Econometric Methodology and Macroeconomics Applications: The Cointegrated VAR Model

Søren Johansen

Department of Economics University of Copenhagen and CREATE, Aarhus University

EIEF: 8 - 14 November 2011
Lecture 1
The cointegrated VAR Chapter 5.3 and 5.4 pp.48-58

- Integration and Cointegration
- The error correction model
- Granger Representation Theorem
Assume in the following $\varepsilon_t$ i.i.d. $(0, \Omega)$, and $C_i$ decreasing exponentially

**Definition 1.** $x_t$ integrated of order 0, $I(0)$, if $x_t = C(L)\varepsilon_t$, with $C(1) \neq 0$

$x_t$ integrated of order 1, $I(1)$, if $\Delta x_t = C(L)\varepsilon_t$, with $C(1) \neq 0$, $\Delta x_t$ is $I(0)$

1. $x_{0t} = \varepsilon_{0t} \sim I(0)$, $C(z) = 1 \neq 0$, ($x_{0t} = \varepsilon_{0t} - \varepsilon_{0t-1}$ not $I(0)$)

2. $x_{1t} = \sum_{i=0}^{\infty} \rho^i \varepsilon_{1t-i} \sim I(0)$, $|\rho| < 1$, $C(z) = \sum_{i=0}^{\infty} \rho^i z^i = \frac{1}{1 - \rho z}$

3. $x_{2t} = \sum_{i=1}^{t} \varepsilon_{2i} \sim I(1)$, $\Delta x_t = \varepsilon_{2t} = C(L)\varepsilon_t$; $C(z) = 1$

4. \[
\begin{pmatrix}
  x_{1t} \\
  x_{2t}
\end{pmatrix} = \begin{pmatrix}
  \sum_{i=1}^{t} \varepsilon_{1i} \\
  \sum_{i=0}^{\infty} \rho^i \varepsilon_{2t-i}
\end{pmatrix} \sim I(1)
\]

\[
\Delta \begin{pmatrix}
  x_{1t} \\
  x_{2t}
\end{pmatrix} = C(L)\varepsilon_t; \quad C(z) = \begin{pmatrix}
  1 & 0 \\
  0 & \sum_{i=0}^{\infty} \rho^i (1 - z) z^i
\end{pmatrix}
\]
**Definition 2.** If $x_t$ is $I(1)$, and $\beta'x_t$ is stationary, then $x_t$ is cointegrated with cointegration vector $\beta$.

**Examples**

\[
x_{1t} = a \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{2t} \sim I(1), \quad \Delta x_{1t} = a \varepsilon_{1t} + \varepsilon_{2t} - \varepsilon_{2,t-1}
\]

\[
x_{2t} = b \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{2,t-1} \sim I(1), \quad \Delta x_{2t} = b \varepsilon_{1t} + \varepsilon_{2,t-1} - \varepsilon_{2,t-2}
\]

\[
x_t \sim I(1) \text{ because } \Delta x_t = C(L)\varepsilon_t; \quad C(z) = \left(\begin{array}{cc} a & 1 - z \\ b & (1 - z)z \end{array}\right)
\]

and $C(1) \neq 0$ but singular. Now consider $bx_{1t} - ax_{2t} = b\varepsilon_{2t} - a\varepsilon_{2t-1}$ is stationary and therefore $x_t$ is cointegrated with $\beta = (b, -a)'$. Note that $a = 0$ means that $x_{1t}$ is stationary.
The Error Correction Model

\[ x_t = \Pi_1 x_{t-1} + \Pi_2 x_{t-2} + \Phi D_t + \epsilon_t \]

\[ x_t - x_{t-1} = (\Pi_1 + \Pi_2 - I_p)x_{t-1} + \Pi_2(x_{t-2} - x_{t-1}) + \Phi D_t + \epsilon_t \]

\[ \Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \Phi D_t + \epsilon_t \]

Note that

\[ \Pi(z) = I_p - z\Pi_1 - z^2\Pi_2 = (1 - z)I_p - \Pi z - \Gamma_1 z(1 - z) \]

If \( \Pi(z) \) has unit root, then

\[ \Pi(1) = -\Pi = -\alpha \beta', \]

for some \( \alpha \) and \( \beta \) of dimension \( p \times r \) and rank \( r < p \)

Error Correction Model:

\[ ECM : \Delta x_t = \alpha \beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \Phi D_t + \epsilon_t \]
Question: If the VAR has unit roots and the other roots are larger than one, what is the moving average representation?

