LECTURE FOUR

JEN-WEN LIN, PhD

July 22, 2014

VECTOR AUTOREGRESSIVE

- Vector autoregressive models are one of the most popular multivariate time series models. It can be seen as a generalization of univariate AR models.
- Their popularity for analyzing the dynamics of economic systems is due to Sims'(1980).

In practice, vector AR model
is easier than
vector ARMA model
due to too many parameters.

VECTOR AUTOREGRESSIVE PROCESS OF ORDER ONE

k-dimensional vector AR(1) model

A multivariate time series $\boldsymbol{r}_t$ is a vector autoregressive process of order one, or VAR(1) for short, if the model is defined as

$$\boldsymbol{r}_t = \Phi_0 + \Phi_1 \boldsymbol{r}_{t-1} + \boldsymbol{a}_t \qquad (1)$$

where Φ_0 is a k-dimensional vector, and Φ_1 is a $k \times k$ matrix, $\boldsymbol{a}_t$ is a multivariate white noise series, usually IID multivariate Gaussian distribution.

BIVARIATE VAR(1) MODEL

* Let $\boldsymbol{r}_t = [r_{1,t}, r_{2,t}]^T$ and $a_t = [a_{1,t}, a_{2,t}]^T$

$$r_{1,t} = \phi_{10} + \phi_{11} r_{1,t-1} + \phi_{12} r_{2,t-1} + a_{1,t}$$

$$r_{2,t} = \phi_{20} + \phi_{21} r_{1,t-1} + \phi_{22} r_{2,t-1} + a_{2,t}$$

* In the matrix form,

$$\underbrace{\begin{bmatrix} r_{1,t} \\ r_{2,t} \end{bmatrix}}_{\mathbf{r_t}} = \underbrace{\begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix}}_{\Phi_0} + \underbrace{\begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}}_{\Phi_1} \underbrace{\begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix}}_{\mathbf{r_{t-1}}} + \underbrace{\begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}}_{\mathbf{a_t}}$$

k-dim vector

k×k matrix

STATIONARY VAR(1) MODEL

Assume that $\Phi_0 = \mathbf{0}$. By repeated substitution, we can express eqn. (2) as

$$\boldsymbol{r}_t = \boldsymbol{a}_t + \Phi_1 \boldsymbol{a}_{t-1} + \cdots + \Phi_1^j \boldsymbol{a}_{t-j} + \cdots = \sum_{j=0}^{\infty} \Phi_1^j \boldsymbol{a}_{t-j},$$

where the impact of past shocks, say $\boldsymbol{a}_{t-j}$, on $\boldsymbol{r}_t$ is given by Φ_1^j.

For eqn. (1) to be stationary and such dependence to be meaningful, Φ_1^j must converge to zero as $j \to \infty$. The convergence of Φ_1^j means that all k eigenvalues of Φ_1 must be less than one in modulus.

$X_t = \sum_{j=0}^{\infty} \Phi^j a_t$

$V(X_t) = \sum_{j=0}^{\infty} \Phi_1^j \Sigma \Phi_1^j$

$\begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right)$

$E(a_{it} \cdot a_{is}) = 0$ if $t \neq s$, $i = 1, 2$

$E(a_{1t} \cdot a_{2s}) = 0$ if $t \neq s$

REMARKS

* In fact, the requirement that all eigenvalues of Φ_1 are less than one in modulus is the necessary and sufficient condition for weak stationarity of $\boldsymbol{r}_t$ provided the covariance matrix of $\boldsymbol{a}_t$ exists
* Cross-covariance matrix of $VAR(1)$ processes:

$$\operatorname{cov}(\mathbf{r}_t, \mathbf{r}_{t-l}) = \boldsymbol{\Gamma}_l = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{l-1}, \forall\, l > 0$$

$$\operatorname{cov}(\mathbf{r}_t, \mathbf{r}_t) = \boldsymbol{\Gamma}_0 = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_1^j \boldsymbol{\Sigma} (\boldsymbol{\Phi}_1^j)^T$$

where $\Phi_1^0 = \boldsymbol{I}_k$ and $\boldsymbol{\Gamma}_l$ is the lag-j cross covariance matrix of $\boldsymbol{r}_t$.

