Econometric Theory, **0**, 2026, 1–51.
doi:10.1017/S0266466626100462

INFERENCE ON COMMON TRENDS IN FUNCTIONAL TIME SERIES

MORTEN ØRREGAARD NIELSEN
Aarhus University

WON-KI SEO
University of Sydney

DAKYUNG SEONG
University of Sydney

We study statistical inference on unit roots and cointegration for time series in a Hilbert space. We develop statistical inference on the number of common stochastic trends embedded in the time series, that is, the dimension of the nonstationary subspace. We also consider tests of hypotheses on the nonstationary and stationary subspaces themselves. The Hilbert space can be of an arbitrarily large dimension, and our methods remain asymptotically valid even when the time series of interest takes values in a subspace of possibly unknown dimension. This has wide applicability in practice; for example, to cointegrated vector time series that are either high-dimensional or of finite dimension, to high-dimensional factor models that include a finite number of nonstationary factors, to cointegrated curve-valued (or function-valued) time series, and to nonstationary dynamic functional factor models. To illustrate our methods, we include two empirical examples.

1. INTRODUCTION

We consider statistical inference on unit roots and common stochastic trends for time series taking values in a Hilbert space of arbitrary dimension or a subspace of possibly unknown dimension. An important first step in the analysis of such time series is the determination of the dimension of the nonstationary subspace, that is, the subspace in which the time series behaves like a unit root process (see Section 2). This dimension is the number of common stochastic trends. Our objectives are to test hypotheses on the dimension of the nonstationary subspace as well as hypotheses on the stationary and nonstationary subspaces themselves.

We are grateful to the co-editor, Yixiao Sun, two anonymous referees, Brendan Beare, James Duffy, Massimo Franchi, Bent Nielsen, Hanlin Shang, and seminar participants at Australian National University, University of Oxford, University of Queensland, University of Sydney, University of Zurich, University of Amsterdam, University of Warwick, Rome Econometrics, as well as the 2022 NBER-NSF Time Series Conference, SETA 2023, ESAM 2023, EcoSta 2023, IAER 2024, IAAE 2024, and ESCW 2025 for helpful comments. A replication package for R is available at https://github.com/sdkseong/Inference-on-common-trends-in-functional-time-series Address correspondence to Morten Ørregaard Nielsen, Aarhus Center for Econometrics, Aarhus University, Denmark, email: mon@econ.au.dk.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

Since we want to examine nonstationary time series in a possibly unknown-dimensional subspace of a Hilbert space $\mathcal{H}$, our tests need to be statistically valid regardless of (i) whether the dimension of the space in which the time series takes values is finite or not and (ii) whether the dimension is known in advance or not. That is, the tests need to be (at least asymptotically) invariant to the dimensionality of the time series, and due to this property they will be called asymptotically dimension invariant (ADI) tests.

This is especially relevant in the recent literature on functional time series, where it is both empirically and theoretically supported that nonstationarity tends to be driven by a finite-dimensional process (e.g., Chang, Kim, and Park, 2016; Beare and Seo, 2020; Franchi and Paruolo, 2020; Li, Robinson, and Shang, 2020, 2023; Nielsen, Seo, and Seong, 2023). Even when a function-valued random element can be accommodated in a finite-dimensional space, and thus can be represented by a finite number of basis functions, its dimension is generally unknown and large. In such cases, most existing cointegration rank tests are not applicable since they require either finite (preferably small) dimensionality or a priori information on the dimensionality of the time series; this is true even for recently developed cointegration tests for high-dimensional time series (e.g., Onatski and Wang, 2018; Bykhovskaya and Gorin, 2022, 2024).

Our testing procedures have wide applicability in practice. For example, they can be used with (i) cointegrated vector time series of finite dimension (e.g., Stock and Watson, 1988; Johansen, 1991), (ii) high-dimensional factor models with a finite number of nonstationary factors (e.g., Nelson and Siegel, 1987; Peña and Poncela, 2004, 2006), (iii) cointegrated curve-valued (or function-valued) time series (e.g., Chang et al., 2016; Nielsen et al., 2023), and (iv) nonstationary dynamic functional factor models (e.g., Martínez-Hernández, Gonzalo, and González-Farías, 2022).

Our approach to developing ADI tests utilizes a dimension-reduction method that projects a high-dimensional or functional time series onto a fixed K-dimensional subspace. This projection is constructed such that, asymptotically, the K-dimensional projected component includes all stochastic trends, while the residual component (of unknown or possibly infinite dimension) contains no information regarding these trends. By constructing test statistics solely from the projected series, we effectively bypass interference from the residual component, ensuring that the resulting inference is inherently ADI. The analysis proceeds by examining generalized eigenvalues of two variance operators associated with the projected time series. Thus, the proposed tests will be classified as variance ratio (VR)-type tests, and some are generalizations of existing tests (Remark 3.4 and Section S.2 of the Supplementary Material). VR tests have desirable properties for the study of nonstationary time series. First, they can avoid estimation of the long-run variance, which implies a consistency property not shared by other tests (Müller, 2007, 2008). Second, the limiting behavior of the tests does not depend on a parametric assumption, such as a vector autoregression. As shown in the finite-dimensional case, finite-sample properties of parametric cointegration rank tests,

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

such as those proposed by Stock and Watson (1988), Ahn and Reinsel (1990), Johansen (1991), Bewley and Yang (1995), and Ahn (1997), depend crucially on the model specification (see Toda, 1995; Haug, 1996; Bewley and Yang, 1998). Third, many parametric assumptions, for example, autoregressive structures, are not preserved under projection, and tests relying on such assumptions cannot generally be ADI (Remark 3.1). In these regards, nonparametric VR tests are appealing.

Our theoretical results can be summarized by the following four points. First, we provide limit theory for general VR-type statistics based on partial summation and/or differencing, and this limit theory is applied to obtain our ADI inferential methods. Second, we apply our ADI tests sequentially to determine the dimension of the nonstationary subspace via either a top-down, bottom-up, or hybrid approach. The bottom-up does not require the choice of an initial hypothesis, but the top-down has better finite-sample properties, while the hybrid combines these advantages. Third, we consider also a direct estimator of the dimension of the nonstationary subspace based on ratios of generalized eigenvalues. This is similar to estimation of the dimension of the "dominant subspace" in Li et al. (2020, 2023) and may be considered complementary to their estimator for curve-valued time series and to those in Zhang and Chan (2018), Zhang, Robinson, and Yao (2019), and Franchi, Georgiev, and Paruolo (2026) for vector-valued time series. Fourth, we consider hypothesis testing on the nonstationary or stationary subspaces.

1.1. Comparisons with Existing Work

First, we view the time series of interest, $\{X_t\}_{t\geq 1}$, as a sequence in a known Hilbert space $\mathcal{H}$, but allow for the possibility that $\{X_t\}_{t\geq 1}$ takes values only in a subspace of $\mathcal{H}$. In the latter case, the variance operator of X_t is singular on $\mathcal{H}$ and may allow only finitely many nonzero eigenvalues even in an infinite-dimensional setting. This violates the common assumption in the literature of having sufficiently many (often infinitely many) nonzero eigenvalues. Of course, the eigenstructure of such a population variance operator is not known in practice, and assuming that infinitely many eigenvalues are nonzero may not only be unrealistic but is also not testable. For example, each functional observation of interest could be constructed by only a finite number of discrete and regularly spaced points on a grid of the domain, and such empirical examples can easily be found in the literature (see, e.g., Nielsen et al., 2023 (age-specific employment rates), Li et al., 2023 (U.S. treasury yield curves, in the working paper version), and Section 7.1). This poses an obstacle to the use of existing methods, and this is precisely where our ADI inference methods have a distinctive advantage over existing ones. As will be detailed, our methods are designed to be asymptotically valid under a more general scenario where the (long-run) variance operator of the stationary part of X_t permits only a few nonzero eigenvalues (we also provide analysis on the exact number of nonzero eigenvalues required for the proposed tests). In comparison, existing methods rule out this possibility by assumption (including our own recent work, Nielsen et al., 2023).

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

Thus, our methodology can accommodate more general and realistic functional time series.

Second, the common requirement of infinitely many nonzero eigenvalues for existing methods distinguishes them from cointegration rank tests developed in a finite-dimensional setting, where any variance operator necessarily permits only finitely many nonzero eigenvalues. We show a natural connection between our ADI tests and well-known tests for cointegration (or stationarity) developed in a finite-dimensional setting, and we find that our ADI tests generalize those tests. Furthermore, we demonstrate that, in specific cases, our ADI tests reduce to some recently developed tests for nonstationary functional time series. These results imply that some well-known tests, developed in both conventional and functional setups, can be understood as special cases of our ADI tests, thereby bridging the gap between them.

Third, the advantages of our methods lie not only in their applicability to a more general setting but also in their practical usefulness for real datasets. Many existing methods require a reasonable conjecture on the maximum possible number of stochastic trends (e.g., Chang et al., 2016; Li et al., 2023; Nielsen et al., 2023). Although some rule-of-thumb choices can be used in practice, such as those based on the scree plot, a more formal procedure may be preferable. The current study complements this aspect by providing a testing procedure based on a generalized version of the functional KPSS test (Horváth, Kokoszka, and Rice, 2014) that can be used in practice to construct an upper bound.

Fourth, we provide statistical inference on the space spanned by stochastic trends based on our ADI tests. Specifically, we can test whether a particular subspace of interest is contained in the nonstationary subspace, spans the nonstationary subspace, or is contained in the stationary subspace. We illustrate the empirical usefulness of these tests in Section 7.2 in the context of the well-known Nelson and Siegel (1987) model.

Compared with our earlier work in Nielsen et al. (2023), all four preceding points apply (see Section S.2.1.2 of the Supplementary Material). Also, in Nielsen et al. (2023), we only considered a particular variance ratio statistic (VR(2,1) under more restrictive assumptions than in Section 3.1 below), whereas in this article, we consider a family of statistics as well as eigenvalue ratio estimators.

Compared with recent work on cointegration testing in high-dimensional vector autoregressive (VAR) processes (Onatski and Wang, 2018; Bykhovskaya and Gorin, 2022, 2024), our setup is fundamentally different. They allow only a finite number of cointegrating vectors, whereas our setup requires that the number of stochastic trends is finite and hence is more suitable when nonstationarity is driven by a small number of factors. Interestingly, under the assumption of compact autoregressive operators widely adopted in the literature, a function-valued VAR implies that the number of stochastic trends is finite (Remark 2.3). Moreover, our setup does not rely on a VAR structure, and it covers intrinsically infinite-dimensional time series which theirs does not. Finally, our tests can be naturally applied to determine the number of stochastic trends while their tests are for the

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

existence of cointegration and cannot immediately be applied as cointegration rank tests. In general, therefore, their tests serve a different purpose and may be viewed as complementary (see also Remark 2.2).

In concurrent work, Li, Li, and Phillips (2025) investigated common stochastic trends within a more complex framework involving a growing number of curve-valued time series. Such data, often referred to as high-dimensional functional time series (e.g., Tang et al., 2025), can be understood as an integration of high-dimensional time series analysis and functional data analysis. They proposed an approximate factor model for such data and developed inferential methods, including a consistent estimator for the low-dimensional stochastic trends that drive the nonstationarity of the high-dimensional functional series. While the present setup does not accommodate their framework, the exploration of a possible extension of our methodology to their setting may be an interesting avenue for future research.

The remainder is organized as follows. Section 2 introduces I(1) time series in Hilbert space, and Section 3 presents our ADI VR tests. Section 4 discusses determination of the dimension of the nonstationary subspace via sequential testing or via eigenvalue ratio estimation. In Section 5, we discuss testing hypotheses about the subspaces. Section 6 presents Monte Carlo simulations, and Section 7 presents two empirical applications to the term structure of interest rates and to labor market indices. Finally, Section 8 concludes. The appendix contains mathematical notation, details, and proofs, and the Supplementary Material includes a practical implementation guide, additional discussion, simulations, and proofs.

2. I(1) TIME SERIES AND STOCHASTIC TRENDS IN HILBERT SPACE

We consider a cointegrated linear I(1) process X_t taking values in a separable Hilbert space, $\mathcal{H}$, equipped with an inner product $\langle v_1, v_2 \rangle$ and the induced norm $\|v_1\| = \langle v_1, v_1 \rangle^{1/2}$ for $v_1, v_2 \in \mathcal{H}$, as defined by Beare, Seo, and Seo (2017). For stationary time series in $\mathcal{H}$, the real-valued sequence $\langle X_t, v \rangle$ must be stationary for all possible choices of $v \in \mathcal{H}$ (see Proposition 3.1 and its proof in Beare et al., 2017). For example, if $\mathcal{H}$ is the standard $L^2[a,b]$ Hilbert space, and hence $X_t(r)$ is a function defined on $[a,b]$, then $\langle X_t, v \rangle = \int_a^b X_t(r)v(r)dr$ may be viewed as a continuous linear combination of $X_t(r)$. For many empirical applications involving economic functional time series, stationarity is too restrictive (e.g., Chang et al., 2016; Seo and Beare, 2019; Nielsen et al., 2023). Thus, we consider a sequence of X_t that allows $\langle X_t, v \rangle$ to be either I(1) nonstationary or I(0) stationary. It is then important to characterize under what choices of $v \in \mathcal{H}$ that $\langle X_t, v \rangle$ becomes stationary or nonstationary, which leads to notions of stationary and nonstationary subspaces. We next discuss this in more detail.

Let $\{X_t\}_{t \geq 1}$ be a sequence whose first difference, denoted ΔX_t, satisfies

$$\Delta X_t = \sum_{j=0}^{\infty} \Phi_j \epsilon_{t-j}, \quad t \geq 1, \tag{2.1}$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is an independent and identically distributed (iid) sequence with $\mathbb{E}[\epsilon_t] = 0$, $\mathbb{E}[\|\epsilon_t\|^4] < \infty$, and positive definite variance C_ϵ. We further assume that $\{\Phi_j\}_{j\geq 0}$ is a sequence of bounded, linear operators satisfying

$$\sum_{j=0}^{\infty} j^2 \|\Phi_j\|_{\text{op}} < \infty \quad \text{and} \quad \Phi(1) = \sum_{j=0}^{\infty} \Phi_j \neq 0, \tag{2.2}$$

where $\|\cdot\|_{\text{op}}$ denotes the operator norm. Then the long-run variance of ΔX_t is well defined as $\Lambda_{\Delta X} = \Phi(1) C_\epsilon \Phi(1)^* \neq 0$ (see Beare et al., 2017). Any stationary sequence with nonzero long-run variance is I(0). Hence, because $\{\Delta X_t\}_{t\geq 1}$ satisfying (2.1) and (2.2) is necessarily stationary and $\Lambda_{\Delta X} \neq 0$, it is I(0) by construction, so that $\{X_t\}_{t\geq 1}$ is I(1). The summability condition in (2.2) is common in the unit root literature (Phillips and Solo, 1992).

Let $\mathcal{H}_{\mathsf{S}} = \ker \Lambda_{\Delta X}$ denote the kernel of $\Lambda_{\Delta X}$ and let $\mathcal{H}_{\mathsf{N}} = \mathcal{H}_{\mathsf{S}}^{\perp}$ be its orthogonal complement. We call $\mathcal{H}_{\mathsf{S}}$ and $\mathcal{H}_{\mathsf{N}}$ the stationary (cointegrating) subspace and nonstationary (attractor) subspace, respectively. These names are related to some distinctive properties possessed by elements in those spaces.

First, it follows from Beare et al. (2017, Remark 3.1) that $\langle X_t, v\rangle$ is stationary for $v \in \mathcal{H}_{\mathsf{S}}$ and nonstationary for $v \notin \mathcal{H}_{\mathsf{S}} = \mathcal{H}_{\mathsf{N}}^{\perp}$. In fact, we will allow X_t to take values only in a strict subspace of unknown dimension. For all v that are in the orthogonal complement of the strict subspace it holds that $\langle X_t, v\rangle = 0$, and hence also $\langle \Delta X_t, v\rangle = 0$ with long-run variance $\langle v, \Lambda_{\Delta X} v\rangle = 0$. Because $\Lambda_{\Delta X}$ is self-adjoint and nonnegative, this implies that $v \in \ker \Lambda_{\Delta X}$ and thus $v \in \mathcal{H}_{\mathsf{S}}$. In general, it therefore follows that if $v \in \mathcal{H}_{\mathsf{S}}$, then $\{\langle X_t, v\rangle\}_{t\geq 1}$ is either a stationary random sequence or equal to zero almost surely (in which case it is also stationary).

Second, regarding $\mathcal{H}_{\mathsf{N}}$, we will assume throughout that

$$\dim(\mathcal{H}_{\mathsf{N}}) = \mathbb{d}_{\mathsf{N}} < \infty. \tag{2.3}$$

If (2.3) is true, we say that $\{X_t\}_{t\geq 1}$ contains $\mathbb{d}_{\mathsf{N}}$ stochastic trends. To see this in detail, we note that $\{X_t\}_{t\geq 1}$ satisfying (2.1) and (2.2) allows the decomposition

$$X_t = \Phi(1) \sum_{s=1}^{t} \epsilon_s + \nu_t + X_0 - \nu_0, \quad t \geq 1, \tag{2.4}$$

where $\nu_t = \sum_{j=0}^{\infty} \tilde{\Phi}_j \epsilon_{t-j}$ and $\tilde{\Phi}_j = -\sum_{s=j+1}^{\infty} \Phi_s$ (see Phillips and Solo, 1992). From the summability condition in (2.2), we also find that $\sum_{j=0}^{\infty} j \|\tilde{\Phi}_j\|_{\text{op}} < \infty$. It is known (Beare et al., 2017, Prop. 3.2) that $\mathcal{H}_{\mathsf{N}}$ is the closure of $\operatorname{ran} \Phi(1)$. Thus, under (2.4), we have

$$\mathcal{H} = \mathcal{H}_{\mathsf{S}} \oplus \mathcal{H}_{\mathsf{N}}, \quad \mathcal{H}_{\mathsf{S}} = \{v \in \mathcal{H} : \Phi(1) C_\epsilon \Phi(1)^* v = 0\}, \quad \mathcal{H}_{\mathsf{N}} = \{\Phi(1) v : v \in \mathcal{H}\}.$$

Since the random walk component in (2.4) clearly takes values in $\mathcal{H}_{\mathsf{N}}$ and $\dim(\mathcal{H}_{\mathsf{N}}) = \mathbb{d}_{\mathsf{N}}$, we say that $\{X_t\}_{t\geq 1}$ contains $\mathbb{d}_{\mathsf{N}}$ stochastic trends.

Since $\dim(\mathcal{H}_{\mathsf{N}}) = \dim(\operatorname{ran} \Phi(1)) = \operatorname{rank}(\Phi(1))$, the projection onto $\mathcal{H}_{\mathsf{N}}$, denoted P_{N}, is a finite-rank operator of rank $\mathbb{d}_{\mathsf{N}}$ under (2.3), regardless of

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

whether $\dim(\mathcal{H})$ is finite or not. However, the rank of the projection onto $\mathcal{H}_{\mathsf{S}}$, denoted $P_{\mathsf{S}} = I - P_{\mathsf{N}}$ with I denoting the identity operator on $\mathcal{H}$, depends on $\dim(\mathcal{H})$. Specifically, the rank is $p - \mathbb{d}_{\mathsf{N}}$ if $\mathcal{H} = \mathbb{R}^p$, while it is ∞ if $\mathcal{H}$ is infinite-dimensional. Importantly, even in the latter case, if X_t only takes values in a strict subspace of dimension p_X, then the (long-run) variance of $\{P_{\mathsf{S}}X_t\}_{t\geq 1}$ is not injective on $\mathcal{H}_{\mathsf{S}}$ and allows at most $p_X - \mathbb{d}_{\mathsf{N}}$ nonzero eigenvalues, which may be only finitely many.

Remark 2.1. The representation in (2.1) and (2.4) is that of a typical cointegrated vector- or curve-valued time series. However, it is also closely related to nonstationary vector- or curve-valued dynamic factor models, where X_t allows the representation

$$X_t = \sum_{j=1}^{\mathbb{d}_{\mathsf{N}}} \beta_{j,t}^{\mathsf{N}} f_j^{\mathsf{N}} + \sum_{j=1}^{\mathbb{d}_{\mathsf{S}}} \beta_{j,t}^{\mathsf{S}} f_j^{\mathsf{S}} + e_t, \quad \operatorname{span}\{f_j^{\mathsf{N}}\}_{j=1}^{\mathbb{d}_{\mathsf{N}}} \subseteq \mathcal{H}, \quad \operatorname{span}\{f_j^{\mathsf{S}}\}_{j=1}^{\mathbb{d}_{\mathsf{S}}} \subseteq \mathcal{H}. \tag{2.5}$$

Here, $\mathbb{d}_{\mathsf{N}}$ and $\mathbb{d}_{\mathsf{S}}$ are nonnegative integers that can be infinite if $\dim(\mathcal{H}) = \infty$, $\{\beta_{j,t}^{\mathsf{N}}\}_{j\geq 1}$ is a real-valued I(1) sequence, $\{\beta_{j,t}^{\mathsf{S}}\}_{j\geq 1}$ is a real-valued stationary sequence, and $\{e_t\}_{t\geq 1}$ is a $\mathcal{H}$-valued centered iid sequence with bounded variance operator. In fact, $\{X_t\}_{t\geq 1}$ satisfying (2.1) with compact, self-adjoint Φ_j can always be written as (2.5) (Martínez-Hernández et al., 2022). If $\mathbb{d}_{\mathsf{N}} < \infty$, $\mathbb{d}_{\mathsf{S}} < \infty$, and e_t is a curve-valued random element, (2.5) is the functional nonstationary dynamic factor model considered by Martínez-Hernández et al. (2022), and in that case, X_t is a nonstationary time series in $\mathcal{H}$, but $\langle X_t, v\rangle$ is stationary if and only if v is orthogonal to every f_j^{N} (i.e., $\mathcal{H}_{\mathsf{N}} = \operatorname{span}\{f_j^{\mathsf{N}}\}_{j=1}^{\mathbb{d}_{\mathsf{N}}}$ and $\mathcal{H}_{\mathsf{S}}$ is its orthogonal complement). See also Peña and Poncela (2004, 2006) for related discussion in a Euclidean space setting.

