Ch. 24 Johansen's MLE for Cointegration

We have so far considered only single-equation estimation and testing for cointegration. While the estimation of single equation is convenient and often consistent, for some purpose only estimation of a system provides sufficient information. This is true, for example, when we consider the estimation of multiple cointegrating vectors, and inference about the number of such vectors. This chapter examines methods of finding the cointegrating rank and derive the asymptotic distributions. To develop these results, we first begin with a discussion of canonical correlation analysis.

1 Canonical Correlation

1.1 Population Canonical Correlations

Let the $(n_1 \times 1)$ vector $\mathbf{y}_t$ and the $(n_2 \times 1)$ vector $\mathbf{x}_t$ denote stationary random vector that are measured as deviation from their population means, so that $E(\mathbf{y}_t\mathbf{y}_t')$ represent the variance-covariance matrix of $\mathbf{y}_t$. In general, there might be complicated correlations among the element of $\mathbf{y}_t$ and $\mathbf{x}_t$, i.e.

$$E\left[\begin{bmatrix} \mathbf{y}_t \\ \mathbf{x}_t \end{bmatrix}\begin{bmatrix} \mathbf{y}_t \\ \mathbf{x}_t \end{bmatrix}'\right] = \begin{bmatrix} E(\mathbf{y}_t\mathbf{y}_t') & E(\mathbf{y}_t\mathbf{x}_t') \\ E(\mathbf{x}_t\mathbf{y}_t') & E(\mathbf{x}_t\mathbf{x}_t') \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma_{yy}} & \mathbf{\Sigma_{yx}} \\ \mathbf{\Sigma_{xy}} & \mathbf{\Sigma_{xx}} \end{bmatrix}.$$

If the two set are very large, the investigator may wish to study only a few of linear combination of $\mathbf{y}_t$ and $\mathbf{x}_t$ which yield most highly correlated. He may find that the interrelation is completely described by the correlation between the first few **canonical variate**.

We now define two new $(n \times 1)$ random vectors, $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$, where n the smaller of n_1 and n_2. These vectors are linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$, respectively:

$$\begin{aligned} \boldsymbol{\eta}_t &\equiv \boldsymbol{\mathcal{K}}'\mathbf{y}_t, \\ \boldsymbol{\xi}_t &\equiv \boldsymbol{\mathcal{A}}'\mathbf{x}_t. \end{aligned}$$

Here, $\boldsymbol{\mathcal{K}}'$ and $\boldsymbol{\mathcal{A}}'$ are $(n \times n_1)$ and $(n \times n_2)$ matrices, respectively. The matrices $\boldsymbol{\mathcal{K}}'$ and $\boldsymbol{\mathcal{A}}'$ are chosen such that the following conditions holds.

(a) $E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \boldsymbol{\mathcal{K}}'\mathbf{\Sigma_{yy}}\boldsymbol{\mathcal{K}} = \mathbf{I}_n$ and $E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \boldsymbol{\mathcal{A}}'\mathbf{\Sigma_{xx}}\boldsymbol{\mathcal{A}} = \mathbf{I}_n$.

(b) $E(\boldsymbol{\xi}_t \boldsymbol{\eta}_t') = \boldsymbol{\mathcal{A}}'\boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\mathcal{K}} = \mathbf{R}$, where

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & . & . & . & 0 \\ 0 & r_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & r_n \end{bmatrix},$$

and $r_i \geq 0, i = 1, 2, ..., n$.
(c) The elements of $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are ordered in such a way that

$$1 \geq r_1 \geq r_2 \geq ... \geq r_n \geq 0.$$

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

The population canonical correlations and the value of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{K}}$ can be calculated as follows.

Theorem 1:
Let

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{yy}} & \boldsymbol{\Sigma}_{\mathbf{yx}} \\ \boldsymbol{\Sigma}_{\mathbf{xy}} & \boldsymbol{\Sigma}_{\mathbf{xx}} \end{bmatrix}$$

be a positive definite symmetric matrix and let $(\lambda_1, \lambda_2, ..., \lambda_{n_1})$ be the eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}}$ ordered $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{n_1}$. Let $(\mathbf{k}_1, \mathbf{k}_2, ..., \mathbf{k}_{n_1})$ be the associated $(n_1 \times 1)$ eigenvectors as normalized by[1]

$$\mathbf{k}_i'\boldsymbol{\Sigma}_{\mathbf{yy}}\mathbf{k}_i = 1 \quad for\ i = 1, 2, ..., n_1.$$

Let $(\mu_1, \mu_2, ..., \mu_{n_2})$ be the eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}}$ ordered $\mu_1 \geq \mu_2 \geq ... \geq \mu_{n_2}$. Let $(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_{n_2})$ be the associated $(n_2 \times 1)$ eigenvectors as normalized by

$$\mathbf{a}_i'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{a}_i = 1 \quad for\ i = 1, 2, ..., n_2.$$

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

$$\begin{aligned} \boldsymbol{\mathcal{K}} &= [\mathbf{k}_1\ \mathbf{k}_2\ ...\ \mathbf{k}_n], \\ \boldsymbol{\mathcal{A}} &= [\mathbf{a}_1\ \mathbf{a}_2\ ...\ \mathbf{a}_n]. \end{aligned}$$

[1] If a computer program has calculated eigenvectors $(\tilde{\mathbf{k}}_1, \tilde{\mathbf{k}}_2, ..., \tilde{\mathbf{k}}_{n_1})$ of the matrix $\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}}$ normalized by $\tilde{\mathbf{k}}_i'\tilde{\mathbf{k}}_i = 1$, it is trival to change these to eigenvectors $(\mathbf{k}_1, \mathbf{k}_2, ..., \mathbf{k}_{n_1})$ by setting $\mathbf{k}_i = \tilde{\mathbf{k}}_i \div \sqrt{\tilde{\mathbf{k}}_i'\boldsymbol{\Sigma}_{\mathbf{yy}}\tilde{\mathbf{k}}_i}$ or $\mathbf{k}_i = \boldsymbol{\Sigma}_{\mathbf{yy}}^{-1/2}\tilde{\mathbf{k}}_i$.

Assuming that $\lambda_1, \lambda_2, ..., \lambda_n$ are distinct, then
(a) $0 \leq \lambda_i < 1$ for $i = 1, 2, ..., n_1$ and $0 \leq \mu_j < 1$ for $j = 1, 2, ..., n_2$;
(b) $\lambda_i = \mu_i$ for $i = 1, 2, ..., n$;
(c) $\boldsymbol{\mathcal{K}}'\boldsymbol{\Sigma}_{\mathbf{yy}}\boldsymbol{\mathcal{K}} = \mathbf{I}_n$ and $\boldsymbol{\mathcal{A}}'\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\mathcal{A}} = \mathbf{I}_n$;
(d) $\boldsymbol{\mathcal{A}}'\boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\mathcal{K}} = \mathbf{R}$, where

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

We may interpret the canonical correlations as follows. The first canonical variates η_{1t} and ξ_{1t} can be interpreted as those linear combination of $\mathbf{y}_t$ and $\mathbf{x}_t$, respectively, such that the correlation between η_{1t} and ξ_{1t} is as large as possible. The variates η_{2t} and ξ_{2t} gives those linear combination of $\mathbf{y}_t$ and $\mathbf{x}_t$ that are uncorrelated with η_{1t} and ξ_{1t} and yield the largest remaining correlation between η_{2t} and ξ_{2t}, and so on.