Generalization: VAR(p) model

A VAR(p) process may be defined as

$$\boldsymbol{r}_t = \mu + \Phi_1 \boldsymbol{r}_{t-1} + \cdots + \Phi_p \boldsymbol{r}_{t-p} + \boldsymbol{a}_t, \qquad (2)$$

- In short notations, we can write eqn. (2) as $\Phi(B)\boldsymbol{r}_t = \mu + \boldsymbol{a}_t$, where $\boldsymbol{a}_t$ is a multivariate white noise series and $\Phi(B) = \boldsymbol{I}_k - \Phi_1 B - \cdots - \Phi_p B^p$
- Question: What are the the stationarity condition for the above VAR(p) model?

STATIONARY $VAR(p)$ MODEL

Theoretically, a stationary $VAR(p)$ model can be checked by evaluating the reverse characteristic polynomial

$$\det\left(\boldsymbol{I}_k - \Phi_1 B - \cdots - \Phi_p B^p\right) \cancel{\neq} 0 \; \forall \, |z| \leq 1.$$

- The calculation of using the above definition is tedious. A simpler approach is to express a $VAR(p)$ model into its companion form and evaluating the corresponding eigenvalues of the companion matrix.
- The companion form and companion matrix are to be defined in the next slide.

$$|I - \Phi Z| = 0$$

$$\left|\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}\Phi_{11} & \Phi_{12}\\ \Phi_{21} & \Phi_{22}\end{bmatrix} Z\right| = 0 \iff 1 - \underline{tr}(\Phi)Z + |\Phi| Z^2 = 0$$

① Sum of the diagonal elements ② Z outside unit circle

$$\Phi = \begin{pmatrix}0.196, & -0.254\\ 0.715, & -0.536\end{pmatrix}$$

$tr(\Phi) = -0.34$ $|\Phi| = 0.07655$

e.g. $\underset{k\times 1}{X_t} = \underset{k\times k}{\Phi_1} \underset{k\times 1}{X_{t-1}} + \cdots + \Phi_p X_{t-p} + a_t$ $k\times 1$

$$X_t = \underset{k\times 1}{\begin{bmatrix} X_{t-p+1} \\ X_{t-p+2} \\ \vdots \\ X_t \end{bmatrix}} \qquad \Phi^* = \begin{bmatrix} 0 & I & & O \\ 0 & & \ddots & \\ \vdots & O & & I \\ \underset{k\times k}{\Phi_p} & \underset{k\times k}{\Phi_{p-1}} & \cdots & \Phi_1 \end{bmatrix} \begin{bmatrix} X_{t-p} \\ X_{t-p+1} \\ \vdots \\ X_{t-1} \end{bmatrix} = X_{t-1}$$

multiple these 2 $\longrightarrow$ get $b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ a_t \end{bmatrix}$

STATIONARY VAR(p) MODEL AND COMPANION MATRIX

Step 1: Transforming the k variate VAR(p) model into a $VAR(1)$ model of $k \times p$ variables , where let

$$\boldsymbol{b}_t = [\mathbf{0}_{k\times 1}, \mathbf{0}_{k\times 1}, \dots, \boldsymbol{a}_t^T]^T$$

and

$$\mathbf{X}_t = \left[\boldsymbol{r}_{t-p+1}^T, \boldsymbol{r}_{t-p+2}^T, \dots, \boldsymbol{r}_t^T\right]^T$$

be two $k \times p$ dimensional time series.

The mean of $\boldsymbol{b}_t$ is zero and the corresponding covariance matrix is a $kp \times kp$ matrix with zero everywhere except for the lower right corner, which is Σ.

