

ORIGINAL ARTICLE OPEN ACCESS

Inference on the Attractor Space via Functional Approximation

Massimo Franchi¹  | Paolo Paruolo² 

¹Sapienza University of Rome, Rome, Italy | ²European Commission, Joint Research Centre (JRC), Ispra, Varese, VA, Italy

Correspondence: Paolo Paruolo (paolo.paruolo@ec.europa.eu)

Received: 26 January 2026 | **Accepted:** 28 January 2026

Keywords: attractor space | canonical correlation analysis | cointegration space | $I(1)$ | semiparametric inference | unit roots

ABSTRACT

This paper discusses semiparametric inference on hypotheses on the cointegration and the attractor spaces for $I(1)$ linear processes with moderately large cross-sectional dimension. The approach is based on sample canonical correlations and functional approximation of Brownian motions, and it can be applied both to the whole system and or to any set of linear combinations of it. The hypotheses of interest are cast in terms of the number of stochastic trends in specified subsystems, and inference is based either on selection criteria or on sequences of tests. This paper derives the limit distribution of these tests in the special one-dimensional case, and discusses asymptotic properties of the derived inference criteria for hypotheses on the attractor space for sequentially diverging sample size and number of basis elements in the functional approximation. Finite sample properties are analysed via a Monte Carlo study and an empirical illustration on exchange rates is provided.

JEL Classification: C32, C33

1 | Introduction

The notion of cointegration introduced in Engle and Granger [1] and its link to the existence of an equilibrium correction mechanism (ECM), see Davidson et al. [2], have greatly influenced the econometric analysis of multiple time series. The introduction in Johansen [3, 4] of the quasi maximum likelihood estimator (QMLE) of the ECM within a vector autoregressive (VAR) model for variables integrated of order 1 ($I(1)$) provided a new set of inferential tools that shaped the understanding of the vector-space nature of the cointegrating space.

Hypotheses on the cointegrating space were introduced together with the QMLE in the $I(1)$ VAR case. Specifically, Johansen [4, equation (3.1)] considered the hypothesis that the cointegrating space is contained in a pre-specified linear subspace, and also

considered similar hypotheses on the matrix of adjustment coefficients. Soon after, Johansen and Juselius [5, equation (15)] considered the converse hypothesis that the cointegrating space contains a pre-specified linear subspace; see Johansen [6, Chapter 7.2] for a summary.

These hypotheses on the cointegrating space are often used to investigate economic-theory implications after determination of the cointegration rank and prior to the identification of the cointegrating matrix, see Johansen [6], Juselius [7] and Section 2 below for some examples. Inclusion restrictions of the same type for the set of linear combinations defining the innovations to the stochastic trends were also discussed in Gonzalo and Granger [8] and subsequent literature. The present paper focuses on this type of hypotheses on the attractor space, which is the complementary orthogonal space with respect to the cointegrating space.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). Oxford Bulletin of Economics and Statistics published by Oxford University and John Wiley & Sons Ltd.

Recent years have seen the surge in availability of datasets with cross-sectional dimension p in the double digits, coupled with a time series dimension T in the double or triple digits; several contributions have recently addressed this case. One approach focuses on the VAR model with p and T diverging proportionally: Onatski and Wang [9, 10] characterise the distribution of the likelihood ratio test of Johansen [3, 4] for the null of no cointegration, Bykhovskaya and Gorin [11, 12] propose a modified version of the likelihood ratio test statistics for the null of no cointegration, and Liang and Schienle [13] discuss Lasso-type estimators of the number of stochastic trends.

Other recent contributions have been cast in a semiparametric framework that encompasses but it is not restricted to VAR processes, see for instance Bierens [14], Cheng and Phillips [15], Barigozzi and Trapani [16], Barigozzi et al. [17], Franchi et al. [18, 19] and references therein. Related approaches based on a fixed number K of cosine functions are presented in Müller and Watson [20–22], who respectively estimate the number of stochastic trends and the cointegrating relations using an IV approach, and analyze long-run variability and covariability of filtered time series.

The present paper considers inclusion restrictions between the attractor space and a given linear subspace, within the semiparametric context of Franchi et al. [19], henceforth FGP, with T diverging and p fixed. The tools proposed in this paper cover the case of moderate cross section dimension p in the sense of Chen and Schienle [23], and can be used together with the estimators proposed in FGP for the estimation of the cointegration rank and of a basis of the attractor space.

The contributions of this paper include: (i) The definition of the hypotheses of interest as hypotheses on the number of stochastic trends on subsets of variables; (ii) the characterisation of the restricted parameter space under the null and alternative hypotheses; (iii) the derivation of decision rules for these hypotheses both in terms of selection criteria for the number of stochastic trends and in terms of sequences of tests of hypotheses; (iv) the derivation of exact formulae for the probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of the limit distribution for the tests of hypothesis in the univariate case; (v) a Monte Carlo study of the finite sample properties of the proposed procedures and an application to a system of $p = 20$ exchange rates.

Following FGP, this paper considers a p -dimensional $I(1)$ linear process $\Delta X_t = C(L)\varepsilon_t$ with Beveridge Nelson decomposition $C(z) = C + (1 - z)C_1(z)$ with $s := \text{rank } C$, see for example, Phillips and Solo [24]. The integer s is the number of Common Trends (CT) and if $s > 0$, C has representation $C = \psi\kappa'$ for ψ , κ of full column rank s , and X_t satisfies the CT representation

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

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 $0 < s \leq p$ linearly independent random walks (stochastic trends), ψ is a $p \times s$ full column rank loading matrix spanning the attractor space $\text{col } \psi$, and $C_1(L)\varepsilon_t$ is an $I(0)$ linear

process; here $C(L)$ and $C_1(L)$ are infinite matrix polynomials and β is a basis of the cointegrating space $(\text{col } \psi)^\perp$, the orthogonal complement of $\text{col } \psi$. When $s = 0$, X_t does not contain stochastic trends.

This paper considers hypotheses $\text{col } b \subseteq \text{col } \beta \subseteq \text{col } B$ that can be reformulated as $\text{col } a \subseteq \text{col } \psi \subseteq \text{col } A$ in terms of the attractor space, with $a = B_\perp$, $\psi = \beta_\perp$, $A = b_\perp$, and it uses estimators of s introduced in FGP. The present study constructs inference criteria for the validity of restrictions of the type $\text{col } a \subseteq \text{col } \psi$ or $\text{col } \psi \subseteq \text{col } A$, as well as appropriate asymptotic tests of these hypotheses.

The approach is based on the sample 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 discretised over the equispaced grid $1/T, 2/T, \dots, 1$. In the asymptotic analysis, the cross-sectional dimension p is fixed while T and K diverge sequentially, K after T , denoted by $(T, K)_{\text{seq}} \rightarrow \infty$.

The proposed inferential procedures are based on the *dimensional coherence* of the 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 [25] for a discussion of this concept in VARs. Dimensional coherence is used here to address inference on the cointegration and attractor subspaces.

The rest of the paper is organised as follows: The remainder of the introduction presents notation, Section 2 contains motivating examples of hypotheses on the attractor space and Section 3 discusses the assumptions on the data generating process (DGP). Section 4 discusses alternative formulations of the hypotheses of interest, Section 5 presents the proposed inference procedures, and 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.

1.1 | Notation

The following notation is employed in the rest of the paper. For any matrix $a \in \mathbb{R}^{p \times q}$, $\text{col } a$ 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 $\text{col } a$ is $\bar{a}a' = aa'$, 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 $\text{col } a$, that is $\text{col } a_\perp = (\text{col } a)^\perp$. The symbol e_j indicates the j th column of the identity matrix I and ι indicates a column vector of ones.

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 one if c is true and zero otherwise. The space of right continuous functions $f : [0, 1] \rightarrow \mathbb{R}^n$ having finite left limits, endowed with the Skorokhod topology, is denoted by $D_n[0, 1]$, see Jacod and Shiryaev [26, Ch. VI] for details, abbreviated to $D[0, 1]$ if $n = 1$. Similarly $L_n^2[0, 1]$ indicates the space of square integrable functions, abbreviated to $L^2[0, 1]$ if $n = 1$. Riemann

integrals $\int_0^1 a(u)b(u)'du$, are abbreviated as $\int ab'$. Finally the complement of set $\mathcal{A}$ is indicated as $\mathcal{A}^c$.

2 | Motivation

This section provides examples of hypotheses on the attractor space related to aggregation, balanced growth and the autonomy of trends for subsets of variables; these examples motivate interest in hypotheses on the attractor space. Here X_t is assumed to satisfy (1) with s common stochastic trends.

2.1 | Aggregation

Consider the following aggregation problem: Assume X_{it} represents value-added in region i in a given country, where $i \in \mathcal{I} := \{1, \dots, p\}$ and $X_t = (X_{1t}, \dots, X_{pt})'$. Let also X_t be $I(1)$; this means that value-added is stochastically trending, at least in some region; for simplicity, assume first that all X_{it} are $I(1)$.

One is interested in an aggregate or average value-added in the country $w'X_t$, where $w = p^{-1}\iota$. Ideally one would like $w'X_t$ to share the same trend with all the regional components X_{it} ; in this case, in fact, the aggregate $w'X_t$ would have the same information content—in terms of trends—as each X_{it} ; call this property ‘aggregation invariance’.

This requirement imposes restrictions on the attractor space in DGP (1), and provides an example of hypothesis that can be tested with the inference tools developed in this paper. In fact, observe that $w'X_t$ and X_{it} have the same trend whenever $(w - e_i)'X_t$ is not trending, that is when $v_i := w - e_i$ is a cointegrating vector. This implies that the hypothesis of aggregation invariance requires that the number of stochastic trends s in (1) is at least equal to one (in order for X_t to be $I(1)$) and that $\mathcal{V} := \text{col}(v_1, \dots, v_p) \subseteq \text{col } \psi$ with $v_i := w - e_i$.

Consider for instance $p = 3$, and $\mathcal{V} = \text{col}(v_1, v_2, v_3)$ with $v_1 = \frac{1}{3}(-2, 1, 1)'$, $v_2 = \frac{1}{3}(1, -2, 1)'$, $v_3 = \frac{1}{3}(1, 1, -2)'$. It is simple to note that $\text{rank}(v_1, v_2, v_3) = 2$ because $v_1 + v_2 + v_3 = 0$, that is, $\dim \mathcal{V} = p - 1$, and that $\text{col}(v_1, v_2, v_3) = \text{col } b$, where $b = (b_1, b_2)$ can be chosen as $b_1 := \frac{3}{2}(v_2 - v_1) = (1, -1, 0)'$, and $b_2 := \frac{3}{2}(v_3 - v_1) = (1, 0, -1)'$.¹ Note that $(\text{col } b)^\perp = \text{col } \iota$, and this implies that $\text{col } b \subseteq \text{col } \beta$ corresponds to $\text{col } \psi \subseteq \text{col } \iota$. This derivation can be easily written for generic p ; aggregation invariance hence implies $\text{col } \psi \subseteq \text{col } \iota$.

Because $s := \dim(\text{col } \psi) \geq 1$ and $\dim(\text{col } \iota) = 1$, the inclusion $\text{col } \psi \subseteq \text{col } \iota$ needs to hold as an equality, that is $\text{col } \psi = \text{col } \iota$ and $s = 1$, and hence one can choose $\psi = \iota$. This shows that aggregation invariance corresponds to a special type of inclusion restriction of the attractor space, of the form $\psi = \iota$.

Proposition 2.1 (a) below shows that this result extends to a more general situation where one considers any weighted sum $w'X_t$ of elements in X_t , where w is a nonzero vector with $w'\iota$ normalised to one, $w'\iota = 1$. Part (b) discusses the case where the regional component X_{it} is non-trending for i in $\mathcal{I} \setminus \mathcal{I}_1$, and shows that the same result holds restricting the elements of ψ to be zero for the non-trending regions.

