Inference on the cointegration and the attractor spaces via functional approximation

Massimo Franchi, Paolo Paruolo

January 16, 2025

Abstract. This paper discusses semiparametric inference on hypotheses on the cointegration and the attractor spaces for $I(1)$ linear processes, using canonical correlation analysis and functional approximation of Brownian Motions. It proposes inference criteria based on the estimation of the number of common trends in various subsets of variables, and compares them to sequences of tests of hypotheses. The exact limit distribution for one of the test statistics is derived in the univariate case. Properties of the inferential tools are discussed theoretically and illustrated via a Monte Carlo study. An empirical analysis of exchange rates is also included.

Key words and phrases. Unit roots; cointegration; $I(1)$; attractor space; cointegrating space; semiparametric inference; canonical correlation analysis

JEL Classification: C32, C33.

Acknowledgments: The views expressed in the paper are those of the authors and do not necessarily reflect the opinion of the institutions of affiliation.

M. Franchi, Sapienza University of Rome, ORCID: 0000-0002-3745-2233 E-mail: massimo.franchi@uniroma1.it, Address: P.le A. Moro 5, 00185 Rome, Italy.

P. Paruolo. European Commission, Joint Research Centre (JRC) ORCID: 0000-0002-3982-4889, Email: paolo.paruolo@ec.europa.eu, Address: Via Enrico Fermi 2749, 21027 Ispra (VA), Italy.

Contents

1. Introduction

The notion of cointegration introduced in Engle and Granger (1987) and its link to the existence of an Equilibrium Correction Mechanism (ECM), see Davidson et al. (1978), have greatly influenced the econometric analysis of multiple time series. The introduction of the Gaussian Maximum Likelihood (ML) estimator of the ECM within a Vector Autoregressive (VAR) model in Johansen (1988, 1991) provided a set of inferential tools and shaped the understanding of the cointegration (CI) space.

Some hypotheses on the CI space were introduced together with the ML estimator. For instance, Johansen (1991, eq. (3.1)) considered the hypothesis that the CI space $\operatorname{col}\beta$ is contained in a pre-specified space $\operatorname{col}H$, where $\operatorname{col}H$ indicates the span of the columns of the matrix H and β is the CI matrix. Soon after, Johansen and Juselius (1992, eq. (15)) considered the converse hypothesis that the cointegration space $\operatorname{col}\beta$ contains a pre-specified subspace $\operatorname{col}h$; see also Johansen (1996, Chapter 7.2) for a summary. These hypotheses can be formulated and tested prior to the identification of the CI parameter matrix β, and tests are informative of economic theory implications that go beyond the ones related to the dimension of $\operatorname{col}\beta$, the cointegration rank.

The present paper considers the hypotheses $\operatorname{col}b \subseteq \operatorname{col}\beta \subseteq \operatorname{col}B$ in the semiparametric context of Franchi et al. (2024), henceforth FGP. They consider a p-dimensional linear process $\Delta X_t = C(L)\varepsilon_t$ with Common Trends (CT) representation

$$X_t = \gamma + \psi\kappa' \sum_{i=1}^{t} \varepsilon_i + C_1(L)\varepsilon_t, \qquad t = 1, 2, \ldots, \tag{1.1}$$

where L (respectively $\Delta = 1 - L$) is the lag (respectively difference) operator, γ is a vector of initial values, ε_t is a vector white noise, $\kappa' \sum_{i=1}^{t} \varepsilon_i$ are $0 \leq s \leq p$ stochastic trends, $\psi := \beta_\perp$ is a $p \times s$ full column rank loading matrix, and $C_1(L)\varepsilon_t$ is an $I(0)$ linear process; here $C(L)$ and $C_1(L)$ are infinite matrix polynomials and $\beta_\perp$ is a basis of the orthogonal complement of $\operatorname{col}\beta$.

This paper exploits the fact that the hypotheses $\operatorname{col}b \subseteq \operatorname{col}\beta \subseteq \operatorname{col}B$ can be reformulated as $\operatorname{col}a \subseteq \operatorname{col}\psi \subseteq \operatorname{col}A$ in terms of the loading matrix ψ in (1.1) with $a = B_\perp$, $A = b_\perp$, and using properties of the estimators and inference procedures introduced in FGP. Inference criteria for the validity of restrictions of the type $\operatorname{col}a \subseteq \operatorname{col}\psi$ or $\operatorname{col}\psi \subseteq \operatorname{col}A$ are constructed, as well as appropriate asymptotic tests of these hypotheses.

The approach is based on the empirical canonical correlations between the $p \times 1$ vector of observables X_t and a $K \times 1$ vector of deterministic variables d_t, constructed as the first K elements of an orthonormal $L^2[0, 1]$ basis discretized over the equispaced grid $1/T, 2/T, \ldots, 1$. In the asymptotic analysis, the cross-sectional dimension p is fixed while T and K diverge sequentially, K after T, denoted by $(T, K)_{seq} \to \infty$.

The inferential procedures proposed in this paper are based on the *dimensional coherence* of the FGP semiparametric approach, which guarantees that assumptions that hold for the whole of X_t also coherently apply to linear combinations of X_t, see Johansen and Juselius (2014) for a discussion of this property in VARs. Dimensional coherence is used here to address inference on the CI and attractor subspaces.

The remainder of the paper is organized as follows. The rest of the introduction defines notation, Section 2 discusses the assumptions on the Data Generating Process (DGP); Section 3 reports the well known weak convergence results of the process to a Brownian motion and its L^2 representation; Section 4 discusses the hypotheses of interest; Section 5 presents the proposed inference procedures; Section 6 discusses their asymptotic properties; Section 7 contains a Monte Carlo experiment; Section 8 reports an empirical application to exchange rates and Section 9 concludes. Appendix A contains proofs.

Notation. The following notation is employed in the rest of the paper. For any matrix $a \in \mathbb{R}^{p\times q}$, $\operatorname{col} a$ indicates the linear subspace of $\mathbb{R}^p$ spanned by the columns of a; for any full column rank matrix a, the orthogonal projection matrix onto $\operatorname{col} a$ is $\bar{a}a' = a\bar{a}'$, where $\bar{a} := a(a'a)^{-1}$, and $a_\perp$ of dimension $p \times (p-q)$ denotes a matrix whose columns span the orthogonal complement of $\operatorname{col} a$, i.e. $\operatorname{col} a_\perp = (\operatorname{col} a)^\perp$. The symbol e_j indicates the j-th column of the identity matrix I.

For any $a \in \mathbb{R}$, $\lfloor a \rfloor$ and $\lceil a \rceil$ denote the floor and ceiling functions; for any condition c, $1(c)$ is the indicator function of c, taking value 1 if c is true and 0 otherwise. The space of right continuous functions $f : [0,1] \mapsto \mathbb{R}^n$ having finite left limits, endowed with the Skorokhod topology, is denoted by $D_n[0,1]$, see Jacod and Shiryaev (2003, Ch. VI) for details, also abbreviated to $D[0,1]$ if $n = 1$. A similar notation is employed for the space of square integrable functions $L^2_n[0,1]$, abbreviated to $L^2[0,1]$ if $n = 1$. Riemann integrals $\int_0^1 a(u)b(u)'\mathrm{d}u$, are abbreviated as $\int ab'$. Finally the complement of set $\mathcal{A}$ is indicated as $\mathcal{A}^{\complement}$.

2. Data Generating Process

This section introduces the assumptions on the DGP, their dimensional coherence and the hypothesis of interest.

2.1. **Assumptions.** Let $\{X_t\}_{t\in\mathbb{Z}}$ be a $p \times 1$ linear process generated by

$$\Delta X_t = C(L)\varepsilon_t, \tag{2.1}$$

where L is the lag operator, $\Delta := 1 - L$, $C(z) = \sum_{n=0}^{\infty} C_n z^n$, $z \in \mathbb{C}$, satisfies Assumption 2.1 below, and the innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are independently and identically distributed (i.i.d.) with expectation $\mathbb{E}(\varepsilon_t) = 0$, finite moments of order $2 + \epsilon$, $\epsilon > 0$, and positive definite variance-covariance matrix $\mathbb{V}(\varepsilon_t) = \Omega_\varepsilon$, indicated as ε_t i.i.d. $(0, \Omega_\varepsilon)$, $\Omega_\varepsilon > 0$.[1]

Assumption 2.1 (Assumptions on $C(z)$)**.** *Let $C(z)$ in (2.1) be of dimension $p \times n_\varepsilon$, $n_\varepsilon \geq p$, and satisfy the following conditions:*

(i) *$C(z) = \sum_{n=0}^{\infty} C_n z^n$ converges for all $|z| < 1 + \delta$, $\delta > 0$;*

(ii) *$\operatorname{rank} C(z) < p$ only at isolated points $z = 1$ or $|z| > 1$; in particular for the point $z = 1$, consider the first order expansion $C(z) = C + C_1(z)(1 - z)$,[2] with matrix $C := C(1)$ of rank s and rank-factorization $C = \psi\kappa'$, where ψ and κ are full column rank s matrices whose columns form bases of $\operatorname{col} C$ and $\operatorname{col}(C')$ respectively;*

(iii) *$\psi'_\perp C_1 \kappa_\perp$ has full row rank $r := p - s$, with $C_1 := C_1(1)$.*

Cumulating (2.1) under Assumption 2.1, one finds the Common Trends (CT) representation

$$X_t = \gamma + \psi\kappa' \sum_{i=1}^{t} \varepsilon_i + C_1(L)\varepsilon_t, \qquad t = 1, 2, \ldots, \tag{2.2}$$

which gives a decomposition of X_t into initial conditions $\gamma := X_0 - C_1(L)\varepsilon_0$, the s-dimensional random walk component $\kappa' \sum_{i=1}^{t} \varepsilon_i$ with loading matrix ψ and the I(0) component $C_1(L)\varepsilon_t$. Initial conditions can be treated as in Johansen (1996) setting $x_t = X_t$ or as in Elliott (1999), see also Müller and Watson (2008), setting $x_t = X_t - X_0$, where x_t are the variables that are used in the statistical analysis.

The matrix ψ in (2.2) is $p \times s$ with columns that form a basis of $\operatorname{col} C$ called the attractor space of X_t. Its orthogonal complement $(\operatorname{col} C)^\perp$ is the cointegrating space of X_t, and the columns of $\beta := \psi_\perp$ (of dimension $p \times r$) form a basis of the cointegrating space. The number $r = \dim \operatorname{col} \beta$ is called the cointegrating rank, $s = \dim \operatorname{col} \psi$ is called the number of CT, and $0 \leq s \leq p$, $p = r + s$.

