Financial Econometrics - Part VI

Vector Autoregressive model (VAR)

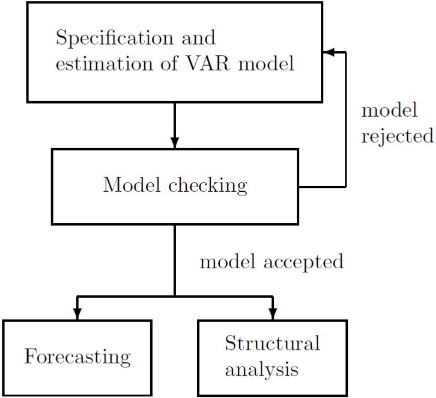
Li Chen
lichen812@xmu.edu.cn

WISE, Xiamen University

Why multivariate models ?

- Univariate models (ARIMA) assume that a time series depends on its own past only
- In practice, many economic time series are inter-dependent dynamically
- A multivariate analysis seems to be more appropriate for a set of economic, financial and business time series
- VAR, a multivariate model, popularized by Sims (1980) " *Macroeconomics and Reality*, **Econometrica** 48, 1-48"

Stable (Stationary) VAR analysis



Specification of VAR(p)

K-dimensional VAR(p) model (VAR model of order p):

$$Y_t = \nu + A_1 Y_{t-1} + ... + A_p Y_{t-p} + \varepsilon_t$$

where:

- $Y_t = (Y_{1t}, ..., Y_{Kt})'$ is a $(K \times 1)$ random vector
- A_i are fixed $(K \times K)$ coefficient matrices
- $\nu = (\nu_1, ..., \nu_K)'$ is a fixed $(K \times 1)$ vector of intercepts
- $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Kt})'$ is a K-dimensional *white noise* process

Basic assumptions of VAR(p)

$\varepsilon_t \sim WN(0_K, \Sigma_\varepsilon)$ means:

- $E(\varepsilon_t) = 0_K$: zero-mean K-dimensional vector
- $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$: nonsingular $(K \times K)$ variance-covariance matrix.
- $E(\varepsilon_t \varepsilon_s') = 0_{K \times K}$ for $s \neq t$: zero autocovariance matrix.

An example: a bivariate VAR(1) model

$Y_t = (Y_{1t}, Y_{2t})'$ and

$Y_t = \nu + A_1 Y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim WN(0_K, \Sigma_\varepsilon)$

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \end{bmatrix} + \begin{bmatrix} 0.7 & 0.5 \\ -0.5 & 0.1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where $\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim \left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} \right)$

So $A_1 = \begin{bmatrix} 0.7 & 0.5 \\ -0.5 & 0.1 \end{bmatrix}$ and $\Sigma_\varepsilon = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$

An Example: In Equation Forms

$$Y_{1t} = 10 + 0.7Y_{1t-1} + 0.5Y_{2t-1} + \varepsilon_{1t}$$

$$Y_{2t} = 5 - 0.5Y_{1t-1} + 0.1Y_{2t-1} + \varepsilon_{2t}$$

So, Y_{1t} does not only depend on its own past value but also on other's (Y_{2t}) and so does Y_{2t}

Polynomial Lag Representation of VAR(p)

K-dimensional VAR(p):

$$Y_t = \nu + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t$$

$$\Rightarrow Y_t - A_1 Y_{t-1} - \dots - A_p Y_{t-p} = \nu + \varepsilon_t$$

$$\Rightarrow (I_K - A_1 L - \dots - A_p L^p) Y_t = \nu + \varepsilon_t$$

$$\Rightarrow A(L) Y_t = \nu + \varepsilon_t$$

Let $A(z) = I_K - A_1 z - \dots - A_p z^p$

I_K: a K-dimensional identity matrix

Stationarity of VAR(p): a formal condition

- A VAR(p) model is stationary if all the roots of

  $$det(A(z)) = det(I_K - A_1 z - ... - A_p z^p) = 0$$

  are greater than 1 in *absolute value*

  (outside the unit circle when roots are *complex numbers*)
- $det(A)$ denotes the determinant of a matrix A

Stationarity of VAR(p): Examples

Find whether the following trivariate VAR(1) is stationary:

$$Y_t = \nu + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} Y_{t-1} + \varepsilon_t$$

Solution: Find the roots of $det(I_K - A_1 z) = 0$

$$det\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} z \right)$$

$$= det \begin{bmatrix} 1 - 0.5z & 0 & 0 \\ -0.1z & 1 - 0.1z & -0.3z \\ 0 & -0.2z & 1 - 0.3z \end{bmatrix} = 0$$

Stationarity of VAR(p): Examples

$$det \begin{bmatrix} 1-0.5z & 0 & 0 \\ -0.1z & 1-0.1z & -0.3z \\ 0 & -0.2z & 1-0.3z \end{bmatrix}$$

$$= (1-0.5z)(1-0.4z-0.03z^2) = 0$$

The roots are: $z_1 = 2, z_2 = 2.1525$, and $z_3 = -15.4858$

All of them are greater than 1 in absolute value $\Rightarrow Y_t$ is stationary.

Stationarity of VAR(p): Examples

Find whether the following bivariate VAR(2) is stationary:

$$Y_t = \nu + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} Y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} Y_{t-2} + \varepsilon_t$$

Solution: Find the roots of $det(I_K - A_1 z - A_2 z^2) = 0$

$$det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} z^2 \right)$$

$$= 1 - z + 0.21z^2 - 0.025z^3 = 0$$

The roots are: $z_1 = 1.3, z_2 = 3.55 + 4.26i$, and $z_3 = 3.55 - 4.26i$

The process Y_t is stationary.

Non-Stationarity of VAR(p): An Informal Condition

- If a K-dimensional VAR(1) model is non-stationary with *at least* one unit root, then $det(A_1 - I_K) = 0$
- In the general K-dimensional VAR(p) model, the condition is: $det(A_1 + A_2 + ... + A_p - I_K) = 0$
- This condition is useful in **Cointegration analysis**, which will be discussed in some next lectures.

Infinite MA Representation of a stable VAR(p)

K-dimensional VAR(p):

$$Y_t = \nu + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \varepsilon_t$$

$$\Rightarrow A(L) Y_t = \nu + \varepsilon_t$$

If VAR(p) is stationary then $A(L)$ is invertible, VAR(p) can be represented as an infinite MA process

Let $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$ such that $\Phi(L)A(L) = I_K$

So

$$Y_t = \Phi(L)\nu + \Phi(L)\varepsilon_t$$

Or $Y_t = A(1)^{-1}\nu + \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i}$

How to find coefficient matrices Φ_i

We have:

$$I_K = \Phi(L)A(L)$$

$$\Rightarrow I_K = (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \ldots)(I_K - A_1 L - \ldots - A_p L^p)$$

$$\Rightarrow I_K = \Phi_0 + (\Phi_1 - \Phi_0 A_1)L + (\Phi_2 - \Phi_1 A_1 - \Phi_0 A_2)L^2 + \ldots + \left(\Phi_i - \sum_{j=1}^{i} \Phi_{i-j} A_j\right) L^i + \ldots$$

Matching coefficient matrices:

$$\begin{aligned} I_K &= \Phi_0 \\ 0 &= \Phi_1 - \Phi_0 A_1 \qquad \ldots \\ 0 &= \Phi_i - \sum_{j=1}^{i} \Phi_{i-j} A_j \qquad \ldots \end{aligned}$$

How to find coefficient matrices Φ_i

So we can obtain Φ_i as:

$$\begin{aligned}\Phi_0 &= I_K \\ \Phi_i &= \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \ldots\end{aligned}$$

where $A_j = 0$ for $j > p$.