Proposition 2.1 (Aggregation invariance). *Let w be a given nonzero vector with $w'\iota = 1$.*

- (a) *The following statements are equivalent:*
 - (i) $w'X_t \sim I(1)$ and $X_{it} \sim I(1)$ share the same common trend for any $i \in \mathcal{I}$;
 - (ii) $\psi = \iota$ with $s = 1$ in (1).
- (b) *The same equivalence between (i) and (ii) holds when substituting $\mathcal{I}$ with any of its nonempty subsets $\mathcal{I}_1$ replacing $\psi = \iota$ with $\psi = \sum_{i \in \mathcal{I}_1} e_i$.*

This example shows that some hypotheses regarding aggregation of trending series translate into hypotheses on $\text{col } \psi$ of the type studied in this paper.

2.2 | Balanced Growth

Another hypothesis of interest concerning the attractor space arises when analyzing the balanced growth of different countries, see King et al. [27]. For simplicity consider the case of two countries, with $(X_{1t}, X_{2t}, X_{3t})'$ (respectively $(X_{4t}, X_{5t}, X_{6t})'$) representing log income, log consumption and log investment in country 1 (respectively country 2). Balanced growth in a given country means that log income, log consumption and log investment share the same common trend. This implies that the number of trends s must be either one or two, depending on whether the two countries are on the same trend path or on different ones.

Assume first that $s = 2$. One may wish to test that each of the two countries is on its own balanced growth path. Balanced growth in the first country corresponds to the hypothesis $\text{col } a_1 \subseteq \text{col } \psi$ with $a_1 := \sum_{i=1}^3 e_i = (1, 1, 1, 0, 0, 0)'$, while balanced growth in the second country is given by $\text{col } a_2 \subseteq \text{col } \psi$ with $a_2 := \sum_{i=4}^6 e_i = (0, 0, 0, 1, 1, 1)'$. These two hypotheses could be tested separately as $\text{col } a_i \subseteq \text{col } \psi$, $i = 1, 2$, or also jointly by considering $\text{col}(a_1, a_2) = \text{col } \psi$. These are three examples of hypotheses of the type $\text{col } a \subseteq \text{col } \psi$ considered in this paper.

Assume now that $s = 1$. Perfectly balanced growth between the two countries corresponds to the hypothesis $\psi = \iota$; under this assumption, country 1 and 2 are on the same growth path with equal loadings (same scale of the trend). Alternatively, proportional balanced growth can be represented as $\psi = (1, 1, 1, c, c, c)'$ with $c > 0$ representing the ratio between the loadings of country 2 and 1. Under this hypothesis, country 1 and 2 are on the same growth path but with different loadings if $c \neq 1$ (and hence different scale of the trend). This is an example of an hypothesis of the type $\text{col } \psi \subseteq \text{col } A$, where $A = (a_1, a_2)$.

2.3 | Autonomous Trends for Subsets

In various applications, including the one in Section 8, one is interested in testing whether a subset of variables X_{it} does not cointegrate, with $i \in \mathcal{I}_1 \subset \mathcal{I} := \{1, 2, \dots, p\}$. This hypothesis can be expressed as $\text{col}\{e_i, i \in \mathcal{I}_1\} \subseteq \text{col } \psi$, an hypothesis of the type $\text{col } a \subseteq \text{col } \psi$. All of the above are examples of hypotheses discussed in this paper.

3 | DGP

This section introduces the assumptions on the DGP, the property of dimensional coherence and the hypotheses of interest.

3.1 | Assumptions

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

$$\Delta X_t = C(L)\varepsilon_t, \quad (2)$$

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 3.1 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. $(0, \Omega_\varepsilon)$, $\Omega_\varepsilon > 0$.²

Assumption 3.1 (Assumptions on $C(z)$). Let $C(z) = C + C_1(z)(1 - z)$ in (2) 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) $\text{rank } C(z) < p$ only at isolated points $z = 1$ or $|z| > 1$;
- (iii) $\psi' C_1 \kappa_\perp$ has full row rank $r := p - s$, where $s := \text{rank } C$, $\psi_\perp$ (respectively $\kappa_\perp$) is a basis of $(\text{col } C)^\perp$ (respectively $(\text{col } C')^\perp$) and $C_1 := C_1(1)$.

Condition (i) is a standard one for integrated linear processes of integer order, see Johansen [6]. In the case of $p = n_\varepsilon$ conditions (ii) and (iii) are necessary and sufficient conditions for the process to have a (infinite order) VAR representation with error correction, see Johansen [6], Chapter 4. In the present context, condition (iii) guarantees that the relevant long run variance is nonsingular, see (A1) and (A2) in the Appendix. When $s > 0$, $C = \psi \kappa'$ is the rank decomposition of C , and cumulating (2), one finds the Common Trends (CT) representation

$$X_t = \gamma + \psi \kappa' \sum_{i=1}^t \varepsilon_i + C_1(L)\varepsilon_t, \quad t = 1, 2, \dots, \quad (3)$$

which gives a decomposition of X_t into initial conditions $\gamma := X_0 - C_1(L)\varepsilon_0$, the s -dimensional linearly independent random walk component $\kappa' \sum_{i=1}^t \varepsilon_i$ with $p \times s$ loading matrix ψ of full column rank, and the $I(0)$ linear process $C_1(L)\varepsilon_t$. Initial conditions can be treated as in Johansen [6] or as in Elliott [28], see also Müller and Watson [20].

Observe that X_t is $I(0)$ if $s = 0$, that is, $r = p$; it is $I(1)$ and cointegrated if $0 < r, s < p$ and it is $I(1)$ and non-cointegrated if $s = p$, that is, $r = 0$. Assumption 3.1 includes possibly nonstationary VARMA processes (for $n_\varepsilon = p$) in line with for example Stock and Watson [29] and Dynamic Factor Models (for $n_\varepsilon > p$) as in for example Bai [30].

If $s > 0$, the columns in the matrix ψ in (3) form a basis of $\text{col } C$, called the attractor space of X_t . Its orthogonal complement $(\text{col } C)^\perp$ 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.

3.2 | 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 (3), linear combinations of X_t also admit a CT representation, with a number of stochastic trends that is at most equal to the one in X_t .

FGP show that when H is a $p \times m$ full column rank matrix and X_t satisfies (2) with $C(z)$ fulfilling Assumption 3.1, then the same holds for $H'X_t$, that is $\Delta H'X_t = G(L)\varepsilon_t$, where $G(z) := H'C(z)$ satisfies Assumption 3.1 with $G(1) = H'C$ of rank $j \leq s$. This property is called 'dimensional coherence' of Assumption 3.1.

3.3 | Hypotheses of Interest

The hypotheses of interest are formulated in terms of β as $\text{col } b \subseteq \text{col } \beta \subseteq \text{col } B$, or equivalently in terms of ψ as

$$\text{col } a \subseteq \text{col } \psi \subseteq \text{col } A, \quad (4)$$

where $a = B_\perp$, $\psi = \beta_\perp$, $A = b_\perp$, and b and B are full column rank matrices. The aim of the paper is to make inferences about (4) from a sample $\{X_t\}_{t=1, \dots, T}$ observed from (3).

4 | Formulation of the Hypotheses

This section discusses the hypotheses of interest and their implications on the DGP. It also relates them to the dimensional coherence property in Section 3.

Let X_t be generated as in (3) with $s > 0$ stochastic trends and let $\Psi := \{\psi \in \mathbb{R}^{p \times s}, \text{rank } \psi = s\}$ indicate the unrestricted parameter space for ψ , that is the set of all DGPs that satisfy (3). Consider hypotheses H_0 on the cointegrating space $\text{col } \beta$ or the attractor space $\text{col } \psi$ of type (4), formulated as follows

$$\begin{aligned} H_0 : \text{col } b \subseteq \text{col } \beta &\Leftrightarrow \text{col } \psi \subseteq \text{col } A \\ &\Leftrightarrow \psi \in \Psi_0 := \{\psi = A\theta, \text{rank } \theta = s\}, \end{aligned} \quad (5)$$

$$\begin{aligned} H_0 : \text{col } \beta \subseteq \text{col } B &\Leftrightarrow \text{col } a \subseteq \text{col } \psi \\ &\Leftrightarrow \psi \in \Psi_0 := \{\psi = (a, a_\perp \varphi), \text{rank } \varphi = s - q\}, \end{aligned} \quad (6)$$

where $A = b_\perp$ is $p \times m$, θ is $m \times s$, $a = B_\perp$ is $p \times q$, $a_\perp$ is $p \times (p - q)$, φ is $(p - q) \times (s - q)$, all of full column rank, and $q \leq s \leq m$.³

The restricted parameter spaces $\Psi_0 \subset \Psi$ in (5) and (6) are next characterised in terms of specific subsystems of variables, on which the dimensional coherence property applies. Under hypothesis H_0 in (5), pre-multiplying (3) by $(\bar{A}, \bar{A}_\perp)'$ one finds

$$\bar{A}'X_t = \theta \kappa' \sum_{i=1}^t \varepsilon_i + \bar{A}'C_1(L)\varepsilon_t + \bar{A}'\gamma \quad (7)$$

$$\bar{A}_\perp'X_t = \bar{A}_\perp' C_1(L)\varepsilon_t + \bar{A}_\perp'\gamma \quad (8)$$

where $\text{rank}(\theta\kappa') = s$. Because $X_t = (A, A_\perp)(\bar{A}, \bar{A}_\perp)'X_t$, (5) is equivalent to the fact that (i) $A'X_t$ contains s stochastic trends and (ii) $A'_\perp X_t$ contains zero stochastic trends.

Similarly, consider hypothesis H_0 in (6) and partition κ conformably with $\psi = (a, a_\perp\varphi)$ as $\kappa = (\kappa_1, \kappa_2)$ with q and $s - q$ columns respectively, so that $\psi\kappa' = a\kappa'_1 + a_\perp\varphi\kappa'_2$. Pre-multiplying (3) by $(\bar{a}, \bar{a}_\perp)'$ one finds

$$\bar{a}'X_t = \kappa'_1 \sum_{i=1}^l \varepsilon_i + \bar{a}'C_1(L)\varepsilon_t + \bar{a}'\gamma \quad (9)$$

$$\bar{a}'_\perp X_t = \varphi\kappa'_2 \sum_{i=1}^l \varepsilon_i + \bar{a}'_\perp C_1(L)\varepsilon_t + \bar{a}'_\perp \gamma \quad (10)$$

where $\text{rank}(\varphi\kappa'_2) = s - q$. Hence (6) is equivalent to the fact that (i) $a'X_t$ contains q stochastic trends and (ii) $a'_\perp X_t$ contains $s - q$ stochastic trends.

Note that both hypotheses (5) and (6) have leading known subspace spanned by a $p \times l$ matrix H , where $(H, l) = (A, m)$ in (5) and $(H, l) = (a, q)$ in (6). Let $n := \min(l, s)$; the next theorem shows the formal equivalence between (5) and (7, 8) and between (6) and (9, 10).

Theorem 4.1 (Parameter space decomposition under the null). *The restricted parameter space Ψ_0 under the null hypotheses (5) or (6) has the following representation*

$$\begin{aligned} H_0 : \psi \in \Psi_0 &= \Psi_{01} \cap \Psi_{02} \\ \Psi_{01} &:= \{\psi \in \Psi : \text{rank}(H'\psi) = n\}, \\ \Psi_{02} &:= \{\psi \in \Psi : \text{rank}(H'_\perp\psi) = s - n\}, \end{aligned} \quad (11)$$

where $(H, n) = (A, s)$ in (5) and $(H, n) = (a, q)$ in (6).