Observe that X_t is $I(0)$ if $s = 0$, i.e. $r = p$; it is $I(1)$ and cointegrated if $0 < r, s < p$ and it is $I(1)$ and non-cointegrated if $s = p$, i.e. $r = 0$. Assumption 2.1 includes possibly nonstationary

[1] The i.i.d. assumption can be relaxed to a martingale difference sequence; this is not done here for simplicity.

[2] This is called the Beveridge-Nelson (BN) decomposition in Phillips and Solo (1992).

VARMA processes (for $n_\varepsilon = p$) in line with e.g. Stock and Watson (1988) and Dynamic Factor Models (for $n_\varepsilon > p$) as in e.g. Bai (2004).

2.2. **Dimensional coherence.** The assumptions on the DGP of X_t imply that they also apply to linear combinations of X_t. In fact whenever X_t has a CT representation (2.2), linear combinations of X_t also admit a CT representation, with a number of stochastic trends that is at most equal to that in X_t; see Johansen and Juselius (2014) for the discussion of this aspect in a VAR context.

Theorem 2.2 (Dimensional coherence, FGP)**.** *Let H be a $p \times m$ full column rank matrix; if X_t satisfies (2.1) with $C(z)$ fulfilling Assumption 2.1, then the same holds for $H'X_t$, i.e. $\Delta H'X_t = G(L)\varepsilon_t$, where $G(z) := H'C(z)$ satisfies Assumption 2.1 with $G(1) = H'C$ of rank $j \leq s$.*

2.3. **Hypotheses of interest.** The hypotheses of interest are formulated in terms of ψ as

$$\operatorname{col} a \subseteq \operatorname{col} \psi \subseteq \operatorname{col} A \tag{2.3}$$

where a and A are full column rank matrices, or equivalently in terms of β as

$$\operatorname{col} b \subseteq \operatorname{col} \beta \subseteq \operatorname{col} B,$$

where $b := A_\perp$ and $B := a_\perp$. The aim of the paper is to make inferences about (2.3) from a sample $\{X_t\}_{t=0,1,\ldots,T}$ observed from (2.2).

3. Functional central limit theorem and L^2-representation

This section reports the weak convergence results to Brownian motions and their L^2 representation. By the functional central limit theorem, see Phillips and Solo (1992), the partial sums $T^{-\frac{1}{2}} \sum_{i=1}^{t} \varepsilon_i$ in the CT representation (2.2) converge weakly to an n_ε-dimensional Brownian Motion $W_\varepsilon(u)$, $u \in [0,1]$, with variance $\Omega_\varepsilon > 0$; this implies that, with $t = \lfloor Tu \rfloor \in \mathbb{N}$ and $u \in [0,1]$, one has

$$T^{-\frac{1}{2}} \begin{pmatrix} \bar{\psi}' x_t \\ \beta' \sum_{i=1}^{t} x_i \end{pmatrix} = T^{-\frac{1}{2}} \begin{pmatrix} \kappa' \\ \beta' C_1 \end{pmatrix} \sum_{i=1}^{\lfloor Tu \rfloor} \varepsilon_i + o_p(1) \underset{T\to\infty}{\overset{w}{\to}} \begin{pmatrix} W_1(u) \\ W_2(u) \end{pmatrix} := \begin{pmatrix} \kappa' \\ \beta' C_1 \end{pmatrix} W_\varepsilon(u), \tag{3.1}$$

where $\overset{w}{\to}$ indicates weak convergence of probability measures on $D_p[0,1]$, $W(u) := (W_1(u)', W_2(u)')'$ is a $p \times 1$ Brownian motion in $u \in [0,1]$ with variance matrix $\Omega := D\Omega_\varepsilon D'$, $D := (\kappa, C_1'\beta)'$; the $o_p(1)$-term is infinitesimal uniformly in $u \in [0,1]$ and Ω is the long run variance of the l.h.s. of (3.1). Note that D is nonsingular by Assumption 2.1.(iii). It is useful to partition Ω conformably with $W(u) = (W_1(u)', W_2(u)')'$ as $\Omega = (\Omega_{ij})_{i,j=1,2}$ and define the standardized version of $W_1(u)$ as $B_1(u) := \Omega_{11}^{-\frac{1}{2}} W_1(u)$.

Recall, see e.g. Phillips (1998), that any n-dimensional standard Brownian motion $B(u)$ admits the representation

$$B(u) \simeq \sum_{k=1}^{\infty} c_k \phi_k(u), \qquad c_k := \int_0^1 B(u)\phi_k(u)\mathrm{d}u = \nu_k \xi_k, \qquad \nu_k \in \mathbb{R}_+, \qquad \xi_k \sim N(0, Q), \tag{3.2}$$

where $Q = I$, $\{\phi_k(u)\}_{k=1}^{\infty}$, $\int_0^1 \phi_j(u)\phi_k(u)\mathrm{d}u = 1_{j=k}$, is an orthonormal basis of $L^2[0,1]$ and $\simeq$ indicates that the series in (3.2) is a.s. convergent in the L^2 sense to the l.h.s.

In the special case where $(\nu_k^2, \phi_k(u))$, $k = 1, 2, \ldots$, is an eigenvalue-eigenvector pair of the covariance kernel of the standard Brownian motion, i.e. $\nu_k^2 \phi_k(u) = \int_0^1 \min(u,v)\phi_k(v)\mathrm{d}v$, (3.2) is the Karhunen-Loève (KL) representation of $B(u)$, for which one has

$$\nu_k = \frac{1}{(k - \frac{1}{2})\pi}, \qquad \phi_k(u) = \sqrt{2}\sin\left(u/\nu_k\right), \qquad \xi_k \text{ i.i.d. } N(0, I_n). \tag{3.3}$$

In this case the series in (3.2) is a.s. uniformly convergent in u and $\simeq$ is replaced by $=$. In the following a basis $\{\phi_k(u)\}_{k=1}^{\infty}$ is indicated as the KL basis when $\phi_k(u)$ is chosen as in (3.3). From (3.2) it follows that the L^2-representation of $W(u)$ in (3.1) is given by (3.2) with $Q = \Omega$, and $\simeq$ is replaced by $=$ when the KL basis is employed.

4. Hypotheses of interest

This section discusses the hypotheses of interest and their implications on the DGP. It also relates them to the dimensional coherence property of the assumptions in Section 2.

4.1. **Formulation.** Let X_t be generated as in (2.2). Inference on the number s of CT can be addressed as in FGP, who define estimators of it, also accompanied by misspecificaton tests. Once s has been determined, one can consider hypotheses on the cointegrating space $\operatorname{col}\beta$ or the attractor space $\operatorname{col}\psi$ of the following form:

$$H_0^1: \qquad \operatorname{col}\psi \subseteq \operatorname{col} A \qquad \Leftrightarrow \qquad \operatorname{col} b \subseteq \operatorname{col}\beta \tag{4.1}$$

$$H_0^2: \qquad \operatorname{col} a \subseteq \operatorname{col}\psi \qquad \Leftrightarrow \qquad \operatorname{col}\beta \subseteq \operatorname{col} B \tag{4.2}$$

where A is $p \times m$, a is $p \times q$, both of full column rank, $q \le s \le m$, and $b := A_\perp$, $B := a_\perp$.

The idea is to test the implication of hypotheses (4.1) and (4.2); these implications are discussed in the following in terms of ψ only.

4.2. **Implications of H_0^1.** One can formulate the restriction H_0^1 in (4.1) as $\psi = A\theta$ where θ is $m \times s$; from (2.2) one finds

$$\bar{A}' X_t = \theta\kappa' \sum_{i=1}^{t} \varepsilon_i + \bar{A}' C_0(L)\varepsilon_t + \bar{A}'\gamma \tag{4.3}$$

$$A'_\perp X_t = \qquad\qquad + A'_\perp C_0(L)\varepsilon_t + A'_\perp\gamma \tag{4.4}$$

where $\text{rank}(\theta\kappa') = s$. Hence the implications of H_0^1 are that (i) $\bar{A}' X_t$ contains s common trends and (ii) $A'_\perp X_t$ contains 0 common trends.

4.3. **Implications of H_0^2.** One can formulate H_0^2 in (4.2) as $\psi = (a, a_\perp\theta)$, where a is $p \times q$, $a_\perp$ is $p \times (p-q)$, and θ is $(p-q) \times (s-q)$, with the convention that empty matrices are absent, i.e. $\psi = a$ when $q = s$. Partition also κ conformably with ψ as $\kappa =: (\kappa_1, \kappa_2)$ with q and $s-q$ columns respectively, so that $\psi\kappa' = a\kappa'_1 + a_\perp\theta\kappa'_2$; one finds

$$\bar{a}' X_t = \kappa'_1 \sum_{i=1}^{t} \varepsilon_i + \bar{a}' C_0(L)\varepsilon_t + \bar{a}'\gamma \tag{4.5}$$

$$\bar{a}'_\perp X_t = \theta\kappa'_2 \sum_{i=1}^{t} \varepsilon_i + \bar{a}'_\perp C_0(L)\varepsilon_t + \bar{a}'_\perp\gamma \tag{4.6}$$

where $\text{rank}(\theta\kappa'_2) = s-q$. Hence the implications of H_0^2 are that (i) $\bar{a}' X_t$ contains q common trends and (ii) $\bar{a}'_\perp X_t$ contains $s-q$ common trends.

4.4. **Formulation of the null and the alternative.** The hypotheses are restated in terms of the subsystems in (4.3), (4.4) and (4.5), (4.6); this gives rise to pairs of subhypotheses H_{01}^j and H_{02}^j, $j = 1, 2$, that need to hold jointly[3]

$$H_0^1: \quad H_{01}^1 \text{ and } H_{02}^1, \qquad\qquad H_0^2: \quad H_{01}^2 \text{ and } H_{02}^2, \tag{4.7}$$

where

$$H_{01}^1: \quad \text{rank}(A'\psi) = s, \qquad\qquad H_{02}^1: \quad \text{rank}(A'_\perp\psi) = 0,$$

$$H_{01}^2: \quad \text{rank}(a'\psi) = q, \qquad\qquad H_{02}^2: \quad \text{rank}(a'_\perp\psi) = s-q.$$

The alternative hypotheses are

$$H_1^1: \text{col}(\psi) \nsubseteq \text{col}(A) \qquad\qquad H_1^2: \text{col}(a) \nsubseteq \text{col}(\psi) \tag{4.8}$$

[3]The word 'and' is used in place of the logical operator $\wedge$.

which can be similarly decomposed into subhypotheses H_{11}^j and H_{12}^j, $j = 1, 2$, where the alternative H_1^j holds if at least one of them holds,[4] i.e.

$$H_1^1: \quad H_{11}^1 \text{ or } H_{12}^1, \qquad\qquad H_1^2: \quad H_{11}^2 \text{ or } H_{12}^2, \tag{4.9}$$

where

$$\begin{aligned} H_{11}^1: &\quad \operatorname{rank}(A'\psi) < s, & H_{12}^1: &\quad \operatorname{rank}(A_\perp'\psi) > 0, \\ H_{11}^2: &\quad \operatorname{rank}(a'\psi) < q, & H_{12}^2: &\quad \operatorname{rank}(a_\perp'\psi) > s - q. \end{aligned}$$

Remark 4.1 (Not all combinations of subhypotheses are possible)*.* Note that H_{02}^1 and H_{11}^1 is not a possible pair of subhypotheses. In fact

$$s = \operatorname{rank}(\psi) = \operatorname{rank}((\bar{A}, \bar{A}_\perp)(A, A_\perp)'\psi) = \operatorname{rank}\begin{pmatrix} A'\psi \\ A_\perp'\psi \end{pmatrix} \leq \operatorname{rank}\left(A'\psi\right) + \operatorname{rank}\left(A_\perp'\psi\right) \tag{4.10}$$

and hence if $\operatorname{rank}(A'\psi) < s$ as in H_{11}^1 and $\operatorname{rank}(A_\perp'\psi) = 0$ as in H_{02}^1 this contradicts (4.10). Note that, on the other hand, $\operatorname{rank}(A'\psi) = s$ as in H_{01}^1 and $\operatorname{rank}(A_\perp'\psi) > 0$ as in H_{12}^1 does not violate (4.10), and hence H_{01}^1 and H_{12}^1 is a possible pair of subhypotheses.

