

Lecture 5: Co-Integration

Unit-Root Nonstationary VARMA Models

3.1 Co-integration Tests

From the error-correction representation, the matrix $\mathbf{\Pi}$ plays an important role in co-integration study. If $\text{Rank}(\mathbf{\Pi}) = 0$, then there is no co-integrating vector, implying that the system is not co-integrated. Testing for co-integration thus focuses on checking the rank of $\mathbf{\Pi}$.

3.1.1 Case of VAR models

Because of simplicity, much work in the literature concerning co-integration tests focuses on VAR models. Johansen's method is perhaps the best known approach in this area. Consider the Gaussian VAR(p) model with a trend component,

$$\mathbf{z}_t = \mathbf{D}(t) + \sum_{i=1}^p \boldsymbol{\phi}_i \mathbf{z}_{t-i} + \mathbf{a}_t, \quad (3.1)$$

where $\mathbf{D}(t) = \mathbf{d}_0 + \mathbf{d}_1 t$, \mathbf{d}_i are constant vectors, and $\{\mathbf{a}_t\}$ is a sequence of *iid* Gaussian random vectors with mean zero and positive definite covariance matrix $\text{Var}(\mathbf{a}_t) = \boldsymbol{\Sigma}$. We assume that $|\boldsymbol{\phi}(x)| = |\mathbf{I} - \sum_{i=1}^p \boldsymbol{\phi}_i x^i| \neq 0$ for $|x| < 1$, but $|\boldsymbol{\phi}(1)| = 0$. Thus, \mathbf{z}_t has some unit roots. In addition, consider the Error-Correction Model (ECM) for \mathbf{z}_t ,

$$\Delta \mathbf{z}_t = \mathbf{D}(t) + \mathbf{\Pi} \mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\phi}_i^* \Delta \mathbf{z}_{t-i} + \mathbf{a}_t, \quad (3.2)$$

where $\mathbf{\Pi} = -\boldsymbol{\phi}(1)$, $\boldsymbol{\phi}_i^* = -\sum_{j=i+1}^p \boldsymbol{\phi}_j$ for $i = 1, \dots, p-1$, and $\Delta \mathbf{z}_t = \mathbf{z}_t - \mathbf{z}_{t-1}$. Note that the coefficient matrices in Eq. (3.1) can be obtained from those in Eq. (3.2) as

$$\begin{aligned} \boldsymbol{\phi}_i &= \boldsymbol{\phi}_i^* - \boldsymbol{\phi}_{i-1}^*, & i = 2, \dots, p \\ \boldsymbol{\phi}_1 &= \mathbf{\Pi} + \mathbf{I} + \boldsymbol{\phi}_1^*. \end{aligned}$$

Let m be the rank of $\mathbf{\Pi}$. There are two cases of interest.

1. $\text{Rank}(\mathbf{\Pi}) = 0$: This implies that $\mathbf{\Pi} = \mathbf{0}$. Thus, there is no co-integrating vector. In this case, \mathbf{z}_t has k unit roots and we can work directly on the differenced series $\Delta \mathbf{z}_t$, which is a VAR($p-1$) process.

2. $\text{Rank}(\mathbf{\Pi}) = m > 0$: In this case, \mathbf{z}_t has m co-integrating vectors and $k - m$ unit roots. As discussed before, there are $k \times m$ full-rank matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'.$$

The vector series $\mathbf{w}_t = \boldsymbol{\beta}'\mathbf{z}_t$ is an $I(0)$ process, which is referred to as the co-integrating series, and $\boldsymbol{\alpha}$ denotes the impact of the co-integrating series on $\Delta\mathbf{z}_t$. Let $\boldsymbol{\beta}_\perp$ be a $k \times (k - m)$ full-rank matrix such that $\boldsymbol{\beta}_\perp'\boldsymbol{\beta} = \mathbf{0}$. Then, $\mathbf{y}_t = \boldsymbol{\beta}_\perp'\mathbf{z}_t$ has $k - m$ unit roots and can be considered as the $k - m$ common trends of \mathbf{z}_t .

3.1.2 Specification of deterministic terms

Johansen (1995, book, Oxford University Press) discusses several forms for $\mathbf{D}(t) = \mathbf{d}_0 + \mathbf{d}_1t$.

