Cointegration Lecture I:
Introduction

Julia Giese
Nuffield College
julia.giese@economics.ox.ac.uk

Hilary Term 2008
Outline

Introduction

Estimation of unrestricted VAR

Non-stationarity

Deterministic components

Appendix: Mis-specification testing
Lectures

When? Hilary 2008: Week 5-8, Friday 11am - 1pm
Where? Seminar Room G
What? Four lectures on Cointegration

1. The unrestricted VAR model:
   Specification and issues with non-stationarity

2. The cointegrated VAR model:
   Estimation and rank determination

3. The cointegrated VAR model:
   Identification of long and short run

4. Extensions:
   I(2), specific to general and general to specific, Global VAR
The course will mainly follow Juselius (2006), aiming to provide the theory needed to use the cointegrated VAR model in applied work. More advanced econometric theory is found in Johansen (1996) which is not required except where explicitly referred to. Further readings will be given during the lectures.

Other Practicalities

- **Office hour**: Friday 10-11am, Desk A4, and by arrangement.
- **Exam**: One question similar to past years.

Many thanks to Bent Nielsen for providing lecture notes on which these are based.
Outline

Introduction

Estimation of unrestricted VAR

Non-stationarity

Deterministic components

Appendix: Mis-specification testing
A VAR in levels

The unrestricted vector autoregressive (VAR) model of order $k$ with $p$ endogenous variables is given by

$$x_t = \Pi_1 x_{t-1} + \ldots + \Pi_k x_{t-k} + \phi D_t + \varepsilon_t, \quad t = 1, 2, \ldots, T$$

where

- $x_t$ is a vector of the $p$ variables at time $t$,
- $\Pi_i$ are $p \times p$ matrices of parameters with $i = 1, \ldots, k$,
- $D_t$ a vector of deterministic components with a vector of coefficients $\phi$; and
- $\varepsilon_t$ a $p \times 1$ vector of errors.
Assumptions

1. The VAR($k$) model is linear in the parameters.
2. The parameters are constant.
3. The error terms are *identically* and *independently distributed* and follow a *Gaussian* (i.e. *Normal*) *distribution*:

$$\varepsilon_t \sim \text{iid } \mathcal{N}_p(\mathbf{0}, \Omega),$$

where $\Omega$ denotes the variance-covariance matrix of the errors.

Need to check these assumptions! Otherwise inference unreliable. See Appendix for details on mis-specification tests.
Maximum likelihood estimation

For simplicity, write unrestricted model as

\[ x_t = B'Z_t + \varepsilon_t, \]

where \( B' = (\Pi_1, \Pi_2, \ldots, \Pi_k, \mu_0) \) and \( Z_t' = (x_{t-1}', x_{t-2}', \ldots, x_{t-k}', 1) \), assuming that only have constant, i.e. \( \phi D_t = 0 \), and that initial conditions are given.

Now consider the log-likelihood function

\[
\ln L(B, \Omega; X) = -T \frac{p}{2} \ln(2\pi) - T \frac{1}{2} \ln|\Omega| - \frac{1}{2} \sum_{t=1}^{T} (x_t - B'Z_t)' \Omega^{-1} (x_t - B'Z_t).
\]

\(^1\)See Juselius (2006), Ch. 4.1
Maximising log-likelihood with respect to $\mathbf{B}'$ and $\Omega^{-1}$ gives the respective ML estimators:

$$
\hat{\mathbf{B}} = \sum_{t=1}^{T} (\mathbf{Z}_t \mathbf{Z}_t')^{-1} \sum_{t=1}^{T} (\mathbf{Z}_t \mathbf{x}_t') \\
= \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{Zx} \\
\hat{\Omega} = T^{-1} \sum_{t=1}^{T} (\mathbf{x}_t - \hat{\mathbf{B}}' \mathbf{Z}_t)(\mathbf{x}_t - \hat{\mathbf{B}}' \mathbf{Z}_t)' \\
= \mathbf{S}_{xx} - \mathbf{S}_{xZ} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{Zx}
$$

where $\mathbf{S}_{ZZ} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{Z}_t \mathbf{Z}_t')$ and $\mathbf{S}_{Zx} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{Z}_t \mathbf{x}_t')$. 
Outline

Introduction

Estimation of unrestricted VAR

Non-stationarity

Deterministic components

Appendix: Mis-specification testing
A VECM in differences

The VAR(k) model can be expressed as error or vector equilibrium correction model (VECM(k – 1)) formulated in differences:

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \ldots + \Gamma_{k-1} \Delta x_{t-k+1} + \phi D_t + \varepsilon_t$$

where $\Pi = -(I - \Pi_1 - \ldots - \Pi_k)$ and $\Gamma_i = -\sum_{j=i+1}^k \Pi_j$.