Similarly, replacing A with a in (4.10) one finds that H_{02}^2 and H_{11}^2 is not a possible pair of subhypotheses, while H_{01}^2 and H_{12}^2 is a possible pair.

4.5. **Example.** The following example illustrate various cases, with reference to H_i^2, $i = 0, 1$. Let

$$\psi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = (e_1, e_2 + e_3)$$

and $a = e_1$, so that $s = 2$ and $q = 1$. One finds $\operatorname{rank}(a'\psi) = \operatorname{rank}(1, 0) = 1 = q$ as in H_{01}^2 and $\operatorname{rank}(a_\perp'\psi) = \operatorname{rank}\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = 1 = s - q$, as in H_{02}^2; hence H_0^2 holds.

Consider next $a = e_2$; one finds $\operatorname{rank}(a'\psi) = \operatorname{rank}(0, 1) = 1 = q$ which still falls under the null H_{01}^2, and $\operatorname{rank}(a_\perp'\psi) = \operatorname{rank}(I_2) = 2 > s - q$, which falls under the alternative H_{12}^2. This illustrates that H_1^2 holds, with only one component of the alternative being verified, namely H_{12}^2.

Finally consider $a = e_2 - e_3$; one finds $\operatorname{rank}(a'\psi) = \operatorname{rank}(0, 0) = 0 < q = 1$ which falls under H_{11}^2 and $\operatorname{rank}(a_\perp'\psi) = \operatorname{rank}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = 2 > s - q = 1$ which falls under the alternative H_{12}^2. In this case H_1^2 holds, with both components of the alternative being verified, namely H_{11}^2 and H_{12}^2. Note that this illustrates all possibilities, see Remark 4.1.

[4]The word 'or' is used in place of the logical operator $\vee$.

4.6. **Dimensional coherence.** The property of dimensional coherence in Theorem 2.2 can be applied to sub-systems (4.3), (4.4), (4.5), (4.6) with $H = \bar{A}, A_\perp, \bar{a}_\perp, \bar{a}$.

5. Statistical analysis

This section reviews the estimators of the number s of stochastic trends (and the cointegrating rank r) proposed in FGP. The estimators are based on canonical correlation analysis, which is introduced first.

5.1. **Canonical correlations.** Let $\varphi_K(u) := (\phi_1(u), \dots, \phi_K(u))'$ be a $K \times 1$ vector function of $u \in [0,1]$, where $\phi_1(u), \dots, \phi_K(u)$ are the first K elements of some fixed orthonormal càdlàg basis of $L^2[0,1]$, see (3.2). Let d_t be the $K \times 1$ vector constructed evaluating $\varphi_K(\cdot)$ at the discrete sample points $1/T, \dots, (T-1)/T, 1$, i.e.

$$d_t := \varphi_K(t/T) := (\phi_1(t/T), \dots, \phi_K(t/T))', \qquad K \geq p, \qquad t = 1, \dots, T. \tag{5.1}$$

For a generic p-dimensional variable y_t observed for $t = 1, \dots, T$, the sample canonical correlation analysis of y_t and d_t in (5.1), denoted as $\texttt{cca}(y_t, d_t)$,[5] consists in solving the following generalized eigenvalue problem, see e.g. Johansen (1996) and references therein,

$$\texttt{cca}(y_t, d_t): \qquad |\lambda M_{yy} - M_{yd} M_{dd}^{-1} M_{dy}| = 0, \qquad M_{ij} := T^{-1} \sum_{t=1}^{T} i_t j_t'. \tag{5.2}$$

This delivers eigenvalues $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, collected in $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$, and corresponding eigenvectors $V = (v_1, \dots, v_p)$, organized as

$$\Lambda_1 := \mathrm{diag}(\lambda_1, \dots, \lambda_s), \quad \Lambda_0 := \mathrm{diag}(\lambda_{s+1}, \dots, \lambda_p), \quad V_1 := (v_1, \dots, v_s), \quad V_0 := (v_{s+1}, \dots, v_p), \tag{5.3}$$

with the convention that empty matrices are absent, i.e. $(\Lambda_1, V_1) = (\Lambda, V)$ when $s = p$ and $(\Lambda_0, V_0) = (\Lambda, V)$ when $s = 0$.

5.2. **Selection criteria on the number of common trends.** This section present selection criteria for the determination of the number of common trends s, as well as its complement r, the cointegration rank.[6] They are functions of the observable variables y_t say, which can equal x_t or some linear combinations of it.

[5]The notation $\texttt{cca}(y_t, d_t)$ is a shorthand for $\texttt{cca}(\{y_t\}_{t=1}^T, \{d_t\}_{t=1}^T)$; similar abbreviations are used in the following.

[6]The selection criteria produce an estimate of the number of common trends; hence they can also be called estimators of the number of CT.

The estimators (both the ones based on criteria and the ones based on sequences of tests) are denoted as $\widehat{s}(y_t)$, $\widetilde{s}(y_t)$, and $\breve{s}(y_t)$, generically indicated as $\check{s}(y_t)$, to specify the observable variables y_t on which the corresponding $\texttt{cca}(y_t, d_t)$ is performed. The symbol $\check{s}$ represent a function of the data y_t (a statistics) that depends on the output of $\texttt{cca}(y_t, d_t)$, and it is an estimator of the number of CT for y_t.

The estimators are introduced as functions of $\texttt{cca}(x_t, d_t)$ as $\check{s}(x_t)$, and in later section they are modified to apply to $\texttt{cca}(H'x_t, d_t)$ for various choices of H; in the latter case x_t in the definition needs to be replaced by $H'x_t$ for the specific choice of H, see Section 4.6.

Definition 5.1 (Selection criteria for s and r)**.** *Define the maximal gap (max-gap) estimator of s as*

$$\widehat{s}(x_t) := \underset{i\in\{0,\dots,p\}}{\operatorname{argmax}}(\lambda_i - \lambda_{i+1}), \qquad \lambda_0 := 1, \qquad \lambda_{p+1} := 0, \tag{5.4}$$

where $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ are the eigenvalues of $\texttt{cca}(x_t, d_t)$ and x_t. Define also the following alternative argmax estimator*:*

$$\widetilde{s}(x_t) := \underset{i\in\{0,1,\dots,p\}}{\operatorname{argmax}} \frac{\prod_{h=1}^{i} \lambda_h}{\prod_{h=i+1}^{p} \left(\frac{T}{K}\lambda_h\right)}, \tag{5.5}$$

where empty products are equal to 1, and $\widetilde{s}$ uses the rates in (6.1) below for $(T, K)_{seq} \to \infty$. Finally define the estimators of r as $\check{r} := p - \check{s}$ for $\check{s} = \widehat{s}, \widetilde{s}$.

Remark 5.2 (Alternative criteria). The criterion (5.5) applies ideas in Bierens (1997) to the set of eigenvalues in $\texttt{cca}(x_t, d_t)$. Alternative criteria inspired by Ahn and Horenstein (2013) when applied to eigenvalues in $\texttt{cca}(x_t, d_t)$ are discussed in FGP; they replace the function in the argmax in (5.5) with

$$\frac{\lambda_i}{\lambda_{i+1}} \qquad \text{or} \qquad \frac{\log\left(1 + \lambda_i / \sum_{h=i+1}^{p} \lambda_h\right)}{\log\left(1 + \lambda_{i+1} / \sum_{h=i+2}^{p} \lambda_h\right)}, \tag{5.6}$$

where the optimization is over the set of integers $0 \leq i \leq p-1$ for the first function and $0 \leq i \leq p-2$ for the second function. Because these criteria do not cover the values p (and also $p-1$ for the last function), they are not considered in the rest of the paper.

5.3. **Test sequences.** This section reviews test sequences in FGP for the determination of the number of common trends s, as well as its complement r, the CI rank, in x_t, based on $\tau^{(i)}$ see (5.8) below.

The test-based estimator $\breve{s}_n$ is similar to the one described in Johansen (1996) for CI rank determination. It depends on n, the order of the n-norm for vectors $\|\cdot\|_n$, and on the significance level η in each test; two special cases are of interest are $n = 1, \infty$. The special case $n = 1$ gives $\|\tau^{(j)}\|_1 = \sum_{k=1}^{j}(1-\lambda_k)$ while $n = \infty$ gives $\|\tau^{(j)}\|_\infty = 1 - \lambda_j$; these forms are similar to the trace and $\lambda_{\max}$ statistics in Johansen (1996).

The procedure can be described as follows: consider hypothesis $H_j : s = j$ versus $H_j^A : s < j$ with test statistics $J_n(\tau^{(j)})$ and rejection region $\mathcal{R}_{j,n,\eta}$ in the right tail, with significance level η. Start with the test of H_p; if the test does not reject, then $\breve{s}_n = p$, otherwise progress further. Test H_j for $j = p-1, \dots, 1$ and proceed as before until H_m is not rejected; in this case $\breve{s}_n = m$. Otherwise all tests of H_j, $j = p, \dots, 1$ reject and one sets $\breve{s}_n = 0$.

The following definition introduces the estimators of the number of CT based on tests.