Remark 2.2. In the context of a high-dimensional cointegrated time series, the existing literature considers vector time series of dimension $p \to \infty$, but $\mathbb{d}_{\mathsf{N}}/p \to 1$ as $p \to \infty$ and $T \to \infty$, while this article considers the case, where $\mathbb{d}_{\mathsf{N}}$ is fixed regardless of the dimension p. Specifically, Onatski and Wang (2018) and Bykhovskaya and Gorin (2022, 2024) studied tests for existence of cointegration in a high-dimensional setup, but not for cointegration rank or the number of stochastic trends, $\mathbb{d}_{\mathsf{N}}$. In contrast with our condition (2.3) that $\mathbb{d}_{\mathsf{N}} < \infty$, they assume that $\mathbb{d}_{\mathsf{N}}$ increases such that $(p - \mathbb{d}_{\mathsf{N}})/p \to 0$ and $\mathbb{d}_{\mathsf{N}}/p \to 1$; that is, the cointegration rank $(p - \mathbb{d}_{\mathsf{N}})$ is negligible relative to $\mathbb{d}_{\mathsf{N}}$ (e.g., (12) of Onatski and Wang, 2018 or Section 3.2.2 of Bykhovskaya and Gorin, 2022). Thus, their methods are not applicable in our setup and should be considered complementary to our methods since they apply in a different context. Despite this distinction, our methodology is in fact applicable to high-dimensional vector time series in some cases. For example, a vector-valued time series X_t of arbitrary dimension can be understood as a Hilbert-valued I(1) time series if the eigenvalues of the variance matrix of ΔX_t are square-summable (Bosq, 2000, Lemmas 1.3 and 1.4). Finally, our methodology can be applied to time series of intrinsically infinite or large unknown dimension, and this is not considered or explored in the above-mentioned papers.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

Remark 2.3. The condition in (2.3) seems reasonable in many empirical examples of functional time series (e.g., in Chang et al., 2016 and Nielsen et al., 2023). From a theoretical point of view, suppose that $\{X_t\}_{t\geq 1}$ is generated by the functional ARMA law of motion:

$$\Theta(L)X_t = \Psi(L)\varepsilon_t,$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an $\mathcal{H}$-valued white noise, $\Theta(z) = I - \sum_{j=1}^{q_1}\Theta_j z^j$, $\Psi(z) = I + \sum_{j=1}^{q_2}\Psi_j z^j$, and q_1, q_2 are allowed to be infinite. In this case, if $\Theta(z)$ is a Fredholm operator-valued function with a unit root, then it can be shown from the Granger–Johansen representation theorem in Hilbert space (Beare and Seo, 2020; Franchi and Paruolo, 2020; Seo, 2023) that (2.3) is always satisfied. A sufficient (but not necessary) condition for $\Theta(z)$ to be a Fredholm operator is that $\Theta_1, \ldots, \Theta_{q_1}$ are compact and $q_1 < \infty$. It is common to assume compactness of autoregressive operators in statistical analysis of the functional ARMA model, and thus finiteness of $\mathbb{d}_{\mathsf{N}}$ arises as a natural consequence. Of course, for vector-valued time series, where the dimension p is fixed, $\mathbb{d}_{\mathsf{N}}$ is inherently finite. However, in high-dimensional settings, where p grows with the sample size as discussed in Remark 2.2, the condition $\mathbb{d}_{\mathsf{N}} < \infty$ may be less natural, and there is no obvious counterpart to the assumption of compact autoregressive operators. Nonetheless, even in high-dimensional settings the condition (2.3) does encompass some relevant examples like factor structures (such as the nonstationary dynamic factor models discussed in Remark 2.1) where nonstationarity is driven by a limited number of factors.

3. ADI VARIANCE RATIO TESTS

We let $\mathcal{H}$ be a separable Hilbert space and let $\{X_t\}_{t\geq 1}$ take values in $\mathcal{H}$ or any subspace. $\mathcal{H}$ can be an infinite-dimensional space of square-integrable functions or sequences depending on whether X_t is a random function or a random vector, but we do not exclude the possibility that $\mathcal{H}$ is finite-dimensional. We assume that $\{X_t\}_{t\geq 1}$ satisfies the following assumption.

Assumption 1. $\{X_t\}_{t\geq 1}$ satisfies the conditions in Section 2; in particular (2.1)–(2.3).

A crucial input to estimation and inference on unit roots and cointegration (Chang et al., 2016; Nielsen et al., 2023; Seo, 2024) is $\mathbb{d}_{\mathsf{N}} = \dim(\mathcal{H}_{\mathsf{N}})$, that is, the number of (linearly independent) stochastic trends. To determine this quantity, we consider the testing problem

$$H_0 : \mathbb{d}_{\mathsf{N}} = \mathbb{d}_0 \quad \text{vs} \quad H_1 : \mathbb{d}_{\mathsf{N}} \leq \mathbb{d}_0 - 1, \tag{3.1}$$

for $\mathbb{d}_0 \in \mathbb{D}_0 = \{1, 2, \ldots, \mathbb{d}_{\max}\}$, where $\mathbb{d}_{\max} < \infty$ is an upper bound on the number of stochastic trends so that $\mathbb{D}_0$ is a finite set. In fact, $\mathbb{d}_{\max}$ can be reasonably chosen from data in such a way that $\mathbb{d}_{\max} \geq \mathbb{d}_{\mathsf{N}}$, but $\mathbb{d}_{\max}$ does not greatly exceed $\mathbb{d}_{\mathsf{N}}$ (Remark 3.13 and Section 4.1).

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

As mentioned above, our tests can accommodate the case, where $\{X_t\}_{t\geq 1}$ only takes values in a strict subspace of possibly unknown dimension p_X, and the tests do not require knowledge of p_X. For this reason, they are called ADI tests. However, it should be noted that this dimension invariance property does not mean that our tests are not affected by p_X at all. Rather, it means that we do not explicitly require *a priori* information on p_X in order to implement the tests.

Our approach to developing ADI tests is based on a dimension-reduction technique that preserves stochastic trends. We call this *slack extraction* of stochastic trends. To elaborate on this, suppose that $\{X_t\}_{t\geq 1}$ is an $\mathcal{H}$-valued cointegrated I(1) process as described in Section 2 and only takes values in a possibly strict subspace of dimension $p_X \leq p_{\mathcal{H}} = \dim(\mathcal{H}) \leq \infty$, where p_X may be much larger than $\mathbb{d}_{\mathsf{N}}$. In addition, assume that we can find a finite integer $\mathrm{K} \in [\mathbb{d}_{\mathsf{N}}, p_X]$ and an orthonormal set $\{f_j\}_{j=1}^{\mathrm{K}}$ in $\mathcal{H}$ such that $\operatorname{span}\{f_j\}_{j=1}^{\mathbb{d}_{\mathsf{N}}} = \mathcal{H}_{\mathsf{N}}$ and $\{\langle X_t, f_j\rangle\}_{t\geq 1}$ is a stationary process for $j = \mathbb{d}_{\mathsf{N}}+1, \ldots, \mathrm{K}$. Define $P_{\mathrm{K}} = \sum_{j=1}^{\mathrm{K}} f_j \otimes f_j$, where $\otimes$ denotes the tensor product and thus $f_j \otimes f_j(\cdot) = \langle f_j, \cdot\rangle f_j$; that is, P_{K} is the unique orthogonal projection onto $\mathcal{H}_{\mathrm{K}} = \operatorname{span}\{f_j\}_{j=1}^{\mathrm{K}}$. In this case, we have $\mathcal{H} = \mathcal{H}_{\mathrm{K}} \oplus \mathcal{H}_{\mathrm{K}}^{\perp}$ and $\mathcal{H}_{\mathrm{K}} = \mathcal{H}_{\mathsf{N}} \oplus \mathcal{H}_{\mathrm{K}}^{\mathsf{S}}$, where $\mathcal{H}_{\mathrm{K}}^{\mathsf{S}} = \operatorname{span}\{f_j\}_{j=\mathbb{d}_{\mathsf{N}}+1}^{\mathrm{K}}$. Then, from the fact that $\{\langle X_t, f_j\rangle\}_{t\geq 1}$ is an I (1) sequence for $j \leq \mathbb{d}_{\mathsf{N}}$, we may deduce that (i) the projected time series $\{P_{\mathrm{K}} X_t\}_{t\geq 1}$ contains $\mathbb{d}_{\mathsf{N}}$ stochastic trends and a $(\mathrm{K} - \mathbb{d}_{\mathsf{N}})$-dimensional stationary component, and (ii) the nonstationary subspace of $\{P_{\mathrm{K}} X_t\}_{t\geq 1}$ is the same as that of $\{X_t\}_{t\geq 1}$. Due to these properties, we call P_{K} a *slack extractor* of stochastic trends. In other words, we do not lose any information about stochastic trends by using the projected time series $\{P_{\mathrm{K}} X_t\}_{t\geq 1}$ in the statistical analysis and ignoring the residual time series $\{(I - P_{\mathrm{K}})X_t\}_{t\geq 1}$ (whose dimension generally depends on p_X). Therefore, a statistical test, which is not dependent on the residual part $\{(I - P_{\mathrm{K}})X_t\}_{t\geq 1}$ but is constructed only from $\{P_{\mathrm{K}} X_t\}_{t\geq 1}$, naturally becomes ADI. Based on this idea, one may readily develop ADI tests by, for example, extending existing tests developed in a finite-dimensional setting (because $\{P_{\mathrm{K}} X_t\}_{t\geq 1}$ itself may be viewed as a K-dimensional cointegrated time series, see Remark 3.2).

Of course, a slack extractor P_{K} satisfying the required conditions is not observable, but it can be replaced with a suitable estimator $\widehat{P}_{\mathrm{K}}$. For now, we impose some high-level conditions on $\widehat{P}_{\mathrm{K}}$ and its rank K (the latter is specified by the practitioner). Define the operator

$$\Lambda_{\mathsf{S},R} = \Lambda_{\mathsf{S},R}(a_R) = \mathbb{E}[P_{\mathsf{S}} X_t \otimes P_{\mathsf{S}} X_t] + 1_{\{a_R > 0\}} \sum_{s=1}^{\infty} (\mathbb{E}[P_{\mathsf{S}} X_t \otimes P_{\mathsf{S}} X_{t-s}] + \mathbb{E}[P_{\mathsf{S}} X_{t-s} \otimes P_{\mathsf{S}} X_t]), \tag{3.2}$$

where $1_{\{\cdot\}}$ is the indicator function, and a_R is a parameter that will be naturally specified (in Assumption 3 below) for each of the proposed tests. Note that $\Lambda_{\mathsf{S},R}$ is the long-run variance (if $a_R > 0$) or variance (if $a_R = 0$) of the stationary component of $\{X_t\}_{t\geq 1}$.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

Assumption 2. (i) $\mathrm{K} \geq \mathrm{d}_{\mathrm{N}}$, (ii) $\|\widehat{P}_{\mathrm{K}} P_{\mathrm{N}} - P_{\mathrm{N}}\|_{\mathrm{op}} = o_{\mathrm{P}}(1)$, and (iii) $\|\widehat{P}_{\mathrm{K}} P_{\mathrm{S}} - P_{\mathrm{S}}^{\mathrm{K}}\|_{\mathrm{op}} = o_{\mathrm{P}}(1)$ for some orthogonal projection $P_{\mathrm{S}}^{\mathrm{K}}$ satisfying $\operatorname{ran} P_{\mathrm{S}}^{\mathrm{K}} \subseteq \mathcal{H}_{\mathrm{S}}$ and $\operatorname{rank}(P_{\mathrm{S}}^{\mathrm{K}} \Lambda_{\mathrm{S},R} P_{\mathrm{S}}^{\mathrm{K}}) = \mathrm{K} - \mathrm{d}_{\mathrm{N}}$.

Assumption 2(i) implies that we need an appropriate choice of K depending on the value of d_{N}. As will be discussed in Section 4, see, in particular, Remarks 4.1 and 4.2 and the subsequent discussion, it is possible to obtain a reasonable upper bound $\mathrm{d}_{\max}$ for d_{N}. By choosing K so that $\mathrm{K} \geq \mathrm{d}_{\max}$, the condition on K required by Assumption 2(i) does not raise any issues in the practical implementation of our tests. For a given choice of K, feasible choices of $\widehat{P}_{\mathrm{K}}$ satisfying Assumption 2(ii) and (iii) are discussed in Section 3.4 under mild low-level conditions.

Remark 3.1. It seems reasonable to ask if one can apply standard (cointegration rank) tests of (3.1) from a Euclidean space setting to the projected time series $\{\widehat{P}_{\mathrm{K}} X_t\}_{t \geq 1}$. This is in fact not always the case, because some properties of the time series are not preserved under projection. In particular, parametric tests based on an AR(p) law of motion, such as those in Johansen (1991), cannot be directly applied since parametric relationships are not generally preserved under projection. To see this in detail, suppose that $\dim(\mathcal{H}) = p_{\mathcal{H}}$, but X_t follows an AR(1) law of motion in a subspace $\mathcal{H}_X$ of (unknown) dimension $p_X \leq p_{\mathcal{H}}$ such that $\Delta X_t = \Theta X_{t-1} + \epsilon_t$ for some bounded linear operator Θ and $t \geq 1$. Then, for $\mathrm{d}_{\mathrm{N}} \leq \mathrm{K} \leq p_X$, we have $\Delta \widehat{P}_{\mathrm{K}} X_t = \widehat{P}_{\mathrm{K}} \Theta \widehat{P}_{\mathrm{K}} X_{t-1} + \widehat{P}_{\mathrm{K}} \Theta (I - \widehat{P}_{\mathrm{K}}) X_{t-1} + \widehat{P}_{\mathrm{K}} \epsilon_t$. The term $\widehat{P}_{\mathrm{K}} \Theta (I - \widehat{P}_{\mathrm{K}}) X_{t-1}$ is dependent on p_X and nonzero in general. Thus, the parametric relationship between $\Delta \widehat{P}_{\mathrm{K}} X_t$ and $\widehat{P}_{\mathrm{K}} X_{t-1}$ is generally different from that between ΔX_t and X_{t-1}, and also dependent on K, $\widehat{P}_{\mathrm{K}}$, and p_X. Even if $\widehat{P}_{\mathrm{K}}$ is a consistent estimator of a slack extractor P_{K}, a parametric test applied to $\{\widehat{P}_{\mathrm{K}} X_t\}_{t \geq 1}$ not only depends on p_X, but is in fact misspecified. This point is illustrated in Section S.3 of the Supplementary Material, where we provide simulation evidence that Johansen's (1991) trace test does not generally work with a consistent estimator of a slack extractor.

A consequence of Remark 3.1 is that, even if we may have a reasonable parametric assumption such as an AR structure for the original time series $\{X_t\}_{t \geq 1}$, such an assumption is not generally preserved under the projection $\widehat{P}_{\mathrm{K}}$. Therefore, when we use the projected time series $\{\widehat{P}_{\mathrm{K}} X_t\}_{t \geq 1}$ for our statistical analysis to obtain ADI tests, these need to be developed without parametric assumptions. To achieve this goal, we consider tests based on various sample (long-run) variance operators that can be computed from the projected time series without any parametric assumptions.

Thus, define $X_{d,t}$ by

$$X_{0,t} = \Delta X_t, \qquad X_{1,t} = X_t, \qquad X_{d,t} = \sum_{s=1}^{t} X_{d-1,s} \text{ for } d \geq 2,$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

so that d denotes the integration order of $X_{d,t}$. Further define the unnormalized sample long-run variance operator of $\{X_{d,t}\}_{t\geq 1}$, denoted $\widehat{\Lambda}_d(h,\mathrm{k})$, as

$$\widehat{\Lambda}_d(h,\mathrm{k}) = \sum_{s=-T+1}^{T-1} \mathrm{k}\left(\frac{s}{h}\right)\widehat{\Gamma}_{d,s}, \qquad \widehat{\Gamma}_{d,s} = \begin{cases} \sum_{t=s+1}^{T} X_{d,t-s}\otimes X_{d,t}, & \text{if } s\geq 0, \\ \sum_{t=-s+1}^{T} X_{d,t}\otimes X_{d,t+s}, & \text{if } s<0, \end{cases} \tag{3.3}$$

where $\mathrm{k}(\cdot)$ is a kernel function and h is the associated bandwidth parameter. From (3.3) and the fact that $\widehat{P}_{\mathrm{K}}$ is self-adjoint, it is readily found that $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}$ is the unnormalized sample (long-run) variance operator of the projected time series $\{\widehat{P}_{\mathrm{K}}X_{d,t}\}_{t\geq 1}$. As a special case, $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(0,\mathrm{k})\widehat{P}_{\mathrm{K}}$ is the unnormalized sample variance operator of $\{\widehat{P}_{\mathrm{K}}X_{d,t}\}_{t\geq 1}$.

Remark 3.2. Throughout, we will use the well-known result that any K-dimensional subspace of $\mathcal{H}$ with orthonormal basis $\{g_j\}_{j=1}^{\mathrm{K}}$ is isomorphic to $\mathbb{R}^{\mathrm{K}}$ via the isomorphism $v\mapsto(\langle v,g_1\rangle,\ldots,\langle v,g_{\mathrm{K}}\rangle)'$ (e.g., Proposition 5.2 and Theorem 5.4 of Conway, 1990). Moreover, any linear operator A acting on the K-dimensional subspace can be represented as the $\mathrm{K}\times\mathrm{K}$ matrix $[A]_{ij}=\langle g_i,Ag_j\rangle$ (see also Nielsen et al., 2023, p. 32). Suppose $\widehat{P}_{\mathrm{K}}=\sum_{j=1}^{\mathrm{K}}\hat{f}_j\otimes\hat{f}_j$ for an orthonormal set $\{\hat{f}_j\}_{j=1}^{\mathrm{K}}$. Then $\widehat{P}_{\mathrm{K}}X_{d,t}=\sum_{j=1}^{\mathrm{K}}\langle X_{d,t},\hat{f}_j\rangle\hat{f}_j$ such that $\widehat{P}_{\mathrm{K}}X_{d,t}$ may be identified as the K-dimensional vector $x_{\mathrm{K},d,t}=(\langle X_{d,t},\hat{f}_1\rangle,\ldots,\langle X_{d,t},\hat{f}_{\mathrm{K}}\rangle)'$. From these results, we find that if $[\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}]$ is the $\mathrm{K}\times\mathrm{K}$ sample long-run variance matrix of $x_{\mathrm{K},d,t}$, then $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}$ can be found as $[\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}]$. That is, $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}$ can also be computed from the vector-valued time series $\{x_{\mathrm{K},d,t}\}_{t\geq 1}$. Finally, for any $v\in\mathcal{H}$, $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}v$ can be computed as $\sum_{j=1}^{\mathrm{K}}a_j\hat{f}_j$, where a_j is the j-th element of the K-dimensional vector $[\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_d(h,\mathrm{k})\widehat{P}_{\mathrm{K}}]v_{\mathrm{K}}$ with $v_{\mathrm{K}}=(\langle v,\hat{f}_1\rangle,\ldots,\langle v,\hat{f}_{\mathrm{K}}\rangle)'$. Additional details and notation are given in Appendix A and in Section S.1.2 of the Supplementary Material.

Our tests are based on the generalized eigenvalue problem

$$\mu_j\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}\widehat{P}_{\mathrm{K}}v_j=\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_R,R}\widehat{P}_{\mathrm{K}}v_j, \tag{3.4}$$

where $\widehat{\Lambda}_{d_m,m}=\widehat{\Lambda}_{d_m}(h_m,\mathrm{k}_m)$ for $m\in\{L,R\}$ (left and right) and choices of integration orders, d_L and d_R. Here, μ_j are the ordered (smallest to largest) eigenvalues of the operators $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}\widehat{P}_{\mathrm{K}}$ and $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_R,R}\widehat{P}_{\mathrm{K}}$, and v_j are the corresponding eigenvectors. These can easily be computed as detailed in Section S.1.2 of the Supplementary Material. Because $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_m,m}\widehat{P}_{\mathrm{K}}$ are sample (long-run) variance operators, we call (3.4) a generalized VR eigenvalue problem, and any test based on it is called a VR(d_L,d_R)-based test. We assume $d_L>d_R$ without loss of generality because $d_L<d_R$ is obtained by simply redefining μ_j as its inverse.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

In the sequel, $\mathrm{k}_m(\cdot)$ and h_m are assumed to satisfy the following assumption.

Assumption 3. $\mathrm{k}_m(\cdot)$ and h_m, for $m \in \{L, R\}$, satisfy:

(i) $\mathrm{k}_m(\cdot)$ is a twice continuously differentiable even function from $\mathbb{R}$ to $[-1, 1]$ such that $\mathrm{k}_m(x) = 0$ for $|x| \geq 1$, $\mathrm{k}_m(0) = 1$, $\mathrm{k}'_m(0) = 0$, $\mathrm{k}''_m(0) \neq 0$, and $\lim_{|x|\to 1} \mathrm{k}_m(x)/(1-|x|)^2 = \text{constant}$.