STATIONARY VAR(p) MODEL AND COMPANION MATRIX

- **Step 2:** Express the VAR(p) model in eqn. (2) as

$$X_t = \Phi^* X_{t-1} + b_t, \qquad (3)$$

where

$$\boldsymbol{\Phi}^*_{kp\times kp} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_k & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_k \\ \boldsymbol{\Phi}_p & \boldsymbol{\Phi}_{p-1} & \cdots & \boldsymbol{\Phi}_1 \end{bmatrix}.$$

  - *Note:* In the literature, $\boldsymbol{\Phi}^*$ is called the companion matrix of $\Phi(B)$.

- **Step 3:** Expressing a $VAR(p)$ model as the form of eqn. (3), we can then apply the stationarity condition that we learnt from the $VAR(1)$ model. That is, that $k \times p$ eigenvalues of $\boldsymbol{\Phi}^*$ must be less than one in modulus.

MODEL BUILDING PROCEDURE (TSAY, 2002)

Suppose that you would like to use a $VAR(j)$ model

$$\boldsymbol{r}_t = \Phi_0 + \Phi_1 \boldsymbol{r}_{t-1} + \cdots + \Phi_j \boldsymbol{r}_{t-j} + \boldsymbol{a}_t \qquad (4),$$

- where j is unknown in practice so the order selection is needed.

We may choose the order of a VAR model by testing $H_0\colon \Phi_j = \mathbf{0}$ against $H_a\colon \Phi_j \neq \mathbf{0}$

- That is, we are testing a $VAR(j-1)$ model versus $VAR(j)$ model.

MODEL BUILDING PROCEDURE (TSAY, 2002)

In final exam

1. Estimate the model in eqn. (4) using OLS.
2. The residuals from a VAR(j) model are given by

$$\hat{\mathbf{a}}_t^{(j)} = r_t - \hat{\Phi}_1^{(j)}\mathbf{r}_{t-1} - \ldots - \hat{\Phi}_j^{(j)}\mathbf{r}_{t-j}$$

3. Calculate the residual covariance matrix

$$\hat{\Sigma}^{(j)} = \frac{1}{T-2j-1}\sum_{t=j+1}^{T}\hat{\mathbf{a}}_t(\hat{\mathbf{a}}_t)^T, j \geq 0$$

4. Calculate the test statistic

$$M(j) = -(T-k-j-\frac{3}{2})\log(\frac{|\hat{\Sigma}^{(j)}|}{|\hat{\Sigma}^{(j-1)}|}) \rightarrow \chi^2_{k^2}$$

$$\underbrace{\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix}}_{y} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \underbrace{\begin{bmatrix} r_{1t-1} \\ r_{2t-1} \end{bmatrix}}_{x} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

MULTIVARIATE PORTMANTEAU TESTS

* Test statistics:

$$Q_{BP} = T\sum_{j=1}^{m} \mathrm{tr}(\hat{C}_j^T \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}) \sim \chi^2_{k^2 m - n^*}$$

$$Q_{LB} = T^2\sum_{j=1}^{m} \frac{1}{T\text{-}j} \mathrm{tr}(\hat{C}_j^T \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}) \sim \chi^2_{k^2 m - n^*}$$

where n^* is the number of coefficients excluding deterministic terms of a $VAR(p)$ model and

$$\hat{C}_i = \frac{1}{T}\sum_{t=i+1}^{T} \hat{\mathbf{a}}_t \hat{\mathbf{a}}_{t-i}^T$$

MODEL SELECTION

and of course, BIC

* The AIC of a $VAR(j)$ model under the normality assumption is defined as

$$AIC(j) = \log(|\widetilde{\boldsymbol{\Sigma}}^{(j)}|) + \frac{2 \cdot j \cdot k^2}{T}$$

where

$$\widetilde{\boldsymbol{\Sigma}}^{(j)} = \frac{1}{T} \sum_{t=j+1}^{T} \hat{\mathbf{a}}_t (\hat{\mathbf{a}}_t)^T.$$

* We select the VAR model of order p such that

$$AIC(p) = \min_{1 \le i \le p_0} AIC(i),$$

where p_0 is a pre-specified integer.