Definition 5.3 (Estimators for s and r based on tests)**.** *The test-based estimator of s at significance level η is defined as*

$$\{\breve{s}_n(x_t) = j\} := \begin{cases} \mathcal{A}_{p,n,\eta} & j = p \\ (\cap_{i=j+1}^{p}\mathcal{R}_{i,n,\eta}) \cap \mathcal{A}_{j,n,\eta} & j = 1, \dots, p-1 \\ \cap_{i=1}^{p}\mathcal{R}_{i,n,\eta} & j = 0 \end{cases}, \tag{5.7}$$

$$\mathcal{R}_{i,n,\eta} := \{J_n(\tau^{(i)}) > c_{i,n,\eta}\}, \qquad \mathcal{A}_{i,n,\eta} := \mathcal{R}_{i,n,\eta}^{\complement} = \{J_n(\tau^{(i)}) \le c_{i,n,\eta}\},$$

$$J_n(\tau^{(i)}) := K\pi^2\|\tau^{(i)}\|_n, \qquad \tau^{(i)} := (1-\lambda_i, 1-\lambda_{i-1}, \dots, 1-\lambda_1)', \tag{5.8}$$

where $1 \ge \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$ are the eigenvalues of $\mathtt{cca}(x_t, d_t)$, $\|\cdot\|_n$ is the n-norm for vectors, and $c_{i,n,\eta}$ are appropriate quantiles derived in Theorem 6.2(iii) below. The corresponding estimator of r is defined as $\breve{r}_n := p - \breve{s}_n$ where $\breve{s}_n$ is as in (5.7).

Note that $\breve{s}_n$ is a function of the eigenvalues of $\mathtt{cca}(x_t, d_t)$.

5.4. **Decision rules on H_0^1 and H_0^2.** This section describes decision rules for rejecting H_0^1 and H_0^2, see Section 4.4. The rules are expressed as functions of a generic estimator of the number of common trends $\breve{s} = \widehat{s}, \widetilde{s}, \breve{s}$, where $\breve{s} = \breve{s}_n$, $n = 1, \infty$.

The decision rule *does not reject* H_0^1 whenever z_1 equals 1, where

$$\breve{z}_1 := \breve{w}_1 \cdot \breve{v}_1, \qquad \breve{w}_1 := 1(\breve{s}(\bar{A}'x_t) = s), \qquad \breve{v}_1 := 1(\breve{s}(A_\perp' x_t) = 0). \tag{5.9}$$

Similarly, the decision rule *does not reject* H_0^2 whenever z_2 equals 1, where

$$\breve{z}_2 := \breve{w}_2 \cdot \breve{v}_2, \qquad \breve{w}_2 := 1(\breve{s}(\bar{a}'x_t) = q), \qquad \breve{v}_2 := 1(\breve{s}(\bar{a}_\perp' x_t) = s - q), \tag{5.10}$$

Note that s, m and q are given in the formulation of the hypothesis (4.1) and (4.2), and they appear in (5.9), (5.10). Note that z_j checks both implications of H_0^j, contained in the indicators w_j and v_j, $j = 1, 2$.

For the special case $\breve{s} = \breve{s}$ one needs to define the significance levels involved in the estimators in (5.9) or (5.10). Let η_1 indicate the significance level that appears in w_j and η_2 be the significance level in the estimator involved in $\breve{v}_j$, $j = 1, 2$.

Remark 5.4 (Indicator structure of decision rules). Note that the indicator structure in (5.9) and (5.10) follows the decomposition of the null hypotheses H_0^j in (4.7) into H_{01}^j and H_{02}^j and of the alternative hypotheses H_1^j in (4.9) into H_{11}^j or H_{12}^j. In fact $\breve{z}_j$, $j = 1, 2$, in (5.9) and (5.10) select H_0^j if *both* H_{01}^j and H_{02}^j are not rejected. Equivalently, $\breve{z}_j$, $j = 1, 2$, rejects H_0^j if *either* H_{11}^j or H_{12}^j is selected, or both.

Asymptotic properties of these decision rules are discussed in the following section.

6. Asymptotic properties

This section discusses the asymptotic properties of the statistics introduced in Section 5. A summary of asymptotic FGP properties is first given, together with a new derivation of the explicit expression of the relevant limit distribution in the univariate case; these results are then used to develop asymptotic properties of the proposed inference procedures.

6.1. **General asymptotic properties.**

This section summarizes some background results taken from FGP that are relevant for inference on the hypotheses of interest.

Let $a_{TK} \overset{p}{\asymp} g(T, K)$ for some function g indicate that $a_{TK}/g(T, K)$ is bounded and bounded away from 0 in probability as $(T, K)_{seq} \to \infty$. For an array a_{TK} of random variables and a real number a, the statement that a_{TK} converges to a in probability as $(T, K)_{seq} \to \infty$ means that $\lim_{K\to\infty} \lim_{T\to\infty} \mathbb{P}(|a_{TK} - a| < \delta) = 1$ for every $\delta > 0$. The asymptotic behavior of canonical correlations is given in the following theorem.

Theorem 6.1 (Limits of the eigenvalues of $\texttt{cca}(x_t, d_t)$, FGP)**.** *Let d_t in* (5.1) *be constructed using any orthonormal càdlàg basis of $L^2[0, 1]$ and let $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ be the eigenvalues of* $\texttt{cca}(x_t, d_t)$*, see* (5.2) *and* (5.3)*; then for $(T, K)_{seq} \to \infty$,*

$$\lambda_i \overset{p}{\asymp} \begin{cases} 1 & i = 1, \dots, s \\ K/T & i = s + 1, \dots, p \end{cases}, \tag{6.1}$$

for $(T, K)_{seq} \to \infty$, and because $K/T \to 0$,

$$\lambda_i \overset{p}{\to} \begin{cases} 1 & i = 1, \dots, s \\ 0 & i = s + 1, \dots, p \end{cases}. \tag{6.2}$$

Moreover $\mathbb{P}(\breve{s} = s) \to 1$ as $(T, K)_{seq} \to \infty$ for any $0 \leq s \leq p$ and $\breve{s} = \widehat{s}, \widetilde{s}$.

Theorem 6.2 (Asymptotic distribution of the eigenvalues with the KL basis, FGP)**.** *Let $B_1(u)$ be defined as in Section 3, let d_t in (5.1) be constructed using the KL basis in (3.3) and let $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ be the eigenvalues of $\texttt{cca}(x_t, d_t)$; then for $(T, K)_{seq} \to \infty$,*

(i) *the s largest eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s$ satisfy*

$$K\pi^2 \tau^{(s)} \overset{w}{\to} \zeta^{(s)}, \tag{6.3}$$

where $\tau^{(s)} := (1-\lambda_s, 1-\lambda_{s-1}, \ldots, 1-\lambda_1)'$, $\zeta^{(s)} := (\zeta_1, \zeta_2, \ldots, \zeta_s)'$, and $\zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_s$ are the eigenvalues of

$$\omega := \left(\int B_1 B_1' \right)^{-1};$$

(ii) *the r smallest eigenvalues $\lambda_{s+1} \geq \lambda_{s+2} \geq \cdots \geq \lambda_p$ satisfy $K\pi^2(1-\lambda_i) \overset{p}{\to} \infty$, $i = s+1, \ldots, p$;*

(iii) *$J_n(\tau^{(s)}) \overset{w}{\to} \|\zeta^{(s)}\|_n$, while $J_n(\tau^{(j)}) \to \infty$ for $j > s$, where $J_n(\tau^{(i)}) := K\pi^2 \|\tau^{(i)}\|_n$, $\tau^{(i)} := (1-\lambda_i, 1-\lambda_{i-1}, \ldots, 1-\lambda_1)'$, see (5.8);*

(iv) *$\mathbb{P}(\breve{s}_n(x_t) = s) \to 1-\eta$ for $0 < s \leq p$ and $\mathbb{P}(\breve{s}_n(x_t) = s) \to 1$ for $s = 0$, where η is the significance level in Definition 5.3;*

(v) *$\mathbb{P}(\breve{s}_n(x_t) = j) \to 0$ for $j > s$.*

Observe that ω does not depend on nuisance parameters, and that tests based on $K\pi^2\tau^{(s)}$ are asymptotically pivotal. The limit distribution in Theorem 6.2 depends on the behaviour of the tail $\sum_{k=K+1}^{\infty} \nu_k^2$, see (3.2), and is therefore not invariant either to the choice of $L_2[0,1]$ basis or the ordering of its elements. This result is also used in FGP to define misspecification tools.

The quantiles of $\|\zeta^{(i)}\|_n$ in part (iii) can be estimated by Monte Carlo simulation; this allows to calculate quantiles $c_{i,n,\eta}$ that enter Definition 5.3 in Section 5.3.

6.2. **Univariate case.** This section contains the explicit form of the p.d.f. and c.d.f. of an asymptotic distribution for the special case $s = 1$, which is a novel result. The following theorem derives the relevant probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of $\zeta^{(1)} = \|\zeta^{(1)}\|_n = J_n(\tau^{(1)})$, $n = 1, \infty$.

Theorem 6.3 (Exact quantiles of $\zeta^{(1)}$)**.** *When $s = 1$, $\zeta := \zeta^{(1)}$ has p.d.f. $f_\zeta(z)$ and c.d.f. $F_\zeta(z)$ of the form*

$$f_\zeta(z) = \frac{1}{\sqrt{\pi z}} \sum_{j=0}^{\infty} \eta_j a_j \exp\left(-\frac{a_j^2}{2} z \right), \qquad F_\zeta(z) = 1 - \left(\frac{2}{\pi} \right)^{\frac{1}{2}} \sum_{j=0}^{\infty} \eta_j \Gamma\left(\frac{1}{2}, \frac{a_j^2}{2} z \right), \qquad z > 0$$

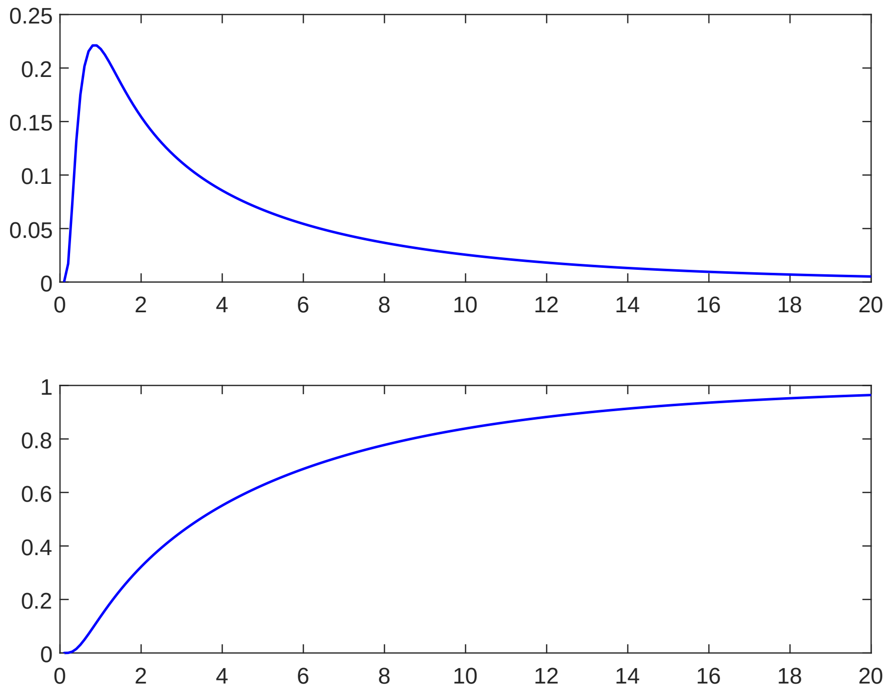

FIGURE 1. p.d.f. and c.d.f. of $\zeta^{(1)}$ in Theorem 6.3.

where $\eta_j := \binom{-\frac{1}{2}}{j}$, $a_j := 2j + \frac{1}{2}$ *and* $\Gamma(k,t) = \int_t^\infty z^{k-1}e^{-z}\mathrm{d}z$ *is the upper incomplete Gamma function.*[7] *Selected quantiles of this distributions are* $F^{-1}(0.90) = 13.06582$, $F^{-1}(0.95) = 17.71180$, $F^{-1}(0.99) = 29.01932$.