Infinite MA Representation: An Example

Find the coefficient matrices Φ_i of a following bivariate VAR(2) model:

$$Y_t = \nu + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} Y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} Y_{t-2} + \varepsilon_t$$

We have:

$$\begin{aligned}
\Phi_0 &= I_2 \\
\Phi_1 &= \Phi_0 A_1 = A_1 \\
\Phi_2 &= \Phi_1 A_1 + A_2 = A_1^2 + A_2 \\
\Phi_3 &= \Phi_2 A_1 + \Phi_1 A_2 = A_1^3 + A_2 A_1 + A_1 A_2
\end{aligned}$$

So, $\Phi_1 = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix}, \Phi_2 = \begin{bmatrix} 0.29 & 0.1 \\ 0.65 & 0.29 \end{bmatrix}, \Phi_3 = \begin{bmatrix} 0.21 & 0.079 \\ 0.566 & 0.21 \end{bmatrix}$

VAR modelling: Advantages and Disadvantages

1. Advantages of VAR modelling
   - All variables are endogenous
   - More general than ARMA modelling
   - Can simply use OLS equation by equation
   - Forecasts often better than "traditional structural" models

2. Problems with VAR modelling
   - Lag length decision
   - Large number of parameters
   - Do all components of VARs need to be stationary?
   - Coefficients' interpretation
   - Robustness

Specification of VAR(p)

K-dimensional VAR(p) model (VAR model of order p):

$$Y_t = \nu + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \varepsilon_t$$

where:

- $Y_t = (Y_{1t}, \ldots, Y_{Kt})'$ is a $(K \times 1)$ random vector
- A_i are fixed $(K \times K)$ coefficient matrices
- $\nu = (\nu_1, \ldots, \nu_K)'$ is a fixed $(K \times 1)$ vector of intercepts
- $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ is a K-dimensional *white noise* process

Estimation

- Parameters can be estimated using the least squares (LS) method. The LS estimator is the estimator that minimizes the variance of the innovation processes (ε_t).
- Multivariate (generalized) LS estimation or equation-by-equation LS estimation
- Other estimation methods
  - Maximum Likelihood
  - Yule-Walker estimation
  - ...
- LS estimation is preferred due to its simplicity and better small sample properties

Multivariate LS estimator

Assumptions:

- T is the sample size for each of the K variables
- p pre–sample values for each variables, $y_{-p+1}, ..., y_0$ are available

Notations:

- $\mathbf{Y} := (Y_1, ..., Y_T) \qquad (K \times T)$
- $B := (\nu, A_1, ..., A_p) \qquad (K \times (Kp + 1))$
- $Z_t := \begin{bmatrix} 1 \\ Y_t \\ \vdots \\ Y_{t-p+1} \end{bmatrix} \qquad ((Kp + 1) \times 1)$

Multivariate LS estimator (con't)

Notations (con't):

- $Z := (Z_0, ..., Z_{T-1}) \quad ((Kp+1) \times T)$
- $U := (\varepsilon_1, ..., \varepsilon_T) \quad (K \times T)$
- $\mathbf{y} := vec(\mathbf{Y}) \quad (KT \times 1)$
- $\beta := vec(B) \quad ((K^2p + K) \times 1)$
- $\mathbf{u} := vec(U) \quad (KT \times 1)$

K**-dimensional VAR(**p**)**:

$$Y_t = \nu + A_1 Y_{t-1} + ... + A_p Y_{t-p} + \varepsilon_t \qquad \varepsilon_t \sim WN(0_K, \Sigma_\varepsilon)$$

can be written as:

$$\mathbf{Y} = BZ + U$$

Multivariate LS estimator (con't)

Using *vec* operator:

$$\begin{aligned} vec(\mathbf{Y}) &= vec(BZ) + vec(U) \\ &= (Z' \otimes I_K) vec(B) + vec(U) \end{aligned}$$

Therefore,

$$\mathbf{y} = (Z' \otimes I_K)\beta + \mathbf{u} \quad \text{where} \quad \Sigma_{\mathbf{u}} = I_T \otimes \Sigma_\varepsilon$$

The multivariate LS estimation of β: find the estimator $\hat{\beta}$ that minimizes

$$\begin{aligned} S(\beta) &= \mathbf{u}'(I_T \otimes \Sigma_\varepsilon)^{-1}\mathbf{u} = \mathbf{u}'(I_T \otimes \Sigma_\varepsilon^{-1})\mathbf{u} \\ &= [\mathbf{y} - (Z' \otimes I_K)\beta]'(I_T \otimes \Sigma_\varepsilon^{-1})[\mathbf{y} - (Z' \otimes I_K)\beta] \end{aligned}$$

Multivariate LS estimator (con't)

Expand the multiplication:

$$S(\beta) = \mathbf{y}'(I_T \otimes \Sigma_\varepsilon^{-1})\mathbf{y} + \beta'(ZZ' \otimes \Sigma_\varepsilon^{-1})\beta - 2\beta'(Z \otimes \Sigma_\varepsilon^{-1})\mathbf{y}$$

Hence,

$$\frac{\partial S(\beta)}{\partial \beta} = 2(ZZ' \otimes \Sigma_\varepsilon^{-1})\beta - 2(Z \otimes \Sigma_\varepsilon^{-1})\mathbf{y}$$

Estimator $\hat{\beta}$ that minimizes $S(\beta)$ makes $\frac{\partial S(\beta)}{\partial \beta} = 0$, so:

$$(ZZ' \otimes \Sigma_\varepsilon^{-1})\hat{\beta} = (Z \otimes \Sigma_\varepsilon^{-1})\mathbf{y}$$

Finally, we have

$$\begin{aligned}\hat{\beta} &= ((ZZ')^{-1} \otimes \Sigma_\varepsilon)(Z \otimes \Sigma_\varepsilon^{-1})\mathbf{y} \\ &= ((ZZ')^{-1}Z \otimes I_K)\mathbf{y}\end{aligned}$$

Ordinary LS (OLS) estimator

The OLS estimator ($\hat{\beta}_{OLS}$) obtained by minimizing

$$\bar{S}(\beta) = \mathbf{u}'\mathbf{u} = [\mathbf{y} - (Z' \otimes I_K)\beta]'[\mathbf{y} - (Z' \otimes I_K)\beta]$$

The $\hat{\beta}_{OLS}$ is identical to the multivariate LS estimator $\hat{\beta}$:

$$\hat{\beta}_{OLS} = \hat{\beta} = ((ZZ')^{-1}Z \otimes I_K)\mathbf{y}$$

Asymptotic properties of the LS estimators

The LS estimators are asymptotically multivariate normal distributed:

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Gamma^{-1} \otimes \Sigma_\varepsilon)$$

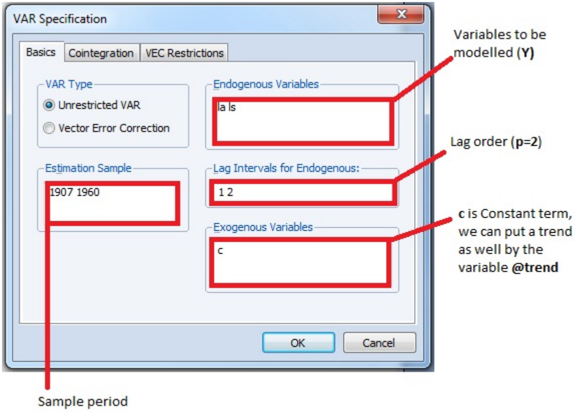
where $\Gamma = plim\ ZZ'/T$ (*plim* means *converge in probability*, see Appendix C.1 Lutkepohl, 2005)

Consistent estimators of Γ and Σ_ε are:

$$\begin{aligned}\hat{\Gamma} &= ZZ'/T \\ \hat{\Sigma}_\varepsilon &= \frac{1}{T - Kp - 1}\mathbf{Y}(I_T - Z'(ZZ')^{-1}Z)\mathbf{Y}'\end{aligned}$$

VAR estimation in Eviews

VAR(p) can be estimated easily in Eviews by option **Quick/Estimate VAR**. Here is how to estimate a bivariate VAR(2)



Order selection by assessing the residuals

- Choose the order (number of lags) so that the residuals (from each equation) mimic a white noise process
- Apply the Ljung-Box (Q) white noise test to residuals from each equation
- Also inspect residual Autocorrelation Function (ACF)
- Start from a low order, for example VAR(1), and stop when residuals are found to be white noise

Order selection by Information Criteria (IC)

- Use one of ICs: AIC, BIC or HQ
- The general form of the IC is:

  $IC = ln|\hat{\Sigma}_{\varepsilon}| + cm/T$

  - m: the number of the parameters to be estimated
  - c: penalty term for m
  - T: sample size

Order selection by Information Criteria (IC) (con't)

- Different criteria adopt different penalty terms
- $AIC : c = 2, \; BIC : c = ln(T), \; HQ : c = 2ln(ln(T))$
- *BIC* gives the heaviest penalty, while *AIC* the lightest
- *AIC* tends to over-estimate, while *BIC* tends to under-estimate the true order
- AIC is widely used in practical research as suggested by Johansen. However, for cases of small sample sizes, BIC seems to be more appropriate because of its heaviest penalty (consequently, lowest order).