* There are other criterion for model selection, such as Hannan and Quinn (1979), and Schuraz (1978).

In final

GRANGER CAUSALITY

- Causality tests are useful to infer whether a variable helps predict another one.
- An operational definition of causality between two time series can be defined in terms of predictability (Granger, 1969).

CAUSALITY

- A variable X is said to cause another variable Y, with respect to a given universe or information set that includes X and Y,
- Ideally, causality may be defined through the concept of the conditional distribution.
  - $\boldsymbol{y_{2t}}$ does not Granger cause $\boldsymbol{y_{1t}}$ if the distribution of y_{1t}, conditional on past values of both y_{1t} and y_{2t}, is the same as the distribution of y1t conditional on its own past values.
- Alternatively, if present Y can be better predicted by using past values of X than by not doing so, all other relevant information (including the past values of Y) is the universe being used in either case.

GRANGR CAUSALITY

- In practice, it would be very difficult to test whether the entire distribution y_{1t} depends on past values of y_{2t}.
- Therefore, we consider an alternative by asking whether the conditional mean of y_{1t} depends on past values of y_{2t}. If this is the case, we can test Granger causality by imposing restrictions on a VAR model.

GRANGR CAUSALITY AND VAR MODELS

$$\underbrace{\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix}}_{y} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \underbrace{\begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix}}_{x} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

* Consider a $VAR(p)$ model as follows:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \phi_{j,11} & \phi_{j,12} \\ \phi_{j,21} & \phi_{j,22} \end{bmatrix} \begin{bmatrix} y_{1,t-j} \\ y_{2,t-j} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

* If y_{2t} does not Granger cause y_{1t}, then all of the $\phi_{j,12}$'s must be zero. Note that $\phi_{j,12}$'s only appear in the equation for y_{1t}. $y_{2t} \not\Rightarrow y_{1t}$. $\phi_{j,12} = 0$

* Similarly, if y_{1t} does not Granger cause y_{2t}, then all of the $\phi_{j,21}$'s must be zero.

$y_{1t} \not\Rightarrow y_{2t}$, $\phi_{j,21} = 0$

TEST PROCEDURE

test questions.

* Obtain ML (or OLS) estimates of the following equations.

$$y_{1t} = \alpha_1 + \sum_{j=1}^{p} \phi_{j,11} y_{1,t-j} + e_{1t}, \qquad (5)$$

$$y_{1t} = \alpha_1 + \sum_{j=1}^{p} \phi_{j,11} y_{1,t-j} + \sum_{j=1}^{p} \phi_{j,12} y_{2,t-j} + \varepsilon_{1t}, \quad (6)$$

* Calculate the values of the log likelihood functions in eqn. (5) and (6). And, the LR statistic is given by

$$n(\log|\tilde{\Sigma}| - \log|\hat{\Sigma}|) \sim \chi_p^2$$

where $\tilde{\Sigma}$ and $\hat{\Sigma}$ denote the residual covariance matrix estimated from eqn. (5) and (6), respectively.