The result in Theorem 6.3 is related to Doornik et al. (2003, Theorem 3), who compute $1 - \frac{1}{2}\mathbb{E}\left(\int_0^1 B(u)^2\mathrm{d}u\right)^{-1} \simeq -1.78143$, which implies $\mathbb{E}(\zeta^{(1)}) \simeq 2(1 + 1.78143) = 5.5629$. This agrees with the analytical expectation that can be derived from the expression $f_\zeta(z)$ above, namely

$$\mathbb{E}(\zeta^{(1)}) = \int_0^\infty z f_\zeta(z)\mathrm{d}z = 2^{\frac{1}{2}} \sum_{j=0}^\infty \eta_j a_j^{-2} \simeq 5.56291.$$

Figure 1 pictures the p.d.f. and c.d.f. in Theorem 6.3.[8]

6.3. **Asymptotic properties of criteria on H_0^1 and H_0^2.** This section presents novel asymptotic properties of the criteria in (5.9), (5.10).

[7] $F_\zeta(0) = 0$.

[8] MATLAB and R code to compute the p.d.f., the c.d.f and the quantile functions in Theorem 6.3 are available on the website of one of the authors.

Theorem 6.4 (Limit behavior of decision rules)**.** *Let the assumptions of Theorem 6.2 hold, and consider $(T, K)_{seq} \to \infty$; then for $\breve{s} = \widehat{s}, \widetilde{s}$ under H_0^j, $j = 1, 2$, one has*

$$\mathbb{P}(\breve{z}_j = 1) \to 1, \qquad \mathbb{P}(\breve{w}_j = 1) \to 1, \qquad \mathbb{P}(\breve{v}_j = 1) \to 1, \tag{6.4}$$

while if H_0^j does not hold one finds $\mathbb{P}(\breve{z}_j = 0) \to 1$ because $\mathbb{P}(\breve{w}_j = 0) \to 1$ under H_{11}^j and $\mathbb{P}(\breve{v}_j = 0) \to 1$ under H_{12}^j. Next consider $\breve{s} = \check{s}$, and let η_1 be the significance level for the estimator in $\breve{w}_j$ and η_2 the significance level for the estimator in $\breve{v}_j$; then if $s > q$ under H_0^j, $j = 1, 2$, one has

$$\mathbb{P}(\breve{z}_1 = 1) \geq \rho_{1T} \to 1 - \eta_1, \qquad \mathbb{P}(\breve{w}_1 = 1) \to 1 - \eta_1, \qquad \mathbb{P}(\breve{v}_1 = 1) \to 1, \tag{6.5}$$

$$\mathbb{P}(\breve{z}_2 = 1) \to 1 - \eta_1 - \eta_2, \qquad \mathbb{P}(\breve{w}_2 = 1) \to 1 - \eta_1, \qquad \mathbb{P}(\breve{v}_2 = 1) \to 1 - \eta_2, \tag{6.6}$$

while if $s = q$ the statements in (6.5) hold for z_1, w_1, v_1 as well as for z_2, w_2, v_2.

Under H_{11}^j, $j = 1, 2$, one has $\mathbb{P}(\breve{w}_j = 0) \to 1$ and hence $\mathbb{P}(\breve{z}_j = 0) \to 1$; under H_{12}^j one has $\mathbb{P}(\breve{v}_j = 0) \geq \rho_{2T} \to 1 - \eta_2$ and hence

$$\mathbb{P}(\breve{z}_j = 0) \geq \rho_{2T} \to 1 - \eta_2. \tag{6.7}$$

A few comments are in order.

- The criteria based on $\breve{s} = \widehat{s}, \widetilde{s}$ are consistent, in the sense that they provide the correct choice asymptotically with probability 1, both when the null H_0^j, $j = 1, 2$ is valid and when it is not.
- The tests based on $\check{s}$ have different asymptotic properties. Eq. (6.5), (6.6), show how to control the size of the test asymptotically. In the case of (6.5) one can choose $\eta_1 = \eta_2 = \eta$ for some pre-specified significance level η, such as 0.05. For the case of (6.6) one can choose $\eta_1 = \eta_2 = \eta/2$ for some pre-specified η.
- Eq. (6.7) gives conservative bounds for the power to reject H_0^j, $j = 1, 2$; these bounds are equal to $1 - \eta_2$, where η_2 is the size of the tests in $\breve{v}_j$. In the special case when H_{11}^j holds, then the power to reject H_0^j, $j = 1, 2$, is asymptotically equal to 1.

7. Simulations

This section reports results from Monte Carlo (MC) simulations. The DGP is taken from Onatski and Wang (2018) and Bykhovskaya and Gorin (2022). The data is generated from

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d.} N(0, I_p), \qquad t = 1, \dots, T, \qquad \Delta X_0 = X_0 = 0, \tag{7.1}$$

Size, $s=1$	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
$H_0^1(m)$ in (7.3)	z	w	v	z	w	v	z	w	v	z	w	v
$m=1$	0	0	0	0	0	0	0.0462	0.0462	0	0.0462	0.0462	0
2	0	0	0	0	0	0	0.0459	0.0459	0	0.0459	0.0459	0
3	0	0	0	0	0	0	0.0459	0.0459	0	0.0459	0.0459	0
4	0	0	0	0	0	0	0.0456	0.0456	0	0.0456	0.0456	0
5	0	0	0	0	0	0	0.0452	0.0452	0	0.0452	0.0452	0
6	0	0	0	0	0	0	0.0451	0.0451	0	0.0451	0.0451	0
7	0	0	0	0	0	0	0.0447	0.0447	0	0.0447	0.0447	0
8	0	0	0	0	0	0	0.0446	0.0446	0	0.0446	0.0446	0
9	0	0	0	0	0	0	0.0441	0.0441	0	0.0441	0.0441	0
10	0	0	0	0	0	0	0.044	0.044	0	0.044	0.044	0
11	0	0	0	0	0	0	0.0438	0.0438	0	0.0438	0.0438	0
12	0	0	0	0	0	0	0.0437	0.0437	0	0.0437	0.0437	0
13	0	0	0	0	0	0	0.0435	0.0435	0	0.0435	0.0435	0
14	0	0	0	0	0	0	0.0433	0.0433	0	0.0433	0.0433	0
15	0	0	0	0	0	0	0.0428	0.0428	0	0.0428	0.0428	0
16	0	0	0	0	0	0	0.0425	0.0425	0	0.0425	0.0425	0
17	0	0	0	0	0	0	0.0425	0.0425	0	0.0425	0.0425	0
18	0	0	0	0	0	0	0.0419	0.0419	0	0.0419	0.0419	0
19	0	0	0	0	0	0	0.0418	0.0418	0	0.0418	0.0418	0
$H_0^2(q)$ in (7.4)	z	w	v	z	w	v	z	w	v	z	w	v
$q=1$	0	0	0	0	0	0	0.0462	0.0462	0	0.0462	0.0462	0
Power, $s=1$	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
$H_0^1(i,0)$ in (7.5)	z	w	v	z	w	v	z	w	v	z	w	v
$i=2$	1	1	1	1	1	1	1	1	0.9579	1	1	0.9579
$H_0^1(i,1)$ in (7.5)	z	w	v	z	w	v	z	w	v	z	w	v
$i=2$	1	0	1	1	0	1	0.994	0.2678	0.7458	0.994	0.2678	0.7458

TABLE 1. Empirical size and power, $s=1$: rejection frequency of (7.3), (7.4) and (7.5) ((7.6) is empty when $s=1$). The sequential tests are conducted at 5% significance level and the number of Monte Carlo replications is set to $N=10^4$.

with $\beta=(0,I_{p-s})'$ and $\alpha=-\beta$, so that

$$\psi=(I_s,0)'=(e_1,\dots,e_s).$$

The values of p,T,K, and s considered in the MC design are

$$p=20,\qquad T=500,\qquad K=\lceil T^{3/4}\rceil=106,\qquad s=1,10,19, \tag{7.2}$$

thus covering a medium ($p=20$) dimensional system with low to high number of stochastic trends ($s=1,p/2,p-1$). The number of Monte Carlo replications is set to $N=10^4$.

First consider the hypotheses

$$H_0^1(m):\qquad \mathrm{col}(\psi)\subseteq\mathrm{col}(e_1,\dots,e_m),\qquad m=s,\dots,p-1, \tag{7.3}$$

$$H_0^2(q):\qquad \mathrm{col}(e_1,\dots,e_q)\subseteq\mathrm{col}(\psi),\qquad q=1,\dots,s, \tag{7.4}$$

and observe that (7.3) and (7.4) hold in the DGP; these designs allow one to investigate properties of the inference procedures under the null.

Next consider the following other hypotheses with $k = 0, 1$

$$H_0^1(i,k): \quad \mathrm{col}(\psi) \subseteq \mathrm{col}(A), \quad A = (e_i, \ldots, e_{s+1}, e_{s+2} + ke_1, \ldots, e_{s+i} + ke_{i-1}), \qquad (7.5)$$
$$i = 2, \ldots, \min(s, p-s),$$
$$H_0^2(h,k): \quad \mathrm{col}(a) \subseteq \mathrm{col}(\psi), \qquad a = (e_{h+2}, \ldots, e_s, e_{s+1} + ke_1, \ldots e_{h+s} + ke_h) \qquad (7.6)$$
$$h = 1, \ldots, \min(s-2, p-s),$$

which do not hold under the DGP; here the number of columns in A is $m = s + 1$ and the one in a is $q = s - 1$.