Order selection by Information Criteria (IC) (con't)

Select the order p in VAR(p) by IC with Eviews: In VAR estimation window, choose **View/Lag Structure/Lag length criteria**

Var: UNTITLED Workfile: ZANIAS0%5B1%5D::Zanias\

View | Proc | Object | Print | Name | Freeze | Estimate | Stats | Impulse | Resids

VAR Lag Order Selection Criteria
Endogenous variables: LA LS
Exogenous variables: C
Date: 04/03/12 Time: 22:23
Sample: 1907 1960
Included observations: 44

Lag	LogL	LR	FPE	AIC	SC	HQ
0	3.664394	NA	0.003178	-0.075654	0.005445	-0.045579
1	46.93681	80.64405	0.000533	-1.860764	-1.617466*	-1.770537
2	53.39796	11.45385*	0.000478	-1.972634	-1.567137	-1.822256
3	58.99517	9.413496	0.000446	-2.045235	-1.477538	-1.834706
4	64.78308	9.208043	0.000414*	-2.126504*	-1.396608	-1.855823*
5	68.56477	5.672523	0.000422	-2.116580	-1.224485	-1.785748
6	70.04843	2.090612	0.000480	-2.002201	-0.947907	-1.611218
7	74.35469	5.676443	0.000483	-2.016122	-0.799629	-1.564988
8	76.04963	2.080155	0.000552	-1.911347	-0.532655	-1.400062
9	76.33823	0.327946	0.000677	-1.742647	-0.201756	-1.171210
10	77.09923	0.795600	0.000823	-1.595420	0.107670	-0.963832

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Model Diagnostics

- Residual correlograms
- Portmanteau test (Ljung-Box Q test)
- LM test for serial correlation
- Normality test (Jarque-Bera)
- White's heteroskedasticity test
- Stationarity check: Eviews provides *VAR stability condition check* (In VAR estimation window, choose **View/Lag Structure/AR Roots table or AR Roots graph**)

Point Forecasts

- The point forecast of variables at time t for horizon h:

$$E_t(Y_{t+h}) := E(Y_{t+h}|\Omega_t)$$

where Ω_t is the *information set* available at time t.

- Optimality of the conditional expectation implies:

$$E_t(Y_{t+h}) = \nu + A_1 E_t(Y_{t+h-1}) + \ldots + A_p E_t(Y_{t+h-p})$$

- Hence, we can calculate h-step forecast starting with $h = 1$ recursively as:

$$\begin{aligned} E_t(Y_{t+1}) &= \nu + A_1 Y_t + \ldots + A_p Y_{t-p+1} \\ E_t(Y_{t+2}) &= \nu + A_1 E_t(Y_{t+1}) + A_2 Y_t + \ldots + A_p Y_{t-p+1} \\ &\vdots \end{aligned}$$

Point Forecasts - Examples

Given a trivariate VAR(1) with $Y_t = (-6, 3, 5)'$:

$$Y_t = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} Y_{t-1} + \varepsilon_t$$

Easily we can calculate the point forecasts as:

$$E_t(Y_{t+1}) = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} -6 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -3 \\ 3.2 \\ 3.1 \end{bmatrix}$$

$$E_t(Y_{t+2}) = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} -3 \\ 3.2 \\ 3.1 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 2.95 \\ 2.57 \end{bmatrix}$$

Point Forecasts – Examples

Given a bivariate VAR(2)

Assume $Y_t = (0.06, 0.03)'$ and $Y_{t-1} = (0.055, 0.03)'$:

$$Y_t = \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} Y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} Y_{t-2} + \varepsilon_t$$

Point Forecasts – Examples

Easily we can calculate the point forecasts as:

$$
\begin{aligned}
E_t(Y_{t+1}) &= \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} \begin{bmatrix} 0.06 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} 0.055 \\ 0.03 \end{bmatrix} \\
&= \begin{bmatrix} 0.053 \\ 0.08275 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
E_t(Y_{t+2}) &= \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} \begin{bmatrix} 0.053 \\ 0.08275 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} 0.06 \\ 0.03 \end{bmatrix} \\
&= \begin{bmatrix} 0.0548 \\ 0.1076 \end{bmatrix}
\end{aligned}
$$

Granger Causality

- A concept of causality is defined in Granger (1969).
- The idea is if a variable x affects another variable z, the former should help improving the predictions of the latter variable.
- x_t *Granger-causes* z_t if an inclusion of the information in the x_t and its past values (i.e., value of $x_t, x_{t-1}, x_{t-2}...$) may help to improve forecasts of z_t (i.e., $E_t(z_{t+1}), E_t(z_{t+2})...$).
- A *feedback system* is a process $(x_t, z_t)'$ where x_t *Granger-causes* z_t and *vice versa*.
- Granger Causality can be easily applied to a stable VAR model.

Granger Causality in a stationary (stable) VAR

- Consider a bivariate time series (Y_{1t}, Y_{2t}).
- If Y_{1t} can be more efficiently predicted when the information in Y_{2t} is taken in account in addition to all other information, then Y_{2t} *Granger-causes* Y_{1t}.
- So, in a VAR model if Y_{2t-i} significantly affects Y_{1t}, then Y_{2t} *Granger-causes* Y_{1t}.
- Specification of a bivariate VAR(p) model:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \ldots + \begin{bmatrix} a_{11,p} & a_{12,p} \\ a_{21,p} & a_{22,p} \end{bmatrix} \begin{bmatrix} Y_{1t-p} \\ Y_{2t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Granger Causality in a stable VAR: A formal test

- Test whether Y_{2t} Granger-causes Y_{1t}:
  $H_0: a_{12,1} = \ldots = a_{12,p} = 0$ (Y_{2t} does not Granger-causes Y_{1t})
  $H_1:$ At least one of $\underset{(i=1,\ldots,p)}{a_{12,i}} \neq 0$ (Y_{2t} Granger-causes Y_{1t})
- Test whether Y_{1t} Granger-causes Y_{2t}:
  $H_0: a_{21,1} = \ldots = a_{21,p} = 0$ (Y_{1t} does not Granger-causes Y_{2t})
  $H_1:$ At least one of $\underset{(i=1,\ldots,p)}{a_{21,i}} \neq 0$ (Y_{1t} Granger-causes Y_{2t})
- This causality testing can be done by performing a F-test (use *Wald test* in Eviews).

Instantaneous Causality

- Instantaneous causality *between* Y_{1t} and Y_{2t} means that at time t, adding Y_{1t+1} to the information set helps to improve the forecast of Y_{2t+1}.
- Hence, instantaneous causality is the causality *between* two time series at the *same time periods*.
- Instantaneous causality may indicate there is a significant contemporaneous relationship *between* two time series.
- This concept of causality is really symmetric (Lutkepohl, 2005, Proposition 2.3), that is:
  - If there is instantaneous causality between Y_{1t} and Y_{2t}, then there is also instantaneous causality between Y_{2t} and Y_{1t}.
  - Hence, we do not use the statement "instantaneous causality *from* Y_{1t} *to* Y_{2t}" at the beginning.

Instantaneous Causality in VAR

- Consider variance-covariance matrix of error terms, Σ_ε, in a bivariate VAR model:

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

- If $\sigma_{12} = \sigma_{21} = 0$, then there is no instantaneous causality between Y_{1t} and Y_{2t}.
- Direction of instantaneous causality is unknown.
- Extension to K-dimensional VAR is similar:
  - If $\sigma_{ij} = \sigma_{ji} = 0 \;\; (i, j = 1, ..., K; \; \textit{and} \; i \neq j)$, then there is no instantaneous causality between Y_{it} and Y_{jt}.