In order to characterise the parameter space under the alternative $H_1 : \psi \in \Psi_0^c := \Psi \setminus \Psi_0$, where c indicates the complement of a set, it is useful to observe that $\text{rank}(H'\psi) \leq n$; this means that Ψ_{01} in (11) requires $\text{rank}(H'\psi)$ to be maximal, that is, equal to n . Moreover, consider the following inequality

$$\begin{aligned} s := \text{rank } \psi &= \text{rank}((\bar{H}, \bar{H}_\perp)(H, H_\perp)'\psi) = \text{rank} \begin{pmatrix} H'\psi \\ H'_\perp\psi \end{pmatrix} \\ &\leq \text{rank}(H'\psi) + \text{rank}(H'_\perp\psi); \end{aligned} \quad (12)$$

this implies $\text{rank}(H'_\perp\psi) \geq s - \text{rank}(H'\psi)$, and hence, substituting $\text{rank}(H'\psi) \leq n$, one finds $\text{rank}(H'_\perp\psi) \geq s - n$. This means that Ψ_{02} in (11) requires $\text{rank}(H'_\perp\psi)$ to be minimal, that is, equal to $s - n$. Summarising, for any $\psi \in \Psi$ one has

$$\text{rank}(H'\psi) \leq n, \quad \text{rank}(H'_\perp\psi) \geq s - n,$$

with equalities under the null hypothesis $\psi \in \Psi_0$, see (11). Under the alternative at least one of these inequalities is strict, that is,

$$\begin{aligned} H_1 : \psi \in \Psi_0^c &= \Psi_{01}^c \cup \Psi_{02}^c \\ \Psi_{01}^c &= \{\psi \in \Psi : \text{rank}(H'\psi) < n\}, \end{aligned}$$

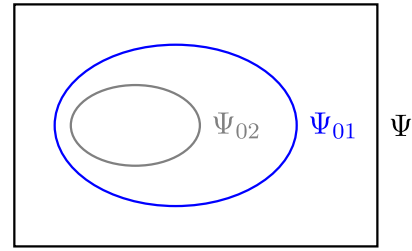


FIGURE 1 | Representation the unrestricted (Ψ) and restricted ($\Psi_0 = \Psi_{01} \cap \Psi_{02}$) parameter space in terms of subsets. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/obes.70054)]

$$\Psi_{02}^c = \{\psi \in \Psi : \text{rank}(H'_\perp\psi) > s - n\}. \quad (13)$$

Inequality (12) has also the following implications about the nesting between parameter spaces under the null H_0 and under the alternative H_1 .

Theorem 4.2 (Nesting of subspaces and decomposition under the alternative). *Let Ψ_{0i} be defined in (11) and Ψ_{0i}^c in (13); one has*

- (i) $\Psi_{02} \subset \Psi_{01}$ (i.e., $\Psi_{02}^c \supset \Psi_{01}^c$),
- (ii) $\Psi_0^c = \Psi_{02}^c$,
- (iii) $\Psi_0^c = \Psi_{01}^c \cup (\Psi_{01} \cap \Psi_{02}^c)$.

Part (i) of the theorem states that there are no parameter values in Ψ for which Ψ_{02} holds and Ψ_{01} fails so that there is no intersection between Ψ_{01}^c and Ψ_{02} ; this is represented in Figure 1.

Part (ii) of the theorem states that the parameter space under the alternative (Ψ_0^c) is equal to the one of the alternative to the second subhypothesis (Ψ_{02}^c) which in term of Figure 1 is everything outside the grey ellipse. Part (iii) states that this region can be partitioned in 2 subregions: the one between the grey and blue ellipse ($\Psi_{01} \cap \Psi_{02}^c$) and the one outside the blue ellipse (Ψ_{01}^c).

As an illustration let

$$\psi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = (e_1, e_2 + e_3) \quad (14)$$

and consider the hypothesis $\text{col } a \subseteq \text{col } \psi$ with $a = e_1$, $s = 2$ and $q = 1$, which is satisfied in (14). This hypothesis is of type (6) and states that the first variable does not cointegrate, that is, that it is autonomous, see Section 2.3. One can illustrate decomposition (11) calculating $\text{rank}(a'\psi) = \text{rank}(1, 0) = 1 = q$ and $\text{rank}(a'_\perp\psi) = \text{rank} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = 1 = s - q$; one can see that $\psi \in \Psi_{01}$ and $\psi \in \Psi_{02}$, as needs to be the case, see (11).

Consider next the hypothesis $\text{col } a \subseteq \text{col } \psi$ with $a = e_2$, which is not satisfied in (14), so that $\psi \in \Psi_0^c$. One can investigate if $\psi \in \Psi_{01}^c$ or $\psi \in \Psi_{02}^c$ (or both); one has $\text{rank}(a'\psi) = \text{rank}(0, 1) = 1 =$

q , which means $\psi \in \Psi_{01}$, and $\text{rank}(a'_\perp \psi) = \text{rank}(I_2) = 2 > s - q$, which means $\psi \in \Psi_{02}^c$. This illustrates that $\mathcal{H}_1$ holds, with only one component of the alternative being true ($\psi \in \Psi_{01}$ and $\psi \in \Psi_{02}^c$).

Finally consider the hypothesis $\text{col } a \subseteq \text{col } \psi$ with $a = e_2 - e_3$ which is also not satisfied in (14), that is, $\psi \in \Psi_0^c$. Again one can verify if $\psi \in \Psi_{01}^c$ or $\psi \in \Psi_{02}^c$; one finds $\text{rank}(a'_\perp \psi) = \text{rank}(0, 0) = 0 < q = 1$ which means $\psi \in \Psi_{01}^c$ and $\text{rank}(a'_\perp \psi) = \text{rank}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = 2 > s - q = 1$, which means $\psi \in \Psi_{02}^c$. In this case $\mathcal{H}_1$ holds, with both components of the alternative being verified ($\psi \in \Psi_{01}^c$ and $\psi \in \Psi_{02}^c$).

This illustrates all possibilities, see Theorem 4.2.

5 | Statistical Analysis

This section defines the selection criteria for hypotheses $\mathcal{H}_0$ based on the estimators of the number s of stochastic trends (and the cointegrating rank r) proposed in FGP. The estimators are based on a sample canonical correlation analysis, which is introduced first.

5.1 | Squared 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 $L^2[0, 1]$. The leading example is the Karhunen-Loève basis for Brownian motion which corresponds to $\phi_k(u) = \sqrt{2} \sin\left((k - \frac{1}{2})\pi u\right)$, see (A3) and (A4) in the Appendix. Let d_t be the $K \times 1$ vector constructed evaluating $\varphi_K(\cdot)$ at the discrete sample points $1/T, \dots, (T-1)/T, 1$, that is,

$$d_t := \varphi_K(t/T) := (\phi_1(t/T), \dots, \phi_K(t/T))',$$

$$K \geq p, \quad t = 1, \dots, T. \quad (15)$$

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 (15), denoted as $\text{cca}(y_t, d_t)$,⁴ consists in solving the following generalised eigenvalue problem, see for example, Johansen [6] and references therein,

$$\text{cca}(y_t, d_t) : |\lambda M_{yy} - M_{yd} M_{dd}^{-1} M_{dy}| = 0,$$

$$M_{ij} := T^{-1} \sum_{t=1}^T i_t j_t'. \quad (16)$$

This delivers eigenvalues $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ (the squared canonical correlations) and corresponding eigenvectors $v_1, v_2, \dots, v_p$ which are used to define canonical variates.

The next two sections present estimators of the number of stochastic trends s , generically indicated by $\hat{s}$. When needed, the notation $\hat{s}(y_t)$ is employed to indicate the dependence on the system of variables y_t used in $\text{cca}(y_t, d_t)$; y_t equals X_t when performing the estimation on the full system, while y_t equals $H'X_t$ and $H'_\perp X_t$ corresponding to subsystems (7), (8) for $H = A$, and (9), (10) for $H = a$.

5.2 | Argmax Estimators

The following definition reports estimators based on the maximisation of functions of the squared canonical correlations.

Definition 5.1 (Argmax estimators $\hat{s}$ of s). Define the maximal gap (max-gap) estimator of s as

$$\underset{i \in \{0, \dots, p\}}{\text{argmax}} (\lambda_i - \lambda_{i+1}), \quad \lambda_0 := 1, \quad \lambda_{p+1} := 0, \quad (17)$$

where $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ are the eigenvalues of $\text{cca}(X_t, d_t)$. Define also the following alternative argmax estimator:

$$\underset{i \in \{0, 1, \dots, p\}}{\text{argmax}} \frac{\prod_{h=1}^i \lambda_h}{\prod_{h=i+1}^p \left(\frac{T}{K} \lambda_h\right)}, \quad (18)$$

where empty products are equal to one, and use is made of the rates in Equation (7) below Theorem 4.1 in FGP for $(T, K)_{seq} \rightarrow \infty$.

Remark 5.2 (Alternative criteria). The criterion (18) applies ideas in Bierens [14] to the set of eigenvalues in $\text{cca}(X_t, d_t)$. Alternative criteria inspired by Ahn and Horenstein [31] when applied to eigenvalues in $\text{cca}(X_t, d_t)$ are discussed in FGP; they replace the function in the argmax in (18) with

$$\frac{\lambda_i}{\lambda_{i+1}} \quad \text{or} \quad \frac{\log(1 + \lambda_i / \sum_{h=i+1}^p \lambda_h)}{\log(1 + \lambda_{i+1} / \sum_{h=i+2}^p \lambda_h)}, \quad (19)$$

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 defines estimators $\hat{s}$ of the number of stochastic trends based on sequences of test similar to the one in Johansen [6] for cointegration rank determination. The procedure is described here for the full system X_t in (3) with s stochastic trends, but can be similarly be applied to subsystems $H'X_t$ and $H'_\perp X_t$ as in (7), (8) for $H = A$, and (9), (10) for $H = a$.

Consider hypothesis $\mathcal{H}_0^* : s = i$ versus $\mathcal{H}_1^* : s < i$ at significance level η with test statistics $K\pi^2 \|\tau^{(i)}\|_h$, where $\tau^{(i)} := (1 - \lambda_i, 1 - \lambda_{i-1}, \dots, 1 - \lambda_1)'$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ are the eigenvalues of $\text{cca}(X_t, d_t)$ and $\|\cdot\|_h$ is the h -norm for vectors. The two special cases of interest are $h = 1, \infty$; the choice $h = 1$ gives $\|\tau^{(i)}\|_1 = \sum_{k=1}^i (1 - \lambda_k)$, while $h = \infty$ yields $\|\tau^{(i)}\|_\infty = 1 - \lambda_i$; these expressions are similar to the trace and $\lambda_{\max}$ statistics in Johansen [6]. The critical values $c_{i,h,\eta}$ for these tests are discussed in Section 6.1 below.

The test sequence is defined as follows. Start with the test of $\mathcal{H}_0^* : s = i$ for $i = p$; if the test does not reject, then $\hat{s} = p$, otherwise progress further. Test $\mathcal{H}_0^* : s = i$ for $i = p-1, p-2, \dots$ and proceed as before until $\mathcal{H}_0^* : s = k$, say, is not rejected; in this case

$\hat{s} = k$. Otherwise, all tests of $\mathcal{H}_0^* : s = i, i = p, \dots, 1$, reject, and $\hat{s} = 0$.

5.4 | Decision Rules on $\mathcal{H}_0$

This section describes decision rules for rejecting $\mathcal{H}_0 : \psi \in \Psi_0$, see (11). The rules are expressed as functions of a generic estimator $\hat{s}$ of the number of stochastic trends defined in Sections 5.2 and 5.3; remark that when $\hat{s}$ is based on test sequences as in Section 5.3, one needs to specify the significance levels used in each test in the sequence.