The designs (7.5) and (7.6) permit one to investigate the performance of the inference procedures under the alternative. In fact, note that under $H_0^1(i, 0)$ one has $\mathrm{rank}(A'\psi) = s - i + 1 < s$ so that H_{11}^1 in (4.9) holds, while under $H_0^1(i, 1)$ one finds $\mathrm{rank}(A'\psi) = s$, and hence H_{01}^1 in (4.7) holds. In both $k = 0, 1$ cases, one finds $\mathrm{rank}(A'_\perp\psi) = i - 1 > 0$, hence H_{12}^1 in (4.9) holds.

Similarly, under $H_0^2(j, 0)$ one has $\mathrm{rank}(a'\psi) = s - j - 1 < q$ and H_{11}^2 in (4.9) holds, while under $H_0^2(j, 1)$ one finds $\mathrm{rank}(a'\psi) = s - 1 = q$ and hence H_{01}^2 in (4.7) holds. In both $k = 0, 1$ cases, one finds $\mathrm{rank}(a'_\perp\psi) = j + 1 > s - q = 1$, hence H_{12}^1 in (4.9) holds.

This shows that with $k = 0$ one has that H_{11}^j and H_{12}^j hold, $j = 1, 2$; on the contrary for $k = 1$ one has that H_{01}^j and H_{12}^j hold, $j = 1, 2$.

Tables 1, 2, 3 and 4 report rejection frequencies of (7.3) and (7.4) (i.e. rejection frequencies under the null) and of (7.5) and (7.6) (i.e. rejection frequencies under the alternative) when $s = 1, 10, 19$. The results can summarized as follows:

- The criteria based on $\breve{s} = \widehat{s}, \widetilde{s}$ are consistent, in the sense that they provide the correct choice with probability 1, both under the null and under the alternative.
- The tests based on $\breve{s}$ of H_0^j, $j = 1, 2$, have empirical size bounded by the nominal level $\eta_1 + \eta_2 = 10\%$, see (6.6), and power equal to one under H_{11}^j and in most cases bounded below by $1 - \eta_2 = 0.95$ under H_{01}^j, see (6.7).
- The results are similar for the different values of $s = 1, 10, 19$, and well agree with the prediction of Theorem 6.4.

Size, $s=10$	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
$H_0^1(m)$ in (7.3)	z	w	v	z	w	v	z	w	v	z	w	v
$m=10$	0	0	0	0	0	0	0.0407	0.0407	0	0.024	0.024	0
11	0	0	0	0	0	0	0.0397	0.0397	0	0.0222	0.0222	0
12	0	0	0	0	0	0	0.039	0.039	0	0.0209	0.0209	0
13	0	0	0	0	0	0	0.0382	0.0382	0	0.0195	0.0195	0
14	0	0	0	0	0	0	0.0367	0.0367	0	0.019	0.019	0
15	0	0	0	0	0	0	0.0352	0.0352	0	0.0182	0.0182	0
16	0	0	0	0	0	0	0.0344	0.0344	0	0.0171	0.0171	0
17	0	0	0	0	0	0	0.0335	0.0335	0	0.016	0.016	0
18	0	0	0	0	0	0	0.0323	0.0323	0	0.0154	0.0154	0
19	0	0	0	0	0	0	0.0312	0.0312	0	0.0144	0.0144	0
$H_0^2(q)$ in (7.4)	z	w	v	z	w	v	z	w	v	z	w	v
$q=1$	0	0	0	0	0	0	0.07	0.0399	0.0315	0.0579	0.0399	0.019
2	0	0	0	0	0	0	0.0737	0.0439	0.0305	0.061	0.0426	0.0188
3	0	0	0	0	0	0	0.0714	0.0408	0.0321	0.0574	0.0364	0.0216
4	0	0	0	0	0	0	0.0779	0.0435	0.0353	0.0644	0.0402	0.0251
5	0	0	0	0	0	0	0.077	0.0429	0.0351	0.0616	0.0356	0.0265
6	0	0	0	0	0	0	0.0774	0.0421	0.0369	0.065	0.0318	0.0342
7	0	0	0	0	0	0	0.0739	0.039	0.0362	0.0656	0.0324	0.0341
8	0	0	0	0	0	0	0.0784	0.0395	0.0404	0.0681	0.0304	0.0386
9	0	0	0	0	0	0	0.0858	0.0444	0.0438	0.075	0.0322	0.0438
10	0	0	0	0	0	0	0.0407	0.0407	0	0.024	0.024	0

TABLE 2. Empirical size, $s=10$: rejection frequency of (7.3), (7.4), (7.5), and (7.6). The sequential tests are conducted at 5% significance level and the number of Monte Carlo replications is set to $N=10^4$.

Power, $s = 10$	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_{\infty}$			$\breve{s}_1$		
$H_0^1(i,0)$ in (7.5)	z	w	v	z	w	v	z	w	v	z	w	v
$i = 2$	1	1	1	1	1	1	1	1	0.9623	1	1	0.9623
3	1	1	1	1	1	1	1	1	0.9998	1	1	0.999
4	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1
$H_0^1(i,1)$ in (7.5)	z	w	v	z	w	v	z	w	v	z	w	v
$i = 2$	1	0	1	1	0	1	0.8313	0.2554	0.7451	0.8088	0.1853	0.7451
3	1	0	1	1	0	1	0.9868	0.4761	0.9653	0.9852	0.499	0.9573
4	1	0	1	1	0	1	0.9992	0.6585	0.9969	0.9993	0.7822	0.9935
5	1	0	1	1	0	1	1	0.8007	0.9999	1	0.9344	0.9992
6	1	0	1	1	0	1	1	0.8858	1	1	0.9851	0.9999
7	1	0	1	1	0	1	1	0.9396	1	1	0.9981	1
8	1	0	1	1	0	1	1	0.9692	1	1	0.9997	1
9	1	0	1	1	0	1	1	0.9856	1	1	0.9999	1
10	1	0	1	1	0	1	1	0.9936	1	1	1	1
$H_0^2(h,0)$ in (7.6)	z	w	v	z	w	v	z	w	v	z	w	v
$h = 1$	1	1	1	1	1	1	1	1	0.9598	1	1	0.962
2	1	1	1	1	1	1	1	1	0.9996	1	1	0.9994
3	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1
$H_0^2(h,1)$ in (7.6)	z	w	v	z	w	v	z	w	v	z	w	v
$h = 1$	1	0	1	1	0	1	0.8282	0.2606	0.7365	0.8129	0.2145	0.7361
2	1	0	1	1	0	1	0.9853	0.4856	0.9584	0.9869	0.5363	0.9554
3	1	0	1	1	0	1	0.9993	0.6677	0.9942	0.9995	0.8076	0.9921
4	1	0	1	1	0	1	1	0.8077	0.9997	1	0.9441	0.999
5	1	0	1	1	0	1	1	0.8885	1	1	0.9872	1
6	1	0	1	1	0	1	1	0.9419	1	1	0.9975	1
7	1	0	1	1	0	1	1	0.9732	1	1	0.9997	1
8	1	0	1	1	0	1	1	0.9861	1	1	0.9999	1

TABLE 3. Empirical power, $s = 10$: rejection frequency of (7.3), (7.4), (7.5), and (7.6). The sequential tests are conducted at 5% significance level and the number of Monte Carlo replications is set to $N = 10^4$.

Size, $s = 19$	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
$H_0^1(m)$ in (7.3)	z	w	v	z	w	v	z	w	v	z	w	v
$m = 19$	0	0	0	0	0	0	0.0402	0.0402	0	0.01	0.01	0
$H_0^2(q)$ in (7.4)	z	w	v	z	w	v	z	w	v	z	w	v
$q = 1$	0	0	0	0	0	0	0.0771	0.0433	0.0356	0.0539	0.0433	0.0108
2	0	0	0	0	0	0	0.0784	0.043	0.0376	0.0532	0.0435	0.01
3	0	0	0	0	0	0	0.0788	0.0427	0.0377	0.054	0.0411	0.0134
4	0	0	0	0	0	0	0.0792	0.0419	0.0386	0.0531	0.0391	0.0144
5	0	0	0	0	0	0	0.0792	0.0416	0.0391	0.0521	0.0348	0.0178
6	0	0	0	0	0	0	0.0803	0.046	0.0359	0.0542	0.0338	0.0211
7	0	0	0	0	0	0	0.0783	0.0432	0.0359	0.0555	0.0347	0.0214
8	0	0	0	0	0	0	0.08	0.043	0.0382	0.057	0.0338	0.0243
9	0	0	0	0	0	0	0.086	0.044	0.0438	0.0577	0.0334	0.0253
10	0	0	0	0	0	0	0.079	0.0422	0.0385	0.0571	0.0284	0.0294
11	0	0	0	0	0	0	0.0823	0.0431	0.0414	0.0572	0.027	0.0309
12	0	0	0	0	0	0	0.0883	0.0439	0.0462	0.0552	0.0224	0.0335
13	0	0	0	0	0	0	0.0873	0.0447	0.0439	0.0554	0.0223	0.0335
14	0	0	0	0	0	0	0.0888	0.044	0.0468	0.0554	0.0187	0.0374
15	0	0	0	0	0	0	0.0828	0.0404	0.0434	0.0564	0.0169	0.0403
16	0	0	0	0	0	0	0.0827	0.0422	0.0423	0.0546	0.0146	0.041
17	0	0	0	0	0	0	0.0865	0.0411	0.0471	0.059	0.0129	0.0464
18	0	0	0	0	0	0	0.0787	0.036	0.044	0.055	0.0115	0.044
19	0	0	0	0	0	0	0.0402	0.0402	0	0.01	0.01	0
Power, $s = 19$	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
$H_0^2(h, 0)$ in (7.6)	z	w	v	z	w	v	z	w	v	z	w	v
$h = 1$	1	1	1	1	1	1	1	1	0.9572	1	1	0.9574
$H_0^2(h, 1)$ in (7.6)	z	w	v	z	w	v	z	w	v	z	w	v
$h = 1$	1	0	1	1	0	1	0.7865	0.1959	0.7201	0.7544	0.0933	0.7226

TABLE 4. Empirical size and power, $s = 19$: rejection frequency of (7.3), (7.4) and (7.6) ((7.5) is empty when $s = 19$). The sequential tests are conducted at 5% significance level and the number of Monte Carlo replications is set to $N = 10^4$.