1.2 Sample Canonical Correlations

The canonical correlations r_i calculated by the procedure just described are population parameters–they are functions of the population moments $\boldsymbol{\Sigma}_{\mathbf{yy}}, \boldsymbol{\Sigma}_{\mathbf{xy}}$, and $\boldsymbol{\Sigma}_{\mathbf{xx}}$. To find their sample analogs, all we have to do is to start from the sample moment of $\boldsymbol{\Sigma}_{\mathbf{yy}}, \boldsymbol{\Sigma}_{\mathbf{xy}}$, and $\boldsymbol{\Sigma}_{\mathbf{xx}}$.

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 moment are given by

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}} = (1/T)\sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t'$$
$$\hat{\boldsymbol{\Sigma}}_{\mathbf{yx}} = (1/T)\sum_{t=1}^{T} \mathbf{y}_t\mathbf{x}_t'$$
$$\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = (1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'.$$

Again, in many applications, $\mathbf{y}_t$ and $\mathbf{x}_t$ would be measured in deviations from their sample means. Then all the sample canonical correlations can be calculated from $\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}}$, $\hat{\boldsymbol{\Sigma}}_{\mathbf{yx}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}$ as the procedures described in Theorem 1.

2 Maximum Likelihood Estimation of a Gaussian VAR for Cointegration and the Test for Cointegration Rank

Consider a general VAR model [2] for the $k \times 1$ vector $\mathbf{y}_t$ with Gaussian error

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + ... + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t) &= \mathbf{0} \\ E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_s') &= \begin{cases} \mathbf{\Omega} & for\ t = s \\ \mathbf{0} & otherwise. \end{cases} \end{aligned}$$

We may rewrite (1) in the error correction form:

$$\triangle\mathbf{y}_t = \boldsymbol{\xi}_1\triangle\mathbf{y}_{t-1} + \boldsymbol{\xi}_2\triangle\mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1}\triangle\mathbf{y}_{t-p+1} + \mathbf{c} + \boldsymbol{\xi}_0\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{2}$$

where

$$\boldsymbol{\xi}_0 \equiv -(\mathbf{I} - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - ... - \mathbf{\Phi}_p) = -\mathbf{\Phi}(1).$$

Suppose that $\mathbf{y}_t$ is $I(1)$ with h cointegrating relationship which implies that

$$\boldsymbol{\xi}_0 = -\mathbf{B}\mathbf{A}' \tag{3}$$

for $\mathbf{B}$ and $\mathbf{A}$ an $(k \times h)$ matrix. That is, under the hypothesis of h cointegrating relations, only h separate linear combination of the level of $\mathbf{y}_{t-1}$ appears in (2).

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

$$\begin{aligned} &\mathcal{L}(\mathbf{\Omega}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1}, \mathbf{c}, \boldsymbol{\xi}_0) = (-Tk/2)\log(2\pi) - (T/2)\log|\mathbf{\Omega}| - (1/2) \\ &\sum_{t=1}^{T}\big[(\triangle\mathbf{y}_t - \boldsymbol{\xi}_1\triangle\mathbf{y}_{t-1} - \boldsymbol{\xi}_2\triangle\mathbf{y}_{t-2} - ... - \boldsymbol{\xi}_{p-1}\triangle\mathbf{y}_{t-p+1} - \mathbf{c} - \boldsymbol{\xi}_0\mathbf{y}_{t-1})' \\ &\times\mathbf{\Omega}^{-1}(\triangle\mathbf{y}_t - \boldsymbol{\xi}_1\triangle\mathbf{y}_{t-1} - \boldsymbol{\xi}_2\triangle\mathbf{y}_{t-2} - ... - \boldsymbol{\xi}_{p-1}\triangle\mathbf{y}_{t-p+1} - \mathbf{c} - \boldsymbol{\xi}_0\mathbf{y}_{t-1})\big]. \end{aligned} \tag{4}$$

The goal is to chose $(\mathbf{\Omega}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1}, \mathbf{c}, \boldsymbol{\xi}_0)$ so as to maximize (4) subject to the constraint that $\boldsymbol{\xi}_0$ can be written in the form of (3).

[2] Here, $\mathbf{y}_t$ in this VAR model are not necessary $I(1)$ variates and are not necessary cointegrated.

2.1 Concentrated Log-likelihood Function

2.1.1 Concentrated Likelihood Function

We often encounter in practice the situation where the parameter vector $\boldsymbol{\theta}_0$ can be naturally partitioned into two sub-vectors $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ as $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0' \quad \boldsymbol{\beta}_0')'$. Let the likelihood function be $L(\boldsymbol{\alpha}\ \boldsymbol{\beta})$. The MLE is obtained by maximizing L simultaneously for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$: i.e.

$$\frac{\partial \ln L}{\partial \boldsymbol{\alpha}} = \mathbf{0}; \tag{5}$$

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \mathbf{0}. \tag{6}$$

However, sometimes it is easier to maximize L in two step. First, maximize it with respect to $\boldsymbol{\beta}$ by taking $\boldsymbol{\alpha}$ as given, insert the maximizing value of $\boldsymbol{\beta}$ back into L; second, maximize L with respect to $\boldsymbol{\alpha}$. More precisely, define

$$L^*(\boldsymbol{\alpha}) = L[\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})], \tag{7}$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is defined as the solution to

$$\left.\frac{\partial \ln L}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = \mathbf{0}, \tag{8}$$

and define $\hat{\boldsymbol{\alpha}}^*$ as the solution to

$$\left.\frac{\partial \ln L^*}{\partial \boldsymbol{\alpha}}\right|_{\hat{\boldsymbol{\alpha}}^*} = \mathbf{0}. \tag{9}$$

We call $L^*(\boldsymbol{\alpha})$ **the concentrated likelihood function of $\boldsymbol{\alpha}$**. It is able to show that the MLE of $\boldsymbol{\alpha}$ from (5) and (6) simultaneously $\hat{\boldsymbol{\alpha}}$, and from concentrated likelihood (9), $\hat{\boldsymbol{\alpha}}^*$, are identical and have the same limiting distribution.

