SEMIPARAMETRIC COINTEGRATING RANK SELECTION FOR CURVED CROSS SECTION TIME SERIES

By

Peter C. B. Phillips

March 2025

COWLES FOUNDATION DISCUSSION PAPER NO. 2432

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Semiparametric Cointegrating Rank Selection for Curved Cross Section Time Series*

Peter C. B. Phillips†

† Yale University, University of Auckland & Singapore Management University,

March 15, 2025

Abstract

Cointegrating rank selection is studied in a function space reduced rank regression where the data are time series of cross section curves. A semiparametric approach to rank selection is employed using information criteria suitably modified to take account of the function space context, extending the linear cointegrating model to accommodate cross section data under general forms of dependence. A parametric formulation is employed analogous to recent work on cross section curve autoregression and cointegrating regression. Consistent cointegrating rank estimation is developed by the use of information criteria methods that are extended to the curve time series environment. The asymptotic theory involves two parameter Gaussian processes that generalize the standard limit processes involved in cointegrating regressions with conventional multiple time series. Simulations provide evidence of the effectiveness of consistent rank selection by the BIC criterion and the tendency of AIC to overestimate order as it does in standard lag order selection in autoregression as well as in reduced rank regression with multiple time series.

Keywords: Cointegrating rank, Curved cross section data, Hilbert space, Gaussian processes, Information criteria.

JEL classification: C21, C23

1 Introduction

Research on nonstationarity has occupied a central position in time series econometrics since the mid 1980s. Formal methods using function space limit theory that were employed in the original developments, most particularly those that appeared during 1985, were at that time completely novel in econometric teaching and research. Several of those initial developments were discussed by David Hendry (Hendry, 1986) in a special issue of the *Oxford Bulletin of Economics and Statistics* that appeared remarkably soon after the work was circulated and submitted. The OBES special issue was devoted partly to these methods and partly to the

*Phillips acknowledges partial research support from the Kelly Fund at the University of Auckland and a KLC Fellowship at Singapore Management University.

doors they had opened for new applications, including those that enabled spurious regressions to be detected by well-founded techniques.

The origins of and especially the need for this line of research in econometrics go back much further to work on modeling macroeconomic variables by way of error correction systems. Models of this type had already been in use in crafting theory specifications with some empirical implementations in a longstanding tradition associated with the London School of Economics primarily by Bill Phillips, Rex Bergstrom, and Denis Sargan (Phillips, 1954; Sargan, 1964; Bergstrom, 1967) in the 1950s and 1960s. That research reflected an intuitive grasp that the nonstationarity so evident in macroeconomic time series might somehow be eliminated or sufficiently attenuated by careful formulation of error correction systems to model differences as dependent variables, giving conventional econometric methods designed for stationary data a possible basis for estimation, inference and prediction. In this intuitive line of thinking there was no apprehension of the fact that nonstationary data, in particular unit roots and other stochastic trends that were present in the levels variables within the correction system itself, had subtle lasting technical impacts on the limit theory needed to guide the design and valid application of econometric methods. It was this discovery, largely through the circulation of several technical papers by the present author distributed during 1985 and cited in David Hendry's overview (Hendry, 1986), that opened the doors to a formal study of the impact of nonstationarity in numerous arenas of empirical econometric work. That same year at the 1985 World Congress of the Econometric Society at MIT, two papers were presented that addressed key foundational aspects of this subject (Engle and Granger, 1987; Phillips, 1987) exposing a wide audience to the new ideas that have subsequently had such a longstanding influence on econometric research. These two papers were later published together as the leading two articles in the March issue of *Econometrica* 1987. Coupled with related work that dealt directly with the development of a rigorous asymptotic theory of cointegrated system estimation (Phillips and Durlauf, 1986; Phillips, 1988a,b; Phillips and Ouliaris, 1988; Park and Phillips, 1988; Johansen, 1988), this research paved a new technical path forward that fostered the emergence of vast fields of empirical applications.

The present paper forms part of an ongoing study of nonstationary time series that is linked to the foundational research discussed above. But while that early technical work and the vast proportion of subsequent research dealt exclusively with multiple time series living in Euclidean space, this research involves time series of curves of cross section data, which we refer to as curved cross section time series or more simply as curve time series. In doing so, it provides new linkages between time series econometrics and microeconometrics by introducing cross section data in a manner that accommodates general forms of cross section dependence, thereby extending more conventional approaches to dynamic panel data analysis with independent or clustered cross section observations. Our approach relates also to general Hilbert and Banach space modeling with nonstationary data that uses operator formulations to capture possible cointegrating relationships between variables in function space (Beare et al., 2017; Beare and Seo, 2020; Chang et al., 2019; Franchi and Paruolo, 2020; Seo and Beare, 2019; Seo, 2023a,b), although our modeling uses simpler parametric cointegrating forms which avoid the challenges of estimating function space operators.

One of the modern approaches to practical cointegration modeling is semiparametric, which allows model users to be agnostic regarding the short memory features of the data and to concentrate attention on long run behavior. The methods that have been developed since 1985 cover several aspects of modeling that are useful in applied work, including estimation and inference concerning the cointegrating space, estimation of cointegrating rank, prediction, and allowance for time series that do not fall strictly within unit root or local unit regimes. The author's recent work in this area (Phillips and Jiang, 2024; Phillips, 2024, 2025) has focused on extending these procedures to time series of curve data that allow cross section curves to evolve over time while maintaining any cointegrating linkages among them.

The present paper's focus is on the estimation of cointegrating rank. The methodology employed involves the use of information criteria. Such criteria are often well suited to order estimation, primary examples being the choice of lag length in autoregressions and variable choice in regression. Application of these methods in cointegrating systems is also natural because rank is itself an order parameter with only a finite number of choices in finite dimensional reduced rank regression. In parametric time series modeling, information criteria have been employed in this context in the past and are known to be consistent under certain conditions (Phillips, 1996; Chao and Phillips, 1999). Those ideas were employed in a semiparametric setting for the purpose of unit root model selection (Phillips, 2008) and cointegrating rank selection (Cheng and Phillips, 2009, 2012). The present paper is most closely related to the latter research, extending the methods of those papers to the context of curved cross section time series models of reduced rank regression.

The following model is a semiparametric formulation of an m-vector of cointegrated curve time series $X_t(r)$ that are assumed to satisfy a reduced rank regression of the form

$$\Delta X_t(r) = \alpha\beta' X_{t-1}(r) + u_t(r), \quad t \in \{1, \ldots, n\}\,,\ r \in [a,b], \tag{1}$$

where α and β are $m \times \ell_0$ matrices of full rank $\ell_0 \leq m$ and the m-vector time series $\{X_t(r), u_t(r)\}_{t=1}^n$ each live in the Hilbert space $\mathcal{H} = L_2\,[a,b]^m$ where $L_2[a,b]^m = \left\{f : \int_a^b f'f < \infty\right\}$ with inner product $\langle f, g\rangle = \int_a^b f'g$ taken over some finite or infinite observational interval $[a,b]$. A common interval for each dimension of the data is used and this can be arranged by suitable choice of the interval $[a,b]$. It is also convenient but not necessary in applications to consider the case where all the functions in (1) are continuously differentiable in r and lie in $C[a,b]^m \subset L_2[a,b]^m$.[1] The error process $u_t(r)$ is a weakly dependent stationary curve time series process with zero mean, satisfying conditions that are detailed below. The series $X_t(r)$ is initialized at $t = 0$ by some (possibly random) quantity $X_0(r) = o_p(\sqrt{n})$ and does not mate-

[1] In practice, this could be achieved by using a smoothing technique such as kernel regression to convert finely observed discrete cross section data (comprising N observations) into a continuously differentiable curve. Provided the error in such a technique passes to zero fast enough as $N \to \infty$ relative to the errors involved in estimation and inference as the sample size $n \to \infty$ this replacement is innocuous asymptotically.

rially affect the limit theory. The model (1) is estimated by reduced rank regression in Hilbert space using the time series curves $\{X_t(r) : t = 1, \cdots, n; r \in [a, b]\}$ leading to estimates of the matrices α and β obtained by ignoring any temporal weak dependence structure or the form of cross section dependence in $u_t(r)$.

The paper is organized as follows. The cointegrating rank selection methodology is given in Section 2. Assumptions and some preliminaries needed for the asymptotic development are in Section 3. Reduced rank regression formulae are given in Section 4. The required limit theory involves two parameter vector Gaussian processes that take account of the cross section curve data used in estimation and rank selection. The main result is given in Section 5 and Section 6 reports simulation findings with reduced rank regression involving curve time series and provides comparisons with rank selection in conventional time series reduced rank regression. Section 7 concludes with some final discussion. Proofs are in Section 8. Tables are in Appendix A (Section 9) and Figures in Appendix B (Section 10).

2 Cointegrating rank selection by information criteria

As discussed in earlier work (Cheng and Phillips, 2009, hereafter CP(2009)) a complete model for statistical purposes is often unnecessary when cointegrating rank determination is needed and asymptotically efficient estimation of a reduced rank regression is the primary focus. Many approaches to econometric estimation and inference are semiparametric in character, allowing practitioners to concentrate on long run behavior without paying attention to specific short memory features of the data. It is then desirable to use methods to evaluate cointegrating rank (or choice of the number of unit roots or near unit roots in a reduced rank system) in a semiparametric context where the short memory component has a general form. This approach extends naturally to systems in which the time series data are cross section curves evolving over time, such as (1).

In such settings, the issue of cointegrating rank choice reduces to the alternative problem of distinguishing the number of unit roots that are present in the system (1), viz., $m - \ell_0$, thereby leading directly to the cointegrating rank ℓ_0. If $\alpha_\perp$ is an $m \times (m - \ell_0)$ matrix orthogonal complement to the full rank matrix α then $\alpha'_\perp \Delta X_t(r) = \alpha'_\perp u_t(r)$ is a weakly dependent curve time series process and the multiple time series process $\alpha'_\perp X_t(r)$ has $m - \ell_0$ unit roots. Since the rank ℓ_0 is unknown, the reduced rank system (1) can be estimated using standard methods for all values of ℓ and the corresponding residual variance matrix calculated in each case. The asymptotic behavior of these fitted residual variance matrices then embodies information about the number of unit roots, thereby producing an indicator of the cointegrating rank in the true system.

More specifically, let $\widehat{\alpha}$ and $\widehat{\beta}$ be reduced rank estimates of the $m \times \ell$ matrices α and β in (1) for some given value of the integer $\ell \leq m$. For each value of $\ell = 0, 1, .., m$, then define the corresponding average residual variance matrices

$$\widehat{\Sigma}\left(\ell\right)=n^{-1}\sum_{t=1}^{n}\int_{a}^{b}\left(\Delta X_{t}(r)-\hat{\alpha}\hat{\beta}^{\prime}X_{t-1}(r)\right)\left(\Delta X_{t}(r)-\hat{\alpha}\hat{\beta}^{\prime}X_{t-1}(r)\right)^{\prime}dr,\quad \ell=1,...,m \tag{2}$$

with $\widehat{\Sigma}\left(0\right)=n^{-1}\sum_{t=1}^{n}\int_{a}^{b}\Delta X_{t}(r)\Delta X_{t}(r)^{\prime}dr$ for $\ell=0$ when a full set of m unit roots are assumed in estimation of the reduced rank regression. Note that the expression for $\widehat{\Sigma}\left(\ell\right)$ involves averaging not only over the historical time series but also across the curved cross section data for $r\in[a,b]$.

The asymptotic behavior of the residual moment matrix $\widehat{\Sigma}\left(\ell\right)$ depends on how the value of ℓ that is employed in the estimation of the (assumed) reduced rank regression with associated estimated matrices $\widehat{\alpha}$ and $\widehat{\beta}$ relates to the true cointegrating rank ℓ_0. The criterion used to evaluate cointegrating rank is based on the multiple time series case of CP(2009) and takes the following simple form that involves a penalty term based on the number of degrees of freedom remaining in the reduced rank regression after estimation and identification

$$IC\left(\ell\right)=\log\left|\widehat{\Sigma}\left(\ell\right)\right|+C_{n}n^{-1}\left(2m\ell-\ell^{2}\right). \tag{3}$$

The penalty term in (3) has coefficient $C_n=\log n$, $2\log\log n$, or 2 corresponding to the BIC (Schwarz, 1978; Rissanen, 1978), Hannan-Quinn (HQ) Hannan and Quinn (1979), and Akaike AIC Akaike (1998) penalties, respectively. In (3) the degrees of freedom term $2m\ell-\ell^2$ is calculated to account for the $2m\ell$ elements in the matrices α and β that have to be estimated, adjusted for the ℓ^2 restrictions that are needed to ensure structural identification of β in a reduced rank regression.[2] The BIC version of (3) in the simpler multiple time series case was earlier given in Phillips and McFarland (1997) and used to determine cointegrating rank in an empirical exchange rate application.

Model evaluation based on $IC\left(\ell\right)$ uses the simple cointegrating rank selection criterion

$$\widehat{\ell}=\underset{0\leq\ell\leq m}{\arg\min}IC\left(\ell\right). \tag{4}$$

As in the pure time series case of CP(2009) and as shown below in Theorem 1, the rank selector $\widehat{\ell}$ turns out to be weakly consistent for selecting the cointegrating rank ℓ_0 provided that the penalty term in (3) satisfies the weak requirements that $C_n\to\infty$ and $C_n/n\to0$ 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 straightforward for practical implementation. Simulations reported in Section 6 reveal that the BIC criterion generally works well for cointegrating rank determination but with a slight tendency to underestimate rank in fully stationary systems, whereas AIC is inconsistent and has a clear tendency to overestimate rank just as it does in estimating lag order in autoregressions.

[2]For further discussion of the formulation of the penalty used in (3) see the Appendix of CP(2009).

3 Assumptions and preliminaries

To develop the limit theory we start with two assumptions. As discussed earlier it will often be convenient to work with the Hilbert space regression model (1) which is parameterized in a finite dimensional way so that the operator $\alpha\beta'$ is simply a finite dimensional $m \times m$ matrix of unknown rank ℓ_0. Multiple time series regressions of this type in Euclidean space coupled with the functional limit theory associated with their data trajectories provided the foundational elements for studying cointegrating regressions among stochastically nonstationary time series (Phillips and Durlauf, 1986; Phillips, 1988a,b; Park and Phillips, 1988; Johansen, 1995) for testing the existence of cointegration, amongst a host of other applications. Our focus in this section is on the development of a procedure for determining cointegrating rank, allowing for curve time series observations in the context of a generating mechanism where there is general weak dependence in the innovations $u_t(r)$. The asymptotic results provide a semiparametric framework for cointegration rank assessment with function space data, which lead to various other potential applications of curve time series cointegrated system estimation and inference. The conditions below allow for general linear process curve innovations within a cointegrated system specification. The subsequent development follows the same general framework as that used in the multiple time series reduced rank regression analysis of CP(2009).