Instantaneous Causality in VAR: A formal test

- Hypothesis:

$$
\begin{aligned}
H_0 : \rho_{12} &= 0 \;\; (\textit{There is no instantaneous causality}) \\
H_1 : \rho_{12} &\neq 0 \;\; (\textit{There is instantaneous causality})
\end{aligned}
$$

- Test statistic: $\chi_{stat} = T|\hat{\rho}_{12}| \sim \chi^2_{(1)}$ under H_0
- Decision rule: Reject H_0 if $\chi_{stat} > \chi^2_{\alpha,1}$ (5% critical value $\chi^2_{0.05,1} = 3.84$)
- Where:
  - ρ_{12} is the correlation between ε_{1t} and ε_{2t}: $\hat{\rho}_{12} = \dfrac{\hat{\sigma}_{12}}{\hat{\sigma}_1 \hat{\sigma}_2}$.
  - T is the sample size.

Impulse response analysis

- Consider a system including **an inflation rate** and **an interest rate**. The effect of an increase in inflation rate, caused by *exogenous positive shock*, to interest rate may be of interest.
    - For example: An increase of the oil price in 1973/1974 when the OPEC agreed on a joint action to raise prices.
    - Such event can be considered as an exogenous price shock to economy which causes an increase in inflation rate.
- The Causality analysis studied so far may not tell a complete story about the interaction between variables of a system regarding the effect of shocks to the system.
- Impulse response analysis helps to understand the *dynamic responses* of one variable to an exogenous shock (*impulse*) to another variable in a system.

Basic IRF

Given a bivariate VAR(1):

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad \text{where} \quad \Sigma_\varepsilon = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Consider a unit shock to Y_{1t} in period $t = 0$

$$\begin{bmatrix} Y_{1,0} \\ Y_{2,0} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1,0} \\ \varepsilon_{2,0} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{then in the next 3 periods we have,}$$

$$\begin{bmatrix} Y_{1,1} \\ Y_{2,1} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix};$$

$$\begin{bmatrix} Y_{1,2} \\ Y_{2,2} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.3 \end{bmatrix}; \text{ similarly, } \begin{bmatrix} Y_{1,3} \\ Y_{2,3} \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.155 \end{bmatrix}$$

Basic IRF (con't)

Consider a unit shock to Y_{2t} in period $t = 0$

$$\begin{bmatrix} Y_{1,0} \\ Y_{2,0} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1,0} \\ \varepsilon_{2,0} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{then in the next 3 periods we have,}$$

$$\begin{bmatrix} Y_{1,1} \\ Y_{2,1} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix};$$

$$\begin{bmatrix} Y_{1,2} \\ Y_{2,2} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.01 \end{bmatrix}; \text{ similarly, } \begin{bmatrix} Y_{1,3} \\ Y_{2,3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0.001 \end{bmatrix}$$

- Note: In this case, the unit shock to Y_{2t} does not have any effect on Y_{1t}.

Basic IRF (con't)

Recall the MA(∞) representation of a stable VAR(p):

$Y_t = A(L)^{-1}\nu + \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i}$ where:

$$\begin{aligned} \Phi_0 &= I_K \\ \Phi_i &= \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \ldots \end{aligned}$$

$\Rightarrow \; Y_t = A(L)^{-1}\nu + \Phi_0 \varepsilon_t + \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2} + \ldots$

In case the Σ_ε is diagonal, the IRF after h period of VAR(p) is:

$$\frac{\partial Y_t}{\partial \varepsilon_{t-h}} = \Phi_h$$

Basic IRF (con't): An example in VAR(1)

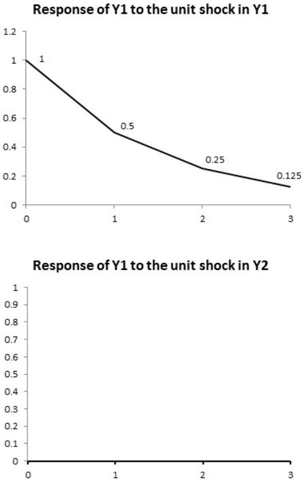
Recall the previous bivariate VAR(1):

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad \text{where} \quad \Sigma_\varepsilon = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Easily, in **VAR(1)** we can obtain $\Phi_h = A^h$. So the impulse response matrices in the next 3 periods are:

$$\begin{aligned} \Phi_1 &= A = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.1 \end{bmatrix} \\ \Phi_2 &= A^2 = \begin{bmatrix} 0.25 & 0 \\ 0.3 & 0.01 \end{bmatrix} \\ \Phi_3 &= A^3 = \begin{bmatrix} 0.125 & 0 \\ 0.155 & 0.001 \end{bmatrix} \end{aligned}$$

Graph of basic IRF



Basic IRF (con't)

- In case of *no instantaneous causality* (Σ_ε *is diagonal*), the IRFs of a stable VAR are the elements of the MA(∞) coefficients.
- Hence, the impulse response analysis is sometimes called *multiplier analysis*.
- The accumulated value of IRFs, $\Psi_\infty = \sum_{i=0}^{\infty} \Phi_i$, represents the long-run multiplier or total responses.
- (i, j) element of Φ_h indicates the h-period ahead response of ith variable to a unit shock to the innovation of jth variable.

Orthogonalized Impulse Response Function

- Basic IRF that we studied so far ignores the instantaneous causality among variables in the VAR.
- This is very unlikely in practice since the VAR does not accommodate contemporaneous relations among variables. The variance-covariance matrix of error terms in the VAR is not diagonal in practice.
- A shock to one variable, therefore, is generally correlated with or dependent on shocks to other variables.
- Sims (1980) solves this problem by introducing the orthogonalized approach for an IRF. So, we often call this as Orthogonalized Impulse Response Function.

Orthogonalized Impulse Response Function (con't)

- Let P be the $(K \times K)$ lower triangular matrix such as, $\Sigma_\varepsilon = PP'$ (Cholesky Decomposition)
- $\Sigma_\varepsilon = PP' \;\Rightarrow\; P^{-1}\Sigma_\varepsilon P'^{-1} = I_K$
- $\varepsilon_t \sim (0, \Sigma_\varepsilon) \;\Rightarrow\; P^{-1}\varepsilon_t \sim (0, I_K)$.
- Set $u_t = P^{-1}\varepsilon_t$, then the variance-covariance matrix of u_t is I_K (i.e., diagonal)
- The MA(∞) representation of a stable VAR(p) now become:

$$Y_t = A(L)^{-1}\nu + \Phi_0 P u_t + \Phi_1 P u_{t-1} + \Phi_2 P u_{t-2} + \ldots$$

Orthogonalized Impulse Response (con't)

- The OIRF after h period can be obtained as:

$$\frac{\partial Y_t}{\partial u_{t-h}} = \Phi_h P$$

- An example: find P in a bivariate system

$$\begin{aligned}
\Sigma_\varepsilon = PP' &= \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{(1-\rho^2)\sigma_2^2} \end{bmatrix} \begin{bmatrix} \sigma_1 & \rho\sigma_2 \\ 0 & \sqrt{(1-\rho^2)\sigma_2^2} \end{bmatrix}
\end{aligned}$$

- P imposes a restriction on the direction of instantaneous causality. Here, (1,2) element of P is 0 so the instantaneous causality runs from Y_{1t} to Y_{2t} but not in the opposite direction.

Wold Causality

- For orthogonalized impulse response analysis in VAR, the variables should be ordered according to the Wold Causality.
- This causality is the instantaneous causal ordering of the variables.
- If $Z = (Y_1, Y_2, Y_3)'$, then the instantaneous causality should run from Y_1 to Y_3.
- The ordering should be determined subjectively or based on economic reasoning.

Generalized Impulse Response

- The Cholesky Decomposition, $\Sigma_\varepsilon = PP'$, requires to determine the direction of the instantaneous causality between variables in the system.

- This could be problematic in case of high dimensional systems and there is no clear economic guidance on how to order the variables.

- Pesaran and Shin (1998) provides an alternative to overcome this potential problem. The approach is called *Generalized Impulse Response* function.

- The Generalized approach does not require to order the variables. In other words, it is *identical* to alternative ordering of the variables.

Generalized Impulse Response (con't)

- The GIRF of Y_t at horizon h:

$$GIR(h, \delta, \Omega_{t-1}) = E(Y_{t+h}|\varepsilon_t = \delta, \Omega_{t-1}) - E(Y_{t+h}|\Omega_{t-1})$$

where $\delta = (\delta_1, \delta_2, ..., \delta_K)'$: $(K \times 1)$ vector of shocks at time t.