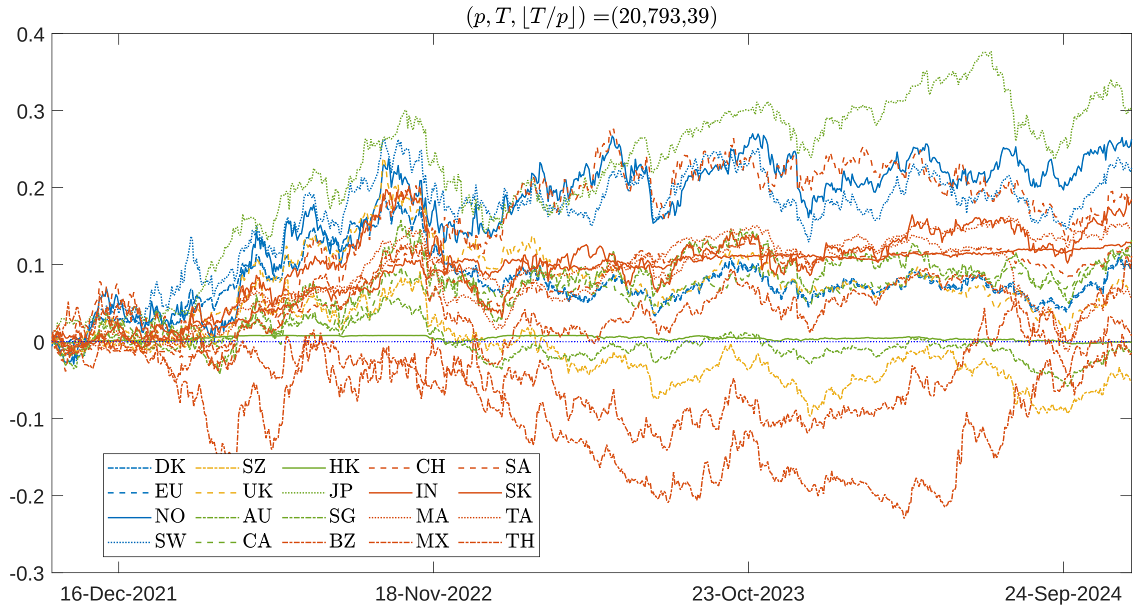

FIGURE 2. 20 World Markets currencies with color codes: blue for Denmark (DK), Europe (EU), Norway (NO), Sweden (SW), yellow for Switzerland (SZ), United Kindom (UK), green Australia (AU), Canada (CA), Hong Kong (HK), Japan (JP), Singapore (SG), red Brazil (BZ), China (CH), India (IN), Malaysia (MA), Mexico (MX), South Africa (SA), South Korea (SK), Taiwan (TA), Thailand (TH). Line types (repeated in alphabetical order): dot-dash, dashes, solid, dots.

8. Empirical application

This section provides an illustration on a panel of daily exchange rates between Oct 5, 2021 and Dec 06, 2024 of the US dollar against 20 World Markets currencies, downloaded from the Federal Reserve Economic Data (FRED) website, https://fred.stlouisfed.org/. See Onatski and Wang (2019) for a similar dataset and a review of the literature on the topic. Data (in logs and normalized to start at 0) are plotted in Figure 2. The sample size is $T = 793$, with a ratio $\lfloor T/p \rfloor = 39$.

The panels in Figure 3 below report results for the $p = 20$ exchange rates with $K = \lceil T^{3/4} \rceil = 150$. The first panel shows the eigenvalues of $\texttt{cca}(x_t, d_t)$ where the first 19 eigenvalues are above 0.75 while λ_{20} drops down to 0.27, so that the largest gap $\lambda_i - \lambda_{i+1}$ is found between λ_{19} and λ_{20}, for $i = 19$. This implies $\widehat{s} = 19$ and indicates the presence of $s = 19$ stochastic trends, i.e. $r = 1$ cointegrating relations, in the $p = 20$ dimensional system.

The second panel reports the test statistic $K\pi^2\tau^{(\widehat{s})}$, see (6.3), and the corresponding 95% asymptotic confidence stripe $\mathcal{B}$, see FGP; given that $K\pi^2\tau^{(\widehat{s})} \in \mathcal{B}$, the model assumptions – including

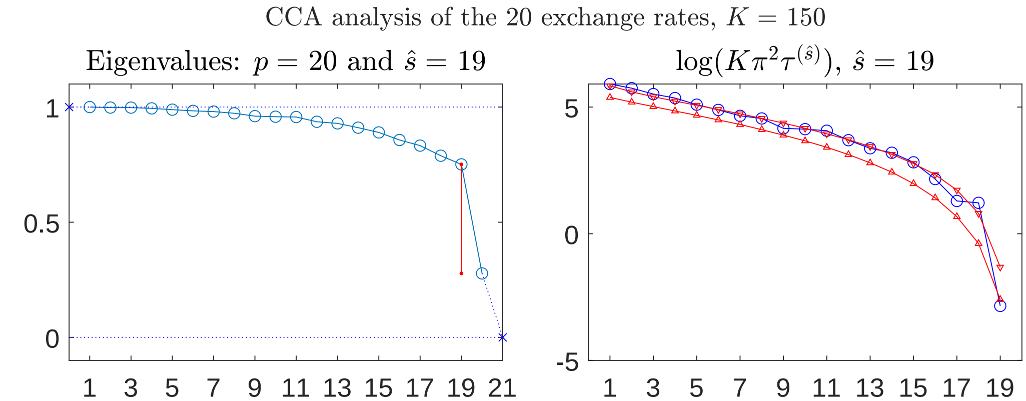

FIGURE 3. Analysis of the 20 World Markets currencies: eigenvalues, max-gap estimator and $\log(K\pi^2\tau^{(\hat{s})})$ for $K = \lceil T^{3/4} \rceil = 150$ and with a 95% stripe around $\mathbb{E}(\log \zeta^{(\hat{s})})$.

$p = 20$	$\widehat{s}$	$\widetilde{s}$	$\check{s}_\infty$	$\check{s}_1$
estimate of s	19	19	18	17
estimate of r	1	1	2	3

TABLE 5. Estimates of s with the argmax estimators and with the sequential tests at 5% significance level.

the selection of s – appear valid, hence supporting the choice $s = 19$. Table 5 reports the results for estimators $\widehat{s}$, $\widetilde{s}$ and the sequential testing procedures based on $\check{s}_\infty$ and $\check{s}_1$; the different criteria tend to agree, with the exception of the testing criteria.

In order to investigate the structure of the cointegrating space and of the attractor space, consider the hypotheses in (4.1) and (4.2),

$$H_0^1 : \operatorname{col} b \subseteq \operatorname{col} \beta, \qquad H_0^2 : \operatorname{col} \beta \subseteq \operatorname{col} B,$$

for

$$b := a_\perp = e_j - e_{EU}, \qquad B := A_\perp = (e_j - e_{EU})_{j \in \mathcal{J}}$$

where j indicates a given country in the list $\mathcal{J} = \{$AU, BZ, CA, CH, DK, HK, IN, JP, MA, MX, NO, SA, SG, SK, SW, SZ, TA, TH, UK$\}$. H_0^1 amounts to testing if there is a currency differential with the Euro that is stationary while H_0^2 checks if a cointegrating relation can be expressed as a linear combination of currency differentials. Results are reported in Table 6.

The top part of Table 6 reports the results for $H_0^2 : \operatorname{col} \beta \subseteq \operatorname{col} B$, which tests if cointegrating relations can be expressed as a linear combination of currency differentials. Given that both statistics w and v are equal to 1, the statistics z is also equal to 1 and hence the hypothesis is not rejected.

H_0^2	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
	z	w	v	z	w	v	z	w	v	z	w	v
	1	1	1	1	1	1	1	1	1	1	1	1
H_0^1	$\widehat{s}$			$\widetilde{s}$			$\breve{s}_\infty$			$\breve{s}_1$		
j	z	w	v	z	w	v	z	w	v	z	w	v
AU	0	0	1	0	0	1	0	0	0	0	0	1
BZ	0	0	1	0	0	1	0	1	0	1	1	1
CA	0	0	1	0	0	1	0	0	0	0	0	1
CH	0	0	1	0	0	1	0	0	0	0	0	1
DK	0	0	1	0	0	1	0	0	0	0	0	1
HK	0	0	1	0	0	1	0	1	0	1	1	1
IN	0	0	1	0	0	1	0	0	0	0	0	0
JP	0	0	1	0	0	1	0	0	0	0	0	1
MA	0	0	1	0	0	1	0	0	0	0	0	1
MX	0	0	1	0	0	1	0	0	0	0	0	1
NO	0	0	1	0	0	1	0	0	0	0	0	1
SA	0	0	1	0	0	1	0	0	0	0	0	1
SG	0	0	1	0	0	1	0	0	0	0	0	1
SK	0	0	1	0	0	1	0	0	0	0	0	1
SW	0	0	1	0	0	1	0	0	0	0	0	1
SZ	0	0	1	0	0	1	0	0	0	0	0	1
TA	0	0	1	0	0	1	0	0	0	0	0	1
TH	0	0	1	0	0	1	0	0	0	0	0	1
UK	0	0	1	0	0	1	0	0	0	0	0	1

TABLE 6. Results on $H_0^1 : B = e_j - e_{EU}$ and $H_0^2 : b = (e_j - e_{EU})_{j \in \mathcal{J}}$; the sequential tests have 5% significance level.

One can further tests if this is due to the presence of a currency differential that is stationary, i.e. $H_0^1 : \operatorname{col} b \subseteq \operatorname{col} \beta$ for each country, the second part of Table 6 shows that this hypothesis is rejected for any country using any of the four criteria (with two exceptions for $\breve{s}_1$). Hence one concludes against the presence of a stationary currency differential.

9. Conclusions

This paper proposes a unified semiparametric framework for inference the cointegrating space and attractor space for $I(1)/I(0)$ processes; this includes VARIMA and Dynamic Factor Model processes. The approach employs implications of the nesting of subspaces and the property of dimensional coherence of the semiparametric approach in FGP for subsets of variables.

The paper reports some novel results on a limit distribution of interest in the special one-dimensional case. The properties of the estimators and tests are compared with alternatives, both theoretically and with a Monte Carlo experiment. They are illustrated via an empirical application to exchange rates.

References

Abadir, K. M. and P. Paruolo (1997). Two mixed normal densities from cointegration analysis. *Econometrica 65*(3), 671–680.

Ahn, S. and A. Horenstein (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica 81*, 1203–1227.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics 122*, 137–183.

Bierens, H. (1997). Nonparametric cointegration analysis. *Journal of Econometrics 77*, 379–404.

Bykhovskaya, A. and V. Gorin (2022). Cointegration in large VARs. *The Annals of Statistics 50*, 1593 – 1617.

Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978, 12). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the united kingdom. *The Economic Journal 88*(352), 661–692.

Doornik, J. A., B. Nielsen, and T. J. Rothenberg (2003). The influence of var dimensions on estimator biases: Comment. *Econometrica 71*(1), 377–383.

