Selecting the Cointegration Rank and the Form of the Intercept when the Time Trends are not cointegrated

by Antonio Aznar and Manuel Salvador
University of Zaragoza

Contents

1.Introduction

- In this paper, we propose a procedure to test the cointegration rank, the number of lags and the form adopted by the deterministic terms of a multivariate dynamic system using an Information Criterion (IC).
- Extension of the paper “Selecting the rank of the cointegration space and the form of the intercept using an information criterion” By A. Aznar and M. Salvador (2002)

  Econometric Theory, 18, 926-947

2. Models

- Consider the following four models

$$M_{3,r,k} : \Delta X_t = \alpha\beta' X_{t-1} + \mu_0 + \alpha\rho_1 t + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

$$M_{2,r,k} : \Delta X_t = \alpha\beta' X_{t-1} + \mu_0 + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

$$M_{1,r,k} : \Delta X_t = \alpha\beta' X_{t-1} + \alpha\rho_0 + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

$$M_{0,r,k} : \Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

2. Models

- X_t is a vector of p I(1) variables.
- β is a p×r matrix of the cointegrating vectors.
- α is a p×r matrix of adjustment coefficients.
- p is the number of variables, r is the cointegration rank and k is the number of lags.
- ρ_0 and ρ_1 are vectors of r elements
- μ_0 is a vector of p elements.

$\varepsilon_t \ i.i.d.\ E[\varepsilon_t] = 0, Cov(\varepsilon_t) = \Omega$

2. Models

Reduced Rank Regression

$$R_{0t} = \alpha\beta' R_{1t} + \tilde{\varepsilon}_t$$

R_{0t} residuals of $\Delta x_t \; on \left(1, \Delta x_{t-1}, \ldots. \Delta x_{t-k+1}\right)$

R_{1t} residuals of $X_{t-1} \; on \left(1, \Delta x_{t-1}, \ldots. \Delta x_{t-k+1}\right)$

Since $\hat{\alpha} = S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}$ we have

$$Var\left(\tilde{\varepsilon}\right) = \frac{1}{T}\sum_{1}^{T}\left(R_{0t} - \hat{\alpha}\beta' R_{1t}\right)\left(R_{0t} - \hat{\alpha}\beta' R_{1t}\right)' =$$

$$= \left(S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}\right)$$

Where $S_{ij} = \frac{1}{T}\sum_{1}^{T} R_{it} R_{jt}^{'}$

3. The Information Criterion (IC)

- We will use an IC that chooses the model that minimizes the expression

$$IC(M) = -2\log \tilde{L}(M) + c_T n_p(M)$$

Where $\tilde{L}(M)$ is the maximized likelihood function, $n_P(M)$ is the number of parameters and c_T is a deterministic sequence that imposes a penalty to encourage the selection of a parsimonious model. We assume

$$c_T \to \infty \ and \ c_T = o(T) \ as \ T \to \infty \qquad (1)$$

Examples: HQ: $c_T = 2\log\log T$

BIC: $c_T = \log T$

3. The Information Criterion

$$M_{2,r,k} : Z_{0t} = \alpha\beta' Z_{1t} + \theta_1' Z_{2t} + \varepsilon_t$$

$$Z_{0t} = \Delta X_t, Z_{1t} = \left(X_{t-1}\right), Z_{2t} = \left(1, \Delta X_{t-1}, \ldots. \Delta X_{t-k+1}\right) \quad \theta_1' = \left(\mu_0, \Gamma_1, \ldots \Gamma_{k-1}\right)$$

$$X_t = C\sum_{i=1}^{t} \varepsilon_i + \tau t + \tau_0 + Y_t \quad C = \beta_\perp \left(\alpha_\perp' \Gamma \beta_\perp\right)^{-1} \alpha_\perp' \; with \; \beta' \tau = 0$$

$$\tilde{L}\left(M_2\right)^{-(2/T)} = \left| S_{00} - S_{01}\tilde{\beta}^* \left(\tilde{\beta}^{*\prime} S_{11} \tilde{\beta}^*\right)^{-1} \tilde{\beta}^{*\prime} S_{10} \right| = \left|S_{00}\right| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i^*\right) = U_{2,r,k}(\tilde{\beta}^*)$$