- It is the difference between the conditional expectation of Y_{t+h} at time $t+h$ after incorporating the shock's effect at time t and that conditional expectation without the shock's effect, given the information set available at time $t-1$, Ω_{t-1}.

- The technique used in GIRF is to shock only one element of ε_t, then integrating out the effects of other shocks.

Generalized Impulse Response (con't)

- If $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$ and $\delta_i = \sigma_i$, then:

$$GIR(h, \delta, \Omega_{t-1}) = \Phi_h \Sigma_\varepsilon \Theta$$

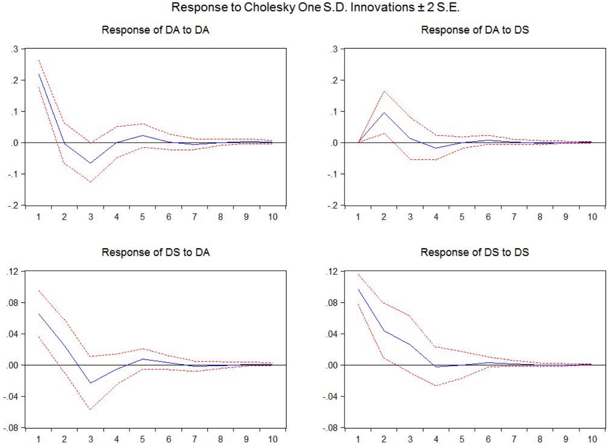
- where:

$$\Theta = \begin{bmatrix} \sigma_1^{-1} & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_K^{-1} \end{bmatrix}$$

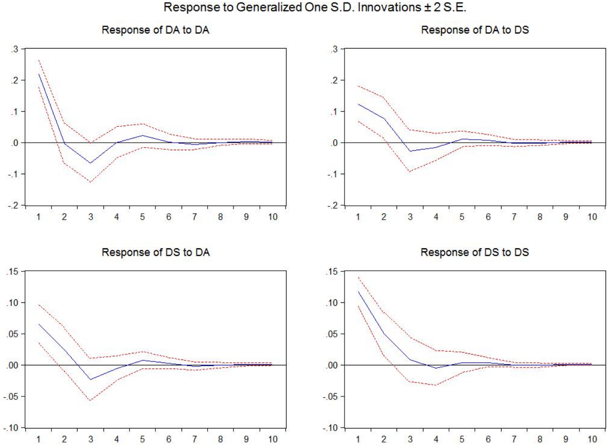
Estimation and Statistical Inference for IRF

- The unknown impulse response values can be estimated using LS estimator for VAR parameters.
- Statistical inference can be conducted using asymptotic normality formula or Monte Carlo simulation.
- Eviews provides 95% confidence band with which statistical significance of impulse response values can be tested.

Graph of Orthogonalized Impulse Response



Graph of Generalized Impulse Response



Variance Decomposition

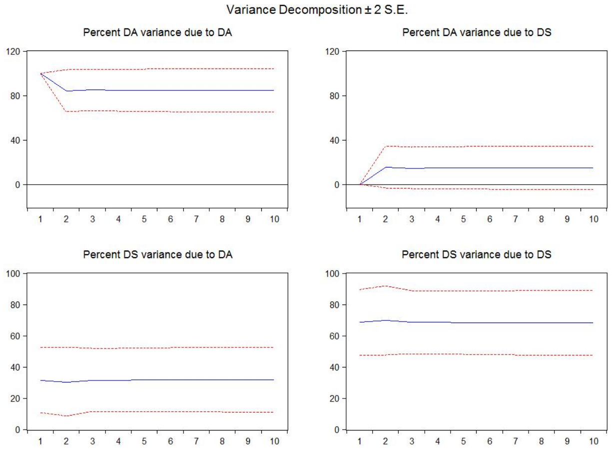
- Variance decomposition explains the proportion that an exogenous shock to one variable contributes to the forecast error variance of other variables in the system.

- If a shock to Y_1 does not explain the forecast error variance of Y_2, then Y_1 does not cause Y_2 (Y_2 is purely exogenous).

- If a shock to Y_1 contributes to the forecast error variance of Y_2, then Y_1 causes Y_2.

- If a shock to Y_1 explains all of the forecast error variance of Y_2, then Y_2 is purely endogenous.

Graph of Variance Decomposition



Reference

- Granger (1969), *"Investigating causal relations by econometrics models and cross-spectral methods"*, **Econometrica** 37, 424-438.
- Sims (1980) *"Macroeconomics and Reality"*, **Econometrica** 48, 1-48.
- Peseran and Shin (1998) *"Generalized impulse response analysis in linear multivariate models"*, **Economics letters** 58, 17-29.
- Lutkepohl (2005) *"New Introduction to Multiple Time Series"*, Springer-Verlag (Part I).

The data. The aggregate US data on consumption, income, investment and interest rate are obtained from *Federal Reserve Economic Data (FRED)*.
We consider a quarterly data set over 1960:1-2009:3 with 199 observations.
Let r_t stand for the real interest rate, and
$c_t = \log(C_t)$, $i_t = \log(I_t)$ and $v_t = \log(V_t)$,
where C_t, I_t and V_t are the consumption expenditures, disposable incomes and investments, respectively, for $t = 1, \cdots, 199$.

The tutor will show you during the tutorial today about how to calculate the ADF test value for each of the data sets.

Figure: The real data

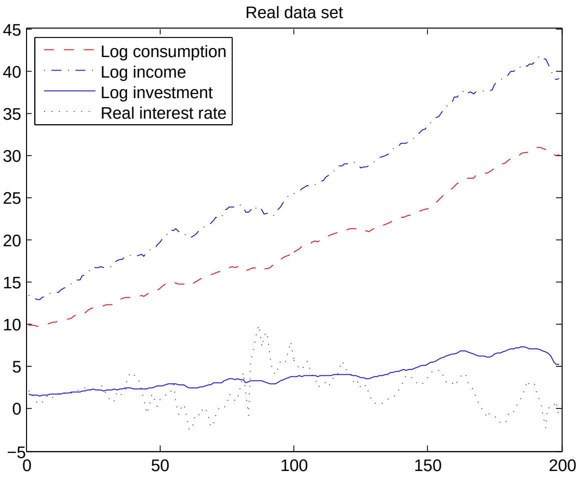
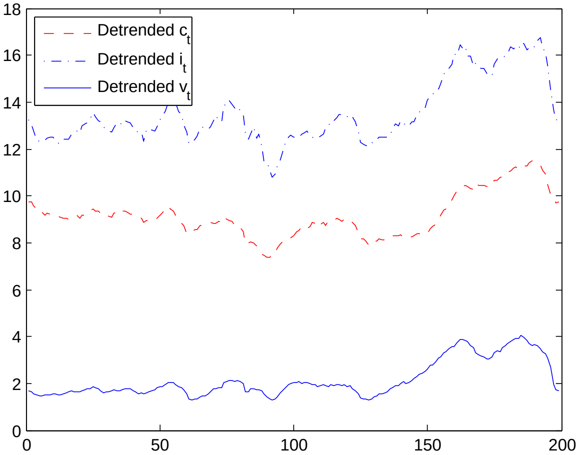


Figure: The de–trended real data



Cointegration analysis: A case of two time series

- Consider a time series $y_t \sim I(d)$, then y_t is said to be *integrated of order d*. That is, process y_t contains d unit roots or stochastic trends.

- If two time series y_{1t} and y_{2t} are *integrated of the same order* and have *common stochastic trends*, then they are *cointegrated*.

- Such common trends leads to co-movements between y_{1t} and y_{2t} to some extent.

- Hence, the cointegration analysis helps to understand the *long-run relationships* among variables and their *short-run adjustments* from the equilibrium.

Cointegration analysis: A case of two time series

- Let y_{1t} and y_{2t} are both $I(1)$. Assume a vector $\beta = (\beta_1, \beta_2)'$ exists such that,

  $$\beta_1 y_{1t} + \beta_2 y_{2t} = u_t, \text{ where } u_t \sim I(0)$$

- Then, y_{1t} and y_{2t} are cointegrated. β is the cointegrating vector, which represents the long-run relationship.
- The long-run equilibrium is: $\beta_1 y_{1t} + \beta_2 y_{2t} = 0$
- At any particular time, the relationship may deviate from the long-run equilibrium. u_t is a stochastic variable representing the deviation from the equilibrium.

