Vector error correction model, VECM
Cointegrated VAR
Chapter 4

Financial Econometrics
Michael Hauser
WS15/16
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 (Johansen)

Bivariate cointegration
Motivation
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 there are 2 processes $x_t$ and $y_t$ are both I(1) and

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

with $\epsilon_t$ 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 \textit{(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.

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

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

$\epsilon_t$ and $\eta_t$ 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
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, $\epsilon_t, \eta_t$.

$$X_t = X_{t-1} + \epsilon_t, \quad Y_t = Y_{t-1} + \eta_t, \quad t = 0, 1, 2, \ldots$$

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.
- $\zeta_t$ is I(1), nonstationary.
- the estimates are inconsistent.
- the $t_\beta$-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 $t_\beta$-values are $t$-distributed,
- the error $\xi_t$ 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. $\zeta_t$ 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 $t_\beta$-statistic is asy normal only if $\zeta_t$ is not serially correlated.
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 \( \beta \) so that

\[
\beta_1 x_{1t} + \ldots + \beta_k x_{kt} = \beta' x_t \quad \ldots \text{ trend-stationary}
\]

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

- \( \beta' x_t \) 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)

- $\beta$ is defined only up to a scale.
  If $\beta'x_t$ is trend-stationary, then also $c(\beta'x_t)$ with $c \neq 0$.
  Moreover, any linear combination of cointegrated variables is stationary.
- More generally we could consider $x \sim l(d)$ and $\beta'x \sim l(d - b)$ with $b > 0$.
  Then the $x$'s are 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): $x_t = \Phi_1 x_{t-1} + \epsilon_t$

We analyze in the following the properties of

$$
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
= 
\begin{bmatrix}
    0.5 & -1. \\
    -0.25 & 0.5
\end{bmatrix}
\begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix}
+ 
\begin{bmatrix}
    \epsilon_{1t} \\
    \epsilon_{2t}
\end{bmatrix}
$$

$\epsilon_t$ are weakly stationary and serially uncorrelated.

We know a VAR(1) is stable, if the eigenvalues of $\Phi_1$ are less 1 in modulus.

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

$\Phi_1$ has a root on the unit circle. So process $x_t$ is not stable.

Remark: $\Phi_1$ is singular; its rank is 1.
Common trend

For all $\Phi_1$ there exists an invertible matrix $L$ so that

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

$\Lambda$ is (for simplicity) diagonal containing the eigenvalues of $\Phi_1$.

We define new variables $y_t = Lx_t$ and $\eta_t = L\epsilon_t$.

Left multiplication of the VAR(1) with $L$ gives

$$Lx_t = L\Phi_1 x_{t-1} + L\epsilon_t$$

$$(Lx_t) = L\Phi_1 L^{-1}(Lx_{t-1}) + (L\epsilon_t)$$

$$y_t = \Lambda y_{t-1} + \eta_t$$
Common trend: x’s are I(1)

In our case $L$ and $\Lambda$ are

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

Then

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

- $\eta_t = L\epsilon_t$: $\eta_{1t}$ and $\eta_{2t}$ 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 $y_{1t}$

$y_t = Lx_t$ with $L$ invertible, so we can express $x_t$ in $y_t$.

Left multiplication by $L^{-1}$ gives

$$L^{-1}y_t = L^{-1}Ay_{t-1} + L^{-1}\eta_t$$

$$x_t = (L^{-1}A)y_{t-1} + \epsilon_t$$

$L^{-1} = \ldots$

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

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

- 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 $x_{2t}$. 
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), an example
A cointegrated VAR(1)

We go back to the system and proceed directly.

\[ x_t = \Phi_1 x_{t-1} + \epsilon_t \]

and subtract \( x_{t-1} \) on both sides (cp. the Dickey-Fuller statistic).

\[
\begin{bmatrix}
\Delta x_{1t} \\
\Delta x_{2t}
\end{bmatrix}
= \begin{bmatrix}
-0.5 & -1. \\
-0.25 & -0.5
\end{bmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

The coefficient matrix \( \Pi, \Pi = -(I - \Phi_1) \), in

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

has only rank 1. It is singular.
Then \( \Pi \) can be factorized as

\[ \Pi = \alpha \beta' \]

\[ (2 \times 2) = (2 \times 1)(1 \times 2) \]
A cointegrated VAR(1)

$k$ the number of endogenous variables, here $k = 2$.