If $x_t$ is integrated of order 1 (I(1), i.e. non-stationary), then:

- $\Delta x_t$ is stationary; but
- right hand side contains both stationary and non-stationary processes.
- Hence $\Pi$ must have reduced rank: only a stationary linear combination of $x_{t-1}$ can allow for stationarity of $\Delta x_t$.

Again, this is testable and we need to look at properties of $\Pi$. 
The roots of the characteristic polynomial\(^2\)

Consider two-dimensional VAR(2):

\[
\begin{align*}
x_t &= \Pi_1 x_{t-1} + \Pi_2 x_{t-2} + \phi D_t + \varepsilon_t \\
(I - \Pi_1 L - \Pi_2 L^2) x_t &= \phi D_t + \varepsilon_t
\end{align*}
\]

Then, roots of \( |\Pi(z)| = |I - \Pi_1 z - \Pi_2 z^2| \) provide information on stationarity of \( x_t \):

- if the roots of \( |\Pi(z)| \) are all outside the unit circle, then \( x_t \) is stationary;
- if some roots are outside and some on the unit circle, then \( x_t \) is non-stationary;
- if any of the roots are inside the unit circle, then \( x_t \) is explosive.

Note that we can also find roots by solving for eigenvalues of companion matrix. These are equal to \( z^{-1} \).

\(^2\)See Juselius (2006), Chapters 3.6 and 5.3.
Interpreting the $\Pi$-matrix

| unit root | $\iff$ | $|\Pi(1)| = 0$ | $\iff$ | $\Pi$ has reduced rank |

Since $\Pi$ is of reduced rank $r \leq p$, it may be written as:

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

where $\alpha$ and $\beta$ are $p \times r$ full-rank matrices. Then:

$$\Delta x_t = \alpha \beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \ldots + \Gamma_{k-1} \Delta x_{t-k+1} + \phi D_t + \varepsilon_t \quad (1)$$

- $\beta' x_{t-1}$ is an $r \times 1$ vector of stationary cointegrating relations.
- All variables in (1) are now stationary.
- $\alpha$ denotes the speed of adjustment to equilibrium.

---

$^3$See Juselius (2006), Ch. 5.2.
Cointegration assumptions

Assumptions

(A) \( \text{rank}(\Pi) = r \leq p \),
(B) number of unit roots is \( p - r \),
(C) remaining \( r \) roots are stationary.

To see (B) and (C), consider a VAR(1) with no deterministics, and decompose into \( p - r \) and \( r \) space:

\[
Ax_t = \left\{ A \left( I_p + \alpha \beta' \right) A^{-1} \right\} Ax_{t-1} + \varepsilon_t
\]

\[
\begin{pmatrix}
\beta'_\perp x_t \\
\beta' x_t
\end{pmatrix}
= \begin{pmatrix}
I_{p-r} & \beta'_\perp \alpha \\
0 & I_r + \beta' \alpha
\end{pmatrix}
\begin{pmatrix}
\beta'_\perp x_{t-1} \\
\beta' x_{t-1}
\end{pmatrix}
+ \begin{pmatrix}
\beta'_\perp \varepsilon_t \\
\beta' \varepsilon_t
\end{pmatrix},
\]

where \( \beta_\perp \in \mathbb{R}^{p \times (p-r)} \) is the orthogonal complement of \( \beta \) such that \( \beta' \beta_\perp = 0 \).
Now find roots by solving eigenvalue problem using $|\rho I - M| = 0$, where $\rho$ is an eigenvalue of square matrix $M$:

$$0 = \begin{vmatrix} \rho I_p - \begin{pmatrix} I_{p-r} & \beta' \perp \alpha \\ 0 & I_r + \beta' \alpha \end{pmatrix} \\ \end{vmatrix}$$

$$= \begin{vmatrix} (\rho - 1) I_{p-r} & \beta' \perp \alpha \\ 0 & \rho I_r - (I_r + \beta' \alpha) \end{vmatrix}$$

$$= (\rho - 1)^{p-r} |\rho I_r - (I_r + \beta' \alpha)| .$$

Hence there are at least $p - r$ unit roots.