Cointegrated process: A K-dimensional case

- A K-dimensional process $Y_t = (Y_{1t}, ..., Y_{Kt})'$ are cointegrated of order (d, b), briefly, $Y_t \sim CI(d, b)$ if:
  - All components of Y_t are $I(d)$
  - Exists $\beta = (\beta_1, ..., \beta_K)'$ such that, $\beta' Y_t \sim I(d - b)$. β is the cointegrating vector, describing the long-run relationships.
  - Example: Assume all components of $Y_t \sim I(1)$ and $\beta' Y_t \sim I(0)$, then $Y_t \sim CI(1, 1)$. This is the only case we will study in this unit.
  - These definitions were introduced by Granger (1981) and Engle and Granger (1987)
- For K variables, there may be up to $(K-1)$ linearly independent cointegrating vectors (or long-run relationship).

Cointegrated process: Important points

- Cointegration relation defined above is a linear relationship. The cointegrating relationship may be non-linear but we will not cover in this unit.
- β is not unique. Need to put *normalization* in β.
- All variables should be integrated of the same order. However, there are some procedures allowing both $I(0)$ and $I(1)$ in one system (e.g., Johansen procedure).
- In most cases we deal with only $CI(1,1)$
- The Engle-Granger method estimates the only one (perhaps economically most sensible) cointegrating vector.
- The Johansen method can estimate up to $(K-1)$ cointegrating relationship, if they exist. (Discussed in the next few lectures)

Rank of a Matrix

To define the multiple cointegrating relationships, we first define the following:

- $Rank(\Pi)$ indicates the number of independent rows or columns of matrix Π.
- Suppose Π is a $(K \times K)$ matrix:
    - $det(\Pi) = 0 \Leftrightarrow Rank(\Pi) < K$.
    - $det(\Pi) \neq 0 \Leftrightarrow Rank(\Pi) = K$.
- If $1 \leq Rank(\Pi) = r < K$, there exist $(K \times r)$ matrices α and β so that $\Pi = \alpha\beta'$.
- Let the number of zero eigen values of Π be n, then:
    - $Rank(\Pi) = r = K - n$

How to find Rank of a Matrix

- **Method 1:** Find the number of independent rows or columns by *Gauss Elimination* technique. For example:

$$Rank \begin{bmatrix} 1 & 0 & 3 \\ 1 & 1 & 3 \\ 2 & 5 & 6 \end{bmatrix} = Rank \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 5 & 0 \end{bmatrix} = 2$$

- **Method 2:**
  - Find eigen values of Π: find λ satisfying $Det(\Pi - \lambda I_K) = 0$.
  - Let n is the number of $\lambda = 0$, then $Rank(\Pi) = r = K - n$.
  - Equivalently, r is the number of $\lambda \neq 0$.

Error Correction Model: A bivariate VAR(1) case

- VAR(1) model for $Y_t = (Y_{1t}, Y_{2t})'$:

$$Y_t = A_1 Y_{t-1} + \varepsilon_t$$

- Error correction model (ECM):

$$\Delta Y_t = \Pi Y_{t-1} + \varepsilon_t \;\; where \;\; \Pi = A_1 - I_2 = -(I_2 - A_1)$$

- If Y_t is non-stationary with unit root(s), then $det(\Pi) = 0$ or $Rank(\Pi) = r < 2$.

- If $r = 1$, then exists (2×1) matrices α and β so that $\Pi = \alpha\beta'$:

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \varepsilon_t$$

A bivariate VAR(1) case: illustration

- Find whether Y_t is non-stationary and write down its ECM:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

$$\Pi = A_1 - I_2 = \begin{bmatrix} -0.5 & 0.3 \\ 0 & 0 \end{bmatrix} \Rightarrow det(\Pi) = 0.$$

- Hence, Y_t is non-stationary with at least 1 unit root. The ECM:

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \end{bmatrix} = \begin{bmatrix} -0.5 & 0.3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- $Rank(\Pi) = 1 < 2$, then $\Pi = \alpha\beta'$. It's easy to obtain:

$$\begin{bmatrix} -0.5 & 0.3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -0.6 \end{bmatrix}$$

Error Correction Model: A bivariate VAR(1) case

- $Y_t = (Y_{1t}, Y_{2t})'$ has unit roots.
- $\Delta Y_t = (\Delta Y_{1t}, \Delta Y_{2t})'$ is stationary.
- ε_t is stationary.
- $\Pi Y_{t-1} = \alpha\beta' Y_{t-1}$ should be stationary.
- $\beta' Y_{t-1} = \beta_1 Y_{1t-1} + \beta_2 Y_{2t-1}$, the long-run relationship, should be stationary.
- $\beta = (\beta_1, \beta_2)'$ is the cointegrating vector describing the long-run relationship.

Error Correction Model: K-dimensional VAR(p) case

- VAR(p) specification for $Y_t = (Y_{1t}, ..., Y_{Kt})'$:

$$Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + \varepsilon_t$$

- ECM specification:

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + ... + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

$$\Pi = A_1 + ... + A_p - I_K = -(I_2 - A_1 - ... - A_p)$$

$$\Gamma_i = -(A_{i+1} + ... + A_p)$$

- If Y_t is non-stationary with unit root(s), then $det(\Pi) = 0$ or $Rank(\Pi) = r < K$.
- If $1 \leq r < K$, then exists $(K \times r)$ matrices α and β so that $\Pi = \alpha\beta'$.

Error Correction Model: K-dimensional VAR(p) case

- If Y_t has a unit root, then ΔY_t is stationary.
- In the ECM form, $\Pi Y_{t-1} = \alpha\beta' Y_{t-1}$, should be stationary.
- β (*cointegrating matrix*) represents r cointegrating vectors. r is called the *cointegrating rank*.
- Hence, if $1 \leq Rank(\Pi) < K$, then $Rank(\Pi)$ is the number of cointegrating vectors (long-run relationships).
- α represents the corresponding matrix of the speed of adjustment coefficient vectors (*adjustment* or *loading matrix*).

K-dimensional VAR(p) case: illustration

- Find whether Y_t is cointegrated, then make your interpretations. Y_t has a specification of bivariate VAR(2) model:

$$A_1 = \begin{bmatrix} 5/8 & 5/16 \\ 3/4 & 3/16 \end{bmatrix}; \text{ and } \; A_2 = \begin{bmatrix} -1/8 & -1/4 \\ -1/4 & 3/4 \end{bmatrix}$$

- $\Pi = A_1 + A_2 - I_2 = \begin{bmatrix} -1/2 & 1/16 \\ 1/2 & -1/16 \end{bmatrix} \Rightarrow det(\Pi) = 0.$

- Hence, Y_t is non-stationary with at least 1 unit root.

- $\Gamma_1 = -(A_2) = \begin{bmatrix} 1/8 & 1/4 \\ 1/4 & -3/4 \end{bmatrix}$. The Vector ECM (VECM) is:

$$\Delta Y_t = \begin{bmatrix} -1/2 & 1/16 \\ 1/2 & -1/16 \end{bmatrix} Y_{t-1} + \begin{bmatrix} 1/8 & 1/4 \\ 1/4 & -3/4 \end{bmatrix} \Delta Y_{t-1} + \varepsilon_t$$

K-dimensional VAR(p) case: illustration (con't)

- $Rank(\Pi) = Rank \begin{bmatrix} -1/2 & 1/16 \\ 1/2 & -1/16 \end{bmatrix} = Rank \begin{bmatrix} -1/2 & 0 \\ 1/2 & 0 \end{bmatrix} = 1$
- There is 1 cointegrating vector and $\Pi = \alpha\beta'$, where α and β are (2×1) matrices.
- $\Pi = \alpha\beta' = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix} \begin{bmatrix} 1 & -1/8 \end{bmatrix}$
- VECM now becomes:

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix} \begin{bmatrix} 1 & -1/8 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} 1/8 & 1/4 \\ 1/4 & -3/4 \end{bmatrix} \begin{bmatrix} \Delta Y_{1t-1} \\ \Delta Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

K-dimensional VAR(p) case: illustration (con't)

- Long run relationship:

$$z_t = \beta' Y_t = \begin{bmatrix} 1 & -1/8 \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = Y_{1t} - 1/8\, Y_{2t}$$

$$\Rightarrow Y_{1t} = 0.125\, Y_{2t} + z_t$$

- Positive relationship in the long run. If Y_{2t} increases by 1 unit, Y_{1t} is expected to increase by 0.125 unit in the long run.

