The adjustment coefficients in α multiplied by the 'errors' $\beta' \boldsymbol{x}_{t-1}$ induce adjustment. They determine Δx_t, so that the x's move in the correct direction in order to bring the system back to 'equilibrium'.

Adjustment to deviations from the long run

- The long run relation is in the example above

$$x_{1,t-1} + 2x_{2,t-1} = \xi_{t-1}$$

  ξ_t is the stationary error.

- The adjustment of $x_{1,t}$ in t to ξ_{t-1}, the deviation from the long run in $(t-1)$, is

$$\Delta x_{1,t} = (-.5)\xi_{t-1} \quad \text{and} \quad x_{1,t} = \Delta x_{1,t} + x_{1,t-1}$$

- If $\xi_{t-1} > 0$, the error is positive, i.e. $x_{1,t-1}$ is too large c.p., then $\Delta x_{1,t}$, the change in x_1, is negative. x_1 decreases to guarantee convergence back to the long run path.
- Similar for $x_{2,t}$ in the 2nd equation.

Cointegrated VAR models (CIVAR)

Model

We consider a VAR(p) with $\boldsymbol{x}_t$ I(1), (unit root) *nonstationary*.

$$\boldsymbol{x}_t = \phi + \boldsymbol{\Phi}_1 \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{x}_{t-p} + \epsilon_t$$

Then

- $\Delta \boldsymbol{x}_t$ is I(0).
- $\boldsymbol{\Pi} = -\boldsymbol{\Phi}(1)$ is singular, i.e. $|\boldsymbol{\Phi}(1)| = 0$

(For weakly stationarity, I(0): $|\boldsymbol{\Phi}(z)| = 0$ only for $|z| > 1$.)

The VEC representation reads with $\boldsymbol{\Pi} = \alpha\beta'$

$$\Delta \boldsymbol{x}_t = \phi + \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \epsilon_t$$

$\boldsymbol{\Pi} \boldsymbol{x}_{t-1}$ is called the **error-correction term.**

3 cases

We distinguish 3 cases for $\mathrm{Rank}(\mathbf{\Pi}) = m$:

I. $m = 0$: $\mathbf{\Pi} = \mathbf{0}$ (all $\lambda(\mathbf{\Pi}) = 0$)

II. $0 < m < k$: $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, $\boldsymbol{\alpha}_{(k \times m)}$, $(\boldsymbol{\beta}')_{(m \times k)}$

III. $m = k$: $|\mathbf{\Pi}| = |-\mathbf{\Phi}(1)| \neq 0$!

I. $\mathrm{Rank}(\mathbf{\Pi}) = 0, \quad m = 0 \quad (\text{all } \lambda(\mathbf{\Pi}) = 0)$:

In case of $\mathrm{Rank}(\mathbf{\Pi}) = 0$, i.e. $m = 0$, it follows

- $\mathbf{\Pi} = \mathbf{0}$, the null matrix.
- There does not exist a linear combination of the I(1) vars, which is stationary.
- The x's are not cointegrated.
- The EC form reduces to a stationary VAR($p - 1$) in differences.

$$\Delta \boldsymbol{x}_t = \phi + \sum_{i=1}^{p-1} \mathbf{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \epsilon_t$$

- $\mathbf{\Pi}$ has $m = 0$ eigenvalues different from 0.

II. $\mathrm{Rank}(\boldsymbol{\Pi}) = m, \;\; 0 < m < k$:

The rank of $\boldsymbol{\Pi}$ is m, $m < k$. We factorize $\boldsymbol{\Pi}$ in two rank m matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$.
$\mathrm{Rank}(\boldsymbol{\alpha}) = \mathrm{Rank}(\boldsymbol{\beta}) = m$.
Both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(k \times m)$.

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \neq \mathbf{0}$$

The VEC form is then

$$\Delta \boldsymbol{x}_t = \phi + \boldsymbol{\alpha}\boldsymbol{\beta}' \, \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \epsilon_t$$

- The x's are integrated, I(1).
- There are m eigenvalues $\lambda(\boldsymbol{\Pi}) \neq 0$.
- The x's are cointegrated. There are m linear combinations, which are stationary.

II. $\text{Rank}(\boldsymbol{\Pi}) = m, \;\; 0 < m < k$:

- There are m linear independent cointegrating (column) vectors in $\boldsymbol{\beta}$.
- The m stationary linear combinations are $\boldsymbol{\beta}' \boldsymbol{x}_t$.
- $\boldsymbol{x}_t$ has $(k - m)$ unit roots, so $(k - m)$ *common stochastic trends*.