2.1.2 Calculate Auxiliary Regressions

The first step involve *concentrating* the likelihood function. This means take $\boldsymbol{\Omega}$ and $\boldsymbol{\xi}_0$ as given and maximization (4) with respect to $(\mathbf{c}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1})$. This restricted maximization problem take the form of seemingly unrelated regression of the elements of the $(k \times 1)$ vector $\triangle \mathbf{y}_t - \boldsymbol{\xi}_0 \mathbf{y}_{t-1}$ on a constant and the explanatory variables $(\triangle \mathbf{y}_{t-1}, \triangle \mathbf{y}_{t-2}, ..., \triangle \mathbf{y}_{t-p+1})$. Since each of the k regressions in this system has the identical explanatory variables, the estimates of $(\mathbf{c}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1})$ would come from *OLS* regression of each regressions of each elements of $\triangle \mathbf{y}_t - \boldsymbol{\xi}_0 \mathbf{y}_{t-1}$ on a constant and $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1})$. Denote the value of $(\mathbf{c}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1})$ that maximize (4) for a given value of $\boldsymbol{\xi}_0$ (and $\boldsymbol{\Omega}$, although it doesn't matter from the properties of $SURE$ model) by

$$\left[\hat{\mathbf{c}}^*(\boldsymbol{\xi}_0), \hat{\boldsymbol{\xi}}_1^*(\boldsymbol{\xi}_0), \hat{\boldsymbol{\xi}}_2^*(\boldsymbol{\xi}_0), ..., \hat{\boldsymbol{\xi}}_{p-1}^*(\boldsymbol{\xi}_0)\right].$$

These values are characterized by the condition that the following residual vector must have sample mean zero and be orthogonal to $(\triangle\mathbf{y}_{t-1}, \triangle\mathbf{y}_{t-2}, ..., \triangle\mathbf{y}_{t-p+1})$:

$$[\triangle\mathbf{y}_t - \boldsymbol{\xi}_0\mathbf{y}_{t-1}] - \left\{\hat{\mathbf{c}}^*(\boldsymbol{\xi}_0) + \hat{\boldsymbol{\xi}}_1^*(\boldsymbol{\xi}_0)\triangle\mathbf{y}_{t-1} + \hat{\boldsymbol{\xi}}_2^*(\boldsymbol{\xi}_0)\triangle\mathbf{y}_{t-2} + ... + \hat{\boldsymbol{\xi}}_{p-1}^*(\boldsymbol{\xi}_0)\triangle\mathbf{y}_{t-p+1}\right\}. \quad (10)$$

To obtain (10) with unknown $\boldsymbol{\xi}_0$ (although we assume it is known at this stage to form concentrated log-likelihood function), we may form two *auxiliary* regressions and estimate them by OLS to get

$$\triangle\mathbf{y}_t = \hat{\boldsymbol{\pi}}_0 + \hat{\boldsymbol{\Pi}}_1\triangle\mathbf{y}_{t-1} + \hat{\boldsymbol{\Pi}}_2\triangle\mathbf{y}_{t-2} + ... + \hat{\boldsymbol{\Pi}}_{p-1}\triangle\mathbf{y}_{t-p+1} + \hat{\mathbf{u}}_t \quad (11)$$

and

$$\mathbf{y}_{t-1} = \hat{\boldsymbol{\theta}}_0 + \hat{\boldsymbol{\Theta}}_1\triangle\mathbf{y}_{t-1} + \hat{\boldsymbol{\Theta}}_2\triangle\mathbf{y}_{t-2} + ... + \hat{\boldsymbol{\Theta}}_{p-1}\triangle\mathbf{y}_{t-p+1} + \hat{\mathbf{v}}_t, \quad (12)$$

where both the residual vector $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{u}}_t$ have sample mean zero and be orthogonal to $(\triangle\mathbf{y}_{t-1}, \triangle\mathbf{y}_{t-2}, ..., \triangle\mathbf{y}_{t-p+1})$ also. Moreover, $\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{u}}_t$ also have sample mean zero and is orthogonal to $(\triangle\mathbf{y}_{t-1}, \triangle\mathbf{y}_{t-2}, ..., \triangle\mathbf{y}_{t-p+1})$. Therefore, the residual vector (10) can be expressed by

$$\begin{aligned} \hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t &= (\triangle\mathbf{y}_t - \hat{\boldsymbol{\pi}}_0 - \hat{\boldsymbol{\Pi}}_1\triangle\mathbf{y}_{t-1} - \hat{\boldsymbol{\Pi}}_2\triangle\mathbf{y}_{t-2} - ..... - \hat{\boldsymbol{\Pi}}_{p-1}\triangle\mathbf{y}_{t-p+1}) \\ &\quad -\boldsymbol{\xi}_0(\mathbf{y}_{t-1} - \hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\Theta}}_1\triangle\mathbf{y}_{t-1} - \hat{\boldsymbol{\Theta}}_2\triangle\mathbf{y}_{t-2} - ..... - \hat{\boldsymbol{\Theta}}_{p-1}\triangle\mathbf{y}_{t-p+1}) \end{aligned}$$

with

$$\begin{aligned} \hat{\mathbf{c}}^*(\boldsymbol{\xi}_0) &= \hat{\boldsymbol{\pi}}_0 - \boldsymbol{\xi}_0\hat{\boldsymbol{\theta}}_0 \\ \hat{\boldsymbol{\xi}}_i^*(\boldsymbol{\xi}_0) &= \hat{\boldsymbol{\Pi}}_i - \boldsymbol{\xi}_0\hat{\boldsymbol{\Theta}}_i, \quad for\ i = 1, 2, ..., p-1. \end{aligned}$$

The concentrated log likelihood function is found by replacing $(\mathbf{c}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{p-1})$ with $(\hat{\mathbf{c}}^*(\boldsymbol{\xi}_0), \hat{\boldsymbol{\xi}}_1^*(\boldsymbol{\xi}_0), \hat{\boldsymbol{\xi}}_2^*(\boldsymbol{\xi}_0), ..., \hat{\boldsymbol{\xi}}_{p-1}^*(\boldsymbol{\xi}_0))$ in (4):

$$\mathbb{L}(\boldsymbol{\Omega}, \boldsymbol{\xi}_0) = (-Tk/2)\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}| - (1/2)\sum_{t=1}^{T}\left[(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)'\boldsymbol{\Omega}^{-1}(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)\right]. \quad (13)$$

We can go one step further to concentrate out $\boldsymbol{\Omega}$. Recall from the analysis of estimation of VAR on p.17 of Chapter 18 that the value of $\boldsymbol{\Omega}$ that maximize (13) (for a given $\boldsymbol{\xi}_0$) is given by

$$\hat{\boldsymbol{\Omega}}^*(\boldsymbol{\xi}_0) = 1/T\sum_{t=1}^{T}\left[(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)'\right]. \quad (14)$$

As in expression (24) of Chapter 18, the value obtained for (13) when evaluated at (14) is then

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\xi}_0) &\equiv \mathbb{L}(\hat{\boldsymbol{\Omega}}^*(\boldsymbol{\xi}_0), \boldsymbol{\xi}_0) \\
&= (-Tk/2)\log(2\pi) - (T/2)\log|\hat{\boldsymbol{\Omega}}^*(\boldsymbol{\xi}_0)| - (Tk/2) \\
&= (-Tk/2)\log(2\pi) - (Tk/2) - (T/2)\log\left|(1/T)\sum_{t=1}^{T}[(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)']\right| \\
&= K_0 - (T/2)\log\left|(1/T)\sum_{t=1}^{T}[(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)']\right|. \qquad (15)
\end{aligned}$$

(15) represents the biggest value one can achieve for the log likelihood for any given value of $\boldsymbol{\xi}_0$. Maximizing the log-likelihood function thus comes down to choosing $\boldsymbol{\xi}_0$ so as to minimize

$$\left|(1/T)\sum_{t=1}^{T}[(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)']\right|. \qquad (16)$$

