The Econometrics Journal

Econometrics Journal (2009), volume **12**, pp. S83–S104.
doi: 10.1111/j.1368-423X.2008.00270.x

Semiparametric cointegrating rank selection

XU CHENG[†] AND PETER C. B. PHILLIPS[†,‡]

[†]*Department of Economics, Yale University, 28 Hillhouse Avenue, New Haven, CT 06511, USA*
E-mail: xu.cheng@yale.edu

[‡]*University of York, University of Auckland and Singapore Management University*
E-mail: peter.phillips@yale.edu

First version received: January 2008; final version accepted: September 2008

Summary Some convenient limit properties of usual information criteria are given for cointegrating rank selection. Allowing for a non-parametric short memory component and using a reduced rank regression with only a single lag, standard information criteria are shown to be weakly consistent in the choice of cointegrating rank provided the penalty coefficient $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$. The limit distribution of the AIC criterion, which is inconsistent, is also obtained. The analysis provides a general limit theory for semiparametric reduced rank regression under weakly dependent errors. The method does not require the specification of a full model, is convenient for practical implementation in empirical work, and is sympathetic with semiparametric estimation approaches to co-integration analysis. Some simulation results on the finite sample performance of the criteria are reported.

Keywords: *Cointegrating rank*, *Consistency*, *Information criteria*, *Model selection*, *Non-parametric*, *Short memory*, *Unit roots*.

1. INTRODUCTION

Information criteria are now widely used in parametric settings for econometric model choice. The methods have been especially well studied in stationary systems. Models that allow for non-stationarity are particularly relevant in econometric work and have been considered by several authors, including Tsay (1984), Pötscher (1989), Wei (1992), Phillips and Ploberger (1996), Phillips (1996) and Nielsen (2006), among others.

Model choice methods are heavily used in empirical work and in forecasting exercises, the most common applications involving choice of lag length in (vector) autoregression and variable choice in regression. The methods have also been suggested and used in the context of cointegrating rank choice where they are known to be consistent under certain conditions, at least in parametric models (Chao and Phillips, 1999). This application is natural because cointegrating rank is an order parameter for which model selection methods are particularly well suited since there are only a finite number of possible choices. Furthermore, rank order may be combined with lag length and intercept and trend degree parameters to provide a wide compass of choice in parametric models that is convenient in practical work, as discussed in Phillips (1996).

When the focus is on co-integration and cointegrating rank selection, it is not necessary to build a complete model for statistical purposes. Indeed, many of the approaches that have been developed for econometric estimation and inference in such contexts are semiparametric in character so that the model user can be agnostic regarding the short memory features of the data and concentrate on long-run behaviour. In such settings, it will often be desirable to perform the evaluation of cointegrating rank (or choice of the number of unit roots in a system) in a semiparametric context allowing for a general short memory component in the time series.

The present paper has this goal and looks specifically at the issue of cointegrating rank choice by information criteria. In the case of a univariate series, this choice reduces to distinguishing unit root time series from stationary series. In such a context, it is known that information criteria provide consistent model choice in a semiparametric framework (Phillips, 2008). The contribution of this paper is to extend that work to the multivariate setting in the context of a semiparametric reduced rank regression of the form

$$\Delta X_t = \alpha\beta' X_{t-1} + u_t, \quad t \in \{1, \ldots, n\}, \tag{1.1}$$

where X_t is an m- vector time series, α and β are $m \times r_0$ full rank matrices and u_t is a weakly dependent stationary time series with zero mean and continuous spectral density matrix $f_u(\lambda)$. The series X_t is initialized at $t = 0$ by some (possibly random) quantity $X_0 = O_p(1)$, although other initialization assumptions may be considered, as in Phillips (2008).

A secondary contribution of the paper that emerges from the analysis is to provide a limit theory for semiparametric reduced rank regressions of the form (1.1) under weakly dependent errors. This limit theory is useful in studying cases where reduced rank regressions are misspecified, possibly through the choice of inappropriate lag lengths in the vector autoregression or incorrect settings of the cointegrating rank.

Under (1.1), the time series X_t is co-integrated with co-integration matrix β of rank r_0, so there are r_0 cointegrating relations in the true model. Of course, r_0 is generally unknown and our goal is to treat (1.1) semiparametrically with regard to u_t and to estimate r_0 directly in (1.1) by information criteria. The procedure we consider is quite simple. Model (1.1) is estimated by conventional reduced rank regression (RRR) for all values of $r = 0, 1, \ldots, m$ just as if u_t were a martingale difference, and r is chosen to optimize the corresponding information criteria as if (1.1) were a correctly specified parametric framework up to the order parameter r. Thus, no explicit account is taken of the weak dependence structure of u_t in the process.

Let $\widehat{\Sigma}(r)$ be the residual covariance matrix from the RRR. The criterion used to evaluate cointegrating rank takes the simple form

$$\mathrm{IC}(r) = \log \left|\widehat{\Sigma}(r)\right| + C_n n^{-1}(2mr - r^2), \tag{1.2}$$

with coefficient $C_n = \log n, 2 \log\log n$, or 2 corresponding to the BIC (Akaike, 1977, Rissanen, 1978, and Schwarz, 1978), Hannan and Quinn (1979) and Akaike (1974) penalties, respectively. Sample information-based versions of the coefficient C_n may also be employed, such as those in Wei's (1992) FIC criterion and Phillips and Ploberger's (1996) PIC criterion. The BIC version of (1.2) was given in Phillips and McFarland (1997) and used to determine cointegrating rank in an exchange rate application. In (1.2) the degrees of freedom term $2mr - r^2$ is calculated to account for the $2mr$ elements of the matrices α and β that have to be estimated, adjusted for the r^2 restrictions that are needed to ensure structural identification of β in reduced rank regression. The effects of other normalization schemes on the information criterion are discussed in the Appendix.

For each $r = 0, 1, \ldots, m$, we estimate the $m \times r$ matrices α and β by reduced rank regression, denoted by $\hat{\alpha}$ and $\hat{\beta}$, and, for use in (1.2), we form the corresponding residual variance matrices

$$\widehat{\Sigma}(r) = n^{-1} \sum_{t=1}^{n} \left(\Delta X_t - \hat{\alpha}\hat{\beta}' X_{t-1}\right) \left(\Delta X_t - \hat{\alpha}\hat{\beta}' X_{t-1}\right)', \quad r = 1, \ldots, m$$

with $\widehat{\Sigma}(0) = n^{-1} \sum_{t=1}^{n} \Delta X_t \Delta X_t'$. Model evaluation based on IC(r) then leads to the cointegrating rank selection criterion

$$\widehat{r} = \arg\min_{0 \le r \le m} \mathrm{IC}(r).$$

As shown below, the information criterion IC(r) is weakly consistent for selecting the cointegrating rank r_0 provided that the penalty term in (1.2) satisfies the weak requirements that $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$. No minimum expansion rate for C_n such as $\log \log n$ is required and no more complex parametric model needs to be estimated. The approach is therefore quite straightforward for practical implementation.

The organization of the paper is as follows. Some preliminaries on estimation and notation are covered in Section 2. The main asymptotic results are given in Section 3. Section 4 briefly reports some simulation findings. Section 5 concludes and discusses some extensions. Proofs and other technical material are in the Appendix.

2. PRELIMINARIES

Reduced rank regression estimates of α and β in (1.1) are obtained ignoring any weak dependence error structure in u_t. To analyze the asymptotic properties of the rank order estimates and the information criterion IC(r) under a general error structure, we start by investigating the asymptotic properties of the various regression components. Using conventional RRR notation, define

$$S_{00} = n^{-1} \sum_{t=1}^{n} \Delta X_t \Delta X_t', \;\; S_{11} = n^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}',$$
$$S_{01} = n^{-1} \sum_{t=1}^{n} \Delta X_t X_{t-1}', \;\text{ and } S_{10} = n^{-1} \sum_{t=1}^{n} X_{t-1} \Delta X_t'. \tag{2.1}$$

For some given r and β, the estimate of α is obtained by regression as

$$\widehat{\alpha}(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1}. \tag{2.2}$$