- Short-run adjustment (expanded from VECM model with z_t):

$$\Delta Y_{1t} = -0.5 z_{t-1} + 0.125 \Delta Y_{1t-1} + 0.25 \Delta Y_{2t-1} + \varepsilon_{1t}$$

$$\Delta Y_{2t} = 0.5 z_{t-1} + 0.25 \Delta Y_{1t-1} - 0.75 \Delta Y_{2t-1} + \varepsilon_{2t}$$

K-dimensional VAR(p) case: illustration (con't)

- z_{t-1} is the past equilibrium error.
- Both ΔY_{1t} and ΔY_{2t} react to the past equilibrium error.
- So, both Y_{1t} and Y_{2t} are equilibrating factors.
- Change in Y_{1t} and Y_{2t} react to the past equilibrium error at the same rate but in opposite direction.

Cointegration and Rank

1. $1 \leq Rank(\Pi) = r < K$:
   - Y_t is cointegrated with r cointegrating vectors.
   - $det(\Pi) = 0$, and α and β exist.
   - $K - r$ unit roots, or $K - r$ common stochastic trends.
2. $r = K$:
   - Y_t is stationary in level. $det(\Pi) \neq 0$.
   - Hence, an informal condition of stationary VAR(p): $det(\Pi) \neq 0$.
3. $r = 0$:
   - Y_t is non-stationary and the VAR can be reformulated entirely in first differences.
   - $det(\Pi) = 0$, but α and β do not exist (Π is a null matrix).
   - K unit roots, K different stochastic trends.

Steps involved in Cointegration analysis

1. Given a VAR(p) model, check whether it is non-stationary:
   - Calculate Π, if Π is a null matrix, then Y_t is non-stationary but not cointegrated. Otherwise go to next steps.
   - Calculate $det(\Pi)$. If $det(\Pi) \neq 0$, Y_t is stationary. If $det(\Pi) = 0$, Y_t is non-stationary and we process to next step.
2. Write down the ECM (or sometimes called Vector ECM).
3. Find $Rank(\Pi) = r$, we will obtain $1 \leq r < K$. Y_t is cointegrated with r cointegrating vectors.
4. Decompose $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices.
   - Put normalisations for β
5. Interpret α and β.

Reference

- Granger (1981), *"Some properties of time series data and their use in econometric model specification"*, **Journal of Econometrics** 16, 121-130.
- Engle and Granger (1987) *"Co-integration and error correction: Representation, Estimation and testing"*, **Econometrica** 55, 251-276.

Introduction of Johansen Procedure

- Recall the VECM specification of K-dimensional VAR(p):

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + ... + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

$$\Pi = A_1 + ... + A_p - I_K;\ \Pi = \alpha\beta' \text{ if } 1 \leq Rank(\Pi) = r < K$$

$$\Gamma_i = -(A_{i+1} + ... + A_p)$$

- Johansen (1988) concentrates in estimating α and β using Full Information Maximum Likelihood (FIML) method.
- FIML is used to estimate simultaneous equations models.
  - Models required to include the full equation system: there are as many equations as endogenous variables.
  - Errors are assumed to be multivariate normally distributed.

Steps involved in **Johansen procedure**:

1. Determine the lag length of a VAR model using information criterion.
2. Identify deterministic terms in the model:
   - Deterministic terms: constant, linear trend or quadratic trend.
   - Trends need to be identified in *Data* and *Cointegration* parts to specify the model.
3. Testing for the cointegrating rank (value of r):
   - The Trace test.
   - The Maximum Eigen value test.
4. Estimate the model using FIML:
   - Impose restrictions on α and β (e.g., normalization, zeros ...).
   - Errors are assumed to be multivariate normally distributed.

Johansen procedure: some remarks

- Popular in applied works. Number of packages provide easy click-button steps for implementation (e.g., Eviews, PCGive, RATS...)
- Johansen estimator of the VECM tends to have least bias among available alternatives (e.g., Ordinary Least Square, non-linear least squares, principal components ...). (see for example Gonzalo, 1994).
- But also produce unlikely estimates, outliers.
- Cauchy type distribution may explain why outliers of the estimates are possible. (see Phillips, 1994).
- Johansen method is quite sensitive to the choice of lag length in VAR model.

Johansen procedure: some remarks (con't)

- AIC is typically used for lag length selection
- If AIC suggests a long lag, Johansen recommends adding one more variable into the models as long lags may be trying to capture the effect of an omitted variable.
- If Johansen method leads to only one cointegrating vector, a comparison with LS (as done for the bivariate case using Engle-Granger procedure) estimates is advised due to robust estimation issues.
- If there is more than one cointegrating vector then there is no better alternative method to Johansen procedure at the moment.

A Summary of Technical Details

The procedure starts with VECM specification:

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + ... + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \Sigma_\varepsilon)$$

The log-likelihood function (LLF) for ε_t, $t = 1, ..., T$ is:

$$l(\Sigma_\varepsilon | \varepsilon_1, ..., \varepsilon_T) = -\frac{TK}{2} ln(2\pi) - \frac{T}{2} ln|\Sigma_\varepsilon| - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t' \Sigma_\varepsilon^{-1} \varepsilon_t$$

Transform $\varepsilon_t = (\Delta Y_t - \Pi Y_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i})$, we get the LLF:

$$l(...) = -\frac{TK}{2} ln(2\pi) - \frac{T}{2} ln|\Sigma_\varepsilon| - \frac{1}{2} \sum_{t=1}^{T} \Big[(\Delta Y_t - \Pi Y_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i})' \Sigma_\varepsilon^{-1} (\Delta Y_t - \Pi Y_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i})\Big]$$

A Summary of Technical Details (con't)

- Parameter values that maximize the LLF are the FIML estimates.
- The Johansen procedure involves what is called the concentrated likelihood function (CLF):
  - Obtain an expression that retains Π (CLF for Π).
  - Then write $\Pi = \alpha\beta'$ and obtain the CLF for β.
- Hence, maximizing the LLF leads to solving for the eigenvalues of Π. In other words, we need to find the cointegrating rank (r) before the maximization.
- Johansen provides the trace and the maximum eigenvalue tests for finding the cointegrating rank r.

Three Testing Principles related to Log-likelihood

$H_0 : \theta = \theta_0; \quad H_1 : \theta \neq \theta_0.$

1. LR (likelihood ratio) test:
   - LR measures $LL(\theta_1) - LL(\theta_0)$. LL denotes log-likelihood.
2. LM (Lagrange Multiplier) test:
   - LM measures the difference in the slope of LL function at θ_1 and θ_0.
3. Wald test
   - Measures the difference between θ_1 and θ_0.
   - t-test and F-test belong to Wald tests.

LR, LM and Wald tests

- All 3 test statistics are asymptotically distributed as χ^2_J. J is the number of restrictions under H_0.
- They are widely used in econometrics.
- They are valid tests when sample size is large enough to justify that the asymptotic theories hold.

LR statistics

- $LR = -2\left[LL(\theta_0) - LL(\theta_1)\right]$
- $LL(\theta_0)$: The log-likelihood from the estimated model under the null hypothesis or restricted model.
- $LL(\theta_1)$: The log-likelihood from the estimated model under the alternative hypothesis or unrestricted model.

LM statistics

- The LM statistic is computed as: $LM = TR_a^2$
- R_a^2: is the value of R^2 from the "auxiliary regression" under the alternative hypothesis.
- Many statistical tests for autocorrelation and heteroskedasticity are based on the LM test.

Likelihood function of the Cointegrated VAR

The log-likelihood function for the cointegrated VAR with $Rank(\Pi) = r$ is a function of:

$$-0.5T\sum_{i=1}^{r} ln(1-\hat{\lambda}_i)$$

where $\hat{\lambda}_i$ are the estimated eigenvalues of Π, and T is the sample size.

