Time Series Analysis

Lecture 11

Review

1. Spurious regression
2. Cointegration
3. Testing for cointegration

Today's Topics

1. Canonical Correlation
2. Maximum Likelihood Estimation
3. Hypothesis Testing

Today's Topics

1. Canonical Correlation
2. Maximum Likelihood Estimation
3. Hypothesis Testing

Today's Topics

1. Canonical Correlation
   - Population Canonical Correlation
   - Sample Canonical Correlation
2. Maximum Likelihood Estimation
3. Hypothesis Testing

Population Canonical Correlation

- Let the $(n_1 \times 1)$ vector $\mathbf{y_t}$ and the $(n_2 \times 1)$ vector $\mathbf{x_t}$ denote stationary random variables.
- Typically, $\mathbf{y_t}$ and $\mathbf{x_t}$ are measured as deviations from their population means, so that $E(\mathbf{y_t}\mathbf{y_t'})$ represents the variance-covariance matrix of $\mathbf{y_t}$.
- In general, there might be complicated correlations among the elements of $\mathbf{y_t}$ and $\mathbf{x_t}$, summarized by the joint variance-covariance matrix

$$\begin{bmatrix} \underset{(n_1 \times n_1)}{\mathbf{E}(\mathbf{y_t}\mathbf{y_t'})} & \underset{(n_1 \times n_2)}{\mathbf{E}(\mathbf{y_t}\mathbf{x_t'})} \\ \underset{(n_2 \times n_1)}{\mathbf{E}(\mathbf{x_t}\mathbf{y_t'})} & \underset{(n_2 \times n_2)}{\mathbf{E}(\mathbf{x_t}\mathbf{x_t'})} \end{bmatrix} = \begin{bmatrix} \underset{(n_1 \times n_1)}{\Sigma_{\mathbf{YY}}} & \underset{(n_1 \times n_2)}{\Sigma_{\mathbf{YX}}} \\ \underset{(n_2 \times n_1)}{\Sigma_{\mathbf{XY}}} & \underset{(n_2 \times n_2)}{\Sigma_{\mathbf{XX}}} \end{bmatrix}.$$

Let $\mathscr{H}'$ and $\mathscr{A}'$ denote $(n \times n_1)$ and $(n \times n_2)$ matrices with $n = min(n_1, n_2)$. Define

$$\boldsymbol{\eta_t} \equiv \mathscr{H}'\mathbf{y}_t, \tag{1}$$

$$\boldsymbol{\xi_t} \equiv \mathscr{A}'\mathbf{x}_t. \tag{2}$$

The matrices $\mathscr{H}$ and $\mathscr{A}$ are chosen so that the following conditions hold.

(1) The individual elements of $\boldsymbol{\eta_t}$ have unit variance and are uncorrelated with one another:

$$E(\boldsymbol{\eta_t}\boldsymbol{\eta_t'}) = \mathscr{H}'\boldsymbol{\Sigma}_{\mathbf{YY}}\mathscr{H} = \mathbf{I_n} \tag{3}$$

(2) The individual elements of $\boldsymbol{\xi_t}$ have unit variance and are uncorrelated with one another:

$$E(\boldsymbol{\xi_t}\boldsymbol{\xi_t'}) = \mathscr{A}'\boldsymbol{\Sigma}_{\mathbf{XX}}\mathscr{A} = \mathbf{I_n} \tag{4}$$

(3) The ith element of $\boldsymbol{\eta_t}$ is uncorrelated with the jth element of $\boldsymbol{\xi_t}$ for $i \neq j$; for $i = j$, the correlation is positive and is given by $\mathbf{r}_i$:

$$E(\boldsymbol{\xi_t}\boldsymbol{\eta_t'}) = \mathscr{A}'\Sigma_{XY}\mathscr{H} = \mathbf{R}, \tag{5}$$

where

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}. \tag{6}$$

(4) The elements of $\boldsymbol{\eta_t}$ and $\boldsymbol{\xi_t}$ are ordered in such a way that

$$(1 \geqslant r_1 \geqslant r_2 \geqslant \cdots \geqslant r_n \geqslant 0). \tag{7}$$

The correlation r_i is known as the ith population canonical correlation between $\mathbf{y}_t$ and $\mathbf{x}_t$.

Implementation

Let $(\lambda_1, \lambda_2, \ldots, \lambda_{n_1})$ denote the eigenvalues of the $(n_1 \times n_1)$ matrix

$$\mathbf{\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}} \tag{8}$$

ordered as

$$(\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{n_1}) \tag{9}$$

with associated eigenvectors $(\mathbf{\bar{k}_1}, \mathbf{\bar{k}_2}, \ldots, \mathbf{\bar{k}_{n_1}})$. Recall that the eigenvalue-eigenvector pair $(\lambda_i, \mathbf{\bar{k}_i})$ satisfies

$$\mathbf{\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\bar{k}_i} = \lambda_i \mathbf{\bar{k}}_i. \tag{10}$$

The usual normalization convention is to set $\mathbf{\bar{k}_i'\bar{k}_i} = 1$. For canonical correlation analysis, however, it is more convenient to ensure that

$$\mathbf{k_i'}\Sigma_{\mathbf{YY}}\mathbf{k_i} = 1 \quad \text{for } i = 1, 2, \ldots, n_1. \tag{11}$$

Condition (11) can be satisfied by setting

$$\mathbf{k_i} = \mathbf{\bar{k}_i} \div \sqrt{\mathbf{\bar{k}_i'}\Sigma_{\mathbf{YY}}\mathbf{\bar{k}_i}}.$$

We further may multiply $\mathbf{k}_i$ by -1 so as to satisfy a certain sign convention (explain later).