(ii) $h_m = [a_m T^{b_m}]$ for some $a_m \geq 0$ and $b_m \in (0, 1/2)$, where $[x]$ is the nearest integer to x.

Assumption 3(i) is adopted from Phillips (1995, Assump. KL). The requirements on $\mathrm{k}_m(\cdot)$ in Assumption 3(i) are not restrictive in practice and are satisfied by many widely used kernel functions, including the Epanechnikov, Parzen, Tukey-Hanning, and quartic kernels. The functional form of h_m depending on T in Assumption 3(ii) is quite standard in practice.

Remark 3.3. Our theory covers all pairs of (d_L, d_R) with $d_R \in \{0, 1\}$, but not $d_R \geq 2$. Our theoretical approach requires a normalizing diagonal operator D_T (acting on $\mathrm{ran}\widehat{P}_{\mathrm{K}}$) and a scaling factor m_T, both depending on T, such that $m_T D_T \widehat{P}_{\mathrm{K}} \widehat{\Lambda}_{d_R,R} \widehat{P}_{\mathrm{K}} D_T$ converges to a nonzero block-diagonal finite-rank operator (see (B.17) and (B.27)). For $d_R \geq 2$, however, this is not generally possible. As a simple illustration with $d_R = 2$, consider a bivariate I(1) dynamic system $z_t = (z_{\mathrm{N},t}, z_{\mathrm{S},t})'$, where $z_{\mathrm{N},t}$ is I(1) and $z_{\mathrm{S},t}$ is I(0), and define $\tilde{z}_t = \sum_{s=1}^{t} z_s$. Then the asymptotic orders of elements of $\widehat{C}_{2,R} = \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t'$ are $\begin{bmatrix} \mathcal{O}_{\mathrm{P}}(T^4) & \mathcal{O}_{\mathrm{P}}(T^3) \\ \mathcal{O}_{\mathrm{P}}(T^3) & \mathcal{O}_{\mathrm{P}}(T^2) \end{bmatrix}$, and they are generally exact. Thus, for $m_T D_T \widehat{C}_{2,R} D_T$ to have nonzero diagonal blocks, we must have $D_T = \begin{bmatrix} T^{-1} & 0 \\ 0 & 1 \end{bmatrix}$ and $m_T = T^{-2}$, but in this case off-diagonal elements are not generally negligible.

Remark 3.4. To motivate some choices of (d_L, d_R), consider a finite-dimensional time series $X_t = (x_{1,t}, \ldots, x_{\mathrm{K},t})'$ for fixed K. Then, for example, (3.4) with $(d_L, d_R) = (2, 1)$ is similar to the generalized eigenvalue problem studied in Breitung (2002) with $a_L = a_R = 0$, and (3.4) with $(d_L, d_R) = (1, 0)$ is similar to that studied in Shintani (2001). The locally best invariant tests studied in Kwiatkowski et al. (1992) and Nyblom and Harvey (2000) can also be seen as special cases of our inverse tests in Section 3.3. Intuitively, these tests are based on the fact that the d_{N} largest eigenvalues of (the finite-dimensional version of) (3.4) exhibit different limiting behavior compared to the remaining $\mathrm{K} - \mathrm{d}_{\mathrm{N}}$ ones when there are $\mathrm{d}_{\mathrm{N}} (\leq \mathrm{K})$ stochastic trends in X_t. In view of the isomorphism in Remark 3.2, it is reasonable to conjecture that similar results hold in a functional time series setup if we choose $\widehat{P}_{\mathrm{K}}$ appropriately. Additional motivation based on comparisons with existing tests is given in Section S.2 of the Supplementary Material.

3.1. Tests Based on VR(d_L,1) with $d_L \geq 2$

In this section, we consider tests based on the VR(d_L,1) eigenvalue problem in (3.4) with $d_L \geq 2$ and $d_R = 1$. We first obtain the asymptotic properties of the eigenvalues of this problem.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

To this end, we let $\{W_{1,q}(r)\}_{r\in[0,1]}$ denote a q-dimensional standard Brownian motion and recursively define the $(d-1)$-fold integrated Brownian motion $W_{d,q}(r) = \int_0^r W_{d-1,q}(u)du$ for $d \geq 2$. For any matrix or compact operator A, we let $\lambda_j\{A\}$ be the j-th smallest eigenvalue of A, that is, $\lambda_1\{A\} \leq \lambda_2\{A\} \leq \cdots$. Given a kernel k_m satisfying Assumption 3, we define

$$c_m = \int_{-1}^{1} \mathrm{k}_m(u)du = 2\int_0^1 \mathrm{k}_m(u)du. \tag{3.5}$$

Moreover, when there is no risk of confusion, we write $\int f$ to denote $\int_0^1 f(u)du$.

THEOREM 3.1. *Suppose that Assumptions 1–3 hold and define $\widetilde{\mu}_j = n_T\mu_j$, where $\{\mu_j\}_{j=1}^{\mathrm{K}}$ are the eigenvalues from (3.4) with $d_L \geq 2$, $d_R = 1$, and*

$$n_T = T^{2d_L-2}(h_L c_L + 1_{\{a_L=0\}})/(h_R c_R + 1_{\{a_R=0\}}), \tag{3.6}$$

where h_m, a_m, and c_m are given in Assumption 3(ii) and (3.5). Then

$$\widetilde{\mu}_j \xrightarrow{\mathrm{d}} \widetilde{\lambda}_j = \lambda_j\left\{\left(\int W_{d_L,\mathbb{d}_\mathsf{N}} W'_{d_L,\mathbb{d}_\mathsf{N}}\right)^{-1}\int W_{1,\mathbb{d}_\mathsf{N}} W'_{1,\mathbb{d}_\mathsf{N}}\right\}$$
$$\text{jointly for } j \leq \mathbb{d}_\mathsf{N} \ (\text{if } \mathbb{d}_\mathsf{N} \geq 1), \tag{3.7}$$

$$\widetilde{\mu}_j \xrightarrow{\mathrm{P}} \infty \quad \text{for } j \geq \mathbb{d}_\mathsf{N} + 1. \tag{3.8}$$

Note that the normalization factor (3.6) of the eigenvalues generally depends on the choice of bandwidth parameters h_m associated with the kernel $\mathrm{k}_m(\cdot)$ for $m \in \{L,R\}$. However, in the case, where $a_L = a_R = 0$ (i.e., $h_L = h_R = 0$), $\widehat{P}_\mathrm{K}\widehat{\Lambda}_{d_L,L}\widehat{P}_\mathrm{K}$ and $\widehat{P}_\mathrm{K}\widehat{\Lambda}_{1,R}\widehat{P}_\mathrm{K}$ reduce to the variances of $\{\widehat{P}_\mathrm{K}X_{d_L,t}\}_{t\geq 1}$ and $\{\widehat{P}_\mathrm{K}X_{1,t}\}_{t\geq 1}$, and the convergence rate is simply $n_T^{-1} = T^{2-2d_L}$.

Theorem 3.1 shows that the eigenvalues in (3.4) have distinct asymptotic properties depending on whether j is greater than the dimension of $\mathcal{H}_\mathsf{N}$ or not, as motivated in Remark 3.4. That is, if we consider the vector $(\widetilde{\mu}_1, \ldots, \widetilde{\mu}_\mathrm{K})$, the first $\mathbb{d}_\mathsf{N}$ elements converge jointly in distribution to the vector $(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{\mathbb{d}_\mathsf{N}})$, while the last $\mathrm{K} - \mathbb{d}_\mathsf{N}$ elements are divergent in probability. If $\mathrm{K} \geq \mathbb{d}_0$, for any continuous map $\mathcal{F} : \mathbb{R}^{\mathbb{d}_0} \to \mathbb{R}$, we have

$$\mathcal{F}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{d}} \mathcal{F}(\{\widetilde{\lambda}_j\}_{j=1}^{\mathbb{d}_0}) \quad \text{under } H_0 \text{ of (3.1)}.$$

If $\mathcal{F}$ additionally satisfies $\mathcal{F}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \to_\mathrm{P} \infty$ under H_1 of (3.1), we can consistently test the hypothesis of interest in an obvious way; hereafter such an $\mathcal{F}$ is called a proper *test functional*. Among many possible choices of proper test functionals, we focus on the two most common choices, $\mathcal{F}_{\max}$ and $\mathcal{F}_{\mathrm{trace}}$, which are defined by

$$\mathcal{F}_{\max}(\{x_j\}_{j=1}^{\mathbb{d}_0}) = \max_{1\leq j\leq \mathbb{d}_0}\{x_j\} \quad \text{and} \quad \mathcal{F}_{\mathrm{trace}}(\{x_j\}_{j=1}^{\mathbb{d}_0}) = \sum_{j=1}^{\mathbb{d}_0} x_j.$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

The following corollary delivers consistent tests of (3.1) based on VR(d_L,1).

COROLLARY 3.1. *Consider the setup of Theorem 3.1 with* $\mathrm{K} \geq \mathbb{d}_0$. *Under* $H_0 : \mathbb{d}_{\mathsf{N}} = \mathbb{d}_0$,

$$\mathcal{F}_{\max}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{d}} \max_{1\leq j\leq \mathbb{d}_0} \{\lambda_j\{\mathcal{A}\}\}, \quad \mathcal{F}_{\mathrm{trace}}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{d}} \sum_{j=1}^{\mathbb{d}_0} \lambda_j\{\mathcal{A}\}, \tag{3.9}$$

where $\mathcal{A} = (\int W_{d_L,\mathbb{d}_0} W'_{d_L,\mathbb{d}_0})^{-1} \int W_{1,\mathbb{d}_0} W'_{1,\mathbb{d}_0}$. *Under* $H_1 : \mathbb{d}_{\mathsf{N}} < \mathbb{d}_0$,

$$\mathcal{F}_{\max}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{P}} \infty, \quad \mathcal{F}_{\mathrm{trace}}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{P}} \infty. \tag{3.10}$$

The test statistics in Corollary 3.1 are only functions of $\{\widehat{P}_{\mathrm{K}} X_t\}_{t\geq 1}$, which may be understood as a K-dimensional vector-valued time series (Remark 3.2), and the associated limiting distributions are functionals of $\mathbb{d}_0$-dimensional standard Brownian motion and do not depend on p_X, K, or any nuisance parameters. These results imply that the proposed tests have the ADI property and that critical values depend only on $\mathbb{d}_0$ and the test functional.

Remark 3.5. The VR(d_L,1) tests allow $a_L = a_R = 0$, and hence they can be completely free from the choice of bandwidth parameters. This seems desirable in practice, since practitioners typically do not want test results to be dependent on the choice of bandwidth parameters.

3.2. Tests Based on VR(d_L,0) with $d_L \geq 1$

We now consider VR(d_L,0)-based tests, which are established from the eigenvalue problem (3.4) with $d_L \geq 1$ and $d_R = 0$. As with VR(d_L,1)-based tests, the tests in this section are based on the asymptotic properties of the sample operators, $\widehat{P}_{\mathrm{K}} \widehat{\Lambda}_{d_L,L} \widehat{P}_{\mathrm{K}}$ and $\widehat{P}_{\mathrm{K}} \widehat{\Lambda}_{0,R} \widehat{P}_{\mathrm{K}}$ for $d_L \geq 1$. The important difference relative to tests based on VR(d_L,1) is that VR(d_L,0)-based tests generally require $a_R > 0$; that is, $\widehat{\Lambda}_{0,R}$ must be the sample *long-run* variance of $X_{0,t}$. This is in contrast with the case in Section 3.1, where $a_R = 0$ can be chosen (see Remarks 3.5 and 3.6).

The asymptotic properties of the VR(d_L,0) eigenvalues with $d_L \geq 1$ are as follows.

THEOREM 3.2. *Suppose that Assumptions 1–3 hold with* $a_R > 0$ *and define* $\widetilde{\mu}_j = n_T \mu_j$, *where* $\{\mu_j\}_{j=1}^{\mathrm{K}}$ *are the eigenvalues from (3.4) with* $d_L \geq 1, d_R = 0$, *and* $n_T = T^{2d_L - 1}(h_L c_L + 1_{\{a_L = 0\}})$, *where* h_L, a_L, *and* c_L *are given in Assumption 3(ii) and (3.5). Then*

$$\widetilde{\mu}_j \xrightarrow{\mathrm{d}} \lambda_j \left\{ \left(\int W_{d_L,\mathbb{d}_{\mathsf{N}}} W'_{d_L,\mathbb{d}_{\mathsf{N}}} \right)^{-1} \right\} \quad \textit{jointly for } j \leq \mathbb{d}_{\mathsf{N}} \ (\textit{if } \mathbb{d}_{\mathsf{N}} \geq 1),$$

$$\widetilde{\mu}_j \xrightarrow{\mathrm{P}} \infty \quad \textit{for } j \geq \mathbb{d}_{\mathsf{N}} + 1.$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

The next corollary delivers consistent tests of (3.1) based on VR(d_L,0) with $d_L \geq 1$.

COROLLARY 3.2. *Consider the setup of Theorem 3.2 with* $\mathrm{K} \geq \mathbb{d}_0$. *Under* $H_0 : \mathbb{d}_{\mathsf{N}} = \mathbb{d}_0$,

$$\mathcal{F}_{\max}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{d}} \max_{1\leq j\leq \mathbb{d}_0} \{\lambda_j\{\mathcal{A}\}\}, \quad \mathcal{F}_{\mathrm{trace}}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{d}} \sum_{j=1}^{\mathbb{d}_0} \lambda_j\{\mathcal{A}\}, \tag{3.11}$$

where $\mathcal{A} = (\int W_{d_L,\mathbb{d}_0} W'_{d_L,\mathbb{d}_0})^{-1}$. *Under* $H_1 : \mathbb{d}_{\mathsf{N}} < \mathbb{d}_0$,

$$\mathcal{F}_{\max}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{P}} \infty, \quad \mathcal{F}_{\mathrm{trace}}(\{\widetilde{\mu}_j\}_{j=1}^{\mathbb{d}_0}) \xrightarrow{\mathrm{P}} \infty. \tag{3.12}$$

Remark 3.6. For the VR(d_L,0) tests, $a_L = 0$ is allowed, which may seem desirable as in Remark 3.5. However, $h_L > 0$ is supported by simulation results for the corresponding finite-dimensional VR(1,0) test in Shintani (2001) (see Section S.2.2.1 of the Supplementary Material), where it was found that using a long-run variance estimator on the left was superior to using a variance estimator. For nonzero h_L, we also note that a larger h_L helps certain asymptotically negligible quantities decay at faster rates (see, e.g., the paragraph following (B.20)). On the other hand, $a_R = 0$ is generally not allowed because $\widehat{\Lambda}_{0,R}$ needs to be a consistent estimator of the long-run variance of ΔX_t, which requires a nonzero bandwidth h_R. Given that ΔX_t is stationary, we may use existing data-dependent choices of h_R that achieve certain mean-square optimality properties (e.g., Rice and Shang, 2017). Since $\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{0,R}\widehat{P}_{\mathrm{K}}$ is the sample long-run variance of the projected time series $\{\widehat{P}_{\mathrm{K}}\Delta X_t\}_{t\geq 1}$, which may be understood as a K-dimensional vector-valued time series (Remark 3.2), we may also apply more conventional methods (e.g., Andrews, 1991). We follow the latter approach in our simulations in Section 6.

3.3. Inverse VR Tests

We showed in Theorem 3.1 that the first $\mathbb{d}_{\mathsf{N}}$ eigenvalues $\{\mu_j\}_{j=1}^{\mathbb{d}_{\mathsf{N}}}$ from the VR(d_L,1) problem with $d_L \geq 2$, properly normalized, converge jointly to a well-defined limit. As we now show, it is also possible to obtain the limiting behavior of the remaining $\mathrm{K} - \mathbb{d}_{\mathsf{N}}$ eigenvalues, $\{\mu_j\}_{j=\mathbb{d}_{\mathsf{N}}+1}^{\mathrm{K}}$. This enables us to examine the inverse testing problem,

$$H_0 : \mathbb{d}_{\mathsf{N}} = \mathbb{d}_0 \quad \text{vs} \quad H_1 : \mathbb{d}_{\mathsf{N}} \geq \mathbb{d}_0 + 1, \tag{3.13}$$

for some $\mathbb{d}_0 \geq 0$. Unlike in the testing problem (3.1), we do not require any prior information on a reasonable upper bound for $\mathbb{d}_{\mathsf{N}}$.

Related to testing (3.13), we will replace Assumption 2 with the following high-level assumption, which allows the possibility that $\mathrm{K} < \mathbb{d}_{\mathsf{N}}$.

Assumption 2'. If $\mathrm{K} \geq \mathbb{d}_{\mathsf{N}}$ then Assumption 2 holds. If $\mathrm{K} < \mathbb{d}_{\mathsf{N}}$ then $\|\widehat{P}_{\mathrm{K}} P_{\mathsf{S}}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1)$.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

THEOREM 3.3. *Suppose that Assumptions 1–3 hold with* $a_R > 0$ *and define* $\widetilde{\mu}_j = n_T \mu_j$, *where* $\{\mu_j\}_{j=1}^{\mathrm{K}}$ *are the eigenvalues from (3.4) with* $d_L \geq 2$, $d_R = 1$, *and* $n_T = T^{2d_L - 3}(h_L c_L + 1_{\{a_L = 0\}})$, *where* h_L, a_L, *and* c_L *are given in Assumption 3(ii) and (3.5). Then*

$$(\widetilde{\mu}_j)^{-1} \xrightarrow{\mathrm{d}} \lambda_{\mathrm{K}-j+1}\{\mathcal{A}\} \quad \textit{jointly for } j = \mathbb{d}_{\mathsf{N}} + 1, \ldots, \mathrm{K} \ (\textit{if } \mathbb{d}_{\mathsf{N}} < \mathrm{K}), \tag{3.14}$$

$$(\widetilde{\mu}_j)^{-1} \xrightarrow{\mathrm{P}} \infty \quad \textit{for } j \leq \mathbb{d}_{\mathsf{N}} \ (\textit{if } \mathbb{d}_{\mathsf{N}} \geq 1), \tag{3.15}$$

where $\mathcal{A} = \int B_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} B'_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} - 1_{\{\mathbb{d}_{\mathsf{N}} \geq 1\}} \int B_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} W'_{d_L, \mathbb{d}_{\mathsf{N}}} (\int W_{d_L, \mathbb{d}_{\mathsf{N}}} W'_{d_L, \mathbb{d}_{\mathsf{N}}})^{-1} \int W_{d_L, \mathbb{d}_{\mathsf{N}}} B'_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}}$ *and* $B_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}}$ *is a* $(\mathrm{K} - \mathbb{d}_{\mathsf{N}})$*-dimensional standard Brownian motion which is independent of* $W_{d_L, \mathbb{d}_{\mathsf{N}}}$.

Theorem 3.3 complements Theorem 3.1, where the asymptotic properties of the first $\mathbb{d}_{\mathsf{N}}$ eigenvalues are presented under similar assumptions. Specifically, the asymptotic results given in Theorem 3.3 are for $\mathbb{d}_{\mathsf{N}} \geq 0$ under the requirement $a_R > 0$, which differs from Theorem 3.1, where the limiting distributions are given under the existence of unit roots ensuring $\mathbb{d}_{\mathsf{N}} \geq 1$ but allowing $a_R = 0$. This means that Theorem 3.3 is not simply a byproduct of Theorem 3.1 that can be obtained with slight modifications, but in fact has its own theoretical and practical justification as we now illustrate.

When $\mathrm{K} > \mathbb{d}_0$, and given a proper test functional $\mathcal{F}$, the asymptotic results in Theorem 3.3 can be used to deliver consistent tests of (3.13) as follows.

COROLLARY 3.3. *Consider the setup of Theorem 3.3 with* $\mathrm{K} > \mathbb{d}_0$. *Under* $H_0 : \mathbb{d}_{\mathsf{N}} = \mathbb{d}_0$,

$$\mathcal{F}_{\max}(\{\widetilde{\mu}_j^{-1}\}_{j=\mathbb{d}_0+1}^{\mathrm{K}}) \xrightarrow{\mathrm{d}} \max_{1 \leq j \leq \mathrm{K}-\mathbb{d}_0} \{\lambda_j\{\mathcal{A}\}\}, \quad \mathcal{F}_{\mathrm{trace}}(\{\widetilde{\mu}_j^{-1}\}_{j=\mathbb{d}_0+1}^{\mathrm{K}}) \xrightarrow{\mathrm{d}} \sum_{j=1}^{\mathrm{K}-\mathbb{d}_0} \lambda_j\{\mathcal{A}\}, \tag{3.16}$$

where $\mathcal{A} = \int B_{\mathrm{K}-\mathbb{d}_0} B'_{\mathrm{K}-\mathbb{d}_0} - 1_{\{\mathbb{d}_0 \geq 1\}} \int B_{\mathrm{K}-\mathbb{d}_0} W'_{d_L, \mathbb{d}_0} (\int W_{d_L, \mathbb{d}_0} W'_{d_L, \mathbb{d}_0})^{-1} \int W_{d_L, \mathbb{d}_0} B'_{\mathrm{K}-\mathbb{d}_0}$. *Under* $H_1 : \mathbb{d}_{\mathsf{N}} > \mathbb{d}_0$,

$$\mathcal{F}_{\max}(\{\widetilde{\mu}_j^{-1}\}_{j=\mathbb{d}_0+1}^{\mathrm{K}}) \xrightarrow{\mathrm{P}} \infty, \quad \mathcal{F}_{\mathrm{trace}}(\{\widetilde{\mu}_j^{-1}\}_{j=\mathbb{d}_0+1}^{\mathrm{K}}) \xrightarrow{\mathrm{P}} \infty. \tag{3.17}$$

The test functionals described in Corollary 3.3 are based on the limiting behavior of the inverse eigenvalues from the VR(d_L,1) problem, so we call them inverse VR tests. Compared to Corollary 3.1 concerning the testing problem (3.1), we allow $\mathbb{d}_0 = 0$ in Corollary 3.3. This means that we can, for example, test the null of stationarity of the time series $\{X_t\}_{t \geq 1}$ against an alternative of unit root nonstationarity. Hence, this test generalizes some existing, and widely used, KPSS-type tests of stationarity that are essentially obtained when $\mathbb{d}_0 = 0$ (see Section S.2.3 of the Supplementary Material for a detailed discussion of related procedures).