$$1 \geq \hat{\lambda}_1^* \geq \ldots\ldots. \geq \hat{\lambda}_p^* \quad roots \; of \quad \left| \hat{\lambda}^* S_{11} - S_{10} S_{00}^{-1} S_{01} \right| = 0$$

$$\hat{\beta}^* = \left(\hat{v}_1^*, \ldots. \hat{v}_r^*\right) is \; the \; MLE \; of \; \beta. The \; v's \; eigenvectors$$

$$S_{ij} = \frac{1}{T} \sum R_{it} R_{jt}' \; i, j = 0,1 \qquad R_{it} \; i = 0,1 \; are \; the \; residuals \; of \; Z_{it} \; on \; Z_{2t}$$

$$n_p(M_2) = \frac{p(p+1)}{2} + \left(k-1\right) p^2 + pr + (p-r)r + p$$

3. The Information Criterion

$$M_{3,r,k} : Z_{0t} = \alpha\beta_1^{'} Z_{11t} + \theta_1^{'} Z_{2t} + \varepsilon_t \ \ with$$

$$Z_{0t} = \Delta X_t , Z_{11t} = \left(X_{t-1}, t \right) \ \ \beta_1^{'} = \left(\beta', \rho_1 \right)_{-1}$$

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + \tau_1 t + \tau_0 + Y_t \quad C = \beta_\perp \left(\alpha_\perp^{'} \Gamma \beta_\perp \right) \ \alpha_\perp^{'} \ with \ \beta' \tau_1 = -\rho_1$$

$$\tilde{L}\left(M_3\right)^{-(2/T)} = \left| S_{00} - S_{0,1b}\tilde{\beta}_1 \left(\tilde{\beta}_1^{'} S_{11,bb} \tilde{\beta}_1 \right)^{-1} \tilde{\beta}_1 S_{1b,0} \right| = \left| S_{00} \right| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i \right) = U_{3,r,k}(\tilde{\beta}_1)$$

$$1 \geq \hat{\lambda}_1 \geq .... \geq \hat{\lambda}_p \quad roots \ of \quad \left| \hat{\lambda} S_{11,bb} - S_{1b,0} S_{00}^{-1} S_{0,1b} \right| = 0$$

$$\hat{\beta} = \left(\hat{v}_1, ...\hat{v}_r \right) is \ the \ MLE \ of \ \beta. The \ v's \ eigenvectors$$

$$Let \ R_{1bt} = \left(R_{1t}, b_t \right) where \ b_t \ residuals \ of \ t \ on \ Z_{2t}$$

$$S_{0,1b} = \frac{1}{T} \sum R_{0t} R_{1bt}^{'} \quad and \quad S_{11,bb} = \frac{1}{T} \sum R_{1bt} R_{1bt}^{'}$$

$$n_p(M_3) = \frac{p(p+1)}{2} + \left(k-1\right) p^2 + p \ \ + (p-r)r + (p+r)$$

4. Proposed Strategy

1. Determine the cointegration rank

$$IC_{3,r_{opt},k} = \min_r IC_{3,r,k} \quad where\ r \in \{0,1,\ldots,p\}$$

2. Determine the number of lags

$$IC_{3,r_{opt},k_{opt}} = \min_k IC_{3,r_{opt},k} \quad where\ k \in \{1,\ldots,k_{\max}\}$$

3. Determine the deterministic terms.

$$IC_{j_{opt},r_{opt},k} = \min_j IC_{j,r_{opt},k_{opt}} \quad where\ j \in \{0,1,2,3\}$$

The model finally chosen will be $M_{j_{opt},r_{opt},k_{opt}}$

5. Cointegration Rank Identificatio Rules

- In this paper, we use alternatively two of the three identification rules considered in chapter 13 of Johansen. The first one satisfies

$$\hat{\beta}' S_{11} \hat{\beta} = I \ or \ \hat{\beta}_1' S_{11,bb} \hat{\beta}_1 = I$$

The second one that is convenient for the mathematical analysis satisfies

$$\tilde{\beta} = \hat{\beta}\left(\overline{\beta}' \hat{\beta}\right)^{-1} \ or \ \tilde{\beta}_1 = \hat{\beta}_1\left(\overline{\beta}_1' \hat{\beta}_1\right)^{-1}$$

5. Cointegration Rank

- **Lemma 4.1.** Let . Assume that 0<r<p and let $Q\left(M_{3,r,k}/M_{3,p,k}\right)$ *be the likelihood ratio of* $M_{3,r,k}$ and $M_{3,p,k}$. We then have $M = M_{i,r,k_0}, i = 0,1,2,3,$ *and let* $M_{i,r,k_0} \subseteq M_{3,r,k_0}$

$$-2\log Q\left(M_{3,r,k}/M_{3,p,k}\right) = O_P\left(1\right)$$

Notice that the result is valid **independently of** k_0 . Next, Theorem 4.1 proves that The IC criterion selects the cointegration rank**, independently of the number of lags,** when the form of the drift is overspecified.