Again, given r, the corresponding RRR estimate of β in (1.1) is an $m \times r$ matrix satisfying

$$\widehat{\beta} = \arg\min_{\beta} \left| S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} \right|, \tag{2.3}$$

subject to the normalization

$$\widehat{\beta}' S_{11} \widehat{\beta} = I_r. \tag{2.4}$$

The estimate $\widehat{\beta}$ is found in the usual way by first solving the determinantal equation

$$\left| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} \right| = 0 \tag{2.5}$$

for the ordered eigenvalues $1 > \widehat{\lambda}_1 > \cdots > \widehat{\lambda}_m > 0$ and corresponding eigenvectors $\widehat{V} = [\widehat{v}_1, \ldots, \widehat{v}_m]$, which are normalized by $\widehat{V}' S_{11} \widehat{V} = I_m$. Estimates of β and α are then obtained as

$$\widehat{\beta} = [\widehat{v}_1, \ldots, \widehat{v}_r], \quad \text{and} \quad \widehat{\alpha} = \widehat{\alpha}(\widehat{\beta}) = S_{01}\widehat{\beta}, \tag{2.6}$$

with $\widehat{\beta}$ formed from the eigenvectors of $\widehat{V}$ corresponding to the r largest roots of (2.5). The residuals from the RRR and the corresponding moment matrix of residuals that appear in the information criterion are

$$\widehat{u}_t = \Delta X_t - \widehat{\alpha}\widehat{\beta}' X_{t-1}, \text{ and} \tag{2.7}$$

$$\widehat{\Sigma}(r) = n^{-1} \sum_{t=1}^{n} \widehat{u}_t \widehat{u}_t' = S_{00} - S_{01}\widehat{\beta}\widehat{\beta}' S_{10}. \tag{2.8}$$

Using (2.8) we have (e.g. Theorem 6.1 of Johansen, 1995)

$$\left|\widehat{\Sigma}(r)\right| = |S_{00}| \, \Pi_{i=1}^{r}(1 - \widehat{\lambda}_i), \tag{2.9}$$

where $\widehat{\lambda}_i$, $1 \leq i \leq r$, are the r largest solutions to (2.5). The criterion (1.2) is then well determined for any given value of r.

3. ASYMPTOTIC RESULTS

The following assumptions make specific the semiparametric and co-integration components of (1.1). Assumption **LP** is a standard linear process condition of the type that is convenient in developing partial sum limit theory. The condition can be relaxed to allow for martingale difference innovations and to allow for some mild heterogeneity in the innovations without disturbing the limit theory in a material way (see Phillips and Solo, 1992). Assumption **RR** gives conditions that are standard in the study of reduced rank regressions with some unit roots (Johansen, 1988, 1995, and Phillips, 1995).

ASSUMPTION **LP**. *Let $D(L) = \sum_{j=0}^{\infty} D_j L^j$, with $D_0 = I$ and full rank $D(1)$, and let u_t have Wold representation*

$$u_t = D(L)\varepsilon_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j}, \quad \text{with} \quad \sum_{j=0}^{\infty} j^{1/2} ||D_j|| < \infty, \tag{3.1}$$

for some matrix norm $||\cdot||$ and where ε_t is i.i.d. $(0, \Sigma_\varepsilon)$ *with $\Sigma_\varepsilon > 0$. We use the notation $\Gamma_{ab}(h) = E(a_t b_{t+h}')$ and $\Lambda_{ab} = \sum_{h=1}^{\infty} \Gamma_{ab}(h)$ for autocovariance matrices and one sided long-run autocovariances and set $\Omega = \sum_{h=-\infty}^{\infty} \Gamma_{uu}(h) = D(1)\Sigma_\varepsilon D(1)' > 0$ and $\Sigma_\varepsilon = E\{\varepsilon_t \varepsilon_t'\}$.*

ASSUMPTION **RR**.

- *(a) The determinantal equation $|I_m - (I_m + \alpha\beta')L| = 0$ has roots on or outside the unit circle, i.e. $|L| \geq 1$.*
- *(b) Set $\Pi = I_m + \alpha\beta'$ where α and β are $m \times r_0$ matrices of full column rank r_0, $0 \leq r_0 \leq m$. (If $r_0 = 0$ then $\Pi = I_m$; if $r_0 = m$ then β has full rank m and $\beta' X_t$ and hence X_t are (asymptotically) stationary)*
- *(c) The matrix $R = I_r + \beta'\alpha$ has eigenvalues within the unit circle.*

Assumption (c) ensures that the matrix $\beta'\alpha$ has full rank. Let $\alpha_\perp$ and $\beta_\perp$ be orthogonal complements to α and β, so that $[\alpha, \alpha_\perp]$ and $[\beta, \beta_\perp]$ are non-singular and $\beta_\perp'\beta_\perp = I_{m-r}$. Then, non-singularity of $\beta'\alpha$ implies the non-singularity of $\alpha_\perp'\beta_\perp$. Under **RR** we have the Wold representation of $\beta' X_t$

$$v_t := \beta' X_t = \sum_{i=0}^{\infty} R^i \beta' u_{t-i} = R(L)\beta' u_t = R(L)\beta' D(L)\varepsilon_t, \tag{3.2}$$

and some further manipulations yield the following useful partial sum representation

$$X_t = C\sum_{s=1}^{t} u_s + \alpha(\beta'\alpha)^{-1}R(L)\beta' u_t + CX_0, \tag{3.3}$$

where $C = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'$. Expression (3.3) reduces to the Granger representation when u_t is a martingale difference (e.g. Johansen, 1995).

Under **LP** a functional law for partial sums of u_t holds, so that $n^{-1/2}\sum_{s=1}^{[n\cdot]} u_s \Rightarrow B_u(\cdot)$ as $n \to \infty$, where B_u is vector Brownian motion with variance matrix Ω. In view of (3.2) and the fact that $R(1) = \sum_{i=0}^{\infty} R^i = (I - R)^{-1} = -(\beta'\alpha)^{-1}$, we further have

$$n^{-1/2}\sum_{s=1}^{[n\cdot]} v_s = n^{-1/2}\sum_{s=1}^{[n\cdot]} \beta' X_s \Rightarrow -(\beta'\alpha)^{-1}\beta' B_u(\cdot), \quad \text{as } n \to \infty. \tag{3.4}$$

These limit laws involve the same Brownian motion B_u and determine the asymptotic forms of the various sample moment matrices involved in the reduced rank regression estimation of (1.1).

Define

$$\operatorname{Var}\begin{bmatrix} \Delta X_t \\ \beta' X_{t-1} \end{bmatrix} = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}. \tag{3.5}$$

Explicit expressions for the submatrices in this expression may be worked out in terms of the autocovariance sequences of u_t and v_t and the parameters of (1.1). These expressions are given in (A.2)–(A.4) in the proof of Lemma A1 in the Appendix. The following result provides some asymptotic limits that are useful in deriving the asymptotic properties of $\widehat{\Sigma}(r)$ in the criterion function (1.2).

LEMMA 3.1. *Under Assumptions **LP** and **RR**,*

$$S_{00} \to_p \Sigma_{00},\ \beta' S_{11}\beta \to_p \Sigma_{\beta\beta},\ \beta' S_{10} \to_p \Sigma_{\beta 0},$$

$$n^{-1}\beta_\perp' S_{11}\beta_\perp \Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left(\int_0^1 B_u B_u'\right)\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1},$$

$$\beta_\perp'(S_{10} - S_{11}\beta\alpha') \Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u' + \Psi_{wu}^1,$$

$$\beta_\perp' S_{11}\beta \Rightarrow -(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u'\beta(\alpha'\beta)^{-1} + \Psi_{wv},$$

$$\beta_\perp' S_{10} \Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u'\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1}\beta_\perp' + \Psi_{wu}^1 + \Psi_{wv}\alpha',$$

where

$$\Psi_{wu}^{1} = \sum_{h=1}^{\infty} E\{(\beta_{\perp}'\Delta X_t)u_{t+h}'\}, \qquad \Psi_{wv} = \sum_{h=0}^{\infty} E\{(\beta_{\perp}'\Delta X_t)(\beta' X_{t+h})'\},$$

and $w_t = \beta_{\perp}'\Delta X_t = \beta_{\perp}' u_t + \beta_{\perp}'\alpha v_{t-1}$.

REMARK 3.1.

(a) When u_t is weakly dependent, it is apparent that the asymptotic limits of $\beta_{\perp}'(S_{10} - S_{11}\beta\alpha')$, $\beta_{\perp}' S_{10}$, and $\beta_{\perp}' S_{11}\beta$ involve bias terms that depend on various one sided long-run covariance matrices associated with the stationary components u_t, v_t, and $w_t = \beta_{\perp}'\Delta X_t$. Explicit values of these one-sided long-run covariance matrices are given in (A.7) and (A.8) in the Appendix.

(b) When u_t is a martingale difference sequence, $\Psi_{wu}^{1} = 0$ and $\Psi_{wv} = \beta_{\perp}' E(u_t v_t')$. Simpler results, such as

$$\beta_{\perp}'(S_{10} - S_{11}\beta\alpha') \Rightarrow (\alpha_{\perp}'\beta_{\perp})^{-1}\alpha_{\perp}' \int_0^1 B_u dB_u', \tag{3.6}$$

then hold for the limits in Lemma 3.1, and these correspond to earlier results given for example in theorem 10.3 of Johansen (1995).