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

Remark 3.7. Note that $a_R > 0$ is generally required in Theorem 3.3 (and Corollary 3.3), whereas $a_R = 0$ is allowed in Theorem 3.1 for the same VR(d_L,1) problem. This is because Theorem 3.3 requires $P_{\mathsf{S}}\widehat{\Lambda}_{1,R}P_{\mathsf{S}}$ to converge in probability in the sense of the operator norm to the true long-run variance of $\{P_{\mathsf{S}}X_t\}_{t\geq 1}$. Also, with $d_R = 1$, $X_{1,t}$ is nonstationary, such that the optimal bandwidths in Remark 3.6 do not appear immediately applicable.

3.4. Estimation of Slack Extractor

To implement the VR-based tests in practice, a slack extractor $\widehat{P}_{\mathrm{K}}$ satisfying either Assumption 2 or 2' is needed. As shown in Sections 3.1–3.3, both the VR and inverse VR tests require Assumption 2 to hold for $\mathrm{K} \geq \mathbb{d}_{\mathsf{N}}$, and when $\mathrm{K} < \mathbb{d}_{\mathsf{N}}$ the inverse VR tests require instead that Assumption 2' holds. In this section, we construct such estimators $\widehat{P}_{\mathrm{K}}$ from observations $\{X_t\}_{t\geq 1}$, and we propose low-level conditions on the eigenstructure of the operator $\Lambda_{\mathsf{S},R}$ defined in (3.2), under which the estimators satisfy the high-level conditions.

We let $\{\tau_j\}_{j\geq 1}$ be the eigenvalues of $\Lambda_{\mathsf{S},R}$, ordered from the largest to the smallest.

Assumption 4. If $\mathrm{K} > \mathbb{d}_{\mathsf{N}}$ then $\tau_j \neq 0$ for $j = 1, \ldots, \mathrm{K} - \mathbb{d}_{\mathsf{N}}$ and $\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} \neq \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}$.

Assumption 4 tells us that the allowable values of K, which is chosen by the practitioner, depend on the number of nonzero eigenvalues of $\Lambda_{\mathsf{S},R}$ defined in (3.2). Specifically, it requires that the first $\mathrm{K} - \mathbb{d}_{\mathsf{N}}$ eigenvalues of $\Lambda_{\mathsf{S},R}$ are nonzero and that the $(\mathrm{K} - \mathbb{d}_{\mathsf{N}})$-th eigenvalue is different from the next eigenvalue, $\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}$. The role of the requirement $\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} \neq \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}$ is subtle (and may be relaxed under some additional assumptions). It is employed to make the slack extractor $\widehat{P}_{\mathrm{K}}$ have a certain desirable property in our proof and can be checked using the eigenvalues of the sample counterpart of $\Lambda_{\mathsf{S},R}$ (see Assumption 5 and Theorem 3.5).

In a high-dimensional Hilbert space, the number of nonzero eigenvalues of $\Lambda_{\mathsf{S},R}$ is generally very large (or possibly infinite). Particularly, in a typical functional time series setting with $\dim(\mathcal{H}) = \infty$ (e.g., the case considered by Nielsen et al., 2023), $\Lambda_{\mathsf{S},R}$ generally has infinitely many nonzero eigenvalues, and Assumption 4 allows K to be any arbitrary finite integer as long as $\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} \neq \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}$. In a finite-dimensional space, a reasonable upper bound on $\mathbb{d}_{\mathsf{N}}$ will be relevant to choose K to satisfy Assumption 4.

Remark 3.8. Consider the simple choice $\mathrm{K} = \mathbb{d}_0 + m$ for $m \geq 0$. Then Assumption 4 becomes more stringent as either $\mathbb{d}_0$ or m increases. It is thus recommended that practitioners select a small m and a $\mathbb{d}_0$ that is not much greater than $\mathbb{d}_{\mathsf{N}}$. Since we can construct a reasonable upper bound on $\mathbb{d}_{\mathsf{N}}$ (Remark 3.13 and Section 4), Assumption 4 does not appear to impose significant restrictions on the use of the VR(d_L, d_R)-based tests in practice.

Remark 3.9. Even if the VR(d_L, d_R)-based tests allow choosing $\mathrm{K} = \mathbb{d}_0$, this is not recommended in practice. If $\mathrm{K} = \mathbb{d}_0$, Assumption 2 requires $\widehat{P}_{\mathrm{K}}$ to be an

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

accurate estimator of P_{N} in the sense that $\widehat{P}_{\mathrm{K}}$ is required to extract precisely the stochastic trends. If this fails, it tends to result in severe over-rejection. On the other hand, $\widehat{P}_{\mathrm{K}}$ for $\mathrm{K} > \mathbb{d}_0$ is less likely to miss relevant stochastic trends and will thus contribute to having correct size in finite samples. This issue is discussed in detail in Nielsen et al. (2023, e.g., Remark 13).

We now describe the construction of the estimator of the slack extractor. Let

$$\widehat{\Lambda}_{1,P} = \widehat{\Lambda}_{1,P}(a_R) = \widehat{\Gamma}_{1,0} + 1_{\{a_R>0\}} \sum_{s=1}^{T-1} \mathrm{k}_P(s/h_P)(\widehat{\Gamma}_{1,s} + \widehat{\Gamma}_{1,-s}) \tag{3.18}$$

denote the sample variance (resp. long-run variance) estimator of $X_{1,t}$ if $a_R = 0$ (resp. $a_R > 0$), but allowing a new choice of bandwidth, h_P, and kernel, $\mathrm{k}_P(\cdot)$. The slack extractor needs to satisfy a rank condition in Assumption 2 that depends on $\Lambda_{\mathrm{S},R}$, which is a variance when $a_R = 0$ and a long-run variance when $a_R > 0$. For this reason, we base our estimator of the slack extractor on $\widehat{\Lambda}_{1,P}$, which depends on a_R in the same way (see also Remark 3.11). The advantage is that the estimator $\widehat{P}_{\mathrm{K}}$ described in Theorem 3.4 is universal in the sense that it can be applied for any of the $\mathrm{VR}(d_L, d_R)$ tests that we consider.

THEOREM 3.4. *Suppose that Assumption 1 holds. If $a_R > 0$ then suppose also that $\mathrm{k}_P(\cdot)$ and h_P in (3.18) satisfy Assumption 3. If $\mathrm{K} > \mathbb{d}_{\mathrm{N}}$ then suppose also that Assumption 4 holds. Let $\{\hat{f}_j\}_{j=1}^{\mathrm{K}}$ be the eigenvectors corresponding to the* K *largest eigenvalues of $\widehat{\Lambda}_{1,P}$. Then $\widehat{P}_{\mathrm{K}} = \sum_{j=1}^{\mathrm{K}} \hat{f}_j \otimes \hat{f}_j$ satisfies Assumption 2 when $\mathrm{K} \geq \mathbb{d}_{\mathrm{N}}$ and satisfies Assumption 2' when $\mathrm{K} < \mathbb{d}_{\mathrm{N}}$.*

Note that practitioners can employ the usual functional principal component analysis (FPCA) in $\mathcal{H}$ to obtain the eigenvectors in Theorem 3.4. Thus, Theorem 3.4 provides an easy-to-implement way to obtain an estimated projection operator $\widehat{P}_{\mathrm{K}}$ that satisfies the high-level conditions in Assumptions 2 and 2'.

Remark 3.10. Of course, the $\widehat{P}_{\mathrm{K}}$ given in Theorem 3.4 is not the only choice that satisfies Assumptions 2 and 2'. In fact, alternative estimators from the literature satisfy Assumptions 2 and 2' under different (but stronger) low-level conditions. For example, regardless of a_R, Chang et al. (2016) proposed using $\widetilde{P}_{\mathrm{K}} = \sum_{j=1}^{\mathrm{K}} \hat{\phi}_j \otimes \hat{\phi}_j$, where $\{\hat{\phi}_j\}_{j\geq 1}$ are the eigenvectors of the sample variance of $\{X_t\}_{t\geq 1}$ (i.e., of $\widehat{\Lambda}_{1,P}(0)$), but that requires additional conditions. If (i) $\mathrm{K} \geq \mathbb{d}_{\mathrm{N}}$, (ii) the variance $\Lambda_{\mathrm{S},R}$ satisfies that $\langle v, \Lambda_{\mathrm{S},R} v\rangle \neq 0$ for every nonzero $v \in \operatorname{ran} P_{\mathrm{S}}$, and (iii) the $(\mathrm{K} - \mathbb{d}_{\mathrm{N}})$-th largest eigenvalue of $\mathbb{E}[P_{\mathrm{S}} X_t \otimes P_{\mathrm{S}} X_t] (= \Lambda_{\mathrm{S},R}(0))$ is distinct from the next one, then we may deduce from Chang et al. (2016, Thm. 3.3) and our proof of Theorem 3.4 that $\sum_{j=1}^{\mathbb{d}_{\mathrm{N}}} \hat{\phi}_j \otimes \hat{\phi}_j \to_{\mathrm{P}} P_{\mathrm{N}}$ and that $\sum_{j=\mathbb{d}_{\mathrm{N}}+1}^{\mathrm{K}} \hat{\phi}_j \otimes \hat{\phi}_j$ converges to a nonrandom projection, $P_{\mathrm{S}}^{\mathrm{K}}$, satisfying $\operatorname{ran} P_{\mathrm{S}}^{\mathrm{K}} \subseteq \operatorname{ran} P_{\mathrm{S}}$ and $\operatorname{rank}(P_{\mathrm{S}}^{\mathrm{K}} \Lambda_{\mathrm{S},R} P_{\mathrm{S}}^{\mathrm{K}}) = \mathrm{K} - \mathbb{d}_{\mathrm{N}}$. That is, Assumptions 2 and 2' hold for $\widetilde{P}_{\mathrm{K}}$. Condition (ii), in particular, may be restrictive in a functional time series setup, but it is necessary. Without it, $\operatorname{rank}(P_{\mathrm{S}}^{\mathrm{K}} \Lambda_{\mathrm{S},R} P_{\mathrm{S}}^{\mathrm{K}})$ could be smaller than $\mathrm{K} - \mathbb{d}_{\mathrm{N}}$, such that the I(0) components

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

have reduced rank (long-run) variance. This is a concern when its inverse is used to construct the test statistic as in, for example, Chang et al. (2016, Sect. 3.3); see also the discussion of Assumption 2 and (B.4) in Nielsen et al. (2023).

Remark 3.11. Our proposed $\widehat{P}_{\mathrm{K}}$ in Theorem 3.4 does not require the additional conditions in Remark 3.10, and it therefore seems preferable for practitioners. Specifically, because of the way $\widehat{\Lambda}_{1,P}$ depends on a_R, the projection $\sum_{j=1}^{\mathrm{d_N}} \hat{f}_j \otimes \hat{f}_j$ is asymptotically equivalent to $\widehat{P}_{\mathrm{K}} P_{\mathrm{N}}$ and converges to P_{N}, while $\sum_{j=\mathrm{d_N}+1}^{\mathrm{K}} \hat{f}_j \otimes \hat{f}_j$ is asymptotically equivalent to $\widehat{P}_{\mathrm{K}} P_{\mathrm{S}}$ and converges to a projection, $P_{\mathrm{S}}^{\mathrm{K}}$, onto the $(\mathrm{K}-\mathrm{d_N})$-dimensional eigensubspace of $\Lambda_{\mathrm{S},R}$ that is spanned by the eigenvectors corresponding to the leading nonzero eigenvalues. It then follows by definition of $\Lambda_{\mathrm{S},R}$ that the rank condition, $\mathrm{rank}(P_{\mathrm{S}}^{\mathrm{K}} \Lambda_{\mathrm{S},R} P_{\mathrm{S}}^{\mathrm{K}}) = \mathrm{K}-\mathrm{d_N}$, is satisfied. In other words, to avoid the additional conditions in Remark 3.10, the projection operator needs to be based on estimators of the eigenelements of $\Lambda_{\mathrm{S},R}$, which depends on a_R, and that is why we base our slack extractor on $\widehat{\Lambda}_{1,P}$, which similarly depends on a_R.

The condition $\tau_{\mathrm{K}-\mathrm{d_N}} \neq \tau_{\mathrm{K}-\mathrm{d_N}+1}$ in Assumption 4 is generally required for the VR(d_L, d_R)-based tests and also required for the inverse VR test when d_0 is close to $\mathrm{d_N}$. Practitioners might be concerned with potential violation of this condition for a chosen K. However, as shown in our proof of Theorem 3.4, the ordered eigenvalues $\{\widehat{\tau}_j\}_{j \geq \mathrm{d_N}+1}$ of $\widehat{\Lambda}_{1,P}$ converge in probability to $\{\tau_j\}_{j \geq 1}$. That is, the first $\mathrm{d_N}$ eigenvalues of $\widehat{\Lambda}_{1,P}$ are associated (at least asymptotically) with the nonstationary subspace, and therefore it is the eigenvalues $\widehat{\tau}_j$ for $j = \mathrm{d_N}+1, \ldots$ that correspond to eigenvalues τ_j for $j = 1, \ldots$ of $\Lambda_{\mathrm{S},R}$.

We may thus avoid violation of the condition $\tau_{\mathrm{K}-\mathrm{d_N}} \neq \tau_{\mathrm{K}-\mathrm{d_N}+1}$ by choosing K such that $\widehat{\tau}_{\mathrm{K}}$ is sufficiently greater than $\widehat{\tau}_{\mathrm{K}+1}$. This approach is pragmatic and also theoretically supported under the following assumption (which is required only when $a_R > 0$).

Assumption 5. (i) $\mathbb{E}[\|P_{\mathrm{S}} X_t\|^8] < \infty$, (ii) $\| \sum_{j=m}^{\infty} \tilde{\Phi}_j \|_{\mathrm{op}} = \mathcal{O}(m^{-\beta})$ for some $\beta > 4$, (iii) there exists $\varphi > 0$ such that $\lim_{x \to 0} (1 - \mathrm{k}_P(x))/|x|^{\varphi} < \infty$, and $\tilde{\varphi} > \varphi$ such that $\sum_{j=-\infty}^{\infty} |j|^{\tilde{\varphi}} \| \mathbb{E}[P_{\mathrm{S}} X_t \otimes P_{\mathrm{S}} X_{t-j}] \|_{\mathrm{op}} < \infty$, and (iv) $T/h_P^{2\varphi+1} = \mathcal{O}(1)$.

The conditions in Assumption 5 are stronger than those in the previous sections. In particular, $\{P_{\mathrm{S}} X_t\}_{t \geq 1}$ needs to satisfy a stronger moment condition than required for $\{X_t\}_{t \geq 1}$ in Section 2, and the coefficients $\{\Phi_j\}_{j \geq 1}$ are required to decay at a faster rate than in Section 2. The latter can be seen from the fact that $\| \sum_{j=m}^{\infty} \tilde{\Phi}_j \|_{\mathrm{op}} = \mathcal{O}(m^{-\gamma})$ for some $\gamma > 1$ implies the summability condition in (2.2). Assumption 5(iii) imposes a regularity condition on $\mathrm{k}_P(\cdot)$ and a summability condition on the autovariance operators of $\{P_{\mathrm{S}} X_t\}_{t \geq 1}$, and Assumption 5(iv) requires that h_P is not too small relative to T. Nonetheless, even if Assumption 5 includes some stronger conditions, we note that they are not essential for our proposed ADI tests but are only needed if practitioners want to check if $\tau_{\mathrm{K}-\mathrm{d_N}} \neq \tau_{\mathrm{K}-\mathrm{d_N}+1}$ based on the sample eigenvalues $\widehat{\tau}_j$. The following theorem establishes the desired result.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

THEOREM 3.5. *Suppose that Assumption 1 holds. If $a_R > 0$, then assume also that $\mathrm{k}_P(\cdot)$ and h_P in (3.18) satisfy Assumption 3 and that Assumption 5 holds. If $\mathrm{K} > \mathbb{d}_{\mathsf{N}}$, then*

$$(\widehat{\tau}_{\mathrm{K}} - \widehat{\tau}_{\mathrm{K}+1}) - (\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} - \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}) = \mathcal{O}_{\mathrm{P}}(M_T^{-1/2}),$$

where $M_T = T/h_P$ if $a_R > 0$ and $M_T = T$ if $a_R = 0$.

From Theorem 3.5, we deduce that, for any η_T satisfying $\eta_T M_T^{-1/2} \to 0$ and $\eta_T \to \infty$,

$$\eta_T(\widehat{\tau}_{\mathrm{K}} - \widehat{\tau}_{\mathrm{K}+1}) \xrightarrow{\mathrm{P}} 0 \qquad \text{if} \quad \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} = \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1},$$

$$\eta_T(\widehat{\tau}_{\mathrm{K}} - \widehat{\tau}_{\mathrm{K}+1}) \xrightarrow{\mathrm{P}} \infty \qquad \text{if} \quad \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} \neq \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}.$$

This result can be applied in practice as an informal way to check whether the condition $\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} \neq \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}$ in Assumption 4 is likely to hold. Since we do not need to impose Assumption 4 in Theorem 3.4 when $\mathrm{K} \leq \mathbb{d}_{\mathsf{N}}$, we do not need to consider $\mathrm{K} \leq \mathbb{d}_{\mathsf{N}}$ in Theorem 3.5.

Remark 3.12. In our proof of Theorem 3.5, Assumption 5 is used to show that, when $a_R > 0$,

$$\|T^{-1}P_{\mathsf{S}}\widehat{\Lambda}_{1,P}P_{\mathsf{S}} - \Lambda_{\mathsf{S},R}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}((T/h_P)^{-1/2}) \tag{3.19}$$

(see Berkes et al., 2016, Theorems 2.2 and 2.3). The desired result then follows from (3.19). Thus, any other conditions that imply (3.19) can replace Assumption 5 in Theorem 3.5 (see also Lemma 4.1 of Rice and Shang, 2017).

Remark 3.13. A potentially useful result that may be deduced from our proofs of Theorems 3.4 and 3.5 is that the eigenvalues $\{\widehat{\tau}_j\}_{j\geq 1}$ of $\widehat{\Lambda}_{1,P}$ satisfy

$$\frac{\widehat{\tau}_j}{\widehat{\tau}_{j+1}} \xrightarrow{\mathrm{P}} \infty \text{ if } j = \mathbb{d}_{\mathsf{N}} \quad \text{and} \quad \frac{\widehat{\tau}_j}{\widehat{\tau}_{j+1}} = \mathcal{O}_{\mathrm{P}}(1) \text{ if } j \neq \mathbb{d}_{\mathsf{N}}. \tag{3.20}$$

This result is analogous to that in Theorem 3.3 of Chang et al. (2016) for the sample variance operator of $\{X_t\}_{t\geq 1}$ (see also Theorem 3.2 of Li et al., 2023). Based on (3.20), we may in practice easily obtain a reasonable upper bound on $\mathbb{d}_{\mathsf{N}}$, say $\mathbb{d}_{\max}$, by computing the ratio given in (3.20) and set $\mathbb{d}_{\max} = \max_{1\leq j\leq \bar{s}}\{\widehat{\tau}_j/\widehat{\tau}_{j+1}\} + k$ for some large integer $\bar{s}$ and a small nonnegative integer k. Of course, in view of the asymptotic properties of $\widehat{\tau}_j$, $\max_{1\leq j\leq \bar{s}}\{\widehat{\tau}_j/\widehat{\tau}_{j+1}\}$ is itself a consistent estimator of $\mathbb{d}_{\mathsf{N}}$. The latter is a slight modification of the eigenvalue ratio estimator of Li et al. (2023). However, we found that the finite-sample properties of this estimator tend to be worse than those of a different eigenvalue ratio-based estimator to be discussed in Section 4.2. Nonetheless, either eigenvalue ratio-based estimator could be useful as a simple way to construct a reasonable upper bound on $\mathbb{d}_{\mathsf{N}}$ in practice.

Remark 3.14. One practical advantage of our methodology is its applicability to both vector-valued and functional time series. In practice, functional data $X_t(r)$ are

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

often observed only at discrete, regular intervals, $\{X_t(r_1), \ldots, X_t(r_M)\}$, as illustrated in Nielsen et al. (2023) and Section 7.1. When the number of sampling points M is small, applying our methodology directly to the vector-valued sequence is often more natural than smoothing the data to reconstruct functional variables. In more complicated cases with irregular or time-varying sampling grids, the process can be modeled as an error-contaminated cointegrated functional time series (Nam and Seo, 2025). Notably, Nam and Seo (2025, Sect. C.2) demonstrate that a version of our VR(2, 1) approach is robust to I(0) measurement errors. A formal extension to accommodate such irregularly sampled and noisy data is left for future research.

3.5. Inclusion of Deterministic Components

So far we have assumed that $\{X_t\}_{t\geq 1}$ has no deterministic component. However, in practice, it is quite common that observed time series contain a deterministic component. In this section, we focus on the case with a nonzero intercept and a linear trend since those seem to be most relevant in practice.