2.2 Canonical Correlations Analysis of MLE

Next, we compute the second-moment matrices of all of these residuals and their cross-products, $\mathbf{S}_{\mathbf{uu}}$, $\mathbf{S}_{\mathbf{uv}}$, $\mathbf{S}_{\mathbf{vu}}$, and $\mathbf{S}_{\mathbf{vv}}$, where

$$\mathbf{S}_{\mathbf{uu}} = T^{-1}\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t'$$
$$\mathbf{S}_{\mathbf{uv}} = T^{-1}\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{v}}_t'$$
$$\mathbf{S}_{\mathbf{vv}} = T^{-1}\sum_{t=1}^{T}\hat{\mathbf{v}}_t\hat{\mathbf{v}}_t'$$

Consequently, from (15) we have the loglikelihood function now become

$$\mathrm{L}(\boldsymbol{\xi}_0) = K_0 - (T/2)\log|\mathbf{S}_{\mathbf{uu}} - \boldsymbol{\xi}_0\mathbf{S}_{\mathbf{vu}} - \mathbf{S}_{\mathbf{uv}}\boldsymbol{\xi}_0' + \boldsymbol{\xi}_0\mathbf{S}_{\mathbf{vv}}\boldsymbol{\xi}_0'|. \qquad (17)$$

If $\boldsymbol{\xi}_0$ were unrestricted, a conventional regression estimator would result. However, we are interested in the class of solution that result from the imposition of the restriction that

$$\boldsymbol{\xi}_0 = \mathbf{B}^*\mathbf{A}'.$$

Hence,[3] from (17),

$$\mathrm{L}^*(\mathbf{B}^*,\mathbf{A}) = K_0 - (T/2)\log\left|\mathbf{S_{uu}} - \mathbf{B}^*\mathbf{A}'\mathbf{S_{vu}} - \mathbf{S_{uv}}\mathbf{A}\mathbf{B}^{*\prime} + \mathbf{B}^*\mathbf{A}'\mathbf{S_{vv}}\mathbf{B}^*\mathbf{A}'\right|. \quad (18)$$

Next, concentrate $\mathrm{L}^*(\mathbf{B}^*,\mathbf{A})$ with respect to $\mathbf{B}^*$, which will deliver an expression for the MLE of $\mathbf{B}^*$ as a function of $\mathbf{A}$, and yields a further concentrated likelihood function which depends only on $\mathbf{A}$. Once the MLE of $\mathbf{A}$ is obtained, we can solve backwards for estimator of all the other unknown parameters as functions of the MLE of $\mathbf{A}$. Thus, from (18), the FOC is

$$\frac{\partial \mathrm{L}^*(\mathbf{B}^*,\mathbf{A})}{\partial \mathbf{B}^*} = \mathbf{0}, \qquad (how?)$$

which implies

$$\hat{\mathbf{B}}^* = \mathbf{S_{uv}}\mathbf{A}(\mathbf{A}'\mathbf{S_{vv}}\mathbf{A})^{-1}. \quad (19)$$

Substituting $\hat{\mathbf{B}}^*$ into (21) yields $\mathrm{L}^{**}(\mathbf{A})$:

$$\mathrm{L}^{**}(\mathbf{A}) = K_1 - (T/2)\log\left|\mathbf{S_{uu}} - \mathbf{S_{uv}}\mathbf{A}(\mathbf{A}'\mathbf{S_{vv}}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S_{vu}}\right|. \quad (20)$$

At first sight, differentiating $\mathrm{L}^{**}(\mathbf{A})$ with respect to $\mathbf{A}$ looks formidable. But we can solve the problem by applying partitioned inversion results [4] to (20) and obtain

$$\begin{aligned}\left|\mathbf{S_{uu}} - \mathbf{S_{uv}}\mathbf{A}(\mathbf{A}'\mathbf{S_{vv}}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S_{vu}}\right| &= \left|\mathbf{A}'\mathbf{S_{vv}}\mathbf{A}\right|^{-1}\left|\mathbf{S_{uu}}\right|\left|\mathbf{A}'\mathbf{S_{vv}}\mathbf{A} - \mathbf{A}'\mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}}\mathbf{A}\right| \\ &= \left|\mathbf{A}'\mathbf{S_{vv}}\mathbf{A}\right|^{-1}\left|\mathbf{S_{uu}}\right|\left|\mathbf{A}'(\mathbf{S_{vv}} - \mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}})\mathbf{A}\right|.\end{aligned}$$

Then maximizing $\mathrm{L}^{**}(\mathbf{A})$ with respect to $\mathbf{A}$ corresponds to minimizing the generalized variances ratio,

$$\left|\mathbf{A}'(\mathbf{S_{vv}} - \mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}})\mathbf{A}\right| / \left|\mathbf{A}'\mathbf{S_{vv}}\mathbf{A}\right|, \quad (21)$$

noting that $|\mathbf{S_{uu}}|$ is a constant.[5]

[3] Here, for neatness of computation, we let $\mathbf{B}^* = -\mathbf{B}$.

[4] Using the results from p.15 of Chapter 1 that

$$\begin{vmatrix}\mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}}\end{vmatrix} \begin{aligned} &= |\mathbf{A_{22}}|\cdot|\mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}^{-1}}\mathbf{A_{21}}| \\ &= |\mathbf{A_{11}}|\cdot|\mathbf{A_{22}} - \mathbf{A_{21}}\mathbf{A_{11}^{-1}}\mathbf{A_{12}}|.\end{aligned}$$

.

[5] By a standard result from the theory of canonical correlations, an expression of the form

$$\left|\boldsymbol{\zeta}'(\mathbf{M}_1 - \mathbf{M}_2)\boldsymbol{\zeta}\right|\left|\boldsymbol{\zeta}'\mathbf{M}_1\boldsymbol{\zeta}\right|^{-1}$$

can be minimized by solving the equation $|\lambda\mathbf{M}_1 - \mathbf{M}_2| = 0$. See Banerjee, p.265 or Johansen 1996, p.92.

This ratio is minimized by the choice of $\hat{\mathbf{A}}_{k\times h} = (\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, ..., \hat{\mathbf{a}}_h)$, where $\hat{\boldsymbol{\mathcal{A}}} = (\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, ..., \hat{\mathbf{a}}_k)$ are the eigenvector of the equation

$$|\lambda \mathbf{S_{vv}} - \mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}}| = 0, \tag{22}$$

or

$$\mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \mathbf{a}_i = \lambda_i \mathbf{S_{vv}} \hat{\mathbf{a}}_i, \tag{23}$$

normalized by

$$\hat{\mathbf{a}}_i' \mathbf{S_{vv}} \hat{\mathbf{a}}_j = \begin{cases} 1 & i = j \\ 0 & otherwise \end{cases} \quad \forall i = 1, 2, ..., k,$$

i.e.

$$\begin{aligned} \hat{\boldsymbol{\mathcal{A}}}' \mathbf{S_{vv}} \hat{\boldsymbol{\mathcal{A}}} &= \mathbf{I}_k, & (24) \\ \hat{\mathbf{A}}' \mathbf{S_{vv}} \hat{\mathbf{A}} &= \mathbf{I}_h. & (25) \end{aligned}$$