From (1.1), $\Delta X_t = u_t + \alpha\beta' X_{t-1} = u_t + \alpha v_{t-1}$, so that (c.f. (A.3)–(A.4)) $\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta} + Eu_t v_{t-1}'$ and

$$\Sigma_{00} = \alpha\Sigma_{\beta 0} + Eu_t v_{t-1}'\alpha' + E(u_t u_t'). \tag{3.7}$$

Define

$$\widetilde{\alpha} = \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1} = \alpha + Eu_t v_{t-1}'\Sigma_{\beta\beta}^{-1}, \tag{3.8}$$

and let $\widetilde{\alpha}_{\perp}$ be an $m \times (m - r)$ orthogonal complement to $\widetilde{\alpha}$ such that $[\widetilde{\alpha}, \widetilde{\alpha}_{\perp}]$ is non-singular.

LEMMA 3.2. *Under Assumptions **LP** and **RR**, when the true co-integration rank is r_0, the $r_0$largest solutions to* (2.5)*, denoted by $\widehat{\lambda}_i$ with $1 \le i \le r_0$, converge to the roots of*

$$\left|\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right| = 0. \tag{3.9}$$

The remaining $m - r_0$ roots, denoted by $\widehat{\lambda}_i$ with $r_0 + 1 \le i \le m$, decrease to zero at the rate n^{-1} and $\{n\widehat{\lambda}_i : i = r_0 + 1, \ldots, m\}$ converge weakly to the roots of

$$\left|\rho\int_0^1 G_u G_u' - \left(\int_0^1 G_u dG_u'\beta_{\perp}' + \Psi\right)\widetilde{\alpha}_{\perp}\left(\widetilde{\alpha}_{\perp}'\Sigma_{00}\widetilde{\alpha}_{\perp}\right)^{-1}\widetilde{\alpha}_{\perp}'\left(\beta_{\perp}\int_0^1 dG_u G_u' + \Psi'\right)\right| = 0, \tag{3.10}$$

where $G_u(r) = (\alpha_{\perp}'\beta_{\perp})^{-1}\alpha_{\perp}' B_u(r)$ is $m - r_0$ dimensional Brownian motion with variance matrix $(\alpha_{\perp}'\beta_{\perp})^{-1}\alpha_{\perp}'\Omega\alpha_{\perp}(\beta_{\perp}'\alpha_{\perp})^{-1}$ and $\Psi = \Psi_{wu}^{1} + \Psi_{wv}\alpha'$.

REMARK 3.2.

(a) Comparing these results with those of the standard RRR case with martingale difference errors (Johansen, 1995, p. 158), we see that, just as in the standard case, the r_0 largest roots of (2.5) are all positive in the limit and the $m - r_0$ smallest roots converge to 0 at the rate n^{-1}, both results now holding under weakly dependent errors. However, when u_t

is weakly dependent, the limit distribution determined by (3.10) is more complex than in the standard case. In particular, the determinantal equation (3.10) involves the composite one-sided long-run covariance matrix Ψ.

(b) When u_t is a martingale difference sequence, we find that $\widetilde{\alpha} = \alpha$, $\widetilde{\alpha}_\perp = \alpha_\perp$, $\Psi^1_{wu} = 0$, $\Psi = \Psi_{wv}\alpha'$, $\Sigma_{00} = \alpha\Sigma_{\beta\beta}\alpha' + \Omega$ and

$$\widetilde{\alpha}_\perp \left(\widetilde{\alpha}'_\perp \Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1} \widetilde{\alpha}'_\perp = \alpha_\perp \left(\alpha'_\perp \Omega\alpha_\perp\right)^{-1} \alpha'_\perp.$$

Then $\alpha'_\perp \beta_\perp G_u(r) = \alpha'_\perp B_u(r)$ is Brownian motion with covariance matrix $\alpha'_\perp \Omega\alpha_\perp$, $\Psi\alpha_\perp = 0$, and the determinantal equation (3.10) reduces to

$$\left|\rho \int_0^1 G_u G'_u - \int_0^1 G_u dG'_u \beta'_\perp \alpha_\perp \left(\alpha'_\perp \Omega\alpha_\perp\right)^{-1} \alpha'_\perp \beta_\perp \int_0^1 dG_u G'_u\right| = 0,$$

which is equivalent to

$$\left|\rho \int_0^1 V_u V'_u - \int_0^1 V_u dV'_u \int_0^1 dV_u V'_u\right| = 0,$$

where $V_u(r)$ is $m - r_0$ dimensional standard Brownian motion, thereby corresponding to the standard limit theory of a parametric reduced rank regression (Johansen, 1995).

THEOREM 3.1.

*(a) Under Assumptions **LP** and **RR**, the criterion IC(r) is weakly consistent for selecting the rank of co-integration provided $C_n \to \infty$ at a slower rate than n.*

(b) The asymptotic distribution of the AIC criterion (IC(r) with coefficient $C_n = 2$) is given by

$$\lim_{n\to\infty} P\left(\hat{r}_{AIC} = r_0\right) = P\left[\bigcap_{r=r_0+1}^{m} \left\{\sum_{i=r_0+1}^{r} \xi_i < 2(r - r_0)(2m - r - r_0)\right\}\right],$$

$$\lim_{n\to\infty} P\left(\hat{r}_{AIC} = r|r > r_0\right) = P\left\{\left(\bigcap_{r'=r+1}^{m} \left\{\sum_{i=r+1}^{r'} \xi_i < 2(r' - r)(2m - r' - r)\right\}\right) \cap \left(\bigcap_{r'=r_0}^{r-1} \left\{\sum_{i=r'+1}^{r} \xi_i > 2\left(r - r'\right)\left(2m - r - r'\right)\right\}\right)\right\},$$

and

$$\lim_{n\to\infty} P\left(\hat{r}_{AIC} = r|r < r_0\right) = 0,$$

where $\xi_{r_0+1}, \ldots, \xi_m$ are the ordered roots of the limiting determinantal equation (3.10).

REMARK 3.3.

(a) BIC, HQ and other information criteria with $C_n \to \infty$ and $C_n/n \to 0$ are all consistent for the selection of cointegrating rank without having to specify a full parametric model.

The same is true for the criterion $\mathrm{IC}^*(r)$ where only the cointegrating space is estimated and structural identification conditions on the cointegrating vector are not imposed or used in rank selection.

(b) AIC is inconsistent, asymptotically never underestimates cointegrating rank, and favours more liberally parametrized systems. This outcome is analogous to the well-known overestimation tendency of AIC in lag length selection in autoregression. Of course, in the present case, maximum rank is bounded above by the order of the system. Thus, the advantages to overestimation in lag length selection that arise when the autoregressive order is infinite might not be anticipated here. However, when cointegrating rank is high (and close to full dimensional), AIC typically performs exceedingly well (as simulations reported below attest) largely because the upper bound in rank restricts the tendency to overestimate.

(c) When $m = 1$, $r_0 = 0$ corresponds to the unit root case and $r_0 = 1$ to the stationary case. Thus, one specialization of the above result is to unit root testing. In this case, the criteria consistently discriminate between unit root and stationary series provided $C_n \to \infty$ and $C_n/n \to 0$, as shown in Phillips (2008). In this case, the limit distribution of AIC is much simpler and involves only the explicit limiting root $\xi_1 = (\int_0^1 B_u dB_u + \lambda)^2/\{(\int_0^1 B_u^2)\Sigma_{00}\}$ where $\lambda = \sum_{h=1}^{\infty} E(u_t u_{t+h})$.

(d) While Theorem 3.1 relates directly to model (1.1), it is easily shown to apply in cases where the model has intercepts and drift. Thus, the result provides a convenient basis for consistent co-integration rank selection in most empirical contexts.

4. SIMULATIONS

Simulations were conducted to evaluate the finite sample performance of the criteria under various generating mechanisms for the short memory component u_t, different settings for the true cointegrating rank, and for various choices of the penalty coefficient C_n. Some illustrative findings for cases of dimension $m = 2$ and 4 are reported here.