The ‘joint decision rule’ does not reject $\psi \in \Psi_0$ whenever z equals one, where

$$\begin{aligned} z &:= w \cdot v, \\ w &:= 1(\hat{s}(H'X_t) = n), \\ v &:= 1(\hat{s}(H'_\perp X_t) = n - s), \end{aligned} \quad (20)$$

and $(H, n) = (A, s)$ for (5) and $(H, n) = (a, q)$ for (6).

Note that z checks both implications $\psi \in \Psi_{01}$ and $\psi \in \Psi_{02}$, via the indicators w and v , and rejects $\psi \in \Psi_0$ if either $\psi \in \Psi_{01}$ or $\psi \in \Psi_{02}$ is rejected, or both. An alternative procedure, called the ‘single decision rule’, does not reject $\psi \in \Psi_0$ if v in (20) equals one, and rejects $\psi \in \Psi_0$ if v equals 0. The rationale for this single decision rule is that if $\psi \in \Psi_{02}$, then also $\psi \in \Psi_{01}$, see Figure 1 and Theorem 4.2. Remark that, given the stochasticity of $\hat{s}$, the decision rules applied to the subsystems may fail to reflect the nesting in Theorem 4.2. As a result, the properties of both the single and the joint decision rules are of interest; they are discussed in the following section.

6 | Asymptotic Properties

This section discusses the asymptotic properties of the statistics introduced in Section 5.

6.1 | Asymptotic Properties of Squared Canonical Correlations

Asymptotics is performed as $(T, K)_{seq} \rightarrow \infty$, which is consistent with $K/T = o(1)$. Let d_t in (15) be constructed using any orthonormal càdlàg basis of $L^2[0, 1]$ and let $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ be the eigenvalues of $\text{cca}(X_t, d_t)$, see (16); FGP show that for $(T, K)_{seq} \rightarrow \infty$,

$$\lambda_i \xrightarrow{p} \begin{cases} 1 & i = 1, \dots, s \\ 0 & i = s + 1, \dots, p \end{cases} \quad (21)$$

which implies $\mathbb{P}(\hat{s} = s) \rightarrow 1$ as $(T, K)_{seq} \rightarrow \infty$ for $\hat{s}$ equal to the max-gap estimator in (17) or the alternative argmax criterion in (18).

Moreover, when d_t in (15) is constructed using the KL basis as in Section 5.1, see also (A4), FGP show that for $(T, K)_{seq} \rightarrow \infty$, the s largest eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s$ satisfy

$$K\pi^2\tau^{(s)} \xrightarrow{w} \zeta^{(s)}, \quad (22)$$

where $\tau^{(s)} := (1 - \lambda_s, 1 - \lambda_{s-1}, \dots, 1 - \lambda_1)'$, $\zeta^{(s)} := (\zeta_1, \zeta_2, \dots, \zeta_s)'$, and $\zeta_1 \geq \zeta_2 \geq \dots \geq \zeta_s$ are the eigenvalues of $(\int B_1 B_1')^{-1}$ and $B_1(u)$ is a standard Brownian motion of dimension s , see the paragraph below (A2) in the Appendix. Additionally, the r smallest eigenvalues $\lambda_{s+1} \geq \lambda_{s+2} \geq \dots \geq \lambda_p$ satisfy $K\pi^2(1 - \lambda_i) \xrightarrow{p} \infty$, $i = s + 1, \dots, p$, which implies, $K\pi^2\|\tau^{(i)}\|_h \rightarrow \infty$ for $i > s$.

Observe that $\zeta^{(s)}$ in (22) does not depend on nuisance parameters, and thus tests based on $K\pi^2\tau^{(s)}$ are asymptotically pivotal. One consequence of this is that the quantiles $c_{i,h,\eta}$ of $K\pi^2\|\tau^{(s)}\|_h \xrightarrow{w} \|\zeta^{(s)}\|_h$ can be estimated by Monte Carlo simulation.⁵ This limit distribution is specific to the KL basis.

Some consequences of these results are collected in the following corollary.

Corollary 6.1 (Asymptotic properties of the eigenvalues with the KL basis). *Let $\hat{s}_h(X_t)$ be the estimator of s based on the test sequence $K\pi^2\|\tau^{(i)}\|_h$ for $\text{cca}(X_t, d_t)$, with significance level η , see Section 5.3;⁶ then as $(T, K)_{seq} \rightarrow \infty$ the following holds:*

- (i) $\mathbb{P}(\hat{s}_h(X_t) = k) \rightarrow 0$ for $k > s$.
- (ii) $\mathbb{P}(\hat{s}_h(X_t) = s) \rightarrow 1 - \eta$ for $0 < s \leq p$ and $\mathbb{P}(\hat{s}_h(X_t) = s) \rightarrow 1$ for $s = 0$;
- (iii) $\mathbb{P}(\hat{s}_h(X_t) = k) \leq \rho_T \rightarrow \eta$ for $k < s$.

6.2 | Univariate Case

This section contains the explicit form of the p.d.f. and c.d.f. of the asymptotic distribution of $K\pi^2\|\tau^{(s)}\|_h$, $h = 1, \infty$, for the special case $s = 1$. This limit distribution is related to the well known functional $\int B^2$ of univariate Brownian motion $B(u)$, which has been studied since Cameron and Martin [32], see [33, 34] and references therein. The following theorem derives the formulae for the p.d.f. and c.d.f. of (22) using these previous results.

Theorem 6.2 (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*

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

where $\eta_j := \left(\frac{-1}{j}\right)$, $a_j := 2j + \frac{1}{2}$ and $\Gamma(k, t) = \int_t^\infty z^{k-1} e^{-z} dz$ 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 expression of the p.d.f. can be shown to match the results in Doornik et al. [35, Theorem 3], who compute $1 - \frac{1}{2}\mathbb{E}(\int B^2)^{-1} \simeq -1.78143$, which implies $\mathbb{E}(\zeta^{(1)}) \simeq 2(1 + 1.78143) = 5.5629$. This

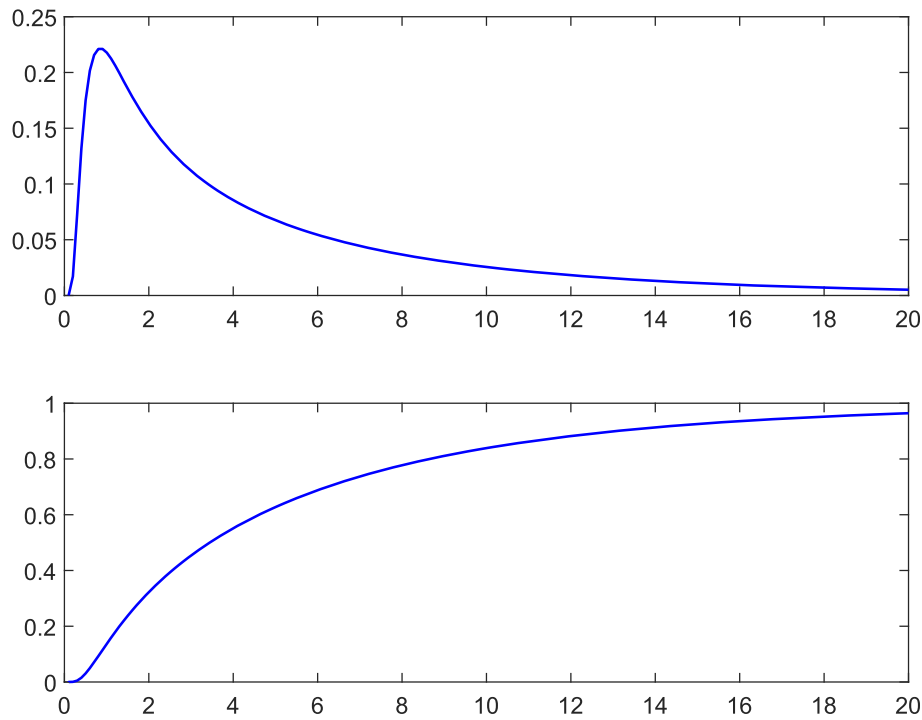


FIGURE 2 | p.d.f. and c.d.f. of $\zeta^{(1)}$ in Theorem 6.2. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/obes.70054)]

agrees with the analytical expectation that can be derived from the expression $f_\zeta(z)$ above, namely

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

Figure 2 pictures the p.d.f. and c.d.f. in Theorem 6.2.⁸

6.3 | Decision Rules on $\mathcal{H}_0$

This section presents asymptotic properties of the joint and single selection criteria, see (20), for the hypothesis $\psi \in \Psi_0$.

Theorem 6.3 (Limit behavior of decision rules). *Let Assumption 3.1 hold and consider $(T, K)_{seq} \rightarrow \infty$. The following holds for $\hat{s}$ equal to the max-gap or the alternative argmax criterion in Section 5.2:*

(i) *Under the null $\psi \in \Psi_0$, one has*

$$\mathbb{P}(z = 1) \rightarrow 1, \quad \mathbb{P}(w = 1) \rightarrow 1, \quad \mathbb{P}(v = 1) \rightarrow 1. \quad (23)$$

(ii) *Under the alternative $\psi \in \Psi_0^c$, one has $\mathbb{P}(z = 0) \rightarrow 1$ because $\mathbb{P}(v = 0) \rightarrow 1$ irrespectively of the behavior of w , which is the following: if $\psi \in \Psi_{01}^c$, one has $\mathbb{P}(w = 0) \rightarrow 1$, whereas if $\psi \in \Psi_{01} \cap \Psi_{02}^c$, one has $\mathbb{P}(w = 1) \rightarrow 1$.*

Next let $\hat{s}$ be equal to the estimator based on the test sequences in Section 5.3, with significance level ν (respectively η) associated with the tests involved in w (respectively v); moreover, define $\varsigma := 0$ for (5) and $\varsigma := \eta$ for (6); the following holds:

(iii) *Under the null, $\psi \in \Psi_0$, one has*

$$\begin{aligned} \mathbb{P}(z = 1) &\geq \rho_{1T} \rightarrow 1 - \nu - \varsigma, \\ \mathbb{P}(w = 1) &\rightarrow 1 - \nu, \\ \mathbb{P}(v = 1) &\rightarrow 1 - \varsigma; \end{aligned} \quad (24)$$

(iv) *under the alternative, if $\psi \in \Psi_{01}^c$, one has $\mathbb{P}(w = 0) \rightarrow 1$ and $\mathbb{P}(z = 0) \rightarrow 1$, whereas if $\psi \in \Psi_{01} \cap \Psi_{02}^c$, one has $\mathbb{P}(w = 0) \rightarrow \nu$ and $\mathbb{P}(v = 0) \geq \rho_{2T} \rightarrow 1 - \eta$ which implies $\mathbb{P}(z = 0) \geq \rho_{2T} \rightarrow 1 - \eta$.*

A few remarks are in order.

Remark 6.4 (Properties of the criteria). Both the joint and single decision rules based on the max-gap or the alternative argmax estimator consistently select the hypothesis under the null and reject it under the alternative with limit probability one. This is different from the case of procedures based on test sequences, where the limit probabilities depend on the significance levels η , ν , as well as on the quantity ς , which is zero in the case of hypothesis (5).

Remark 6.5 (Non uniformity). The behavior of the procedures under the alternative $\psi \in \Psi_0^c$ depends on which part of the alternative holds, in terms of the decomposition in Theorem 4.2. The bounds given in Theorem 6.3 (iv) are conservative, that is, they consider the worst case under the alternative; see the empirical power in Table 1.

Remark 6.6 (Properties of the test sequences and selection of ν , η .). The criteria based on tests have different asymptotic

TABLE 1 | Top panel: Frequency of incorrect selection of s . Middle panel: Rejection frequency under the null (27) and (28). Bottom panel: Rejection frequency under the alternative (29) and (30). The sequential tests are conducted at 5% significance level and the number of Monte Carlo replications is $N = 10^4$.