What about roots of $|\rho I_r - (I_r + \beta' \alpha)|$?

- Given assumption B, there are $p - r$ unit roots in total, and hence 1 cannot be a root here. This implies $|\beta' \alpha| \neq 0$.
- Given assumption C, there are $r$ stationary roots. Hence the absolute eigenvalues of $(I_r + \beta' \alpha)$ are $< 1$. 
Deriving the Granger-Johansen representation\(^4\)

**Cointegrating relations:** Pre-multiply with \(\beta'\):

\[
\Delta (\beta' x_t) = \beta' \alpha \beta' x_{t-1} + \beta' \epsilon_t, \quad \text{then}
\]

\[
\beta' x_t = (I_r + \beta' \alpha) \beta' x_{t-1} + \beta' \epsilon_t
\]

\[
= \sum_{s=0}^{t-1} (I_r + \beta' \alpha)^s (\beta' \epsilon_{t-s}) + (I_r + \beta' \alpha)^t \beta' x_0
\]

is approximately stationary if absolute eigenvalues of \(I_r + \beta' \alpha < 1\).

**Common trends:** Pre-multiply with \(\alpha'_{\perp}\):

\[
\alpha'_{\perp} \Delta x_t = (\alpha'_{\perp} \alpha) \beta' x_{t-1} + \alpha'_{\perp} \epsilon_t = \alpha'_{\perp} \epsilon_t
\]

and cumulate to see

\[
\alpha'_{\perp} x_t = \alpha'_{\perp} \sum_{s=1}^{t} \epsilon_s + \alpha'_{\perp} x_0
\]

---

\(^4\)See Juselius (2006), Ch. 5.4
Use beautiful identity for Granger-Johansen representation of VAR(1):

\[
x_t = (\alpha (\beta' \alpha)^{-1} \beta' + \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp) x_t
\approx \alpha (\beta' \alpha)^{-1} (\text{stationary process})
\]

\[
+ \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \left( \alpha'_\perp \sum_{s=1}^{t} \varepsilon_s + \alpha'_\perp x_0 \right).
\]

**Theorem**

*Given assumptions A, B and C, the Granger-Johansen (or MA) representation of the VAR(k) is given by:*

\[
x_t \approx C \sum_{s=1}^{t} \varepsilon_s + C^*(L) \varepsilon_t + \tilde{X}_0,
\]

*where* \( C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp \) *with* \( \Gamma = -(I - \Gamma_1 - \ldots - \Gamma_{k-1}) \), \( \beta' \tilde{X}_0 = 0 \) *and* \( C^*(L) \varepsilon_t \) *is a stationary process.*
Pulling and pushing forces

Figure: The process $x_t = [b_t^{120}, b_t^1]$ is pushed along the attractor set by the common trends and pulled towards the attractor set by the adjustment coefficients.

5See Juselius (2006), Ch. 5.5.
Outline

Introduction

Estimation of unrestricted VAR

Non-stationarity

Deterministic components

Appendix: Mis-specification testing
Introducing constant and trend

Consider the VAR(1) with a trend coefficient $\mu_1$ and constant $\mu_0$:

$$\Delta x_t = \alpha \beta' x_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t.$$  

$\mu_0$ and $\mu_1$ may be decomposed into mean and trend of $\beta'x_t$ and $\Delta x_t$. Using beautiful identity,

$$\mu_0 = (\alpha (\beta' \alpha)^{-1} \beta' + \beta_{\perp} (\alpha'_{\perp} \beta_{\perp})^{-1} \alpha'_{\perp}) \mu_0 \equiv \alpha \beta_0 + \gamma_0$$

$$\mu_1 = (\alpha (\beta' \alpha)^{-1} \beta' + \beta_{\perp} (\alpha'_{\perp} \beta_{\perp})^{-1} \alpha'_{\perp}) \mu_1 \equiv \alpha \beta_1 + \gamma_1$$

This gives

$$\Delta x_t = \alpha \beta' x_{t-1} + \alpha \beta_0 + \alpha \beta_1 t + \gamma_0 + \gamma_1 t + \varepsilon_t$$

$$= \alpha (\beta', \beta_0, \beta_1) \begin{pmatrix} x_{t-1} \\ 1 \\ t \end{pmatrix} + \gamma_0 + \gamma_1 t + \varepsilon_t$$

---

$^6$See Juselius (2006), Ch. 6.2
Five cases

1. $\mu_1 = \mu_0 = 0$. No deterministic components in data.