$m = \text{Rank}(\Pi) = 1$, is the number of cointegrating relations.

A solution for $\Pi = \alpha\beta'$ is

$$
\begin{bmatrix}
-0.5 & -1. \\
-0.25 & -0.5
\end{bmatrix}
= \begin{pmatrix}
-0.5 \\
-0.25
\end{pmatrix}
\begin{pmatrix}
1 \\
2
\end{pmatrix}'
= \begin{pmatrix}
-0.5 \\
-0.25
\end{pmatrix}
\begin{pmatrix}
1 & 2
\end{pmatrix}
$$

Substituted in the model

$$
\begin{bmatrix}
\Delta x_{1t} \\
\Delta x_{2t}
\end{bmatrix}
= \begin{pmatrix}
-0.5 \\
-0.25
\end{pmatrix}
\begin{pmatrix}
1 & 2
\end{pmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
$$
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)

$$x_t = \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + \epsilon_t$$

There always exists an error correction representation of the form (trick $x_t = x_{t-1} + \Delta x_t$)

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + \epsilon_t$$

where $\Pi$ and the $\Phi^*$ are functions of the $\Phi$’s. Specifically,

$$\Phi_j^* = - \sum_{i=j+1}^{p} \Phi_i, \quad j = 1, \ldots, p - 1$$

$$\Pi = -(I - \Phi_1 - \ldots - \Phi_p) = -\Phi(1)$$

The characteristic polynomial is $I - \Phi_1 z - \ldots - \Phi_p z^p = \Phi(z)$. 
Interpretation of

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta x_{t-i} + \epsilon_t$$

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

$$\Pi = \alpha \beta'$$

$$(k \times k) = (k \times m)[(k \times m)']$$

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

$$\text{Rank}(\Pi) = \min[\text{Rank}(\alpha), \text{Rank}(\beta)]$$
Long term relationship in

\[ \Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{p-1} \Phi^*_i \Delta x_{t-i} + \epsilon_t \]

There is an adjustment to the 'equilibrium' \( x^* \) or long term relation described by the cointegrating relation.

- Setting \( \Delta x = 0 \) we obtain the long run relation, i.e.

\[ \Pi x^* = 0 \]

This may be written as

\[ \Pi x^* = \alpha (\beta' x^*) = 0 \]

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

\[ \beta' x^* = 0_{m \times 1} \]
The long run relation does not hold perfectly in \((t - 1)\). There will be some deviation, an *error*,

\[ \beta' x_{t-1} = \xi_{t-1} \neq 0 \]

The adjustment coefficients in \(\alpha\) multiplied by the ’errors’ \(\beta' x_{t-1}\) induce adjustment. They determine \(\Delta x_t\), 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
  \[ x_{1,t-1} + 2x_{2,t-1} = \xi_t \]

  \( \xi_t \) is the stationary error.
- The adjustment of \( x_{1,t} \) in \( t \) to \( \xi_{t-1} \), 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 $x_t \ I(1)$, (unit root) nonstationary.

$$x_t = \phi + \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + \epsilon_t$$

Then

- $\Delta x_t$ is $I(0)$.
- $\Pi = -\Phi(1)$ is singular, i.e. $|\Phi(1)| = 0$

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

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

$$\Delta x_t = \phi + \Pi x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + \epsilon_t$$

$\Pi x_{t-1}$ is called the error-correction term.
We distinguish 3 cases for \( \text{Rank}(\Pi) = m \):

1. \( m = 0 \): \( \Pi = 0 \)
2. \( 0 < m < k \): \( \Pi = \alpha\beta', \quad \alpha_{(k \times m)}, (\beta')_{(m \times k)} \)
3. \( m = k \): \( |\Pi| = | - \Phi(1)| \neq 0! \)
1. \( \text{Rank}(\Pi) = 0, \ m = 0 : \)

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

- \( \Pi = 0, \) 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.

\[
\Delta x_t = \phi + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + \epsilon_t
\]

- \( \Pi \) has \( m = 0 \) eigenvalues different from 0.
II. Rank(\(\Pi\)) = m, \(0 < m < k\):

The rank of \(\Pi\) is \(m, m < k\). We factorize \(\Pi\) in two rank \(m\) matrices \(\alpha\) and \(\beta'\).

\[ \text{Rank}(\alpha) = \text{Rank}(\beta) = m. \]

Both \(\alpha\) and \(\beta\) are \((k \times m)\).