		Max-gap	Altern. argmax	$K\pi^2\ \tau^{(i)}\ _\infty$	$K\pi^2\ \tau^{(i)}\ _1$	Max-gap	Altern. argmax	$K\pi^2\ \tau^{(i)}\ _\infty$	$K\pi^2\ \tau^{(i)}\ _1$
$\mathcal{H}_0$	s	$T = 150$				$T = 300$			
		Frequency of incorrect selection of s							
	1	0	0.98	0.02	0.02	0	0.06	0.04	0.04
	10	0.04	0.01	0.01	0.66	0	0	0.02	0.01
	19	0.30	0.99	0.39	0.93	0	0	0.03	0.06
		Empirical size							
	1	0	0.95	0.02	0.02	0	0.04	0.04	0.04
(27)	10	0.03	0.01	0.01	0.58	0	0	0.02	0.01
	19	0.01	0.99	0	0	0	0	0.03	0
	1	0	0.98	0.04	0.04	0	0.08	0.05	0.05
(28)	10	0.02	0.02	0.05	0.46	0	0	0.07	0.06
	19	0.25	0.95	0.21	0.90	0	0	0.07	0.07
		Empirical power							
	1	1	1	1	1	1	1	1	1
(29)	10	1	1	1	0.98	1	1	1	1
	19	1	1	0.99	0.96	1	1	1	1
	1	1	1	1	1	1	1	1	1
(30)	10	1	1	1	1	1	1	1	1
	19	1	1	1	1	1	1	1	1

properties. Equation (24) shows how to control the size of the test asymptotically. Take for instance the case of the single decision rule for (5), with $\zeta = 0$; one can choose η equal to some significance level, such as 5%, and in this case the single decision rule selects the null correctly with limit probability one and rejects it when false with limit probability 95%.

With this setting, the joint criterion makes the correct choice with limit probability at least equal to 95% under the null and under the alternative. Similar reasoning applies for the hypothesis (6), where one can choose $\nu = \eta = 0.025$; in this case the joint (respectively single) criterion makes the correct choice with limit probability 95% (respectively 97.5%) under the null and the alternative.

7 | Simulations

This section reports a Monte Carlo (MC) simulation study; results show that the proposed methods perform well in systems with moderately large cross-sectional and time series dimension. The DGP is taken from Onatski and Wang [9] and Bykhovskaya and Gorin [11]. The data is generated from

$$\begin{aligned}
 \Delta X_t &= \alpha \beta' X_{t-1} + \varepsilon_t, \\
 \varepsilon_t &\sim \text{i.i.d. } N(0, I_p), \\
 t &= 1, \dots, T, \\
 X_0 &= 0,
 \end{aligned} \tag{25}$$

with $\beta = (0, I_{p-s})'$ and $\alpha = -\beta$, so that $\psi = (I_s, 0)'$. The values of p, T, K , and s considered in the MC design are

$$p = 20, \quad T = 150, 300, \quad K = \lceil T^{3/4} \rceil, \quad s = 1, 10, 19, \tag{26}$$

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, the small sample performance of the estimators in detecting the correct number s of stochastic trends in the p -dimensional system (25) is considered. The top panel in Table 1 reports the frequency of wrong selection of s based on the different $\hat{s}$ estimators.

For $T = 150$, the max-gap and the sequential test based on $K\pi^2\|\tau^{(i)}\|_\infty$ behave similarly (upper-left part of the table): For $s = 1$ and $s = 10$ the frequency of correct detections is close to one, while for $s = 19$ it drops down to 70% and 60%, respectively. The alternative argmax estimator performs well only for $s = 10$ (99% of correct detections) and basically always fails when $s = 1, 19$. The sequential test based on $K\pi^2\|\tau^{(i)}\|_1$ works well only for $s = 1$. This shows that a ratio of $T/p = 150/20 = 7.5$ does not guarantee a uniformly good performance of all estimators, which is instead achieved for $T = 300$ (upper-right part of the table) with a ratio of $T/p = 300/20 = 15$.

Next, in order to investigate the small sample performance of the inference procedures in Theorem 6.3 under the null, consider the hypotheses

$$H_0 \text{ of type (4.1) : } \text{col } \psi \subseteq \text{col}(e_1, \dots, e_{p-1}) \\ (\text{i.e., } \text{col } e_p \subseteq \text{col } \beta), \quad (27)$$

$$H_0 \text{ of type (4.2) : } \text{col } e_1 \subseteq \text{col } \psi \\ (\text{i.e., } \text{col } \beta \subseteq \text{col}(e_2, \dots, e_p)), \quad (28)$$

which hold in the DGP. Hypothesis (27) is of type (5) with $A = b_\perp = (e_1, \dots, e_{p-1})$ and $A_\perp = b = e_p$, while hypothesis (28) is of type (6) with $a = B_\perp = e_1$, $q = 1$, and $a_\perp = B = (e_2, \dots, e_p)$. Inference on them is performed via the z statistic in (20), which checks for (27) whether (i) $A'X_t$ (the first $p-1 = 19$ variables) contains $s = 19$ stochastic trends and (ii) $A'_\perp X_t$ (the last variable) is stationary. For (28) the z statistic checks whether (i) $a'X_t$ (the first variable) contains $q = 1$ stochastic trends and (ii) $a'_\perp X_t$ (the last $p-1 = 19$ variables) contains $s-q = 19-1 = 18$ stochastic trends.

The second block of rows in Table 1 (corresponding to (27)) reports the frequency of $z = 0$ (i.e., of rejection) under the null, given that (27) holds in the DGP, using different estimators $\hat{s}$. Similarly to above, $T = 150$ does not guarantee a uniformly good performance of all estimators, with the max-gap and the sequential test based on $K\pi^2\|\tau^{(i)}\|_\infty$ performing very well uniformly in s , the alternative argmax estimator performing well only for $s = 10$ and the sequential test based on $K\pi^2\|\tau^{(i)}\|_1$ for $s = 1$ and $s = 19$. For $T = 300$ a uniformly good performance of all estimators is instead achieved. Similar results hold for hypothesis (28), see the third block of rows in Table 1 (corresponding to (28)), with a slightly less satisfactory behavior of the alternative argmax estimator and the sequential test based on $K\pi^2\|\tau^{(i)}\|_1$ when $T = 150$.

Finally, in order to investigate the small sample performance of the inference procedures in Theorem 6.3 under the alternative, consider the hypotheses

$$H_0 \text{ of type (4.1) : } \text{col } \psi \subseteq \text{col}(e_2, \dots, e_p) \\ (\text{i.e., } \text{col } e_1 \subseteq \text{col } \beta), \quad (29)$$

$$H_0 \text{ of type (4.2) : } \text{col } e_p \subseteq \text{col } \psi \\ (\text{i.e., } \text{col } \beta \subseteq \text{col}(e_1, \dots, e_{p-1})), \quad (30)$$

which do not hold in the DGP. Hypothesis (29) is of type (5) with $A = b_\perp = (e_2, \dots, e_p)$ and $A_\perp = b = e_1$, while

hypothesis (30) is of type (6) with $a = B_\perp = e_p$, $q = 1$, and $a_\perp = B = (e_1, \dots, e_{p-1})$.

The bottom panel of Table 1 reports the frequency of $z = 0$, that is, the rejection frequency under the alternative, using different estimators: The empirical power of z in (20) appears uniformly good for all of them already for $T = 150$.

8 | Empirical Application

This section provides an illustration of the methods on a panel of daily exchange rates from January 04, 2022 to August 30, 2024 of the US dollar against 20 World Markets (WM) currencies, downloaded from the Federal Reserve Economic Data (FRED) website, <https://fred.stlouisfed.org/>; see Onatski and Wang [10] for a similar dataset and a review of the literature on the topic.

The sample size is $T = 667$ and the total number of series is $p = 20$, with a ratio $\lfloor T/p \rfloor = 33$. Data (in logs and normalised to start at zero) are plotted in Figure 3, where the 20 WM currencies are grouped into six European (EU) exchange rates (Denmark (*DK*), Eurozone (*Euro*), Norway (*NK*), Sweden (*SK*), Switzerland (*SF*), United Kingdom (*BP*)) and 14 Non-European (Non-EU) exchange rates (Australia, Brazil, Canada, China, Hong Kong, India, Japan, Malaysia, Mexico, Singapore, South Africa, South Korea, Taiwan, Thailand).

The top panels in Figure 4 report results for the $p = 20$ WM exchange rates (X_t) with $K = \lceil T^{3/4} \rceil = 132$. The first panel shows the eigenvalues of $\text{cca}(X_t, d_t)$ where the first 19 eigenvalues are above 0.75 while λ_{20} drops down to 0.25 approximately, so that the largest gap $\lambda_i - \lambda_{i+1}$ is found between λ_{19} and λ_{20} , for $i = 19$. This indicates the presence of $s = 19$ stochastic trends, that is, $r = 1$ cointegrating relation, in the $p = 20$ -dimensional WM system. The second panel reports the test statistic $K\pi^2\tau^{(s)}$, see (22), and the corresponding 95% asymptotic confidence stripe $\mathcal{B}$, see FGP; given that $K\pi^2\tau^{(s)} \in \mathcal{B}$, the model assumptions—including the selection of s —appear approximately valid, hence supporting the choice $s = 19$.

Table 2 reports results for the argmax estimators and the sequential tests at 5% significance level; the different criteria agree, with the exception of the test $K\pi^2\|\tau^{(i)}\|_1$. Given these overall results,

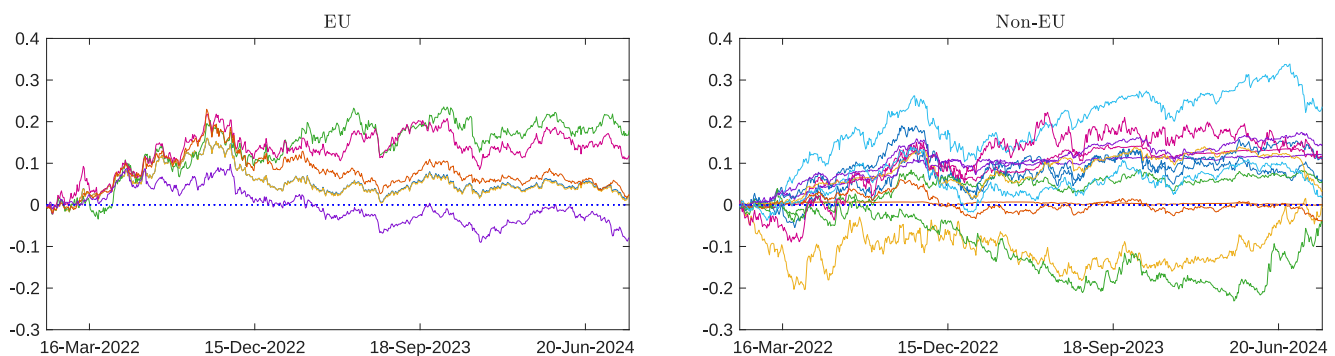


FIGURE 3 | Daily exchange rates (in logs and normalised to start at zero) from January 04, 2022 to August 30, 2024 of the 20 WM currencies: 6 EU (left panel) and 14 Non-EU (right panel). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