Assumption LP.

(i) *The $\mathcal{H}$-valued error sequence $u_t(r)$ follows the linear process $u_t(r) = D(L)(\varepsilon_t(r)) := \sum_{j=0}^{\infty} d_j(\varepsilon_{t-j}(r))$, where $\varepsilon_t(r) \sim mds(0, K_\varepsilon)$ is an m-dimensional martingale difference process with natural filtration and covariance matrix kernel $K_\varepsilon(r,s) = \mathbb{E}\varepsilon_t(r)\varepsilon_t(s)' \in L_2[a,b]^{m\times m}$. Fourth moments of ε_t exist and the d_j are bounded linear $m \times m$ matrix operators with nonsingular $D(1) = \sum_{j=1}^{\infty} d_j$ and $\sum_{j=1}^{\infty} j^2 \|d_j\|_{op} < \infty$, where $\|d_j\|_{op}$ is the operator norm. The long run covariance matrix operator of $u_t(r)$ is $C_u = D(1)C_\varepsilon D(1)^*$, where $D(1)^*$ is the adjoint matrix operator of $D(1)$ and C_ε is the covariance matrix operator of ε_t.*

(ii) *$X_0(r)$ is an $L_2[a,b]^m$-valued initial condition and satisfies $\sup_{r\in[a,b]} |X_0(r)| = o_p(n^{1/2})$.*

In Assumption LPi covariance operators such as C_ε are defined by their action on the function space, here $L_2[a,b]^m$ with inner product $\langle x, y\rangle = \int_a^b x(s)'y(s)ds$ and norm $\|x\| = \left(\int_a^b x(s)'x(s)ds\right)^{1/2}$, so that for any $x \in L_2[a,b]^m$, $C_\varepsilon(x) := \mathbb{E}(\varepsilon_t\langle\varepsilon_t, x\rangle)$. Then

$$C_\varepsilon(x)(r) = \mathbb{E}(\varepsilon_t(r)\langle\varepsilon_t, x\rangle) = \mathbb{E}\Big[\varepsilon_t(r)\int_a^b \varepsilon_t(s)'x(s)ds\Big]$$
$$= \int_a^b \mathbb{E}\Big[\varepsilon_t(r)\varepsilon_t(s)'\Big]x(s)ds = \int_a^b K_\varepsilon(r,s)x(s)ds, \tag{5}$$

where $K_\varepsilon(r,s) = \mathbb{E}\Big[\varepsilon_t(r)\varepsilon_t(s)'\Big]$ is the matrix covariance kernel of the process $\varepsilon_t(r)$. The matrix covariance kernel of the error process $u_t(r)$ is $K_u(r,s) = \mathbb{E}\Big[u_t(r)u_t(s)'\Big]$. Under LPi, the spectral density matrix operator of $u_t(r)$ is given by

$$f_u(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_h e^{-i\lambda h} = \frac{1}{2\pi} \left(\sum_{j=-\infty}^{\infty} d_j e^{-i\lambda j} \right) C_\varepsilon \left(\sum_{j=-\infty}^{\infty} d_j e^{-i\lambda j} \right)^*,$$

where Γ_h is the autocovariance matrix operator of $u_t(r)$. The long run covariance matrix operator of $u_t(r)$ is then $C_u = 2\pi f_u(0) = \sum_{h=-\infty}^{\infty} \Gamma_h$. Further, the coefficient operators d_j in the linear process $u_t(r) = \sum_{j=0}^{\infty} d_j(\varepsilon_{t-j})(r)$ are defined by $d_j(x)(r) = \int_a^b d_{f_j}(s,r)x(s)ds$ for some functions $d_{f_j(\cdot,\cdot)} \in L_2[a,b]^{m^2}$, m-vector of functions $x \in L_2[a,b]^m$ and the operator norm is $\|d_j\|_{op} = \left(\sup_{\|x\| \leq 1} \| \int_a^b d_{f_j}(s,r)x(s)ds\|^2 \right)^{1/2}$.

Assumption LPi is a Hilbert space linear process condition of the type that is convenient in developing partial sum functional limit theory (Bosq, 2000; Phillips, 2024). The methods of Phillips and Solo (1992) apply in the Hilbert space context under suitable summability conditions and are convenient for establishing results of this type, just as they are in Euclidean space.

Assumption RR.

(i) The determinantal equation $|I_m - (I_m + \alpha\beta') L| = 0$ has roots on or outside the unit circle, i.e., $|L| \geq 1$.

(ii) Define $\Pi = I_m + \alpha\beta'$ where α and β are $m \times \ell_0$ matrices of full column rank ℓ_0, $0 \leq \ell_0 \leq m$. If $\ell_0 = 0$ then $\Pi = I_m$; if $\ell_0 = m$ then β has full rank m and so both $\beta' X_t(r)$ and X_t are (asymptotically) stationary.

(iii) The matrix $R = I_{\ell_0} + \beta'\alpha$ has eigenvalues within the unit circle.

Assumption RR gives conditions that are standard in the study of reduced rank regressions with some unit roots (Johansen, 1988, 1995; Phillips, 1996). Assumption RRiii ensures that the matrix $\beta'\alpha$ has full rank. Define $\alpha_\perp$ and $\beta_\perp$ to be orthogonal complements to α and β, so that $[\alpha, \alpha_\perp]$ and $[\beta, \beta_\perp]$ are nonsingular and $\beta_\perp' \beta_\perp = I_{m-\ell}$. Then, nonsingularity of $\beta'\alpha$ implies the nonsingularity of $\alpha_\perp' \beta_\perp$. Under RR we have a curve time series Wold representation of the stationary transform $v_t(r) := \beta' X_t(r) = R\beta' X_{t-1}(r) + \beta' u_t(r)$

$$v_t(r) = \sum_{i=0}^{\infty} R^i \beta' u_{t-i}(r) = R(L) \beta' u_t(r) = R(L) \beta' D(L) \varepsilon_t(r). \tag{6}$$

Average covariance matrices of the stationary curve time series $\{u_t(r), v_t(r) = \beta' X_t(r)\}$ are written as $\Gamma_h(u,u) = \int_a^b \mathbb{E} u_t(r) u_{t+h}(r)' dr$. $\Gamma_h(u,v) = \int_a^b \mathbb{E} u_t(r) v_{t+h}(r)' dr$, and $\Gamma_h(v,v) = \int_a^b \mathbb{E} v_t(r) v_{t+h}(r)' dr$.

The Wold representation (6) yields the following useful partial sum representation of the curve time series $X_t(r)$ after some further manipulations

$$X_t(r) = C\sum_{s=1}^{t} u_s(r) + \alpha\left(\beta'\alpha\right)^{-1} R\left(L\right)\beta' u_t(r) + CX_0(r), \tag{7}$$

where $C = \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\alpha'_\perp$. Expression (7) extends the Granger-Johansen representation by allowing for weakly dependent time series of curve innovations $u_t(r)$ rather than martingale differences, as in reduced rank regressions that are assumed to be correctly specified VARs. Note that (7) also delivers a decomposition of the Phillips and Solo (1992) type suited for functional central limit theory for partial sums of curve time series innovations.

Under assumption LP a Hilbert space functional law for partial sums of the curve error processes $u_t(r)$ holds. In particular, using the limit theory in Phillips (2025) for scalar and vector curve time series, we have as $n \to \infty$

$$G_n(p,r) := \frac{1}{\sqrt{n}}\sum_{s=1}^{\lfloor np\rfloor} u_s(r) \rightsquigarrow \mathcal{G}_u\left(p,r\right), \tag{8}$$

where $\mathcal{G}_u\left(p,r\right)$ is a two parameter m-vector Gaussian process with $p \in [0,1]$, $r \in [a,b]$. The process $\mathcal{G}_u(p,r) = (\mathcal{G}_{u,h}(p,r))$ lives on the product space $C[0,1]^m \times C[a,b]$ and the limit theory (8) follows by virtue of multi-dimensional Hilbert space weak convergence. Functional weak convergence results of this general type in the scalar case $m = 1$ are given in recent work (Berkes et al., 2013; Jirak, 2013; Phillips, 2024). In the present multiple curve time series case with $m > 1$ the Gaussian process $\mathcal{G}_u(p,r)$ has long run covariance matrix kernel function $\Omega(r,s) = \sum_{j=-\infty}^{\infty}\mathbb{E}(u_t(r)u_{t+j}(s))$.

The component processes $\mathcal{G}_{u,h}(p,r)$ may each be represented in terms of a coordinatewise limit involving independent standard Wiener processes $\{W_{h,i}(\cdot)\}_{i=1}^{\infty}$ by a Karhunen Loève (KL) representation. In fact, each element $\mathcal{G}_{u,h}(p,r)$ can be written in the form

$$\mathcal{G}_{u,h}(p,r) = \sum_{i=1}^{\infty}\lambda_{h,i}^{1/2}\phi_{h,i}(r)W_{h,i}(p), \tag{9}$$

where the $\{W_{h,i}(p)\}_{i=1}^{\infty}$ are independent standard Wiener processes on $C[0,1]$ and where $\{\lambda_{h,i},\phi_{h,i}(r)\}_{i=1}^{\infty}$ are the eigenvalues and orthonormal eigenfunctions of the covariance kernel $K_{u_h}(r,s) = \mathbb{E}u_{h,t}(r)u_{h,t}(s)$ with Mercer representation $K_{u_h}(r,s) = \sum_{j=1}^{\infty}\lambda_{h,i}\phi_{h,i}(r)\phi_{h,j}(s)$.

In view of (6) and the fact that $R\left(1\right) = \sum_{i=0}^{\infty} R^i = (I-R)^{-1} = -\left(\beta'\alpha\right)^{-1}$, we further have

$$n^{-1/2}\sum_{s=1}^{\lfloor np\rfloor} v_s(r) = n^{-1/2}\sum_{s=1}^{\lfloor np\rfloor}\beta' X_s(r) \rightsquigarrow -\left(\beta'\alpha\right)^{-1}\beta'\mathcal{G}_u\left(p,r\right), \quad \text{as } n\to\infty. \tag{10}$$

The limit laws (8) and (10) involve the same two-dimensional vector Gaussian process $\mathcal{G}_u\left(p,r\right)$ and they combine to determine the asymptotic forms of the various sample moment matrices involved in the reduced rank regression estimation of (1). We finish these preliminaries by defining the following partitioned cross section average variance matrix of the stationary components of the curve time series $X_t(r)$, viz., $\Delta X_t(r) = \alpha\beta' X_{t-1}(r) + u_t(r)$ and $v_{t-1}(r) = \beta' X_{t-1}(r)$

$$\int_a^b \mathbb{E} \begin{bmatrix} \Delta X_t(r) \\ \beta' X_{t-1}(r) \end{bmatrix} \left[\Delta X_t(r)', X_{t-1}(r)'\beta\right] dr = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}. \tag{11}$$

These covariance matrices involve cross section averaged expectations of the key stationary elements in the reduced rank regression. Explicit expressions for the submatrices in the above expression are worked out in terms of the autocovariance matrix sequences of the component curve time series $u_t(r)$ and $v_t(r)$ and the parameters of (1). These are given in the Appendix.

4 Reduced rank regression formulae for curve time series

Our procedure is to perform a reduced rank regression (RRR) with estimates of α and β in (1) obtained by ignoring any weak dependence error structure in $u_t(r)$. To analyze the asymptotic properties of the rank order estimates and the information criterion $IC(\ell)$ under a general error structure, we start by investigating the asymptotic properties of the various regression components. Using conventional RRR notation but allowing for curve time series data, define the following average sample moment matrices

$$S_{00} = n^{-1}\sum_{t=1}^{n}\int_a^b \Delta X_t(r)\Delta X_t(r))' dr, \; S_{11} = n^{-1}\sum_{t=1}^{n}\int_a^b X_{t-1}(r)X_{t-1}(r)' dr, \tag{12}$$

$$S_{01} = n^{-1}\sum_{t=1}^{n}\int_a^b \Delta X_t(r)X_{t-1}(r)' dr, \text{ and } S_{10} = n^{-1}\sum_{t=1}^{n}\int_a^b X_{t-1}(r)\Delta X_t(r)' dr. \tag{13}$$

These formulae extend those in CP(2009) by the use of cross section as well time series averaging of the data. For some given ℓ and β, the estimate of α is obtained by regression as

$$\widehat{\alpha}(\beta) = S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}. \tag{14}$$

Again, given ℓ, the corresponding RRR estimate of β in (1) is an $m \times \ell$ matrix satisfying

$$\widehat{\beta} = \arg\min_{\beta}\left|S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}\right|, \tag{15}$$

subject to a normalization such as[3]

$$\widehat{\beta}' S_{11}\widehat{\beta} = I_r. \tag{16}$$

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

$$\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0 \tag{17}$$

[3] Quadratic form normalizations like (16) are commonly used in reduced rank regression applications. But such normalization, as distinct from normalization on a particular element of the vector $X_t(r)$, has the property that no finite moments of the corresponding reduced rank regression estimates of the elements of β exist, which is partly responsible for the heavy tailed feature of these finite sample distributions Phillips (1994).

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

$$\widehat{\beta} = [\widehat{v}_1, \cdots, \widehat{v}_\ell], \quad \text{and } \widehat{\alpha} = \widehat{\alpha}(\widehat{\beta}) = S_{01}\widehat{\beta}, \tag{18}$$

with $\widehat{\beta}$ formed from the eigenvectors of $\widehat{V}$ corresponding to the ℓ largest roots of (17). The residuals from the RRR and the corresponding moment matrix of residuals that appear in the information criterion are

$$\widehat{u}_t(r) = \Delta X_t(r) - \widehat{\alpha}\widehat{\beta}' X_{t-1}(r), \text{ and} \tag{19}$$

$$\widehat{\Sigma}(\ell) = n^{-1}\sum_{t=1}^{n}\int_a^b \widehat{u}_t(r)\widehat{u}_t(r)'dr = S_{00} - S_{01}\widehat{\beta}\widehat{\beta}'S_{10}. \tag{20}$$

Using (20) we have from standard RRR manipulations, e.g., Johansen (1995, theorem 6.1),

$$\left|\widehat{\Sigma}(\ell)\right| = |S_{00}|\, \Pi_{i=1}^{\ell}\left(1 - \widehat{\lambda}_i\right), \tag{21}$$

where $\widehat{\lambda}_i$, $1 \leq i \leq \ell$, are the ℓ largest solutions to (17). The criterion (4) is then well determined for any given value of ℓ.

