Cointegration Lecture I: Introduction

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

Hilary Term 2008

Outline

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

Readings

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.

- Hansen, P. and Johansen, S. (1998) *Workbook on Cointegration*, Oxford: Oxford University Press.
- Harris, R. (1995) *Using Cointegration Analysis in Econometric Modelling*. Hemel Hampstead: Prentice Hall.
- Johansen, S. (1996) *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.
- Juselius, K. (2006) *The Cointegrated VAR Model - Methodology and Applications*. Oxford: Oxford University Press.

Other Practicalities

- **Lecture notes**: These slides. More detail in Juselius (2006).
- **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

A VAR in levels

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

$$\mathbf{x}_t = \Pi_1 \mathbf{x}_{t-1} + ... + \Pi_k \mathbf{x}_{t-k} + \phi \mathbf{D}_t + \varepsilon_t, \qquad t = 1, 2, ..., T$$

where

- $\mathbf{x}_t$ is a vector of the p variables at time t,
- Π_i are $p \times p$ matrices of parameters with $i = 1, ...k$,
- $\mathbf{D}_t$ a vector of deterministic components with a vector of coefficients ϕ; and
- ε_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*:

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

where Ω 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[1]

For simplicity, write unrestricted model as

$$\mathbf{x}_t = \mathbf{B}'\mathbf{Z}_t + \varepsilon_t,$$

where $\mathbf{B}' = (\Pi_1, \Pi_2, ..., \Pi_k, \mu_0)$ and $\mathbf{Z}_t' = (\mathbf{x}_{t-1}', \mathbf{x}_{t-2}', ..., \mathbf{x}_{t-k}', 1)$, assuming that only have constant, i.e. $\phi\mathbf{D}_t = 0$, and that initial conditions are given.

Now consider the log-likelihood function

$$\ln L(\mathbf{B}, \Omega; \mathbf{X}) = -T\frac{p}{2}\ln(2\pi) - T\frac{1}{2}\ln|\Omega| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \mathbf{B}'\mathbf{Z}_t)'\Omega^{-1}(\mathbf{x}_t - \mathbf{B}'\mathbf{Z}_t).$$

[1] See Juselius (2006), Ch. 4.1

Maximising log-likelihood with respect to $\mathbf{B}'$ and Ω^{-1} gives the respective ML estimators:

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

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

A VECM in differences

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

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \ldots + \Gamma_{k-1} \Delta \mathbf{x}_{t-k+1} + \phi \mathbf{D}_t + \varepsilon_t$$

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

If $\mathbf{x}_t$ is integrated of order 1 (I(1), i.e. non-stationary), then:

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

Again, this is testable and we need to look at properties of Π.

The roots of the characteristic polynomial[2]

Consider two-dimensional VAR(2):

$$\begin{aligned} \mathbf{x}_t &= \Pi_1 \mathbf{x}_{t-1} + \Pi_2 \mathbf{x}_{t-2} + \phi \mathbf{D}_t + \varepsilon_t \\ (I - \Pi_1 L - \Pi_2 L^2)\mathbf{x}_t &= \phi \mathbf{D}_t + \varepsilon_t \end{aligned}$$

Then, roots of $|\Pi(z)| = |I - \Pi_1 z - \Pi_2 z^2|$ provide information on stationarity of $\mathbf{x}_t$:

- if the roots of $|\Pi(z)|$ are all outside the unit circle, then $\mathbf{x}_t$ is stationary;
- if some roots are outside and some on the unit circle, then $\mathbf{x}_t$ is non-stationary;
- if any of the roots are inside the unit circle, then $\mathbf{x}_t$ 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 Π-matrix[3]

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

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

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

where α and β are $p \times r$ full-rank matrices. Then:

$$\Delta \mathbf{x}_t = \alpha\beta'\mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \ldots + \Gamma_{k-1} \Delta \mathbf{x}_{t-k+1} + \phi \mathbf{D}_t + \varepsilon_t \quad (1)$$

- $\beta'\mathbf{x}_{t-1}$ is an $r \times 1$ vector of stationary cointegrating relations.
- All variables in (1) are now stationary.
- α 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:

$$\begin{aligned} A\mathbf{x}_t &= \left\{A\left(I_p + \alpha\beta'\right)A^{-1}\right\}A\mathbf{x}_{t-1} + \varepsilon_t \\ \begin{pmatrix} \beta_\perp'\mathbf{x}_t \\ \beta'\mathbf{x}_t \end{pmatrix} &= \begin{pmatrix} I_{p-r} & \beta_\perp'\alpha \\ 0 & I_r + \beta'\alpha \end{pmatrix}\begin{pmatrix} \beta_\perp'\mathbf{x}_{t-1} \\ \beta'\mathbf{x}_{t-1} \end{pmatrix} + \begin{pmatrix} \beta_\perp'\varepsilon_t \\ \beta'\varepsilon_t \end{pmatrix}, \end{aligned}$$

where $\beta_\perp \in \mathbf{R}^{p\times(p-r)}$ is the orthogonal complement of β such that $\beta'\beta_\perp = 0$.