Error correction formulation:

\[ \Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t \]

\[ \Pi(z) = (1 - z)I_p - \alpha \beta' z - \sum_{i=1}^{k-1} (1 - z)z^i \Gamma_i \]

\( I(1) \) condition:

\[ \det(\Pi(z)) = 0 \implies z = 1 \text{ or } |z| > 1 \]

\[ \Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i, \quad \det(\alpha' \Gamma \beta') \neq 0 \]
The Granger Representation Theorem

\[
\det(\Pi(z)) = 0 \iff z = 1 \text{ or } |z| > 1
\]

\[
\Gamma = l_p - \sum_{i=1}^{k-1} \Gamma_i, \quad \det(\alpha' \Gamma \beta_\perp) \neq 0
\]

**Theorem:** If \( I(1) \) condition is satisfied then

\[
(1 - z) \Pi^{-1}(z) = C + \sum_{i=0}^{\infty} C_i^* (1 - z) z^i \quad \text{or} \quad \Pi^{-1}(z) = \frac{1}{1 - z} C + \sum_{i=0}^{\infty} C_i^* z^i
\]

and the solution of the ECM is

\[
x_t = C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i} + A, \quad \beta' A = 0, \quad \text{where} \ C = \beta_\perp (\alpha' \Gamma \beta_\perp)^{-1} \alpha_\perp
\]

1. \( \Delta x_t \) is \( I(0) \): \( x_t \) is \( I(1) \)
2. \( \beta' x_t \) is stationary: \( x_t \) has \( r = \text{rank}(\beta) \) cointegrating or long-run relations
3. There are \( p - r = \text{rank}(\alpha_\perp) \) common trends \( \alpha'_\perp \sum_{i=1}^{t} \varepsilon_i \)
An example of the solution

\[ \Delta x_{1t} = \alpha_1(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} \]
\[ \Delta x_{2t} = \alpha_2(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t} \]

Subtracting we find an AR(1) process

\[ \Delta(x_{1t} - x_{2t}) = (\alpha_1 - \alpha_2)(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} - \varepsilon_{2t} \]

\[ x_{1t} - x_{2t} = \sum_{i=0}^{\infty} (1 + \alpha_1 - \alpha_2)^i (\varepsilon_{1t-i} - \varepsilon_{2t-i}) (= y_t) \]

which is stationary if \(|1 + \alpha_1 - \alpha_2| < 1|.

Note that the \(I(1)\) condition involves

\[ \Pi = \begin{pmatrix} \alpha_1 & -\alpha_1 \\ \alpha_2 & -\alpha_2 \end{pmatrix}, \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \beta = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \]

\[ \alpha_\perp = \begin{pmatrix} \alpha_1 \\ -\alpha_2 \end{pmatrix}, \beta_\perp = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Gamma = I_2, \alpha_\perp' \Gamma \beta_\perp = \alpha_1 - \alpha_2 \neq 0 \]

(Note: For \(\alpha_1 = \alpha_2\) we get \(I(2)\))
Similarly we find a random walk (common trend)

\[ \alpha_2 \Delta x_{1t} - \alpha_1 \Delta x_{2t} = \alpha_2 \varepsilon_{1t} - \alpha_1 \varepsilon_{2t} \]

\[ \alpha_2 x_{1t} - \alpha_1 x_{2t} = \alpha_2 x_{10} - \alpha_1 x_{20} + \sum_{i=1}^{t} (\alpha_2 \varepsilon_{1i} - \alpha_1 \varepsilon_{2i}) (= S_t) \]

\[ x_{1t} - x_{2t} = y_t \text{ (stationary cointegrating relation)} \]

\[ x_{1t} = \frac{1}{\alpha_2 - \alpha_1} (S_t - \alpha_1 y_t) \]

\[ x_{2t} = \frac{1}{\alpha_2 - \alpha_1} (S_t - \alpha_2 y_t) \]

Thus if \(|1 + \alpha_1 - \alpha_2| < 1\) then

1. \(x_{1t} - x_{2t}\) is stationary
2. \(\alpha_2 x_{1t} - \alpha_1 x_{2t}\) is random walk
3. \(x_t\) is \(I(1)\)
4. \(x_t\) cointegrated with cointegration vector \((1, -1)\).
The movement of two cointegrated processes

\[ \begin{align*}
\dot{x}_t &= \left[ m_t, y_t \right] \\
&= \beta_0 + \beta_t \sum_{i=1}^{t} \varepsilon_i + \alpha \cdot x_{\infty,t}
\end{align*} \]

\[ sp(\beta_\perp) \]