There are

- k I(1) variables,
- m cointegrating relations (eigenvalues of $\boldsymbol{\Pi}$ different from 0), and
- $(k - m)$ stochastic trends.

$$k = m + (k - m)$$

III. $\text{Rank}(\mathbf{\Pi}) = m, \quad m = k$:

Full rank of $\mathbf{\Pi}$ implies

- that $|\mathbf{\Pi}| = |-\mathbf{\Phi}(1)| \neq 0$.
- $\boldsymbol{x}_t$ has no unit root. That is $\boldsymbol{x}_t$ is I(0).
- There are $(k - m) = 0$ stochastic trends.
- As consequence we model the relationship of the x's in levels, not in differences.
- There is no need to refer to the error correction representation.

II. $\mathrm{Rank}(\boldsymbol{\Pi}) = m, \;\; 0 < m < k$: (cont) common trends

A general way to obtain the $(k-m)$ common trends is to use the *orthogonal complement matrix* $\boldsymbol{\alpha}_\perp$ of $\boldsymbol{\alpha}$.

$$\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha} = \mathbf{0}$$
$$\{k \times (k-m)\}'\{k \times m\} = \{(k-m) \times m\}$$

If the ECM is left multiplied by $\alpha'_\perp$ the error correction term vanishes,

$$\boldsymbol{\alpha}'_\perp \boldsymbol{\Pi} = (\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha})\boldsymbol{\beta}' = \mathbf{0}_{(k-m)\times k}$$

with $\boldsymbol{\alpha}'_\perp \Delta \boldsymbol{x}_t = \Delta(\boldsymbol{\alpha}'_\perp \boldsymbol{x}_t)$

$$\Delta(\boldsymbol{\alpha}'_\perp \boldsymbol{x}_t) = (\boldsymbol{\alpha}'_\perp \phi) + \sum_{i=1}^{p-1} \boldsymbol{\Phi}^*_i \Delta(\boldsymbol{\alpha}'_\perp \boldsymbol{x}_{t-i}) + (\boldsymbol{\alpha}'_\perp \epsilon_t)$$

II. $\mathrm{Rank}(\Pi) = m, \;\; 0 < m < k$: (cont) common trends

The resulting system is a $(k - m)$ dimensional system of first differences, corresponding to $(k - m)$ independent RWs

$$\boldsymbol{\alpha}_{\perp}' \boldsymbol{x}_t$$

which are the common trends.

Example (from above): $\boldsymbol{\alpha} = (-1, -.5)'$ then $\boldsymbol{\alpha}_{\perp} = (1, -2)'$.

Non uniqueness of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ in $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$

For any orthogonal matrix $\boldsymbol{\Omega}_{m\times m}$, $\boldsymbol{\Omega}\boldsymbol{\Omega}' = \boldsymbol{I}$,

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\,\boldsymbol{\Omega}\boldsymbol{\Omega}'\,\boldsymbol{\beta}' = (\boldsymbol{\alpha}\boldsymbol{\Omega})(\boldsymbol{\beta}\boldsymbol{\Omega})' = \boldsymbol{\alpha}^*(\boldsymbol{\beta}^*)'$$

where both $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ are of rank m.

Usually the structure

$$\boldsymbol{\beta}' = [\boldsymbol{I}_{m\times m},\ (\boldsymbol{\beta}_1')_{m\times(k-m)}]$$

is imposed.
Each of the first m variables belong only to one equation and their coeffs are 1.

Economic interpretation is helpful when structuring $\boldsymbol{\beta}'$. Also, a reordering of the vars might be necessary.

Inclusion of deterministic functions

There are several possibilities to specify the deterministic part, ϕ, in the model.

1. $\phi = \mathbf{0}$: All components of $\boldsymbol{x}_t$ are I(1) without drift. The stationary series $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$ has a zero mean.
2. $\phi = (\phi_0)_{k\times 1} = \boldsymbol{\alpha}_{k\times m}\,\boldsymbol{c}_{0,m\times 1}$: This is the special case of a restricted constant. The ECM is
$$\Delta\boldsymbol{x}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{c}_0) + \ldots$$
$\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$ has a mean of $(-\boldsymbol{c}_0)$.
There is only a constant in the cointegrating relation, but the x's are I(1) without a drift.
3. $\phi = \phi_0 \neq \mathbf{0}$: The x's are I(1) with drift. The coint rel may have a nonzero mean. Intercept ϕ_0 may be spilt in a drift component and a const vector in the coint eq's.