Since $\mathbf{S_{vv}}$ is symmetric and positive definite matrix, from the eigenvalues equation (23) we have[6]

$$\mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\boldsymbol{\mathcal{A}}} = \mathbf{S_{vv}} \hat{\boldsymbol{\mathcal{A}}} \boldsymbol{\Lambda}_k \tag{26}$$

and

$$\mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\mathbf{A}} = \mathbf{S_{vv}} \hat{\mathbf{A}} \boldsymbol{\Lambda}_h, \tag{27}$$

this implies

$$\hat{\mathbf{A}}' \mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\mathbf{A}} = \hat{\mathbf{A}}' \mathbf{S_{vv}} \hat{\mathbf{A}} \boldsymbol{\Lambda}_h = \boldsymbol{\Lambda}_h, \tag{28}$$

[6] To see this,

$$\begin{aligned} \mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\boldsymbol{\mathcal{A}}} &= \mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, ..., \hat{\mathbf{a}}_k] \\ &= [\mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\mathbf{a}}_1,\ \mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\mathbf{a}}_2, ..., \mathbf{S_{vu}} \mathbf{S_{uu}^{-1}} \mathbf{S_{uv}} \hat{\mathbf{a}}_k] \\ &= [\lambda_1 \mathbf{S_{vv}} \hat{\mathbf{a}}_1, \lambda_2 \mathbf{S_{vv}} \hat{\mathbf{a}}_2, ..., \lambda_k \mathbf{S_{vv}} \hat{\mathbf{a}}_k] \end{aligned}$$

and

$$\begin{aligned} \mathbf{S_{vv}} \hat{\boldsymbol{\mathcal{A}}} \boldsymbol{\Lambda}_k &= [\mathbf{S_{vv}} \hat{\boldsymbol{\mathcal{A}}} \Lambda_1,\ \mathbf{S_{vv}} \hat{\boldsymbol{\mathcal{A}}} \Lambda_2, ..., \mathbf{S_{vv}} \hat{\boldsymbol{\mathcal{A}}} \Lambda_k] \\ &= [\lambda_1 \mathbf{S_{vv}} \hat{\mathbf{a}}_1, \lambda_2 \mathbf{S_{vv}} \hat{\mathbf{a}}_2, ..., \lambda_k \mathbf{S_{vv}} \hat{\mathbf{a}}_k]. \end{aligned}$$

where

$$\mathbf{\Lambda}_k = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \lambda_k \end{bmatrix} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & . & . & .\Lambda_k \end{bmatrix}$$

with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k \geq 0$.

We see $\hat{\mathbf{A}}$ indeed minimize the generalized variances ratio (21) by substituting (25) and (28) into (21) to get

$$\begin{aligned}
\left|\hat{\mathbf{A}}'(\mathbf{S_{vv}} - \mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}})\hat{\mathbf{A}}\right| / \left|\hat{\mathbf{A}}'\mathbf{S_{vv}}\hat{\mathbf{A}}\right| &= \left|\hat{\mathbf{A}}'(\mathbf{S_{vv}} - \mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}})\hat{\mathbf{A}}\right| \\
&= \left|\hat{\mathbf{A}}'\mathbf{S_{vv}}\hat{\mathbf{A}} - \hat{\mathbf{A}}'\mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}}\hat{\mathbf{A}}\right| \\
&= |\mathbf{I}_h - \mathbf{\Lambda}_h| \\
&= \begin{bmatrix} 1-\lambda_1 & 0 & . & . & . & 0 \\ 0 & 1-\lambda_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1-\lambda_h \end{bmatrix} \\
&= \begin{bmatrix} 1-\hat{r}_1^2 & 0 & . & . & . & 0 \\ 0 & 1-\hat{r}_2^2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1-\hat{r}_h^2 \end{bmatrix}.
\end{aligned} \tag{29}$$

Therefore, $\hat{\mathbf{A}}$ have chosen the highest h canonical correlations among the k element of $\mathbf{u}_t$ and $\mathbf{v}_t$ so that the $|\mathbf{A}'(\mathbf{S_{vv}} - \mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}})\mathbf{A}| / |\mathbf{A}'\mathbf{S_{vv}}\mathbf{A}|$ is minimization.

Substituting (29) back into (20), it follows that the maximum value achieved for the likelihood function is given by

$$\begin{aligned}
\mathrm{L}^{**}(\mathbf{A}) &= K_1 - (T/2)\log|\mathbf{S_{uu}}| - (T/2)\log\left[\prod_{i=1}^{h}(1-\lambda_i)\right] &(30)\\
&= K_1 - (T/2)\log|\mathbf{S_{uu}}| - (T/2)\sum_{i=1}^{h}\log(1-\lambda_i). &(31)
\end{aligned}$$

2.3 Likelihood Ratio Test for Cointegration Rank

Denote by H_h the hypothesis that there are the rank of $\boldsymbol{\xi}_0$ is h in the system (1). When $\boldsymbol{\xi}_0$ is unrestricted, all k eigenvalue are retained and the unrestricted maximum of the likelihood function is given by

$$\mathrm{L}^{**}(\boldsymbol{\mathcal{A}}) \;=\; K_1 - (T/2)\log|\mathbf{S_{uu}}| - (T/2)\sum_{i=1}^{k}\log(1-\lambda_i). \qquad (32)$$

The likelihood ratio test statistic for the hypothesis H_h in H_k ($\mathbf{y}_t$ is stationary) can be based on twice the difference the log-likelihood in (31) and that in (32); that is,

$$\eta_h = -2\log(H_h|H_k) = -T\sum_{i=h+1}^{k}\log(1-\lambda_i). \qquad (33)$$

Similarly the likelihood ratio test statistic for testing H_h in H_{h+1} is given by

$$\kappa_h = -2\log(H_h|H_{h+1}) = -T\log(1-\lambda_{h+1}). \qquad (34)$$

Both η_h and κ_h have non-standard distributions which are functionals of multivariate Brownian motion.

Theorem 1:
The statistics η_h $(= -2\log(H_h|H_k))$ has a limiting distribution which, if $\mathbf{B}'_{\perp}\mathbf{c} \neq \mathbf{0}$, can be expressed in terms of a $(k-h)$-dimensional Wiener process $\boldsymbol{w}$ with *i.i.d.* components as

$$tr\left\{\int(d\boldsymbol{w})\boldsymbol{f}'\left[\int \boldsymbol{f}\boldsymbol{f}'du\right]^{-1}\int \boldsymbol{f}(d\boldsymbol{w})'\right\}, \qquad (35)$$

where $\boldsymbol{f}' = (\boldsymbol{f}'_1, F'_2)$, and

$$\boldsymbol{f}_{1i}(t) = W_i(t) - \int W_i(u)du \qquad i = 1, 2, ..., k-h-1$$

and

$$F_2(t) = t - \frac{1}{2}.$$

The statistics $\kappa_h (= -2 \log(H_h | H_{h+1}))$ is asymptotically distributed as the maximum eigenvalues of the matrix in (35). The asymptotic distribution of η_h is given is that of the variables in the **case 3** section of Table B.10, while the asymptotic distribution of κ_h is given in the **case 3** of Table B.11 of Hamilton (1994).