**Figure:** The process \( x'_t = [m'_t, y'_t] \) is pushed along the attractor set by the common trends and pulled towards the attractor set, \( sp(\beta_\perp) \), by the adjustment coefficients.
An example of a simple model

\[ \Delta x_{1t} = -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} \]

\[ \Delta x_{2t} = \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t} \]

gives \( I(1) \) integration and cointegration

\( 1 + \alpha_1 - \alpha_2 = 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2} < 1 \)

Another example

\[ \Delta x_{1t} = \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} \]

\[ \Delta x_{2t} = -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t} \]

is explosive and not cointegrated \( 1 + \alpha_1 - \alpha_2 = 1 + \frac{1}{4} + \frac{1}{4} = \frac{3}{2} > 1 \)

\[ \Delta(x_{1t} - x_{2t}) = \frac{1}{2}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} - \varepsilon_{2t}, \text{ implies that } x_{1t} - x_{2t} \text{ explosive} \]
A strange example

\begin{align*}
\Delta x_{1t} &= \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \frac{9}{4}\Delta x_{2t-1} + \varepsilon_{1t} \\
\Delta x_{2t} &= -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t}
\end{align*}

is \(I(1)\) and cointegrated.

The sign of the adjustment is not intuitive

The processes do not adjust properly, yet are \(I(1)\).

\[
\det(\Pi(z)) = \det\left(\begin{array}{cc}
1 - z & \frac{1}{4}z \\
\frac{1}{4}z & \frac{1}{4}z - \frac{9}{4}z(1 - z)
\end{array}\right) = 0
\]
implies \(|z| > 1\) or \(z = 1\) and \(\alpha'\Gamma\beta' \neq 0\).

Cointegration is a system property and requires a careful analysis of the characteristic polynomial.
Conclusion:

The cointegrated vector autoregressive model

\[ \Delta x_t = \alpha (\beta' x_{t-1} - \beta_0) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t \]

is a dynamic stochastic model for all the variables, that allows the simultaneous modelling of the long-run relations \( \beta' x = \beta_0 \), and the adjustment towards the disequilibrium errors.

1. The long-run relations \( \beta' x = \beta_0 \) define the attractor set

\[ \{ x \in \mathbb{R}^p | Cx = \alpha (\beta' \alpha)^{-1} \beta_0 \} = \{ x | \beta' x = \beta_0 \} \]

the set of equilibria or steady states. The coefficients are long-run elasticities.

2. The adjustment coefficients \( \alpha \) define the direction of adjustment, the 'pulling forces'

3. The common trends are given by \( \alpha' \sum_{i=1}^{t} \varepsilon_i \) define the 'pushing forces'

The Granger Representation Theorem gives the solution of the autoregressive equations and is useful for deterministics and asymptotics.
A Constant and Linear Term in the AR(1) model and in the VAR
A Constant and Linear Term in the AR(1) model and in the VAR

A simple example

\[ y_t = \gamma t + \mu + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } N(0, \sigma^2) \]

lag equation one period, multiply by \( \rho \) and subtract

\[
\begin{align*}
y_t &= \gamma t + \mu + u_t, \text{ and } y_{t-1} = \gamma(t-1) + \mu + u_{t-1}, \\
y_t - \rho y_{t-1} &= \gamma t + \mu - \rho \gamma(t-1) - \rho \mu + (u_t - \rho u_{t-1}) \\
y_t &= \rho y_{t-1} + \gamma(1-\rho)t + \rho \gamma + (1-\rho)\mu + \varepsilon_t \\
y_t &= b_1 y_{t-1} + b_2 t + b_0 + \varepsilon_t \quad \text{(if } \rho = 1, \text{ we have } \Delta y_t = \gamma + \varepsilon_t) \\
\end{align*}
\]