Inclusion of deterministic functions

4. $\phi = \phi_t = \phi_0 + (\alpha \boldsymbol{c}_1)t$:
   Analogous, ϕ_0 enters the drift of the x's. $\boldsymbol{c}_1$ becomes the trend in the coint rel.

$$\Delta \boldsymbol{x}_t = \phi_0 + \alpha(\beta' \boldsymbol{x}_{t-1} + \boldsymbol{c}_1 t) + \ldots$$

5. $\phi = \phi_t = \phi_0 + \phi_1 t$:
   Both constant and slope of the trend are unrestricted. The trending behavior in the x's is determined both by a drift and a quadratic trend.
   The coint rel may have a linear trend.

Case 3, $\phi = \phi_0$, is relevant for asset prices.

Remark: The assignment of the const to either intercept or coint rel is not unique.

ML estimation: Johansen (1)

Estimation is a 3-step procedure:

- *1st step:* We start with the VEC representation and extract the effects of the lagged $\Delta \boldsymbol{x}_{t-j}$ from the lhs $\Delta \boldsymbol{x}_t$ and from the rhs $\boldsymbol{x}_{t-1}$. (Cp. Frisch-Waugh). This gives the residuals $\hat{\boldsymbol{u}}_t$ for $\Delta \boldsymbol{x}_t$ and $\hat{\boldsymbol{v}}_t$ for $\boldsymbol{x}_{t-1}$, and the model

$$\hat{\boldsymbol{u}}_t = \boldsymbol{\Pi} \hat{\boldsymbol{v}}_t + \boldsymbol{\epsilon}_t$$

- *2nd step:* All variables in the cointegration relation are dealt with symmetrically. There are no endogenous and no exogeneous variables. We view this system as

$$(\tilde{\alpha})^{-1} \boldsymbol{u}_t = \tilde{\boldsymbol{\beta}}' \boldsymbol{v}_t$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are $(k \times k)$. The solution is obtained by *canonical correlation*.

Johansen (2): canonical correlation

- We determine *vectors* $\check{\alpha}_j$, $\check{\beta}_j$ so that the linear combinations

$$\check{\alpha}_j' \boldsymbol{u}_t \quad \text{and} \quad \check{\beta}_j' \boldsymbol{v}_t$$

  correlate
  - maximal for $j = 1$,
  - maximal subjcet to orthogonality wrt the solution for $j = 1$ ($\rightarrow j = 2$),
  - etc.

For the largest correlation we get a largest eigenvalue, λ_1, for the second largest a smaller one, $\lambda_2 < \lambda_1$, etc. The eigenvalues are the *squared* (canonical) correlation coefficients.

The columns of β are the associated normalized eigenvectors.

The λ's are *not* the eigenvalues of $\mathbf{\Pi}$, but have the same zero/nonzero properties.

Johansen (2)

Actually we solve a generalized eigenvalue problem

$$|\lambda \boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}| = 0$$

with the sample covariance matrices

$$S_{00} = \frac{1}{T-p}\sum \hat{\boldsymbol{u}}_t \hat{\boldsymbol{u}}_t', \qquad S_{01} = \frac{1}{T-p}\sum \hat{\boldsymbol{u}}_t \hat{\boldsymbol{v}}_t'$$

$$S_{11} = \frac{1}{T-p}\sum \hat{\boldsymbol{v}}_t \hat{\boldsymbol{v}}_t'$$

The number of eigenvalues λ larger 0 *determines* the rank of $\boldsymbol{\beta}$, resp. $\boldsymbol{\Pi}$, and so the number of cointegrating relations:

$$\lambda_1 > \ldots > \lambda_m > 0 = \ldots = 0 = \lambda_k$$

Johansen (3)

3rd step: In this final step the adjustment parameters $\boldsymbol{\alpha}$ and the $\boldsymbol{\Phi}^*$'s are estimated.

$$\Delta \boldsymbol{x}_t = \phi + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \boldsymbol{\epsilon}_t$$

The maximized likelihood function based on m cointegrating vectors is

$$L_{max}^{-2/T} \propto |S_{00}| \prod_{i=1}^{m} (1 - \hat{\lambda}_i)$$

Under Gaussian innovations and the model is true, the estimates of the $\boldsymbol{\Phi}_j^*$ matrices are **asy normal and asy efficient.**

Remark: S_{00} depends only on $\Delta \boldsymbol{x}_t$ and $\Delta \boldsymbol{x}_{t-j}$, $j = 1, \ldots, p$.