If $\mathbf{B}'_{\perp}\mathbf{c} = \mathbf{0}$ and assumes that no restriction are imposed on the constant term in the estimation of auxiliary regression (11) and (12), then the asymptotic distribution of η_h and κ_h are given as the trace and the maximum eigenvalues respectively of the matrix in (35) with $\boldsymbol{f}(t) = \boldsymbol{w}(t) - \int \boldsymbol{w}(u)du$. The asymptotic distribution of η_h is given in the **case 2** section of Table B.10, while the asymptotic distribution of κ_h is given in the **case 2** panel of Table B.11.

If $\mathbf{c}$ is not included in the estimated model (1), then the asymptotic distribution of η_h and κ_h are given as the trace and the maximum eigenvalues respectively of the matrix in (35) with $\boldsymbol{f}(t) = \boldsymbol{w}(t)$ and the asymptotic distribution of η_h is given in the **case 1** section of Table B.10, while the asymptotic distribution of κ_h is given in the **case 1** panel of Table B.11.

2.4 MLE estimation of Parameters

Given the cointegration rank (h) have been inferenced from the likelihood ratio test discussed above, the parameters are estimated from

$$\hat{\mathbf{B}} = -\hat{\mathbf{B}}^* = -\mathbf{S_{uv}}\hat{\mathbf{A}}(\hat{\mathbf{A}}'\mathbf{S_{vv}}\hat{\mathbf{A}})^{-1} = -\mathbf{S_{uv}}\hat{\mathbf{A}}. \tag{36}$$

and therefore

$$\hat{\boldsymbol{\xi}}_0 = -\hat{\mathbf{B}}\hat{\mathbf{A}}' = \mathbf{S_{uv}}\hat{\mathbf{A}}\hat{\mathbf{A}}'. \tag{37}$$

The other parameters then is obtained as following

$$\begin{aligned}
\hat{\mathbf{c}}^*(\hat{\boldsymbol{\xi}}_0) &= \hat{\boldsymbol{\pi}}_0 - \hat{\boldsymbol{\xi}}_0\hat{\boldsymbol{\theta}}_0, && (38)\\
\hat{\boldsymbol{\xi}}_i^*(\hat{\boldsymbol{\xi}}_0) &= \hat{\boldsymbol{\Pi}}_i - \hat{\boldsymbol{\xi}}_0\hat{\boldsymbol{\Theta}}_i, \quad for\ i = 1, 2, ..., p-1, && (39)\\
\hat{\boldsymbol{\Omega}}^*(\hat{\boldsymbol{\xi}}_0) &= 1/T\sum_{t=1}^{T}\left[(\hat{\mathbf{u}}_t - \hat{\boldsymbol{\xi}}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \hat{\boldsymbol{\xi}}_0\hat{\mathbf{v}}_t)'\right]. && (40)
\end{aligned}$$

Johansen (1988) prove that these estimator are consistent.

Example:

See the example on p.647 of Hamilton.

3 Extension

3.1 Linear Restriction on Cointegrating Vector

A different set of generalization concerns testing linear restriction on $\mathbf{A}$ and $\mathbf{B}$. These would correspond to investigating *a priori* theories about the cointegrating vectors, and about their role in different equations. Consider a system of k variables that is assumed to be characterized h cointegrating relations. We might then want to test a restriction on theses cointegrating vector, such as only q of the variables are involved in the cointegration relations. For example, we might be interested in whether the middle coefficient in the PPP theory is zero, that is, in whether the cointegrating relation involved solely the price levels between two countries. For this case, $h = 1$, $q = 2$, and $k = 3$.

Consider testing linear restriction on $\mathbf{A}$ of the form

$$H_q : \mathbf{A} = \mathbf{DE}, \tag{41}$$

where $\mathbf{D}$ is a known $k \times q$ matrix and $\mathbf{E}$ is an $q \times h$ matrix of unknown parameters and $h \leq q < k$.

The error correction term in (1) will takes the form

$$\boldsymbol{\xi}_0 \mathbf{y}_t = -\mathbf{B}\mathbf{A}'\mathbf{y}_t = -\mathbf{B}\mathbf{E}'\mathbf{D}'\mathbf{y}_t.$$

This is equivalently to say that the cointegrating relations are restricted to involve only $\mathbf{D}'\mathbf{y}_t$. For the preceding example,

$$\mathbf{D}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

such that

$$\begin{aligned} \boldsymbol{\xi}_0 \mathbf{y}_t &= -\mathbf{B}\mathbf{A}'\mathbf{y}_t = -\mathbf{B}\mathbf{E}'\mathbf{D}'\mathbf{y}_t \\ &= -\mathbf{B}\mathbf{E}' \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_t \\ s_t \\ p_t^* \end{bmatrix} = -\mathbf{B}\mathbf{E}' \begin{bmatrix} p_t \\ p_t^* \end{bmatrix}, \end{aligned}$$

only p_t and p_t^* enter the cointegrating relations.

Maximum likelihood estimation ($\mathbf{E}$) is unaltered until equation (21), which become

$$\begin{aligned} &\left|\mathbf{E}'\mathbf{D}'(\mathbf{S_{vv}} - \mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}})\mathbf{DE}\right| / |\mathbf{E}'\mathbf{D}'\mathbf{S_{vv}}\mathbf{DE}| \\ = &\left|\mathbf{E}'(\mathbf{D}'\mathbf{S_{vv}}\mathbf{D} - \mathbf{D}'\mathbf{S_{vu}}\mathbf{S_{uu}^{-1}}\mathbf{S_{uv}}\mathbf{D})\mathbf{E}\right| / |\mathbf{E}'(\mathbf{D}'\mathbf{S_{vv}}\mathbf{D})\mathbf{E}| \\ = &\left|\mathbf{E}'(\tilde{\mathbf{S}}_{vv} - \tilde{\mathbf{S}}_{vu}\mathbf{S_{uu}^{-1}}\tilde{\mathbf{S}}_{uv})\mathbf{E}\right| / \left|\mathbf{E}'\tilde{\mathbf{S}}_{vv}\mathbf{E}\right| \end{aligned}$$

where $\tilde{\mathbf{S}}_{vv} = \mathbf{D}'\mathbf{S_{vv}}\mathbf{D}$ and $\tilde{\mathbf{S}}_{vu} = \mathbf{D}'\mathbf{S_{vu}}$.

The minimization is by the choice of $\hat{\mathbf{E}}_{q\times h} = (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, ..., \hat{\mathbf{e}}_h)$, where $\hat{\boldsymbol{\mathcal{E}}} = (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, ..., \hat{\mathbf{e}}_q)$ are the eigenvector of the equation

$$\tilde{\mathbf{S}}_{vv}^{-1}\tilde{\mathbf{S}}_{vu}\mathbf{S_{uu}^{-1}}\tilde{\mathbf{S}}_{uv}\mathbf{e}_i = \lambda_i^*\mathbf{e}_i, \tag{42}$$

normalized by

$$\mathbf{e}_i'\tilde{\mathbf{S}}_{vv}\mathbf{e}_j = \begin{cases} 1 & i=j \\ 0 & otherwise \end{cases} \quad \forall i = 1, 2, ..., q.$$

A likelihood-ratio test against the unrestricted value of $\mathbf{A}$ can be calculated and amounts to testing H_q within H_h, and is therefore based on

$$\eta_{qh} = T\sum_{i=1}^{h} \log[(1-\lambda_i^*)/(1-\lambda_i)]. \tag{43}$$

In this case, the null hypothesis involves only coefficients on $I(0)$ variables, and standard asymptotic distribution theory turns out to apply. Johansen (1988, 1991) showed that the likelihood ratio statistic (43) has an asymptotic χ^2 distribution with $h(k-p)$ degree of freedom.