<table>
<thead>
<tr>
<th>( b_1 )</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>( \rho )</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>( \gamma(1-\rho) )</td>
</tr>
<tr>
<td>( \rho b_1 )</td>
<td>( \gamma(1-\rho) )</td>
</tr>
<tr>
<td>( b_0 )</td>
<td>( \rho \gamma + (1-\rho) \mu )</td>
</tr>
<tr>
<td>( \rho \gamma + (1-\rho) \mu )</td>
<td></td>
</tr>
<tr>
<td>( \mu )</td>
<td>( \frac{(1-b_1) b_0 - b_2 b_1}{(1-b_1)^2} )</td>
</tr>
</tbody>
</table>

Thus a "regression with autocorrelated errors" is the same as a "regression on lagged dependent variable"
The linear ’innovation term’

Model

\[ \Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + (\mu_0 + \mu_1 t + \epsilon_t) \]

Granger Representation Theorem

\[ x_t = C \sum_{i=1}^{t} (\epsilon_i + \mu_0 + \mu_1 i) + \sum_{i=0}^{\infty} C_i^* (\epsilon_{t-i} + \mu_0 + \mu_1 (t-i)) + A, \quad C = \beta (\alpha' \Gamma) \]

Thus in the process we have

1. Quadratic trend, \( \frac{1}{2} C \mu_1 t^2 \) in general
2. If \( \alpha' \mu_1 = 0 \), only linear trend, \( (C \mu_0 + \sum_{i=0}^{\infty} C_i^* \mu_1) t \)
3. If \( \mu_1 = 0 \), still linear trend, \( C \mu_0 t \), but \( \beta' x_t \) no trend because \( \beta' C = 0 \)
4. If \( \mu_1 = 0 \), \( \alpha' \mu_0 = 0 \) no linear trend but constant term \( \sum_{i=0}^{\infty} C_i^* \mu_0 \)
5. If \( \mu_1 = \mu_0 = 0 \) (no deterministics).
Expectations of stationary processes $\Delta x_t$ and $\beta' x_t$

\[
x_t = C \sum_{i=1}^{t} \varepsilon_i + C \mu_0 \, t + \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i} + \sum_{i=0}^{\infty} C_i^* \mu_0 + A
\]

\[
\Delta x_t = C \varepsilon_t + C \mu_0 + \Delta \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i} \quad \text{implies} \quad E(\Delta x_t) = C \mu_0
\]

\[
\Delta x_t = \alpha \beta' x_{t-1} + \mu_0 + \varepsilon_t
\]

\[
E(\Delta x_t) = \alpha E(\beta' x_{t-1}) + \mu_0
\]

\[
C \mu_0 = \alpha E(\beta' x_{t-1}) + \beta_0 \quad \text{implies} \quad E(\beta' x_{t-1}) = -(\beta' \alpha)^{-1} \beta' \mu_0
\]

\[
\Delta x_t - \underbrace{C \mu_0}_{\text{growth rate}} = \alpha (\beta' x_{t-1} - \underbrace{-(\beta' \alpha)^{-1} \beta' \mu_0}_{\text{disequilibrium mean}}) + \varepsilon_t
\]
The 'linear additive term'

\[ x_t = \tau_0 + \tau_1 t + y_t \]

\[ \Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \]

\[ \Delta x_t - \tau_1 = \alpha \beta'(x_{t-1} - \tau_0 - \tau_1(t-1)) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} - \Gamma_i \tau_1 + \varepsilon_t \]

In 'innovation' form with \( \alpha' \mu_1 = 0 \)

\[ \Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \mu_0 + \mu_1 t + \varepsilon_t \]

\[ \mu_0 = \alpha \beta'(\tau_1 - \tau_0) + (l_p - \sum_{i=1}^{k-1} \Gamma_i) \tau_1 \]

\[ \mu_1 = -\alpha \beta' \tau_1 \]
Other deterministics

The 'innovation' dummy

\[ d_t = 1_{\{t=t_0\}} = \begin{cases} 
1, & t = t_0 \\
0, & t \neq t_0 
\end{cases} \]

Model

\[ \Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{-i} + \Phi d_t + \epsilon_t \]

GRT

\[ x_t = C \sum_{i=1}^{t} (\epsilon_i + \Phi d_i) + \sum_{i=0}^{\infty} C_i^* (\epsilon_{t-i} + \Phi d_{t-i}) + A \]