ALTERNATIVE TEST

- Pierce and Haugh (1977) expanded up the work of Granger (1969) and gave a comprehensive survey regarding research on causality in temporal systems.
- For simplicity, in what follows, we consider the case of two time series $\{X_t\}$ and $\{Y_t\}$.
- Let $\{X_t\}$ and $\{Y_t\}$ be causal and invertible univariate $ARMA$ processes and be given by

$$\phi_X(B)(X_t - \mu_X) = \theta_X(B)u_t, \quad u_t \sim \mathrm{WN}(0, \sigma_u^2)$$

$$\phi_Y(B)(Y_t - \mu_Y) = \theta_Y(B)v_t, \quad v_t \sim \mathrm{WN}(0, \sigma_v^2)$$

PIERCE AND HAUGH (1977)

- The cross-correlation function at lag k between u_t and v_t series is given by

$$\rho_{uv}(k) = \frac{E(u_t, v_{t+k})}{\sqrt{E(u_t^2)E(v_t^2)}}$$

- Pierce and Haugh (1977) explained that there are many possible types of causal interpretation between $\{X_t\}$ and $\{Y_t\}$ which can be characterized by the properties of $\rho_{uv}(k)$.

CAUSAL RELATIONSHIPS BETWEEN TWO VARIABLES

very useful

RELATIONSHIPS	RESTRICTIONS ON $\rho_{uv}(k)$
X causes Y	$\rho_{uv}(k) \neq 0$ for *some* $k > 0$
Y causes X	$\rho_{uv}(k) \neq 0$ for *some* $k < 0$
Instantaneous Causality	$\rho_{uv}(0) \neq 0$
Feedback	$\rho_{uv}(k) \neq 0$ for *some* $k > 0$ and for some $k < 0$
X causes Y but not instantaneously	$\rho_{uv}(k) \neq 0$ for *some* $k > 0$ and $\rho_{uv}(0) = 0$
Y does not cause X	$\rho_{uv}(k) = 0$ for *all* $k < 0$
Y does not cause X at all	$\rho_{uv}(k) = 0$ for *all* $k \leq 0$
Unidirectional causality from X to Y	$\rho_{uv}(k) \neq 0$ for some $k > 0$ and $\rho_{uv}(k) = 0$ for either (a) all $k < 0$ or (b) all $k \leq 0$
X and Y are only related instantaneously	$\rho_{uv}(0) \neq 0$ and $\rho_{uv}(k) = 0$ for all $k \neq 0$
X and Y are independent	$\rho_{uv}(k) = 0$ for *all* k

Portmanteau tests for causality:

- H_0: X does not cause Y
- $Q_L = n^2 \sum_{k=0}^{L} (n-k)^{-1} r_{uv}^2(k) \sim \chi_{L+1}^2$

SPURIOUS REGRESSION

- Consider a simple regression model with the following assumptions:

$$y_t = \alpha + \beta X_t + \varepsilon_t, \qquad (1)$$

want X_t, Y_t to be "random walk"

  - Innovations are IID random variables
  - X_t and Y_t are independent $I(1)$ processes

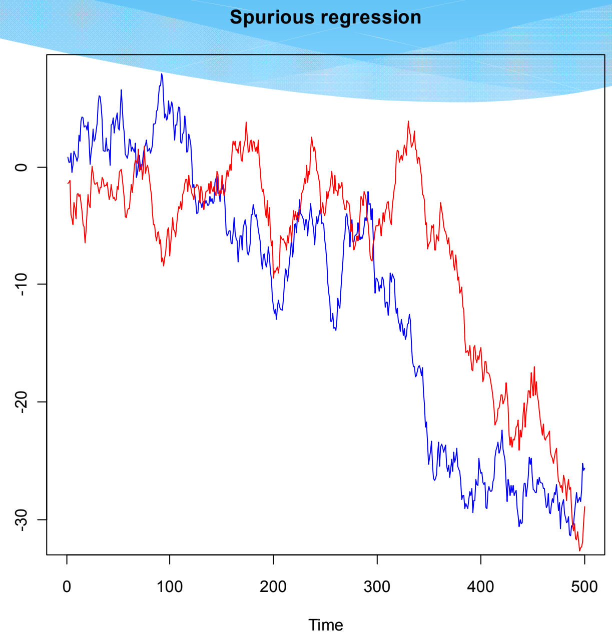
- What statistical inference can we draw from cross-section regression analysis?
  - $\hat{\beta} \to 0$ in probability
  - $t_\beta = \frac{\hat{\beta}}{se(\hat{\beta})} \sim$ student t distribution
  - $R^2 \to 0$ in probability