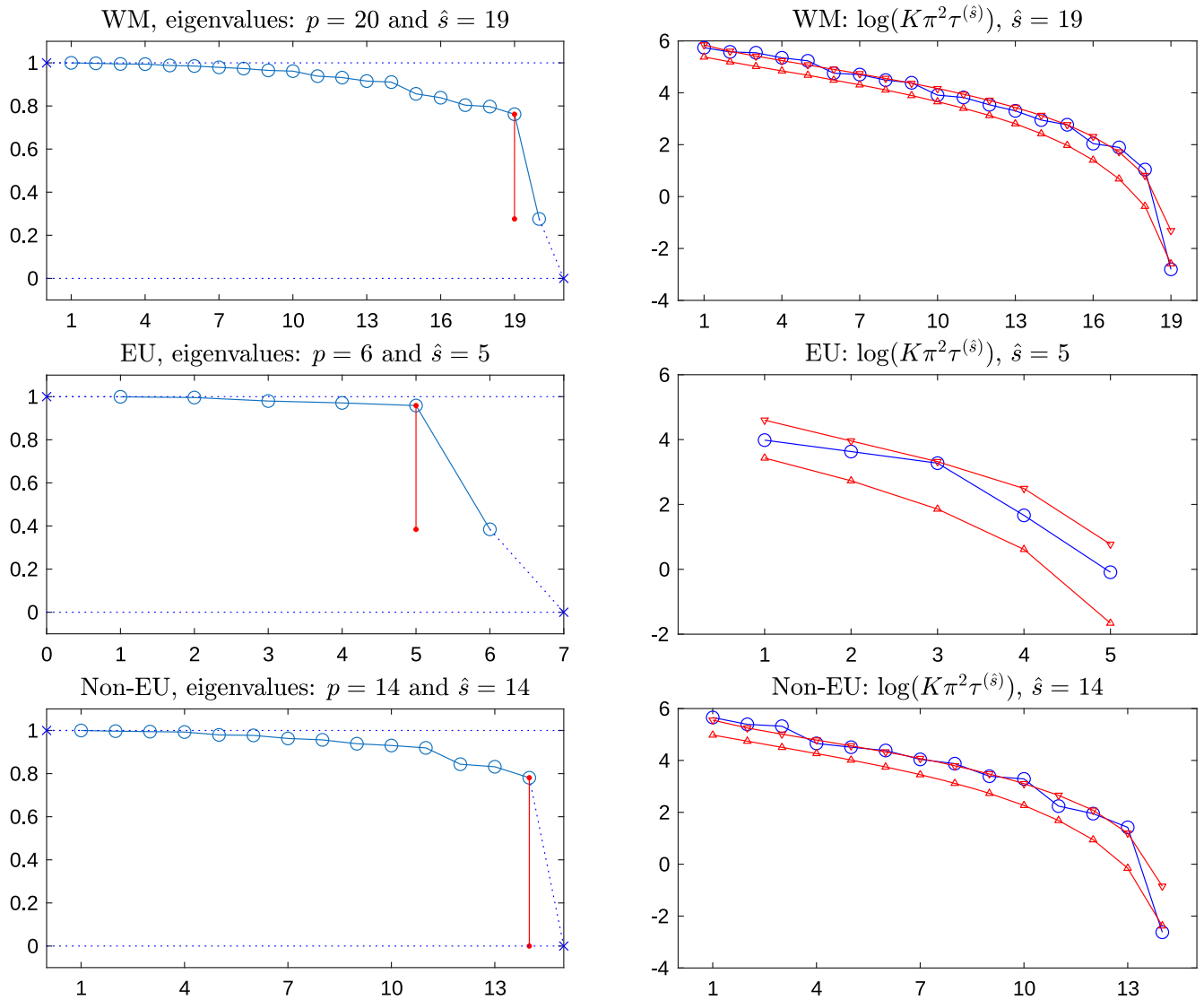


FIGURE 4 | Analysis of the 20 WM (top panels), the six EU (middle panels) and the 14 Non-EU (bottom panels) currencies: Eigenvalues, max-gap estimator and $\log(K\pi^2\tau^{(\hat{s})})$ for $K = \lceil T^{3/4} \rceil = 132$ and with a 95% stripe around $\mathbb{E}(\log \zeta^{(\hat{s})})$. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/obes.70054)]

TABLE 2 | Analysis of the 20 WM, the six EU and the 14 Non-EU currencies: Estimates of s in with the argmax estimators and with the sequential tests at 5% significance level.

Estimate of s	Alternative			
	Max-gap	argmax	$K\pi^2\ \tau^{(i)}\ _\infty$	$K\pi^2\ \tau^{(i)}\ _1$
WM ($p = 20$)	19	19	19	17
EU ($p = 6$)	5	5	5	5
Non-EU ($p = 14$)	14	14	13	12

s is fixed at 19 for the $p = 20$ -dimensional WM system in the rest of the analysis.

It is next investigated whether the cointegrating relation can be found in the 6-dimensional EU system. This is an hypothesis of type (6), with $a = (0, I_q)'$, $q = 14$, and $a_\perp = (I_{p-q}, 0)'$, $p - q = 6$ and it is tested by checking whether (i) $a'X_t$ (the 14-dimensional Non-EU system) contains $q = 14$ stochastic trends and (ii) $a'_\perp X_t$

(the 6-dimensional EU system) contains $s - q = 19 - 14 = 5$ stochastic trends, see Theorem 4.1.

The results are reported in the middle and bottom panels in Figure 4: The first five eigenvalues of the 6-dimensional EU system are close to one while λ_6 drops down to 0.4 approximately, which indicates the presence of $s = 5$ stochastic trends, that is, $r = 1$ cointegrating relation, in the $p = 6$ -dimensional EU system. Hence (ii) is not rejected using the max-gap estimator.

The $p = 14$ -dimensional Non-EU system, see the bottom panels in Figure 4, displays eigenvalues all above 0.75, and indicates the presence of $s = 14$ stochastic trends, that is, no cointegration. Hence (i) is also not rejected and thus the hypothesis that the cointegrating relation is found in the 6-dimensional EU system is not rejected using the max-gap estimator.

The same conclusion, see Table 2, is found using the alternative argmax estimator. On the contrary, with the sequential tests one rejects the hypothesis because (ii) is not rejected ($\hat{s} = 5$) but (i)

TABLE 3 | Analysis of the six EU currencies: Estimates of s in (i) $A'X_t$ and (ii) $A'_\perp X_t$ with the argmax estimators and with the sequential tests at 5% significance level.

Estimate of s in (i), (ii)	Max-gap	Alternative argmax	$K\pi^2\ \tau^{(i)}\ _\infty$	$K\pi^2\ \tau^{(i)}\ _1$
$j = DK$	5, 1	5, 1	5, 0	5, 0
NK	5, 1	5, 1	5, 1	5, 1
SK	5, 1	5, 1	5, 1	5, 1
SF	5, 1	5, 1	5, 1	5, 1
BP	5, 1	5, 1	5, 1	5, 1

is rejected ($\hat{s} < 14$). Given these overall results, in the rest of the analysis s is fixed at five for the $p = 6$ -dimensional EU system and at 14 for the $p = 14$ -dimensional Non-EU system.

In order to investigate the structure of the cointegrating space and of the attractor space, consider the hypothesis that a given currency differential with the Euro in the 6-dimensional EU system is stationary. This is an hypothesis of type (5) with $b = A_\perp = e_j - e_{Euro}$, where $j \in \{DK, NK, SK, SF, BP\}$, and $A = b_\perp = (e_{DK} + e_{Euro}, e_{NK}, e_{SK}, e_{SF}, e_{BP})$ for DK and similarly for the other currencies NK, SK, SF, BP . This is tested by checking whether (i) $A'X_t$ (the sum of the given currency and the Euro, and the other currencies) contains $s = 5$ stochastic trends and (ii) $A'_\perp X_t$ (the given currency differential with the Euro) contains 0 stochastic trends, see Theorem 4.1.

The results are reported in Table 3: For NK, SK, SF, BP , (i) is not rejected (because $\hat{s} = 5$) and (ii) is always rejected (because $\hat{s} = 1$); hence the hypothesis that each of those currency differentials is stationary is rejected using any estimator. The same conclusion is found for DK using the argmax estimators. On the contrary, with the sequential tests one does not reject the hypothesis for DK because both (i) and (ii) are not rejected ($\hat{s} = 5$ in (i) and $\hat{s} = 0$ in (ii)). Given these overall results, also the DK currency differential with the Euro is considered non-stationary in the rest of the analysis.

It is next investigated whether the cointegrating relation can be expressed as a linear combination of the five currency differentials with the Euro in the 6-dimensional EU system. This is an hypothesis of type (6), with $B = a_\perp = (e_{DK} - e_{Euro}, e_{NK} - e_{Euro}, e_{SK} - e_{Euro}, e_{SF} - e_{Euro}, e_{BP} - e_{Euro})$, $a = B_\perp = I$, $q = 1$, and it is tested by checking whether (i) $a'X_t$ (the sum of the currencies in the EU system) contains $q = 1$ stochastic trends and (ii) $a'_\perp X_t$ (the five currency differentials with the Euro) contains $s - q = 5 - 1 = 4$ stochastic trends, see Theorem 4.1.

The results are reported in the first row (all differentials) of Table 4: (i) is not rejected (because $\hat{s} = 1$) and (ii) is not rejected (because $\hat{s} = 4$) using any estimator; hence one concludes that the cointegrating relation can be expressed as a linear combination of the five currency differentials with the Euro.

It is finally investigated whether the cointegrating relation can be expressed as a linear combination of only the DK and the NK currency differentials, as found in FGP. This is an hypothesis of type (6), with $B = a_\perp = (e_{DK} - e_{Euro}, e_{NK} - e_{Euro})$, $a = B_\perp =$

TABLE 4 | Analysis of the six EU currencies: Estimates of s in (i) $a'X_t$ and (ii) $a'_\perp X_t$ with the argmax estimators and with the sequential tests at 5% significance level.

Estimate of s in (i), (ii)	Max-gap	Alternative argmax	$K\pi^2\ \tau^{(i)}\ _\infty$	$K\pi^2\ \tau^{(i)}\ _1$
All differentials	1, 4	1, 4	1, 4	1, 4
DK, NK differentials	4, 1	4, 2	4, 1	4, 1

$(e_{DK} + e_{Euro} + e_{NK}, e_{SK}, e_{SF}, e_{BP})$, $q = 4$, and it is tested by checking whether (i) $a'X_t$ (the sum of DK, NK and the Euro, and the other currencies) contains $q = 4$ stochastic trends and (ii) $a'_\perp X_t$ (the DK and the NK differentials with the Euro) contains $s - q = 5 - 4 = 1$ stochastic trends, see Theorem 4.1.

The results in the second row (DK, NK differentials) of Table 4 show that, using the max-gap estimator and the sequential tests, one concludes that the cointegrating relation can be expressed as a linear combination of only the DK and the NK currency differentials; indeed, (i) is not rejected (because $\hat{s} = 4$) and (ii) is not rejected (because $\hat{s} = 1$). On the contrary, with the alternative argmax estimator the hypothesis is rejected because $\hat{s} = 2$ in (ii). Given these overall results, one concludes that the cointegrating relation can be expressed as a linear combination of only the DK and the NK currency differentials.

9 | Conclusions

This paper proposes a unified semiparametric framework for inference on hypotheses on the cointegrating space and attractor space for $I(1)/I(0)$ processes; this includes VARIMA and dynamic factor model processes.

The equivalence between hypotheses on the attractor space and the number of common stochastic trends in specific subsystems is used together with dimensional coherence of the semiparametric approach to define selection criteria and sequences of tests of hypotheses. In the special one-dimensional case, the limit distribution of interest is found analytically.

It is shown that the selection criteria provide consistent decisions on the hypotheses of interest while test sequences correctly select the null and alternative with limit probabilities that depend on the significance level used in the sequence. The properties of the estimators and tests are compared with alternatives, both theoretically and with a Monte Carlo experiment, and results are illustrated on an empirical application to exchange rates. The tools appear useful for discovering properties of the cointegrating space and the attractor space in systems with moderately large cross-sectional and time series dimension.

