**Università di Pavia**

# Cointegrated VARs 2

Eduardo Rossi

The normalization ensures identified parameters $\boldsymbol{\beta}_{(n-r)}$ and thus inference becomes possible.
The ML estimator of $\boldsymbol{\beta}_{(n-r)}$ may be obtained from the ML estimator of $\boldsymbol{\beta}$. By denoting the first $r$ rows of $\widehat{\boldsymbol{\beta}}$ by $\widehat{\boldsymbol{\beta}}_{(n-r)}$.

If uniqueness restrictions are imposed it can be shown that $T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{T}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ converge in distribution. The estimator of $\boldsymbol{\beta}$ converges with the fast rate $T$ and is called *superconsistent*.

The estimators for the parameters $\boldsymbol{\beta}_{(n-r)}$ have an asymptotic distribution that is multivariate normal upon appropriate normalization. Partitioning $\mathbf{y}_t$:

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{y}_t^{(2)} \end{bmatrix} \quad \begin{matrix} (r \times 1) \\ ((n-r) \times 1) \end{matrix}$$

Defining

$$\mathbf{Y}_{-1}^{(2)} = \left[\mathbf{y}_0^{(2)}, \ldots, \mathbf{y}_{T-1}^{(2)}\right]$$

$$vec\left\{\left(\widehat{\boldsymbol{\beta}}'_{(n-r)} - \boldsymbol{\beta}'_{(n-r)}\right)\left(\mathbf{Y}_{-1}^{(2)}\mathbf{M}\mathbf{Y}_{-1}^{(2)'}\right)^{1/2}\right\}$$
$$= \left[\left(\mathbf{Y}_{-1}^{(2)}\mathbf{M}\mathbf{Y}_{-1}^{(2)'}\right)^{1/2} \otimes \mathbf{I}_{(n-r)}\right] \mathrm{vec}\left(\widehat{\boldsymbol{\beta}}'_{(n-r)} - \boldsymbol{\beta}'_{(n-r)}\right)$$
$$\xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{(n-r)} \otimes \left(\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}\right)^{-1})$$

$$\mathrm{vec}\left(\widehat{\boldsymbol{\beta}}'_{(n-r)}\right) \approx N\left(\mathrm{vec}\left(\boldsymbol{\beta}'_{(n-r)}\right), \left(\mathbf{Y}_{-1}^{(2)}\mathbf{M}\mathbf{Y}_{-1}^{(2)'}\right)^{-1} \otimes \left(\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}\right)^{-1}\right)$$

t-ratios are obtained by dividing the elements of $\widehat{\boldsymbol{\beta}}_{(n-r)}$ by

$$\sqrt{\mathbf{u}_i' \left[ \left( \mathbf{Y}_{-1}^{(2)} \mathbf{M} \mathbf{Y}_{-1}^{(2)\prime} \right)^{-1} \otimes (\boldsymbol{\alpha}' \boldsymbol{\Omega}^{-1} \boldsymbol{\alpha})^{-1} \right] \mathbf{u}_i}$$

For a given fixed $(J \times r(n-r))$ matrix $\mathbf{R}$, with $rk(\mathbf{R}) = J$, and $\mathbf{r}$, $(J \times 1)$

$$H_0 : \mathbf{R}\text{vec}(\boldsymbol{\beta}'_{(n-r)}) = \mathbf{r}$$

The Wald test statistic has an asymptotic distribution $\chi^2$ distribution with $J$ degrees of freedom under $H_0$:

$$\xi_W = \left[ \left( \mathbf{R}\text{vec}(\boldsymbol{\beta}'_{(n-r)}) - \mathbf{r} \right)' (\mathbf{R}\widehat{\boldsymbol{\Omega}}\mathbf{R}')^{-1} \left( \mathbf{R}\text{vec}(\boldsymbol{\beta}'_{(n-r)}) - \mathbf{r} \right) \right] \xrightarrow{d} \chi^2_J$$

- Determining the Autoregressive order
- Specifying the cointegrating rank

Determining the lag order:

1. Sequential testing procedures
2. Model selection criteria (information criteria)

Fit VAR(m) models with orders $m = 0, 1, \ldots, p_{max}$ and choose an estimator that minimizes the preferred criterion.
The general form of the criteria:

$$Cr(m) = \log|\widehat{\mathbf{\Omega}}(m)| + c_T\varphi(m)$$

$$\widehat{\mathbf{\Omega}}(m) = T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{\epsilon}}_t\widehat{\boldsymbol{\epsilon}}_t'$$

the residual covariance matrix estimator for a model of order $m$, $c_T$ is a sequence that depends on $T$, $\varphi(m)$ is a function that penalizes large $m$.

The term $\log|\widehat{\mathbf{\Omega}}(m)|$ measures the fit of the model with order $m$. This decreases when $m$ increases.

$$
\begin{aligned}
AIC(m) &= \log|\widehat{\mathbf{\Omega}}(m)| + \frac{2}{T}mn^2 \\
HQ(m) &= \log|\widehat{\mathbf{\Omega}}(m)| + \frac{2\log\log T}{T}mn^2 \\
SC(m) &= \log|\widehat{\mathbf{\Omega}}(m)| + \frac{\log T}{T}mn^2
\end{aligned}
$$

The AIC asymptotically overestimates the order with positive probability. HQ and SC estimate the order consistently under quite general conditions if the actual DGP has a finite VAR order and the maximum order is larger than the true order.

These results not only hold for $I(0)$ processes but also for $I(1)$ processes with cointegrated variables (Paulsen (1984)).

For $T \geq 16$, the orders selected by the three criteria

$$\widehat{p}(SC) \leq \widehat{p}(HQ) \leq \widehat{p}(AIC)$$

VECM without deterministic part:

$$\Delta \mathbf{y}_t = \mathbf{\Pi} \mathbf{y}_{t-1} + \mathbf{\Gamma}_1 \Delta \mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t$$

The cointegrating rank $r$ has to be chosen in addition to the lag-order.

The log-likelihood function is:

$$-2 \log L(\boldsymbol{\beta}) = T \log |\mathbf{S}_{00}| + T \sum_{i=1}^{n} \log (1 - \lambda_i)$$

$\lambda_i, i = 1, \ldots, n$ can be found by solving the eigenvalue problem

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

$$\text{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r) = \widehat{\boldsymbol{\beta}}' \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{01} \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\alpha}}' \mathbf{S}_{00}^{-1} \widehat{\boldsymbol{\alpha}}$$

When $\lambda_i = 0$, $\boldsymbol{\beta}_i' \mathbf{y}_t$ is non stationary and there is no equilibrium correction, $\boldsymbol{\alpha}_i$.
Test procedure to discriminate between those $\lambda_i, i = 1, \ldots, r$ which correspond to stationary relations.
The maximum is exclusively a function of the non-zero eigenvalue

$$L^{-2/T} = |\mathbf{S}_{00}| \prod_{i=1}^{r} (1 - \lambda_i)$$

**Trace test.** LR test for the determination of the cointegration of the cointegration rank. Suppose we wish to test

$$H_0 : \text{rank}(\boldsymbol{\Pi}) = r_0 \qquad \text{against} \qquad H_1 : r_0 < \text{rank}(\boldsymbol{\Pi}) \leq n$$

The LR test statistic

$$
\begin{aligned}
\mathcal{LR}(r_0, r_1) &= 2[L(r_1) - L(r_0)] \\
&= T\left[-\sum_{i=1}^{r_1} \log(1-\lambda_i) + \sum_{i=1}^{r_0} \log(1-\lambda_i)\right] \\
&= -T \sum_{i=r_0+1}^{r_1} \log(1-\lambda_i)
\end{aligned}
$$

**Example.** $n = 5$

$$H_0 : \text{rank}(\mathbf{\Pi}) = 2, n - r = 3$$

$$H_1 : \text{rank}(\mathbf{\Pi}) = 5, n - r = 0$$

$$\mathcal{LR}(2,5) = -T\{\log(1-\lambda_3) + \log(1-\lambda_4) + \log(1-\lambda_5)\}$$

the test LR of $\lambda_3 = \lambda_4 = \lambda_5 = 0$ is correctly accepted also when $\lambda_2 = 0$ or $\lambda_2 = \lambda_1 = 0$.
If $H_0$ is accepted we conclude that there are al least three unit roots and at most two stationary relations.