Suppose that the observed time series $\{Y_t\}_{t\geq 1}$ is

$$Y_t = \zeta_1 + \zeta_2 t + X_t, \quad t \geq 1, \tag{3.21}$$

where $\{X_t\}_{t\geq 1}$ satisfies Assumption 1. We let $U_t^{(1)}$ and $U_t^{(2)}$ be defined by

$$U_t^{(1)} = Y_t - \bar{Y}, \quad U_t^{(2)} = Y_t - \bar{Y} - (t-\bar{t})\frac{\sum_{t=1}^T (t-\bar{t})Y_t}{\sum_{t=1}^T (t-\bar{t})^2}, \tag{3.22}$$

where $\bar{Y} = T^{-1}\sum_{t=1}^T Y_t$ and $\bar{t} = T^{-1}\sum_{t=1}^T t = (T+1)/2$. Thus, $U_t^{(1)}$ and $U_t^{(2)}$ are, respectively, the mean-adjusted and the trend-adjusted residuals of Y_t (Kokoszka and Young, 2016).

Our ADI VR tests can be adjusted to accommodate deterministic terms by replacing the sample operators computed from $\{X_t\}_{t\geq 1}$ with the corresponding operators computed from $\{U_t^{(1)}\}_{t\geq 1}$ or $\{U_t^{(2)}\}_{t\geq 1}$. When the model for Y_t is given by (3.21) with $\zeta_2 = 0$ then we use $U_t^{(1)}$, and otherwise we use $U_t^{(2)}$.

This adjustment to accommodate deterministic terms produces some obvious changes in the asymptotic results given in Theorems 3.1 and 3.2 as in the existing literature (e.g., Nielsen et al., 2023, Sect. 3.4 and Seo, 2024, Sect. S3). For example, the extension for the VR $(d_L, 1)$ test can easily be done as in Theorem 3 of Nielsen et al. (2023). More specifically, in Theorems 3.1 and 3.2, $W_{1,\mathbb{d}_N}$ needs to be replaced by $\mathbb{d}_N$-dimensional demeaned standard Brownian motion (if $\zeta_2 = 0$) or detrended standard Brownian motion (if $\zeta_2 \neq 0$), and the definition of $W_{d_L,\mathbb{d}_N}$ needs to change accordingly. Similar changes have to be made in Theorem 3.3, but where, as may be deduced from our proof of Theorem 3.3, $B_{K-\mathbb{d}_N}$ needs to be replaced with the standard Brownian bridge (if $\zeta_2 = 0$) or the second-level standard Brownian bridge (if $\zeta_2 \neq 0$), while $W_{d_L,\mathbb{d}_N}$ is the $(d_L - 1)$-fold integrated demeaned (if $\zeta_2 = 0$) or detrended (if $\zeta_2 \neq 0$) Brownian motion. The asymptotic results given in Theorems 3.4 and 3.5 do not require any changes once $\widehat{\Lambda}_{1,P}$ is constructed

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

from the relevant residuals (our proofs of Theorems 3.4 and 3.5 contain related discussions). Following these changes to the theorems, it is quite obvious to make relevant changes to the corollaries given in the previous sections, and hence the details are omitted.

4. DETERMINATION OF DIMENSION OF NONSTATIONARY SUBSPACE

Practitioners may mostly be interested in determining the dimension of $\mathcal{H}_{\mathsf{N}}$ rather than testing a specific hypothesis on the dimension. In this section, we first propose sequential testing procedures for doing so, and then we consider an eigenvalue ratio estimation approach.

4.1. Estimation via Sequential Testing

First, we propose a "top-down" (TD) sequential procedure to estimate $\mathbb{d}_{\mathsf{N}}$. Here, we test (3.1) sequentially (using any of the tests given in Corollaries 3.1 and 3.2) for $\mathbb{d}_0 = \mathbb{d}_{\max}, \mathbb{d}_{\max} - 1, \ldots$, where $\mathbb{d}_{\max}$ is an upper bound on the dimension of $\mathcal{H}_{\mathsf{N}}$ that can reasonably be chosen to be (slightly) greater than $\mathbb{d}_{\mathsf{N}}$ in practice (e.g., Remark 3.13). We let $\widehat{\mathbb{d}}_{\mathrm{TD}}$ denote the value of $\mathbb{d}_0$ under the first non-rejected null hypothesis; if the null hypothesis is rejected for every $\mathbb{d}_0 = \mathbb{d}_{\max}, \mathbb{d}_{\max} - 1, \ldots, 1$, then $\widehat{\mathbb{d}}_{\mathrm{TD}} = 0$.

Second, we consider a "bottom-up" (BU) procedure, where we test (3.13) sequentially for $\mathbb{d}_0 = 0, 1, \ldots$ using any of the inverse VR tests in Corollary 3.3. We then let $\widehat{\mathbb{d}}_{\mathrm{BU}}$ denote the first non-rejected null hypothesis. This procedure is attractive because no $\mathbb{d}_{\max}$ needs to be specified, and the condition $\mathbb{d}_{\max} \geq \mathbb{d}_{\mathsf{N}}$ is not required (see also Remark 4.1).

The sequential procedures can be applied with a different K and $\widehat{P}_{\mathrm{K}}$ for each value of $\mathbb{d}_0$, that is, in each step of the algorithm. Then K should be understood as a function of $\mathbb{d}_0$, denoted $\mathrm{K} = \mathrm{K}(\mathbb{d}_0)$, although we mostly suppress the dependence on $\mathbb{d}_0$ to simplify the notation. In practice, K can simply be set to $\mathrm{K} = \mathbb{d}_{\max} + m$ or $\mathrm{K} = \mathrm{K}(\mathbb{d}_0) = \mathbb{d}_0 + m$ for some integer $m \geq 1$ as in Nielsen et al. (2023) (see Remark 4.1 for details).

The consistency results for both procedures are given in the following theorem.

THEOREM 4.1. *Suppose that Assumptions 1 and 3 hold and that the nominal level of each individual test is* α.

(i) *If* $\mathbb{d}_{\max} \geq \mathbb{d}_{\mathsf{N}}$ *and, in each step,* $\mathrm{K} \geq \mathbb{d}_0$ *and* $\widehat{P}_{\mathrm{K}}$ *satisfies Assumption 2, then*

$$\mathrm{P}(\widehat{\mathbb{d}}_{\mathrm{TD}} = \mathbb{d}_{\mathsf{N}}) \to 1 - \alpha \quad \textit{and} \quad \mathrm{P}(\widehat{\mathbb{d}}_{\mathrm{TD}} > \mathbb{d}_{\mathsf{N}}) \to 0.$$

(ii) *If, in each step,* $\mathrm{K} > \mathbb{d}_0$ *and* $\widehat{P}_{\mathrm{K}}$ *satisfies Assumption 2', then*

$$\mathrm{P}(\widehat{\mathbb{d}}_{\mathrm{BU}} = \mathbb{d}_{\mathsf{N}}) \to 1 - \alpha \quad \textit{and} \quad \mathrm{P}(\widehat{\mathbb{d}}_{\mathrm{BU}} < \mathbb{d}_{\mathsf{N}}) \to 0.$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

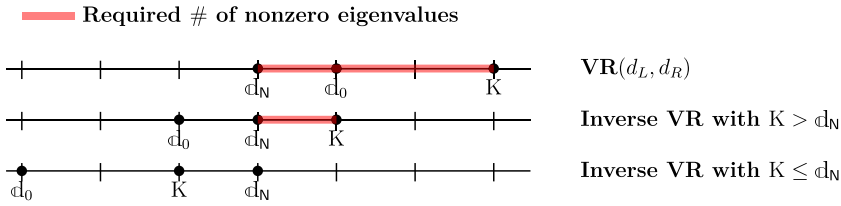


FIGURE 1. Eigenvalue requirements for the proposed tests, $\mathrm{K} = \mathrm{K}(\mathrm{d}_0) = \mathrm{d}_0 + 2$.

The proof of Theorem 4.1 follows directly from earlier results and is therefore omitted. For both procedures, we note that if $\alpha \to 0$ as $T \to \infty$ then the procedures are consistent; that is, $\mathrm{P}(\widehat{\mathrm{d}}_{\mathrm{TD}} = \mathrm{d_N}) \to 1$ and $\mathrm{P}(\widehat{\mathrm{d}}_{\mathrm{BU}} = \mathrm{d_N}) \to 1$.

Remark 4.1. The conditions stated in Theorem 4.1 for each of the sequential procedures simplify nicely if simple choices of K are applied. For example, if either $\mathrm{K} = \mathrm{d}_{\max} + m$ or $\mathrm{K} = \mathrm{K}(\mathrm{d}_0) = \mathrm{d}_0 + m$ for some integer $m \geq 1$, then the TD procedure in Theorem 4.1(i) only requires Assumption 2 to hold for $\mathrm{K} = \mathrm{d}_{\max} + m$. If $\mathrm{K} = \mathrm{K}(\mathrm{d}_0) = \mathrm{d}_0 + m$, then the BU procedure in Theorem 4.1 only requires Assumption 2' to hold for $\mathrm{K} = \mathrm{d_N} + m$, that is, Assumption 2 with $\mathrm{K} = \mathrm{d_N} + m$. Because $\mathrm{d}_{\max} \geq \mathrm{d_N}$, the condition required for the TD procedure is stronger than that for the BU procedure.

Remark 4.2. To implement the proposed sequential procedures under the low-level conditions discussed in Section 3.4, in each step we require Assumption 4 with $\mathrm{K} = \mathrm{K}(\mathrm{d}_0) \geq \mathrm{d}_0$ for the TD procedure or $\mathrm{K} = \mathrm{K}(\mathrm{d}_0) > \mathrm{d}_0$ for the BU procedure. In either case, for each d_0, our tests do not require knowledge of the dimension of the (possibly strict sub)space in which X_t takes values, but do require $\Lambda_{\mathsf{S},R}$ to have $\mathrm{K} - \mathrm{d_N}$ nonzero eigenvalues. It may be deduced from Theorem 3.4 and Assumption 4 that the TD procedures based on VR(d_L, d_R) tests require a stronger condition on $\Lambda_{\mathsf{S},R}$ than the BU procedure based on inverse VR tests. To see this, suppose that a common choice of K (as a function of d_0) is used for both of the procedures, such as $\mathrm{K} = \mathrm{d}_0 + m$ for some $m \geq 1$. Because the BU procedure starts from $\mathrm{d}_0 \leq \mathrm{d_N}$ and the TD procedure from $\mathrm{d}_0 \geq \mathrm{d_N}$, the number of nonzero eigenvalues that we require for the TD procedure is always larger than that required for the BU procedure; this is illustrated for $m = 2$ in Figure 1. Moreover, the number of nonzero eigenvalues of $\Lambda_{\mathsf{S},R}$ required by the TD procedure is always positive when $\mathrm{d}_0 > \mathrm{d_N}$, and this number increases as d_0 gets larger. This suggests that it is important to have a reasonable starting point $\mathrm{d}_{\max} (\geq \mathrm{d_N})$ for the TD procedure. Similarly, it should also be noted that, all else equal, $\Lambda_{\mathsf{S},R}$ is required to have more nonzero eigenvalues as K increases for both procedures, so it is preferable to choose K not too much larger than d_0. A similar suggestion for the choice of K can be found in Nielsen et al. (2023), but their argument was based on computational advantages in a more restrictive setting than ours. Our finding thus gives a theoretical justification for their choice in our more general setting.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

As discussed in Remarks 4.1 and 4.2, the conditions required for the TD procedure are stronger than those required for the BU procedure. The BU procedure has the additional advantage that it can be implemented without specifying an upper bound $\mathbb{d}_{\max}$ on $\mathbb{d}_{\mathrm{N}}$ and, specifically, the choices of K for the BU procedure to consistently estimate $\mathbb{d}_{\mathrm{N}}$ (with shrinking significance levels as discussed above) depend only on $\mathbb{d}_0$ (not $\mathbb{d}_{\mathrm{N}}$). However, Monte Carlo simulations have shown that the TD procedure has better finite-sample properties. Thus, we combine the above two sequential procedures into an "up-down" (UD) hybrid procedure that enjoys the advantages of both the TD and BU procedures. Specifically, we first apply the BU procedure to obtain $\widehat{\mathbb{d}}_{\mathrm{BU}}$ and use this to specify a data-driven value of the upper bound as $\mathbb{d}_{\max} = \widehat{\mathbb{d}}_{\mathrm{BU}} + m_{\mathbb{d}}$ for an integer $m_{\mathbb{d}} \geq 0$. We then apply the TD procedure initiated at this upper bound to obtain the UD hybrid estimate, $\widehat{\mathbb{d}}_{\mathrm{UD}}$. Note that the upper bound obtained in the first (BU) part of the UD procedure can also be used to choose the same $\mathrm{K} = \mathbb{d}_{\max} + m$ for an integer $m \geq 0$ in all steps in the second (TD) part of the UD procedure. The finite-sample properties of this UD hybrid procedure, along with those of the other procedures, will be investigated by simulations in Section 6.

4.2. Eigenvalue Ratio Estimator

Recent work concerning vector-valued time series provides ways to consistently estimate the number of stochastic trends using the fact that the eigenvalues of the sample autocovariance matrices have different stochastic orders if the time series contains components of different integration orders (e.g., Zhang and Chan, 2018; Zhang et al., 2019; Chang, Cheng, and Yao, 2021). In a functional time series setting, the eigenvalue ratio estimator proposed by Li et al. (2020, 2023) can be used to estimate the dimension of the component of the highest integration order. A slight modification of their estimator was discussed in Remark 3.13.

Alternatively, based on the asymptotic results in Section 3, we may use the generalized eigenvalue problem (3.4) to construct an eigenvalue ratio estimator for $\mathbb{d}_{\mathrm{N}}$. Let μ_j be an eigenvalue of (3.4). As we prove in Theorem 4.2 below, we can use Theorems 3.1–3.3 and arguments from their proofs to show that $\mu_j^{-1}\mu_{j+1} = \mathcal{O}_{\mathrm{P}}(1)$ if $j < \mathbb{d}_{\mathrm{N}}$ (resp. $j > \mathbb{d}_{\mathrm{N}}$) since both $n_T\mu_j$ and $n_T\mu_{j+1}$ converge (resp. diverge at the same rate), while $\mu_j^{-1}\mu_{j+1} \to_{\mathrm{P}} \infty$ if $j = \mathbb{d}_{\mathrm{N}}$. Hence, the ratio $\mu_j^{-1}\mu_{j+1}$ will be maximized at $j = \mathbb{d}_{\mathrm{N}}$. This motivates the estimator

$$\ddot{\mathbb{d}} = \underset{1 \leq j \leq \mathbb{d}_{\max}}{\arg\max} \left\{ \frac{\mu_{j+1}}{\mu_j} \right\}, \tag{4.1}$$

where $\mathbb{d}_{\max}$ is a pre-specified integer indicating an upper bound on $\mathbb{d}_{\mathrm{N}}$. We note that these eigenvalue ratio-type estimators are only applicable when $\mathbb{d}_{\mathrm{N}} \geq 1$; hence this needs to be examined first in practical implementation of these estimators. A convenient way to do this is to test (3.13) with $\mathbb{d}_0 = 0$ using the tests in Corollary 3.3.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

THEOREM 4.2. *Suppose that (i)* $\mathbb{d}_{\max} \geq \mathbb{d}_{\mathsf{N}} \geq 1$*, (ii) Assumptions 1–3 hold with* $\mathrm{K} \geq \mathbb{d}_{\max} + 1$*, and (iii)* $a_R \geq 0$ *(resp.* $a_R > 0$*) if* μ_j *is obtained from (3.4) with* $d_R = 1$ *(resp.* $d_R = 0$*) and* $d_L \geq d_R + 1$*. Then* $\mathrm{P}(\ddot{\mathbb{d}} = \mathbb{d}_{\mathsf{N}}) \to 1$.

The estimator in (4.1) is very similar to the eigenvalue ratio estimator proposed by Li et al. (2020, 2023), see also Remark 3.13, but there are some important differences. First, their estimator is constructed directly from the eigenvalues of a sample variance operator, while we use the generalized eigenvalue problem (3.4). Second, their estimator requires a tuning parameter, say $\kappa > 0$, to deal with estimation error associated with small eigenvalues; specifically, they set $\widehat{\mathbb{d}}_{\mathrm{LRS}} = \arg\min_{1 \leq j \leq \mathbb{d}_{\max}} \{\hat{\kappa}_{j+1}/\hat{\kappa}_j\}$, where $\hat{\kappa}_j$ is the j-th largest eigenvalue of the sample variance operator and $\hat{\kappa}_j$ is regarded as zero if $\hat{\kappa}_j/\hat{\kappa}_1$ is smaller than κ, and they define $0/0 = 1$. However, our estimator in (4.1) does not require such a tuning parameter.

5. INFERENCE ON SUBSPACES

In practical applications, it may be of interest to test hypotheses about $\mathcal{H}_{\mathsf{N}}$ or $\mathcal{H}_{\mathsf{S}}$. For a specified subspace $\mathcal{H}_0$, consider the following hypotheses:

$$H_0 : \mathcal{H}_0 \subseteq \mathcal{H}_{\mathsf{N}} \text{ (or equivalently, } \mathcal{H}_0^{\perp} \supseteq \mathcal{H}_{\mathsf{S}}) \quad \text{vs.} \quad H_1 : H_0 \text{ is not true,} \tag{5.1}$$

$$H_0 : \mathcal{H}_0 \supseteq \mathcal{H}_{\mathsf{N}} \text{ (or equivalently, } \mathcal{H}_0^{\perp} \subseteq \mathcal{H}_{\mathsf{S}}) \quad \text{vs.} \quad H_1 : H_0 \text{ is not true,} \tag{5.2}$$

$$H_0 : \mathcal{H}_0 \subseteq \mathcal{H}_{\mathsf{S}} \text{ (or equivalently, } \mathcal{H}_0^{\perp} \supseteq \mathcal{H}_{\mathsf{N}}) \quad \text{vs.} \quad H_1 : H_0 \text{ is not true.} \tag{5.3}$$

For example, we may be interested in testing if a specific element $v \in \mathcal{H}$ is included in $\mathcal{H}_{\mathsf{N}}$, and hence can be interpreted as one of the common stochastic trends. In that case, we can use (5.1) with $\mathcal{H}_0 = \mathrm{span}\{v\}$. We can also test if a specified subspace, for example, $\mathcal{H}_0 = \mathrm{span}\{v_1, v_2\}$, contains the entire nonstationary subspace by using (5.2). Finally, it may be of interest to examine if a specified element or subspace is in $\mathcal{H}_{\mathsf{S}}$, and then we can use (5.3). In a finite-dimensional setting, (5.3) corresponds to testing if a vector or matrix is cointegrating.

Let $P_{\mathcal{H}_0}$ denote the projection onto $\mathcal{H}_0$, let $p_0 = \dim(\mathcal{H}_0)$, and let $\mathbb{d}_{\mathsf{N}}$ be known (or replaced by a consistent estimator from Section 4). The hypotheses (5.1) and (5.2) can then be tested by investigating the dimension of the nonstationary subspace associated with the residual series $\{(I - P_{\mathcal{H}_0})X_t\}_{t \geq 1}$, and the hypothesis (5.3) can be tested by investigating the dimension of the nonstationary subspace associated with the projected series $\{P_{\mathcal{H}_0}X_t\}_{t \geq 1}$. Specifically, if H_0 of (5.1) is true, then $\{(I - P_{\mathcal{H}_0})X_t\}_{t \geq 1}$ contains $\mathbb{d}_{\mathsf{N}} - p_0$ stochastic trends, if H_0 of (5.2) is true then $\{(I - P_{\mathcal{H}_0})X_t\}_{t \geq 1}$ contains zero stochastic trends, and if H_0 of (5.3) is true then $\{P_{\mathcal{H}_0}X_t\}_{t \geq 1}$ contains zero stochastic trends. Under H_1 in (5.1)–(5.3), there are more stochastic trends in the relevant time series than under the null. Thus, by analyzing $\{(I - P_{\mathcal{H}_0})X_t\}_{t \geq 1}$ or $\{P_{\mathcal{H}_0}X_t\}_{t \geq 1}$, (5.1)–(5.3) may be readily tested by our inverse VR test, assuming the projected time series satisfy the conditions for the inverse VR test.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

THEOREM 5.1. *Suppose that* $\{P_{\mathcal{H}_0}X_t\}_{t\geq 1}$ *and* $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$ *satisfy the conditions imposed in Theorem 3.3.*

(i) *Consider the hypotheses (5.1). The test statistics in Corollary 3.3 computed from* $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$ *with* $\mathbb{d}_0 = \mathbb{d}_{\mathsf{N}} - p_0 \geq 0$ *satisfy (3.16) under* H_0 *and (3.17) under* H_1.
(ii) *Consider the hypotheses (5.2). The test statistics in Corollary 3.3 computed from* $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$ *with* $\mathbb{d}_0 = 0$ *and* $p_0 \geq \mathbb{d}_{\mathsf{N}}$ *satisfy (3.16) under* H_0 *and (3.17) under* H_1.
(iii) *Consider the hypotheses (5.3). The test statistics in Corollary 3.3 computed from* $\{P_{\mathcal{H}_0}X_t\}_{t\geq 1}$ *with* $\mathbb{d}_0 = 0$ *and* $\mathrm{K} = p_0$ *satisfy (3.16) under* H_0 *and (3.17) under* H_1.