Test for cointegration: trace test

Given the specification of the deterministic term we test for the rank m of $\mathbf{\Pi}$. There are 2 *sequential tests*

the trace test, and

the maximum eigenvalue test.

- **trace test**:

$$H_0 : \mathrm{Rank}(\mathbf{\Pi}) = m \quad \text{against} \quad H_A : \mathrm{Rank}(\mathbf{\Pi}) > m$$

The likelihood ratio statistic is

$$LK_{tr}(m) = -(T - p) \sum_{i=m+1}^{k} \ln(1 - \hat{\lambda}_i)$$

We start with $m = 0$ – that is $\mathrm{Rank}(\mathbf{\Pi}) = 0$, there is no cointegration – against $m \geq 1$, that there is at least one coint rel. Etc.

Test for cointegration: trace test

$LK_{tr}(m)$ takes large values (i.e. H_0 is rejected) when the 'sum' of the remaining eigenvalues $\lambda_{m+1} \geq \lambda_{m+2} \geq \ldots \geq \lambda_k$ is large.

If λ is

- large (say ≈ 1), then $\quad -\ln(1-\hat{\lambda}_i) \quad$ is large.
- small (say ≈ 0), then $\quad -\ln(1-\hat{\lambda}_i) \approx 0$.

Test for cointegration: max eigenvalue statistic

- **maximum eigenvalue test**:

  $$H_0 : \mathrm{Rank}(\mathbf{\Pi}) = m \quad \text{against} \quad H_A : \mathrm{Rank}(\mathbf{\Pi}) = m + 1$$

  The statistic is

  $$LK_{max}(m) = -(T - p)\ln(1 - \hat{\lambda}_{m+1})$$

  We start with $m = 0$ – that is $\mathrm{Rank}(\mathbf{\Pi}) = 0$, there is no cointegration – against $m = 1$, that there is one coint rel. Etc.

In case we reject $m = k - 1$ coint rel, we should have to conclude that there are $m = k$ coint rel. But this would not fit to the assumption of I(1) vars.

The critical values of both test statistics are nonstandard and are obtained via Monte Carlo simulation.

Forecasting, summary

The fitted ECM can be used for forecasting $\Delta \boldsymbol{x}_{t+\tau}$. The forecasts of $\boldsymbol{x}_{t+\tau}$ (τ-step ahead) are obtained recursively.

$$\hat{\boldsymbol{x}}_{t+\tau} = \widehat{\Delta \boldsymbol{x}}_{t+\tau} + \hat{\boldsymbol{x}}_{t+\tau-1}$$

A summary:

- If all vars are stationary / the VAR is stable, the adequate model is a VAR in levels.
- If the vars are integrated of order 1 but not cointegrated, the adequate model is a VAR in first differences (no level components included).
- If the vars are integrated and cointegrated, the adequate model is a cointegrated VAR. It is estimated in the first differences with the cointegrating relations (the levels) as explanatory vars.

Bivariate Cointegration

Estimation and testing: Engle and Granger

- **Engle-Granger**: $\boldsymbol{x}_t, y_t \;\sim\; I(1)$

$$y_t = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + u_t$$

MacKinnon has tabulated critical values for the test of the LS residuals $\hat{u}_t$ under the null of no cointegration (of a unit root), similar to the augmented Dickey-Fuller test.

$$H_0: \; u_t \sim \mathrm{I}(1), \text{ no coint} \qquad H_A: \; u_t \sim \mathrm{I}(0), \text{ coint}$$

The test distribution depends on the inclusion of an intercept or a trend. Additional lagged differences may be used.

If u is stationary, x's and y are cointegrated.

Phillips-Ouliaris test

- **Phillips-Ouliaris**: Two residuals are compared.
  $\hat{u}_t$ from the Engle-Granger test and $\hat{\xi}_t$ from

  $$\boldsymbol{z}_t = \boldsymbol{\Pi}\boldsymbol{z}_{t-1} + \boldsymbol{\xi}_t$$

  estimated via LS, where $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t')'$.

  $\hat{\xi}_{1,t}$ is stationary, $\hat{u}_t$ only if the vars are cointegrated.
  Intuitively the ratio $(s^2_{\xi_1}/s^2_u)$ is small under no coint and large under coint (due to the superconsistency associated with s^2_u).

  $$H_0: \text{ no coint} \qquad H_A: \text{ coint}$$

  Two test statisticis $\hat{P}_u$ and $\hat{P}_z$ are available in `ca.po {urca}`.

Remark: If z_t is a RW, then $z_t = 1z_{t-1} + \xi_t$ and ξ_t stationary.