Theorem 4.1. Let $M = M_{i,r,k_0}, i = 0,1,2,3,$ *be the true DGP with* $M_{i,r,k_0} \subseteq M_{3,r,k_0}$ Then, for $r' \neq r$ and k, the following holds

$$P\left(IC_{3,r,k} < IC_{3,r',k}\right) \rightarrow 1 \; as \; T \rightarrow \infty$$

if c_T satisfies (21).

6. Number of Lags

- **Lemma 4.2** Let $U_{3,r,k}(\beta_1)$ be defined as before assuming that β_1 is known. Then, for any k', we have that if $M_{i,r,k'}, i=0,1,2,3,$ is the true DGP , then $U_{3,r,k'}(\tilde{\beta}_1) = U_{3,r,k'}(\theta_i^*) + o_P\left(\frac{1}{T}\right)$, where $\theta_i^* = (\beta,0)'$ if $i \in \{0,2\}$, $\theta_1^{*'} = (\beta', \rho_0)$ and $\theta_3^{*'} = (\beta', \rho_1)$. **Lemma 4.2 permits to continue the analysis assuming that θ_i is known.** Note

$$U_{3,r,k}(\tilde{\beta}_1) = \left| S_{00} - S_{0,1b}\tilde{\beta}_1 \left(\tilde{\beta}_1' S_{11,bb}\tilde{\beta}_1\right)^{-1} \tilde{\beta}_1 S_{1b,0} \right|$$

Proof: First, let $M_{2,r,k}$ be the true DGP. In this case, $\beta_1 = \beta^* = \begin{pmatrix} \beta \\ 0_{1\times r} \end{pmatrix}$

Define $A_{32T} = \left(\beta^*, T^{-1/2}\overline{\gamma}^*, T^{-1}\overline{\tau}^*\right)$ γ^* *is orthogonal to* β^* *and* τ^*

With $\gamma^* = \begin{pmatrix} \gamma \\ 0_{1\times(p-r)} \end{pmatrix}, \tau^* = \begin{pmatrix} \tau \\ -1 \end{pmatrix}$ $\overline{\beta}^* = \beta^*\left(\beta^{*'}\beta^*\right)^{-1}, etc.$

6. Number of Lags

$$\tilde{\beta}_1 = \beta_1 + \overline{\gamma}^* \gamma^{*'} \tilde{\beta}_1 + \overline{\tau}^* \tau^{*'} \tilde{\beta}_1 = \beta_1 + B_{32T} U_{32T}$$

(10)

$$B_{32T} = \left(\overline{\gamma}^*, T^{-1/2} \overline{\tau}^* \right) and\ U_{32T} = \left(\gamma^*, T^{1/2} \tau^* \right)' \tilde{\beta}_1$$

- The roots of $\left| \lambda A_{32T}^{'} S_{11,bb} A_{32T} - A_{32T}^{'} S_{1b,0} S_{00}^{-1} S_{0,1b} A_{32T} \right|$ (2) are those of $\left| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} \right|$ (3) and the eigenvectors of (2) are $A_{32T}^{-1} V$ where V is the matrix of eigenvectors of (3).

Remember that $\tilde{\beta}_1$ are the first eigenvectors of (3) and that the space spanned by the first eigenvectors of (2) is $A^{-1} \hat{\beta}_1 = A^{-1} \tilde{\beta}_1$.

6. Number of Lags

- We have

$$A_{32T}^{-1}\tilde{\beta}_1 = \left(\bar{\beta}^*, T^{1/2}\gamma^*, T\tau^*\right)' \tilde{\beta}_1 = \left(I, T^{1/2}U_{32T}'\right)'$$

$$so\ that \quad T^{1/2}U_{32T} \xrightarrow{P} 0$$

because the space spanned by the r first eigenvectors of (2) converges to the space spanned by the first r unit vectors or equivalently the space spanned by vectors with zeros in the last p-r+1 coordinates.