Of course, the high-level conditions in Assumption 2', which are applied in Theorem 5.1, may be replaced by appropriate low-level conditions as discussed in Theorems 3.4 and 3.5.

The approach to testing the hypotheses (5.1) and (5.2) based on $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$ is parallel to that in Seo (2024) who relied on an FPCA-based test for examining the number of stochastic trends embedded in a cointegrated functional time series. However, Seo (2024, Assump. W) requires the condition that the long-run variance of $\{P_{\mathsf{S}}X_t\}_{t\geq 1}$ is positive definite on $\mathcal{H}_{\mathsf{S}}$, and this may be restrictive in cases, where X_t only takes values on an unknown (and possibly finite-dimensional) subspace. This condition is not required under our setup.

It is not possible, in general, to investigate the hypothesis (5.1) (resp. (5.2)) by testing if there are p_0 (resp. $\mathbb{d}_{\mathsf{N}}$) stochastic trends in the projected time series $\{P_{\mathcal{H}_0}X_t\}_{t\geq 1}$. For example, let $\{f_j\}_{j\geq 1}$ be an orthonormal basis of $\mathcal{H}$ and suppose $\mathcal{H}_{\mathsf{N}} = \operatorname{span}\{f_1\}$, that is, $\mathbb{d}_{\mathsf{N}} = 1$. Consider $\mathcal{H}_0 = \operatorname{span}\{g\}$ for $g = 0.5f_1 + 0.5f_2$, which can be a null hypothesis in either (5.1) or (5.2). Both null hypotheses are false since neither $\operatorname{span}\{g\} \subseteq \operatorname{span}\{f_1\}$ nor $\operatorname{span}\{g\} \supseteq \operatorname{span}\{f_1\}$ is true. However, the projected time series $P_{\mathcal{H}_0}X_t = \langle X_t, g\rangle = \langle X_t, f_1\rangle + \langle X_t, f_2\rangle$ is nonstationary and has one stochastic trend. Similarly, H_0 in (5.3) cannot be investigated by testing if there are $\mathbb{d}_{\mathsf{N}}$ stochastic trends in $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$. These impossibilities are discussed in Theorem 4.2 in Franchi and Paruolo (2026), which characterizes the structure of the parameter space under the null and alternative of hypotheses of type (5.1) and (5.2).

Remark 5.1. In addition to testing the hypotheses (5.1)–(5.3), it may be of independent interest to examine the number of stochastic trends in the residual series $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$ for a specified subspace $\mathcal{H}_0$. This is particularly useful when $\mathcal{H}_0$ may be understood as a finite-dimensional "model" of X_t, as in our empirical example in Section 7.2. In such cases, we may be interested in determining how well the model can capture the dominant (nonstationary) variation in the data. The number of stochastic trends in $\{(I-P_{\mathcal{H}_0})X_t\}_{t\geq 1}$, denoted $\mathbb{d}_{\mathcal{H}_0}$, can be interpreted as the dimension of the nonstationary subspace that remains unexplained by the model. We can determine $\mathbb{d}_{\mathcal{H}_0}$ by sequential testing as in Section 4.1. Specifically,

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

we can apply inverse VR tests on

$$H_0 : \mathrm{d}_{\mathcal{H}_0} = \mathrm{d}_0 \quad \text{vs} \quad H_1 : \mathrm{d}_{\mathcal{H}_0} > \mathrm{d}_0 \tag{5.4}$$

for $\mathrm{d}_0 = \max\{\mathrm{d}_{\mathsf{N}} - p_0, 0\}, \max\{\mathrm{d}_{\mathsf{N}} - p_0, 0\} + 1, \ldots, \mathrm{d}_{\mathsf{N}}$, or we can apply VR tests on

$$H_0 : \mathrm{d}_{\mathcal{H}_0} = \mathrm{d}_0 \quad \text{vs} \quad H_1 : \mathrm{d}_{\mathcal{H}_0} < \mathrm{d}_0 \tag{5.5}$$

for $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}}, \mathrm{d}_{\mathsf{N}} - 1, \ldots, \max\{\mathrm{d}_{\mathsf{N}} - p_0, 1\}$.

6. MONTE CARLO STUDY

We study the finite-sample performance of the proposed ADI tests using a Monte Carlo simulation setup similar to the functional AR(1) model in Li et al. (2023), Nielsen et al. (2023), and Seo (2024). Let $\{g_j\}_{j\geq 1}$ be the Fourier basis functions on $[0,1]$. For each d_{N}, we let $P_{\mathsf{N}}\Delta X_t$ and $P_{\mathsf{S}}X_t$ be generated by the following stationary functional AR(1) models:

$$P_{\mathsf{N}}\Delta X_t = \sum_{j=1}^{\mathrm{d}_{\mathsf{N}}} \alpha_j \langle g_j^N, P_{\mathsf{N}}\Delta X_{t-1}\rangle g_j^N + P_{\mathsf{N}}\varepsilon_t,$$

$$P_{\mathsf{S}}X_t = \sum_{j=1}^{12} (0.9)^{j-1}\beta_j \langle g_j^S, P_{\mathsf{S}}X_{t-1}\rangle g_j^S + P_{\mathsf{S}}\varepsilon_t,$$

where $\{g_j^N\}_{j=1}^{\mathrm{d}_{\mathsf{N}}}$ are randomly drawn from $\{g_1, \ldots, g_{\mathrm{d}_{\mathsf{N}}+3}\}$ ($\mathrm{d}_{\mathsf{N}} \leq 8$ in our simulation experiments), $\{g_j^S\}_{j=1}^{12}$ are randomly drawn from $\{g_{15}, \ldots, g_{30}\}$, and ε_t is given by $\sum_{j=1}^{40}(0.9)^{j-1}\theta_{j,t}g_j$ for standard normal random variables $\{\theta_{j,t}\}_{j\geq 1}$ that are independent across j and t. We let α_j and β_j be independent uniform random variables on $[-0.8, 0.8]$ for each realization of the data-generating process (DGP). In practice, it is common to have a nonzero intercept or a linear time trend as in Section 3.5. We here consider the former case and add an intercept ζ_1, which is also randomly chosen, to each realization of the DGP. Specifically, $\zeta_1 = \sum_{j=1}^{30}(0.9)^{j-1}\tilde{\theta}_j g_j$, where $\{\tilde{\theta}_j\}_{j=1}^{30}$ are independent standard normal random variables. Viewed as a time series taking values in the $L^2[0,1]$ Hilbert space, we may obtain the eigenelements of the sample (long-run) variance of X_t using FPCA (see, e.g., Chang et al., 2016; Li et al., 2023; Nielsen et al., 2023). To this end, we represent X_t, which is assumed to be observed on 200 regularly spaced points of $[0,1]$, using the first 40 Legendre basis functions.

To choose bandwidths, we follow Remarks 3.5–3.7 and 3.11. For VR tests with $d_R = 1$, we let $h_P = h_L = h_R = 0$ (Remarks 3.5 and 3.11). For the remaining tests, we let $h_P = [T^{1/3}]$ because a large bandwidth is supported by Assumption 5. For VR tests with $d_R = 0$, we let $h_R = h_{\mathrm{opt}}$, which is the Andrews (1991) optimal bandwidth (Remark 3.6), but for the inverse tests the series are nonstationary, so we let $h_R = [T^{1/5}]$, which is the same rate as h_{opt} (Remark 3.7). We consider both $h_L = 0$ and

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

$h_L = [T^{1/3}]$ (Remark 3.6). Throughout, we let $\mathrm{k}_m(\cdot)$ be the Tukey–Hanning kernel for $m \in \{L, R, P\}$, all test statistics apply the trace test functional $\mathcal{F}_{\text{trace}}$, and the slack extraction uses $\mathrm{K} = \mathrm{d}_0 + 2$ following Remark 4.1.

We present simulation results for the VR(2,1), VR(2,0), VR(1,0), and inverse VR tests (the latter based on the (2,1) problem). The VR choices (2,1), (1,0), and the inverse VR are motivated by well-known tests from finite-dimensional Euclidean space (see Remark 3.4 and Section S.2 of the Supplementary Material, and the VR(2,0) is in-between). We also performed simulations for VR tests based on larger values of d_L, but found that their performances were inferior to the tests reported here (results available upon request).

Table 1 reports rejection frequencies. We consider true dimension $\mathrm{d}_{\mathsf{N}} \in \{0, 1, 3, 5, 7\}$. For the VR tests, we consider the null $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}}$ and alternatives $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}} + 1$ and $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}} + 2$, and for the inverse VR test we consider the null $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}}$ and alternatives $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}} - 1$ and $\mathrm{d}_0 = \mathrm{d}_{\mathsf{N}} - 2$. The results show that the VR(1,0) and VR(2,0) tests are seriously under-sized when $a_L = 0$, as expected based on Shintani (2001), especially for the larger values of d_{N} considered, and both tests suffer from lower power in those cases as a consequence. Size control is much better for $h_L = [T^{1/3}]$ for the VR(1,0) test, but not for the VR(2,0) test. The VR(2,1) test has very good size properties in all cases considered except the largest value of d_{N} with the smallest sample size, and it has excellent power. Finally, we see that the inverse VR test is slightly over-sized in some cases. Its power is nearly as impressive as that of the VR(2,1) test, but of course the alternatives are different for the inverse VR test, and it does not have the natural upper bound $\mathrm{d}_{\max}$ that the VR tests have.

In Table 2, we investigate the finite-sample performance of the tests when they are applied to estimate d_{N} using one of the sequential procedures described in Section 4.1. For comparison, we also consider the eigenvalue ratio estimator $\ddot{\mathrm{d}}$ computed from the VR(2,1) problem (see Theorem 4.2), and Li et al.'s (2023) estimator, which is denoted $\widehat{\mathrm{d}}_{\text{LRS}}$. For each method and each value of d_{N}, the table reports the frequencies of correctly estimating the true dimension d_{N}. Overall, the direct estimators $\ddot{\mathrm{d}}$ and $\widehat{\mathrm{d}}_{\text{LRS}}$ perform worse than the VR(2,1) and inverse VR sequential testing procedures, and between the two eigenvalue ratio estimators, $\ddot{\mathrm{d}}$ tends to perform better than $\widehat{\mathrm{d}}_{\text{LRS}}$. The sequential TD procedure based on the VR(2,1) tests performs very well, and the BU procedure (based on the inverse VR test) performs nearly as well. On the other hand, the TD procedures based on the VR(2,0) and VR(1,0) tests perform very poorly for $\mathrm{d}_{\mathsf{N}} \geq 3$, which was expected based on their poor power observed in Table 1. The choice of h_L seems to significantly affect the performance of the VR(1,0) tests, especially when d_{N} is large. Of course, this is not the case for the TD procedure based on the VR(2,1) test, and indeed, this method appears to perform the best overall.

Results for the UD hybrid procedure are also reported in Table 2. This enjoys the advantage of a data-driven selection of $\mathrm{d}_{\max} = \widehat{\mathrm{d}}_{\text{BU}} + 5$. This $\mathrm{d}_{\max}$ is random, but the good performance of the BU procedure, the addition of 5 from $\widehat{\mathrm{d}}_{\text{BU}}$, and the strong robustness of the TD procedures to the choice of $\mathrm{d}_{\max}$, imply that in fact the

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

TABLE 1. Rejection frequencies of VR(2,1), VR(2,0), VR(1,0), and inverse VR tests.

		$T = 250$					$T = 500$				
	d_0	$d_N = 0$	$d_N = 1$	$d_N = 3$	$d_N = 5$	$d_N = 7$	$d_N = 0$	$d_N = 1$	$d_N = 3$	$d_N = 5$	$d_N = 7$
VR(2,1)	d_N		0.049	0.053	0.064	0.105		0.051	0.050	0.059	0.061
	$d_N + 1$	1.000	0.998	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000
	$d_N + 2$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					$h_L = 0$						
VR(2,0)	d_N		0.043	0.022	0.008	0.004		0.046	0.031	0.018	0.009
	$d_N + 1$	1.000	0.899	0.599	0.374	0.213	1.000	0.982	0.857	0.678	0.498
	$d_N + 2$	1.000	0.915	0.667	0.451	0.286	1.000	0.992	0.919	0.770	0.608
VR(1,0)	d_N		0.027	0.005	0.000	0.000		0.036	0.010	0.003	0.000
	$d_N + 1$	1.000	0.636	0.144	0.024	0.002	1.000	0.709	0.374	0.155	0.036
	$d_N + 2$	1.000	0.671	0.223	0.093	0.010	1.000	0.729	0.417	0.246	0.122
Inv.VR	d_N	0.092	0.070	0.070	0.065	0.080	0.084	0.073	0.075	0.082	0.097
	$d_N - 1$		0.982	0.940	0.923	0.890		0.998	0.998	0.999	0.999
	$d_N - 2$			1.000	1.000	1.000			1.000	1.000	1.000
					$h_L = [T^{1/3}]$						
VR(2,0)	d_N		0.043	0.024	0.010	0.005		0.046	0.032	0.021	0.011
	$d_N + 1$	1.000	0.902	0.613	0.402	0.252	1.000	0.982	0.865	0.698	0.534
	$d_N + 2$	1.000	0.920	0.689	0.491	0.330	1.000	0.992	0.928	0.799	0.646
VR(1,0)	d_N		0.052	0.046	0.047	0.050		0.058	0.046	0.052	0.049
	$d_N + 1$	1.000	0.879	0.638	0.480	0.362	1.000	0.954	0.795	0.668	0.547
	$d_N + 2$	1.000	0.899	0.698	0.552	0.433	1.000	0.967	0.850	0.743	0.637
Inv.VR	d_N	0.089	0.066	0.054	0.032	0.025	0.082	0.070	0.064	0.057	0.051
	$d_N - 1$		0.982	0.930	0.882	0.773		0.998	0.997	0.998	0.996
	$d_N - 2$			1.000	1.000	0.998			1.000	1.000	1.000

Note: Based on 10,000 Monte Carlo replications. The DGP true dimension is d_N, the null hypothesized value is d_0, and the nominal size is 5%.

UD procedures enjoy nearly identical frequencies of correct estimation of d_N as the corresponding TD procedures. Thus, the advantage of the data-driven selection of d_{max} comes without cost, at least for these sample sizes, and consequently the UD procedure based on VR(2,1) is the preferred method in our Monte Carlo simulations.

Finally, recall that the functional observations are constructed from 200 regularly spaced points on $[0, 1]$ by smoothing using the first 40 Legendre basis functions. However, as discussed earlier, our methodology can also be applied to the time series of 200-dimensional vectors of discrete realizations of X_t with no additional theoretical modification. We thus repeated the analysis in Table 2

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

TABLE 2. Frequencies of correctly estimating d_{N}.

	$T = 250$					$T = 500$				
Method	$\mathrm{d}_{\mathrm{N}} = 0$	$\mathrm{d}_{\mathrm{N}} = 1$	$\mathrm{d}_{\mathrm{N}} = 3$	$\mathrm{d}_{\mathrm{N}} = 5$	$\mathrm{d}_{\mathrm{N}} = 7$	$\mathrm{d}_{\mathrm{N}} = 0$	$\mathrm{d}_{\mathrm{N}} = 1$	$\mathrm{d}_{\mathrm{N}} = 3$	$\mathrm{d}_{\mathrm{N}} = 5$	$\mathrm{d}_{\mathrm{N}} = 7$
$\ddot{\mathrm{d}}$		0.916	0.622	0.543	0.356		0.958	0.832	0.862	0.830
$\widehat{\mathrm{d}}_{\mathrm{LRS}}$		0.998	0.275	0.062	0.016		0.996	0.478	0.190	0.071
$\widehat{\mathrm{d}}_{\mathrm{TD}}$: VR(2,1)	1.000	0.949	0.946	0.935	0.894	1.000	0.949	0.950	0.941	0.939
$\widehat{\mathrm{d}}_{\mathrm{UD}}$: VR(2,1)	1.000	0.949	0.946	0.935	0.894	1.000	0.949	0.950	0.941	0.939
				$h_L = 0$						
$\widehat{\mathrm{d}}_{\mathrm{TD}}$: VR(2,0)	1.000	0.810	0.552	0.357	0.207	1.000	0.925	0.805	0.650	0.487
$\widehat{\mathrm{d}}_{\mathrm{TD}}$: VR(1,0)	1.000	0.602	0.134	0.023	0.002	1.000	0.668	0.363	0.152	0.036
$\widehat{\mathrm{d}}_{\mathrm{BU}}$: Inv.VR	0.908	0.915	0.872	0.860	0.812	0.916	0.925	0.923	0.917	0.901
$\widehat{\mathrm{d}}_{\mathrm{UD}}$: VR(2,0)	1.000	0.810	0.552	0.358	0.207	1.000	0.925	0.804	0.650	0.487
$\widehat{\mathrm{d}}_{\mathrm{UD}}$: VR(1,0)	1.000	0.602	0.133	0.023	0.002	1.000	0.668	0.362	0.152	0.036
				$h_L = [T^{1/3}]$						
$\widehat{\mathrm{d}}_{\mathrm{TD}}$: VR(2,0)	1.000	0.817	0.568	0.384	0.243	1.000	0.928	0.815	0.668	0.519
$\widehat{\mathrm{d}}_{\mathrm{TD}}$: VR(1,0)	1.000	0.823	0.591	0.432	0.312	1.000	0.893	0.748	0.615	0.498
$\widehat{\mathrm{d}}_{\mathrm{BU}}$: Inv.VR	0.911	0.918	0.878	0.851	0.749	0.918	0.928	0.933	0.941	0.945
$\widehat{\mathrm{d}}_{\mathrm{UD}}$: VR(2,0)	1.000	0.817	0.569	0.384	0.243	1.000	0.928	0.815	0.668	0.519
$\widehat{\mathrm{d}}_{\mathrm{UD}}$: VR(1,0)	1.000	0.822	0.591	0.432	0.312	1.000	0.893	0.748	0.615	0.498

Note: Based on 10,000 Monte Carlo replications. The DGP true dimension is d_{N}, the nominal size is 5%, and $\mathrm{d}_{\max} = \mathrm{d}_{\mathrm{N}} + 5$ for the TD methods.

letting $\{X_t\}_{t \geq 1}$ be viewed as a 200-dimensional vector-valued time series, but the results were nearly identical and are hence omitted.

Remark 6.1. Based on our simulation evidence, the hybrid procedure, combining the inverse VR (with $h_L = 0$ and $h_R > 0$) and the VR(2, 1) sequential tests, appears to perform well with $\mathrm{K} = \mathrm{d}_0 + m$ (see Remark 4.1). We provide a detailed practical implementation guide in Section S.1 of the Supplementary Material. In that section, we also discuss how to implement the procedure when the VR(2, 1) test is replaced by other variants.

7. EMPIRICAL APPLICATIONS

In this section, we illustrate our methodology with two real-world data examples.

7.1. Corporate Bond Yield Curves

We apply the proposed tests to examine the number of common stochastic trends in the monthly time series of high quality market corporate bond yield curves (a.k.a. the HQM yield curve). The data are from the U.S. Department

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

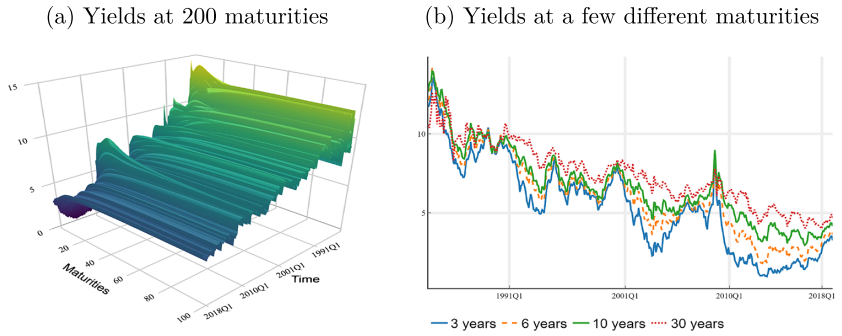


FIGURE 2. HQM yield curves.

of Treasury available at https://home.treasury.gov/ and spans January 1984 to December 2018. At each time t, we observe a corporate bond with 200 different maturities ranging from six months to 100 years. This data were studied by Barigozzi and Trapani (2022), who assumed that this high-dimensional time series is driven by a few factors of deterministic trends and I(1) and I(0) processes, and estimated the number of those factors using their own testing procedure. However, it may be more natural to view this time series as realizations of yield curves that are either finite-dimensional (of high and unknown dimension) or curve-valued. With this point of view, we apply our methodology to determine the number of stochastic trends. To implement our methodology, we represent the HQM yield data by the first 40 Legendre basis functions as in Section 6.

Figure 2a shows the time series of HQM yield curves (observed at 200 different maturities) and Figure 2b shows the time series of yields at fixed maturities of 3, 6, 10, and 30 years. The figures suggest that the HQM yield data includes a linear time trend. The existence of a linear trend is tested and confirmed by Barigozzi and Trapani (2022) for the data span of January 1985 to September 2017. Thus, we apply the trend-adjusted testing procedures in Section 3.5.

The empirical results are reported in Table 3. We implement all procedures as in Section 6 using $h_L = [T^{1/3}]$ for the VR(2,0), VR(1,0), and inverse VR tests, and using $\mathrm{K} = \mathrm{d}_0 + 2$ for slack extraction. The eigenvalue ratio estimates in the first part of the table are $\ddot{\mathrm{d}} = 6$ and $\widehat{\mathrm{d}}_{\mathrm{LRS}} = 5$ (using $\kappa = 10^{-4}$), although the ratios are close to suggesting $\ddot{\mathrm{d}} = 2$ and $\widehat{\mathrm{d}}_{\mathrm{LRS}} = 3$.