1. $\mathbf{D}(t) = \mathbf{0}$: In this case, there is no constant term in the ECM model of Eq. (3.2). Thus, the components of \mathbf{z}_t are $I(1)$ processes without drift and $\mathbf{w}_t = \boldsymbol{\beta}'\mathbf{z}_t$ has mean zero.
2. $\mathbf{d}_0 = \boldsymbol{\alpha}\mathbf{c}_0$ and $\mathbf{d}_1 = \mathbf{0}$: This is a case of restricted constant. The ECM model becomes

$$\Delta\mathbf{z}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{z}_{t-1} + \mathbf{c}_0) + \sum_{i=1}^{p-1} \phi_i^* \Delta\mathbf{z}_{t-i} + \mathbf{a}_t.$$

Here the components of \mathbf{z}_t are $I(1)$ series without drift and the co-integrating series \mathbf{w}_t has a non-zero mean \mathbf{c}_0 .

3. $\mathbf{d}_1 = \mathbf{0}$ and \mathbf{d}_0 is unrestricted: This is a case of unrestricted constant. The ECM becomes

$$\Delta\mathbf{z}_t = \mathbf{d}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta\mathbf{z}_{t-i} + \mathbf{a}_t.$$

The \mathbf{z}_t series are $I(1)$ with drift \mathbf{d}_0 and \mathbf{w}_t may have a non-zero mean.

4. $\mathbf{D}(t) = \mathbf{d}_0 + \boldsymbol{\alpha}\mathbf{c}_1t$. This is a case of restricted trend. The ECM becomes

$$\Delta\mathbf{z}_t = \mathbf{d}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{z}_{t-1} + \mathbf{c}_1t) + \sum_{i=1}^{p-1} \phi_i^* \Delta\mathbf{z}_{t-i} + \mathbf{a}_t.$$

The \mathbf{z}_t series are $I(1)$ with drift \mathbf{d}_0 and the co-integrating series \mathbf{w}_t has a linear trend term \mathbf{c}_1t .

5. $\mathbf{D}(t) = \mathbf{d}_0 + \mathbf{d}_1t$: In this case, the component series of \mathbf{z}_t are $I(1)$ with a quadratic trend and \mathbf{w}_t has a linear trend.

The limiting distributions of co-integrating tests depend on the specification of $\mathbf{D}(t)$.

3.1.3 Review of Likelihood Ratio Tests

We start with a brief review of likelihood ratio test under multivariate normal distribution. Consider a random vector \mathbf{z} that follows a multivariate normal distribution with mean zero and positive-definite covariance matrix Σ_z . Suppose that $\mathbf{z}' = (\mathbf{x}', \mathbf{y}')$, where the dimensions of \mathbf{x} and \mathbf{y} is p and q , respectively. Without loss of generality, assume that $q \leq p$. We partition Σ_z accordingly as

$$\Sigma_z = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

Suppose also that the null hypothesis of interest is that \mathbf{x} and \mathbf{y} are uncorrelated. That is, we are interested in testing $H_o : \Sigma_{xy} = \mathbf{0}$ versus $H_a : \Sigma_{xy} \neq \mathbf{0}$. This is equivalent to testing the coefficient matrix $\mathbf{\Pi}$ being zero in the multivariate linear regression

$$\mathbf{x}_i = \mathbf{\Pi} \mathbf{y}_i + \mathbf{e}_i,$$

where \mathbf{e}_i denotes the error term. The likelihood ratio test of $\mathbf{\Pi} = \mathbf{0}$ has certain optimal properties and is the test of choice in practice.

Assume that the available random sample is $\{\mathbf{z}_i\}_{i=1}^T$. Under the null hypothesis $\Sigma_{xy} = 0$ so that the maximum likelihood estimate of Σ_z is

$$\hat{\Sigma}_o = \begin{bmatrix} \hat{\Sigma}_{xx} & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_{yy} \end{bmatrix},$$

where $\hat{\Sigma}_{xx} = \frac{1}{T} \sum_{i=1}^T \mathbf{x}_i \mathbf{x}_i'$ and $\hat{\Sigma}_{yy} = \frac{1}{T} \sum_{i=1}^T \mathbf{y}_i \mathbf{y}_i'$. If $E(\mathbf{z}_t) \neq \mathbf{0}$, the mean-correction is needed in the above covariance matrix estimators. The maximized likelihood function under the null hypothesis is

$$\ell_o \propto |\hat{\Sigma}_o|^{-T/2} = (|\hat{\Sigma}_{xx}| |\hat{\Sigma}_{yy}|)^{-T/2},$$

see, for example, Johnson and Wichern (2007, Eq. (4.18) on page 172). On other other hand, under the alternative, there is no constraint on the covariance matrix so that the maximum likelihood estimate of Σ_z is

$$\hat{\Sigma}_a = \frac{1}{T} \sum_{i=1}^T \begin{bmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{bmatrix} [\mathbf{x}_i', \mathbf{y}_i'] \equiv \begin{bmatrix} \hat{\Sigma}_{xx} & \hat{\Sigma}_{xy} \\ \hat{\Sigma}_{yx} & \hat{\Sigma}_{yy} \end{bmatrix}.$$