Lemma 1. *Under Assumptions LP and RR and using the cross section average variance matrix expressions in (11), the following limit theory holds for the stationary and nonstationary components:*

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

$$n^{-1}\beta_\perp' S_{11}\beta_\perp \rightsquigarrow \left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\left(\int_a^b\int_0^1 \mathcal{G}_u(p,r)\mathcal{G}_u(p,r)'dpdr\right)\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1},$$

$$\beta_\perp'(S_{10} - S_{11}\beta\alpha') \rightsquigarrow \left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\int_a^b\int_0^1 \mathcal{G}_u(p,r)d\mathcal{G}_u(p,r)'dr + \Psi_{wu}^1,$$

$$\beta_\perp' S_{11}\beta \rightsquigarrow -\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\int_a^b\int_0^1 \mathcal{G}_u(p,r)d\mathcal{G}_u(p,r)'dr\beta(\alpha'\beta)^{-1} + \Psi_{wv},$$

$$\beta_\perp' S_{10} \rightsquigarrow \left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\int_a^b\int_0^1 \mathcal{G}_u(p,r)d\mathcal{G}_u(p,r)'dr\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1}\beta_\perp' + \Psi_{wu}^1 + \Psi_{wv}\alpha',$$

where the two parameter limit process $\mathcal{G}_u(p,r)$ is as in (8), and

$$\Psi_{wu}^1 = \sum_{h=1}^{\infty}\int_a^b \mathbb{E}\left\{\left(\beta_\perp'\Delta X_t(r)\right)u_{t+h}(r)'\right\}, \Psi_{wv} = \sum_{h=0}^{\infty}\int_a^b \mathbb{E}\left\{\left(\beta_\perp'\Delta X_t(r)\right)\left(\beta' X_{t+h}(r)\right)'\right\}, \tag{22}$$

are one-sided average long run variance matrices, and $w_t(r) = \beta_\perp'\Delta X_t(r) = \beta_\perp' u_t(r) + \beta_\perp'\alpha v_{t-1}(r)$.

Remarks:

(a) When the curve time series errors $u_t(r)$ are weakly dependent the asymptotic limits of

$\beta'_\perp(S_{10} - S_{11}\beta\alpha')$, $\beta'_\perp S_{10}$, and $\beta'_\perp S_{11}\beta$ each involve bias terms dependent on one-sided long run average covariance matrices in (22) that are associated with the stationary curve time series components $u_t(r)$, $v_t(r)$, and $w_t(r) = \beta'_\perp \Delta X_t(r)$. These one-sided long run covariance matrices are given explicitly in (40)-(42) in the proofs.

(b) In the special case of martingale difference errors $u_t(r)$, we have $\Psi^1_{wu} = 0$, $\Psi_{wv} = \beta'_\perp \int_a^b \mathbb{E}\left(u_t(r)v_t(r)'\right) dr$, and simpler results apply, such as

$$\beta'_\perp(S_{10} - S_{11}\beta\alpha') \rightsquigarrow \left(\alpha'_\perp\beta_\perp\right)^{-1} \alpha'_\perp \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr, \tag{23}$$

for the various limits given in Lemma 1. These deliver curve time series extensions of results for RRR in explicit VAR models with martingale difference errors such as those in Johansen (1995, theorem 10.3) and the weakly dependent innovation results in Cheng and Phillips (2009, Lemma 3.1). The matrix $\int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr$ in the limit (23) is a matrix stochastic integral taken over $p \in [0,1]$) and integrated over the curve $r \in [a,b]$. Stochastic integrals of this type are discussed in Phillips (2025) and play a major role in cointegrated regression limit theory with curve time series observations. When the data are multiple time series instead of multiple time series of cross section curves the stochastic integrals $\int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr$ and $\int_a^b \int_0^1 \mathcal{G}_u(p,r)\mathcal{G}_u(p,r)' dp dr$ reduce to the integrals $\int_0^1 B_u dB'_u$ and $\int_0^1 B_u B'_u$ involving the simpler process $B_u(p)$, vector Brownian motion with variance matrix $\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E} u_t u'_{t+h}$. In this event, the functional limit theory (8) is replaced by $G_n(p) := \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor np \rfloor} u_s \rightsquigarrow B_u(p)$, thereby reducing to the findings in CP(2009), Lemma 3.1 after similar reductions in the formulae for the one-sided long run covariances in (22).

5 Main results

Rewriting the model (1) as $\Delta X_t(r) = u_t(r) + \alpha\beta' X_{t-1}(r) = u_t(r) + \alpha v_{t-1}(r)$ in terms of the stationary components $\{u_t(r), v_{t-1}(r)\}$, the following relationships are obtained among the submatrix covariances in (11)

$$\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta} + \int_a^b \mathbb{E} u_t(r) v_{t-1}(r)' dr, \;\; \Sigma_{\beta 0} = \Sigma'_{0\beta}, \tag{24}$$

$$\Sigma_{\beta\beta} = \int_a^b \mathbb{E} v_t(r) v_t(r)' dr, \;\; \Sigma_{uu} = \int_a^b \mathbb{E} u_t(r) u_t(r)' dr, \tag{25}$$

$$\Sigma_{00} = \alpha\Sigma_{\beta 0} + \int_a^b \mathbb{E} u_t(r) v_{t-1}(r)' dr \alpha' + \Sigma_{uu}. \tag{26}$$

Define

$$\widetilde{\alpha} = \Sigma_{0\beta}\Sigma^{-1}_{\beta\beta} = \alpha + \int_a^b \mathbb{E} u_t(r) v_{t-1}(r)' dr \Sigma^{-1}_{\beta\beta}, \tag{27}$$

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

Lemma 2. *Under Assumptions LP and RR, when the true cointegration rank is ℓ_0, the ℓ_0 largest solutions to (17), denoted by $\widehat{\lambda}_i$ with $1 \leq i \leq \ell_0$, converge to the ℓ_0 roots $0 < \lambda_i < 1$ of the following determinantal equation in λ*

$$\left|\lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right| = 0. \tag{28}$$

The remaining $m - \ell_0$ roots, denoted by $\widehat{\lambda}_i$ with $\ell_0 + 1 \leq i \leq m$, decrease to zero at the rate $O\left(n^{-1}\right)$ and the rescaled roots $\{n\widehat{\lambda}_i : i = \ell_0 + 1, ..., m\}$ converge weakly to the roots of the following determinantal equation in ρ

$$\left|\rho \int_a^b \int_0^1 \widetilde{G}_u(p,r)\widetilde{G}_u(p,r)' dp dr - \widetilde{A}_u \widetilde{\alpha}_{\perp}\left(\widetilde{\alpha}_{\perp}' \Sigma_{00} \widetilde{\alpha}_{\perp}\right)^{-1} \widetilde{\alpha}_{\perp}' \widetilde{A}_u'\right| = 0, \tag{29}$$

where $\widetilde{A}_u = \left(\int_a^b \int_0^1 \widetilde{G}_u(p,r) d\widetilde{G}_u(p,r)' dr \beta_{\perp}' + \Psi\right)$, $\widetilde{G}_u(p,r) = \left(\alpha_{\perp}'\beta_{\perp}\right)^{-1}\alpha_{\perp}' \mathcal{G}_u(p,r)$ is a two parameter $(m-\ell_0)$-vector process with long run covariance matrix kernel

$$\left(\alpha_{\perp}'\beta_{\perp}\right)^{-1}\alpha_{\perp}' \sum_{j=-\infty}^{\infty} \mathbb{E}(u_t(r)u_{t+j}(s))\alpha_{\perp}(\beta_{\perp}'\alpha_{\perp})^{-1} = \left(\alpha_{\perp}'\beta_{\perp}\right)^{-1}\alpha_{\perp}'\Omega(r,s)\alpha_{\perp}(\beta_{\perp}'\alpha_{\perp})^{-1}, \tag{30}$$

and $\Psi = \Psi_{wu}^1 + \Psi_{wv}\alpha'$ is a composite one-sided long run covariance matrix.

Remarks:

(c) The findings in Lemma 2 relate closely to those obtained in CP(2009) for the multiple time series case with weakly dependent errors and to those of a standard RRR in a VAR with martingale difference errors (Johansen, 1995, p.158). In particular, as in the standard case, the ℓ_0 largest roots of (17) are all positive in the limit and the $m - \ell_0$ smallest roots converge to 0 at the rate n^{-1}, with both results now holding under weakly dependent errors. Since $u_t(r)$ is a weakly dependent curve time series, the limit distribution determined by (29) is considerably more complex than in the standard correctly specified RRR case and also the more general case of weakly dependent time series without curve data. In particular, the determinantal equation (29) now involves the composite one-sided long run covariance matrix Ψ as well as the two parameter Gaussian process $\widetilde{G}_u(p,r)$ in place of a transformed vector Brownian motion.

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

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

Then $\alpha_{\perp}'\beta_{\perp}\widetilde{G}_u(p,r) = \alpha_{\perp}'\mathcal{G}_u(p,r) =: G_{\alpha_{\perp}}(p,r)$ is a Gaussian process with long run covariance matrix kernel $\alpha_{\perp}'\Omega(r,s)\alpha_{\perp}$, and the one-sided long run covariance $\Psi\alpha_{\perp} = 0$.

In this case the determinantal equation (29) has the following form (omitting the (p, r) arguments for simplicity)

$$\left| \rho \int_a^b \int_0^1 \widetilde{G}_u \widetilde{G}_u' - \int_a^b \int_0^1 \widetilde{G}_u d\widetilde{G}_u' \beta_\perp' \alpha_\perp \left(\alpha_\perp' \Sigma_{uu} \alpha_\perp \right)^{-1} \alpha_\perp' \beta_\perp \int_a^b \int_0^1 d\widetilde{G}_u \widetilde{G}_u' \right| = 0.$$

Scaling this determinantal equation on left and right by the determinant $|\alpha_\perp' \beta_\perp|$ leaves the roots of the equation unaffected, giving the equivalent determinantal equation (again omitting the (p, r) arguments)

$$\left| \rho \int_a^b \int_0^1 G_{\alpha_\perp} G_{\alpha_\perp}' - \int_a^b \int_0^1 G_{\alpha_\perp} dG_{\alpha_\perp}' \left(\alpha_\perp' \Sigma_{uu} \alpha_\perp \right)^{-1} \int_a^b \int_0^1 dG_{\alpha_\perp} G_{\alpha_\perp}' \right| = 0. \tag{31}$$

The system can be further simplified to the multiple time series case by imposing full cross section curve dependence so that $u_t(r) = u_t$ *a.s.* with $\Sigma_{uu,0} = \mathbb{E} u_t(r) u_t(r)'$ for all $r \in [a, b]$. The stochastic process $G_{\alpha_\perp}(p, r) = G_{\alpha_\perp}(p)$ is then fixed over the cross section dimension $[a, b]$. The vector Gaussian process $G_{\alpha_\perp}(p)$ is now simply $m - \ell_0$ vector Brownian motion with covariance matrix $\alpha_\perp' \Sigma_{uu} \alpha_\perp = (b - a) \alpha_\perp' \Sigma_{uu,0} \alpha_\perp$ since, in view of (25), $\Sigma_{uu} = \int_a^b \mathbb{E} u_t(r) u_t(r)' dr = (b - a) \Sigma_{uu,0}$ with $\Sigma_{uu,0} = \mathbb{E} u_t u_t'$. We may now write $G_{\alpha_\perp}(p) = \left(\alpha_\perp' \Sigma_{uu} \alpha_\perp \right)^{1/2} V(p) = (b - a)^{1/2} \Sigma_{uu,0}^{1/2} V(p)$ in terms of the standard vector Brownian motion process $V(p)$ with variance matrix $I_{m - \ell_0}$. Then, rescaling (31) by the determinant $|\alpha_\perp' \Sigma_{uu} \alpha_\perp|$ leads to the equivalent determinantal equation

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

in this special case of martingale difference errors $u_t(r) = u_t$ with full cross section dependence. These simplifications reduce the general result of Lemma 2 for weakly dependent curve time series errors to the standard limit theory of a strictly well-specified parametric reduced rank regression (Johansen, 1995).

Theorem 1. (a) *Under Assumptions LP and RR, the rank estimator $\widehat{\ell}$ in (4) based on the criterion IC(ℓ) is weakly consistent for selecting the true cointegrating rank ℓ_0 provided the penalty coefficient $C_n \to \infty$ at a slower rate than n.*

(b) *The asymptotic distribution of the AIC estimator $\widehat{\ell}$ using the penalty criterion IC(ℓ) with coefficient $C_n = 2$ is*

$$\begin{aligned} &\lim_{n \to \infty} \mathbb{P} \left(\widehat{\ell}_{AIC} = \ell_0 \right) \\ &= \mathbb{P} \left[\bigcap_{\ell = \ell_0 + 1}^{m} \left\{ \sum_{i = \ell_0 + 1}^{\ell} \xi_i < 2 (\ell - \ell_0)(2m - \ell - \ell_0) \right\} \right], \end{aligned} \tag{32}$$

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = \ell | \ell > \ell_0\right) \\
&= \mathbb{P}\left\{\left(\bigcap_{\ell'=\ell+1}^{m}\left\{\sum_{i=\ell+1}^{\ell'} \xi_i < 2\left(\ell' - \ell\right)\left(2m - \ell' - \ell\right)\right\}\right) \cap \right. \\
&\left.\left(\bigcap_{\ell'=\ell_0}^{r-1}\left\{\sum_{i=\ell'+1}^{\ell} \xi_i > 2\left(\ell - \ell'\right)\left(2m - \ell - \ell'\right)\right\}\right)\right\},
\end{aligned} \tag{33}
$$

and

$$
\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = \ell | \ell < \ell_0\right) = 0, \tag{34}
$$

where $\xi_{\ell_0+1}, ..., \xi_m$ *are the ordered roots of the limiting determinantal equation* (29).