In the second part of Table 3, we report the VR and inverse VR test statistics for a range of d_0. The inverse VR tests imply that $\widehat{\mathrm{d}}_{\mathrm{BU}} = 3$ (at 1% level) or $\widehat{\mathrm{d}}_{\mathrm{BU}} = 6$ (at 5% level). Given this result, the choice $\mathrm{d}_{\max} = 7$ seems like a reasonable input to the UD procedures. The latter give either $\widehat{\mathrm{d}}_{\mathrm{UD}} = 2$ or $\widehat{\mathrm{d}}_{\mathrm{UD}} = 3$ depending on the choice of (d_L, d_R) and level. However, the BU and TD procedures (with (2,1) and (2,0)) only agree at 1% level, in which case all three find $\widehat{\mathrm{d}} = 3$. Overall, the evidence thus supports $\mathrm{d}_{\mathrm{N}} = 3$ stochastic trends.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

TABLE 3. Empirical results for HQM yield curve data.

Eigenvalue ratio estimates								
Statistic		$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
μ_{j+1}/μ_j		2.89	3.00	2.02	1.91	1.15	3.25	2.15
$\hat{\kappa}_{j+1}/\hat{\kappa}_j$		0.42	0.14	0.08	0.29	0.00	1	1
Variance ratio test statistics								
Statistic	$\mathrm{d}_0=0$	$\mathrm{d}_0=1$	$\mathrm{d}_0=2$	$\mathrm{d}_0=3$	$\mathrm{d}_0=4$	$\mathrm{d}_0=5$	$\mathrm{d}_0=6$	$\mathrm{d}_0=7$
Inv.VR	0.94**	0.60**	0.21**	0.12*	0.10*	0.07*	0.05 †	0.03
VR(2,1)		151.61	409.16	1,355.09*	3,484.67**	7,175.13**	10,623.79**	23,154.51**
VR(2,0)		7,007.65	13,261.54	117,284.83*	522,860.44**	1,208,358.03**	2,215,359.22**	6,936,298.27**
VR(1,0)		24.25	77.66 †	259.58**	599.41**	1,050.22**	1,836.25**	3,825.57**

Note: †, *, and ** indicate significance at 10%, 5%, and 1% level, respectively.

TABLE 4. Test results for $(I - P_{\mathcal{H}_0})X_t$ with $\mathcal{H}_0 = \text{span}\{\gamma_0, \gamma_1, \gamma_2\}$ and $\varsigma = 1.37$.

Statistic	$\mathbb{d}_0 = 0$	$\mathbb{d}_0 = 1$	$\mathbb{d}_0 = 2$	$\mathbb{d}_0 = 3$
Inv.VR	1.47**	0.27**	0.24**	0.12*
VR(2,1)		113.48	629.06 †	1,577.24*
VR(2,0)		6,842.81	39,832.64*	123,243.78**
VR(1,0)		72.34**	167.27**	346.79**

Note: †, *, and ** indicate significance at 10%, 5%, and 1% level, respectively.

7.2. The Nelson–Siegel Term Structure Model

Given an estimated dimension, it is possible to examine various hypotheses on the nonstationary subspace as in Section 5. A popular model for the term structure of interest rates was proposed by Nelson and Siegel (1987); henceforth the N–S model. They represent the term structure of yield curves by a linear combination of three parametric functions:

$$\gamma_0 = 1, \quad \gamma_1 = \gamma_1(\varsigma, \tau) = \varsigma(1 - e^{-\tau/\varsigma})/\tau,$$
$$\gamma_2 = \gamma_2(\varsigma, \tau) = \varsigma(1 - e^{-\tau/\varsigma})/\tau - e^{-\tau/\varsigma},$$

where τ is time to maturity, ς is a shape parameter, and γ_0, γ_1, and γ_2 denote level, slope, and curvature factors. Because γ_1 and γ_2 are nearly collinear for large τ, we subsequently consider only data for $\tau \in [0, 30]$ years. This has almost no impact on the results reported in Table 3. We now investigate to what extent the N–S model can explain the dominant (nonstationary) variation in the data. To this end, we need to specify the shape parameter, ς. Since our data are annualized, we set $\varsigma = 1.37$ following Diebold and Li (2006).

We first test the hypothesis H_0 in (5.3) with $\mathcal{H}_0 = \text{span}\{\gamma_0, \gamma_1, \gamma_2\}$ using Theorem 5.1(iii). That is, we test whether "fitted values," $\{P_{\mathcal{H}_0} X_t\}_{t \geq 1}$, from the N–S model are stationary. Not surprisingly, this hypothesis is strongly rejected. The same conclusion holds for tests of (5.3) when $\mathcal{H}_0$ is specified as the span of any one or two of the factors γ_i, $i = 0, 1, 2$. This confirms that none of the factors are stationary.

Next, we test if the level, slope, and curvature factors span the nonstationary subspace as in (5.2). From our earlier results in Table 3, we set the dimension of the latter to $\mathbb{d}_N = 3$ (testing at 1% level). Thus, there is no possibility that the N–S model with only one or two factors can yield stationary residuals. With three factors, it is of interest to test if the N–S model can explain all the nonstationary variation in the yield curve data and, as a result, can yield stationary residuals. In this sense, the hypothesis $H_0 : \mathcal{H}_0 = \text{span}\{\gamma_0, \gamma_1, \gamma_2\} \supseteq \mathcal{H}_N$ in (5.2) is very interesting, and its non-rejection would be the strongest conclusion that we could obtain. However, implementing the test as in Theorem 5.1(ii) with $\mathbb{d}_0 = 0$ and $p_0 = 3$, the inverse VR statistic is 1.47 (see Table 4) and rejects the null at 1% level.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

We proceed to test H_0 in (5.1), where $\mathcal{H}_0$ is specified as various choices of factor(s). That is, we test whether any one factor can reduce the dimension of the nonstationary subspace by one, and also whether any two factors can reduce the dimension by two. Of course, the latter is less likely to be accepted than the former. For one factor, implementing the test as in Theorem 5.1(i), the inverse VR statistics with $\mathrm{d}_0 = 2$ are 0.24, 0.21, and 0.22 for $\mathcal{H}_0 = \text{span}\{\gamma_0\}$, $\mathcal{H}_0 = \text{span}\{\gamma_1\}$, $\mathcal{H}_0 = \text{span}\{\gamma_2\}$, respectively, in each case rejecting the null at 1% level. For two factors, the inverse VR statistics with $\mathrm{d}_0 = 1$ are 0.40 and 0.38 for $\mathcal{H}_0 = \text{span}\{\gamma_0, \gamma_1\}$ and $\mathcal{H}_0 = \text{span}\{\gamma_1, \gamma_2\}$, respectively, in both cases rejecting the null at 1% level. Thus, there is no evidence to suggest that any one factor is in the nonstationary subspace, that is, that any one factor can be interpreted as one of the common stochastic trends.

Finally, following Remark 5.1, we investigate if the nonstationary subspace of the residual series, $\{(I - P_{\mathcal{H}_0})X_t\}_{t\geq 1}$ with $\mathcal{H}_0 = \text{span}\{\gamma_0, \gamma_1, \gamma_2\}$, is of smaller dimension than that of X_t itself. Letting the former be denoted $\mathrm{d}_{\mathcal{H}_0}$, we apply the sequential procedures in (5.4) and (5.5). The results are reported in Table 4. The BU procedure yields $\widehat{\mathrm{d}}_{\mathcal{H}_0,\mathrm{BU}} = 3$ at 1% level and rejects $\mathrm{d}_0 = 3$ at 5% level. Applying the TD sequential procedure starting from $\mathrm{d}_{\max} = 3$, we find $\widehat{\mathrm{d}}_{\mathcal{H}_0,\mathrm{TD}} = 2$ with the preferred VR(2,1) test at 5% level (but $\widehat{\mathrm{d}}_{\mathcal{H}_0,\mathrm{TD}} = 1$ at 10% level with the VR(2,1) test, $\widehat{\mathrm{d}}_{\mathcal{H}_0,\mathrm{TD}} = 1$ at 5% level with the VR(2,0) test, and $\widehat{\mathrm{d}}_{\mathcal{H}_0,\mathrm{TD}} = 0$ at 1% level with the VR(1,0) test). Thus, we do find some evidence that three factors in the N–S model explain some of the nonstationary variation in the data.

The empirical results for the N–S model are specific to the value of ς, at least for the slope and curvature factors. As a robustness check, we repeated the analysis for $\varsigma = 1, 2, 3, 4, 5$, and the results were qualitatively the same. In conclusion, the three parametric factors of the N–S model cannot fully explain the nonstationarity in the corporate yield curve data, so the residuals from the model will still exhibit nonstationarity. However, the model can capture some of the nonstationarity, in the sense that the dimension of the nonstationary subspace of the residuals is smaller than that of the original yield curve data.

7.3. Labor Market Indices

In empirical (macro) analysis, high-dimensional observations are often given by a collection of potentially nonstationary variables that are closely related to each other but do not have a natural ordering. In such cases, not only is each observation naturally understood as a realization of a high-dimensional random vector that cannot generally be smoothed to a curve (unlike in our previous example of yield curve data), but also the number of linearly independent I(1) stochastic trends may be substantially smaller than the total number of variables. As discussed, in this case as well, our proposed methodology can be applied to study the number of stochastic trends and/or stationarity without any modifications. To illustrate this versatility of our methodology, we now consider another empirical example analyzing 29 monthly labor market indices provided in the `FRED-MD`

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

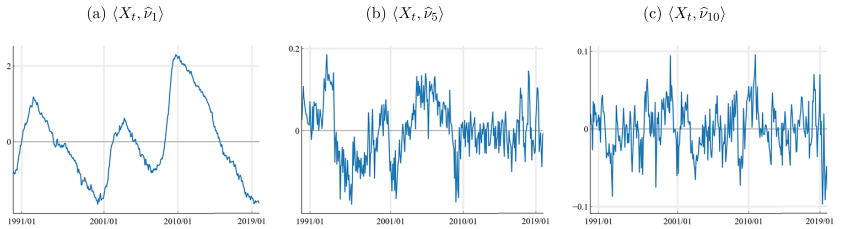


FIGURE 3. Time series of $\langle X_t, \widehat{v}_j \rangle$.
Note: Time series of $\langle X_t, \widehat{v}_j \rangle$ for some selected eigenvectors $\widehat{v}_j$ of $\widehat{\Lambda}_{0,R}$.

data set (McCracken and Ng, 2016). The indices include the civilian labor force, employment/unemployment rate, hourly wage, and other labor-related economic variables; a detailed list can be found in the table labeled "Group 2" in the appendix of McCracken and Ng (2016). We apply log-transformation to the data.[1] The time span used in this analysis is January 1990 to December 2019.

In Figure 3, we display time series of $\langle X_t, \widehat{v}_j \rangle$ for some eigenvectors $\widehat{v}_j$ that are computed from $\widehat{\Lambda}_{0,R}$. These are the j^{th} score processes (see Chang et al., 2016 and Section 5 of Nielsen et al., 2023). It is clear from Figure 3 that there is some nonstationary behavior (e.g., $j = 1$), but this seems to eventually disappear leaving stationary processes (e.g., $j = 10$).

McCracken and Ng (2016) suggested that some of these labor market indices may be I(2) variables.[2] To examine the existence of this type of higher-order stochastic trends, we first display in Figure 4 the first three score processes, $\{\langle \Delta X_t, \widehat{v}_j \rangle\}_{t=1}^T$. Some of these must behave as a unit root process if the original time series is I(2), and that does not appear to be the case. Next, we apply our testing procedures to the first differenced time series, $\{\Delta X_t\}_{t=1}^T$. The bandwidth parameters are set to those applied in Section 7.1, and in this part of the analysis, we use mean-adjusted statistics (Section 3.5). The results are reported in Table 5, where we see that $\ddot{\mathrm{d}} = 2$, $\widehat{\mathrm{d}}_{\text{LRS}} = 1$ (but recall that they cannot be zero), whereas $\widehat{\mathrm{d}}_{\text{BU}} = \widehat{\mathrm{d}}_{\text{TD}} = \widehat{\mathrm{d}}_{\text{UD}} = 0$ regardless of the choice of test statistic and level (except inv.VR at 10% level). From Figure 4 and Table 5, it seems unlikely that there are any I(2) stochastic trends.

We therefore proceed to apply our methodology to the level time series, $\{X_t\}_{t=1}^T$, to determine the number of I(1) stochastic trends. We now correct for a mean and linear time trend. The results are presented in Table 6. For the eigenvalue ratio estimators, we find $\ddot{\mathrm{d}} = \widehat{\mathrm{d}}_{\text{LRS}} = 1$. However, the sequential testing results are quite different. Although the inverse VR test results are reported only for $\mathrm{d}_0 = 6, \ldots, 11$,

[1] The log transformation was suggested by McCracken and Ng (2016) for some variables. We found that the remaining variables are often very large in scale (e.g., "HWI") or have only positive values. Thus, those variables are also log-transformed in our testing procedure.

[2] Specifically, FRED ID variables CES0600000008, CES2000000008, and CES3000000008.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

TABLE 5. Empirical results for labor indices in first differences

Eigenvalue ratio estimates						
Statistic		$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
μ_{j+1}/μ_j		2.06	2.77	2.67	2.32	1.88
$\hat{\kappa}_{j+1}/\hat{\kappa}_j$		0.57	0.82	0.96	0.62	0.70
Variance ratio test statistics						
Statistic	$\mathrm{d}_0=0$	$\mathrm{d}_0=1$	$\mathrm{d}_0=2$	$\mathrm{d}_0=3$	$\mathrm{d}_0=4$	$\mathrm{d}_0=5$
Inv.VR	0.72†	0.07	0.04	0.01	0.01	0.01
VR(2,1)		423.84**	10,347.35**	39,933.38**	101,141.36**	139,311.88**
VR(2,0)		17,922.13**	481,545.37**	812,958.94**	2,993,476.52**	5,194,233.83**
VR(1,0)		117.55**	558.39**	1,291.18**	1,695.46**	3,269.15**

Note: †, *, and ** indicate significance at 10%, 5%, and 1% level, respectively.

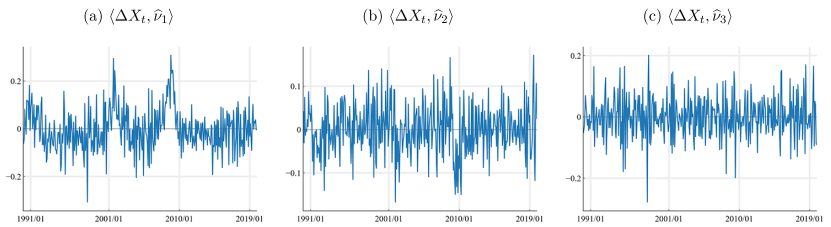


FIGURE 4. Time series of $\langle \Delta X_t, \widehat{\nu}_j \rangle$.
Note: Time series of the first three score functions of ΔX_t. The estimated eigenvectors $\widehat{\nu}_1, \widehat{\nu}_2$, and $\widehat{\nu}_3$ are computed from $\widehat{\Lambda}_{1,R}$.

the null hypotheses $\mathrm{d}_0 = 0, 1, \ldots, 5$ are all rejected at 1% level. Thus, we find that $\widehat{\mathrm{d}}_{\mathrm{BU}} = 10$ at 5% level and $\widehat{\mathrm{d}}_{\mathrm{BU}} = 9$ at 1% level, and we use this to set $\mathrm{d}_{\max} = 15$. We report results for the VR tests for $\mathrm{d}_0 = 6, 7, \ldots, 11$. The larger values of d_0 are rejected at 1% level for VR(2,1) and VR(1,0), but not for VR(2,0). This is consistent with the simulation findings in Table 1. We then find $\widehat{\mathrm{d}}_{\mathrm{UD}} = 8$ based on VR(1,0) at 5% level, $\widehat{\mathrm{d}}_{\mathrm{UD}} = 9$ based on VR(2,1) at 5% level or VR(1,0) at 1% level, and $\widehat{\mathrm{d}}_{\mathrm{UD}} = 10$ based on VR(2,1) at 1% level. On balance, we conclude that there are likely nine stochastic trends. This means that the number of cointegrating relationships is about 2/3 of the dimension of the original data. Given this result, one should be careful in the application of cointegration rank tests that require a small number of cointegrating relationships (Remark 2.2).

Finally, for both empirical applications, we investigate the robustness of $\widehat{\mathrm{d}}_{\mathrm{TD}}$ to the initial hypothesis, $\mathrm{d}_{\max}$, by computing $\widehat{\mathrm{d}}_{\mathrm{TD}}$ using the VR(2,1), VR(2,0), and VR(1,0) tests at 5% level for a range of $\mathrm{d}_{\max}$. In the yield example, $\widehat{\mathrm{d}}_{\mathrm{TD}} = 2$ for $\mathrm{d}_{\max} = 2, \ldots, 20$ using either of the tests. In the labor index example in levels, $\widehat{\mathrm{d}}_{\mathrm{TD}}$ is unchanged for $\mathrm{d}_{\max} = 9, \ldots, 20$ using VR(2,1) and VR(1,0). The VR(2,0) test is

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

TABLE 6. Empirical results for labor indices in levels

Eigenvalue ratio estimates						
Statistic	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
μ_{j+1}/μ_j	2.06	2.02	1.74	1.37	1.34	1.42
$\hat{\kappa}_{j+1}/\hat{\kappa}_j$	0.07	0.27	0.34	0.82	0.80	0.50
Variance ratio test statistics						
Statistic	$\mathrm{d}_0=6$	$\mathrm{d}_0=7$	$\mathrm{d}_0=8$	$\mathrm{d}_0=9$	$\mathrm{d}_0=10$	$\mathrm{d}_0=11$
Inv.VR	0.15**	0.11**	0.06**	0.03*	0.03	0.02
VR(2,1)	2,557.11	4,531.11	6,119.56	10,730.08	16,691.44*	22,714.94**
VR(2,0)	74,227.09	157,057.93	235,784.03	738,700.29	1,581,152.06	2,481,609.80*
VR(1,0)	169.35	229.45	300.84	564.11*	762.54**	898.93**

Note: †, *, and ** indicate significance at 10%, 5%, and 1% level, respectively.

sensitive to $\mathrm{d}_{\max}$ as expected given the simulation results in Table 1. The former two tests are remarkably robust to the choice of $\mathrm{d}_{\max}$.

8. CONCLUSION

We have considered statistical inference on unit roots and cointegration for time series taking values in a Hilbert space of an arbitrarily large dimension or a subspace of possibly unknown dimension. This has wide applicability in practice; for example, in cointegrated vector time series of finite dimension, in a high-dimensional factor model that includes a finite number of nonstationary factors, in cointegrated curve-valued (or function-valued) time series, and in nonstationary dynamic functional factor models. We considered mainly determination of the dimension of the nonstationary subspace (number of common stochastic trends), but we also considered hypothesis testing on the stationary and nonstationary subspaces themselves.

We provided limit theory for general VR-type statistics based on partial summation and/or differencing, and we demonstrated how to apply these VR statistics to sequentially test for the dimension of the nonstationary subspace or to test hypotheses on the subspaces. Our theory covers many possible tests, and choosing among them is not trivial. Our preferred test is the VR(2,1) test due to several theoretical and practical advantages. Most importantly, it does not require the choice of any bandwidth parameters as it can be implemented only with estimated sample variance operators and not long-run variance operators. This is not only a practical advantage, but also implies a consistency property not shared by tests that need to estimate a long-run variance (Müller, 2007, 2008). Also, critical values are readily available (Breitung, 2002) by the isomorphism with $\mathbb{R}^K$. To illustrate our methods, we included a small Monte Carlo simulation study, wherein

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

the VR(2,1) test also showed superior performance, as well as two empirical illustrations to the term structure of interest rates and labor market indices, respectively.

A. APPENDIX A: NOTATION, LINEAR OPERATORS, AND RANDOM ELEMENTS IN $\mathcal{H}$

For a real, separable Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, let $\mathcal{L}_{\mathcal{H}}$ denote the collection of bounded linear operators acting on $\mathcal{H}$ equipped with the uniform operator norm, $\|A\|_{\text{op}} = \sup_{\|v\| \leq 1} \|Av\|$. For any $A \in \mathcal{L}_{\mathcal{H}}$, let $\ker A = \{v \in \mathcal{H} : Av = 0\}$ denote the kernel of A and let $\text{ran} A = \{Av : v \in \mathcal{H}\}$ denote the range of A. The dimensions of $\ker A$ and $\text{ran} A$ are, respectively, called the nullity and rank of A. We let A^* denote the adjoint operator of A, which is uniquely defined by the property $\langle Av_1, v_2 \rangle = \langle v_1, A^* v_2 \rangle$ for all $v_1, v_2 \in \mathcal{H}$. An operator $A \in \mathcal{L}_{\mathcal{H}}$ is compact if it is the limit of a sequence of finite-rank operators in $\mathcal{L}_{\mathcal{H}}$. Moreover, whenever it is convenient, we will let $[A]$ denote the matrix representation for any operator A with respect to some orthonormal set of vectors $\{g_j\}_{j=1}^m$ for some $m > 0$ that will be specified depending on the context; that is, $[A]_{ij} = \langle g_i, Ag_j \rangle$ as discussed in Remark 3.2.