It is easy to verify that $\mathbf{k}_i$ is a eigenvector of the matrix in (8).

The canonical correlations $(r_1, r_2, \ldots, r_n)$ turn out to be given by the square roots of the corresponding first n eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ of (8).

The associated $(n \times 1)$ eigenvectors $\mathbf{k_1}, \mathbf{k_2}, \ldots, \mathbf{k_n}$, when normalized by (11) and a sign convention, turn out to make up the rows of the $(n \times n_1)$ matrix $\mathscr{H}'$.

Proposition 1 *Let*

$$\underset{(n_1+n_2)\times(n_1+n_2)}{\boldsymbol{\Sigma}} \equiv \begin{bmatrix} \underset{(n_1\times n_1)}{\boldsymbol{\Sigma}_{\mathbf{YY}}} & \underset{(n_1\times n_2)}{\boldsymbol{\Sigma}_{\mathbf{YX}}} \\ \underset{(n_2\times n_1)}{\boldsymbol{\Sigma}_{\mathbf{XY}}} & \underset{(n_2\times n_2)}{\boldsymbol{\Sigma}_{\mathbf{XX}}} \end{bmatrix}$$

be a positive definite symmetric matrix and let $(\lambda_1, \lambda_2, \ldots, \lambda_{n_1})$ *be the eigenvalues of the matrix in (8), ordered* $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{n_1}$*. Let* $\mathbf{k_1}, \mathbf{k_2}, \ldots, \mathbf{k_{n_1}}$ *be the associated* $(n_1 \times 1)$ *eigenvectors as normalized by (11).*
Let $(\mu_1, \mu_2, \ldots, \mu_{n_2})$ *be the eigenvalues of the* $(n_2 \times n_2)$ *matrix*

$$\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}} \tag{12}$$

ordered $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_{n_2}$. Let $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n_2})$ be the eigenvectors of (12):

$$\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}}\mathbf{a}_i = \mu_i\mathbf{a_i} \tag{13}$$

normalized by

$$\mathbf{a_i'}\boldsymbol{\Sigma}_{\mathbf{XX}}\mathbf{a_i} = 1 \qquad \textit{for } i = 1, 2, \ldots, n_2. \tag{14}$$

Let n be the smaller of n_1 and n_2, and collect the first n vectors $\mathbf{k}_i$ and the first n vectors $\mathbf{a}_j$ in matrices

$$\underset{(n_1 \times n)}{\mathscr{H}} = [\mathbf{k_1}\ \mathbf{k_2}\ \cdots\ \mathbf{k_n}], \tag{15}$$

$$\underset{(n_2 \times n)}{\mathscr{A}} = [\mathbf{a_1}\ \mathbf{a_2}\ \cdots\ \mathbf{a_n}]. \tag{16}$$

Assuming that $\lambda_1, \lambda_2, \ldots, \lambda_n$ *are distinct, then*

(a) $0 \leqslant \lambda_i < 1$ *for* $i = 1, 2, \ldots, n_1$ *and* $0 \leqslant \mu_j < 1$ *for* $j = 1, 2, \ldots, n_2$*;*

(b) $\lambda_i = \mu_i$ *for* $i = 1, 2, \ldots, n$*;*

(c) $\mathscr{H}'\mathbf{\Sigma_{YY}}\mathscr{H} = \mathbf{I_n}$ *and* $\mathscr{A}'\mathbf{\Sigma_{XX}}\mathscr{A} = \mathbf{I_n}$*;*

(d) $\mathscr{A}'\mathbf{\Sigma_{XY}}\mathscr{H} = \mathbf{R}$,

where

$$\mathbf{R^2} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \tag{17}$$

Remark From (d) we know $r_i = \mathbf{a_i'}\Sigma_{\mathbf{XY}}\mathbf{k_i}$. But the magnitude $\mathbf{a_i'}\Sigma_{\mathbf{XY}}\mathbf{k_i}$ need not to be positive. If $\mathbf{a_i'}\Sigma_{\mathbf{XY}}\mathbf{k_i} < 0$ for some i, one can replace $\mathbf{k_i}$ with $-\mathbf{k_i}$, so that the r_i will correspond to the positive square root of λ_i.

Special case: $n = n_1 = 1$

In this case, the matrix (8) is just a scalar, a (1×1) "matrix" that is equal to its own eigenvalue, i.e. $\lambda_1 = \Sigma_{YY}^{-1} \mathbf{\Sigma}_{Y\mathbf{X}} \mathbf{\Sigma}_{\mathbf{XX}}^{-1} \mathbf{\Sigma}_{\mathbf{X}Y}$. Thus, the squared population canonical correlation between a scalar y_t and a set of n_2 explanatory variables $\mathbf{x}_t$ is given by

$$r_1^2 = \lambda_1 = \frac{\mathbf{\Sigma}_{Y\mathbf{X}} \mathbf{\Sigma}_{\mathbf{XX}}^{-1} \mathbf{\Sigma}_{\mathbf{X}Y}}{\Sigma_{YY}}.$$

To interpret this expression, recall that the mean squared error of a linear projection of y_t on $\mathbf{x_t}$ is given by

$$MSE = \Sigma_{YY} - \mathbf{\Sigma}_{Y\mathbf{X}} \mathbf{\Sigma}_{\mathbf{XX}}^{-1} \mathbf{\Sigma}_{\mathbf{X}Y},$$

and so

$$1 \ - \ r_1^2 = \frac{\Sigma_{YY}}{\Sigma_{YY}} \ - \ \frac{\boldsymbol{\Sigma}_{Y\mathbf{X}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{X}Y}}{\Sigma_{YY}} \ = \ \frac{MSE}{\Sigma_{YY}}.$$

Thus, for this simple case, r_1^2 is the population squared multiple correlation coefficient, commonly denoted R^2.