Remarks:

(e) The findings in Theorem 1(a) match those in CP(2009) for the consistency of BIC, HQ (Hannan and Quinn, 1979) and other information criteria with $C_n \to \infty$ and $C_n/n \to 0$. Importantly, consistency in determining cointegrating rank holds under these two simple rate conditions without having to specify a full parametric model. This means that cointegrating regression analysis for curve time series can be conducted conditional on the selected rank in the RRR or using alternative IVX methods such as those discussed in Phillips (2025), which allow for the same generality concerning weak dependence in the innovations $u_t(r)$ as well as the presence of some roots local to unity as well as other roots at unity.

(f) The findings for AIC also match those in CP(2009). AIC is inconsistent, overestimating cointegrating rank asymptotically in favor of more liberally parametrized systems in terms of higher cointegrating rank. This compares with AIC's well-known overestimation tendency in lag length selection in autoregression. But in cointegrating rank selection maximum rank is bounded above by the order of the system. So the potential advantage (e.g. in terms of size control in inference) of overestimation in lag length selection in autoregressions might not be anticipated here. However, when cointegrating rank is high (and close to full dimension), AIC can be expected to perform well largely because the upper bound in rank restricts the tendency to overestimate. This tendency is confirmed in simulations reported below and matches what was found in the multiple time series case in CP(2009), although the findings differ in terms of degree because of additional cross section curve information that is brought to bear in model selection.

(g) When $m = 1$, the rank $\ell_0 = 0$ corresponds to the unit root case and $\ell_0 = 1$ is the stationary case. Thus, Theorem 1 specializes to unit root testing and the selection criteria are consistent in discriminating between unit root and stationary curve time series provided

$C_n \to \infty$ and $C_n/n \to 0$, matching the finding in Phillips (2008) for simple scalar time series. In this case, as shown in (68) in the proofs, the limit distribution of AIC is itself much simpler and involves only the limiting root with the following explicit form

$$\xi_1 = \left(\int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r) dr + \psi \right)^2 \Big/ \left\{ \left(\int_a^b \int_0^1 \mathcal{G}_u(p,r)^2 dr \right) \Sigma_{00} \right\}, \tag{35}$$

where the limit process $\mathcal{G}_u(p,r)$ is now a scalar two-dimensional Gaussian process, $\Sigma_{00} = \int_a^b \mathbb{E} u_t(r)^2 dr$ is the average variance of $u_t(r)$ and $\psi = \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left(u_t(r) u_{t+h}(r) dr\right)$ is the one-sided average long run covariance of $u_t(r)$. The distribution of ξ_1 is a curve time series version of the usual limiting unit root distribution and reduces to that simpler limit distribution when there is full cross section dependence and $\mathcal{G}_u(p,r) = B_u(p)$, Brownian motion with variance $\omega^2 = \sum_{h=-\infty 1}^{\infty} \mathbb{E}\left(u_t u_{t+h}\right)$. This 'curve unit root limit distribution' of ξ_1 in (35) is supported on the entire real line, just as the distribution of the usual unit root statistic. Using the fact that the penalty coefficient $C_n = 2$ for AIC we have $\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = 0\right) = \mathbb{P}(\xi_1^2 < 2)$ and $\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = 1\right) = 1 - \mathbb{P}(\xi_1^2 < 2)$, when $\ell_0 = 0$ and there is a unit root in the data. So, AIC is inconsistent with an asymptotic bias towards the stationary case.

(h) Theorem 1 relates to the model (1). But, as in CP(2009), the findings also apply in cases where the model has intercepts and drift. So consistent cointegration rank selection is possible using this approach in most empirical contexts with curve time series data. The only change in the limit theory involves the use of limit processes that are constructed using the appropriate L_2-space residual projections of the type introduced in Park and Phillips (1988).

(i) A further extension of Theorem 1 applies to the case where the model (1) may have local unit roots rather than (or in addition to) strict unit roots in the specification. In such cases, the model would then have $m - \ell_0$ roots local to unity and take the following form in the direction $\alpha'_\perp$ of nonstationarity

$$\Delta_c \alpha'_\perp X_t(r) = \alpha'_\perp u_t(r), \quad t \in \{1, \ldots, n\}, \; r \in [a,b], \;\; \Delta_c = I_{m-\ell_0} - \frac{c}{n} L \tag{36}$$

with $c = diag\{c_1, \cdots, c_{m-\ell_0}\}$, fixed localizing coefficients c_i and lag operator L. The Gaussian process $\widetilde{G}_u(p,r) = \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp \mathcal{G}_u(p,r)$ in (29) would then be replaced by a corresponding two parameter $(m - \ell_0)$-vector Gaussian diffusion process.[4] The main conclusions of Theorem 1 in terms of consistent cointegrating rank selection by BIC and inconsistent selection by AIC would be expected to continue to hold. In this respect, the order selection approach to cointegrating rank estimation would provide a considerable

[4]Such a two parameter $(m - \ell_0)$-vector Gaussian diffusion process may be defined as the vector process $\widetilde{\mathcal{J}}_c(p,r)$ satisfying the stochastic differential equation $d\widetilde{\mathcal{J}}_c(p,r) = c\widetilde{\mathcal{J}}_c(p,r)dp + d\widetilde{G}_u(p,r)$ with solution $\widetilde{\mathcal{J}}_c(p,r) = \int_0^p e^{c(p-q)} d\widetilde{G}_u(q,r)$ for all $r \in [a,b]$ given zero initial conditions at the origin $p = 0$.

advantage over sequential testing methods which rely on explicit unit root specifications for the computation of critical values. Detailed investigation of these properties and the finite sample performance characteristics of rank determination in the presence of local unit roots are topics for future research.

6 Simulations

This section reports simulations conducted to evaluate the finite sample performance of four cointegrating rank selection criteria: AIC ($C_n = 2$), BIC ($C_n = \log n$), HQ ($C_n = 2\log\log n$), and Log(HQ) ($C_n = \log(2\log\log n)$). The models chosen are Hilbert space versions of the multiple time series RRR models used in the simulation study reported in CP(2009). These choices enable a comparison that reveals the impact of cross section curve time series data on the effectiveness of the main model selection criteria. The model designs include various generating mechanisms for the short memory component u_t, different settings for cointegrating rank, and the various choices of the penalty coefficient C_n. As in CP(2009) we report findings for systems of dimension $m = 2$ and $m = 4$ variables. Several generating mechanisms (VAR, VMA, VARMA) were employed for the short memory component $u_t(r)$, different settings for the true cointegrating rank ℓ, and for the penalty coefficient C_n.

The simulation design is based on (1) and employs curve time series formulations that extend the designs used in CP(2009). In particular, Brownian motion, Brownian bridge and segmented Brownian bridge curves are used in modeling the cross section data in conjunction with the reduced rank time series specification. When the model dimension $m = 2$, the designs allow for three different cointegrating ranks. For $\ell_0 = 0$ with a full set of m unit roots we have $\alpha'\beta = 0$; for $\ell_0 = 1$ the reduced rank coefficient structure is set so that

$$\alpha'\beta = R_1 = (1, 0.5)\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \text{ with stationary root } \lambda_1\left[I + \beta'\alpha\right] = 0.5,$$

and for $\ell_0 = 2$

$$\alpha'\beta = R_2 = \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.4 \end{pmatrix}, \text{ with stationary roots } \lambda_i\left[I + \beta'\alpha\right] = \{0.7, 0.4\}, i = 1, 2.$$

The error processes $u_t(r)$ were formulated as curve time series generated from VAR(1), VMA(1), and VARMA(1,1) models as follows

$$u_t(r) = Au_{t-1}(r) + \varepsilon_t(r),\ u_t(r) = \varepsilon_t(r) + B\varepsilon_{t-1}(r),\ \text{and } u_t(r) = Au_{t-1}(r) + \varepsilon_t(r) + B\varepsilon_{t-1}(r), \tag{37}$$

with coefficient matrices $A = \psi I_m$, $B = \phi I_m$, where $\psi = \phi = 0.4$, and with innovation processes $\varepsilon_t(r) \sim_{iid} \{W(r) : r \in [0, 1]\}$, for each $t = 1, \cdots, n$, and with standard Brownian motion curves $W(r)$ over the interval $[0, 1]$. The time series models in (37) take Brownian motion curves as inputs and generate weakly dependent time series of curve cross section data.

The performance of the four rank selection criteria (AIC, BIC, HQ, log(HQ)) were investigated for various time series sample sizes ($n \in \{50, 100, 250, 1000\}$ to show the capabilities of these criteria in determining cointegrating rank with small and large time series samples and $n = 1000$ used primarily to assess the degree of inconsistency in AIC and the relative success of BIC, HQ and log(HQ). All cases included 20 additional observations to eliminate start-up effects from the initializations $X_0(r) = 0$ and $\varepsilon_0(r) = 0$. The findings are based on $3,000$ replications. Summary results for time series regressions are reported here in Tables 1-2 with correct selections (i.e., correctly selected rank ℓ_0) shown in bold type; and for curve time series regressions in Tables 3-4. Results for the primary cases of interest are shown in color with AIC (gold) and BIC (blue). The findings are given for models driven by time series and curve time series VARMA(1) errors. Similar results were obtained for VAR and VMA error generating schemes in (37) and are not reported. To assist in understanding the main effects of curve time series on cointegrating rank selection, Tables 1 and 2 show results for a simple time series reduced rank regression with VARMA(1,1) errors; and Tables 3 and 4 show corresponding results for curve times series reduced rank regressions with VARMA(1,1) errors for the same sample sizes. Bar graphics of the respective results for both time series and curve time series are shown in Figures 2 - 7.

The main findings for the time series models are given in (a) and (b) below. These broadly match those reported in CP(2009) for models with VAR errors u_t. The corresponding findings for curve time series data are given in (c) and (d).

(a) For the smaller sample sizes $n = 50, n = 100$ in Table 1 the poor performance of AIC, HQ and Log(HQ) in estimating the correct number of unit roots is apparent. Each of these methods favor the existence of some degree of cointegration (and hence a greater number of fitted parameters) over a full set of 2 unit roots when $\ell_0 = 0$. The penalties in each case for excessive parameterization are therefore too low for satisfactory rank selection. Note that for $n = 50$ the HQ penalty coefficient ($2log(logn) = 2.7281$) barely exceeds 2 and for Log(HQ) the penalty is even smaller ($2log(log(logn)) = 1.0036$). Furthermore, AIC shows no improvement in detecting two unit roots when $n = 100$ or even for the larger values $n = 250, 1000$ as shown in Table 2. These findings strongly corroborate the inconsistency of AIC in cointegrating rank estimation.

(b) By contrast, the BIC criterion shows substantial improvement over the other criteria in selecting two unit roots ($\ell_0 = 0$) when $n = 50$ and shows further improvement when $n = 100$. These improvements in BIC performance continue with the higher sample size $n = 250$ and for $n = 1000$, where BIC has more than 90% success rate in selecting the right number of unit roots in the system, with 99% for cointegrating rank $\ell_0 = 1$ and 100% for cointegrating rank $\ell_0 = 2$.

(c) The most notable change with curve time series data shown in Tables 3 and 4 is the substantial improvement in rank determination by the AIC criterion. For the smaller sample sizes $n = 50, 100$ the AIC correct detection rate for a full set of unit roots is 77%, much larger than the 20% rate for time series data. But the rate of correct detection of $\ell_0 = 0$ remains around the 77% level for both larger sample sizes $n = 250, 1000$ revealing no tendency to rise, which corroborates the limit theory of inconsistency. The correct detection rates by AIC for positive cointegrating ranks also rise considerably: to 86% with $n = 50$ and 96% with $n = 100$ for $\ell_0 = 1$; to 96% for $n = 250$ and 97% $n = 1000$; and to 100% for $n \geq 100$. These findings show that the additional information embodied in the curve time series data (even heavily cross section dependent data like Brownian motion curves) are put to good use in overcoming some of the deficiencies of the AIC criterion in overestimating order.

(d) The gains from the use of curve time series data are also evident in the performance of BIC. The correct detection rates for BIC are at least 97% for $\ell_0 = 0$ for all sample sizes with essentially the same rates for cointegrating rank $\ell_0 = 1$; and when $n \geq 100$ the correct detection rates for BIC are 100%.

(e) The results for both HQ and Log(HQ) show clear tendencies to overestimate cointegrating rank, particularly when $\ell_0 = 0$ and the system has a full set of unit roots. The findings for models with curve time series data show some improvement over those with time series but overestimation of cointegrating rank is still strongly evident when $\ell_0 = 0$. As pointed out in the proof of Theorem 1, when $C_n \to \infty$ very slowly as $n \to \infty$ very large samples may be needed to prevent overestimation of cointegrating rank.

The simulation findings reported above are all given for the case of Brownian motion cross section curve time series errors $u_t(r)$.[5] Segmented Brownian bridge curves (Phillips and Jiang, 2024; Phillips, 2024) may attract some special interest because this class of segmented curve provides a specification that relates more closely to typical dynamic panel modeling (under independence or clustering) where there is more limited cross section dependence than Brownian motion and quite different asymptotic theory with potentially higher rates of convergence that reflect the presence of some degree of independence in the cross section. Nonetheless, empirical data such as Engel curves for many commodity groups such as transportation and education demonstrate a considerable degree of cross section dependence – see Figure 1 for illustrative data drawn from the Singapore Life panel for ageing seniors.[6] Other data such as lifetime income quantiles with specific educational qualifications and gender across the population display similar dependence characteristics (Cho et al., 2022). So there seems to be empirical relevance in considering systems that permit substantial cross section dependence and in using this information constructively in estimation and inference.

[5]Other models for cross section curves are currently under study and will be reported in later work.

[6]The SLP is a continuing longitudinal study at Singapore Management University of consumption behavior of ageing seniors in Singapore, see https://rosa.smu.edu.sg/singapore-life-panel/about-singapore-life-panelr.

7 Concluding remarks

When the Oxford Bulletin of Economics and Statistics published its special issue on cointegration in 1986 it showed remarkable prescience of the importance of unit root and cointegration theory to the future of econometrics. This importance was underlined by the longstanding absence of any formal asymptotic basis suited to the manifest nonstationarity of the time series data used in much econometric research. While there were prior findings on scalar unit root regressions and testing in the late 1970s, notably Dickey and Fuller (1979) and Hasza and Fuller (1979), that work relied on independent, identically distributed normally distributed errors and no central limit theory, invariance principle, functional limit theory or multivariate limit theory was established, which delimited the applicability of the results. Only in the mid-1980s were suitably general asymptotics developed that were robust to distributional characteristics, heterogeneity, potential weak dependence in the errors and multivariate regression applications, not to mention departures from strict unit roots. The Oxford Bulletin recognized these developments almost as soon as they appeared as an advance on existing statistical theory that would bring vast benefits to applied research. Econometrics as a discipline and particularly time series econometrics benefited immediately from the provision of this new foundation for estimation, inference and prediction designed particularly for the type of nonstationary data available for empirical work and the linkages between the relevant variables that economic theory suggested.