The data generating process follows (1.1). When $m = 2$, the design is as follows. For $r_0 = 0$ we have $\alpha'\beta = 0$. For $r_0 = 1$ the reduced rank coefficient structure is set so that

$$\alpha'\beta = R_1 = (1, 0.5)\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

For $r_0 = 2$, two different designs (A and B) were simulated, one with smaller and the other with larger stationary roots as follows:

$$\text{A: } \alpha'\beta = R_2 = \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.4 \end{pmatrix}, \quad \text{with stationary roots } \lambda_i[I + \beta'\alpha] = \{0.7, 0.4\};$$

$$\text{B: } \alpha'\beta = R_3 = \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.15 \end{pmatrix}, \quad \text{with stationary roots } \lambda_i[I + \beta'\alpha] = \{0.9, 0.45\}.$$

When the dimension $m = 4$, the matrix $\alpha'\beta$ was constructed to have a block diagonal form reflecting the true cointegrating rank. We call the four dimensional set up design C in what follows. For $r_0 = 0$ we have $\alpha'\beta = 0$. Let

$$R_4 = \begin{pmatrix} 2 \\ 0.5 \end{pmatrix} (-1, 1) \text{ and } R_5 = \begin{pmatrix} -0.7 & 0.1 \\ 0.2 & -0.6 \end{pmatrix}.$$

For $r_0 = 1$, the reduced rank coefficient structure is set to

$$\alpha'\beta = diag\{R_{4,} 0, 0\}, \quad \text{with stationary root } \lambda_i[I + \beta'\alpha] = -0.5.$$

For $r_0 = 2$,

$$\alpha'\beta = diag\{R_{5,} 0, 0\}, \quad \text{with stationary roots } \lambda_i[I + \beta'\alpha] = \{0.2, 0.5\}.$$

For $r_0 = 3$,

$$\alpha'\beta = diag\{R_{2,} R_4\}, \quad \text{with stationary roots } \lambda_i[I + \beta'\alpha] = \{0.4, 0.7, -0.5\}.$$

For $r_0 = 4$,

$$\alpha'\beta = diag\{R_{2,} R_3\}, \quad \text{with stationary roots } \lambda_i[I + \beta'\alpha] = \{0.4, 0.7, 0.45, 0.9\}.$$

Simulations were conducted with AR(1), MA(1) and ARMA(1,1) errors, corresponding to the models

$$u_t = Au_{t-1} + \varepsilon_t,\ u_t = \varepsilon_t + B\varepsilon_{t-1} \quad \text{and } u_t = Au_{t-1} + \varepsilon_t + B\varepsilon_{t-1}, \tag{4.1}$$

respectively, with coefficient matrices $A = \psi I_m$, $B = \phi I_m$, where $|\psi| < 1$, $|\phi| < 1$, and with innovations $\varepsilon_t =$ i.i.d. $N(0, \Sigma_\varepsilon)$, where

$$\Sigma_\varepsilon = diag\{1 + \theta, 1 - \theta\} > 0 \text{ when } m = 2 \text{ and}$$
$$\Sigma_\varepsilon = diag\{1 + \theta_1, 1 - \theta_1, 1 + \theta_2, 1 - \theta_2\} \text{ when } m = 4.$$

The parameters for these models were set to $\psi = \phi = 0.4$, $\theta = 0.25$, $\theta_1 = 0.25$ and $\theta_2 = 0.4$.

The performance of the criteria AIC, BIC, HQ and log(HQ) was investigated for sample sizes $n = 100$ in design A and $n = 100$, 400 in design B and $n = 100$, 250 and 400 in design C.[1] All cases including 50 additional observations to eliminate start-up effects from the initializations $X_0 = 0$ and $\varepsilon_0 = 0$. The results are based on 20,000 replications and are summarized in Fig. 1, which shows the results for design A, in Table 1, which shows the results for design B, and in Table 2, which shows the results for design C, with correct selections in bold type. The results displayed are for the model with AR(1) errors. Similar results were obtained for the other error generating schemes in (4.1).

As is evident in Fig. 1, the BIC criterion generally performs very well when $n \geq 100$. For design B, where the stationary roots of the system are closer to unity, BIC has a tendency to underestimate rank when $n = 100$ and $r_0 = 2$, thereby choosing more parsimoniously parameterized systems in this case, just as it does in lag length selection in autoregressions. But BIC performs well when $n = 400$, as seen in Table 1.

The tendency of AIC to overestimate rank is also clear in Fig. 1, but this tendency is noticeably attenuated when the true rank is 1 and is naturally delimited when the true rank is 2 because of the upper bound in rank choice. For design B, AIC performs better than BIC when

[1] log(HQ) has penalty coefficient $C_n = \log(2 \log\log n)$.

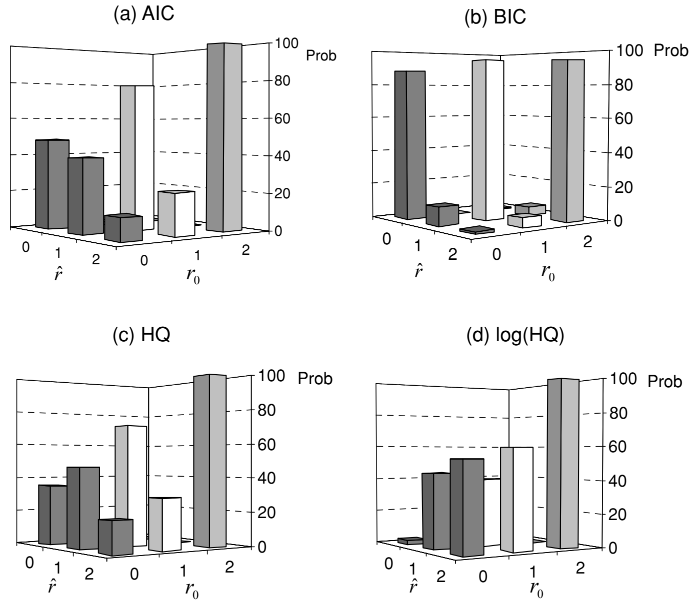

Figure 1. Cointegrating rank selection in design A when u_t is AR(1) and $n = 100$.

the cointegrating rank is 2 and the system is stationary, as does HQ, for which the penalty is $C_n < 2$ when $n = 100, 400$. Criteria with weaker penalties, such as log(HQ) with $C_n = \log(2 \log \log n)$, also do better in this case, although for other cases they perform much less satisfactorily than AIC and HQ, showing a strong tendency to overestimate cointegrating rank.

Design C is a more extensive set-up for the higher dimensional system with $m = 4$. As shown in Table 2, BIC generally performs well when $n = 250$ and the pattern follows that of the two dimensional set-up. When $r_0 = 4$ and the system is stationary, BIC tends to underestimate cointegrating rank because one stationary root is close to unity. BIC also performs better when some stationary roots are negative than it does when all stationary roots are positive. We found, for example, that when the cointegrating rank is 3 with the three positive stationary roots $\{0.4, 0.5, 0.7\}$ BIC can perform poorly even for $n = 250$. However, if the distribution of the stationary roots is more balanced—for example if one of the roots is negative as in $\{-0.5, 0.4, 0.7\}$ in Design C—performance improves significantly.

Based on overall performance, it seems that BIC can be recommended for practical work in choosing cointegrating rank and it gives generally very sharp results when $n \geq 250$. The main weakness of BIC is its tendency to choose more parsimonious models (i.e. models with more unit roots) in the following conditions: (i) when the system is stationary and has a root near unity; (ii) when there are many positive stationary roots and (iii) when the sample size is small and the system dimension is large.[2]

Wang and Bessler (2005) reported some related simulation work under the assumption that it is known that the time series are already transformed into a form where the observed variables are

[2] Simulations (not reported here) showed that the tendency for BIC to select models with more unit roots is exacerbated when the criterion IC* (r), which has a stronger penalty, is used.

Table 1. Cointegrating rank selection in design B when u_t follows an AR(1) process.

	$n = 100$			$n = 400$		
$r_0 = 0$						
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$
AIC	**0.48**	0.40	0.12	**0.53**	0.36	0.11
BIC	**0.88**	0.11	0.01	**0.95**	0.05	0.00
HQ	**0.35**	0.47	0.18	**0.47**	0.40	0.13
Log(HQ)	**0.03**	0.44	0.53	**0.07**	0.49	0.44
$r_0 = 1$						
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$
AIC	0.00	**0.78**	0.22	0.00	**0.76**	0.24
BIC	0.00	**0.94**	0.06	0.00	**0.97**	0.03
HQ	0.00	**0.71**	0.29	0.00	**0.74**	0.26
Log(HQ)	0.00	**0.40**	0.60	0.00	**0.46**	0.54
$r_0 = 2$						
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$
AIC	0.00	0.25	**0.75**	0.00	0.00	**1.00**
BIC	0.05	0.74	**0.21**	0.00	0.02	**0.98**
HQ	0.00	0.14	**0.86**	0.00	0.00	**1.00**
Log(HQ)	0.00	0.02	**0.98**	0.00	0.00	**1.00**

either stationary or integrated. In the present context, this is equivalent to setting $\alpha\beta'$ to a diagonal matrix with elements of either zero or unity. The problem of cointegrating rank selection in this simpler framework is equivalent to direct unit root testing on each variable. We may therefore use the selection method of Phillips (2008) to estimate the co-integration rank by conducting a unit root test on each time series and simply counting the number of unit roots obtained. Simulations (not reported here) indicate that this procedure works well. However, since the transformation which takes the model into a canonical form where the observed variables are either stationary or integrated is seldom known, this procedure is generally not practical for estimating cointegrating rank.

5. CONCLUSION

Model selection for cointegrating rank treats rank as an order parameter and provides the convenience of consistent estimation of this parameter under weak conditions on the expansion rate of the penalty coefficient. The approach is easy to implement in practice and is sympathetic with other semiparametric approaches to estimation and inference in cointegrating systems where the focus is on long-run behaviour. Information criteria such as (1.2) provide a useful

Table 2. Cointegrating rank selection in design C when u_t follows an AR(1) process.