Elliott, G. (1999). Efficient Tests for a Unit Root When the Initial Observation is Drawn from Its Unconditional Distribution. *International Economic Review 40*, 767–783.

Engle, R. and C. Granger (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica 55*, 251–276.

Franchi, M., I. Georgiev, and P. Paruolo (2024). Canonical correlation analysis of stochastic trends via functional approximation. Working Paper arXiv: 2411.19572.

Jacod, J. and A. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. Springer.

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control 12*, 231–254.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica 59*, 1551–1580.

Johansen, S. (1996). *Likelihood-based Inference in Cointegrated Vector Auto-Regressive Models*. Oxford University Press.

Johansen, S. and K. Juselius (1992). Testing structural hypotheses in a multivariate cointegration analysis of the ppp and the uip for uk. *Journal of Econometrics 53*(1), 211–244.

Johansen, S. and K. Juselius (2014). An asymptotic invariance property of the common trends under linear transformations of the data. *Journal of Econometrics 178*, 310–315.

Müller, U. and M. Watson (2008). Testing Models of Low-Frequency Variability. *Econometrica 76*, 979–1016.

Onatski, A. and C. Wang (2018). Alternative Asymptotics for Cointegration Tests in Large VARs. *Econometrica 86*, 1465–1478.

Onatski, A. and C. Wang (2019). Extreme canonical correlations and high-dimensional cointegration analysis. *Journal of Econometrics 212*, 307–322.

Phillips, P. C. B. (1998). New Tools for Understanding Spurious Regressions. *Econometrica 66*, 1299–1326.

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

Stock, J. and M. Watson (1988). Testing for common trends. *Journal of the American Statistical Society 83*, 1097–1107.

Appendix A. Proofs

Proof of Theorem 6.3 Let $S = \int_0^1 B(u)^2 \mathrm{d}u$. By (10) in Abadir and Paruolo (1997), one has

$$f_S(s) = \frac{s^{-\frac{3}{2}}}{\sqrt{\pi}} \sum_{j=0}^{\infty} \eta_j a_j \exp\left(-\frac{a_j^2}{2s}\right)$$

where $\eta_j := \binom{-\frac{1}{2}}{j}$, $a_j := 2j + \frac{1}{2}$. Next define $\zeta := 1/S = g(S)$; by the transformation theorem $S = g^{-1}(\zeta) = 1/\zeta$, $|\mathrm{d}g^{-1}(\zeta)/\mathrm{d}\zeta| = \zeta^{-2}$, and hence, for $z > 0$

$$f_\zeta(z) = f_S(h^{-1}(z))z^{-2} = z^{-\frac{1}{2}} \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \eta_j a_j \exp\left(-\frac{a_j^2}{2} z\right).$$

Note also that

$$\begin{aligned} F_\zeta(t) &= 1 - \int_t^\infty f_\zeta(z)\mathrm{d}z = 1 - \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \eta_j a_j \int_t^\infty z^{-\frac{1}{2}} \exp\left(-\frac{a_j^2}{2} z\right) \mathrm{d}z \\ &= 1 - \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \eta_j a_j \Gamma\left(\frac{1}{2}, \frac{a_j^2}{2} t\right) \left(\frac{a_j^2}{2}\right)^{-\frac{1}{2}} = 1 - \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \sum_{j=0}^{\infty} \eta_j \Gamma\left(\frac{1}{2}, \frac{a_j^2}{2} t\right), \end{aligned}$$

where the following identity has been used, setting $y = az = h(z)$, $z = h^{-1}(y) = y/a$, $\mathrm{d}z = a^{-1}\mathrm{d}y$:

$$\int_t^\infty z^{k-1} e^{-az} \mathrm{d}z = \int_{at}^\infty \left(\frac{y}{a}\right)^{k-1} e^{-y} a^{-1} \mathrm{d}y = a^{-k} \int_{at}^\infty y^{k-1} e^{-y} \mathrm{d}y = a^{-k} \Gamma(k, at).$$

Here $\Gamma(k, t) = \int_t^\infty z^{k-1} e^{-z} \mathrm{d}z$ is the upper incomplete Gamma function. ∎

Proof of Theorem 6.4. (i) $\breve{s} = \widehat{s}, \widetilde{s}$. Assume H_0^j holds $j = 1, 2$, and define

$$\mathcal{B}_1 := \{\breve{s}(\bar{A}'x_t) = s\}, \quad \mathcal{C}_1 := \{\breve{s}(A'_\perp x_t) = 0\}, \tag{A.1}$$
$$\mathcal{B}_2 := \{\breve{s}(\bar{a}'x_t) = q\}, \quad \mathcal{C}_2 := \{\breve{s}(a'_\perp x_t) = s - q\},$$

and note that $\mathbb{P}(\mathcal{B}_1) \to 1$ and $\mathbb{P}(\mathcal{C}_1) \to 1$ under H_0^1, $\mathbb{P}(\mathcal{B}_2) \to 1$ and $\mathbb{P}(\mathcal{C}_2) \to 1$ under H_0^2 by Theorem 6.1; this proves (6.4).

Under H_{11}^1 one has $\text{rank}(A'\psi) = s - h$, $h > 0$, and $\mathbb{P}(\breve{s}(\bar{A}'x_t) = s - h) \to 1$, which implies $\mathbb{P}(\breve{s}(\bar{A}'x_t) = s) \to 0$; hence $\mathbb{P}(\breve{w}_1 = 0) \to 1$. Under H_{12}^1 one has $\text{rank}(A'_\perp\psi) = k$, $k > 0$, and $\mathbb{P}(\breve{s}(A'_\perp x_t) = k) \to 1$, which implies $\mathbb{P}(\breve{s}(\bar{A}'_\perp x_t) = 0) \to 0$; hence $\mathbb{P}(\breve{v}_1 = 0) \to 1$. Hence under H_1^1, one finds $\mathbb{P}(\breve{z}_1 = 0) \to 1$.

Similarly under H_{11}^2 one has $\text{rank}(\bar{a}'\psi) = q - h$, $h > 0$, $\mathbb{P}(\breve{s}(\bar{a}'x_t) = q - h) \to 1$, and hence $\mathbb{P}(\breve{s}(\bar{a}'x_t) = q) \to 0$, i.e. $\mathbb{P}(\breve{w}_2 = 0) \to 1$. Under H_{12}^2 one has $\text{rank}(a'_\perp\psi) = s - q + k$, $k > 0$, $\mathbb{P}(\breve{s}(a'_\perp x_t) = s - q + k) \to 1$, which implies $\mathbb{P}(\breve{s}(a'_\perp x_t) = s - q) \to 0$; hence $\mathbb{P}(\breve{v}_2 = 0) \to 1$. Thus under H_1^2, one finds $\mathbb{P}(\breve{z}_2 = 0) \to 1$.

(ii) $\breve{s} = \check{s}$. Under H_0^1 one has $\mathbb{P}(\mathcal{C}_1) \to 1$ thanks to Theorem 6.2.(iv) case $s = 0$ because the number of CT under the null is 0. This implies $\mathbb{P}(\breve{v}_1 = 1) \to 1$. Moreover $\mathbb{P}(\mathcal{B}_1) \to 1 - \eta_1$ thanks to Theorem 6.2.(iv) case $s > 0$. This implies $\mathbb{P}(\breve{w}_1 = 1) \to 1 - \eta_1$. Finally, for $j = 1$

$$\mathbb{P}(\mathcal{B}_j \cap \mathcal{C}_j) = 1 - \mathbb{P}(\mathcal{B}_j^\complement \cap \mathcal{C}_j^\complement) \geq 1 - \mathbb{P}(\mathcal{B}_j^\complement) - \mathbb{P}(\mathcal{C}_j^\complement) =: \rho_{1T} \tag{A.2}$$

where, from above, $\rho_{1T} \to 1 - \eta_1$. This proves $\mathbb{P}(\breve{z}_1 = 1) \geq \rho_{1T} \to 1 - \eta_1$ and completes the proof of (6.5).

Under H_0^2 with $s > q$ one has $\mathbb{P}(\mathcal{B}_2) \to 1 - \eta_1$ and $\mathbb{P}(\mathcal{C}_2) \to 1 - \eta_2$ thanks to Theorem 6.2.(iv) case $s > 0$. This implies $\mathbb{P}(\breve{w}_2 = 1) \to 1 - \eta_1$ and $\mathbb{P}(\breve{v}_2 = 1) \to 1 - \eta_2$. Finally using (A.2) for $j = 2$ one finds $\mathbb{P}(\breve{z}_2 = 1) \geq \rho_{1T} \to 1 - \eta_1 - \eta_2$. This proves (6.6). Under H_0^2 with $s = q$ one has $\mathbb{P}(\mathcal{C}_2) \to 1$ thanks to Theorem 6.2.(iv) case $s = 0$.

Under H_{11}^1 one has $\text{rank}(A'\psi) = s - h$, $h > 0$, and $\mathbb{P}(\breve{s}(A'\psi) = s) \to 0$, see Theorem 6.2(v); this implies $\mathbb{P}(\breve{w}_1 = 0) \to 1$. Under H_{12}^1 one has $\text{rank}(A'_\perp\psi) = k$, $k > 0$, and $\mathbb{P}(\breve{s}(A'_\perp x_t) = 0) \leq \nu_1 \to \eta_2$, which implies $\mathbb{P}(\breve{v}_1 = 0) \geq \rho_{2T} \to 1 - \eta_2$. Hence under H_1^1, one finds $\mathbb{P}(\breve{z}_1 = 0) \geq \min(\mathbb{P}(\breve{v}_1 = 0), \mathbb{P}(\breve{w}_1 = 0))$ which is asymptotically equal to $\rho_{2T} \to 1 - \eta_2$.

Under H_{11}^2 one has $\text{rank}(\bar{a}'\psi) = q - h$, $h > 0$, and $\mathbb{P}(\breve{s}(\bar{a}'\psi) = q) \to 0$, see Theorem 6.2(v); this implies $\mathbb{P}(\breve{w}_2 = 0) \to 1$. Under H_{12}^2 one has $\text{rank}(a'_\perp\psi) = s - q + k$, $k > 0$, $\mathbb{P}(\breve{s}(a'_\perp x_t) = s - q + k) \to 1 - \eta_2$, which implies $\mathbb{P}(\breve{s}(a'_\perp x_t) = s - q) \leq \nu_2 \to \eta_2$; hence $\mathbb{P}(\breve{v}_2 = 0) \geq \rho_{2T} \to 1 - \eta_2$. Thus under H_1^2, one finds $\mathbb{P}(\breve{z}_2 = 0) \geq \min(\mathbb{P}(\breve{v}_2 = 0), \mathbb{P}(\breve{w}_2 = 0))$ which is asymptotically equal to $\rho_{2T} \to 1 - \eta_2$.

This completes the proof. ■