The maximized likelihood function under the alternative hypothesis is

$$\ell_a \propto |\hat{\Sigma}_a|^{-T/2} = (|\hat{\Sigma}_{xx}| |(\hat{\Sigma}_{yy} - \hat{\Sigma}_{yx} \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy})|)^{-T/2}.$$

The likelihood ratio test statistic is therefore

$$L = \frac{\ell_o}{\ell_a} = \left(\frac{|\hat{\Sigma}_a|}{|\hat{\Sigma}_o|} \right)^{T/2} = \left(|\mathbf{I} - \hat{\Sigma}_{yy}^{-1} \hat{\Sigma}_{yx} \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy}| \right)^{T/2}.$$

One rejects the null hypothesis if L is small. Next, let $\{\lambda_i\}_{i=1}^q$ be the eigenvalues of the matrix $\hat{\Sigma}_{yy}^{-1}\hat{\Sigma}_{yx}\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xy}$. Then, $\{1 - \lambda_i\}_{i=1}^q$ are the eigenvalues of $\mathbf{I} - \hat{\Sigma}_{yy}^{-1}\hat{\Sigma}_{yx}\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xy}$. Consequently, the negative log likelihood ratio statistic is

$$\begin{aligned} \text{LR} &= -\frac{T}{2} \ln \left(|\mathbf{I} - \hat{\Sigma}_{yy}^{-1}\hat{\Sigma}_{yx}\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xy}| \right) \\ &= -\frac{T}{2} \ln \left(\prod_{i=1}^q (1 - \lambda_i) \right) = -\frac{T}{2} \sum_{i=1}^q \ln(1 - \lambda_i). \end{aligned} \quad (3.3)$$

One rejects the null hypothesis if the test statistic LR is large.

Note that $\{\lambda_i\}$ are the squared sample canonical correlation coefficients between \mathbf{x} and \mathbf{y} . Thus, the likelihood ratio test for $\mathbf{\Pi} = \mathbf{0}$ is based on the canonical correlation analysis between \mathbf{x} and \mathbf{y} .

3.1.4 Co-integration tests

Return to co-integration test, which essentially is to test the rank of the matrix $\mathbf{\Pi}$ in Eq. (3.2). Since $\mathbf{\Pi}$ is related to the covariance matrix between \mathbf{z}_{t-1} and $\Delta\mathbf{z}_t$, we can apply the likelihood ratio test to check its rank. To simplify the procedure, it pays to concentrate out the effects of $\mathbf{D}(t)$ and $\Delta\mathbf{z}_{t-i}$ from Eq. (3.2) before estimating $\mathbf{\Pi}$. To this end, consider the next two linear regressions

$$\Delta\mathbf{z}_t = \mathbf{D}(t) + \sum_{i=1}^{p-1} \phi_i^* \Delta\mathbf{z}_{t-i} + \mathbf{u}_t, \quad (3.4)$$

$$\mathbf{z}_{t-1} = \mathbf{D}(t) + \sum_{i=1}^{p-1} \Phi_i \Delta\mathbf{z}_{t-i} + \mathbf{v}_t, \quad (3.5)$$

where it is understood that the form of $\mathbf{D}(t)$ is pre-specified, and \mathbf{u}_t and \mathbf{v}_t denote the error terms. These two regressions can be estimated by the least squares method. Let $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ are the residuals of Eqs. (3.4) and (3.5), respectively. Then, we have the regression

$$\hat{\mathbf{u}}_t = \mathbf{\Pi}\hat{\mathbf{v}}_t + \mathbf{e}_t, \quad (3.6)$$

where \mathbf{e}_t denotes the error term. The LS estimates of $\mathbf{\Pi}$ are identical between Eqs. (3.2) and (3.6). Let

$$H(0) \subset H(1) \subset \dots \subset H(k)$$

be the nested models such that under $H(m)$ there are m co-integrating vectors in \mathbf{z}_t . That is, under $H(m)$, $\text{rank}(\mathbf{\Pi}) = m$. In particular, we have $\mathbf{\Pi} = \mathbf{0}$ under $H(0)$.