Since 1986 a vast body of econometric research in the general field of nonstationarity has emerged, influencing subject areas throughout the social and business sciences, as well as major areas of climate science and paleobiology in the natural sciences, where rich high-dimensional data sources are abundant. The present paper contributes to this body of econometric work by introducing curve time series analysis to the study of reduced rank regression. The framework employed allows for quite general forms of cross section dependence in contrast to much existing work on panel data and clustering. Although the use of heavily dependent data does not typically raise convergence rates it is evident from the simulations reported here that curve time series data does sharpen inference on cointegrating rank determination. The semiparametric approach taken in the paper also allows for general forms of weak temporal dependence in the innovations and broader specifications that accommodate deterministic components and local unit roots in reduced rank regressions. The present work provides a useful beginning on rank determination in this line of multivariate curve time series analysis of cointegration. Estimation, inference and prediction with cointegrating curve time series is ongoing and will be reported elsewhere. Further extensions to the more general operator framework in full function space settings are possible and would help to complete much of the present research agenda.

References

Akaike, H. (1998): "Information theory and an extension of the maximum likelihood principle," in *Selected papers of Hirotugu Akaike*, Springer, 199–213. 5

Beare, B. K., J. Seo, and W.-K. Seo (2017): "Cointegrated linear processes in Hilbert space," *Journal of Time Series Analysis*, 38, 1010–1027. 2

Beare, B. K. and W.-K. Seo (2020): "Representation of I (1) and I (2) autoregressive Hilbertian processes," *Econometric Theory*, 36, 773–802. 2

Bergstrom, A. R. (1967): *The Construction and Use of Economic Models*, English Universities Press. 2

Berkes, I., L. Horváth, and G. Rice (2013): "Weak invariance principles for sums of dependent random functions," *Stochastic Processes and their Applications*, 123, 385–403. 8

Bosq, D. (2000): *Linear processes in function spaces: theory and applications*, vol. 149, Springer Science & Business Media. 7

Chang, Y., B. Hu, and J. Y. Park (2019): "Econometric analysis of functional dynamics in the presence of persistence," Tech. rep., Discussion Paper, Department of Economics, Indiana University. 2

Chao, J. C. and P. C. B. Phillips (1999): "Model selection in partially nonstationary vector autoregressive processes with reduced rank structure," *Journal of Econometrics*, 91, 227–271. 3

Cheng, X. and P. C. B. Phillips (2009): "Semiparametric cointegrating rank selection," *The Econometrics Journal*, 12, S83–S104. 3, 4, 5, 6, 9, 11, 12, 14, 15, 16, 17, 22, 27

——— (2012): "Cointegrating rank selection in models with time-varying variance," *Journal of Econometrics*, 169, 155–165. 3

Cho, J. S., P. C. B. Phillips, and J. Seo (2022): "Parametric conditional mean inference with functional data applied to lifetime income curves," *International Economic Review*, 63, 391–456. 18

Dickey, D. A. and W. A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427–431. 19

Engle, R. F. and C. W. Granger (1987): "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, 251–276. 2

Franchi, M. and P. Paruolo (2020): "Cointegration in functional autoregressive processes," *Econometric Theory*, 36, 803–839. 2

Hannan, E. J. and B. G. Quinn (1979): "The determination of the order of an autoregression," *Journal of the Royal Statistical Society: Series B (Methodological)*, 41, 190–195. 5, 14

Hasza, D. P. and W. A. Fuller (1979): "Estimation for autoregressive processes with unit roots," *The Annals of Statistics*, 7, 1106–1120. 19

Hendry, D. F. (1986): "Econometric Modelling with Cointegrated Variables: An Overview." *Oxford Bulletin of Economics & Statistics*, 48. 1, 2

Jirak, M. (2013): "On weak invariance principles for sums of dependent random functionals," *Statistics & Probability Letters*, 83, 2291–2296. 8

Johansen, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231–254. 2, 7

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press. 6, 7, 10, 11, 12, 13

Park, J. Y. and P. C. B. Phillips (1988): "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468–497. 2, 6, 15

Phillips, A. W. H. (1954): "Stabilization policy in a closed economy," *Economic Journal*, 64, 290–333. 2

Phillips, P. C. B. (1987): "Time series regression with a unit root," *Econometrica: Journal of the Econometric Society*, 277–301. 2, 31

——— (1988a): "Multiple regression with integrated time series," *Contemporary Mathematics*, 80, 79–105. 2, 6

——— (1988b): "Regression theory for near-integrated time series," *Econometrica*, 1021–1043. 2, 6

——— (1994): "Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models," *Econometrica*, 73–93. 9

——— (1996): "Econometric model determination," *Econometrica*, 763–812. 3, 7

——— (2008): "Unit root model selection," *Journal of the Japan Statistical Society*, 38, 65–74. 3, 15, 22, 30

——— (2024): "Cross section curved data autoregression Part 2: unit roots," *Working Paper in process, Yale University*. 3, 7, 8, 18, 30

——— (2025): "Cross section curved data cointegrating regression," *Working Paper in process, Yale University*. 3, 8, 11, 14

Phillips, P. C. B. and S. N. Durlauf (1986): "Multiple time series regression with integrated processes," *The Review of Economic Studies*, 53, 473–495. 2, 6

Phillips, P. C. B. and L. Jiang (2024): "Cross section curved data autoregression Part 1," *Working Paper in process, Yale University.* 3, 18

Phillips, P. C. B. and J. W. McFarland (1997): "Forward exchange market unbiasedness: the case of the Australian dollar since 1984," *Journal of International Money and Finance*, 16, 885–907. 5

Phillips, P. C. B. and S. Ouliaris (1988): "Testing for cointegration using principal components methods," *Journal of Economic Dynamics and Control*, 12, 205–230. 2

Phillips, P. C. B. and V. Solo (1992): "Asymptotics for linear processes," *The Annals of Statistics*, 971–1001. 7, 8

Rissanen, J. (1978): "Modeling by shortest data description," *Automatica*, 14, 465–471. 5

Sargan, J. D. (1964): "Wages and prices in the United Kingdom: A study in econometric methodology," *Econometric Analysis for National Economic Planning, Colston Papers*, 16, 25–63. 2

Schwarz, G. (1978): "Estimating the dimension of a model," *The annals of statistics*, 461–464. 5

Seo, W.-K. (2023a): "Cointegration and representation of cointegrated autoregressive processes in Banach spaces," *Econometric Theory*, 39, 737–788. 3

——— (2023b): "Functional principal component analysis for cointegrated functional time series," *Journal of Time Series Analysis.* 3

Seo, W.-K. and B. K. Beare (2019): "Cointegrated linear processes in Bayes Hilbert space," *Statistics & Probability Letters*, 147, 90–95. 2

8 Proofs

Our approach throughout this section follows similar lines to those employed in Phillips (2008) and CP(2009), but uses the functional limit theory and cross section averaging involved in dealing with sample moments of curve time series. We start with some preliminary algebraic representations that are useful in the proofs below and are established in CP(2009), Lemma 3, which we follow in laying out the formulae below. First, under (1) and Assumption LP, we have

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}\left(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{\beta 0}\right)^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1} = \Sigma_{00}^{-1/2}c_{\perp}\left(c_{\perp}'c_{\perp}\right)^{-1}c_{\perp}'\Sigma_{00}^{-1/2}, \tag{38}$$

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

$$\Sigma_{00}^{-1/2}c_\perp \left(c_\perp' c_\perp\right)^{-1} c_\perp' \Sigma_{00}^{-1/2} = \widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1} \widetilde{\alpha}_\perp'. \tag{39}$$

When $\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta}$, (39) then becomes

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

as in Johansen (1995, Lemma 10.1).

In the present semiparametric curve time series case we have $\Delta X_t(r) = u_t(r) + \alpha\beta' X_{t-1}(r) = u_t(r) + \alpha v_{t-1}(r)$ and the covariance matrix of the two stationary components $\{v_{t-1}(r), u_t(r)\}$ is $\Gamma_{vu}(1) = \int_a^b \mathbb{E}\left(v_{t-1}(r)u_t(r)\right)' dr$, which is generally nonzero. Then

$$\Sigma_{\beta\beta} = \int_a^b \mathbb{E}v_t(r)v_t(r)'dr = \Gamma_{vv}(0), \tag{40}$$

$$\Sigma_{\beta 0} = \int_a^b \mathbb{E}v_{t-1}(r)\left(u_t(r) + \alpha v_{t-1}(r)\right)' dr = \Gamma_{vu}(1) + \Gamma_{vv}(0)\alpha' = \Gamma_{vu}(1) + \Sigma_{\beta\beta}\alpha', \tag{41}$$

$$\Sigma_{00} = \alpha\Sigma_{\beta\beta}\alpha' + \alpha\Gamma_{vu}(1) + \Gamma_{uv}(-1)\alpha' + \Gamma_{uu}(0). \tag{42}$$

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

$$\Sigma_{00}^{-1/2}c_\perp \left(c_\perp' c_\perp\right)^{-1} c_\perp' \Sigma_{00}^{-1/2} = \widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1} \widetilde{\alpha}_\perp', \tag{43}$$

as given in (39).

Proof of Lemma 1: Both $\Delta X_t(r) = u_t(r) + \alpha\beta' X_{t-1}(r) = u_t(r) + \alpha v_t(r)$ and $v_t(r) = \beta' X_t(r)$ are stationary curve time series that satisfy Assumption LP. Hence, by the law of large numbers

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

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

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

In view of (7) we have

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

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

so that the standardized process $n^{-1/2}\beta_\perp' X_{[n\cdot]}(r) \rightsquigarrow \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \mathcal{G}_u(\cdot, r)$ by virtue of the functional law (8), and then from (10) we have

$$n^{-1/2} \sum_{s=1}^{[n\cdot]} \beta' X_s(r) \rightsquigarrow -\left(\beta' \alpha\right)^{-1} \beta' \mathcal{G}_u(\cdot, r). \tag{44}$$

By conventional weak convergence methods

$$n^{-1}\beta_\perp' S_{11}\beta_\perp \rightsquigarrow \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \left(\int_a^b \int_0^1 \mathcal{G}_u(p,r)\mathcal{G}_u(p,r)' dp dr \right) \alpha_\perp \left(\beta_\perp' \alpha_\perp\right)^{-1},$$

$$\begin{aligned}
\beta_\perp'(S_{10} - S_{11}\beta\alpha') &= \beta_\perp' \left\{ n^{-1} \sum_{t=1}^{n} \int_a^b X_{t-1}(r) \left(\Delta X_t(r) - \alpha\beta' X_{t-1}(r)\right)' dr \right\} \\
&= \int_a^b \sum_{t=1}^{n} \frac{\beta_\perp' X_{t-1}(r)}{\sqrt{n}} \frac{u_t(r)}{\sqrt{n}} dr \rightsquigarrow \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr + \Psi_{wu}^1,
\end{aligned}$$

$$\beta_\perp' S_{11}\beta = \int_a^b \sum_{t=1}^{n} \frac{\beta_\perp' X_{t-1}(r)}{\sqrt{n}} \frac{\left(\beta' X_{t-1}(r)\right)'}{\sqrt{n}} dr \rightsquigarrow -\left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr \beta(\alpha'\beta)^{-1} + \Psi_{wv},$$

where

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

are one-sided long run covariance matrices involving $w_t(r) := \beta_\perp'(r)\Delta X_t(r) = \beta_\perp' u_t + \beta_\perp' \alpha v_{t-1}$, $u_t(r)$ and $v_t(r)$. We thereby deduce the explicit form

$$\begin{aligned}
\Psi_{wu}^1 &= \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ \left(\beta_\perp' \Delta X_t(r)\right) u_{t+h}(r)' \right\} = \beta_\perp' \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ u_t(r) u_{t+h}(r)' \right\} dr + \beta_\perp' \alpha \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ v_{t-1}(r) u_{t+h}(r)' \right\} dr \\
&= \beta_\perp' \Lambda_{uu} + \beta_\perp' \alpha \left[\Lambda_{vu} - \Gamma_{vu}\left(1\right)\right],
\end{aligned} \tag{45}$$

where $\Lambda_{uu} = \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ u_t(r) u_{t+h}(r)' \right\} dr = \sum_{h=1}^{\infty} \Gamma_{uu}(h)$ and $\Lambda_{vu} = \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ v_t(r) u_{t+h}(r)' \right\} dr = \sum_{h=1}^{\infty} \Gamma_{vu}(h)$; and further

$$\begin{aligned}
\Psi_{wv} &= \sum_{h=0}^{\infty} \int_a^b \mathbb{E}\left\{ \left(\beta_\perp' \Delta X_t(r)\right) \left(\beta' X_{t+h}(r)\right)' \right\} dr \\
&= \beta_\perp' \sum_{h=0}^{\infty} \int_a^b \mathbb{E}\left\{ u_t(r) v_{t+h}(r)' \right\} dr + \beta_\perp' \alpha \sum_{h=0}^{\infty} \int_a^b \mathbb{E}\left\{ v_{t-1}(r) v_{t+h}(r)' \right\} dr \\
&= \beta_\perp'(\Lambda_{uv} + \Gamma_{uv}(0)) + \beta_\perp' \alpha \Lambda_{vv}.
\end{aligned} \tag{46}$$