Today's Topics

1. Canonical Correlation
   - Population Canonical Correlation
   - Sample Canonical Correlation
2. Maximum Likelihood Estimation
3. Hypothesis Testing

Sample Canonical Correlations

Suppose we have a sample of T observations on the $(n_1 \times 1)$ vector $\mathbf{y_t}$ and the $(n_2 \times 1)$ vector $\mathbf{x_t}$, whose sample moments are given by

$$\hat{\Sigma}_{\mathbf{YY}} = (1/T)\sum_{t=1}^{T} \mathbf{y_t y_t'} \tag{18}$$

$$\hat{\Sigma}_{\mathbf{YX}} = (1/T)\sum_{t=1}^{T} \mathbf{y_t x_t'} \tag{19}$$

$$\hat{\Sigma}_{\mathbf{XX}} = (1/T)\sum_{t=1}^{T} \mathbf{x_t x_t'}. \tag{20}$$

Again, in many applications, $\mathbf{y_t}$ and $\mathbf{x_t}$ would be measured in deviations from their sample means.

To calculate sample canonical correlations, the objective is to generate a set of T observations on a new $(n \times 1)$ vector $\hat{\boldsymbol{\eta}}_{\mathbf{t}}$, where $n = min(n_1, n_2)$. The vector $\hat{\boldsymbol{\eta}}_{\mathbf{t}}$ is a linear combination of the observed value of $\mathbf{y_t}$:

$$\hat{\boldsymbol{\eta}}_{\mathbf{t}} = \hat{\mathscr{H}}' \mathbf{y_t} \tag{21}$$

for $\hat{\mathscr{H}}$ an $(n_1 \times n)$ matrix to be estimated from the data. The task will be to choose $\hat{\mathscr{H}}$ so that the ith generated series $(\hat{\eta}_{it})$ has unit sample variance and is orthogonal to the jth generated series:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_{\mathbf{t}} \hat{\boldsymbol{\eta}}_{\mathbf{t}}' = \mathbf{I_n}. \tag{22}$$

Similarly, we will generate an $(n \times 1)$ vector $\hat{\boldsymbol{\xi}}_t$ from the elements of $\mathbf{x_t}$:

$$\hat{\boldsymbol{\xi}}_\mathbf{t} = \hat{\mathscr{A}}'\mathbf{x_t}. \tag{23}$$

Each of the variables $\hat{\xi}_{it}$ has unit sample variance and is orthogonal to $\hat{\xi}_{jt}$ for $i \neq j$:

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_\mathbf{t}\hat{\boldsymbol{\xi}}_\mathbf{t}' = \mathbf{I_n}. \tag{24}$$

Finally, $\hat{\eta}_{it}$ is orthogonal to $\hat{\xi}_{jt}$ for $i \neq j$, while the sample correlation between $\hat{\eta}_{it}$ and $\hat{\xi}_{it}$ is called the sample canonical correlation coefficient:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_{\mathbf{t}} \hat{\boldsymbol{\eta}}_{\mathbf{t}}' = \hat{\mathbf{R}} \tag{25}$$

for

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}_1 & 0 & \cdots & 0 \\ 0 & \hat{r}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{r}_n \end{bmatrix}. \tag{26}$$

Finding matrices $\hat{\mathscr{H}}, \hat{\mathscr{A}}$ and $\hat{\mathbf{R}}$ satisfying (22), (24), and (25) involves exactly the same calculations as did finding matrices $\mathscr{H}, \mathscr{A}$ and $\mathbf{R}$ satisfying (3) through (5).

For example, (21) allows us to write (22) as

$$\mathbf{I_n} = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_{\mathbf{t}}\hat{\boldsymbol{\eta}}_{\mathbf{t}}' = \hat{\mathscr{H}}'(1/T)\sum_{t=1}^{T} \mathbf{y_t}\mathbf{y_t}'\hat{\mathscr{H}} = \hat{\mathscr{H}}'\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\hat{\mathscr{H}}, \quad (27)$$

where the last line follows from (18). Expression (27) is identical to (3) with hats placed over the variables.

Similarly, substituting (23) into (24) gives $\hat{\mathscr{A}}'\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\hat{\mathscr{A}} = \mathbf{I_n}$, which corresponds to (4).

Equation (25) becomes $\hat{\mathscr{A}}'\hat{\boldsymbol{\Sigma}}_{\mathbf{XY}}\hat{\mathscr{H}} = \hat{\mathbf{R}}$, as in (5).

Again, we can replace $\hat{\mathbf{k}}_{\mathbf{i}}$ with $-\hat{\mathbf{k}}_{\mathbf{i}}$ if any of the elements of $\hat{\mathbf{R}}$ should turn out negative.

Thus, to calculate the sample canonical correlations, the procedure described in **Proposition 1** is simply applied to the sample moments $(\hat{\Sigma}_{YY}, \hat{\Sigma}_{YX}, \text{and} \hat{\Sigma}_{XX})$ rather than to the population moments.