2. $\mu_1 = \gamma_0 = 0$ but $\beta_0 \neq 0$. A constant restricted to be in cointegrating relations.

3. $\mu_1 = 0$ but $\mu_0$ is unrestricted. A constant in cointegrating relations, and linear trend in levels.

4. $\gamma_1 = 0$ but $(\gamma_0, \beta_0, \beta_1) \neq 0$. A trend restricted to be in cointegrating relations, and unrestricted constant.

5. No restrictions on $\mu_0$ or $\mu_1$. Unrestricted trend and constant. Trend cumulates to quadratic trend in levels.

---

See Juselius (2006), Ch. 6.3.
Granger-Johansen representation\(^8\)

Inverting the VAR(1) to give the MA form:

\[
\mathbf{x}_t = \mathbf{C} \sum_{i=1}^{\infty} (\mathbf{Q}_i + \mu_0 + \mu_1 i) + \mathbf{C}^* (L) (\mathbf{Q}_t + \mu_0 + \mu_1 t)
\]

\[
= \mathbf{C} \sum_{i=1}^{t} \mathbf{Q}_t + \mathbf{C} \mu_0 t + \frac{1}{2} \mathbf{C} \mu_1 t + \frac{1}{2} \mathbf{C} \mu_1 t^2 + \mathbf{C}^* (L) \mathbf{Q}_t + \mathbf{C}^* (L) \mu_0
\]

\[+ \mathbf{C}^* (L) \mu_1 t + \tilde{\mathbf{X}}_0\]

when summing over finite sample 1 to \(T\). But

\[
\alpha'_\perp \mu_0 t = \alpha'_\perp \alpha \beta_0 t + \alpha'_\perp \gamma_0 t = \alpha'_\perp \gamma_0 t
\]

\[
\alpha'_\perp \frac{1}{2} \mu_1 t = \frac{1}{2} (\alpha'_\perp \alpha \beta_1 t + \alpha'_\perp \gamma_1 t) = \frac{1}{2} \alpha'_\perp \gamma_1 t
\]

\[
\alpha'_\perp \frac{1}{2} \mu_1 t^2 = \frac{1}{2} (\alpha'_\perp \alpha \beta_1 t^2 + \alpha'_\perp \gamma_1 t^2) = \frac{1}{2} \alpha'_\perp \gamma_1 t^2
\]

\(^8\)See Juselius (2006), Ch. 6.4.
Then

\[
x_t = C \sum_{i=1}^{t} \epsilon_t + C\gamma_0 t + \frac{1}{2} C\gamma_1 t + \frac{1}{2} C\gamma_1 t^2 + C^*(L)\epsilon_t + C^*(L)\mu_0 \\
+ C^*(L)\mu_1 t + \tilde{X}_0
\]

Hence linear trends may originate from three different sources in the VAR model:

1. From the term \(C^*(L)\mu_1 t\) of a restricted or unrestricted linear trend \(\mu_1 t\).
2. From the term \(\gamma_1 t\) of the unrestricted linear trend \(\mu_1 t\).
3. From the term \(\gamma_0 t\) of the unrestricted constant \(\mu_0\).

More compact the Granger-Johansen is given by:

\[
x_t = \tau_0 + \tau_1 t + \tau_2 t^2 + C \sum_{i=1}^{t} \epsilon_t + C^*(L)\epsilon_t + \tilde{X}_0.
\]
Outline

Introduction

Estimation of unrestricted VAR

Non-stationarity

Deterministic components

Appendix: Mis-specification testing
Test for lag length$^9$

Fit model of order $k + 1$

$$X_t = A_1 X_{t-1} + \cdots + A_k X_{t-k} + A_{k+1} X_{t-k-1} + \varepsilon_t$$

Compute likelihood ratio ($LR$) to test $A_{k+1} = 0$:

$$LR(\mathcal{H}_k|\mathcal{H}_{k+1}) = -2 \ln Q(\mathcal{H}_k/\mathcal{H}_{k+1}) = T(\ln |\hat{\Omega}_k| - \ln |\hat{\Omega}_{k+1}|), \quad (2)$$

where $\mathcal{H}_k$ is null hypothesis of $k$ lags, while $\mathcal{H}_{k+1}$ is alternative hypothesis that $k + 1$ lags are needed.