6. Number of Lags

- Now we have

$$\tilde{\beta}_1' S_{11,bb} \tilde{\beta}_1 = \beta_1' S_{11,bb} \beta_1 + U_{32T}' B_{32T}' S_{11,bb} \beta_1 + \beta_1' S_{11,bb} B_{32T} U_{32T} + U_{32T}' B_{32T}' S_{11,bb} B_{32T} U_{32T} =$$

$$= \beta_1' S_{11,bb} \beta_1 + o_P(1)$$

Because $B_{32T}' S_{11,bb} \beta_1$ $is\ O_P(1)$ **,** $T^{-1} B_{32T}' S_{11,bb} B_{32T}$ $is\ O_P(1)$

and $T^{1/2} U_{32T} \xrightarrow{P} 0$

Similarly, $\tilde{\beta}_1' S_{1b,0} = \beta_1' S_{1b,0} + U_{32T}' B_{32T}' S_{1b,0} = \beta_1' S_{1b,0} + o_P(T^{-1/2})$

The analysis can continue assuming β_1 **known**

- Let $M_{3,r,k}$ be the true DGP. Then,

$$A_{33T} = \left(\beta_1, T^{-1/2}\overline{\gamma}_1^*, T^{-1}\overline{\tau}_1^*\right), B_{33T} = \left(\overline{\gamma}_1^*, T^{-1/2}\overline{\tau}_1^*\right), U_{33T} = \left(\gamma_1^*, T^{1/2}\tau_1^*\right)' \tilde{\beta}_1$$

with $\gamma_1^* = \begin{pmatrix} \gamma \\ 0 \end{pmatrix}, \tau_1^* = \begin{pmatrix} \tau_1 \\ 1 \end{pmatrix}$

In this case, γ *is orthogonal to* β *and* τ_1

For the same reasons given previously

$$T^{1/2} U_{33T} \xrightarrow{P} 0$$

- Then, we have

$$\tilde{\beta}_1' S_{11,bb} \tilde{\beta}_1 = \beta_1' S_{11,bb} \beta_1 + U_{33T}' B_{33T}' S_{11,bb} \beta_1 + \beta_1' S_{11,bb} B_{33T} U_{33T} + U_{33T}' B_{33T}' S_{11,bb} B_{33T} U_{33T}$$
$$= \beta_1' S_{11,bb} \beta_1 + o_P(1)$$

Because $B_{32T}' S_{11,bb} \beta_1 \quad is\ O_P(1) \quad T^{-1} B_{32T}' S_{11,bb} B_{32T} \quad is\ O_P(1)$

and $T^{1/2} U_{32T} \xrightarrow{P} 0$

Similarly, $\tilde{\beta}_1' S_{1b,0} = \beta_1' S_{1b,0} + U_{32T}' B_{32T}' S_{1b,0} = \beta_1' S_{1b,0} + o_P(T^{-1/2})$

and the result follows.

6. Number of Lags

– **Theorem 4.2.** Let $M_0 = M_{3,r,k}$ *and* $M_1 = M_{3\ r,,k'}$ with a matrix of cointegrating vectors that is known and k<k'. If c_T satisfies (21), then

a) If M_0 is the true DGP then $P\left(IC_0 < IC_1\right) \to 1\ as\ T \to \infty$

b) If M_1 is the true DGP then $P\left(IC_0 > IC_1\right) \to 1\ as\ T \to \infty$

The number of lags is consistently estimated

7. Deterministic Terms

- In this section, we will show how the use of the IC criterion allows us to consistently estimate the form of the drift of the model.
- To discriminate between $M_{0,r,k}, M_{1,r,k}, M_{2,r,k}$ *against* $M_{3,r,k}$ when 0<r<p we are going to introduce an slight modification

  of the IC criterion when this is calculated for the three more

  restrictive models .
- For these three models, the likelihood is estimated substituting their parameters by their estimators obtained from $M_{3,r,k}$ and we will use $IC^*(M)$.