Applying the results of subsection 3.1.3, we define

$$\hat{\Sigma}_{00} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t', \quad \hat{\Sigma}_{11} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_t \hat{\mathbf{v}}_t', \quad \hat{\Sigma}_{01} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{v}}_t'.$$

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ be the ordered eigenvalues of the matrix $\hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{10} \hat{\Sigma}_{00}^{-1} \hat{\Sigma}_{01}$, and \mathbf{g}_i be the eigenvector associated with eigenvalue λ_i . That is,

$$\hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{10} \hat{\Sigma}_{00}^{-1} \hat{\Sigma}_{01} \mathbf{g}_i = \lambda_i \mathbf{g}_i. \quad (3.7)$$

Equivalently, we have

$$\hat{\Sigma}_{10} \hat{\Sigma}_{00}^{-1} \hat{\Sigma}_{01} \mathbf{g}_i = \lambda_i \hat{\Sigma}_{11} \mathbf{g}_i, \quad i = 1, \dots, k.$$

Furthermore, the eigenvectors can be normalized such that

$$\mathbf{G}' \hat{\Sigma}_{11}^{-1} \mathbf{G} = \mathbf{I},$$

where $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_k]$ is the matrix of eigenvectors.

Remark: The eigenvalues are non-negative and the eigenvectors are orthogonal to each other with respect to $\hat{\Sigma}_{11}$ because

$$\left[\hat{\Sigma}_{11}^{-1/2} \hat{\Sigma}_{10} \hat{\Sigma}_{00}^{-1} \hat{\Sigma}_{01} \hat{\Sigma}_{11}^{-1/2} \right] \left(\hat{\Sigma}_{11}^{1/2} \mathbf{g}_i \right) = \lambda_i \left(\hat{\Sigma}_{11}^{1/2} \mathbf{g}_i \right), \quad i = 1, \dots, k.$$

In the prior equation, the matrix inside the square brackets is symmetric.

Consider the nested hypotheses:

$$H_o : m = m_0 \quad \text{vs.} \quad H_a : m > m_0,$$

where $m = \text{Rank}(\mathbf{\Pi})$ and m_0 is a given integer between 0 and $k - 1$ with k being the dimension of \mathbf{z}_t . Johansen's trace statistic is defined as

$$L_{tr}(m_0) = -(T - kp) \sum_{i=m_0+1}^k \ln(1 - \lambda_i), \quad (3.8)$$

where λ_i are the eigenvalues defined in Eq. (3.7). If $\text{rank}(\mathbf{\Pi}) = m_0$, then the m_0 smallest eigenvalues should be zero, i.e. $\lambda_{m_0+1} = \dots = \lambda_k = 0$, and the test statistic should be small. On the other hand, if $\text{rank}(\mathbf{\Pi}) > m_0$, then some of the eigenvalues in $\{\lambda_i\}_{i=m_0+1}^k$ are non-zero and the test statistic should be large. Because of the existence of unit roots, the limiting distribution of L_{tr} statistic is not chi-squared. It is a function of the standard Brownian motion. Specifically, the limiting distribution of the test statistic $L_{tr}(m_0)$ has been derived and is given by

$$L_{tr}(m_0) \rightarrow_d \text{tr} \left\{ \left[\int_0^1 \mathbf{W}_v(\mathbf{u}) d\mathbf{W}_v(\mathbf{u})' \right]' \left[\int_0^1 \mathbf{W}_v(\mathbf{u}) \mathbf{W}_v(\mathbf{u})' d\mathbf{u} \right]^{-1} \left[\int_0^1 \mathbf{W}_v(\mathbf{u}) d\mathbf{W}_v(\mathbf{u})' \right] \right\}, \quad (3.9)$$

where $v = m_0$ and $\mathbf{W}_v(\mathbf{u})$ is a v -dimensional standard Brownian motion process. This distribution depends only on m_0 and not on any parameters or the order of the VAR model.

Thus, the critical values can be tabulated via simulation. See, for instance, Johansen (1988) and Reinsel and Ahn (1992).