The asymptotic distribution of the $\mathcal{LR}$ under the $H_0$ for given $r_0$ and $r_1$ is nonstandard. It is not a $\chi^2$-distribution.
It depends on the number of common trends $n - r_0$ under $H_0$ and on the alternative hypothesis.

If $\mathcal{LR}(n, r^*) > C_{n-r^*}$ we reject the hypothesis of $n - r^*$ unit roots (common trends) in the model, and conclude that there are fewer unit roots than assumed.
If $\mathcal{LR}(n, r^*) < C_{n-r^*}$ we accept the hypothesis of at least $n - r^*$ unit roots in the model but conclude that there may be more.

$$H_0 : \text{rank}(\mathbf{\Pi}) = r_0 \qquad \text{against} \qquad H_1 : r_0 + 1$$

with $\mathcal{LR}(r_0, r_0 + 1)$ is called the *maximum eigenvalue statistic*.

Sequential procedures based on LR-type tests. Top-bottom. The sequence of hypotheses:

$$\begin{array}{lcl} H_0(0): rk(\mathbf{\Pi}) = r = 0 & vs & H_1(0): rk(\mathbf{\Pi}) = r > 0 \\ H_0(1): rk(\mathbf{\Pi}) = r = 1 & vs & H_1(1): rk(\mathbf{\Pi}) = r > 1 \\ \vdots & \vdots & \vdots \\ H_0(n-1): rk(\mathbf{\Pi}) = r = n-1 & vs & H_1(n-1): rk(\mathbf{\Pi}) = r = n \end{array}$$

The testing procedures terminates when the null hypothesis cannot be rejected for the first time.

If $H_0(0)$ cannot be rejected a VAR in first differences is considered.

If $H_0(n-1)$ cannot be rejected a levels VAR should be considered.

Under Gaussian assumptions the LR statistic under $H_0(r_0)$ is nonstandard. It depends on the difference $n - r_0$. The deterministic trend terms and shift dummy variables in the DGP have an impact on the distribution of the LR test statistic under the null.

LR-type tests have been derived under different assumptions regarding the deterministic trend. On the assumption that the lag order is specified correctly, the limiting null distributions do not depend on the short-term dynamics.

Consider the model

$$\mathbf{y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \mathbf{x}_t$$

$\mathbf{x}_t \sim \text{VAR}(p)$. Three cases of practical interest:

- $\boldsymbol{\mu}_1 = \mathbf{0}$, only constant mean no deterministic trend term.
- $\boldsymbol{\mu}_1 \neq \mathbf{0}$, linear trend confined to some individual variables but is absent from the cointegrating relations, $\boldsymbol{\beta}'\boldsymbol{\mu}_1 = \mathbf{0}$
- fully unrestricted linear trend.

$\boldsymbol{\mu}_1 = \mathbf{0}$

$$\mathbf{y}_t - \boldsymbol{\mu}_0 = \mathbf{x}_t$$

$$\Delta \mathbf{y}_t = \Delta \mathbf{x}_t$$

from the VECM of $\mathbf{x}_t$, the mean adjusted $\mathbf{y}_t$ have the VECM form

$$\Delta \mathbf{y}_t = \mathbf{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_0) + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

or if an intercept term is used

$$\begin{aligned} \Delta \mathbf{y}_t &= \boldsymbol{\nu}_0^* + \mathbf{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_0) + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \mathbf{\Pi}^* \begin{bmatrix} \mathbf{y}_{t-1} \\ 1 \end{bmatrix} + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \end{aligned}$$

where

$$\mathbf{\Pi}^* = \left[ \begin{array}{ccc} \mathbf{\Pi} & | & \boldsymbol{\nu}_0^* \end{array} \right] : (n \times (n+1))$$

$$\boldsymbol{\nu}_0^* = -\mathbf{\Pi}\boldsymbol{\mu}_0$$

It follows from the absence of an intercept term that the intercept can be absorbed into the cointegrating relations

$$\mathbf{\Pi}^* = \boldsymbol{\alpha}\boldsymbol{\beta}^*$$

has rank $r$.
Both versions can be used for testing the cointegrating rank.
Johansen (1995) considers the intercept version and provides critical values for the LR test, which is known as the *trace test*.