The deterministic part of \( x_t \) is

\[ C \Phi \sum_{i=1}^{t} d_i + \sum_{i=0}^{\infty} C_i^* \Phi d_{t-i} = C \Phi 1_{\{t \geq t_0\}} + C_{t-t_0}^* \Phi 1_{\{t \geq t_0\}} \]
The 'additive' dummy

\[ x_t = \phi d\{t \geq t_0\} + y_t \]

\[ \Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \]

\[ \Delta x_t - \phi d\{t=t_0\} = \alpha \beta' (x_{t-1} - \phi d\{t-1 \geq t_0\}) \]

\[ + \sum_{i=1}^{k-1} (\Gamma_i \Delta x_{t-i} - \Gamma_i \phi d\{t-i=t_0\}) + \varepsilon_t \]

\[ \Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} - \alpha \beta' \phi d\{t-1 \geq t_0\} \]

\[ + \phi d\{t=t_0\} - \sum_{i=1}^{k-1} \Gamma_i \phi d\{t-i=t_0\} + \varepsilon_t \]

Note the many lagged dummies.
The Granger Representation Theorem

\[ x_t = C \sum_{i=1}^{t} (\varepsilon_i + \mu_0 + \mu_1 i) + \sum_{i=0}^{\infty} C^*_i (\varepsilon_{t-i} + \mu_0 + \mu_1 (t-i)) + A, \]

\[ C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp, \beta' A = 0 \]

gives the solution of the autoregressive equations and is useful for understanding the role of deterministic terms.
\[ \Delta x_t = \alpha' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \Phi D_t + \varepsilon_t, \quad \varepsilon_t \ i.i.d. \ (0, \Omega) \]

Structural form:

\[
A_0 \Delta x_t = a' x_{t-1} + A_1 \Delta x_{t-1} + \tilde{\Phi} D_t + \varepsilon_t^*, \quad \varepsilon_t \ i.i.d. \ (0, \Sigma) \\
a = A_0 \alpha, \ A_1 = a_0 \Gamma_1, \ \tilde{\Phi} = A_0 \Phi, \ \varepsilon_t^* = A_0 \varepsilon_t, \ \Sigma = A_0 \Omega A'_0
\]

Note that \( \beta \) is the same but coefficient to \( \beta' x_{t-1}, \Delta x_{t-1}, D_t \) have changed.
Therefore

1. First identify the long-run parameter \( \beta \) by suitable restrictions (can be done in reduced form). Then \( \alpha, \Gamma_1, \Phi, \Omega \) are identified.
2. Next identify the short-run parameters \( (A_0, a, A_1, \tilde{\Phi}, \Sigma) \)
3. If need be identify the shocks
Definition of identification

**Definition** The vector \( \beta_1 \) is identified by restrictions \( R_1' \beta_1 = 0 \) if there is no linear combination \( \sum_{i=1}^{r} a_i \beta_i \) satisfying the restrictions \( R_1' \sum_{i=1}^{r} a_i \beta_i = 0 \), other then if \( a_i = 0, \ i = 2, \ldots, r \)

**Three concepts**
1. generic identification (mathematical)
2. empirical identification (statistics)
3. economic identification (economics)
The rank condition

Identifying restrictions on $\beta$

$$\beta = (H_1\phi_1, \ldots, H_r\phi_2) \text{ or } R_i'\beta_i = 0, i = 1, \ldots, r$$

The rank condition (Abraham Wald) for identification of $\beta_1$ by $R_1$ in the system is that the matrix

$$R_1'\beta = (R_1'\beta_1, R_1'\beta_2, \ldots, R_1'\beta_r)$$

has rank $r - 1$, or if the $r \times r$ matrix $\beta'R_1R_1'\beta$ has rank $r - 1$

$$\text{rank}(\beta'R_1R_1'\beta) = r - 1.$$ 

If there is an $(a_1, \ldots, a_r)$ for which $R_1'\beta a = 0$, we consider the vector $\beta_1^* = \beta_1 + \sum_{i=1}^r a_i\beta_i = \beta_1 + \beta a$. If $\beta_1$ is identified, then $a_i = 0, i = 2, \ldots, r$, which shows that $a$ is unique and that

$$\text{rank}(R_1'\beta) = r - 1.$$
An example 1

\[ x_t = (m_t^r, y_t^r, \Delta p_t, i_t^{\text{deposit}}, i_t^{\text{bond}})', \quad r = 2 \]