\[ \Pi = \alpha \beta' \neq 0 \]

The VEC form is then

\[ \Delta x_t = \phi + \alpha \beta' x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + \epsilon_t \]

- The \(x\)'s are integrated, \(I(1)\).
- There are \(m\) eigenvalues \(\lambda(\Pi) \neq 0\).
- The \(x\)'s are cointegrated. There are \(m\) linear combinations, which are stationary.
II. Rank(Π) = m, 0 < m < k:

- There are m linear independent cointegrating (column) vectors in β.
- The m stationary linear combinations are \( \beta'x_t \).
- \( x_t \) has \((k - m)\) unit roots, so \((k - m)\) common stochastic trends.

There are

- \( k \) I(1) variables,
- \( m \) cointegrating relations (eigenvalues of Π different from 0), and
- \((k - m)\) stochastic trends.

\[
k = m + (k - m)
\]
III. $\text{Rank}(\Pi) = m, \ m = k$:

Full rank of $\Pi$ implies

- that $|\Pi| = | - \Phi(1)| \neq 0$.
- $x_t$ has no unit root. That is $x_t$ 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. Rank(\(\Pi\)) = m, 0 < m < k : (cont) common trends

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

\[
\alpha_{\perp}'\alpha = 0
\]

\[
\{k \times (k - m)\}'\{k \times m\} = \{(k - m) \times m\}
\]

If the ECM is left multiplied by \(\alpha_{\perp}'\), the error correction term vanishes,

\[
\alpha_{\perp}'\Pi = (\alpha_{\perp}'\alpha)\beta' = 0_{(k-m)\times k}
\]

with \(\alpha_{\perp}'\Delta x_t = \Delta(\alpha_{\perp}'x_t)\)

\[
\Delta(\alpha_{\perp}'x_t) = (\alpha_{\perp}'\phi) + \sum_{i=1}^{p-1} \Phi_i^*\Delta(\alpha_{\perp}'x_{t-i}) + (\alpha_{\perp}'\epsilon_t)
\]
II. 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

\[
\alpha'_{\perp} x_t
\]

which are the common trends.

Example (from above): \(\alpha = (-1, -0.5)'\) then \(\alpha_{\perp} = (1, -2)'.\)
Non uniqueness of $\alpha, \beta$ in $\Pi = \alpha\beta'$

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

$$\alpha\beta' = \alpha \Omega \Omega' \beta' = (\alpha \Omega)(\beta \Omega)' = \alpha^* (\beta^*)'$$

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

Usually the structure

$$\beta' = [I_{m \times m}, (\beta_1')_{m \times (k-m)}]$$

is imposed.

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

Economic interpretation is helpful when structuring $\beta'$. Also, a reordering of the vars might be necessary.
Inclusion of deterministic functions

There are several possibilities to specify the deterministic part, \( \phi \), in the model.

1. \( \phi = 0 \): All components of \( x_t \) are I(1) without drift. The stationary series \( w_t = \beta' x_t \) has a zero mean.

2. \( \phi = (\phi_0)_{k \times 1} = \alpha_{k \times m} c_{0,m \times 1} \): This is the special case of a restricted constant. The ECM is

\[
\Delta x_t = \alpha (\beta' x_{t-1} + c_0) + \ldots
\]

\( w_t = \beta' x_t \) has a mean of \((-c_0)\).

There is only a constant in the cointegrating relation, but the \( x \)'s are I(1) without a drift.

3. \( \phi = \phi_0 \neq 0 \): The \( x \)'s are I(1) with drift. The coint rel may have a nonzero mean. Intercept \( \phi_0 \) may be split in a drift component and a const vector in the coint eq’s.
Inclusion of deterministic functions

4 $\phi = \phi_t = \phi_0 + (\alpha c_1) t$:
  Analogous, $\phi_0$ enters the drift of the $x$’s. $c_1$ becomes the trend in the coint rel.
  
  $$\Delta x_t = \phi_0 + \alpha (\beta' x_{t-1} + c_1 t) + \ldots$$

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.
Estimation is a 3-step procedure:

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

$$\hat{u}_t = \Pi \hat{v}_t + \epsilon_t$$

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

$$(\tilde{\alpha})^{-1} u_t = \tilde{\beta}' v_t$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are $(k \times k)$. The solution is obtained by canonical correlation.
We determine vectors $\alpha_j$, $\beta_j$ so that the linear combinations

$$\alpha_j'u_t \quad \text{and} \quad \beta_j'v_t$$

correlate

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

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

*The columns of $\beta$ are the associated normalized eigenvectors.*

The $\lambda$'s are not the eigenvalues of $\Pi$, but have the same zero/nonzero properties.
Actually we solve a generalized eigenvalue problem

\[ |\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0 \]

with the sample covariance matrices

\[ S_{00} = \frac{1}{T - p} \sum \hat{u}_t \hat{u}_t', \quad S_{01} = \frac{1}{T - p} \sum \hat{u}_t \hat{v}_t' \]

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

The number of eigenvalues \( \lambda \) larger 0 determines the rank of \( \beta \), resp. \( \Pi \), and so the number of cointegrating relations:

\[ \lambda_1 > \ldots > \lambda_m > 0 = \ldots = 0 = \lambda_k \]
3rd step: In this final step the adjustment parameters $\alpha$ and the $\Phi^*$’s are estimated.

$$\Delta x_t = \phi + \alpha \beta' x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + \epsilon_t$$

The maximized likelihood function based on $m$ cointegrating vectors is

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

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

Remark: $S_{00}$ depends only on $\Delta x_t$ and $\Delta x_{t-j}$, $j = 1, \ldots, p$. 

Johansen (3)
Test for cointegration: trace test

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

- the trace test, and
- the maximum eigenvalue test.

**trace test:**

\[ H_0 : \text{Rank}(\Pi) = m \quad \text{against} \quad H_A : \text{Rank}(\Pi) > m \]

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 $\text{Rank}(\Pi) = 0$, 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 $\lambda$ is

- large (say $\approx 1$), then $-\ln(1 - \hat{\lambda}_i)$ is large.
- small (say $\approx 0$), then $-\ln(1 - \hat{\lambda}_i) \approx 0$. 
Test for cointegration: max eigenvalue statistic

- **maximum eigenvalue test:**

  \[ H_0 : \text{Rank}(\Pi) = m \quad \text{against} \quad H_A : \text{Rank}(\Pi) = m + 1 \]

  The statistic is

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

  We start with \( m = 0 \) – that is \( \text{Rank}(\Pi) = 0 \), 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.
The fitted ECM can be used for forecasting $\Delta x_{t+\tau}$. The forecasts of $x_{t+\tau}$ ($\tau$-step ahead) are obtained recursively.

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

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:** $x_t, y_t \sim I(1)$

$$y_t = \alpha + x_t'\beta + u_t$$

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.

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

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 
  \[
  z_t = \Pi z_{t-1} + \xi_t
  \]
  estimated via LS, where \( z_t = (y_t, x_t')' \).

\( \hat{\xi}_{1,t} \) is stationary, \( \hat{u}_t \) only if the vars are cointegrated. 

*Intuitively* the ratio \( (s_{\hat{\xi}_1}^2 / s_{\hat{u}}^2) \) is small under no coint and large under coint (due to the superconsistency associated with \( s_{\hat{u}}^2 \)).

\[
H_0: \text{no coint} \quad H_A: \text{coint}
\]

Two test statistics \( \hat{P}_u \) and \( \hat{P}_z \) are available in \texttt{ca.po} \{\texttt{urca}\}.

*Remark:* If \( z_t \) is a RW, then \( z_t = 1z_{t-1} + \xi_t \) and \( \xi_t \) stationary.
Exercises and references
Exercises

Choose 2 of (1G, 2) and 1 out of (3G, 4G).

1G Use Ex4_SpurReg_R.txt to generate and comment the small sample distribution of the $t$-statistic and $R^2$ for
(a) the spurious regression problem,
(b) for the model in $\Delta X_t$ and $\Delta Y_t$.

2 Given a VAR(2) in standard form. Derive the VEC representation. Show the equivalence of both representations. Use $x_t = x_{t-1} + \Delta x_t$.

3G Investigate the cointegration properties of stock indices.
(a) For 2 stock exchanges,
(b) for 3 or more stock exchanges.
Use Ex4_R.txt.
Exercises

4G  Investigate the price series of black and white pepper, \texttt{PepperPrices} from the \texttt{R} library ("AER") wrt cointegration and give the VECM.
References

Tsay 8.5-6
Geyer 3.1
Johnson and Wichern: Multivariate Analysis, for canonical correlation
Phillips and Ouliaris (1990): Asymptotic properties of residual based tests for
cointegration, Econometrica 58, 165-193