The trace statistic L_{tr} can be applied sequentially, starting with $m_0 = 0$, to identify the rank of $\mathbf{\Pi}$, i.e. the number of co-integrating vectors. Johansen also proposes a likelihood ratio test based on the maximum eigenvalue to identify the number of co-integrating vectors. Consider the hypotheses

$$H_o : m = m_0 \quad \text{vs.} \quad H_a : m = m_0 + 1,$$

where m and m_0 are defined as before in Eq. (3.8). The test statistic is

$$L_{max}(m_0) = -(T - kp) \ln(1 - \lambda_{m_0+1}). \quad (3.10)$$

Similarly to the case of $L_{tr}(m_0)$ statistic, the limiting distribution of $L_{max}(m_0)$ is not chi-squared. It is also a function of the standard Brownian motion, and the critical values of the test statistic have been obtained via simulation.

Since the eigenvectors \mathbf{g}_i give rise to linear combinations of \mathbf{z}_t , they provide an estimate of the co-integrating vectors. In other words, for a given $m_0 = \text{Rank}(\mathbf{\Pi})$, an estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = [\mathbf{g}_1, \dots, \mathbf{g}_{m_0}]$.

3.1.5 An Illustration

Consider the monthly yields of Moody's seasoned corporate AAA and BAA bonds from 1954.7 to 2005.2. The data are obtained from Federal Reserve Bank of St Louis. Figure 3.1 shows the time plots of the yield data. The two series move in a parallel manner.

Splus Demonstration

```
> da=read.table("m-bnd.txt")
> zt=da[,4:5]
> zt=data.frame(zt)
> var.fit=VAR(zt,max.ar=10,criterion="AIC")    <== Find the order p
> var.fit$ar.order
[1] 8
> coint.uc=coint(zt,lag=7)    % lag = p-1.
> coint.uc
Call:
coint(Y = zt, lags = 7)
```

```
Trend Specification:
H1(r): Unrestricted constant
```

```
Trace tests significant at the 5% level are flagged by ' + '.
Trace tests significant at the 1% level are flagged by ' ++ '.
Max Eigenvalue tests significant at the 5% level are flagged by ' * '.
Max Eigenvalue tests significant at the 1% level are flagged by ' ** '.
```

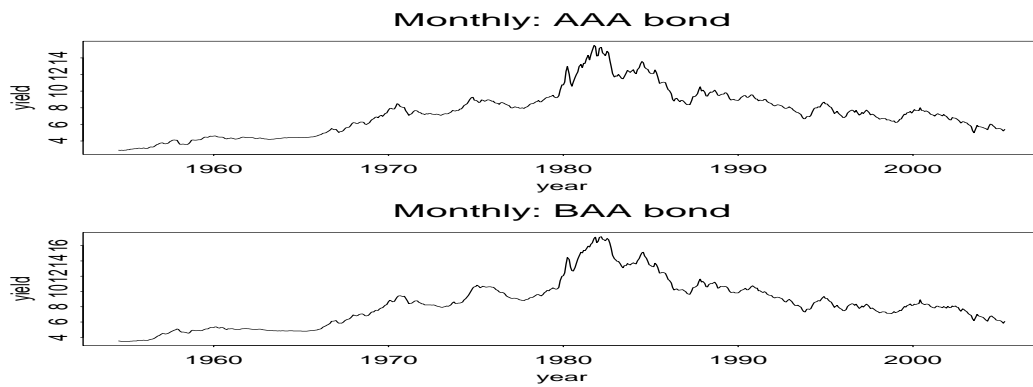


Figure 3.1: Time plots of Monthly Moody's Seasoned Corporate AAA and BAA Bond Yields from July 1954 to February 2005.

```

Tests for Cointegration Rank:
      Eigenvalue Trace Stat  95% CV  99% CV Max Stat  95% CV  99% CV
H(0)+++*  0.0322   23.0683  15.4100 20.0400 19.6985  14.0700 18.6300
H(1)      0.0056    3.3698   3.7600  6.6500  3.3698   3.7600  6.6500
% The result indicates there is a co-integrating vector.
> coint.uc$coint.vectors[1,]
      V4      V5
-4.667266 4.17414
>
> coint.rc=coint(zt,trend='rc',lag=7)
> coint.rc
Call:
coint(Y = zt, lags = 7, trend = "rc")

```

```

Trend Specification:
H1*(r): Restricted constant

```

```

Trace tests significant at the 5% level are flagged by ' +'.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged by ' *'.
Max Eigenvalue tests significant at the 1% level are flagged by '**'.