We want to identify the two relations as

1. One relation has homogeneity between money and income
2. Another has coefficient to inflation rate zero

\[(1, 1, 0, 0, 0)' \beta_1 = 0 \]
\[(0, 0, 1, 0, 0)' \beta_2 = 0 \]

\[ \beta' x_t = \left( \begin{array}{c}
\phi_{11} m_t^r - \phi_{11} y_t^r + \phi_{13} \Delta p_t + \phi_{14} i_t^{\text{deposit}} + \phi_{15} i_t^{\text{bond}} \\
\phi_{21} m_t^r + \phi_{22} y_t^r + 0 \Delta p_t + \phi_{24} i_t^{\text{deposit}} + \phi_{25} i_t^{\text{bond}}
\end{array} \right) \]

or

\[ \beta_1 = \begin{pmatrix}
1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \quad \phi_1 = \begin{pmatrix}
\phi_{11} \\
-\phi_{11} \\
\phi_{13} \\
\phi_{14} \\
\phi_{15}
\end{pmatrix} \]
An example 2

Check identification: apply $R_1' = (1,1,0,0,0)$ and $R_2' = (0,0,1,0,0)$ to $\beta$

$$\beta' = \left( \begin{array}{c} \phi_{11}, -\phi_{11}, \phi_{13}, \phi_{14}, \phi_{15} \\ \phi_{21}, \phi_{22}, 0, \phi_{24}, \phi_{25} \end{array} \right)$$

$$R_1'\beta = (0, \phi_{21} + \phi_{22}) \text{ rank 1 (in general)}$$
$$R_2'\beta = (\phi_{13}, 0) \text{ rank 1 (in general)}$$

Both are identified \textbf{generically}. (Only if $\phi_{13} = 0$, the second is unidentified and only if $\phi_{22} + \phi_{23} = 0$ the first is unidentified)

\textbf{Empirical} identification involves showing that, for a given data set, that in fact $\phi_{13} \neq 0$ and $\phi_{22} + \phi_{23} \neq 0$

\textbf{Economic} identification involves "making sense" of these relations
The first has interpretation that velocity is a function of $\Delta p_t, i_t^{\text{dep}}, i_t^{\text{bond}}$
The second is just a relation between the variables $(m_t, y_t, i_t^{\text{deposit}}, i_t^{\text{bond}})$
Another criterion

Another condition for generic identification (independent of the parameter) is that \( \beta_1 \) is generically identified in the system of \((\beta_1, \beta_2, \beta_3)\) if

\[
\begin{align*}
\text{rank}(R'_1 H_2) &\geq 1, \quad \text{rank}(R'_1 H_3) \geq 1 \\
\text{rank}(R'_1 (H_2, H_3)) &\geq 2
\end{align*}
\]

**Theorem** If \( \beta_1, \ldots, \beta_r \) are identified by \( m_i \) restrictions on \( \beta_i \), then 

\[-2 \log Q(\beta = (H_1 \phi_1, \ldots, H_r \phi_r)) \]

converges in distribution to a \( \chi^2 \) distribution with 

\[
f = \sum_{i=1}^{r} (m_i - r + 1)
\]

degrees of freedom.
\[ \beta' x_t = \begin{pmatrix} \beta_{11} & -\beta_{11} & 0 & \beta_{12} & -\beta_{12} \\ 0 & \beta_{21} & \beta_{22} & 0 & \beta_{23} \\ 0 & 0 & 0 & \beta_{31} & \beta_{32} \end{pmatrix} x_t = \begin{pmatrix} \beta_{11} (m_t^r - y_t^r) + \beta_{12} (i_{t}^{dept}) \\ \beta_{21} y_t^r + \beta_{22} \Delta p_t + \beta_{23} i_t^{dept} \\ \beta_{31} i_t^{dept} + \beta_{32} i_t^{bor} \end{pmatrix} \]
\[ \beta' x_t = \begin{pmatrix}
\beta_{11} & -\beta_{11} & 0 & \beta_{12} & -\beta_{12} \\
0 & \beta_{21} & \beta_{22} & 0 & \beta_{23} \\
0 & 0 & 0 & \beta_{31} & \beta_{32}
\end{pmatrix} x_t \]

\[ R_1' = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix} \]

\[ R_1' H_2 \phi = R_1'(0, \beta_{21}, \beta_{22}, 0, \beta_{23}) = \begin{pmatrix}
\beta_{21} \\
\beta_{22} \\
\beta_{23}
\end{pmatrix} : \text{rank}(R_1' H_2) = 3 \]

\[ R_1'(H_2, H_3) \phi = \begin{pmatrix}
\beta_{21} \\
\beta_{22} \\
\beta_{31} + \beta_{23} + \beta_{32}
\end{pmatrix} : \text{rank}(R_1'(H_2, H_3)) = 3 \]
Asymptotic distribution of the identified $\hat{\beta}$

Let $r = 2$ and assume $\beta$ is identified by normalization and linear restrictions $\beta = (h_1 + H_1\phi_1, h_2 + H_2\phi_2)$.