Acknowledgments

The views expressed in the paper are those of the authors and do not necessarily reflect the opinion of the institutions of affiliation. The authors thank the associate editor and two referees for their valuable comments and suggestions that helped improve the paper. Thanks also go to Sophocles Mavroeidis, Anindya Banerjee, Peter Boswijk, David Hendry, Søren Johansen, and participants in the 'Workshop to celebrate 40 years of Unit

Roots and Cointegration' (7–9 April 2025, Oxford) and the 'Eleventh Italian Congress of Econometrics and Empirical Economics' (29–31 May 2025, Palermo). Massimo Franchi gratefully acknowledges partial financial support from PRIN 20223725WE 'Methodological and computational issues in large-scale time series models for economics and finance.'

Funding

This work was partially supported by the Ministero dell'Università e della Ricerca (Grant No. PRIN 20223725WE 'Methodological and computational issues in large-scale time series models for economics and finance').

Endnotes

¹Of course, other choices of b are also possible, such as for instance $b = (v_2 - v_1, v_3 - v_2)$.

²The i.i.d. assumption can be relaxed to a martingale difference sequence; this is not done here for simplicity.

³In the following, the convention is that empty matrices are absent, for example $\psi = a$ when $q = s$ in (6).

⁴The notation $\text{cca}(y_t, d_t)$ is a shorthand for $\text{cca}(\{y_t\}_{t=1}^T, \{d_t\}_{t=1}^T)$; similar abbreviations are used in the following.

⁵Note that these critical values enter Definition A.1 in the Appendix.

⁶Definition A.1 in the Appendix reports the precise definition of the test sequence.

⁷ $F_\zeta(0) = 0$.

⁸MATLAB and R code to compute the p.d.f., the c.d.f and the quantile functions in Theorem 6.2 are available at the authors' personal websites.

References

1. R. Engle and C. Granger, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55 (1987): 251–276.
2. J. E. H. Davidson, D. F. Hendry, F. Srba, and S. Yeo, "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal* 88, no. 352 (1978): 661–692.
3. S. Johansen, "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12 (1988): 231–254.
4. S. Johansen, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59 (1991): 1551–1580.
5. S. Johansen and K. Juselius, "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the Ppp and the Uip for UK," *Journal of Econometrics* 53, no. 1 (1992): 211–244.
6. S. Johansen, *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models* (Oxford University Press, 1996).
7. K. Juselius, *The co-Integrated VAR Model: Econometric Methodology and Macroeconomic Applications* (Oxford University Press, 2006).
8. J. Gonzalo and C. Granger, "Estimation of Common Long-Memory Components in Cointegrated Systems," *Journal of Business & Economic Statistics* 13, no. 1 (1995): 27–35.
9. A. Onatski and C. Wang, "Alternative Asymptotics for Cointegration Tests in Large VARs," *Econometrica* 86 (2018): 1465–1478.
10. A. Onatski and C. Wang, "Extreme Canonical Correlations and High-Dimensional Cointegration Analysis," *Journal of Econometrics* 212 (2019): 307–322.
11. A. Bykhovskaya and V. Gorin, "Cointegration in Large VARs," *Annals of Statistics* 50 (2022): 1593–1617.
12. A. Bykhovskaya and V. Gorin, "Asymptotics of Cointegration Tests for High-Dimensional Var (k)," *Review of Economics and Statistics*, Early view (2024): 1–38.
13. C. Liang and M. Schienle, "Determination of Vector Error Correction Models in High Dimensions," *Journal of Econometrics* 208 (2019): 418–441.
14. H. Bierens, "Nonparametric Cointegration Analysis," *Journal of Econometrics* 77 (1997): 379–404.
15. X. Cheng and P. C. B. Phillips, "Semiparametric Cointegrating Rank Selection," *Econometric Journal* 12, no. s1 (2009): S83–S104.
16. M. Barigozzi and L. Trapani, "Testing for Common Trends in Nonstationary Large Datasets," *Journal of Business & Economic Statistics* 40 (2022): 1107–1122.
17. M. Barigozzi, G. Cavaliere, and L. Trapani, "Inference in Heavy-Tailed Nonstationary Multivariate Time Series," *Journal of the American Statistical Association* 119, no. 545 (2024): 565–581.
18. M. Franchi, I. Georgiev, and P. Paruolo, "Estimating the Number of Common Trends in Large Tand N Factor Models via Canonical Correlations Analysis," *Econometrics and Statistics*, Early view (2023).
19. M. Franchi, I. Georgiev, and P. Paruolo, "Canonical Correlation Analysis Of Stochastic Trends via Functional Approximation, 2025", Working Paper arXiv: 2411.19572.
20. U. Müller and M. Watson, "Testing Models of Low-Frequency Variability," *Econometrica* 76 (2008): 979–1016.
21. U. Müller and M. Watson, "Low-Frequency Econometrics," in *Advances in Economics and Econometrics: Eleventh World Congress, Econometric Society Monographs*, ed. B. Honoré, A. Pakes, M. Piazzesi, and L. E. Samuelson (Cambridge University Press, 2017), 53–94.
22. U. Müller and M. Watson, "Long-Run Covariability," *Econometrica* 86 (2018): 775–804.
23. S. Chen and M. Schienle, "Large Spillover Networks of Nonstationary Systems," *Journal of Business & Economic Statistics* 42 (2024): 422–436.
24. P. C. B. Phillips and V. Solo, "Asymptotics for Linear Processes," *Annals of Statistics* 20 (1992): 971–1001.
25. S. Johansen and K. Juselius, "An Asymptotic Invariance Property of the Common Trends Under Linear Transformations of the Data," *Journal of Econometrics* 178 (2014): 310–315.
26. J. Jacod and A. Shiryaev, *Limit Theorems for Stochastic Processes* (Springer, 2003).
27. R. G. King, C. I. Plosser, J. H. Stock, and M. W. Watson, "Stochastic Trends and Economic Fluctuations," *American Economic Review* 81, no. 4 (1991): 819–840.
28. G. Elliott, "Efficient Tests for a Unit Root When the Initial Observation is Drawn From Its Unconditional Distribution," *International Economic Review* 40 (1999): 767–783.
29. J. Stock and M. Watson, "Testing for Common Trends," *Journal of the American Statistical Society* 83 (1988): 1097–1107.
30. J. Bai, "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics* 122 (2004): 137–183.
31. S. Ahn and A. Horenstein, "Eigenvalue Ratio Test for the Number of Factors," *Econometrica* 81 (2013): 1203–1227.
32. R. Cameron and W. Martin, "Transformations for Wiener Integrals Under a General Class of Linear Transformations," *Transactions of the American Mathematical Society* 58 (1945): 184–219.
33. K. M. Abadir, "On the Asymptotic Power of Unit Root Tests," *Econometric Theory* 9, no. 2 (1993): 189–221.
34. S. T. Jensen and B. Nielsen, "The Joint Laplace Transform of a Quadratic Function and a Non Symmetric Function of a Brownian

Motion, 1995", Working paper, Institute of Mathematical Statistics, University of Copenhagen.

35. J. A. Doornik, B. Nielsen, and T. J. Rothenberg, "The Influence of Var Dimensions on Estimator Biases: Comment," *Econometrica* 71, no. 1 (2003): 377–383.

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

37. K. M. Abadir and P. Paruolo, "Two Mixed Normal Densities From Cointegration Analysis," *Econometrica* 65, no. 3 (1997): 671–680.

Appendix A

Proofs

This Appendix contains analytical results and proofs. Some well-known results on the functional central limit theorem and the L^2 -representation of Brownian motion are first reported.

By the functional central limit theorem, see Phillips and Solo [24], the partial sums $T^{-\frac{1}{2}} \sum_{i=1}^t \varepsilon_i$ in the CT representation (3) 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 $t = [Tu] \in \mathbb{N}$ and $u \in [0, 1]$, one has

$$T^{-\frac{1}{2}} \begin{pmatrix} \bar{\psi}' X_t \\ \beta' \sum_{i=1}^t X_i \end{pmatrix} = T^{-\frac{1}{2}} \begin{pmatrix} \kappa' \\ \beta' C_1 \end{pmatrix} \sum_{i=1}^{[Tu]} \varepsilon_i + o_p(1) \\ \xrightarrow[T \rightarrow \infty]{w} \begin{pmatrix} W_1(u) \\ W_2(u) \end{pmatrix} := DW_\varepsilon(u), \quad (A1)$$

where the $o_p(1)$ -term is infinitesimal uniformly in $u \in [0, 1]$, $\xrightarrow{w}$ 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 nonsingular long-run variance $\Omega := D\Omega_\varepsilon D'$, $D := (\kappa, C_1' \beta)'$. Indeed,

$$\text{rank } D = \text{rank} \begin{pmatrix} \kappa' \\ \beta' C_1 \end{pmatrix} (\bar{\kappa}, \kappa_\perp) \\ = \text{rank} \begin{pmatrix} I_s & 0 \\ \beta' C_1 \bar{\kappa} & \beta' C_1 \kappa_\perp \end{pmatrix} \\ = s + \text{rank } \beta' C_1 \kappa_\perp, \quad (A2)$$

where $\beta = \psi_\perp$ and $\psi_\perp' C_1 \kappa_\perp$ is full rank under Assumption 3.1 (iii). It is useful to partition Ω conformably with $W(u)$ as $\Omega = (\Omega_{ij})_{i,j=1,2}$ and define the standardised version of $W_1(u)$ as $B_1(u) := \Omega_{11}^{-\frac{1}{2}} W_1(u)$.

Recall, see for example, Phillips [36], that any n -dimensional standard Brownian motion $B(u)$ admits the representation

$$B(u) \simeq \sum_{k=1}^{\infty} c_k \phi_k(u), \\ c_k := \int_0^1 B(u) \phi_k(u) du = v_k \xi_k, \\ v_k \in \mathbb{R}_+, \quad \xi_k \sim N(0, Q), \quad (A3)$$

where $Q = I_n$, $\{\phi_k(u)\}_{k=1}^{\infty}$ is an orthonormal basis of $L^2[0, 1]$ and $\simeq$ indicates that the series in (A3) is a.s. convergent in the L^2 sense to the l.h.s.

In the special case where $(v_k^2, \phi_k(u))$, $k = 1, 2, \dots$, is an eigenvalue-eigenvector pair of the covariance kernel of the standard Brownian motion, that is, $v_k^2 \phi_k(u) = \int_0^1 \min(u, v) \phi_k(v) dv$, (A3) is the Karhunen-Loève (KL) representation of $B(u)$, for which one has

$$v_k = \frac{1}{(k - \frac{1}{2})\pi}, \quad \phi_k(u) = \sqrt{2} \sin(u/v_k), \quad \xi_k \text{ i.i.d. } N(0, Q). \quad (A4)$$

In this case the series in (A3) 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 (A4). It follows that the L^2 -representation of $W(u)$ in (A1) is given by (A3) and (A4) with $Q = \Omega$ and $\simeq$ replaced by $=$ when the KL basis is employed.

Next, proofs are reported.

Proof of Proposition 2.1. (a). Aggregation invariance requires $s \geq 1$ (because $X_t \sim I(1)$) and $\text{col}(v_1, \dots, v_p) \subseteq \text{col } \beta$ with $v_i := w - e_i$, $i \in \mathcal{I}$. Recall also that $(\text{col } \beta)^\perp = \text{col } \psi$.

Consider the square $p \times p$ matrices $V = (v_1, \dots, v_p) = (w - e_1, \dots, w - e_p) = wI - I_p$, $B = (w, e_1 - e_p, \dots, e_{p-1} - e_p)$, $C := (e_1, \dots, e_{p-1})'$. Observe that C is nonsingular because it is a rearrangement of a triangular matrix with ones on the diagonal. Similarly CB is nonsingular because it is a triangular matrix with ones on the diagonal; this implies that B is full rank as well. Next consider $VB = (0, e_1 - e_p, \dots, e_{p-1} - e_p)$ and hence $\text{rank } V = \text{rank}(VB) = p - 1$ and $\text{col}(v_1, \dots, v_p) = \text{col}(e_1 - e_p, \dots, e_{p-1} - e_p)$. Finally $(\text{col}(e_1 - e_p, \dots, e_{p-1} - e_p))^\perp = \text{col } i$. One hence has $\text{col}(v_1, \dots, v_p) \subseteq \text{col } \beta$ if and only if $\text{col } \psi \subseteq \text{col } i$. Because $s \geq 1$, one has $\text{col } \psi = \text{col } i$ and one can choose $\psi = i$.