```

```

Tests for Cointegration Rank:
      Eigenvalue Trace Stat  95% CV  99% CV Max Stat  95% CV  99% CV
H(0)+ *  0.0323   23.1942  19.9600 24.6000 19.7042  15.6700 20.2000
H(1)     0.0058    3.4899   9.2400 12.9700  3.4899   9.2400 12.9700
> coint.rc$coint.vectors[1,]
      V4      V5 Intercept*
4.666782 -4.173484  0.3058871

```

3.2 Maximum likelihood estimation of ECM

For a given $m_0 = \text{Rank}(\mathbf{\Pi})$, the ECM becomes a reduced rank multivariate regression and can be estimated accordingly. As mentioned before, an estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = [\mathbf{g}_1, \dots, \mathbf{g}_{m_0}]$. From this estimate, one can form the normalized estimator $\hat{\boldsymbol{\beta}}_c$ by imposing the appropriate normalization and identifying constraints.

Next, given $\hat{\boldsymbol{\beta}}_c$, we can obtain estimates of other parameters by the multivariate linear regression

$$\Delta \mathbf{z}_t = \boldsymbol{\alpha}_c (\hat{\boldsymbol{\beta}}_c' \mathbf{z}_{t-1}) + \sum_{i=1}^{p-1} \boldsymbol{\phi}_i^* \Delta \mathbf{z}_{t-i} + \mathbf{a}_t,$$

where the subscript of $\boldsymbol{\alpha}$ is used to signify the dependence on $\hat{\boldsymbol{\beta}}_c$. The maximized likeli-

hood function based on the m_0 cointegrating vectors is

$$\ell^{-2/T} \propto |\hat{\Sigma}_{00}| \prod_{i=1}^{m_0} (1 - \lambda_i).$$

The estimates of the orthogonal complements of α_c and β_c are given by

$$\begin{aligned} \hat{\alpha}_{c,\perp} &= \hat{\Sigma}_{00}^{-1} \hat{\Sigma}_{11} [\mathbf{g}_{m_0+1}, \dots, \mathbf{g}_k] \\ \hat{\beta}_{c,\perp} &= \hat{\Sigma}_{11} [\mathbf{g}_{m_0+1}, \dots, \mathbf{g}_k]. \end{aligned}$$

Finally, pre-multiplying the ECM by the orthogonal complement α'_{\perp} , we obtain

$$\alpha'_{\perp}(\Delta z_t) = \alpha'_{\perp} \alpha \beta' z_{t-1} + \sum_{i=1}^p \alpha'_{\perp} \phi_i^* \Delta z_{t-i} + \alpha'_{\perp} a_t.$$

Since $\alpha'_{\perp} \alpha = \mathbf{0}$, we see that

$$\alpha'_{\perp}(\Delta z_t) = \sum_{i=1}^p \alpha'_{\perp} \phi_i^* (\Delta z_{t-i}) + \alpha'_{\perp} a_t,$$

which does not contain any error-correction term. Consequently, $\mathbf{y}_t = \alpha'_{\perp} z_t$ has $k - m$ unit roots and represents the common trends of z_t .

Splus Demonstration

```
> vecm.fit=VECM(coint.uc)
> summary(vecm.fit)
Call:
VECM(test = coint.uc)
```

Cointegrating Vectors:

```
coint.1
1.0000
```

```
V5 -0.8943
(std.err) 0.0164
(t.stat) -54.6718
```

VECM Coefficients:

```
          V4      V5
coint.1 0.0028 0.0676
(std.err) 0.0378 0.0335
(t.stat) 0.0740 2.0196

V4.lag1 0.5146 0.3633
(std.err) 0.0938 0.0831
```