The Trace test: a LR test

$$H_0 : r \leq r_0; \;\; H_1 : r_0 < r \leq K; \quad r_0 = 0, 1, ..., K - 1.$$

$$\begin{aligned} LR_{statistic} &= -2\left[LL(H_0) - LL(H_1)\right] \\ &= -2\left[-0.5T\sum_{i=1}^{r_0} ln(1-\hat{\lambda}_i) + 0.5T\sum_{i=1}^{K} ln(1-\hat{\lambda}_i)\right] \end{aligned}$$

Therefore:

$$\lambda_{trace}(r_0) = -T\sum_{i=r_0+1}^{K} ln(1-\hat{\lambda}_i)$$

The Maximum Eigenvalues Test: another LR test

$$H_0 : r \leq r_0; \quad H_1 : r = r_0 + 1; \quad r_0 = 0, 1, ..., K - 1.$$

$$\lambda_{max}(r_0) = -T ln(1 - \hat{\lambda}_{r_0+1})$$

- The trace and maximum eigenvalue test statistics do not follow the usual χ^2 distribution. Instead, they follow the non-standard distribution and its critical values tabulated by Johansen.
- H_0 is rejected at a level of significance if statistic is greater than the critical value.
- Eviews provides critical values and p-values.

Sequential Testing: An example for $K=2$

Stage 1:

The Trace Test	The Maximum Eigenvalues Test
$H_0 : r \leq 0; \quad H_1 : 0 < r \leq 2$	$H_0 : r \leq 0; \quad H_1 : r = 1$

Stage 2:

The Trace Test	The Maximum Eigenvalues Test
$H_0 : r \leq 1; \quad H_1 : 1 < r \leq 2$	$H_0 : r \leq 1; \quad H_1 : r = 2$

- Accept H_0 in both stages: $r = 0$ (no cointegration).
- Reject H_0 in both stages: $r = 2$ (no cointegration).
- Reject H_0 in stage 1 and accept H_0 in stage 2: $r = 1$ (cointegration).

I(0) variables

- Johansen procedure can accommodate $I(0)$ variables in the cointegrating relations.
- Example: suppose we have 3 variables, $I(1), I(1)$ and $I(0)$.
  - May extend r to 2 to accommodate the $I(0)$ variable.
  - i.e., 2 cointegrating vectors may take the form $(1, 1, 0)$ and $(0, 0, 1)$.
  - So, two $I(1)$ variables form one cointegrating relationship and the $I(0)$ variable is cointegrated with itself.

I(0) variables

- Allowing $I(0)$ variable in a cointegrating relationship is important because:
  - Selection of variables does not depend on their order of integration.
  - They are rather chosen because of their importance from a theoretical point of view.
  - E.g., In a money demand equation, $m = ln(real\ money\ balances)$, $y = (real\ income)$, $i = \ interest\ rate$. m and y are likely to be $I(1)$ but i is likely to be $I(0)$.

Deterministic terms: the case of Cointegrated VAR(1)

$$\Delta Y_t = \nu_1 + \delta_1 t + \alpha(\beta' Y_{t-1} - \nu_2 - \delta_2 t) + \varepsilon_t$$

- Linear trend in data and constant in cointegration:

  $$\Delta Y_t = \nu_1 + \alpha(\beta' Y_{t-1} - \nu_2) + \varepsilon_t$$

  ν_2: long-run equilibrium level of Y's

- Some or all time series show trend:

  $$\Delta Y_t = \nu_1 + \alpha(\beta' Y_{t-1} - \nu_2 - \delta_2 t) + \varepsilon_t$$

  $\nu_2 + \delta_2 t$: long-run equilibrium level of Y's

- Allowing for quadratic trend in data:

  $$\Delta Y_t = \nu_1 + \delta_1 t + \alpha(\beta' Y_{t-1} - \nu_2 - \delta_2 t) + \varepsilon_t$$

Deterministic terms: Alternative models

	Data	Cointegration
1	No linear trend, no quadratic trend $\nu_1 = \delta_1 = 0$	No constant, no linear trend $\nu_2 = \delta_2 = 0$
2	No linear trend, no quadratic trend $\nu_1 = \delta_1 = 0$	Constant, no linear trend $\nu_2 \neq 0,\ \delta_2 = 0$
3	Linear trend, no quadratic trend $\nu_1 \neq 0,\ \delta_1 = 0$	Constant, no linear trend $\nu_2 \neq 0,\ \delta_2 = 0$
4	Linear trend, no quadratic trend $\nu_1 \neq 0,\ \delta_1 = 0$	Constant, linear trend $\nu_2 \neq 0,\ \delta_2 \neq 0$
5	Linear trend, quadratic trend $\nu_1 \neq 0,\ \delta_1 \neq 0$	Constant, linear trend $\nu_2 \neq 0,\ \delta_2 \neq 0$

Deterministic terms: Choice of Models

- Model 1 is too restrictive.
- If none of time series show linear trend, Model 2 should be used.
- If some or all time series show linear trend, use either Model 3 or Model 4.
- If quadratic trend in the data, use Model 5.

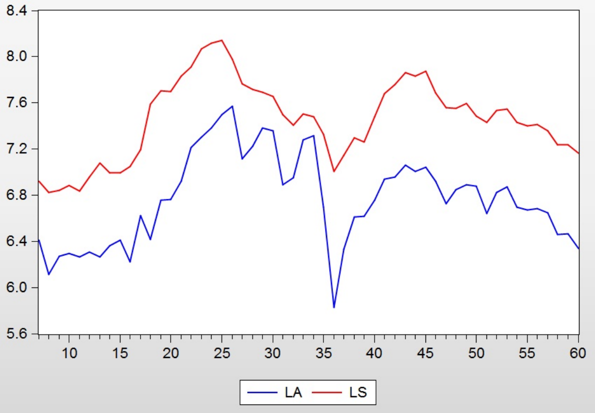
Case Study: Lydia Pinkham data

- A classical data set in the Advertising - Sales relationship literature.
- Annual data from 1907 to 1960.
- Investigate the inter-relationship between Advertising and Sales.
- Data in natural logs are estimated:
    - LA: log of Advertising.
    - LS: log of Sales.

Case Study: Why this data set?

- Company produced one product (a compound for women's health)
- No competitors in the market.
- Advertising was the only marketing instrument.
- Advertising expenditure is set at a proportion of sales.
- Very "clean" data set for the advertising and sales relationship.

Case study: Series Plots



Case study: Cointegrating rank test

Trend assumption: No deterministic trend (restricted constant)
Series: LA LS
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.284574	23.83590	20.26184	0.0154
At most 1	0.116184	6.422313	9.164546	0.1605

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.284574	17.41359	15.89210	0.0287
At most 1	0.116184	6.422313	9.164546	0.1605

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Model 2 is fitted (no trend in data)

Trace and Max Eigenvalue test statistics indicate that they are cointegrated

Case study: VECM estimation

Vector Error Correction Estimates
Date: 04/12/07 Time: 20:27
Sample (adjusted): 1909 1960
Included observations: 52 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	
LA(-1)	1.000000	
LS(-1)	-0.796038 (0.11364) [-7.00477]	
C	-0.844051 (0.84956) [-0.99351]	
Error Correction:	**D(LA)**	**D(LS)**
CointEq1	-0.695664 (0.16551) [-4.20307]	-0.292733 (0.08872) [-3.29957]
D(LA(-1))	0.180466 (0.17702) [1.01946]	0.185968 (0.09489) [1.95990]
D(LS(-1))	0.031054 (0.32266) [0.09624]	0.078159 (0.17295) [0.45190]
R-squared	0.317918	0.287379

Long run relationship

Estimates of β vector

Error correction models

Case Study: Implications

- 1% increase in Sales will increase 0.8% of Advertising expenditure in the long-run.
- In the short-run, both are equilibrating factors.
- Change in Advertising reacts to the past equilibrium error more sensitively than the change in Sales.

Reference

- Johansen (1988), *"Statistical analysis of cointegration vectors"*, **Journal of Economic Dynamics and Control** 12, 231-254.
- Phillips (1994) *"Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models"*, **Econometrica** 62, 73-93.
- Gonzalo (1994) *"Five alternative methods of estimating long-run equilibrium relationships"*, **Journal of Econometrics** 60, 203-233.