Example;

See the example on p.649 of Hamilton.

3.2 MLE in the Absence of Deterministic Time trend

The preceding analysis assumed that $\mathbf{c}$, the $(k \times 1)$ vector of constant terms in the VAR, was unrestricted. The value of $\mathbf{c}$ contributes h constant terms for the cointegrating relations,[7] along with $k-h$ deterministic trends that are common

[7]See Johansen's Granger Representation Theorem on the dimension of $E(\mathbf{A}'\mathbf{y}_t)$.

to each of the k elements of $\mathbf{y}_t$.[8] In some applications it might be of interest to allow constant terms in the cointegrating relations but to rule out deterministic time trend for any of the variables. This would require that $\mathbf{B}'_{\perp}\mathbf{c} = \mathbf{0}$ or[9]

$$\mathbf{c} = \mathbf{B}\boldsymbol{\mu}_1^*,$$

where $\boldsymbol{\mu}_1^*$ is an $(h \times 1)$ vector corresponding to the unconditional mean of $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$. Thus, for this restricted case, we want to estimate only the h elements of $\boldsymbol{\mu}_1^*$ rather than all k elements of $\mathbf{c}$.

3.2.1 Calculate Auxiliary Regressions

To maximize the likelihood function subject to the restrictions that there are h cointegrating relations and no deterministic time trends in any of the series, Johansen's (1991) first step was to concentrated out $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ...,$ and $\boldsymbol{\xi}_{p-1}$ (but not $\mathbf{c}$) For given $\mathbf{c}$ and $\boldsymbol{\xi}_0$, this is achieved by *OLS* regression of $\triangle\mathbf{y}_t - \mathbf{c} - \boldsymbol{\xi}_0\mathbf{y}_{t-1}$ on the explanatory variables $(\triangle\mathbf{y}_{t-1}, \triangle\mathbf{y}_{t-2}, ..., \triangle\mathbf{y}_{t-p+1})$. The residuals from three separated regressions:

$$\triangle\mathbf{y}_t = \tilde{\mathbf{\Pi}}_1\triangle\mathbf{y}_{t-1} + \tilde{\mathbf{\Pi}}_2\triangle\mathbf{y}_{t-2} + ... + \tilde{\mathbf{\Pi}}_{p-1}\triangle\mathbf{y}_{t-p+1} + \tilde{\mathbf{u}}_t,$$

$$\mathbf{y}_{t-1} = \tilde{\mathbf{\Theta}}_1\triangle\mathbf{y}_{t-1} + \tilde{\mathbf{\Theta}}_2\triangle\mathbf{y}_{t-2} + ... + \tilde{\mathbf{\Theta}}_{p-1}\triangle\mathbf{y}_{t-p+1} + \tilde{\mathbf{v}}_t,$$

and

$$1 = \tilde{\boldsymbol{w}}'_1\triangle\mathbf{y}_{t-1} + \tilde{\boldsymbol{w}}'_2\triangle\mathbf{y}_{t-2} + ... + \tilde{\boldsymbol{w}}'_{p-1}\triangle\mathbf{y}_{t-p+1} + \tilde{w}_t,$$

The concentrated log likelihood function is then

$$\begin{aligned} \mathbb{L}(\mathbf{\Omega}, \mathbf{c}, \boldsymbol{\xi}_0) &= (-Tk/2)\log(2\pi) - (T/2)\log|\mathbf{\Omega}| \\ &\quad -(1/2)\sum_{t=1}^{T}\left[(\tilde{\mathbf{u}}_t - \mathbf{c}\tilde{w}_t - \boldsymbol{\xi}_0\tilde{\mathbf{v}}_t)'\mathbf{\Omega}^{-1}(\tilde{\mathbf{u}}_t - \mathbf{c}\tilde{w}_t - \boldsymbol{\xi}_0\tilde{\mathbf{v}}_t)\right]. \end{aligned} \tag{44}$$

Further concentrating out $\mathbf{\Omega}$ results in

$$\begin{aligned} \mathrm{L}(\mathbf{c}, \boldsymbol{\xi}_0) &= (-Tk/2)\log(2\pi) - (Tk/2) \\ &\quad -(T/2)\log\left|(1/T)\sum_{t=1}^{T}[(\tilde{\mathbf{u}}_t - \mathbf{c}\tilde{w}_t - \boldsymbol{\xi}_0\tilde{\mathbf{v}}_t)(\tilde{\mathbf{u}}_t - \mathbf{c}\tilde{w}_t - \boldsymbol{\xi}_0\hat{\mathbf{v}}_t)']\right| \end{aligned} \tag{45}$$

[8]See Johansen's Granger Representation Theorem on the dimension of $E(\triangle\mathbf{y}_t)$.

[9]This would satisfy $\mathbf{B}'_{\perp}\mathbf{c} = \mathbf{0}$.

Imposing the constraints $\mathbf{c} = \mathbf{B}\boldsymbol{\mu}_1^*$ and $\boldsymbol{\xi}_0 = -\mathbf{B}\mathbf{A}' = \mathbf{B}^*\mathbf{A}'$, the magnitude in (45) can be written as

$$\begin{aligned} \mathrm{L}(\mathbf{c}, \boldsymbol{\xi}_0) &= (-Tk/2)\log(2\pi) - (Tk/2) \\ &\quad -(T/2)\log\left|(1/T)\sum_{t=1}^{T}\left[(\tilde{\mathbf{u}}_t - \mathbf{B}^*\tilde{\mathbf{A}}'\tilde{\mathbf{w}}_t)(\tilde{\mathbf{u}}_t - \mathbf{B}^*\tilde{\mathbf{A}}'\tilde{\mathbf{w}}_t)'\right]\right|, \quad (46) \end{aligned}$$

where

$$\begin{aligned} \tilde{\mathbf{w}}_t &= \begin{bmatrix} \tilde{w}_t \\ \tilde{\mathbf{v}}_t \end{bmatrix}_{(k+1)\times 1} \\ \tilde{\mathbf{A}}' &= \begin{bmatrix} -\boldsymbol{\mu}_1^* & \mathbf{A}' \end{bmatrix}_{h\times(k+1)}. \quad (47) \end{aligned}$$

By constructing

$$\tilde{\mathbf{S}}_{\mathbf{ww}} = T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{w}}_t\tilde{\mathbf{w}}_t',$$

$$\tilde{\mathbf{S}}_{\mathbf{uu}} = T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{u}}_t\tilde{\mathbf{v}}_t',$$

$$\tilde{\mathbf{S}}_{\mathbf{uw}} = T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{u}}_t\tilde{\mathbf{w}}_t',$$