7. Deterministic Terms

Lemma 5.1 Let $U_{i,r,k}(\theta_i), i = 0,1,2,3$ be defined as in Section 2 assuming that the values of the parameters are known. Denote $\theta_i = \beta$ if $i \in \{0,2\}$, $\theta_1' = (\beta', \rho_0)$ and $\theta_3' = (\beta', \rho_1)$ .Let $\tilde{\theta}_{i,3,r,k}$ the estimator of θ_i calculated from $M_{3,r,k}$. Then, if $M_{i,r,k'}, i = 0,1,2,3,$ is the true DGP we have

$$U_{j,r,k}\left(\tilde{\theta}_{j3,r,k}\right) = U_{j,r,k}\left(\theta_j\right) + o_P\left(\frac{1}{T}\right)$$

Proof: Similar to that presented for Lemma 4.2

7. Deterministic Terms

- **Theorem 5.1.** Let $M_0 = M_{0,r,k}$ *and* $M_1 = M_{3,r,k}$ *with* r *and* k *fixed*

  If c_T satisfies (21) then

  a) If M_0 is the true DGP then $P\left(IC_0^* < IC_1\right) \to 1 \text{ as } T \to \infty$

  b) If M_1 is the true DGP then $P\left(IC_0^* > IC_1\right) \to 1 \text{ as } T \to \infty$

Theorem 5.2 Let $M_0 = M_{1,r,k}$ *and* $M_1 = M_{3,r,k}$ *with* r *and* k *fixed*

If c_T satisfies (21) then

a) If M_0 is the true DGP then $P\left(IC_0^* < IC_1\right) \to 1 \text{ as } T \to \infty$

b) If M_1 is the true DGP then $P\left(IC_0^* > IC_1\right) \to 1 \text{ as } T \to \infty$

7. Deterministic Terms

- **Theorem 5.3** Let $M_0 = M_{2,r,k}$ and $M_1 = M_{3,r,k}$ with r and k fixed
  If c_T satisfies (21), then

  a) If M_0 is the true DGP then $P\left(IC_0^* < IC_1\right) \to 1 \text{ as } T \to \infty$
  b) If M_1 is the true DGP then $P\left(IC_0^* > IC_1\right) \to 1 \text{ as } T \to \infty$

Proof of Theorem 5.3: a) $M_{2,r,k}$ is the true DGP. If we estimate using the same model it is the case treated by Johansen (1995) Lemma 13.1. If we use $M_{3,r,k}$ to estimate the likelihood the results have been derived in Lemma 4.2. We have to compare $L(M_1)^{-2/T}$ and $L(M_0)^{-2/T}$
Since, in this case, $\rho_1 = 0$ we have

7. Deterministic Terms

$$L\left(M_1\right)^{-2/T} = \left|S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}\right| = L\left(M_0\right)^{-2/T}$$

and hence $$IC^*\left(M_0\right) - IC\left(M_1\right) = -rc_T \to -\infty$$

b) $M_{3,r,k}$ **is the true DGP.** We have

$$L\left(M_0\right)^{-2/T} = \left|S_{00} - T^{-1/2}S_{01}\beta\frac{1}{T}\left(\frac{1}{T^2}\beta' S_{11}\beta\right)^{-1}T^{-1/2}\beta' S_{10}\right| \xrightarrow{p} \left|\Sigma_{00}\right|$$

7.Deterministic Terms

- On the other hand

$$L\left(M_1\right)^{-2/T} = \left|S_{00} - S_{01}\beta_1\left(\beta_1^{'}S_{11}\beta_1\right)^{-1}\beta_1^{'}S_{10}\right| \xrightarrow{P} \left|\Sigma_{00} - \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0}\right|$$

We obtain

$$\begin{aligned} IC_0^* - IC_1 &= -2\log L\left(M_0\right) + 2\log L\left(M_1\right) + o_P\left(T\right) - rc_T \\ &= -T\log\frac{\left|\Sigma_{00} - \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0}\right|}{\left|\Sigma_{00}\right|} + o_p\left(T\right) - rc_T \to \infty \end{aligned}$$

using

$$\left|A - BB'\right| = \left|A\right|\left|I - B'A^{-1}B\right| < \left|A\right| \; if \;\; B \neq 0$$

Remaining questions

1. To carry out simulation exercises to confirm the theoretical results.
2. **Key question**: In the first step, is the determination of the cointegration rank really independent of the number of lags?
3. To derive the results in the third step estimating the maximum likelihood of each model with the estimators calculated with the corresponding model and not with the estimators derived from the less restrictive model.