In particular, the square of the ith sample canonical correlation ($\hat{r}_i^2$) is given by the ith largest eigenvalue of the matrix

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{XY}} = \left\{(1/T)\sum_{t=1}^{T}\mathbf{y_t}\mathbf{y_t'}\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\mathbf{y_t}\mathbf{x_t'}\right\} \times \left\{(1/T)\sum_{t=1}^{T}\mathbf{x_t}\mathbf{x_t'}\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\mathbf{x_t}\mathbf{y_t'}\right\}. \tag{28}$$

The ith column of $\hat{\mathscr{H}}$ is given by the eigenvector associated with this ith eigenvalue, normalized so that

$$\mathbf{\hat{k}_i'}\left\{(1/T)\sum_{t=1}^{T}\mathbf{y_t}\mathbf{y_t'}\right\}\mathbf{\hat{k}_i} = 1. \tag{29}$$

The ith column of $\hat{\mathscr{A}}$ is given by the eigenvector associated with the eigenvalue $\hat{\lambda}_i$ for the matrix $\mathbf{\hat{\Sigma}_{XX}^{-1}\hat{\Sigma}_{XY}\hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YX}}$ normalized by the condition that $\mathbf{\hat{a}_i'\hat{\Sigma}_{XX}\hat{a}_i} = 1$.

Special case: $n = n_1 = 1$

Then the matrix in (28) is a scalar equal to its own eigenvalue. Hence, the sample squared canonical correlation between the scalar y_t and a set of n_2 explanatory variables $\mathbf{x_t}$ is given by

$$\begin{aligned}
\hat{r}_1^2 &= \frac{\{T^{-1}\sum y_t \mathbf{x_t'}\}\{T^{-1}\sum \mathbf{x_t}\mathbf{x_t'}\}^{-1}\{T^{-1}\sum \mathbf{x_t} y_t\}}{\{T^{-1}\sum y_t^2\}} \\
&= \frac{\{\sum y_t \mathbf{x_t'}\}\{\sum \mathbf{x_t}\mathbf{x_t'}\}^{-1}\{\sum \mathbf{x_t} y_t\}}{\{\sum y_t^2\}},
\end{aligned}$$

which is just the squared sample multiple correlation coefficient R^2.

Today's Topics

1. Canonical Correlation

2. Maximum Likelihood Estimation

3. Hypothesis Testing

The Error Correction Model

Let $\mathbf{y_t}$ denote an $(n \times 1)$ vector. The maintained hypothesis is that $\mathbf{y_t}$ follows a $VAR(p)$ in levels. Recall that any pth-order VAR can be written in the form

$$\mathbf{\Delta y_t} = \boldsymbol{\zeta_1} \mathbf{\Delta y_{t-1}} + \boldsymbol{\zeta_2} \mathbf{\Delta y_{t-2}} + \cdots + \boldsymbol{\zeta_{p-1}} \mathbf{\Delta y_{t-p+1}} + \boldsymbol{\alpha} + \boldsymbol{\zeta_0} \mathbf{y_{t-1}} + \boldsymbol{\varepsilon_t}, \tag{30}$$

with

$$E(\boldsymbol{\varepsilon_t}) = 0,$$
$$E(\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon'_\tau}) = \begin{cases} \mathbf{\Omega} & \text{for } t = \tau, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Suppose that each individual variable y_{it} is $I(1)$, although h linear combinations of $\mathbf{y_t}$ are stationary. This implies that $\boldsymbol{\zeta_0}$ can be written in the form

$$\zeta_{\mathbf{0}} = -\mathbf{BA}' \tag{31}$$

for $\mathbf{B}$ an $(n \times h)$ matrix and $\mathbf{A}'$ an $(h \times n)$ matrix.

Consider a sample of $T + p$ observations on $\mathbf{y_t}$ denoted $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \cdots, \mathbf{y}_T)$. If the disturbances ε_t are Gaussian, then the log likelihood of $(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T)$ conditional on $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \cdots, \mathbf{y}_0)$ is given by

$$
\begin{aligned}
& \mathscr{L}(\boldsymbol{\Omega}, \zeta_{\mathbf{1}}, \zeta_{\mathbf{2}}, \cdots, \zeta_{\mathbf{p-1}}, \boldsymbol{\alpha}, \zeta_{\mathbf{0}}) \\
= \; & (-Tn/2)\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}| \\
& -(1/2)\sum_{t=1}^{T} \big[(\boldsymbol{\Delta}\mathbf{y_t} - \zeta_{\mathbf{1}}\boldsymbol{\Delta}\mathbf{y_{t-1}} - \zeta_{\mathbf{2}}\boldsymbol{\Delta}\mathbf{y_{t-2}} - \cdots \\
& -\zeta_{\mathbf{p-1}}\boldsymbol{\Delta}\mathbf{y_{t-p+1}} - \boldsymbol{\alpha} - \zeta_{\mathbf{0}}\mathbf{y_{t-1}})'\boldsymbol{\Omega}^{-1} \\
& \times(\boldsymbol{\Delta}\mathbf{y_t} - \zeta_{\mathbf{1}}\boldsymbol{\Delta}\mathbf{y_{t-1}} - \zeta_{\mathbf{2}}\boldsymbol{\Delta}\mathbf{y_{t-2}} - \cdots \\
& -\zeta_{\mathbf{p-1}}\boldsymbol{\Delta}\mathbf{y_{t-p+1}} - \boldsymbol{\alpha} - \zeta_{\mathbf{0}}\mathbf{y_{t-1}})\big]. \qquad (32)
\end{aligned}
$$

The goal is to choose $(\boldsymbol{\Omega}, \zeta_{\mathbf{1}}, \zeta_{\mathbf{2}}, \cdots, \zeta_{\mathbf{p-1}}, \boldsymbol{\alpha}, \zeta_{\mathbf{0}})$ so as to maximize (32) subject to the constraint that $\zeta_{\mathbf{0}} = -\mathbf{BA'}$.