(46) is an expression of exactly the same as (18) with $\mathbf{A}$ replaced by $\tilde{\mathbf{A}}$ and $\hat{\mathbf{v}}_t$ replaced by $\tilde{\mathbf{w}}_t$. Thus, the restricted log likelihood function is

$$\mathrm{L}^*(\mathbf{B}^*, \tilde{\mathbf{A}}) = K_0 - (T/2)\log\left|\tilde{\mathbf{S}}_{\mathbf{uu}} - \mathbf{B}^*\tilde{\mathbf{A}}'\tilde{\mathbf{S}}_{\mathbf{wu}} - \tilde{\mathbf{S}}_{\mathbf{uw}}\tilde{\mathbf{A}}\mathbf{B}^{*\prime} + \mathbf{B}^*\tilde{\mathbf{A}}'\tilde{\mathbf{S}}_{\mathbf{ww}}\mathbf{B}^*\tilde{\mathbf{A}}'\right|.$$

This ratio is minimized by the choice of $\tilde{\mathbf{A}}_{k\times h} = (\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, ..., \tilde{\mathbf{a}}_h)$, where $\tilde{\boldsymbol{\mathcal{A}}} = (\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, ..., \tilde{\mathbf{a}}_{k+1})$ are the eigenvector of the equation

$$\tilde{\mathbf{S}}_{\mathbf{wu}}\tilde{\mathbf{S}}_{\mathbf{uu}}^{-1}\tilde{\mathbf{S}}_{\mathbf{uw}}\tilde{\mathbf{a}}_i = \tilde{\lambda}_i\tilde{\mathbf{S}}_{\mathbf{ww}}\tilde{\mathbf{a}}_i,$$

normalized by

$$\tilde{\mathbf{a}}_i'\tilde{\mathbf{S}}_{\mathbf{ww}}\tilde{\mathbf{a}}_j = \begin{cases} 1 & i = j \\ 0 & otherwise \end{cases} \quad \forall i = 1, 2, ..., k+1.$$

The maximum value achieved for the log likelihood function subject to the constraints that there are h cointegrating relations and no deterministic trend is

$$\begin{aligned} \mathrm{L}^{**}(\mathbf{A}) &= K_1 - (T/2)\log|\mathbf{S}_{\mathbf{uu}}| - (T/2)\log\left[\prod_{i=1}^{h}(1-\tilde{\lambda}_i)\right] \quad (48) \\ &= K_1 - (T/2)\log|\mathbf{S}_{\mathbf{uu}}| - (T/2)\sum_{i=1}^{h}\log(1-\tilde{\lambda}_i). \quad (49) \end{aligned}$$

3.2.2 Likelihood Ratio Test for Cointegration Rank

Theorem 2:
Under the hypothesis $H_h^* : \boldsymbol{\xi}_0 = \mathbf{BA}'$ *and* $\mathbf{c} = \mathbf{B}\boldsymbol{\mu}_1^*$, the likelihood ratio test statistics $-2\log(H_h^*|H_k)$ and $-2\log(H_h^*|H_{h+1}^*)$ are distributed as the trace and maximal eigenvalues respectively of the matrix in (38), with $\boldsymbol{f} = (\boldsymbol{w}(t)', 1)'$. **This distribution is tabulated on Table A3** of Johansen and Juselius (1990), p.209.

Finally we test the hypothesis H_h^* in H_h, by a likelihood ratio test, i.e., test that the trend is absent under the assumption that there are h cointegrating relations.

Theorem 3:
The asymptotic distribution of the likelihood ratio test $-2\log(H_h^*|H_h)$ for the hypothesis H_h^* given the hypothesis H_h, i.e., $\mathbf{B}'_\perp\mathbf{c} = \mathbf{0}$, when there are h cointegrating vectors, is asymptotically distributed as χ^2 with $k-h$ degree of freedom.

3.2.3 MLE estimation of Parameters

Given the cointegration rank (h) have been inferred from the likelihood ratio test discussed above, the parameters are estimated from

$$\tilde{\mathbf{B}} = -\tilde{\mathbf{B}}^* = -\tilde{\mathbf{S}}_{\mathbf{uw}}\tilde{\mathbf{A}}(\tilde{\mathbf{A}}'\tilde{\mathbf{S}}_{\mathbf{ww}}\tilde{\mathbf{A}})^{-1} = -\tilde{\mathbf{S}}_{\mathbf{uw}}\tilde{\mathbf{A}} \tag{50}$$

and therefore

$$\tilde{\mathbf{B}}\tilde{\mathbf{A}}' = -\tilde{\mathbf{S}}_{\mathbf{uw}}\tilde{\mathbf{A}}\tilde{\mathbf{A}}'. \tag{51}$$

Recall from (47) we have

$$\begin{aligned}\tilde{\mathbf{B}}\tilde{\mathbf{A}}' &= \begin{bmatrix} -\tilde{\mathbf{B}}\boldsymbol{\mu}_1^* & -\tilde{\mathbf{B}}\mathbf{A}' \end{bmatrix} \\ &= \begin{bmatrix} -\mathbf{c} & -\boldsymbol{\xi}_0 \end{bmatrix}\end{aligned}$$

Thus, (51) implies that the maximum likelihood estimate of $\mathbf{c}$ and $\boldsymbol{\xi}_0$ are given by

$$\begin{bmatrix} \tilde{\mathbf{c}} & \tilde{\boldsymbol{\xi}}_0 \end{bmatrix} = \tilde{\mathbf{S}}_{\mathbf{uw}}\tilde{\mathbf{A}}\tilde{\mathbf{A}}'.$$

The MLE of $\boldsymbol{\xi}_i$ and $\boldsymbol{\Omega}$ are given by

$$\tilde{\boldsymbol{\xi}}_i = \tilde{\boldsymbol{\Pi}}_i - \tilde{\mathbf{c}}\tilde{\boldsymbol{w}}_t' - \tilde{\boldsymbol{\xi}}_0\tilde{\boldsymbol{\Theta}}_i, \quad for\ i = 1, 2, ..., p-1, \tag{52}$$

$$\tilde{\boldsymbol{\Omega}} = 1/T\sum_{t=1}^{T}\left[(\tilde{\mathbf{u}}_t - \tilde{\mathbf{c}}\tilde{w}_t - \tilde{\boldsymbol{\xi}}_0\tilde{\mathbf{v}}_t)(\tilde{\mathbf{u}}_t\tilde{\mathbf{c}}\tilde{w}_t - \tilde{\boldsymbol{\xi}}_0\tilde{\mathbf{v}}_t)'\right]. \tag{53}$$

4 Simulation of Johansen's LR test Statistics's Critical Values: Functional of Multivariate Brownian Motion

The limit distribution are expressed as functions of the stochastic matrix $\boldsymbol{W}$ with *i.i.d.* components as

$$\left\{\int (d\boldsymbol{W})F' \left[\int FF'du\right]^{-1} \int F(d\boldsymbol{W})'\right\}. \tag{54}$$

The $(k-h)$-dimensional Brownian motion $\mathbf{w}(t) = (W_1(t), W_2(t), ..., W_{k-h}(t))'$ can be approximated by a random walk with $T = 400$ (say) steps.