| | | | |
|----------------------------------|----------|---------|--------|
| | (t.stat) | 5.4845 | 4.3721 |
| V5.lag1 | -0.0573 | 0.1417 | |
| (std.err) | 0.1039 | 0.0920 | |
| (t.stat) | -0.5516 | 1.5403 | |
| V4.lag2 | -0.2172 | -0.0803 | |
| (std.err) | 0.0958 | 0.0848 | |
| (t.stat) | -2.2679 | -0.9465 | |
| V5.lag2 | -0.0525 | -0.0775 | |
| (std.err) | 0.1057 | 0.0936 | |
| (t.stat) | -0.4967 | -0.8282 | |
| (*Lags 3 & 4 are insignificant*) | | | |
| V4.lag5 | 0.3788 | 0.2391 | |
| (std.err) | 0.0956 | 0.0847 | |
| (t.stat) | 3.9607 | 2.8236 | |
| V5.lag5 | -0.2998 | -0.1787 | |
| (std.err) | 0.1052 | 0.0931 | |
| (t.stat) | -2.8507 | -1.9188 | |
| V4.lag6 | -0.2449 | -0.2021 | |
| (std.err) | 0.0926 | 0.0820 | |
| (t.stat) | -2.6460 | -2.4658 | |
| V5.lag6 | 0.2293 | 0.1766 | |
| (std.err) | 0.1037 | 0.0919 | |
| (t.stat) | 2.2101 | 1.9223 | |
| V4.lag7 | -0.1282 | -0.0453 | |
| (std.err) | 0.0893 | 0.0791 | |
| (t.stat) | -1.4358 | -0.5735 | |
| V5.lag7 | 0.1515 | 0.1057 | |
| (std.err) | 0.0955 | 0.0846 | |
| (t.stat) | 1.5859 | 1.2497 | |
| Intercept | 0.0030 | 0.0066 | |
| (std.err) | 0.0085 | 0.0075 | |
| (t.stat) | 0.3469 | 0.8801 | |

Regression Diagnostics:

| | | |
|-----------|--------|--------|
| | V4 | V5 |
| R-squared | 0.2276 | 0.2754 |

Adj. R-squared 0.2078 0.2568
 Resid. Scale 0.1985 0.1757

Information Criteria:

| logL | AIC | BIC | HQ |
|----------|------------|------------|------------|
| 817.6215 | -1603.2431 | -1532.8656 | -1575.8486 |

| | total residual | |
|--------------------|----------------|-----|
| Degree of freedom: | 601 | 585 |

3.3 Discussion

Co-integration is an interesting concept and has attracted much interest in various scientific fields. However, co-integration has its shares of weakness. First, it does not address the rate of achieving long-term equilibrium. For instance, if the co-integrating series $\mathbf{w}_t = \boldsymbol{\beta}'\mathbf{z}_t$ has a characteristic root that is close to unit circle, then the co-integration relationship may take a long time period to achieve. Second, co-integration tests are scale invariant. For instance, multiplying the component series of \mathbf{z}_t by any non-singular matrix should not change the result of a co-integration test. However, scaling can be very important in practice. Consider the following simple examples.

Example 1. Suppose $\{\mathbf{a}_t\}$ is a sequence of iid bivariate normal random vectors with mean zero and covariance matrix $\text{Var}(\mathbf{a}_t) = \mathbf{I}$. Let $\mathbf{y}_0 = \mathbf{0}$ and define $\mathbf{y}_t = \mathbf{y}_{t-1} + \mathbf{a}_t$. That is, the components of \mathbf{y}_t are univariate random walk processes. Let $z_{1t} = 10000y_{1t} + \frac{1}{10000}y_{2t}$ and $z_{2t} = 10000y_{1t} - \frac{1}{10000}y_{2t}$. Clearly, the \mathbf{z}_t series has two unit roots, that is, z_{1t} and z_{2t} are not co-integrated. However, because of the scaling effects, for any reasonable sample size commonly encountered in practice z_{1t} and z_{2t} should move closely together. In fact, for moderate T , the point forecasts $z_{1,T}(\ell)$ and $z_{2,T}(\ell)$ should be very close to each other for all ℓ . Consequently, a rejection of co-integration does not necessarily imply that the two forecast series are far apart.

Example 2. Suppose that $\{\mathbf{a}_t\}$ are the same as those of Example 1. However, construct the \mathbf{y}_t series via $y_{1t} = 0.9y_{1,t-1} + a_{1t}$ and $y_{2t} = y_{2,t-1} + a_{2t}$. Define $z_{1t} = \frac{1}{10000}y_{2t} - 10000y_{1t}$ and $z_{2t} = \frac{1}{10000}y_{2t} + 10000y_{1t}$. In this case, \mathbf{z}_t has a single unit root and its two components series are co-integrated with co-integrating $(1, -1)'$. However, the forecasts $z_{1,T}(\ell)$ and $z_{2,T}(\ell)$ can be very different for moderate T and ℓ . This example demonstrates that a co-integrated system may take a long period of time to show the co-integrating relationship in forecasting.