Step 1: Calculate Auxiliary Regressions

The first step is to estimate a $(p-1)$th-order *VAR* for $\mathbf{\Delta y_t}$:

$$\mathbf{\Delta y_t} = \hat{\boldsymbol{\pi}}_\mathbf{0} + \hat{\boldsymbol{\Pi}}_\mathbf{1}\mathbf{\Delta y_{t-1}} + \hat{\boldsymbol{\Pi}}_\mathbf{2}\mathbf{\Delta y_{t-2}} + \cdots + \hat{\boldsymbol{\Pi}}_\mathbf{p-1}\mathbf{\Delta y_{t-p+1}} + \hat{\mathbf{u}}_\mathbf{t}, \tag{33}$$

where $\hat{\boldsymbol{\Pi}}_\mathbf{i}$ denotes an $(n \times n)$ matrix of *OLS* coefficient estimates and $\hat{\mathbf{u}}_\mathbf{t}$ denotes the $(n \times 1)$ vector of *OLS* residuals.

We also estimate a second battery of regressions,

$$\mathbf{y_{t-1}} = \hat{\theta} + \hat{\aleph}_1\mathbf{\Delta y_{t-1}} + \hat{\aleph}_2\mathbf{\Delta y_{t-2}} + \cdots + \hat{\aleph}_\mathbf{p-1}\mathbf{\Delta y_{t-p+1}} + \hat{\mathbf{v}}_\mathbf{t}, \tag{34}$$

with $\hat{\mathbf{v}}_\mathbf{t}$ the $(n \times 1)$ vector of residuals from this second battery of regressions.

Step 2: Calculate Canonical Correlations

Next calculate the sample variance-covariance matrices of the *OLS* residuals $\hat{\mathbf{u}}_{\mathbf{t}}$ and $\hat{\mathbf{v}}_{\mathbf{t}}$:

$$\hat{\Sigma}_{\mathbf{VV}} \equiv (1/T) \sum_{t=1}^{T} \mathbf{v_t} \mathbf{v_t'}, \tag{35}$$

$$\hat{\Sigma}_{\mathbf{UU}} \equiv (1/T) \sum_{t=1}^{T} \mathbf{u_t} \mathbf{u_t'}, \tag{36}$$

$$\hat{\Sigma}_{\mathbf{UV}} \equiv (1/T) \sum_{t=1}^{T} \mathbf{u_t} \mathbf{v_t'}, \tag{37}$$

$$\hat{\Sigma}_{\mathbf{VU}} \equiv \hat{\Sigma}'_{\mathbf{UV}}.$$

From these, find the eigenvalues of the matrix

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{VV}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{VU}}\hat{\boldsymbol{\Sigma}}_{\mathbf{UU}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{UV}} \tag{38}$$

with the eigenvalues ordered $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$.

The maximum value attained by the log likelihood function subject to the constraint that there are h cointegrating relations is given by

$$\begin{aligned} \mathscr{L}^* &= -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log\left|\hat{\boldsymbol{\Sigma}}_{UU}\right| \\ &\quad -(T/2)\sum_{i=1}^{h}\log(1-\hat{\lambda}_i). \end{aligned} \tag{39}$$

Step 3: Calculate Maximum Likelihood Estimates of Parameters

If maximum likelihood estimates of parameters are also desired, these can be calculated as follows.

Let $\hat{\mathbf{a}}_{\mathbf{1}}, \hat{\mathbf{a}}_{\mathbf{2}}, \ldots, \hat{\mathbf{a}}_{\mathbf{h}}$ denote the $(n \times 1)$ eigenvectors of (38) associated with the h largest eigenvalues, and are normalized by the conditions $\hat{\mathbf{a}}_{\mathbf{i}}' \hat{\Sigma}_{\mathbf{VV}} \hat{\mathbf{a}}_{\mathbf{i}} = 1$. Collect the first h normalized vectors in an $(n \times h)$ matrix $\hat{\mathbf{A}}$:

$$\hat{\mathbf{A}} \equiv [\hat{\mathbf{a}}_{\mathbf{1}} \ \hat{\mathbf{a}}_{\mathbf{2}} \ \cdots \ \hat{\mathbf{a}}_{\mathbf{h}}]. \tag{40}$$

Then the *MLE* of $\boldsymbol{\zeta}_{\mathbf{0}}$ is given by

$$\hat{\boldsymbol{\zeta}}_{\mathbf{0}} = \hat{\Sigma}_{\mathbf{UV}} \hat{\mathbf{A}} \hat{\mathbf{A}}'. \tag{41}$$

The *MLE* of $\zeta_{\mathbf{i}}$ for $i = 1, 2, \ldots, p-1$ is

$$\hat{\zeta}_{\mathbf{i}} = \hat{\mathbf{\Pi}}_{\mathbf{i}} - \hat{\zeta}_{\mathbf{0}}\hat{\aleph}_{\mathbf{i}}, \tag{42}$$

and the *MLE* of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\pi}}_{\mathbf{0}} - \hat{\zeta}_{\mathbf{0}}\hat{\boldsymbol{\theta}}. \tag{43}$$