Now find roots by solving eigenvalue problem using $|\rho I - M| = 0$, where ρ is an eigenvalue of square matrix M:

$$\begin{aligned} 0 &= \left| \rho I_p - \begin{pmatrix} I_{p-r} & \beta_\perp' \alpha \\ 0 & I_r + \beta' \alpha \end{pmatrix} \right| \\ &= \begin{vmatrix} (\rho - 1) I_{p-r} & \beta_\perp' \alpha \\ 0 & \rho I_r - (I_r + \beta' \alpha) \end{vmatrix} \\ &= (\rho - 1)^{p-r} \left| \rho I_r - (I_r + \beta' \alpha) \right|. \end{aligned}$$

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 β':

$$\begin{aligned}\Delta\left(\beta'\mathbf{x}_t\right) &= \beta'\alpha\beta'\mathbf{x}_{t-1} + \beta'\varepsilon_t, \quad \text{then} \\ \beta'\mathbf{x}_t &= \left(I_r + \beta'\alpha\right)\beta'\mathbf{x}_{t-1} + \beta'\varepsilon_t \\ &= \sum_{s=0}^{t-1}\left(I_r + \beta'\alpha\right)^s\left(\beta'\varepsilon_{t-s}\right) + \left(I_r + \beta'\alpha\right)^t\beta'\mathbf{x}_0\end{aligned}$$

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

Common trends: Pre-multiply with $\alpha'_\perp$:

$$\alpha'_\perp\Delta\mathbf{x}_t = \left(\alpha'_\perp\alpha\right)\beta'\mathbf{x}_{t-1} + \alpha'_\perp\varepsilon_t = \alpha'_\perp\varepsilon_t$$

and cumulate to see

$$\alpha'_\perp\mathbf{x}_t = \alpha'_\perp\sum_{s=1}^{t}\varepsilon_s + \alpha'_\perp\mathbf{x}_0$$

[4] See Juselius (2006), Ch. 5.4

Use *beautiful identity* for Granger-Johansen representation of VAR(1):

$$\begin{aligned}\mathbf{x}_t &= (\alpha\left(\beta'\alpha\right)^{-1}\beta' + \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\alpha'_\perp)\mathbf{x}_t \\ &\approx \alpha\left(\beta'\alpha\right)^{-1}\text{(stationary process)} \\ &\quad +\beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\left(\alpha'_\perp\sum_{s=1}^{t}\varepsilon_s + \alpha'_\perp\mathbf{x}_0\right).\end{aligned}$$

Theorem

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

$$\mathbf{x}_t \approx \mathbf{C}\sum_{s=1}^{t}\varepsilon_s + \mathbf{C}^*(L)\varepsilon_t + \tilde{\mathbf{X}}_0,$$

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

Pulling and pushing forces[5]

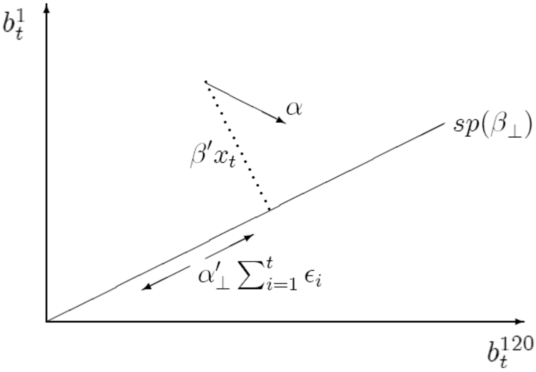

Figure: The process $\mathbf{x}_t = [b_t^{120}, b_t^1]$ is pushed along the attractor set by the common trends and pulled towards the attractor set by the adjustment coefficients.

[5] See Juselius (2006), Ch. 5.5.

Outline

Introducing constant and trend[6]

Consider the VAR(1) with a trend coefficient μ_1 and constant μ_0:

$$\Delta \mathbf{x}_t = \alpha\beta'\mathbf{x}_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t.$$

μ_0 and μ_1 may be decomposed into mean and trend of $\beta'\mathbf{x}_t$ and $\Delta\mathbf{x}_t$. Using *beautiful identity*,

$$\begin{aligned} \mu_0 &= (\alpha\left(\beta'\alpha\right)^{-1}\beta' + \beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp')\mu_0 \equiv \alpha\beta_0 + \gamma_0 \\ \mu_1 &= (\alpha\left(\beta'\alpha\right)^{-1}\beta' + \beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp')\mu_1 \equiv \alpha\beta_1 + \gamma_1 \end{aligned}$$

This gives

$$\begin{aligned} \Delta\mathbf{x}_t &= \alpha\beta'\mathbf{x}_{t-1} + \alpha\beta_0 + \alpha\beta_1 t + \gamma_0 + \gamma_1 t + \varepsilon_t \\ &= \alpha\left(\beta', \beta_0, \beta_1\right)\begin{pmatrix} \mathbf{x}_{t-1} \\ 1 \\ t \end{pmatrix} + \gamma_0 + \gamma_1 t + \varepsilon_t \end{aligned}$$

[6] See Juselius (2006), Ch. 6.2

Five cases[7]

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 μ_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 μ_0 or μ_1. Unrestricted trend and constant. Trend cumulates to *quadratic* trend in levels.