$\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, linear deterministic trend.
If the trend is confined to some individual variables but is absent from the cointegrating relations, the cointegrating relations are drifting along a common linear trend. This situation can arise if *the trend slope is the same for all variables which have a linear trend.* This occurs if

$$\boldsymbol{\beta}'\boldsymbol{\mu}_1 = 0$$

$$\boldsymbol{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1(t-1)) = \boldsymbol{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_0)$$

using the VECM form of $\mathbf{x}_t = \mathbf{y}_t - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 t$

$$\Delta(\mathbf{y}_t - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 t) = \boldsymbol{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1(t-1)) + \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j \Delta(\mathbf{y}_{t-j} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1(t-j)) + \boldsymbol{\epsilon}_t$$

Structural shift in the level of DGP captured by adding variables to the deterministic part of the process leads to a change in the asymptotic distributions of the Johansen-type tests for the cointegaritng rank.

The small sample properties of the tests depend quite strongly on the lag order chosen. Working with a low-order model that does not capture the serial dependence in the data may lead to size distortions, large lag order may spoil the power of the tests.

It is a good strategy to perform tests for different lag orders and to check the robustness of the results.

Doornik, Hendry and Nielsen (1998) small sample and asymptotic evidence that, not taking into account a deterministic trend actually present in the DGP results in size distortions. Including an unnecessary trend term may result in a loss of power.

Johansen (1994, 1995) proposed test regarding the deterministic trends in the Gaussian framework.

| $H_0$ | $H_1$ | asymp. distribution |
|---|---|---|
| $\boldsymbol{\mu}_0$ arbitrary, $\boldsymbol{\mu}_1 = \mathbf{0}$ | $\boldsymbol{\mu}_0$arbitrary, $\boldsymbol{\mu}_1 \neq \mathbf{0}, \boldsymbol{\beta}'\boldsymbol{\mu}_1 = \mathbf{0}$ | $\chi^2_{n-r}$ |
| $\boldsymbol{\mu}_0$ arbitrary, $\boldsymbol{\mu}_1 \neq \mathbf{0}, \boldsymbol{\beta}'\boldsymbol{\mu}_1 = \mathbf{0}$ | $\boldsymbol{\mu}_0$arbitrary, $\boldsymbol{\mu}_1$ arbitrary | $\chi^2_{n-r}$ |

These tests introduce an additional layer of uncertainty into the overall procedure. The tests assume a specific cointegrating rank. Ideally the cointegrating rank has to be determined before the deterministic terms are tested.
A useful strategy is to check the robustness of the testing results for the cointegrating rank to different specifications of the deterministic terms.

# Example

Three interest rates are driven by a common stochastic trend so that there are two cointegrating relations ($r = 2$). The normalized cointegration matrix

$$\boldsymbol{\beta}' = \begin{bmatrix} 1 & 0 & \boldsymbol{\beta}_{31} \\ 0 & 1 & \boldsymbol{\beta}_{32} \end{bmatrix}$$

Suppose one wants to test that the interest rate spreads are stationary so that the cointegrating relations are

$$i_1 - i_3$$

$$i_2 - i_3$$

to test that the cointegration matrix has the form

$$\boldsymbol{\beta}' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

The null hypothesis of interest is then

$$H_0 : \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{31} \\ \boldsymbol{\beta}_{32} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

Overidentifying restrictions for the cointegration matrix

$$\boldsymbol{\beta} = \mathbf{H}\boldsymbol{\varphi}$$

$\mathbf{H}$ is known, fixed $(n \times s)$ matrix, $\boldsymbol{\varphi}$ is $(s \times r)$, $s \geq r$.
$n = 3, r = 1$, if $\boldsymbol{\beta}_{31} = -\boldsymbol{\beta}_{21}$, we have

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{11} \\ \boldsymbol{\beta}_{21} \\ -\boldsymbol{\beta}_{21} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{11} \\ \boldsymbol{\beta}_{21} \end{bmatrix} = \mathbf{H}\boldsymbol{\varphi}$$

$$\boldsymbol{\varphi} = (\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{21})'$$

free parameters.