The *MLE* of $\mathbf{\Omega}$ is

$$\hat{\mathbf{\Omega}} = (1/T)\sum_{t=1}^{T}[(\hat{\mathbf{u}}_{\mathbf{t}} - \hat{\zeta}_{\mathbf{0}}\hat{\mathbf{v}}_{\mathbf{t}})(\hat{\mathbf{u}}_{\mathbf{t}} - \hat{\zeta}_{\mathbf{0}}\hat{\mathbf{v}}_{\mathbf{t}})']. \tag{44}$$

Today's Topics

1. Canonical Correlation
2. Maximum Likelihood Estimation
3. Hypothesis Testing

Today's Topics

1. Canonical Correlation

2. Maximum Likelihood Estimation

3. Hypothesis Testing
   Testing the Null Hypothesis of h Cointegratng Relations
   Examples

H_0: there are exactly h cointegrating relations

Suppose that the $(n \times 1)$ vector $\mathbf{y_t}$ can be characterized by a $VAR(p)$ in levels:

$$\boldsymbol{\Delta}\mathbf{y_t} = \boldsymbol{\zeta_1}\boldsymbol{\Delta}\mathbf{y_{t-1}} + \boldsymbol{\zeta_2}\boldsymbol{\Delta}\mathbf{y_{t-2}} + \cdots + \boldsymbol{\zeta_{p-1}}\boldsymbol{\Delta}\mathbf{y_{t-p+1}} + \boldsymbol{\alpha} + \boldsymbol{\zeta_0}\mathbf{y_{t-1}} + \boldsymbol{\varepsilon_t}. \tag{45}$$

Under the null hypothesis H_0 that there are exactly h cointegrating relations among the elements of $\mathbf{y_t}$, this VAR is restricted by the requirement that $\boldsymbol{\zeta_0}$ can be written in the form $\boldsymbol{\zeta_0} = -\mathbf{BA'}$, for $\mathbf{B}$ an $(n \times h)$ matrix and $\mathbf{A'}$ an $(h \times n)$ matrix.

The largest value that can be achieved for the log likelihood function under this constraint was given by (39):

$$\mathscr{L}_0^* = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\mathbf{\Sigma}}_{\mathbf{UU}}| \\ - (T/2)\sum_{i=1}^{h}\log(1-\hat{\lambda}_i). \tag{46}$$

H_{A1}: there are n cointegrating relations

Consider the alternative hypothesis that there are n cointegrating relations. This amounts to the claim that every linear combination of $\mathbf{y_t}$ is stationary, in which case no restrictions are imposed on $\boldsymbol{\zeta_0}$. The value for the log likelihood function in the absence of constraints is given by

$$\mathscr{L}_A^* = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\boldsymbol{\Sigma}}_{\mathbf{UU}}| - (T/2)\sum_{i=1}^{n}\log(1-\hat{\lambda}_i). \quad (47)$$

A likelihood ratio test of H_0 against H_{A1} can be based on

$$2(\mathscr{L}_A^* - \mathscr{L}_0^*) = -T \sum_{i=h+1}^{n} log(1 - \hat{\lambda}_i). \tag{48}$$

This is also called the trace statistic.

- Case 1: $\boldsymbol{\alpha} = 0$, and no constant term is included in the auxiliary regressions (33) and (34)
- Case 2: $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*$, where $\boldsymbol{\mu}_1^* = E[\mathbf{z}_t]$, $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$, and no restrictions are imposed on the constant term in the estimation of auxiliary regressions (33) and (34)
- Case 3: one or more elements of $\boldsymbol{\alpha} - \mathbf{B}\boldsymbol{\mu}_1^*$ are nonzero

Note that $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*$ implies $E[\Delta \mathbf{y}_t] = 0$.

TABLE B.10
Critical Values for Johansen's Likelihood Ratio Test of the Null Hypothesis of h Cointegrating Relations Against the Alternative of No Restrictions

Number of random walks $(g = n - h)$ (g)	*Sample size* (T)	*Probability that* $2(\mathcal{L}_A - \mathcal{L}_0)$ *is greater than entry*					
		0.500	*0.200*	*0.100*	*0.050*	*0.025*	*0.001*
				Case 1			
1	400	0.58	1.82	2.86	3.84	4.93	6.51
2	400	5.42	8.45	10.47	12.53	14.43	16.31
3	400	14.30	18.83	21.63	24.31	26.64	29.75
4	400	27.10	33.16	36.58	39.89	42.30	45.58
5	400	43.79	51.13	55.44	59.46	62.91	66.52
				Case 2			
1	400	2.415	4.905	6.691	8.083	9.658	11.576
2	400	9.335	13.038	15.583	17.844	19.611	21.962
3	400	20.188	25.445	28.436	31.256	34.062	37.291
4	400	34.873	41.623	45.248	48.419	51.801	55.551
5	400	53.373	61.566	65.956	69.977	73.031	77.911
				Case 3			
1	400	0.447	1.699	2.816	3.962	5.332	6.936
2	400	7.638	11.164	13.338	15.197	17.299	19.310
3	400	18.759	23.868	26.791	29.509	32.313	35.397
4	400	33.672	40.250	43.964	47.181	50.424	53.792
5	400	52.588	60.215	65.063	68.905	72.140	76.955

The probability shown at the head of the column is the area in the right-hand tail. The number of random walks under the null hypothesis (g) is given by the number of variables described by the vector autoregression (n) minus the number of cointegrating relations under the null hypothesis (h). In each case the alternative is that $g = 0$.