SPURIOUS STATISTICAL INFERENCE

* The absolute value of t_β tends to become larger and larger as the series length T increases;
* We will eventually rejects the null hypothesis that $\beta = 0$ with probability one as $T \to \infty$
* Empirical observations shows that a spurious regression is usually characterized by a high R-square.
  * Rule of thumb: a model is suspicious if the R-square is greater than the Durbin-Watson statistics.
* When a regression model as in eqn. (1) appears to find relationship that do not really exist, it is called spurious regression.
* Granger (2001) addressed that spurious regression can occur even when all variables are stationary. The risk can be far from negligible with stationary series that exhibit substantial series correlation.

SIMULATION EXAMPLE

- library(lmtest)
- set.seed(1112)
- e1 <- rnorm(500)
- e2 <- rnorm(500)
- y1 <-cumsum(e1)
- y2 <-cumsum(e2)
- sr.reg<- lm(y1 ~ y2)
- sr.dw <- dwtest(sr.reg1)$statistic



- R-square is 0.58 and the Durbin-Watson statistic 0.0507 is close to zero, as expected.

REGRESSION ON TWO I(1) PROCESSES

* It seems the statistical inference for spurious regression is incorrect. Can we solve the spurious regression problem under I(1) series ??
* If we are able to do so, then we can regress an integrated series on another
* How to detect the I(1) processes?
  * Using the unit root test
* How to convert the spurious regression under I(1) series into a valid regression?
  * Differencing
* Can differencing solve the problem?
  * Differencing solves the statistical problems
  * but not the economic interpretation of the regression.
  * After differencing, we are loosing information.

REGRESSION ON TWO I(1) PROCESSES

- Cons:
  - I(1) series tend to diverge as $n \to \infty$ because their unconditional variances are proportional to n.
  - Thus, it might seem that two or more such variables would never be expected to obey any sort of long-run relationship.
- Pros:
  - Multivariate time series that are all individually I(1) may in a certain sense diverge together.
- In general, it is possible for some linear combinations of a set of I(1) series to be I(0). If that is the case, the set of series is said to be *cointegrated*.

BALANCE OF REGRESSION

* Econometricians have been aware of the importance of balance since at least Box and Jenkins (1976).
* Consider the issues of balance in mean.
  * A constant-mean regressor can't fully explain a variable with a linear trend
  * A variable with a linear trend cannot explain a series with an exponential trend.
* An I(d-1) process cannot adequately explain the behavior of an I(d) process.
  * For example, an I(0) process (which has a constant mean and variance) cannot explain all of the behavior of an I(1) process whose mean and variance move upward overtime.

SOME USEFUL RESULTS

<u>Linear combinations of I(0) and I(1) processes</u>

1. $X_t \rightarrow I(0) \Rightarrow a + bX_t \rightarrow I(0)$

   $X_t \rightarrow I(1) \Rightarrow a + bX_t \rightarrow I(1)$

2. $X_t, Y_t \rightarrow I(0) \Rightarrow aX_t + bY_t \rightarrow I(0)$

3. $X_t \rightarrow I(0), Y_t \rightarrow I(1) \Rightarrow aX_t + bY_t \rightarrow I(1)$

4. $X_t, Y_t \rightarrow I(1) \Rightarrow aX_t + bY_t \rightarrow I(1),\text{ in general}$

COINTEGRATION

* In 1981, Granger introduced the concept of cointegration and the general case was publicized by Engle and Granger (1987). The general idea behind cointegration is to find a linear combination between two I(d)-variables yields a variable with a lower order of integration.

* **Definition:** If X_t and Y_t are $I(1)$ processes, and there exists a linear combination, say $Z_t = m + aX_t + bY_t$ such that Z_t is $I(0)$, then $X_t, \mathrm{Y_t}$ are said to be cointegrated.