[7] See Juselius (2006), Ch. 6.3.

Granger-Johansen representation[8]

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

$$\begin{aligned} \mathbf{x}_t &= \mathbf{C}\sum_{i=1}^{\infty}(\varepsilon_i + \mu_0 + \mu_1 i) + \mathbf{C}^*(L)(\varepsilon_t + \mu_0 + \mu_1 t) \\ &= \mathbf{C}\sum_{i=1}^{t}\varepsilon_t + \mathbf{C}\mu_0 t + \frac{1}{2}\mathbf{C}\mu_1 t + \frac{1}{2}\mathbf{C}\mu_1 t^2 + \mathbf{C}^*(L)\varepsilon_t + \mathbf{C}^*(L)\mu_0 \\ &\quad + \mathbf{C}^*(L)\mu_1 t + \tilde{\mathbf{X}}_0 \end{aligned}$$

when summing over finite sample 1 to T. But

$$\begin{aligned} \alpha'_\perp \mu_0 t &= \alpha'_\perp \alpha \beta_0 t + \alpha'_\perp \gamma_0 t = \alpha'_\perp \gamma_0 t \\ \alpha'_\perp \frac{1}{2}\mu_1 t &= \frac{1}{2}(\alpha'_\perp \alpha \beta_1 t + \alpha'_\perp \gamma_1 t) = \frac{1}{2}\alpha'_\perp \gamma_1 t \\ \alpha'_\perp \frac{1}{2}\mu_1 t^2 &= \frac{1}{2}(\alpha'_\perp \alpha \beta_1 t^2 + \alpha'_\perp \gamma_1 t^2) = \frac{1}{2}\alpha'_\perp \gamma_1 t^2 \end{aligned}$$

[8] See Juselius (2006), Ch. 6.4.

Then

$$\begin{aligned}\mathbf{x}_t &= \mathbf{C}\sum_{i=1}^{t}\varepsilon_t + \mathbf{C}\gamma_0 t + \frac{1}{2}\mathbf{C}\gamma_1 t + \frac{1}{2}\mathbf{C}\gamma_1 t^2 + \mathbf{C}^*(L)\varepsilon_t + \mathbf{C}^*(L)\mu_0 \\ &\quad + \mathbf{C}^*(L)\mu_1 t + \tilde{\mathbf{X}}_0\end{aligned}$$

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

1. From the term $\mathbf{C}^*(L)\mu_1 t$ 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 μ_0.

More compact the Granger-Johansen is given by:

$$\mathbf{x}_t = \tau_0 + \tau_1 t + \tau_2 t^2 + \mathbf{C}\sum_{i=1}^{t}\varepsilon_t + \mathbf{C}^*(L)\varepsilon_t + \tilde{\mathbf{X}}_0.$$

Outline

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

[9] See Juselius (2006), Ch. 4.3.2, and Lütkepohl (1991), Nielsen (2001)

Test for residual autocorrelation[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:

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

which is approximately distributed as χ^2 with p^2 degrees of freedom.

[10] See Juselius (2006), Ch. 4.3.3.

Test for ARCH[11]

Compute R^2 from auxiliary regression

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

If $R^2 = 1 - \sum \tilde{u}_{i,t}^2 / \sum [\hat{\varepsilon}_{i,t}^2 - avg_t(\hat{\varepsilon}_{i,t}^2)]^2$ 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 $(T+k-m) \times R^2 \xrightarrow{D} \chi^2(m)$.

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

[11] See Juselius (2006), Ch. 4.3.4, and Engle (1982).

Test for normality[12]

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

$$\text{skewness}_i = T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_i/\hat{\sigma}_i)_t^3, \quad \text{and} \quad \text{kurtosis}_i = T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_i/\hat{\sigma}_i)_t^4$$

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:

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

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

[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 $\ln L_{\max} = -\frac{1}{2}T\ln|\hat{\Omega}| +$ constants.
- Derive expression for τ_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 4.12, 1.-3., Johansen (1996).
- Exercise 5.1, Johansen (1996).
- Exercise 6.1, 1., 2., Johansen (1996).

References

- Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1008.
- Doornik, J.A. and Hansen, H. (1994) An omnibus test for univariate and multivariate normality. Nuffield College economics preprint No. 91.
- Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.
- Nielsen, B. (2001). Order determination in general vector autoregressions. Discussion Paper. (and Exam 2001, 7)
- Rahbek, A., Hansen, E., and Dennis, J.G. (2002) ARCH innovations and their impact on cointegration rank testing. Preprint no. 12, 1998, Department of Theoretical Statistics. Working paper no. 22, Centre for Analytical Finance.