H_{A2}: there are $h+1$ cointegrating relations

The test statistic of this test is given by

$$2(\mathscr{L}_A^* - \mathscr{L}_0^*) = -T \log(1 - \hat{\lambda}_{h+1}). \tag{49}$$

This is also called the eigenvalue statistic.

When $g = 1$, these two statistics (48) and (49) are identical.

Same conditions apply to define three different cases. The asymptotic distribution of (49) is given in Table B.11.

TABLE B.11
Critical Values for Johansen's Likelihood Ratio Test of the Null Hypothesis of h Cointegrating Relations Against the Alternative of $h+1$ Relations

Number of random walks ($g = n - h$) (g)	Sample size (T)	Probability that $2(\mathcal{L}_A - \mathcal{L}_0)$ is greater than entry					
		0.500	0.200	0.100	0.050	0.025	0.001
				Case 1			
1	400	0.58	1.82	2.86	3.84	4.93	6.51
2	400	4.83	7.58	9.52	11.44	13.27	15.69
3	400	9.71	13.31	15.59	17.89	20.02	22.99
4	400	14.94	18.97	21.58	23.80	26.14	28.82
5	400	20.16	24.83	27.62	30.04	32.51	35.17
				Case 2			
1	400	2.415	4.905	6.691	8.083	9.658	11.576
2	400	7.474	10.666	12.783	14.595	16.403	18.782
3	400	12.707	16.521	18.959	21.279	23.362	26.154
4	400	17.875	22.341	24.917	27.341	29.599	32.616
5	400	23.132	27.953	30.818	33.262	35.700	38.858
				Case 3			
1	400	0.447	1.699	2.816	3.962	5.332	6.936
2	400	6.852	10.125	12.099	14.036	15.810	17.936
3	400	12.381	16.324	18.697	20.778	23.002	25.521
4	400	17.719	22.113	24.712	27.169	29.335	31.943
5	400	23.211	27.899	30.774	33.178	35.546	38.341

The probability shown at the head of the column is the area in the right-hand tail. The number of random walks under the null hypothesis (g) is given by the number of variables described by the vector autoregression (n) minus the number of cointegrating relations under the null hypothesis (h). In each case the alternative is that there are $h + 1$ cointegrating relations.

Today's Topics

1. Canonical Correlation
2. Maximum Likelihood Estimation
3. Hypothesis Testing
   - Testing the Null Hypothesis of h Cointegratng Relations
   - Examples

Example 1

We illustrate Johansens likelihood method with the U.S. Treasury real yield curve (daily) rates at fixed maturities 5, 7, 10, 20, and 30 years in Janunary 2, 2013 - February 11, 2014.

There are in total 278×5 observations.

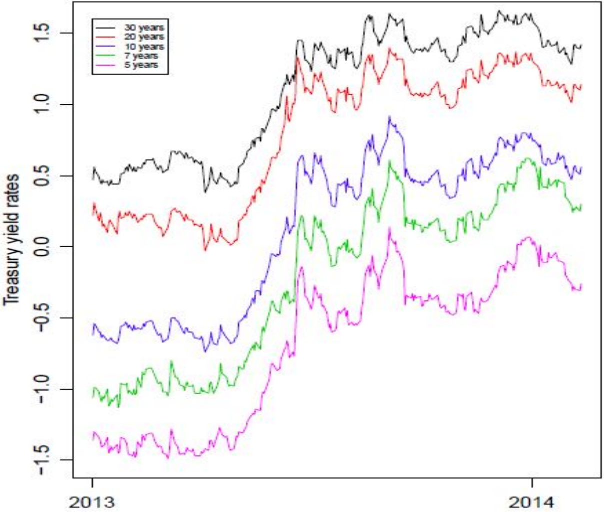


Figure 1 : The daily U.S. Treasury real yield curve rates at fixed maturities of five, senven, ten, twenty and thirty years in the period of Janunary 2, 2013 - February 11, 2014.

We first check if all five component series are I(1) processes.

This can be done by applying the Augmented Dickey-Fuller (ADF) test for each component series and its differenced series, which can be carried out using ur.df in the R-package urca.

Tests confirm that all the five yields are I(1) series.

trace test

```
> m1=ca.jo(tbill, type="trace", ecdet="none", K=2,
          spec="transitory")
> summary(m1)
Eigenvalues (lambda):
[1] 0.124620 0.116170 0.056899 0.025242 0.004652

Values of teststatistic and critical values of test:
          test 10pct  5pct  1pct
r <= 4 |  1.29  6.50  8.18 11.65
r <= 3 |  8.34 15.66 17.95 23.52
r <= 2 | 24.51 28.71 31.52 37.22
r <= 1 | 58.60 45.23 48.28 55.43
r = 0  | 95.33 66.49 70.60 78.87
```

The test rejects $H_0 : h \leq 1$ at the 1% significant level, but cannot reject the hypothesis for $h = 2$ even at the 10% level. This indicates that there exists two cointegrating relations among the five yield series.