Replace $\mathbf{Y}_{-1}$ with $\mathbf{H}\mathbf{Y}_{-1}$ in the generalized eigenvalue problem

$$det\left(\lambda\mathbf{H}'\mathbf{S}_{11}\mathbf{H}-\mathbf{H}'\mathbf{S}'_{01}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\mathbf{H}\right)=0$$

the eigenvectors corresponding to $\lambda_1^H, \lambda_2^H, \ldots, \lambda_r^H$ are the estimators of the columns of $\boldsymbol{\varphi}$. The restricted estimator

$$\widehat{\boldsymbol{\beta}}=\mathbf{H}\widehat{\varphi}$$

the estimatos for $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$ follow.

In general the restrictions may be available in the form

$$\boldsymbol{\beta}=[\mathbf{H}_1\boldsymbol{\varphi}_1,\mathbf{H}_2\boldsymbol{\varphi}_2]$$

$$\begin{aligned}\mathbf{H}_j &: (n\times s_j)\\ \varphi_j &: (s_j\times r_j)\end{aligned}$$

with $r=r_1+r_2$.

$n = 3$ and $r = 2$ one zero restriction on the last element of the second cointegrating vector with $\boldsymbol{\beta}_{32} = 0$ can be represented

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{11} & \boldsymbol{\beta}_{12} \\ \boldsymbol{\beta}_{21} & \boldsymbol{\beta}_{22} \\ \boldsymbol{\beta}_{31} & 0 \end{bmatrix} = [\mathbf{H}_1 \boldsymbol{\varphi}_1, \mathbf{H}_2 \boldsymbol{\varphi}_2]$$

with $\mathbf{H}_1 = \mathbf{I}_3$, $\boldsymbol{\varphi}_1 = (\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{21}, \boldsymbol{\beta}_{31})'$

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$\varphi_2 = (\boldsymbol{\beta}_{12}, \boldsymbol{\beta}_{22})'$

Linear restrictions on the loading matrix $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = \mathbf{G}\boldsymbol{\psi}$$

$$\mathbf{G} \quad : \quad (n \times s)$$
$$\boldsymbol{\psi} \quad : \quad (s \times r)$$

$\boldsymbol{\psi}$ parameter matrix with $s \geq r$ can also be imposed easily.

Example: consider the restriction that some or all of the cointegration relations do not enter a particular equation. This restriction implies *weak exogeneity* of specific variables for the cointegrating parameters.

A variable is weakly exogenous for the cointegrating parameters if none of the cointegrating relations enter the equation for that variable.

$n = 3$ with $r = 2$ we may wish to consider the case that the third variable is weakly exogenous:

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \mathbf{G}\boldsymbol{\psi}$$

To impose the restriction $\boldsymbol{\alpha} = \mathbf{G}\boldsymbol{\psi}$ with $rk(\mathbf{G})$. Premultiplying

$$\Delta \mathbf{YM} = \mathbf{\Pi Y}_{-1}\mathbf{M} + \widehat{\epsilon}$$

by $(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$

$$(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\Delta \mathbf{YM} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{\Pi Y}_{-1}\mathbf{M} + (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\widehat{\epsilon}$$

and performing an RR regression on a trasformed model. The restricted estimator of $\boldsymbol{\alpha}$

$$\widehat{\boldsymbol{\alpha}} = \mathbf{G}\widehat{\boldsymbol{\psi}}$$

The estimator of $\boldsymbol{\psi}$ may be obtained via the solution of an appropriately modified generalized eigenvalue problem.