$n = 250$					
$r_0 = 0$					
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 3$	$\widehat{r} = 4$
AIC	**0.13**	0.40	0.31	0.12	0.04
BIC	**0.94**	0.06	0.00	0.00	0.00
HQ	**0.58**	0.33	0.08	0.01	0.00
Log(HQ)	**0.00**	0.09	0.31	0.40	0.20
$r_0 = 1$					
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 3$	$\widehat{r} = 4$
AIC	0.00	**0.34**	0.43	0.18	0.05
BIC	0.00	**0.96**	0.04	0.00	0.00
HQ	0.00	**0.75**	0.21	0.03	0.00
Log(HQ)	0.00	**0.05**	0.30	0.45	0.20
$r_0 = 2$					
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 3$	$\widehat{r} = 4$
AIC	0.00	0.00	**0.54**	0.36	0.10
BIC	0.00	0.02	**0.93**	0.05	0.00
HQ	0.00	0.00	**0.80**	0.17	0.03
Log(HQ)	0.00	0.00	**0.23**	0.50	0.26
$r_0 = 3$					
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 3$	$\widehat{r} = 4$
AIC	0.00	0.00	0.00	**0.82**	0.18
BIC	0.00	0.00	0.10	**0.88**	0.02
HQ	0.00	0.00	0.00	**0.93**	0.07
Log(HQ)	0.00	0.00	0.00	**0.66**	0.34
$r_0 = 4$					
	$\widehat{r} = 0$	$\widehat{r} = 1$	$\widehat{r} = 2$	$\widehat{r} = 3$	$\widehat{r} = 4$
AIC	0.00	0.00	0.00	0.00	**1.00**
BIC	0.00	0.00	0.04	0.16	**0.80**
HQ	0.00	0.00	0.00	0.02	**0.98**
Log(HQ)	0.00	0.00	0.00	0.00	**1.00**

diagnostic check on system cointegrating rank and proceed as if there is no prior information on cointegrating rank. If prior information were available and could be formulated as prior probabilities on the models of different rank then this could, of course, be incorporated into a Bayes factor.

In subsequent work Cheng and Phillips (2008) show that consistent cointegrating rank selection by information criteria continues to hold in models where there is unconditional heterogeneity in the error variance of unknown form, including breaks in the variance or smooth transition functions in the variance over time. Such permanent changes in variance are known to invalidate both unit root tests and likelihood ratio tests for cointegrating rank because of their effects on the limit distribution theory under the null (see Cavaliere, 2004, Beare, 2007, and Cavaliere and Taylor, 2007). Since consistency of the information criteria is unaffected by the presence of this form of variance induced non-stationarity, the approach offers an additional degree of robustness in cointegrating rank determination that is useful in empirical applications. Cheng and Phillips (2008) give an empirical application of this theory to exchange rate dynamics.

Some applications of the methods outlined here are possible in other models. First, rather than work with reduced rank regression formulations within a vector autoregressive framework, it is possible to use reduced rank formulations in regressions of the time series on a fixed (or expanding) number of deterministic basis functions such as time polynomials or sinusoidal polynomials (Phillips, 2005). In a similar way to the present analysis, it can be shown that information criteria such as BIC and HQ will be consistent for cointegrating rank in such coordinate systems. The coefficient matrix in such systems turns out to have a random limit, corresponding to the matrix of random variables that appear in the Karhunen–Loève representation (Phillips, 1998), but has a rank that is the same as the dimension of the cointegrating space, which enables consistent rank estimation by information criteria. A second application is to dynamic factor panel models with a fixed number of stochastically trending unobserved factors, as in Bai and Ng (2004). Again, these models have reduced rank structure (this time with non-random coefficients) and the number of factors may be consistently estimated using model selection criteria of the same type as those considered here but in the presence of an increasing number of incidental loading coefficients. In such cases, the BIC penalty, as derived from the asymptotic behaviour of the Bayes factor, has a different form from usual and typically involves both cross section and time series sample sizes. Some extensions of the present methods to these models will be reported in later work.

ACKNOWLEDGMENTS

Our thanks go to the Editor and a referee for helpful comments on the original version. Cheng acknowledges support from an Anderson Fellowship. Phillips acknowledges support from the NSF under Grant No. SES 06-47086.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory*, 267–281. Budapest: Akademiai Kiado.

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnarah (Ed.), *Applications of Statistics*, 27–41. Amsterdam: North–Holland.

Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica 72*, 1127–77.

Beare, B. (2007). Robustifying unit root tests to permanent changes in innovation variance. Working paper, Yale University.

Cavaliere, G. (2004). Unit root tests under time-varying variance shifts. *Econometric Reviews 23*, 259–92.

Cavaliere, G. and A. M. R. Taylor (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics 140*, 919–47.

Chao, J. and P. C. B. Phillips (1999). Model selection in partially non-stationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics 91*, 227–71.

Cheng, X. and P. C. B. Phillips (2008). Cointegrating rank selection in models with time-varying variance. Working paper, Yale University.

Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, Series B *41*, 190–5.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*, 231–54.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Kapetanios, G. (2004). The asymptotic distribution of the cointegration rank estimation under the Akaike information criterion. *Econometric Theory 20*, 735–42.

Nielsen, B. (2006). Order determination in general vector autoregressions. *IMS Lecture Notes- Monograph Series 52*, 93–112.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica 59*, 283–306.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica 63*, 1023–78.

Phillips, P. C. B. (1996). Econometric model determination. *Econometrica 64*, 763–812.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica 66*, 1299–326.

Phillips, P. C. B. (2005). Challenges of trending time series econometrics. *Mathematics and Computers in Simulation 68*, 401–16.

Phillips, P. C. B. (2008). Unit root model selection. *Journal of the Japan Statistical Society 38*, 65–74.

Phillips, P. C. B. and J. McFarland (1997). Forward exchange market unbiasedness: the case of the Australian dollar since 1984. *Journal of International Money and Finance 16*, 885–907.

Phillips P. C. B. and W. Ploberger (1996). An asymptotic theory of bayesian inference for time series. *Econometrica 64*, 381–413.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *Annals of Statistics 20*, 971–1001.

Pötscher, B. M. (1989). Model selection under nonstationarity: autoregressive models and stochastic linear regression models. *Annals of Statistics 17*, 1257–74.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica 14*, 465–71.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–4.

Tsay, R. S. (1984). Order selection in nonstationary autoregressive models. *Annals of Statistics 12*, 1425–33.

Wang, Z. and D. A. Bessler (2005). A Monte Carlo Study on the selection of cointegration rank using information criteria. *Econometric Theory 21*, 593–620.

Wei, C. Z. (1992). On predictive least squares principles. *Annals of Statistics 20*, 1–42.

APPENDIX

A.1. Normalization restrictions and degrees of freedom

In triangular system specifications (Phillips, 1991) the cointegrating matrix β' in (1.1) takes the form $[I_r, -B]$ for some unrestricted $r \times (m - r)$ matrix B, which involves r^2 restrictions and leads to degrees of freedom $2mr - r^2$ in A. Under normalization restrictions of the form (2.4) on β that are conventionally employed in empirical reduced rank regression modeling, the degrees of freedom term would be $2mr - r(r + 1)/2$, leading to the alternate criterion

$$\text{IC}^*(r) = \log|\widehat{\Sigma}(r)| + C_n n^{-1}(2mr - r(r + 1)/2).$$

In this case the outer product form of the coefficient matrix in (1.1) implies that $A = \alpha\beta' = \alpha CC'\beta'$ for an arbitrary orthogonal matrix C, so that α and β are not uniquely identified even though the likelihood is well defined. In such cases, only the cointegrating rank and the cointegrating space are identified and consistently estimable.

Correspondingly, under this normalization there are more degrees of freedom in the system. However, the usual justification for the BIC criterion (Schwarz, 1978 and Ploberger and Phillips, 1996) involves finding an asymptotic approximation to the Bayesian data density (and hence the posterior probability of the model), which is obtained by Laplace approximation methods using a Taylor series expansion of the log likelihood around a consistent parameter estimate. In the reduced rank regression case, r^2 restrictions on β are required to identify the structural parameters as in the above formulation $\beta' = [I_r, -B]$. If only normalization restrictions such as $\beta'\beta = I_r$ are imposed, then we can write

$$A = \alpha\beta' = \alpha CC'(I_r + BB')^{-1/2}[I_r, -B],$$

with $\beta' = (I_r + BB')^{-1/2}[I_r, -B]$ and where C is an arbitrary orthogonal matrix. In this case, C is unidentified and if C has a uniform prior distribution on the orthogonal group $O(r)$ independent of the prior on (α, B), then C may be integrated out of the Bayesian data density or marginal likelihood. The data density then has the same form as it does for the case where

$$A = \alpha(I_r + BB')^{-1/2}[I_r, -B] = \bar{\alpha}^{-1/2}[I_r, -B],$$

and where $\bar{\alpha}$ and B are identified. In this event, the model selection criterion is the same as that given in (1.2).