eigenvalue test

$> ca.jo(tbill, type = \text{“}eigen\text{”}, ecdet = \text{“}none\text{”}, K = 2, spec = \text{“}transitory\text{”})$

```
Values of teststatistic and critical values of test:
          test 10pct  5pct  1pct
r <= 4 |  1.29  6.50  8.18 11.65
r <= 3 |  7.06 12.91 14.90 19.19
r <= 2 | 16.17 18.90 21.07 25.75
r <= 1 | 34.08 24.78 27.14 32.14
r = 0  | 36.73 30.84 33.32 38.78
```

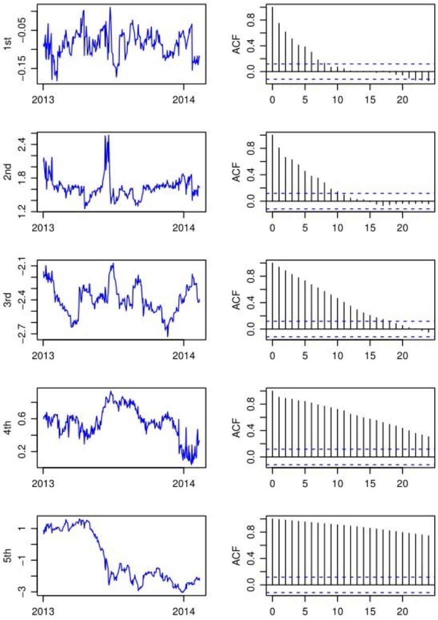
This test also indicates two cointegration relations.

The obtained transformation are now collected in y:
$> y = as.matrix(tbill\% * \%as.matrix(m1@V))$

The two estimated error correction terms are:

$$\begin{aligned} ect1 &= 0.58 * 5Year.l1 - 0.44 * 7Year.l1 - 0.17 * 10Year.l1 \\ &\quad -0.39 * 20Year.l1 + 0.53 * 30Year.l1, \\ ect2 &= 0.11 * 5Year.l1 + 0.14 * 7Year.l1 - 0.72 * 10Year.l1 \\ &\quad +0.66 * 20Year.l1 - 0.13 * 30Year.l1. \end{aligned}$$

To refit the ECM with $r = 2$ fixed, run R-function $cajorls(m1, r = 2)$.


1st
−0.15
−0.05
2013
2014
ACF
0.0
0.4
0.8
0
5
10
15
20
2nd
1.2
1.8
2.4
3rd
−2.7
−2.4
−2.1
4th
0.2
0.6
5th
−3
−1
1

Example 2

Consider two weekly U.S. short-term interest rates. The series are the 3-month Treasury bill (TB) rate and 6-month Treasury bill rate from December 12, 1958, to August 6, 2004, for 2383 observations.

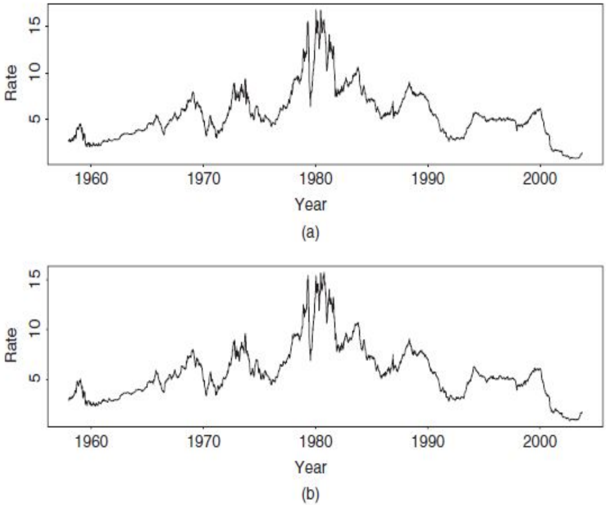


Figure 2 : Time plots of weekly U.S. interest rate from December 12, 1958, to August 6, 2004. (a) The 3-month Treasury bill rate and (b) 6-month Treasury bill rate. Rates are from secondary market.

Denote the two series by $tb3m$ and $tb6m$ and define the vector series $\mathbf{x_t} = (tb3m_t, tb6m_t)'$. The augmented Dickey-Fuller unit-root tests fail to reject the hypothesis of a unit root in the individual series.

Thus, we proceed to VAR modeling. For the bivariate series $\mathbf{x_t}$, the BIC criterion selects a VAR(3) model:

To perform a cointegration test, we choose a restricted constant because there is no reason a priori to believe the existence of a drift in the U.S. interest rate. Both Johansens tests confirm that the two series are cointegrated with one cointegrating vector when a VAR(3) model is entertained.

Tests for Cointegration Rank:

	Eigenvalue	Trace Stat	95% CV	99% CV
H(0)++**	0.0322	83.2712	19.96	24.60
H(1)	0.0023	5.4936	9.24	12.97

	Max Stat	95% CV	99% CV
H(0)++**	77.7776	15.67	20.20
H(1)	5.4936	9.24	12.97

Next, we perform the maximum-likelihood estimation of the specified cointegrated VAR(3) model using an ECM presentation. The results are as follows:

$$\Delta \mathbf{x}_t = \begin{bmatrix} -0.09 \\ -0.02 \end{bmatrix} (z_{t-1} + 0.23) + \begin{bmatrix} 0.05 & 0.27 \\ -0.04 & 0.32 \end{bmatrix} \Delta \mathbf{x}_{t-1} + \begin{bmatrix} -0.21 & 0.25 \\ -0.03 & 0.10 \end{bmatrix} \Delta \mathbf{x}_{t-2} + \mathbf{a}_t,$$

where the error correction term $z_{t-1} \approx tb3m_{t-1} - tb6m_{t-1}$.