**THEOREM** In the model without deterministic terms where $\epsilon_t$ are i.i.d. $(0, \Omega)$, the asymptotic distribution of

$$Tvec(\hat{\beta} - \beta) = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} T\hat{\phi}_1 \\ T\hat{\phi}_2 \end{pmatrix} = T \begin{pmatrix} H_1\hat{\phi}_1 \\ H_2\hat{\phi}_2 \end{pmatrix}$$

is given by

$$\begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} \rho_{11} H'_1 G H_1 & \rho_{12} H'_1 G H_2 \\ \rho_{21} H'_2 G H_1 & \rho_{22} H'_2 G H_2 \end{pmatrix}^{-1} \begin{pmatrix} H'_1 \int_0^1 G(dV_1) \\ H'_2 \int_0^1 G(dV_2) \end{pmatrix},$$

where

$$T^{-1/2}X_{[T_u]} \xrightarrow{w} G = CW, \quad T^{-1}S_{11} \xrightarrow{w} G = C \int_0^1 WW' duC', \quad V = \alpha'\Omega^{-1}W = (V_1, V_2)', \quad \rho_{ij} = \alpha'_i\Omega^{-1}\alpha_j.$$

The estimators of the remaining parameters are asymptotically Gaussian and asymptotically independent of $\hat{\beta}$. 
An illustration of the mixed Gaussian distribution in cointegration.

\[ x_{1t} = \theta x_{2t-1} + \varepsilon_{1t}, \]
\[ \Delta x_{2t} = \varepsilon_{2t}. \]

where \( \varepsilon_t \) not only i.i.d. but also \( \varepsilon_{1t} \) and \( \varepsilon_{2t} \) independent then the maximum likelihood estimator satisfies

\[
\hat{\theta} = \frac{\sum_{t=1}^{T} x_{1t} x_{2t-1}}{\sum_{t=1}^{T} x_{2t-1}^2} = \theta + \frac{\sum_{t=1}^{T} \varepsilon_{1t} x_{2t-1}}{\sum_{t=1}^{T} x_{2t-1}^2}.
\]

The distribution of \( \hat{\theta} \) conditional on the regressor \( \{x_{2t}\} \) is \( N(\theta, \sigma_1^2 / \sum_{t=1}^{T} x_{2t-1}^2) \). Hence \( \hat{\theta} \) is mixed Gaussian with mixing parameter \( 1 / \sum_{t=1}^{T} x_{2t-1}^2 \), and hence has mean \( \theta \) and variance \( \sigma_1^2 E(1 / \sum_{t=1}^{T} x_{2t-1}^2) \). Inference is \( \chi^2 \).
When constructing a test for $\theta = \theta_0$ we do not base our inference on the Wald test
\[
\frac{\hat{\theta} - \theta}{\sqrt{\text{Var}(\hat{\theta})}} = \frac{\hat{\theta} - \theta}{\sqrt{E(\hat{\sigma}^2_1 / \sum_{t=1}^{T} x_{2t-1}^2)}},
\]
but rather on the Wald test which comes from an expansion of the likelihood function and is based on the observed information:
\[
t = \frac{\hat{\theta} - \theta}{\sqrt{\hat{\sigma}^2_1 / \sum_{t=1}^{T} x_{2t-1}^2}},
\]
which is distributed as $N(0, 1)$. Thus we normalize by the observed information not the expected information often used when analyzing stationary processes.
Figure: The joint distribution of $\hat{\theta}$ and the observed information $(\sum_{i=1}^{T} x_{2t-1}^{2} / \hat{\sigma}^{2})$ in the model $x_{1t} = \theta x_{2t-1} + \epsilon_t$, and $\Delta x_{2t} = \epsilon_{2t}$
Conclusion

The identification problem for $\beta$ is solved as the classical identification problem by the rank criterion. Various forms of identification were discussed and another criterion for identification, which does not depend on parameters, was given. A few comments on the application of the mixed Gaussian distribution for inference were given.