A.2. Proofs

LEMMA A1. *Under (1.1) and Assumption* ***LP****,*

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}\left(\Sigma_{\beta0}\Sigma_{00}^{-1}\Sigma_{\beta0}\right)^{-1}\Sigma_{\beta0}\Sigma_{00}^{-1} = \Sigma_{00}^{-1/2}c_\perp(c_\perp'c_\perp)^{-1}c_\perp'\Sigma_{00}^{-1/2},$$

where $c = \Sigma_{00}^{-1/2}\Sigma_{0\beta}$ *and* $c_\perp$ *is an orthogonal complement to* c*. Defining* $\widetilde{\alpha} = \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1} = \Sigma_{00}^{1/2}c\Sigma_{\beta\beta}^{-1}$ *and* $\widetilde{\alpha}_\perp = \Sigma_{00}^{-1/2}c_\perp$*, we have the alternate form*

$$\Sigma_{00}^{-1/2}c_\perp(c_\perp'c_\perp)^{-1}c_\perp'\Sigma_{00}^{-1/2} = \widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp'. \tag{A.1}$$

When $\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta}$*, (A.1) reduces to*

$$\Sigma_{00}^{-1/2}c_\perp(c_\perp'c_\perp)^{-1}c_\perp'\Sigma_{00}^{-1/2} = \alpha_\perp\left(\alpha_\perp'\Sigma_{00}\alpha_\perp\right)^{-1}\alpha_\perp.$$

Proof: Since $[c, c_\perp]$ is non-singular we have

$$I = c(c'c)^{-1}c' + c_\perp(c_\perp'c_\perp)^{-1}c_\perp',$$

and then

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{\beta 0})^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1}$$
$$= \Sigma_{00}^{-1/2}\{I - c(c'c)^{-1}c'\}\Sigma_{00}^{-1/2} = \Sigma_{00}^{-1/2}c_\perp(c_\perp' c_\perp)^{-1}c_\perp'\Sigma_{00}^{-1/2},$$

as required.

Observe that when $\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta}$, we have $c = \Sigma_{00}^{-1/2}\Sigma_{0\beta} = \Sigma_{00}^{-1/2}\alpha\Sigma_{\beta\beta}$ and we may choose $c_\perp = \Sigma_{00}^{1/2}\alpha_\perp$ where $\alpha_\perp$ is an orthogonal complement to α. In that case we have

$$\Sigma_{00}^{-1/2}c_\perp(c_\perp' c_\perp)^{-1}c_\perp'\Sigma_{00}^{-1/2} = \alpha_\perp\left(\alpha_\perp'\Sigma_{00}\alpha_\perp\right)^{-1}\alpha_\perp,$$

as stated. This corresponds with the result in Johansen (1995, lemma 10.1) where u_t is a martingale difference. In the present semiparametric case, $\Delta X_t = u_t + \alpha\beta' X_{t-1} = u_t + \alpha v_{t-1}$ and the covariance $\Gamma_{vu}(1) = E(v_{t-1}u_t)'$ is generally non-zero, so that

$$\Sigma_{\beta\beta} = Ev_t v_t' = \Gamma_{vv}(0), \tag{A.2}$$

$$\Sigma_{\beta 0} = Ev_{t-1}(u_t + \alpha v_{t-1})' = \Gamma_{vu}(1) + \Gamma_{vv}(0)\alpha' = \Gamma_{vu}(1) + \Sigma_{\beta\beta}\alpha', \tag{A.3}$$

$$\Sigma_{00} = \alpha\Sigma_{\beta\beta}\alpha' + \alpha\Gamma_{vu}(1) + \Gamma_{uv}(-1)\alpha' + \Gamma_{uu}(0). \tag{A.4}$$

Note that $\widetilde{\alpha} = \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1} = \Sigma_{00}^{1/2}c\Sigma_{\beta\beta}^{-1}$ and we may choose $\widetilde{\alpha}_\perp = \Sigma_{00}^{-1/2}c_\perp$. In this notation, we may write in the general case

$$\Sigma_{00}^{-1/2}c_\perp(c_\perp' c_\perp)^{-1}c_\perp'\Sigma_{00}^{-1/2} = \widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp', \tag{A.5}$$

as given in (A.1). □

Proof of Lemma 3.1: Since both $\Delta X_t = u_t + \alpha\beta' X_{t-1}$ and $v_t = \beta' X_t$ are stationary and satisfy Assumption LP, the law of large numbers gives

$$S_{00} = n^{-1}\sum_{t=1}^{n}\Delta X_t\Delta X_t' \to_p \Sigma_{00} = \Gamma_{uu}(0) + \alpha\Gamma_{vv}(0)a' + \alpha\Gamma_{vu}(0) + \Gamma_{uv}(0)a',$$

$$\beta' S_{11}\beta = n^{-1}\sum_{t=1}^{n}\beta' X_{t-1}\left(\beta' X_{t-1}\right)' \to_p \Sigma_{\beta\beta} = \Gamma_{vv}(0), \text{ and}$$

$$\beta' S_{10} = n^{-1}\sum_{t=1}^{n}\beta' X_{t-1}\Delta X_t' \to_p \Sigma_{\beta 0} = \Gamma_{vu}(1) + \Gamma_{vv}(0)a'.$$

In view of (3.3) we have

$$\beta_\perp' X_t = \beta_\perp' C\sum_{s=1}^{t}u_s + \beta_\perp'\alpha(\beta'\alpha)^{-1}R(L)\beta' u_t + \beta_\perp' CX_0$$
$$= (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left\{\sum_{s=1}^{t}u_s + X_0\right\} + \beta_\perp'\alpha(\beta'\alpha)^{-1}R(L)\beta' u_t,$$

so that the standardized process $n^{-1/2}\beta_\perp' X_{[n\cdot]} \Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' B_u(\cdot)$, and from (3.4) we have

$$n^{-1/2}\sum_{s=1}^{[n\cdot]}\beta' X_s \Rightarrow -(\beta'\alpha)^{-1}\beta' B_u(\cdot). \tag{A.6}$$

It follows by conventional weak convergence methods that

$$n^{-1}\beta_\perp' S_{11}\beta_\perp \Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left(\int_0^1 B_u B_u'\right)\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1},$$

$$\begin{aligned}\beta_\perp'(S_{10}-S_{11}\beta\alpha') &= \beta_\perp'\left\{n^{-1}\sum_{t=1}^n X_{t-1}\left(\Delta X_t-\alpha\beta' X_{t-1}\right)'\right\}\\ &= \sum_{t=1}^n \frac{\beta_\perp' X_{t-1}}{\sqrt{n}}\frac{u_t}{\sqrt{n}} \Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u' + \Psi_{wu}^1,\\ \beta_\perp' S_{11}\beta &= \sum_{t=1}^n \frac{\beta_\perp' X_{t-1}}{\sqrt{n}}\frac{(\beta' X_{t-1})'}{\sqrt{n}} \Rightarrow -(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u'\beta(\alpha'\beta)^{-1}+\Psi_{wv},\end{aligned}$$

where

$$\Psi_{wu}^1=\sum_{h=1}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)u_{t+h}'\right\} \quad \text{and } \Psi_{wv}=\sum_{h=0}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)(\beta' X_{t+h})'\right\}$$

are one-sided long-run covariance matrices involving $w_t=\beta_\perp'\Delta X_t$, u_t and v_t. Note that

$$w_t := \beta_\perp'\Delta X_t = \beta_\perp' u_t + \beta_\perp'\alpha v_{t-1}$$

so we may deduce the explicit form

$$\begin{aligned}\Psi_{wu}^1 &= \sum_{h=1}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)u_{t+h}'\right\} = \beta_\perp'\sum_{h=1}^\infty E\left\{u_t u_{t+h}'\right\} + \beta_\perp'\alpha\sum_{h=1}^\infty E\left\{v_{t-1}u_{t+h}'\right\}\\ &= \beta_\perp'\Lambda_{uu}+\beta_\perp'\alpha\left[\Lambda_{vu}-\Gamma_{vu}(1)\right], \end{aligned} \tag{A.7}$$

and

$$\begin{aligned}\Psi_{wv} &= \sum_{h=0}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)\left(\beta' X_{t+h}\right)'\right\}\\ &= \beta_\perp'\sum_{h=0}^\infty E\left\{u_t v_{t+h}'\right\} + \beta_\perp'\alpha\sum_{h=0}^\infty E\left\{v_{t-1}v_{t+h}'\right\}\\ &= \beta_\perp'(\Lambda_{uv}+\Gamma_{uv}(0))+\beta_\perp'\alpha\Lambda_{vv}. \end{aligned} \tag{A.8}$$