Then, using the limit theory (44) and standard methods again, we obtain

$$
\begin{aligned}
\beta_\perp' S_{10} &= \sum_{t=1}^{n} \int_a^b \frac{\beta_\perp' X_{t-1}(r)}{\sqrt{n}} \frac{\Delta X_t(r)'}{\sqrt{n}} dr = \sum_{t=1}^{n} \int_a^b \frac{\beta_\perp' X_{t-1}(r)}{\sqrt{n}} \left(\frac{u_t(r) + \alpha v_{t-1}(r)}{\sqrt{n}} \right)' dr \\
&\rightsquigarrow \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr - \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr \beta \left(\alpha' \beta\right)^{-1} \alpha' \\
&+ \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ \left(\beta_\perp' \Delta X_t(r)\right) u_{t+h}(r)' \right\} dr + \sum_{h=0}^{\infty} \int_a^b \mathbb{E}\left\{ \left(\beta_\perp' \Delta X_t(r)\right) v_{t+h}(r)' \alpha' \right\} dr \\
&= \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr \left\{ I - \beta \left(\alpha' \beta\right)^{-1} \alpha' \right\} \\
&+ \sum_{h=1}^{\infty} \int_a^b \mathbb{E}\left\{ \left(\beta_\perp' \Delta X_t(r)\right) u_{t+h}(r)' \right\} dr + \sum_{h=0}^{\infty} \int_a^b \mathbb{E}\left\{ \left(\beta_\perp' \Delta X_t(r)\right) v_{t+h}(r)' \alpha' \right\} dr \\
&= \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r)' dr \alpha_\perp \left(\beta_\perp' \alpha_\perp\right)^{-1} \beta_\perp' + \Psi_{wu}^1 + \Psi_{wv} \alpha',
\end{aligned}
$$

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

Proof of Lemma 2: Let $S(\lambda) = \lambda S_{11} - S_{10} S_{00}^{-1} S_{01}$, so that the determinantal equation (17) is $|S(\lambda)| = 0$. Defining $P_n = [\beta, n^{-1/2}\beta_\perp]$ and using Lemma 1, we have

$$
\begin{aligned}
&\left| P_n' \left(S(\lambda)\right) P_n \right| \\
&= \left| \begin{bmatrix} \lambda \beta' S_{11} \beta & \lambda n^{-1/2} \beta' S_{11} \beta_\perp \\ \lambda n^{-1/2} \beta_\perp' S_{11} \beta & \lambda n^{-1} \beta_\perp' S_{11} \beta_\perp \end{bmatrix} - \begin{bmatrix} \beta' S_{10} S_{00}^{-1} S_{01} \beta & n^{-1/2} \beta' S_{10} S_{00}^{-1} S_{01} \beta_\perp \\ n^{-1/2} \beta_\perp' S_{10} S_{00}^{-1} S_{01} \beta & n^{-1} \beta_\perp' S_{10} S_{00}^{-1} S_{01} \beta_\perp \end{bmatrix} \right| \\
&\rightsquigarrow \left| \begin{bmatrix} \lambda \Sigma_{\beta\beta} & 0 \\ 0 & \lambda \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \left(\int_a^b \int_0^1 \mathcal{G}_u(p,r) \mathcal{G}_u(p,r)' dp dr \right) \alpha_\perp \left(\beta_\perp' \alpha_\perp\right)^{-1} \end{bmatrix} - \begin{bmatrix} \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix} \right| \\
&= \left| \lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} \right| \left| \lambda \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \left(\int_a^b \int_0^1 \mathcal{G}_u(p,r) \mathcal{G}_u(p,r)' dp dr \right) \alpha_\perp \left(\beta_\perp' \alpha_\perp\right)^{-1} \right|.
\end{aligned}
\tag{47}
$$

The determinantal equation

$$
\left| \lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} \right| \left| \lambda \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \left(\int_a^b \int_0^1 \mathcal{G}_u(p,r) \mathcal{G}_u(p,r)' dp dr \right) \alpha_\perp \left(\beta_\perp' \alpha_\perp\right)^{-1} \right| = 0
$$

has $m - \ell_0$ zero roots and ℓ_0 positive roots given by the solutions of

$$
\left| \lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} \right| = 0. \tag{48}
$$

Thus, the ℓ_0 largest roots of (17) converge to the roots of (48) and the remainder converge to zero. Defining $\overline{P} = [\beta, \beta_\perp]$, we have

$$\begin{aligned} \left|\overline{P}'\left(S\left(\lambda\right)\right)\overline{P}\right| &= \left|\left[\begin{array}{cc} \beta' S\left(\lambda\right)\beta & \beta' S\left(\lambda\right)\beta_\perp \\ \beta_\perp' S\left(\lambda\right)\beta & \beta_\perp' S\left(\lambda\right)\beta_\perp \end{array}\right]\right| \\ &= \left|\beta' S\left(\lambda\right)\beta\right| \left|\beta_\perp'\left\{S\left(\lambda\right) - S\left(\lambda\right)\beta\left[\beta' S\left(\lambda\right)\beta\right]^{-1}\beta' S\left(\lambda\right)\right\}\beta_\perp\right|. \end{aligned} \tag{49}$$

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 1, we have

$$\begin{aligned} \beta' S\left(\lambda\right)\beta &= \rho n^{-1}\beta' S_{11}\beta - \beta' S_{10}S_{00}^{-1}S_{01}\beta = -\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} + o_p\left(1\right), \\ \beta_\perp' S\left(\lambda\right)\beta_\perp &= \rho n^{-1}\beta_\perp' S_{11}\beta_\perp - \beta_\perp' S_{10}S_{00}^{-1}S_{01}\beta_\perp, \text{ and} \\ \beta_\perp' S\left(\lambda\right)\beta &= \rho n^{-1}\beta_\perp' S_{11}\beta - \beta_\perp' S_{10}S_{00}^{-1}S_{01}\beta \\ &= -\beta_\perp' S_{10}S_{00}^{-1}S_{01}\beta + o_p\left(1\right). \end{aligned} \tag{50}$$

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 Lemma 1, (38) and (39), we have

$$\begin{aligned} N_n &= \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}\left(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right)\Sigma_{\beta 0}\Sigma_{00}^{-1} + o_p\left(1\right) \\ &= \widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp' + o_p\left(1\right). \end{aligned} \tag{51}$$

By (50) and (51), the second factor in (49) becomes

$$\begin{aligned} &\beta_\perp'\left\{S\left(\lambda\right) - S\left(\lambda\right)\beta\left[\beta' S\left(\lambda\right)\beta\right]^{-1}\beta' S\left(\lambda\right)\right\}\beta_\perp \\ &= \rho n^{-1}\beta_\perp' S_{11}\beta_\perp - \beta_\perp' S_{10}N_n S_{01}\beta_\perp + o_p\left(1\right) \\ &= \rho n^{-1}\beta_\perp' S_{11}\beta_\perp - \beta_\perp' S_{10}\widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp' S_{01}\beta_\perp + o_p\left(1\right). \end{aligned} \tag{52}$$

By Lemma 1, we have

$$\begin{aligned} &\beta_\perp'\left\{S\left(\lambda\right) - S\left(\lambda\right)\beta\left[\beta' S\left(\lambda\right)\beta\right]^{-1}\beta' S\left(\lambda\right)\right\}\beta_\perp \\ &\rightsquigarrow \rho\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\left(\int_a^b\int_0^1 \mathcal{G}_u(p,r)\mathcal{G}_u(p,r)' dp dr\right)\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1} \\ &- \left\{\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\int_a^b\int_0^1 \mathcal{G}_u(p,r)d\mathcal{G}_u(p,r)' dr\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1}\beta_\perp' + \Psi\right\} \\ &\times \quad \widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp'\left\{\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\int_a^b\int_0^1 d\mathcal{G}_u(p,r)\mathcal{G}_u(p,r)' dr\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1} + \Psi'\right\} \\ &= \rho\int_a^b\int_0^1 \widetilde{G}_u(p,r)\widetilde{G}_u(p,r)' dp dr - \left(\int_a^b\int_0^1 \widetilde{G}_u(p,r)d\widetilde{G}_u(p,r)' dr\beta_\perp' + \Psi\right) \end{aligned}$$

$$\times \quad \widetilde{\alpha}_{\perp}\left(\widetilde{\alpha}_{\perp}'\Sigma_{00}\widetilde{\alpha}_{\perp}\right)^{-1}\widetilde{\alpha}_{\perp}'\left(\beta_{\perp}\int_{a}^{b}\int_{0}^{1}d\widetilde{G}_{u}(p,r)\widetilde{G}_{u}(p,r)'dr+\Psi'\right), \tag{53}$$

where $\Psi = \Psi_{wu}^{1} + \Psi_{wv}\alpha'$ and $\widetilde{G}_{u}(p,r) = \left(\alpha_{\perp}'\beta_{\perp}\right)^{-1}\alpha_{\perp}'\mathcal{G}_{u}(p,r)$ is the two parameter $(m-\ell_0)$-vector process defined in (29) with long run covariance matrix kernel given in (30).

Equations (49), (52) and (53) reveal that when normalized by n the $m-\ell_0$ smallest solutions of (17) converge to those of the equation

$$\left|\rho\int_{a}^{b}\int_{0}^{1}\widetilde{G}_{u}(p,r)\widetilde{G}_{u}(p,r)'dpdr-\widetilde{A}_{u}\widetilde{\alpha}_{\perp}\left(\widetilde{\alpha}_{\perp}'\Sigma_{00}\widetilde{\alpha}_{\perp}\right)^{-1}\widetilde{\alpha}_{\perp}'\widetilde{A}_{u}'\right|=0, \tag{54}$$

where $\widetilde{A}_{u} = \left(\int_{a}^{b}\int_{0}^{1}\widetilde{G}_{u}(p,r)d\widetilde{G}_{u}(p,r)'dr\beta_{\perp}'+\Psi\right)$ and $\widetilde{G}_{u}\left(p,r\right) = \left(\alpha_{\perp}'\beta_{\perp}\right)^{-1}\alpha_{\perp}'\mathcal{G}_{u}(p,r)$ as stated. ■

Proof of Theorem 1

The proof follows the same lines as CP(2009) with modifications to take account of the limit theory and selection criteria based on the use of curve time series regression.

Part (a) Let $IC\left(\ell_0\right)$ denote the information criterion defined in (3) when the true cointegration rank is ℓ_0. Cointegrating rank is estimated by minimizing $IC\left(\ell\right)$ for $0 \leq \ell \leq m$; and consistency is checked by comparing $IC\left(\ell\right)$ with $IC\left(\ell_0\right)$ for any $\ell \neq \ell_0$ as $n \to \infty$.

When $\ell > \ell_0$, using (3) and (21), we have

$$\begin{aligned} IC\left(\ell\right)-IC\left(\ell_0\right) &= \sum_{i=\ell_0+1}^{\ell}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left[\left(2m\ell-\ell^2\right)-\left(2m\ell_0-\ell_0^2\right)\right] \\ &= \sum_{i=\ell_0+1}^{\ell}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left(\ell-\ell_0\right)\left(2m-\ell-\ell_0\right). \end{aligned} \tag{55}$$

For consistent order selection with $\widehat{\ell} \to \ell_0$ with probability 1 as $n \to \infty$ we need

$$\sum_{i=\ell_0+1}^{\ell}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left(r-\ell_0\right)\left(2m-r-\ell_0\right)>0, \tag{56}$$

with probability 1 as $n \to \infty$ for any $\ell_0 < \ell < m$. From (29) we know that $\hat{\lambda}_i$ is $O_p\left(n^{-1}\right)$ for all $i = \ell_0+1,...,r$. Expanding $\log\left(1-\hat{\lambda}_i\right)$, we have

$$\sum_{i=\ell_0+1}^{\ell}\log\left(1-\hat{\lambda}_i\right) = -\sum_{i=\ell_0+1}^{\ell}\hat{\lambda}_i+o_p\left(n^{-1}\right)=O_p\left(n^{-1}\right). \tag{57}$$

Using (57) and Lemma 2, we then have

$$
\begin{aligned}
&n\left(\sum_{i=\ell_0+1}^{\ell}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left(\ell-\ell_0\right)\left(2m-\ell-\ell_0\right)\right)\\
&=-\sum_{i=\ell_0+1}^{\ell} n\hat{\lambda}_i + C_n\left(\ell-\ell_0\right)\left(2m-\ell-\ell_0\right)+o_p\left(1\right), \qquad (58)
\end{aligned}
$$

where $n\hat{\lambda}_i$ for $i=\ell_0+1,...,\ell$ are $O_p\left(1\right)$. Hence, as long as $C_n \to \infty$ as $n\to\infty$, the second term on the right side of (58) dominates, which leads to (56) as $n \to \infty$. So whenever the penalty coefficient satisfies $C_n \to \infty$, cointegrating rank $\ell > \ell_0$ will never be selected asymptotically. 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 parameterized system. Note further that if $C_n \to \infty$ very slowly as $n\to\infty$ it may be that extremely large samples are needed to ensure (56) holds.

When $\ell < \ell_0$,

$$
\begin{aligned}
&IC\left(\ell\right)-IC\left(\ell_0\right)\\
&=-\sum_{i=\ell+1}^{\ell_0}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left(\left(2m\ell-\ell^2\right)-\left(2m\ell_0-\ell_0^2\right)\right)\\
&=-\sum_{i=\ell+1}^{\ell_0}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left(\ell-\ell_0\right)\left(2m-\ell-\ell_0\right). \qquad (59)
\end{aligned}
$$

To consistently select ℓ_0 with probability 1 as $n \to \infty$, we need

$$
-\sum_{i=\ell+1}^{\ell_0}\log\left(1-\hat{\lambda}_i\right)+C_n n^{-1}\left(r-\ell_0\right)\left(2m-r-\ell_0\right)>0, \quad \text{as } n\to\infty, \qquad (60)
$$

whenever $\ell < \ell_0$. By definition in (17) and from Lemma 2, we know that $0 < \hat{\lambda}_i < 1$ for $i = \ell+1,...,\ell_0$. So the first term on the right side of (59) is a positive number that is bounded away from 0 and the second term on the right side of (59) is a negative number of order $O\left(C_n n^{-1}\right)$. In order for (60) to hold as $n \to \infty$, we therefore require only that $C_n/n = o\left(1\right)$, 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. These three information criteria therefore select models with insufficient cointegrating rank (or excess unit roots) with probability zero asymptotically.