(b). Rearrange X_t so as to list regions in $\mathcal{I}_1$ first, corresponding to the first m elements of X_t . Apply part (a) to $(X_{1t}, \dots, X_{mt})'$, and note that the stochastic trends representation for the whole X_t is obtained by padding the loading matrix ψ with zeros for the remaining variables X_{it} , $i = m + 1, \dots, p$. $\square$

Proof of Theorem 4.1. The statement of the theorem is rewritten as follows for better readability:

- (i). $\psi = A\theta$ with $\text{rank } \theta = s$ if and only if (ii.1) $\text{rank}(A'\psi) = s$ and (ii.2) $\text{rank}(A_\perp'\psi) = 0$;
- (iii). $\psi = (a, a_\perp \varphi)$ with $\text{rank } \varphi = s - q$ if and only if (iv.1) $\text{rank}(a'\psi) = q$ and (iv.2) $\text{rank}(a_\perp'\psi) = s - q$.

Consider each step separately.

(i) implies (ii.1) and (ii.2): Pre-multiply $\psi = A\theta$ by $(\bar{A}, \bar{A}_\perp)'$ and obtain $s = \text{rank}(\bar{A}, \bar{A}_\perp)' \psi = (\theta', \theta')'$ which implies (ii.1) and (ii.2).

(ii.1) and (ii.2) imply (i): By orthogonal projections $\psi = A\bar{A}'\psi + A_\perp \bar{A}_\perp'\psi$; next substitute from (ii.1) and (ii.2) to find $\psi = A\theta$ with $\theta := \bar{A}'\psi$ of rank s by (ii.1).

(iii) implies (iv.1) and (iv.2): Pre-multiply $\psi = (a, a_\perp \varphi)$ by $(\bar{a}, \bar{a}_\perp)'$ and obtain $\bar{a}'\psi = (I_q, 0)$ of rank q and $\bar{a}_\perp'\psi = \varphi$ of rank $s - q$.

(iv.1) and (iv.2) imply (iii): Because $\text{rank}(a'\psi) = q$ one can rotate a so that $\bar{a}'\psi = (I_q, 0)$. By orthogonal projections $\psi = a\bar{a}'\psi + a_\perp \bar{a}_\perp'\psi$; next substitute from (iv.1) and (iv.2) to find $\psi = (a, a_\perp \varphi)$ with $\varphi := \bar{a}_\perp'\psi$ of rank $s - q$.

This completes the proof. $\square$

Proof of Theorem 4.2. Note that $\psi \in \Psi_{01}^c$ means $\text{rank}(H'\psi) =: n - i < n$ for some integer $i > 0$, and that $\psi \in \Psi_{02}^c$ means $\text{rank}(H_\perp'\psi) = s - n$. Note that these joint statements would contradict (12) because this would imply $s \leq n - i + s - n = s - i$, with $i > 0$. This proves (i) $\Psi_{01}^c \cap \Psi_{02}^c = \emptyset$. In turn this implies $\Psi_{01}^c \cap \Psi_{02}^c = \Psi_{01}^c$, that is, $\Psi_{01}^c \subset \Psi_{02}^c$, and hence (ii) $\Psi_0^c = \Psi_{02}^c$, see Figure 1. Finally one finds decomposition (iii) using (i) and (ii), see Figure 1. $\square$

The next definition provides the precise description of the selection procedure $\hat{s}$ based on test sequences using $\text{cca}(X_t, d_t)$, see Section 6.1.

Definition A.1 (Estimators for s based on tests). The test-based estimator of s at significance level η is defined as

$$\{\hat{s}(X_t) = j\} := \begin{cases} \mathcal{A}_{p,h,\eta} & j = p \\ (\cap_{i=j+1}^p \mathcal{R}_{i,h,\eta}) \cap \mathcal{A}_{j,h,\eta} & j = 1, \dots, p-1, \\ \cap_{i=1}^p \mathcal{R}_{i,h,\eta} & j = 0 \end{cases} \\ \mathcal{R}_{i,h,\eta} := \{K\pi^2 \|\tau^{(i)}\|_h > c_{i,h,\eta}\},$$

$$\mathcal{A}_{i,h,\eta} := \mathcal{R}_{i,h,\eta}^c = \{K\pi^2 \|\tau^{(i)}\|_h \leq c_{i,h,\eta}\},$$

$$\tau^{(i)} := (1 - \lambda_i, 1 - \lambda_{i-1}, \dots, 1 - \lambda_1)', \quad (\text{A5})$$

where $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ are the eigenvalues of $\text{cca}(X_t, d_t)$, $\|\cdot\|_h$ is the h -norm for vectors, and $c_{i,h,\eta}$ are appropriate quantiles discussed prior to Corollary 6.1. The corresponding estimator of r is defined as $\hat{r}_h := p - \hat{s}_h$.

Proof of Corollary 6.1.

- (i). Assume $s < k$, and observe that $\mathbb{P}((\cap_{i=k+1}^p \mathcal{R}_{i,h,\eta}) \cap \mathcal{A}_{k,h,\eta}) \rightarrow 0$ in (A5); this implies $\mathbb{P}(\hat{s}_h(X_t) = k) \rightarrow 0$ for $k > s$.
- (ii). For $0 < s \leq p$ one has $\mathbb{P}((\cap_{i=s+1}^p \mathcal{R}_{i,h,\eta}) \cap \mathcal{A}_{s,h,\eta}) \rightarrow 1 - \eta$ in (A5); this implies $\mathbb{P}(\hat{s}_h(X_t) = s) \rightarrow 1 - \eta$ for $0 < s \leq p$. If $s = 0$ one has $\mathbb{P}(\cap_{i=1}^p \mathcal{R}_{i,h,\eta}) \rightarrow 1$ and hence $\mathbb{P}(\hat{s}_h(X_t) = s) \rightarrow 1$ for $s = 0$.
- (iii). Assume $s > k$; one has $\mathbb{P}(\hat{s}_h(X_t) = k) \leq \mathbb{P}(\hat{s}_h(X_t) \in \{s - 1, \dots, 1, 0\}) =: \rho_T \rightarrow \eta$ for $k < s$. $\square$

Proof of Theorem 6.2. Let $S = \int_0^1 B(u)^2 du$. By (10) in Abadir and Paruolo [37], 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 $S = g^{-1}(\zeta) = 1/\zeta$, $|\text{d}g^{-1}(\zeta)/\text{d}\zeta| = \zeta^{-2}$, and hence, for $z > 0$

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

Note also that

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

where $\Gamma(k, t) = \int_t^\infty z^{k-1} e^{-z} dz$ is the upper incomplete Gamma function and the following identity has been used, setting $y = az = h(z)$, $z = h^{-1}(y) = y/a$, $dz = a^{-1} dy$:

$$\begin{aligned} \int_t^\infty z^{k-1} e^{-az} dz &= \int_{at}^\infty \left(\frac{y}{a}\right)^{k-1} e^{-y} a^{-1} dy \\ &= a^{-k} \int_{at}^\infty y^{k-1} e^{-y} dy \\ &= a^{-k} \Gamma(k, at). \end{aligned}$$

$\square$

Proof of Theorem 6.3. Observe that

$$\mathbb{P}(z = 1) \geq 1 - \mathbb{P}(w = 0) - \mathbb{P}(v = 0) := \rho_{1T}, \quad (\text{A6})$$

$$\mathbb{P}(z = 0) \geq \max(\mathbb{P}(w = 0), \mathbb{P}(v = 0)) := \rho_{2T}. \quad (\text{A7})$$

To see (A6), define $\mathcal{B} := \{w = 1\}$, $\mathcal{C} := \{v = 1\}$ (and hence $\mathcal{B}^c = \{w = 0\}$, $\mathcal{C}^c = \{v = 0\}$) and note that $\mathbb{P}(z = 1) = 1 - \mathbb{P}(z = 0)$ where $\{z = 0\} = (\mathcal{B} \cap \mathcal{C})^c = \mathcal{B}^c \cup \mathcal{C}^c$ so that $\mathbb{P}(z = 0) \leq \mathbb{P}(\mathcal{B}^c) + \mathbb{P}(\mathcal{C}^c)$ by Boole's inequality, which gives (A6). For (A7), observe that $\mathbb{P}(z = 0) = \mathbb{P}(\mathcal{B}^c \cup \mathcal{C}^c)$, from which (A7) follows by additivity. In the following, Equation (A6) is used when $\psi \in \Psi_0$, and Equation (A7) is employed when $\psi \in \Psi_0^c$.

Let $\hat{s}$ be the max-gap or the alternative argmax criterion. If $\psi \in \Psi_0$, then $\mathbb{P}(\mathcal{B}) \rightarrow 1$ and $\mathbb{P}(\mathcal{C}) \rightarrow 1$ by (21); this proves (23) using (A6). If $\psi \in \Psi_0^c$, then $\mathbb{P}(\mathcal{B}^c) \rightarrow 1$ and $\mathbb{P}(\mathcal{C}^c) \rightarrow 1$ by (21) and Theorem 4.2 (i); this shows $\mathbb{P}(z = 0) \rightarrow 1$ by (A7). If $\psi \in \Psi_{01} \cap \Psi_{02}^c$ then $\mathbb{P}(\mathcal{B}) \rightarrow 1$ and $\mathbb{P}(\mathcal{C}^c) \rightarrow 1$; this shows $\mathbb{P}(z = 0) \rightarrow 1$ by (A7).

Consider now the case of $\hat{s}$ equal the $\hat{s}_h$ estimator based on test sequences, $h = 1, \infty$. If $\psi \in \Psi_0$, then $\mathbb{P}(\mathcal{B}^c) \rightarrow \nu$ and $\mathbb{P}(\mathcal{C}^c) \rightarrow \zeta$ by Corollary 6.1 (ii); this proves (24) using (A6). Note that $\zeta = 0$ when $H = A$ because $s - n = 0$, and the test sequence has limit probability of correct selection equal to one when the number of stochastic trends is zero; in case $H = a$, ζ equals the size of the tests η because $s - q > 0$ (unless $q = s$, in which case hypothesis $H = a$ coincides with $H = A$).

Next assume $\psi \in \Psi_0^c = \Psi_{02}^c$, with $\text{rank}(H'_\perp \psi) = j > s - n$. Observe that $\mathbb{P}(v = 0) \geq \rho_{3T} := \mathbb{P}(\hat{s}_h(H'_\perp X_t) = j) \rightarrow 1 - \eta$ by Corollary 6.1 (ii). Next separate two cases, $\psi \in \Psi_{01} \cap \Psi_{02}^c$ and $\psi \in \Psi_{01}^c$; if $\psi \in \Psi_{01} \cap \Psi_{02}^c$ then $\mathbb{P}(w = 0) \rightarrow \nu$ by Corollary 6.1 (ii) and this, together with $\mathbb{P}(v = 0) \geq \rho_{3T} \rightarrow 1 - \eta$, implies $\mathbb{P}(z = 0) \geq \rho_{2T} \rightarrow 1 - \eta$ by (A7). Instead if $\psi \in \Psi_{01}^c$ with $\text{rank}(H'_\perp \psi) = k < n$, by Corollary 6.1 (i) one finds $\mathbb{P}(w = 0) \rightarrow 1$ and hence $\mathbb{P}(z = 0) \rightarrow 1$ by (A7). This completes the proof. $\square$