Finally, using (A.6) and standard limit theory again, we obtain

$$\begin{aligned}\beta_\perp' S_{10} = &\sum_{t=1}^n \frac{\beta_\perp' X_{t-1}}{\sqrt{n}}\frac{\Delta X_t'}{\sqrt{n}} = \sum_{t=1}^n \frac{\beta_\perp' X_{t-1}}{\sqrt{n}}\left(\frac{u_t+\alpha v_{t-1}}{\sqrt{n}}\right)'\\ &\Rightarrow (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u' - (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u'\beta\left(\alpha'\beta\right)^{-1}\alpha'\\ &\quad+\sum_{h=1}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)u_{t+h}'\right\}+\sum_{h=0}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)v_{t+h}'\alpha'\right\}\\ &= (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u'\{I-\beta(\alpha'\beta)^{-1}\alpha'\}\\ &\quad+\sum_{h=1}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)u_{t+h}'\right\}+\sum_{h=0}^\infty E\left\{\left(\beta_\perp'\Delta X_t\right)v_{t+h}'\alpha'\right\}\\ &= (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1 B_u dB_u'\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1}\beta_\perp' + \Psi_{wu}^1+\Psi_{wv}\alpha',\end{aligned}$$

since $\beta\,(\alpha'\beta)^{-1}\alpha'+\alpha_\perp(\beta_\perp'\,\alpha_\perp)^{-1}\beta_\perp'=I$ (e.g. Johansen, 1995, p. 39). □

Proof of Lemma 3.2: Let $S(\lambda) = \lambda S_{11} - S_{10} S_{00}^{-1} S_{01}$, so that the determinantal equation (2.5) is $|S(\lambda)| = 0$. Defining $P_n = [\beta, n^{-1/2}\beta_\perp]$ and using Lemma 3.1, we have

$$\begin{aligned}
&\left|P_n' \left(S\left(\lambda\right)\right) P_n\right| \\
=& \left|\begin{bmatrix} \lambda\beta' S_{11}\beta & \lambda n^{-1/2}\beta' S_{11}\beta_\perp \\ \lambda n^{-1/2}\beta_\perp' S_{11}\beta & \lambda n^{-1}\beta_\perp' S_{11}\beta_\perp \end{bmatrix}\right. \\
&\left. - \begin{bmatrix} \beta' S_{10} S_{00}^{-1} S_{01}\beta & n^{-1/2}\beta' S_{10} S_{00}^{-1} S_{01}\beta_\perp \\ n^{-1/2}\beta_\perp' S_{10} S_{00}^{-1} S_{01}\beta & n^{-1}\beta_\perp' S_{10} S_{00}^{-1} S_{01}\beta_\perp \end{bmatrix}\right| \\
\Rightarrow& \left|\begin{bmatrix} \lambda\Sigma_{\beta\beta} & 0 \\ 0 & \lambda(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \left(\int_0^1 B_u B_u'\right)\alpha_\perp \left(\beta_\perp'\alpha_\perp\right)^{-1} \end{bmatrix} - \begin{bmatrix} \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix}\right| \\
=& \left|\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right| \left|\lambda(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \left(\int_0^1 B_u B_u'\right)\alpha_\perp \left(\beta_\perp'\alpha_\perp\right)^{-1}\right|. \qquad \text{(A.9)}
\end{aligned}$$

The determinantal equation

$$\left|\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right| \left|\lambda(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \left(\int_0^1 B_u B_u'\right)\alpha_\perp \left(\beta_\perp'\alpha_\perp\right)^{-1}\right| = 0$$

has $m - r_0$ zero roots and r_0 positive roots given by the solutions of

$$\left|\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right| = 0. \qquad \text{(A.10)}$$

Thus, the r_0 largest roots of (2.5) converge to the roots of (A.10) and the remainder converge to zero.

Defining $\overline{P} = [\beta, \beta_\perp]$, we have

$$\begin{aligned}
\left|\overline{P}' \left(S\left(\lambda\right)\right) \overline{P}\right| &= \left|\begin{bmatrix} \beta' S\left(\lambda\right)\beta & \beta' S\left(\lambda\right)\beta_\perp \\ \beta_\perp' S\left(\lambda\right)\beta & \beta_\perp' S\left(\lambda\right)\beta_\perp \end{bmatrix}\right| \\
&= |\beta' S(\lambda)\beta| |\beta_\perp' \{S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\}\beta_\perp|. \qquad \text{(A.11)}
\end{aligned}$$

As in Johansen (1995, theorem 11.1), we let $n \to \infty$ and $\lambda \to 0$ such that $\rho = n\lambda = O_p(1)$. Using Lemma 3.1, we have

$$\begin{aligned}
\beta' S\left(\lambda\right)\beta &= \rho n^{-1}\beta' S_{11}\beta - \beta' S_{10} S_{00}^{-1} S_{01}\beta = -\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} + o_p(1), \\
\beta_\perp' S\left(\lambda\right)\beta_\perp &= \rho n^{-1}\beta_\perp' S_{11}\beta_\perp - \beta_\perp' S_{10} S_{00}^{-1} S_{01}\beta_\perp, \text{ and} \\
\beta_\perp' S\left(\lambda\right)\beta &= \rho n^{-1}\beta_\perp' S_{11}\beta - \beta_\perp' S_{10} S_{00}^{-1} S_{01}\beta \\
&= -\beta_\perp' S_{10} S_{00}^{-1} S_{01}\beta + o_p(1). \qquad \text{(A.12)}
\end{aligned}$$

Define

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

Using Lemmas 3.1 and A1, we have

$$\begin{aligned}
N_n &= \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}\left(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right)\Sigma_{\beta 0}\Sigma_{00}^{-1} + o_p(1) \\
&= \widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1} \widetilde{\alpha}_\perp' + o_p(1). \qquad \text{(A.13)}
\end{aligned}$$

By (A.12) and (A.13), the second factor in (A.11) becomes

$$\begin{aligned}
&\beta_\perp'\{S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\}\beta_\perp \\
&= \rho n^{-1}\beta_\perp' S_{11}\beta_\perp - \beta_\perp' S_{10} N_n S_{01}\beta_\perp + o_p(1) \\
&= \rho n^{-1}\beta_\perp' S_{11}\beta_\perp - \beta_\perp' S_{10}\widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1} \widetilde{\alpha}_\perp' S_{01}\beta_\perp + o_p(1). \qquad (A.14)
\end{aligned}$$

By Lemma 3.1, we have

$$\begin{aligned}
&\beta_\perp'\{S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\}\beta_\perp \\
&\sim \rho(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \left(\int_0^1 B_u B_u'\right)\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1} \\
&\quad - \left\{(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \int_0^1 B_u dB_u'\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp' + \Psi\right\} \\
&\quad \times \widetilde{\alpha}_\perp(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp)^{-1}\widetilde{\alpha}_\perp' \left\{\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \int_0^1 dB_u B_u'\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1} + \Psi'\right\} \\
&= \rho\int_0^1 G_u G_u' - \left(\int_0^1 G_u dG_u'\beta_\perp' + \Psi\right)\widetilde{\alpha}_\perp(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp)^{-1}\widetilde{\alpha}_\perp'\left(\beta_\perp\int_0^1 dG_u G_u' + \Psi'\right),
\end{aligned}$$

where $\Psi = \Psi_{wu}^1 + \Psi_{wv}\alpha'$ and $G_u(r) = (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' B_u(r)$ is Brownian motion with variance matrix

$$(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\Omega\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}.$$

Equations (A.11), (A.14) and Lemma 3.1 reveal that the $m - r_0$ smallest solutions of (2.5) normalized by n converge to those of the equation

$$\left|\rho\int_0^1 G_u G_u' - \left(\int_0^1 G_u dG_u'\beta_\perp' + \Psi\right)\widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp'\left(\beta_\perp\int_0^1 dG_u G_u' + \Psi'\right)\right| = 0, \quad (A.15)$$

as stated. □

Proof of Theorem 3.1:

Part (a) Let $\mathrm{IC}_{r_0}(r)$ denote the information criterion defined in (1.2) when the true co-integration rank is r_0. Cointegrating rank is estimated by minimizing $\mathrm{IC}_{r_0}(r)$ for $0 \le r \le m$. To check the consistency of this estimator, we need to compare $\mathrm{IC}_{r_0}(r)$ with $\mathrm{IC}_{r_0}(r_0)$ for any $r \neq r_0$.