Combining these results for $\ell > \ell_0$ and $\ell < \ell_0$, we deduce that use of the information criterion leads to consistent estimation of cointegration 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 $AIC(\ell_0) \leq AIC(\ell)$ for some $\ell \leq \ell_0$ is given by

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left\{AIC(\ell_0) \leq AIC(\ell)\right\} \\
&= \lim_{n\to\infty} \mathbb{P}\left\{-\sum_{i=\ell+1}^{\ell_0} \log\left(1-\hat{\lambda}_i\right) + 2n^{-1}\left(\ell-\ell_0\right)\left(2m-\ell-\ell_0\right) > 0\right\} \\
&= \lim_{n\to\infty} \mathbb{P}\left\{\sum_{i=\ell+1}^{\ell_0} \log\left(1-\hat{\lambda}_i\right) < 2n^{-1}\left(\ell-\ell_0\right)\left(2m-\ell-\ell_0\right)\right\} = 1, \qquad (61)
\end{aligned}$$

because the λ_i with $0 < \lambda_i < 1$ for $i = \ell+1, ..., \ell_0$ are the $\ell_0 - \ell$ smallest solutions to (28) and then $\sum_{i=\ell+1}^{\ell_0} \log\left(1-\lambda_i\right) < 0$, giving (61). Hence, when ℓ_0 is the true rank, AIC will not select any $\ell < \ell_0$ as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = \ell | \ell < \ell_0\right) = 0. \qquad (62)$$

Let $\xi_{\ell_0+1} > ... > \xi_m$ be the ordered roots of the limiting determinantal equation (29). When $\ell' > \ell \geq \ell_0$, $AIC(\ell) < AIC(\ell')$ iff

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

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

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left\{AIC(\ell) < AIC(\ell')\right\} &= \lim_{n\to\infty} \mathbb{P}\left\{-\sum_{i=\ell+1}^{\ell'} n\hat{\lambda}_i + 2\left(\ell'-\ell\right)\left(2m-\ell'-\ell\right) > 0\right\} \\
&= \mathbb{P}\left\{\sum_{i=\ell+1}^{\ell'} \xi_i < 2\left(\ell'-\ell\right)\left(2m-\ell'-\ell\right)\right\}. \qquad (63)
\end{aligned}$$

The probability that AIC will select rank ℓ is then equivalent to the probability that ℓ is chosen over any other $\ell' \geq \ell_0$ for which (63) holds. This probability is

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = \ell > \ell_0\right) \\
&= \mathbb{P}\left\{\left(\bigcap_{\ell'=\ell+1}^{m}\left\{\sum_{i=\ell+1}^{\ell'} \xi_i < 2\left(\ell'-\ell\right)\left(2m-\ell'-\ell\right)\right\}\right) \cap \right. \\
&\left. \left(\bigcap_{\ell'=\ell_0}^{\ell-1}\left\{\sum_{i=\ell'+1}^{\ell} \xi_i > 2\left(\ell-\ell'\right)\left(2m-\ell-\ell'\right)\right\}\right)\right\}, \qquad (64)
\end{aligned}$$

where the first part is the limiting probability that ℓ is chosen over all $\ell' > \ell$ and the other part is the probability that ℓ is chosen over all $\ell_0 \leq \ell' < \ell$. Any rank less than ℓ_0 is not taken into account here because those ranks are always dominated in the limit by ℓ_0 from (62).

The probability that the true cointegrating rank ℓ_0 is consistently estimated by AIC is

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = \ell_0\right) = \mathbb{P}\left[\bigcap_{\ell=\ell_0+1}^{m} \left\{\sum_{i=\ell_0+1}^{\ell} \xi_i < 2\left(\ell - \ell_0\right)\left(2m - \ell - \ell_0\right)\right\}\right]. \tag{65}$$

This is a special case of (64) with $\ell = \ell_0$. ■

The curve time series unit root case

When the system order is $m = 1$, the model (1) is simply a scalar curve time series process $X_t(r) = \gamma X_{t-1} + u_t(r)$ with parameter $\gamma = 1 + \alpha\beta$. The procedure then reduces to a mechanism for unit root testing of $\gamma = 1$ for the curve time series observations $X_t(r)$. In this case the model has a unit root when $\alpha = \beta = \alpha\beta = 0$, i.e., when $\ell_0 = 0$ and there is no stationary root. From (65) we have

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = 1|\ell_0 = 0\right) = \mathbb{P}\left\{\xi_1 > 2\right\} = 1 - \mathbb{P}\left\{\xi_1 < 2\right\} \text{ and} \tag{66}$$

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = 0|\ell_0 = 0\right) = \mathbb{P}\left\{\xi_1 < 2\right\}, \tag{67}$$

where ξ_1 is the solution to (29) when $m = 1$ and $\ell_0 = 0$. In this scalar curve time series case under the null of a unit root with $\gamma = 1$, the limiting root ξ_1 takes the explicit form

$$\xi_1 = \frac{\left(\int_a^b \int_0^1 \widetilde{G}_u(p,r) d\widetilde{G}_u(p,r) dr + \Psi\right)^2}{\int_a^b \int_0^1 \widetilde{G}_u(p,r)^2 dp dr \Sigma_{00}} = \frac{\left(\int_a^b \int_0^1 \mathcal{G}_u(p,r) d\mathcal{G}_u(p,r) dr + \psi\right)^2}{\int_a^b \int_0^1 \mathcal{G}_u(p,r)^2 dp dr \Sigma_{00}}, \tag{68}$$

as reported in (35). We note the following simplifications that apply in this scalar unit root case: (i) with $\alpha = \beta = 0$ the orthonormal complements become $\alpha_\perp = 1, \beta_\perp = 1$ so that $\widetilde{G}_u(p,r) = \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp' \mathcal{G}_u(p,r) = \mathcal{G}_u(p,r)$, which is now a scalar two parameter Gaussian process; (ii) the bias parameter in the numerator of (68) is $\psi = \sum_{j=1}^{\infty} \int_a^b \mathbb{E} u_t(r) u_{t+h}(r) dr$, a scalar version of Ψ_{wu}^1 in (22) and the composite one-sided average long run covariance matrix $\Psi = \Psi_{wu}^1 + \Psi_{wv}\alpha'$ in (29); and (iii) the matrix $\widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \Sigma_{00} \widetilde{\alpha}_\perp\right)^{-1} \widetilde{\alpha}_\perp'$ in (29) reduces to the simple scalar Σ_{00}^{-1}, with $\Sigma_{00} = \int_a^b \mathbb{E} u_t(r)^2 dr$ being the average variance of the scalar curve error process $u_t(r)$.

If $\ell_0 = 1$ and the model is stationary with $|\gamma| = |1 + \alpha\beta| < 1$, we have

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = 0|\ell_0 = 1\right) = 0 \text{ and } \lim_{n\to\infty} \mathbb{P}\left(\widehat{\ell}_{AIC} = 1|\ell_0 = 1\right) = 1, \tag{69}$$

using (61). The results (66) – (69) for the scalar case case $m = 1$ are curve time series versions of those in Phillips (2008) for simple time series autoregression and unit root testing by use of AIC selection. Direct methods for testing the presence of a unit root in curve time series autoregression have been developed in recent work by the author (Phillips, 2024), which studies curve time series unit root autoregression limit theory and provides extensions to such regressions of the coefficient and t-ratio semiparametric unit root tests in Phillips (1987) as well as conventional ADF regression tests. ■

9 A: Simulation results – tables

Table 1: Cointegrating rank selection in a time series reduced rank regression with ARMA(1,1) equation errors u_t for sample sizes $n = 50$ and $n = 100$

	$n = 50$			$n = 100$		
$\ell_0 = 0$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	**0.22**	0.68	0.09	**0.20**	0.70	0.10
BIC	**0.68**	0.33	0.01	**0.75**	0.25	0.00
HQ	**0.09**	0.73	0.17	**0.11**	0.73	0.16
Log(HQ)	**0.00**	0.46	0.54	**0.00**	0.52	0.48
$\ell_0 = 1$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	**0.83**	0.17	0.00	**0.82**	0.18
BIC	0.00	**0.96**	0.04	0.00	**0.97**	0.03
HQ	0.00	**0.73**	0.27	0.00	**0.76**	0.24
Log(HQ)	0.00	**0.37**	0.63	0.00	**0.43**	0.57
$\ell_0 = 2$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	0.14	**0.86**	0.00	0.00	**1.00**
BIC	0.05	0.40	**0.54**	0.00	0.03	**0.97**
HQ	0.00	0.08	**0.92**	0.00	0.00	**1.00**
Log(HQ)	0.00	0.01	**0.99**	0.00	0.00	**1.00**

Table 2: Cointegrating rank selection in a time series reduced rank regression with ARMA(1,1) equation errors u_t for sample sizes $n = 250$ and $n = 1000$

	$n = 250$			$n = 1000$		
$\ell_0 = 0$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	**0.20**	0.71	0.09	**0.20**	0.72	0.08
BIC	**0.82**	0.18	0.00	**0.91**	0.09	0.00
HQ	**0.15**	0.73	0.12	**0.19**	0.73	0.08
Log(HQ)	**0.00**	0.59	0.41	**0.01**	0.62	0.37
$\ell_0 = 1$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	**0.83**	0.17	0.00	**0.82**	0.18
BIC	0.00	**0.98**	0.02	0.00	**0.99**	0.01
HQ	0.00	**0.79**	0.21	0.00	**0.82**	0.18
Log(HQ)	0.00	**0.49**	0.51	0.00	**0.53**	0.47
$\ell_0 = 2$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	0.00	**1.00**	0.00	0.00	**1.00**
BIC	0.00	0.00	**1.00**	0.00	0.00	**1.00**
HQ	0.00	0.00	**1.00**	0.00	0.00	**1.00**
Log(HQ)	0.00	0.00	**1.00**	0.00	0.00	**1.00**

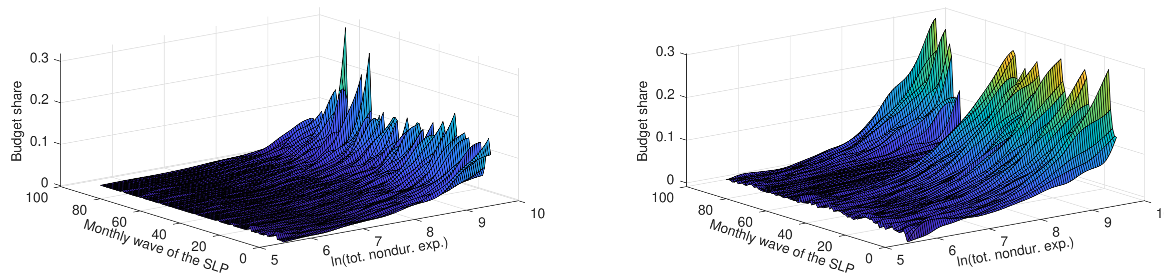
Table 3: Cointegrating rank selection in a curve time series reduced rank regression with ARMA(1,1) cross section curve errors $u_t(r), r \in [0,1]$ for sample sizes $n = 50$ and $n = 100$

	$n = 50$			$n = 100$		
$\ell_0 = 0$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	**0.77**	0.22	0.01	**0.77**	0.22	0.01
BIC	**0.97**	0.03	0.00	**0.98**	0.02	0.00
HQ	**0.56**	0.42	0.02	**0.63**	0.34	0.03
Log(HQ)	**0.04**	0.60	0.36	**0.07**	0.67	0.26
$\ell_0 = 1$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	**0.83**	0.16	0.00	**0.96**	0.04
BIC	0.00	**0.96**	0.04	0.00	**1.00**	0.00
HQ	0.00	**0.73**	0.27	0.00	**0.93**	0.07
Log(HQ)	0.00	**0.37**	0.63	0.00	**0.57**	0.43
$\ell_0 = 2$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	0.14	**0.86**	0.00	0.00	**1.00**
BIC	0.05	0.40	**0.54**	0.00	0.01	**0.99**
HQ	0.00	0.08	**0.92**	0.00	0.00	**1.00**
Log(HQ)	0.00	0.01	**0.99**	0.00	0.00	**1.00**

Table 4: Cointegrating rank selection in a curve time series reduced rank regression with ARMA(1,1) cross section curve errors $u_t(r), r \in [0,1]$ for sample sizes $n = 250$ and $n = 1000$

	$n = 250$			$n = 1000$		
$\ell_0 = 0$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	**0.77**	0.22	0.01	**0.78**	0.21	0.01
BIC	**0.99**	0.01	0.00	**1.00**	0.00	0.00
HQ	**0.70**	0.28	0.02	**0.75**	0.23	0.02
Log(HQ)	**0.10**	0.68	0.22	**0.14**	0.69	0.17
$\ell_0 = 1$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	**0.96**	0.04	0.00	**0.97**	0.03
BIC	0.00	**1.00**	0.00	0.00	**1.00**	0.00
HQ	0.00	**0.94**	0.06	0.00	**0.96**	0.04
Log(HQ)	0.00	**0.66**	0.34	0.00	**0.71**	0.29
$\ell_0 = 2$						
	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$	$\widehat{\ell} = 0$	$\widehat{\ell} = 1$	$\widehat{\ell} = 2$
AIC	0.00	0.00	**1.00**	0.00	0.00	**1.00**
BIC	0.00	0.00	**1.00**	0.00	0.00	**1.00**
HQ	0.00	0.00	**1.00**	0.00	0.00	**1.00**
Log(HQ)	0.00	0.00	**1.00**	0.00	0.00	**1.00**

10 B: Figures



(a) Education Engel Curves

(b) Leisure Engel Curves

Figure 1: Smoothed Engle curves for ageing seniors in the Singapore Life Panel

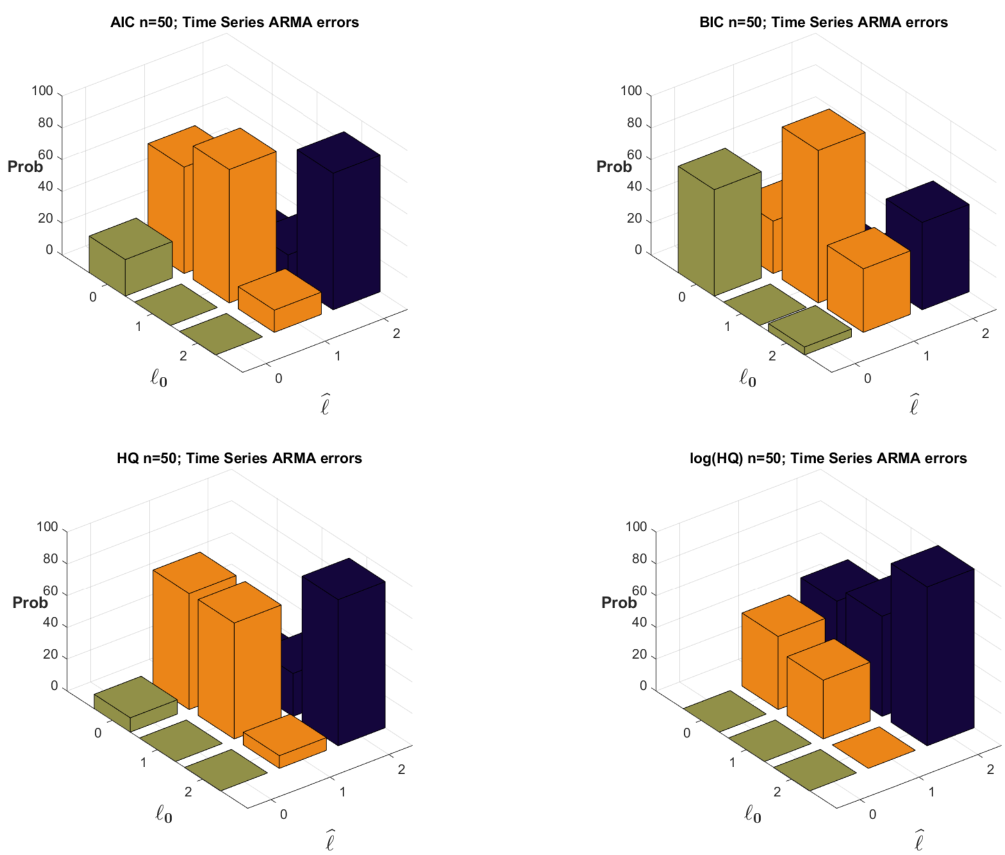


Figure 2: Selection probabilities by AIC, BIC, HQ and log(HQ) for cointegrating rank in a time series reduced rank regression with ARMA(1) time series errors and $n = 50$