Then $LR \xrightarrow{D} \chi^2(p^2)$. 

---

$^9$See Juselius (2006), Ch. 4.3.2, and Lütkepohl (1991), Nielsen (2001)
Test for residual autocorrelation\textsuperscript{10}

Regress estimated VAR residuals on $k$ lagged variables and $j^{th}$ lagged VAR residual:

$$\hat{\varepsilon}_t = A_1 X_{t-1} + \cdots + A_k X_{t-k} + A_{\varepsilon}\hat{\varepsilon}_{t-j} + \tilde{\varepsilon}_t.$$  

We want $\hat{\varepsilon}_t \approx \tilde{\varepsilon}_t$ and use a Lagrange Multiplier ($LM$) test (calculated as a Wilks’ ratio test) with a small-sample correction:

$$LM(j) = -(T - p(k + 1) - \frac{1}{2})\ln \left( \frac{|\Omega(j)|}{|\hat{\Omega}|} \right),$$

which is approximately distributed as $\chi^2$ with $p^2$ degrees of freedom.

\textsuperscript{10}See Juselius (2006), Ch. 4.3.3.
Test for ARCH\textsuperscript{11}

Compute $R^2$ from auxiliary regression

$$\hat{\varepsilon}_{i,t}^2 = \gamma_0 + \sum_{j=1}^{m} \gamma_j \hat{\varepsilon}_{i,t-j} + u_{i,t}.$$  

If $R^2 = 1 - \sum \tilde{u}_{i,t}^2 / \sum [\hat{\varepsilon}_{i,t}^2 - \text{avg}_t(\hat{\varepsilon}_{i,t}^2)]^2$ is small, variances are likely not autocorrelated.

The $m^{th}$ order ARCH test is calculated as $(T + k - m) \times R^2$, where $T$ is sample size and $k$ the lag length in VAR, and $(T + k - m) \times R^2 \overset{D}{\rightarrow} \chi^2(m)$.

Note that Rahbek et al. (2002) have shown that the cointegration rank tests are robust against moderate residual ARCH effects.

\textsuperscript{11}See Juselius (2006), Ch. 4.3.4, and Engle (1982).
Test for normality\textsuperscript{12}

Fit model of order $k$, and check third and fourth moments of residuals (no skewness and kurtosis of 3 for normality). Calculate:

\[
\text{skewness}_i = T^{-1} \sum_{t=1}^{T} \left( \frac{\hat{\epsilon}_i}{\hat{\sigma}_i} \right)_t^3, \quad \text{and} \quad \text{kurtosis}_i = T^{-1} \sum_{t=1}^{T} \left( \frac{\hat{\epsilon}_i}{\hat{\sigma}_i} \right)_t^4
\]

The test statistic is calculated and asymptotically distributed as

\[
\eta_i^{as} = \frac{T}{6} (\text{skewness}_i)^2 + \frac{T}{24} (\text{kurtosis}_i - 3)^2 \overset{a}{\sim} \chi^2(2).
\]

If the sample is large, one can use the asymptotic multivariate test:

\[
m\eta_i^{as} = \sum_{i=1}^{p} \eta_i^{as} \overset{a}{\sim} \chi^2(2p).
\]

But in small samples skewness and kurtosis are neither asymptotically normal nor independent and one needs to use transformations.

\textsuperscript{12}See Juselius (2006), Ch. 4.3.5, and Doornik and Hansen (1994)
Exercises

• Derive log-likelihood function and estimators of unrestricted VAR in matrix notation. Show $\ln L_{\text{max}} = -\frac{1}{2}T\ln|\hat{\Omega}| + \text{constants}$.

• Derive expression for $\tau_0$ in Granger-Johansen representation when $k = 1$.

• Exam 2007, Question 6 (i),(ii).

• Exam 2006, Question 5 (i)-(iii).

• Exam 2002, Question 7.

• Exercise 3.8, Johansen (1996).

• Exercise 4.1, Johansen (1996).

• Exercise 4.6, Johansen (1996).


• Exercise 5.1, Johansen (1996).

References