When $r > r_0$, using (1.2) and (2.9), we have

$$\begin{aligned}
&\mathrm{IC}_{r_0}(r) - \mathrm{IC}_{r_0}(r_0) \\
&= \sum_{i=r_0+1}^{r} \log(1-\hat{\lambda}_i) + C_n n^{-1}\left((2mr - r^2) - \left(2mr_0 - r_0^2\right)\right) \\
&= \sum_{i=r_0+1}^{r} \log(1-\hat{\lambda}_i) + C_n n^{-1}(r-r_0)(2m - r - r_0). \qquad (A.16)
\end{aligned}$$

In order to consistently select r_0 with probability 1 as $n \to \infty$ we need

$$\sum_{i=r_0+1}^{r} \log(1-\hat{\lambda}_i) + C_n n^{-1}(r-r_0)(2m-r-r_0) > 0, \qquad (A.17)$$

with probability 1 as $n \to \infty$ for any $r_0 < r < m$. From (3.10), we know that $\hat{\lambda}_i$ is $O_p(n^{-1})$ for all $i = r_0 + 1, \ldots, r$. Expanding $\log(1 - \hat{\lambda}_i)$, we have

$$\sum_{i=r_0+1}^{r} \log(1 - \hat{\lambda}_i) = -\sum_{i=r_0+1}^{r} \hat{\lambda}_i + o_p(n^{-1}) = O_p(n^{-1}). \tag{A.18}$$

Using (A.18) and Lemma 3.2, we then have

$$\begin{aligned} & n\left(\sum_{i=r_0+1}^{r} \log(1 - \hat{\lambda}_i) + C_n n^{-1}(r - r_0)(2m - r - r_0)\right) \\ &= -\sum_{i=r_0+1}^{r} n\hat{\lambda}_i + C_n(r - r_0)(2m - r - r_0) + o_p(1), \end{aligned} \tag{A.19}$$

where $n\hat{\lambda}_i$ for $i = r_0 + 1, \ldots, r$ are $O_p(1)$. As such, as long as $C_n \to \infty$ as $n \to \infty$, the second term on the right-hand side of (A.19) dominates, which leads to (A.17) as $n \to \infty$. Hence, if the penalty coefficient $C_n \to \infty$, cointegrating rank $r > r_0$ will never be selected. So, too few unit roots will never be selected in the system in such cases. Thus, the criteria BIC and HQ will never select excessive cointegrating rank as $n \to \infty$. On the other hand the AIC penalty is fixed at $C_n = 2$ for all n, so we may expect AIC to select models with excessive cointegrating rank with positive probability as $n \to \infty$. This corresponds to a more liberally parametrized system.

When $r < r_0$,

$$\begin{aligned} & \mathrm{IC}_{r_0}(r) - \mathrm{IC}_{r_0}(r_0) \\ &= -\sum_{i=r+1}^{r_0} \log(1 - \hat{\lambda}_i) + C_n n^{-1}\left((2mr - r^2) - \left(2mr_0 - r_0^2\right)\right) \\ &= -\sum_{i=r+1}^{r_0} \log(1 - \hat{\lambda}_i) + C_n n^{-1}(r - r_0)(2m - r - r_0). \end{aligned} \tag{A.20}$$

In order to consistently select r_0 with probability 1 as $n \to \infty$, we need

$$-\sum_{i=r+1}^{r_0} \log(1 - \hat{\lambda}_i) + C_n n^{-1}(r - r_0)(2m - r - r_0) > 0, \quad \text{as } n \to \infty. \tag{A.21}$$

From Lemma 3.2, we know that $0 < \hat{\lambda}_i < 1$ for $i = r + 1, \ldots, r_0$. So the first term on the right-hand side of (A.20) is a positive number that is bounded away from 0 and the second term on the right-hand side of (A.20) is a negative number of order $O(C_n n^{-1})$. In order for (A.21) to hold as $n \to \infty$, we therefore require only that $C_n/n = o(1)$, i.e. that the penalty coefficient must pass to infinity slower than n. For each of the criteria AIC, BIC and HQ, the penalty coefficient $C_n \to \infty$ slower than n. Hence, these three information criterion all select models with insufficient cointegrating rank (or excess unit roots) with probability zero asymptotically.

Combining the conditions on C_n for $r > r_0$ and $r < r_0$, it follows the information criterion will lead to consistent estimation of the co-integration rank provided the penalty coefficient satisfies $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$.

Part (b) Under AIC, $C_n = 2$. The limiting probability that $AIC(r_0) \leq AIC(r)$ for some $r \leq r_0$ is given by

$$\begin{aligned}
&\lim_{n\to\infty} P\left\{AIC(r_0) \leq AIC(r)\right\} \\
&= \lim_{n\to\infty} P\left\{-\sum_{i=r+1}^{r_0} \log(1-\hat{\lambda}_i) + 2n^{-1}(r-r_0)(2m-r-r_0) > 0\right\} \\
&= \lim_{n\to\infty} P\left\{\sum_{i=r+1}^{r_0} \log(1-\hat{\lambda}_i) < 2n^{-1}(r-r_0)(2m-r-r_0)\right\} = 1,
\end{aligned} \tag{A.22}$$

because $0 < \lambda_i < 1$ for $i = r+1, \ldots, r_0$ are the $r_0 - r$ smallest solutions to (3.9) and then $\sum_{i=r+1}^{r_0} \log(1 - \lambda_i) < 0$, giving (A.22). Hence, when r_0 is the true rank, AIC will not select any $r < r_0$ as $n \to \infty$, i.e.

$$\lim_{n\to\infty} P\left(\hat{r}_{AIC} = r | r < r_0\right) = 0. \tag{A.23}$$

Let $\xi_{r_0+1} > \cdots > \xi_m$ be the ordered roots of the limiting determinantal equation (3.10). When $r' > r \geq r_0$, $AIC(r) < AIC(r')$ iff

$$\sum_{i=r+1}^{r'} \log(1-\hat{\lambda}_i) + C_n n^{-1}(r'-r)(2m-r'-r) > 0,$$

so that the limiting probability that r will be chosen over r' is

$$\begin{aligned}
&\lim_{n\to\infty} P\left\{AIC(r) < AIC(r')\right\} \\
&= \lim_{n\to\infty} P\left\{-\sum_{i=r+1}^{r'} n\hat{\lambda}_i + 2(r'-r)(2m-r'-r) > 0\right\} \\
&= P\left\{\sum_{i=r+1}^{r'} \xi_i < 2(r'-r)(2m-r'-r)\right\}.
\end{aligned} \tag{A.24}$$

Accordingly, the probability that AIC will select rank r is equivalent to the probability that r is chosen over any other $r' \geq r_0$. This probability is

$$\begin{aligned}
&\lim\nolimits_{n\to\infty} P\left(\hat{r}_{AIC} = r | r > r_0\right) \\
&= P\left\{\left(\bigcap_{r'=r+1}^{m}\left\{\sum_{i=r+1}^{r'} \xi_i < 2(r'-r)(2m-r'-r)\right\}\right) \cap \right. \\
&\left.\left(\bigcap_{r'=r_0}^{r-1}\left\{\sum_{i=r'+1}^{r} \xi_i > 2\left(r-r'\right)\left(2m-r-r'\right)\right\}\right)\right\},
\end{aligned} \tag{A.25}$$

where the first part is the limiting probability that r is chosen over all $r' > r$ and the other part is the probability that r is chosen over all $r_0 \leq r' < r$. Any rank less than r_0 is not taken into account here because those ranks are always dominated in the limit by r_0 from (A.23).

The probability that the co-integration rank r_0 is consistently estimated by AIC as $n \to \infty$ is

$$\begin{aligned}
&\lim\nolimits_{n\to\infty} P\left(\hat{r}_{AIC} = r_0\right) \\
&= P\left[\bigcap_{r=r_0+1}^{m}\left\{\sum_{i=r_0+1}^{r} \xi_i < 2(r-r_0)(2m-r-r_0)\right\}\right].
\end{aligned} \tag{A.26}$$

This is a special case of (A.25) with $r = r_0$. □

The unit root case. When the system order is $m = 1$, the procedure provides a mechanism for unit root testing. If $r_0 = 0$, i.e. the model has a unit root, we have by (A.26)

$$\lim_{n\to\infty} P(\hat{r}_{AIC} = 1|r_0 = 0) = P\{\xi_1 > 2\} = 1 - P\{\xi_1 < 2\} \text{ and}$$
$$\lim_{n\to\infty} P(\hat{r}_{AIC} = 0|r_0 = 0) = P\{\xi_1 < 2\}, \tag{A.27}$$

where ξ_1 is the solution to (3.10) when $m = 1$ and $r_0 = 0$. In this case, we see that

$$\xi_1 = \frac{\left(\int_0^1 G_u dG_u' \beta_\perp' + \Psi\right)^2}{\left(\int_0^1 G_u G_u'\right)\Sigma_{00}} = \frac{\left(\int_0^1 B_u dB_u + \lambda\right)^2}{\left(\int_0^1 B_u^2\right)\Sigma_{00}}$$

since $G_u = B_u$, $\widetilde{\alpha}_\perp = 1$, $\beta_\perp = 1$, and $\Psi = \lambda$ in this case.

If $r_0 = 1$, when the model is stationary, we have

$$\lim_{n\to\infty} P(\hat{r}_{AIC} = 0|r_0 = 1) = 0 \text{ and } \lim_{n\to\infty} P(\hat{r}_{AIC} = 1|r_0 = 1) = 1, \tag{A.28}$$

using (A.22). These results for the scalar case $m = 1$ are consistent with those in Phillips (2008).