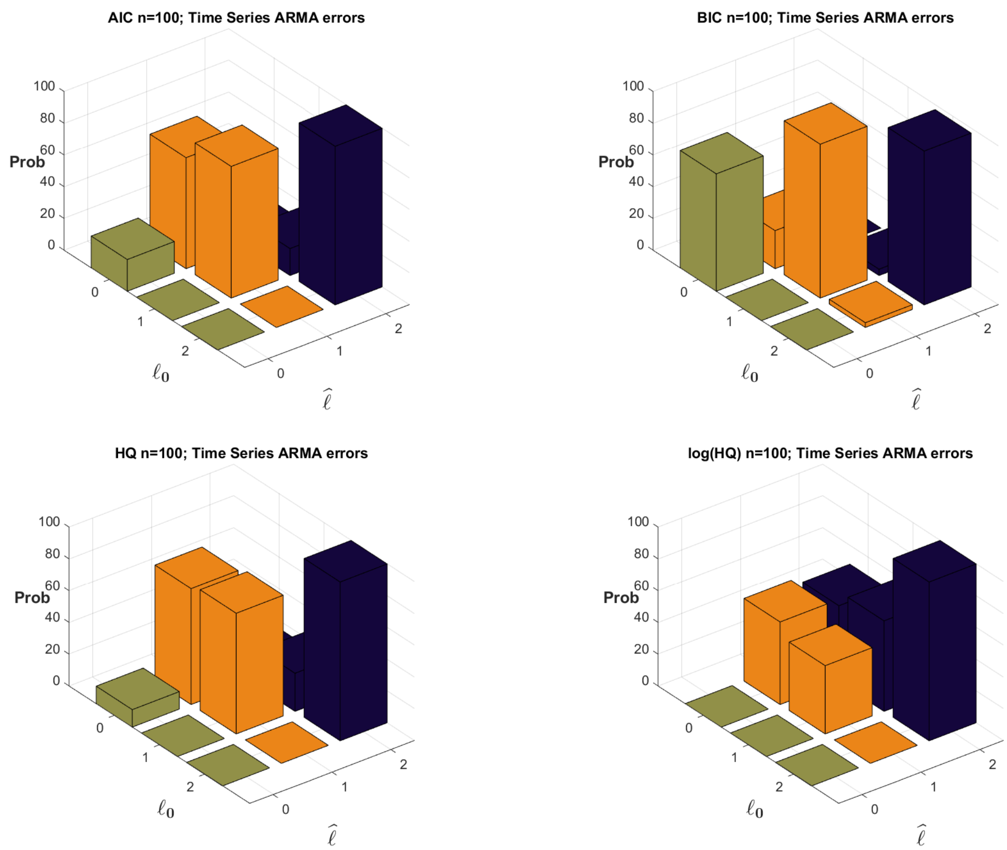


Figure 3: Selection probabilities by AIC, BIC, HQ and log(HQ) for cointegrating rank in a time series reduced rank regression with ARMA(1,1) time series errors and $n = 100$

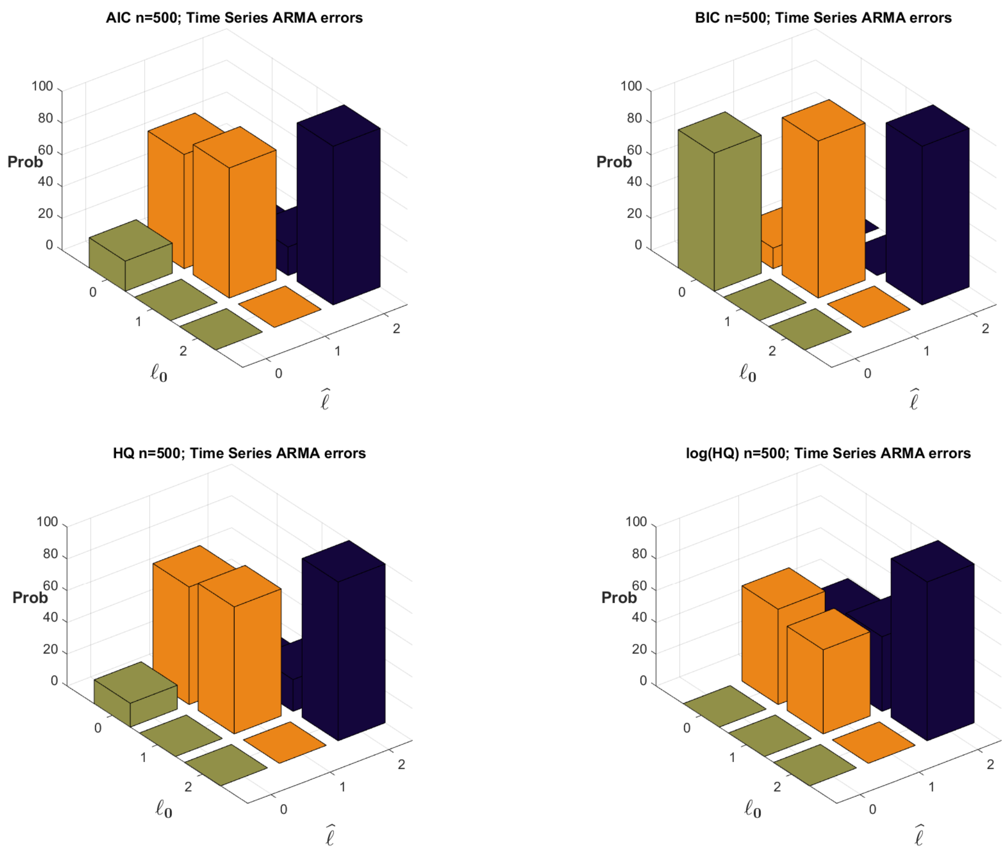


Figure 4: Selection probabilities by AIC, BIC, HQ and log(HQ) for cointegrating rank in a time series reduced rank regression with ARMA(1,1) time series errors and $n = 500$

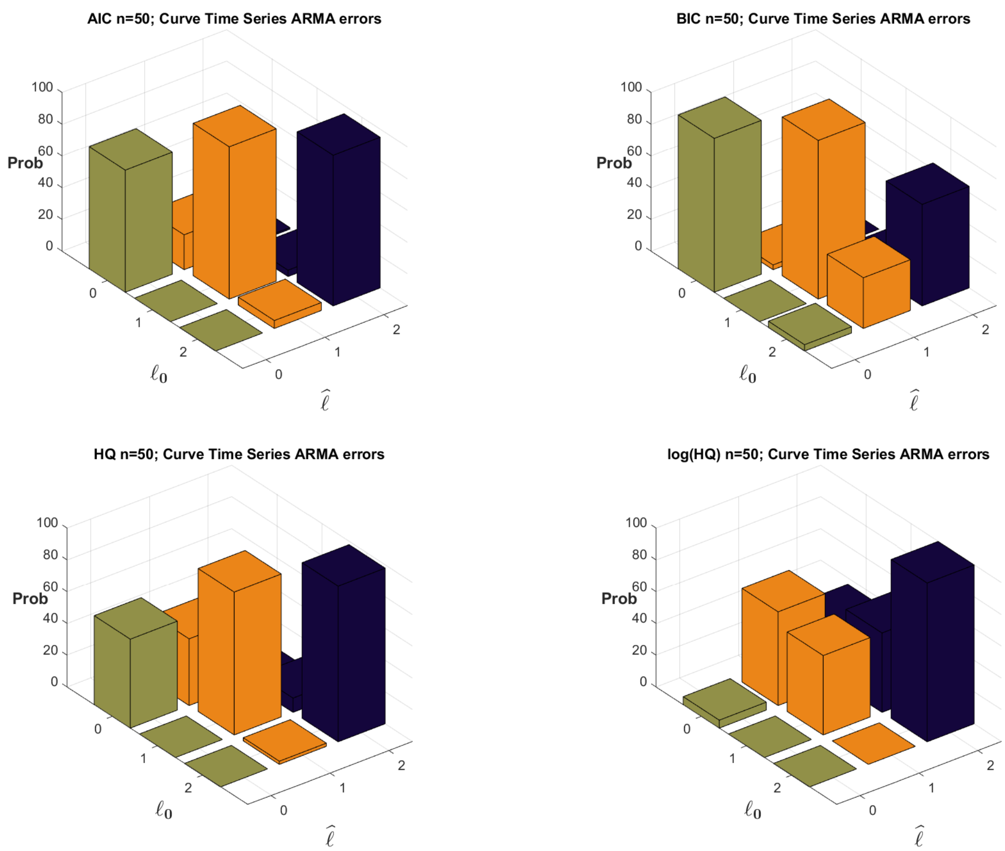


Figure 5: Selection probabilities by AIC, BIC, HQ and log(HQ) for cointegrating rank in a curve time series reduced rank regression with BM curve ARMA(1,1) time series errors and $n = 50$

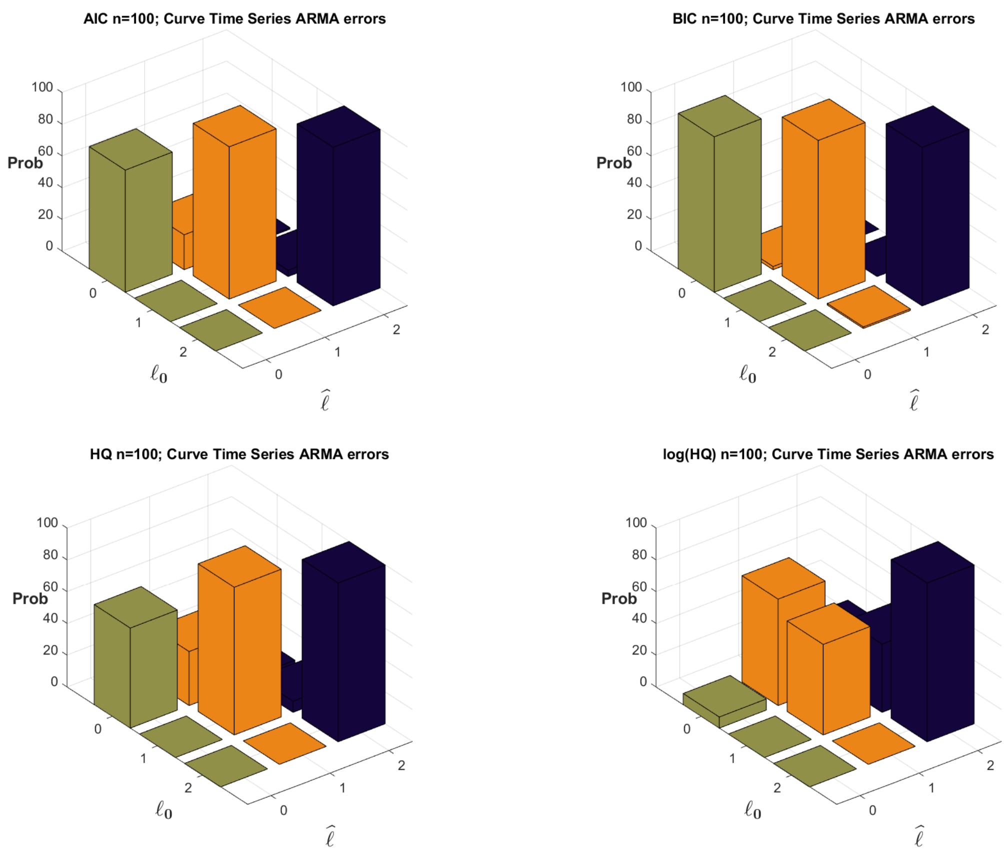


Figure 6: Selection probabilities by AIC, BIC, HQ and log(HQ) for cointegrating rank in a curve time series reduced rank regression with BM curve ARMA(1,1) time series errors and $n = 100$

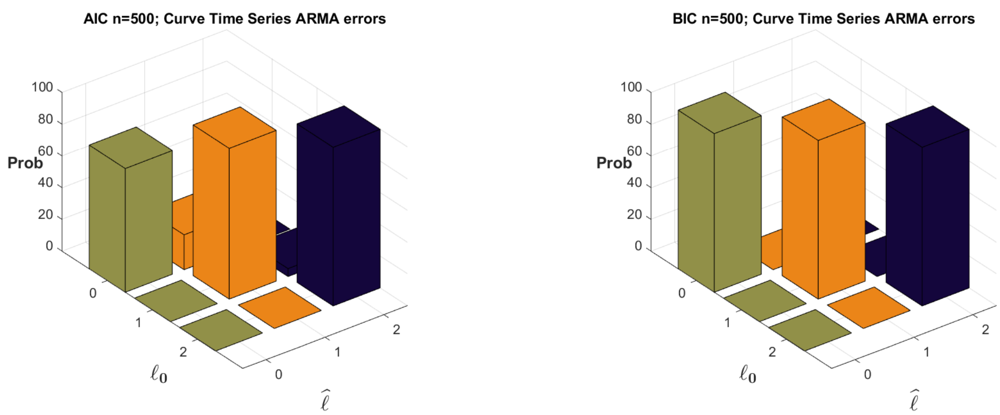


Figure 7: Selection probabilities by AIC, BIC, HQ and log(HQ) for cointegrating rank in a curve time series reduced rank regression with BM curve ARMA(1,1) time series errors and $n = 500$