COMMON TREND

* The observation of Stock and Watson (1988) that cointegrated variables share common stochastic trends provides a very useful way to understand cointegration relationships.
* Consider the following example—X_t and Y_t are I(1) processes and

$$X_t \equiv \alpha \cdot W_t + \tilde{X}_t$$

$$Y_t \equiv W_t + \tilde{Y}_t$$

X_t and Y_t share the same nonstationary sources W_t--an ARIMA(p,1,q) process, or I(1) process

Stationary ARMA(p,q) process, or I(0) process

•X_t and Y_t are both I(1) process since I(1)+I(0)=I(1) by rule (3)

COMMON TREND

X_t and Y_t have a common *I(1)* factor w_t,
and define the linear combination Z_t as follows:

$$Z_t \equiv X_t - \alpha \cdot Y_t = \cancel{\alpha \cdot W_t} + \boxed{\widetilde{X}_t} - \cancel{\alpha \cdot W_t} - \boxed{\alpha \cdot \widetilde{Y}_t}$$

$$Z_t \equiv \widetilde{X}_t - \alpha \cdot \widetilde{Y}_t \rightarrow I(0) \quad \text{(rule 2)}$$

Result 1.

If two I(1) series have a common I(1) factor and idiosincratic I(0) components, then they are cointegrated.

It can be proved that Result 1 is an IF and ONLY IF result.

COINTEGRATION WITH MORE TWO SERIES

Example 1. $k = 3, \boxed{h = 2}$

three I(1) processes

$$\left.\begin{aligned} Y_t &\equiv W_t + u_t \\ X_t &\equiv W_t + v_t \\ Z_t &\equiv W_t + s_t \end{aligned}\right\} W_t \to I(1) \quad u_t, v_t, s_t \to I(0)$$

1 common stochastic trend $\to W_t$

2 cointegrating vectors: $\boxed{(1 \quad -1 \quad 0)' \quad (0 \quad 1 \quad -1)'}$

Two cointegration relationships

Example 2. $k = 3, h = 1$

$$\left.\begin{aligned} Y_t &\equiv W_t + u_t \\ X_t &\equiv W_t + R_t + v_t \\ Z_t &\equiv R_t + s_t \end{aligned}\right\} W_t, R_t \to I(1) \quad u_t, v_t, s_t \to I(0)$$

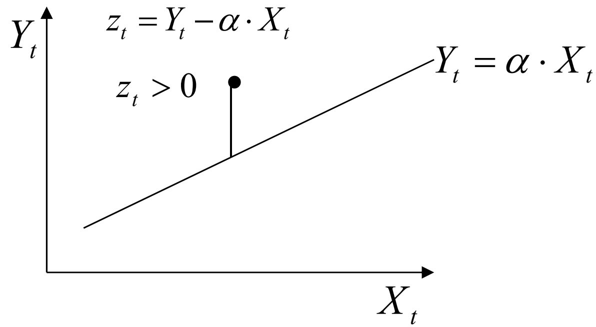
2 common stochastic trends $\to W_t, R_t$

1 cointegrating vector: $(1 \quad -1 \quad 1)'$

ERROR CORRECTION MODEL (ECM)

* A principal feature of cointegrated variables is that their time paths are influenced by the extend of any deviation from long-run equilibrium.

•In the long run, or when the system is in an equilibrium state, we have $z_t = 0$. However, this is usually not the case but there is a belief that the system will return to the equilibrium state eventually.



•Short-run dynamics are modeled as the Error Correction Mechanism (ECM) that guide the economy towards the Long-run equilibrium

GRANGER REPRESENTATION THEOREM

Consider a bivariate Vector Autoregressive process, where X_t and Y_t are both I(1) processes and co-integrated.

$$\boxed{\begin{aligned} \Delta X_t &= \alpha_1 + \gamma_1 z_{t-1} + \sum_{i=1}^{m_1} \beta_{1i} \Delta X_{t-i} + \sum_{i=1}^{m_2} \beta_{2i} \Delta Y_{t-i} + \varepsilon_{1t} \\ \Delta Y_t &= \alpha_2 + \gamma_2 z_{t-1} + \sum_{i=1}^{m_3} \beta_{3i} \Delta X_{t-i} + \sum_{i=1}^{m_4} \beta_{4i} \Delta Y_{t-i} + \varepsilon_{2t} \end{aligned}} \qquad (9)$$

where $(\varepsilon_{1t}, \varepsilon_{2t})'$ is a bivariate white noise and $z_t = X_t - \alpha \cdot Y_t \rightarrow I(0)$.

Granger Representation Theorem: If X_t and Y_t are co-integrated, then exists an ECM representation. Co-integration is a necessary condition for ECM and vice versa.

$$\boxed{\begin{array}{l} \text{If } X_t, Y_t \text{ are not cointegrated} \Rightarrow z_t \rightarrow I(1) \quad \text{(rule 4)} \\ \text{In the ECM, I(1) cannot explain I(0), i.e. } \Delta X_t, \Delta Y_t \Rightarrow \gamma_1 = \gamma_2 = 0 \end{array}}$$

What's z_t?
$z_t = Y_t - \alpha X_t$

MODELING ECM

* Consider the following example in Enders (2010), Applied Econometric Time Series

$$
\begin{aligned}
\Delta r_{S,t} &= a_{10} + \alpha_s (r_{L,t-1} - \beta \cdot r_{S,t-1}) + \sum a_{11}(i) \Delta r_{s,t-i} + \sum a_{12}(i) \Delta r_{L,t-i} + \varepsilon_{S,t} \\
\Delta r_{L,t} &= a_{20} - \alpha_L (r_{L,t-1} - \beta \cdot r_{S,t-1}) + \sum a_{21}(i) \Delta r_{s,t-i} + \sum a_{22}(i) \Delta r_{L,t-i} + \varepsilon_{L,t} \\
&\alpha_s, \alpha_L > 0, \quad \varepsilon_{i,t} \sim \mathrm{WN}(0, \sigma_i^2), \quad i = s, L
\end{aligned}
$$

* This two variable error-correction model is a bivariate VAR in first differences augmented by the error-correction terms. Need to understand the following concepts
  * Speed of adjustment parameters
  * Granger representation theorem
  * Co-integration coefficient restrictions in a VAR model
  * (Weakly) exogenous processes in a co-integrated system

MODELING PROCEDURE

* Engle and Granger (1987) proposed a procedure for modeling and testing co-integration based on an ordinary least squares regression.
* Generally, modeling co-integration or ECM includes two stages of work.
  * The first stage is 1 to find co-integrated relationships in a long-run equilibrium (or disequilibrium term) for our data.
  * The disequilibrium term identified in the first stage is to include the ECM (for the short-run analysis)
  * E.g. Consider two I(1) series Y_t and X_t (checked by unit root tests).

2 Find the co-integrating relation using OLS regression. If the residuals are stationary, i.e., Y_t and X_t is co-integrated, then the long-run equilibrium relationship between Y_t and X_t is given by

$$Y_t = c + \alpha \cdot X_t\,; \quad \mathrm{Y_t}\,, \mathrm{X_t} \sim \mathrm{I(1)}\ \text{processes}$$

$(1, -\alpha)$ are referred to as the co-integration vector between Y_t and X_t.

WHY ENGLE-GRANGER METHOD

* It is very straightforward to implement and to interpret the Engle-Granger procedure.
* From the risk management point of view, the Engle-Granger criterion that minimizes variance is usually more important than the Johansen criterion that maximizes stationarity.
* Sometimes there is a natural choice of dependent variable in the cointegrating regressions, for example, in equity index tracking.