An $\mathcal{H}$-valued random variable X is defined as a measurable map from the underlying probability space, say $\mathcal{S}$, to $\mathcal{H}$, where $\mathcal{H}$ is equipped with the usual Borel σ-field. The random element X is square-integrable if $\mathbb{E}\|X\|^2 < \infty$. For any square-integrable X, its expectation $\mathbb{E}[X]$ is defined as a unique element in $\mathcal{H}$ satisfying $\mathbb{E}\langle X, v \rangle = \langle \mathbb{E}[X], v \rangle$ for any $v \in \mathcal{H}$. The variance operator of X is defined as $C_X = \mathbb{E}[(X - \mathbb{E}[X]) \otimes (X - \mathbb{E}[X])]$, where $\otimes$ denotes the tensor product on $\mathcal{H}$. The operator C_X is self-adjoint (i.e., $C_X = C_X^*$) and allows only nonnegative eigenvalues. We also note the well-known property $\|X_{t-s} \otimes X_t\|_{\text{op}} \leq \|X_{t-s}\| \|X_t\|$.

We will sometimes consider convergence of a sequence of random bounded linear operators. For such a sequence $\{A_j\}_{j \geq 1}$, we write $A_j \to_{\text{P}} A$ if $\|A_j - A\|_{\text{op}} \to_{\text{P}} 0$. We also write $A_j = \mathcal{O}_{\text{P}}(a_n)$ (resp. $A_j = o_{\text{P}}(a_n)$) if $\|A_j\|_{\text{op}} = \mathcal{O}_{\text{P}}(a_n)$ (resp. $\|A_j\|_{\text{op}} = o_{\text{P}}(a_n)$). For example, under Assumption 2, $\widehat{P}_{\text{K}}$ converges in probability and in operator norm to $P_{\text{N}} + P_{\text{S}}^{\text{K}} = P_{\text{K}}$.

When Assumption 1 is satisfied, Beare et al. (2017, Props. 3.1 and 3.2) showed that the long-run variance operator $\Lambda_{\Delta X} (= \Phi(1) C_\epsilon \Phi^*(1))$ of ΔX_t is a finite-rank operator of rank $\mathbb{d}_{\text{N}}$. Because $\Lambda_{\Delta X}$ is also compact and self-adjoint, it allows the spectral decomposition $\Lambda_{\Delta X} = \sum_{j=1}^{\mathbb{d}_{\text{N}}} r_j \varpi_j \otimes \varpi_j$, where $\{r_j, \varpi_j\}_{j=1}^{\mathbb{d}_{\text{N}}}$ denote the pairs of eigenvalues and eigenvectors of $\Lambda_{\Delta X}$ (see Bosq, 2000, pp. 34–35). Obviously, $\{\varpi_j\}_{j=1}^{\mathbb{d}_{\text{N}}}$ is an orthonormal basis of $\mathcal{H}_{\text{N}}$. We then let $\{\mathcal{W}_{1, \mathbb{d}_{\text{N}}}(r)\}_{r \in [0,1]}$ denote the $\mathcal{H}$-valued random element given by $\mathcal{W}_{1, \mathbb{d}_{\text{N}}}(r) = \sum_{j=1}^{\mathbb{d}_{\text{N}}} W_j(r) \varpi_j$ (e.g., Berkes, Horváth, and Rice, 2013, p. 386), where $\{W_j(r)\}_{r \in [0,1]}$ are independent standard Brownian motions. Noting that $\Lambda_{\Delta X}^{1/2}$ is well-defined in terms of its eigenvalues and eigenvectors, $\{r_j^{1/2}, \varpi_j\}_{j=1}^{\mathbb{d}_{\text{N}}}$, we find that $\Lambda_{\Delta X}^{1/2} \mathcal{W}_{1, \mathbb{d}_{\text{N}}}(r) = \sum_{j=1}^{\mathbb{d}_{\text{N}}} r_j^{1/2} W_j(r) \varpi_j$ is $\mathcal{H}$-valued Brownian motion with variance $\Lambda_{\Delta X}$ (e.g., Chen and White, 1998, p. 268). We also recursively define the $(d-1)$-fold integrated Brownian motion $\mathcal{W}_{d, \mathbb{d}_{\text{N}}}(r) = \int_0^r \mathcal{W}_{d-1, \mathbb{d}_{\text{N}}}(u) du$ for $d \geq 2$.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

Under Assumption 1, it is well known that

$$\sup_{0\le r\le 1}\|T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta X_t-\Lambda_{\Delta X}^{1/2}\mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}}(r)\|\xrightarrow{\mathrm{P}}0 \tag{A.1}$$

(see Berkes et al., 2013, Theorem 1. We will repeatedly, and sometimes implicitly, use this result in the subsequent sections. Moreover, in the case where the unobserved components model (3.21) is considered, and thus $U_t^{(1)}$ or $U_t^{(2)}$ is used in the analysis, we note that

$$\sup_{0\le r\le 1}\|T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta U_t^{(1)}-\Lambda_{\Delta X}^{1/2}\mathcal{W}^{(1)}_{1,\mathbb{d}_{\mathsf{N}}}(r)\|\xrightarrow{\mathrm{P}}0, \tag{A.2}$$

$$\sup_{0\le r\le 1}\|T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta U_t^{(2)}-\Lambda_{\Delta X}^{1/2}\mathcal{W}^{(2)}_{1,\mathbb{d}_{\mathsf{N}}}(r)\|\xrightarrow{\mathrm{P}}0, \tag{A.3}$$

where $\mathcal{W}^{(1)}_{1,\mathbb{d}_{\mathsf{N}}}$ (resp. $\mathcal{W}^{(2)}_{1,\mathbb{d}_{\mathsf{N}}}$) is the demeaned (resp. detrended) standard Brownian motion taking values in $\mathcal{H}_{\mathsf{N}}$ (see Nielsen et al., 2023, Lemma 3. Results (A.2) and (A.3) are essential to extend our theoretical results to the unobserved components model (3.21).

In the sequel, we will often consider operators that can be written as $\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}$ for some $A\in\mathcal{L}_{\mathcal{H}}$. Let $\mathcal{H}_{\mathrm{K}}=\operatorname{ran}\widehat{P}_{\mathrm{K}}$. Given that $\widehat{P}_{\mathrm{K}}v=0$ and $\langle\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}v,w\rangle=0$ for any $v,w\in\mathcal{H}_{\mathrm{K}}^{\perp}$, $\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}$ can be understood as a map on $\mathcal{H}_{\mathrm{K}}$. Let Q_1 be any orthogonal projection on $\mathcal{H}$ and let $Q_2=I-Q_1$. Then it holds that $\mathcal{H}_{\mathrm{K}}=\mathcal{H}_{\mathrm{K},1}\oplus\mathcal{H}_{\mathrm{K},2}$, where $\mathcal{H}_{\mathrm{K},1}=[\ker Q_1\widehat{P}_{\mathrm{K}}]^{\perp}$ and $\mathcal{H}_{\mathrm{K},2}=\ker Q_1\widehat{P}_{\mathrm{K}}\cap\mathcal{H}_{\mathrm{K}}$. Note that $\ker Q_1\widehat{P}_{\mathrm{K}}$ includes $\mathcal{H}_{\mathrm{K}}^{\perp}=[\operatorname{ran}\widehat{P}_{\mathrm{K}}]^{\perp}$, and hence it needs to be restricted to be a subspace of $\mathcal{H}_{\mathrm{K}}$, whereas $[\ker Q_1\widehat{P}_{\mathrm{K}}]^{\perp}=\operatorname{ran}\widehat{P}_{\mathrm{K}}Q_1$ (Conway, 1990, Theorem 2.19) and thus is already a subspace of $\mathcal{H}_{\mathrm{K}}$. It follows that any $v\in\mathcal{H}_{\mathrm{K}}$ can be written as $v=v_1+v_2$, where $v_1\in\mathcal{H}_{\mathrm{K},1}$ and $v_2\in\mathcal{H}_{\mathrm{K},2}$, and we may understand v as the tuple $\binom{v_1}{v_2}$. Let $\widehat{\Pi}$ be the orthogonal projection onto $\mathcal{H}_{\mathrm{K},2}$ (i.e., $[\operatorname{ran}\widehat{P}_{\mathrm{K}}Q_1]^{\perp}=\ker Q_1\widehat{P}_{\mathrm{K}}$ in $\mathcal{H}_{\mathrm{K}}$) such that $(I-\widehat{\Pi})\widehat{P}_{\mathrm{K}}Q_1=\widehat{P}_{\mathrm{K}}Q_1=\mathcal{H}_{\mathrm{K},1}$. Then any $w\in\mathcal{H}_{\mathrm{K}}$ can be written as $(I-\widehat{\Pi})w+\widehat{\Pi}w$, where $(I-\widehat{\Pi})w\in\mathcal{H}_{\mathrm{K},1}$ and $\widehat{\Pi}w\in\mathcal{H}_{\mathrm{K},2}$, and thus w may be understood as the tuple $\binom{(I-\widehat{\Pi})w}{\widehat{\Pi}w}$. From this result and properties of $\widehat{\Pi}$, we may understand $w=\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}v=\widehat{P}_{\mathrm{K}}Q_1AQ_1\widehat{P}_{\mathrm{K}}v_1+\widehat{P}_{\mathrm{K}}Q_1AQ_2\widehat{P}_{\mathrm{K}}v_2+\widehat{P}_{\mathrm{K}}Q_2AQ_1\widehat{P}_{\mathrm{K}}v_1+\widehat{P}_{\mathrm{K}}Q_2AQ_2\widehat{P}_{\mathrm{K}}v_2$ as the tuple $\binom{w_1}{w_2}$, where $w_1=(I-\widehat{\Pi})w=(\widehat{P}_{\mathrm{K}}Q_1AQ_1\widehat{P}_{\mathrm{K}}+\widehat{\Upsilon}_1)v_1+(\widehat{P}_{\mathrm{K}}Q_1AQ_2\widehat{P}_{\mathrm{K}}+\widehat{\Upsilon}_2)v_2$ and $w_2=\widehat{\Pi}w=\widehat{\Pi}\widehat{P}_{\mathrm{K}}Q_2AQ_1\widehat{P}_{\mathrm{K}}v_1+\widehat{\Pi}\widehat{P}_{\mathrm{K}}Q_2AQ_2\widehat{P}_{\mathrm{K}}v_2$ with $\widehat{\Upsilon}_1=(I-\widehat{\Pi})\widehat{P}_{\mathrm{K}}Q_2AQ_1\widehat{P}_{\mathrm{K}}$ and $\widehat{\Upsilon}_2=(I-\widehat{\Pi})\widehat{P}_{\mathrm{K}}Q_2AQ_2\widehat{P}_{\mathrm{K}}$. To sum up, the operator $\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}$ on $\mathcal{H}_{\mathrm{K}}$ may be understood as a transformation via the (operator) matrix $\mathcal{A}$, viewing the aforementioned tuples as vectors:

$$\begin{pmatrix}w_1\\w_2\end{pmatrix}=\mathcal{A}\begin{pmatrix}v_1\\v_2\end{pmatrix},\quad\text{where}\quad\mathcal{A}=\begin{pmatrix}\widehat{P}_{\mathrm{K}}Q_1AQ_1\widehat{P}_{\mathrm{K}}+\widehat{\Upsilon}_1 & \widehat{P}_{\mathrm{K}}Q_1AQ_2\widehat{P}_{\mathrm{K}}+\widehat{\Upsilon}_2\\ \widehat{\Pi}\widehat{P}_{\mathrm{K}}Q_2AQ_1\widehat{P}_{\mathrm{K}} & \widehat{\Pi}\widehat{P}_{\mathrm{K}}Q_2AQ_2\widehat{P}_{\mathrm{K}}\end{pmatrix}. \tag{A.4}$$

Note that $\mathcal{A}$ is a standard characterization of the map $\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}$, viewed as a map from $\mathcal{H}_{\mathrm{K}}$ to itself when $\mathcal{H}_{\mathrm{K}}$ allows the bipartite decomposition $\mathcal{H}_{\mathrm{K}}=\mathcal{H}_{\mathrm{K},1}\oplus\mathcal{H}_{\mathrm{K},2}$. Obviously, the (i,j)-th entry of $\mathcal{A}$ may be understood as a map from $\mathcal{H}_{\mathrm{K},j}$ to $\mathcal{H}_{\mathrm{K},i}$ for $i,j=1,2$.

For any operator $\widehat{P}_{\mathrm{K}}A\widehat{P}_{\mathrm{K}}$ whose operator matrix form is $\begin{pmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{pmatrix}$, such as $\mathcal{A}$ in (A.4), we let D_m be the diagonal operator corresponding to the operator matrix $\begin{pmatrix}m^{-1/2}I_{\dim(\mathcal{H}_{\mathrm{K},1})}&0\\0&I_{\dim(\mathcal{H}_{\mathrm{K},2})}\end{pmatrix}$ for appropriately defined identity operators $I_{\dim(\mathcal{H}_{\mathrm{K},1})}$ on $\mathcal{H}_{\mathrm{K},1}$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

and $I_{\dim(\mathcal{H}_{K,2})}$ on $\mathcal{H}_{K,2}$ so that $D_m\widehat{P}_K A\widehat{P}_K$ and $\widehat{P}_K A\widehat{P}_K D_m$, respectively, result in the operator matrices

$$\begin{pmatrix} m^{-1/2}A_{11} & m^{-1/2}A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} m^{-1/2}A_{11} & A_{12} \\ m^{-1/2}A_{21} & A_{22} \end{pmatrix}. \tag{A.5}$$

Subsequent proofs with such a D_m will often have $\dim(\mathcal{H}_{K,1}) = \mathbb{d}_N$ and $\dim(\mathcal{H}_{K,2}) = K - \mathbb{d}_N$.

B. APPENDIX B: PROOFS OF THEOREMS

In this appendix, we give the proofs of all theorems. The corollaries follow by the continuous mapping theorem. We first state a lemma, which is proven in the Supplementary Material.

LEMMA B.1. *Under Assumptions 1–3, the following hold:*

(i) $P_N\widehat{\Lambda}_{d,m}P_N = \mathcal{O}_P(h_m T^{2d})$ *for* $d \geq 1$ *and* $P_N\widehat{\Lambda}_{0,m}P_N = \mathcal{O}_P(T)$.
(ii) $P_N\widehat{\Lambda}_{d,m}P_S^K = \mathcal{O}_P(h_m T^{2d-1})$ *for* $d \geq 1$ *and* $P_N\widehat{\Lambda}_{0,m}P_S^K = \mathcal{O}_P(T/h_m)$.
(iii) $P_S^K\widehat{\Lambda}_{d,m}P_S^K = \mathcal{O}_P(h_m T^{2d-2})$ *for* $d \geq 2$, $P_S^K\widehat{\Lambda}_{1,m}P_S^K = \mathcal{O}_P(T)$, *and* $P_S^K\widehat{\Lambda}_{0,m}P_S^K = \mathcal{O}_P(T/h_m^2)$.

Finally, of course, $P_S^K\widehat{\Lambda}_{d,m}P_N$ *is the same order as* $P_N\widehat{\Lambda}_{d,m}P_S^K$ *for* $d \geq 0$.

B.1. Proof of Theorem 3.1

We only give the proof for the case with $a_m > 0$ for $m \in \{L, R\}$. If $a_m = 0$, such that $\widehat{\Lambda}_{d_m,m}$ is a sample variance operator, then we set $h_m = 1$ and $k_m(u) = 1_{\{u=0\}}$ and use the same proof.

First note that the eigenvalues μ_j from the problem (3.4) with $d_L \geq 2$ and $d_R = 1$ are identical to those in

$$(T^{2d_L-2}h_L h_R^{-1}\mu_j)\left(\frac{h_R}{h_L T^{2d_L-1}}D_{h_R T}\widehat{P}_K\widehat{\Lambda}_{2,L}\widehat{P}_K D_{h_R T}\right)v_j = \left(\frac{1}{T}D_{h_R T}\widehat{P}_K\widehat{\Lambda}_{1,R}\widehat{P}_K D_{h_R T}\right)v_j, \tag{B.1}$$

where $D_{h_R T}$ is defined as in (A.5). Next, consider the decomposition $\widehat{P}_K = \widehat{P}_K P_N + \widehat{P}_K P_S = \widehat{P}_N^K + \widehat{P}_S^K$. Then, for $d \geq 1$ and $m \in \{L, R\}$,

$$\widehat{P}_K\widehat{\Lambda}_{d,m}\widehat{P}_K = \widehat{P}_N^K\widehat{\Lambda}_{d,m}(\widehat{P}_N^K)^* + \widehat{P}_S^K\widehat{\Lambda}_{d,m}(\widehat{P}_N^K)^* + \widehat{P}_N^K\widehat{\Lambda}_{d,m}(\widehat{P}_S^K)^* + \widehat{P}_S^K\widehat{\Lambda}_{d,m}(\widehat{P}_S^K)^*. \tag{B.2}$$

As in Appendix A, $\widehat{P}_K\widehat{\Lambda}_{d,m}\widehat{P}_K$ may be understood as the operator matrix given by

$$\begin{pmatrix} \widehat{P}_N^K\widehat{\Lambda}_{d,m}(\widehat{P}_N^K)^* + \widehat{\Upsilon}_{d,m}^N & \widehat{P}_N^K\widehat{\Lambda}_{d,m}(\widehat{P}_S^K)^* + \widehat{\Upsilon}_{d,m}^S \\ \widehat{\Pi}\widehat{P}_S^K\widehat{\Lambda}_{d,m}(\widehat{P}_N^K)^* & \widehat{\Pi}\widehat{P}_S^K\widehat{\Lambda}_{d,m}(\widehat{P}_S^K)^* \end{pmatrix}, \tag{B.3}$$

where $\widehat{\Pi}$ is defined so that $I - \widehat{\Pi}$ becomes the orthogonal projection onto $\operatorname{ran}\widehat{P}_N^K$. Hence,

$$\widehat{\Upsilon}_{d,m}^N = (I - \widehat{\Pi})\widehat{P}_S^K\widehat{\Lambda}_{d,m}(\widehat{P}_N^K)^* \quad \text{and} \quad \widehat{\Upsilon}_{d,m}^S = (I - \widehat{\Pi})\widehat{P}_S^K\widehat{\Lambda}_{d,m}(\widehat{P}_S^K)^*, \tag{B.4}$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

see (A.4), where we replaced A, Q_1, and Q_2 with $\widehat{\Lambda}_{d,m}$, P_{N}, and P_{S}, respectively. Noting that $\|\widehat{P}^{\mathsf{K}}_{\mathsf{N}} - P_{\mathsf{N}}\|_{\mathrm{op}} = \|\widehat{P}_{\mathsf{K}} P_{\mathsf{N}} - P_{\mathsf{N}}\|_{\mathrm{op}} \to_{\mathrm{P}} 0$ and $\|A\|_{\mathrm{op}} = \|A^*\|_{\mathrm{op}}$ for any $A \in \mathcal{L}_{\mathcal{H}}$, it follows that $\|(I - \widehat{\Pi})P_{\mathsf{S}}\|_{\mathrm{op}} = \|P_{\mathsf{S}}(I - \widehat{\Pi})\|_{\mathrm{op}} = o_{\mathrm{P}}(1)$, and hence $\|\widehat{\Pi}P_{\mathsf{S}} - P_{\mathsf{S}}\|_{\mathrm{op}} = o_{\mathrm{P}}(1)$ and also

$$\|\widehat{\Pi}P^{\mathsf{K}}_{\mathsf{S}} - P^{\mathsf{K}}_{\mathsf{S}}\|_{\mathrm{op}} = o_{\mathrm{P}}(1), \tag{B.5}$$

where $P^{\mathsf{K}}_{\mathsf{S}}$ is defined in Assumption 2(iii). We next obtain the limits of the operators in (B.1).

First, $h_L^{-1}T^{1-2d_L}h_R D_{h_R T}\widehat{P}_{\mathsf{K}}\widehat{\Lambda}_{2,L}\widehat{P}_{\mathsf{K}}D_{h_R T}$ is understood as the operator matrix (Appendix A)

$$\widehat{\mathfrak{D}}_L = \begin{pmatrix} h_L^{-1}T^{-2d_L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* + \widehat{\Upsilon}^{\mathsf{N}}_{d_L,L}) & h_L^{-1}h_R^{1/2}T^{1/2-2d_L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* + \widehat{\Upsilon}^{\mathsf{S}}_{d_L,L}) \\ h_L^{-1}h_R^{1/2}T^{1/2-2d_L}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* & h_L^{-1}h_R T^{1-2d_L}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* \end{pmatrix}.$$

Let $\widehat{\delta}_{\mathsf{K}} = \widehat{P}_{\mathsf{K}} - P_{\mathsf{K}}$, where $P_{\mathsf{K}} = P_{\mathsf{N}} + P^{\mathsf{K}}_{\mathsf{S}}$. Using the fact that $P_{\mathsf{K}}P_{\mathsf{N}} = P_{\mathsf{N}}$ and the definition $\widehat{P}^{\mathsf{K}}_{\mathsf{N}} = \widehat{P}_{\mathsf{K}}P_{\mathsf{N}}$, we find $\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* = P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}} + \widehat{\delta}_{\mathsf{K}}P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}}\widehat{P}_{\mathsf{K}} + P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}}\widehat{\delta}^*_{\mathsf{K}}$, where Assumption 2 implies $\widehat{\delta}_{\mathsf{K}} = o_{\mathrm{P}}(1)$ and $\widehat{P}_{\mathsf{K}} = \mathcal{O}_{\mathrm{P}}(1)$. Combining these results with Lemma B.1, we find that

$$\begin{aligned} \frac{1}{h_L T^{2d_L}}\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* &= \frac{1}{h_L T^{2d_L}}P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}} + o_{\mathrm{P}}(1) \\ &= \frac{1}{h_L T^{2d_L}}\sum_{s=1-T}^{T-1}\mathrm{k}_L\left(\frac{s}{h_L}\right)\widehat{\Gamma}^{\mathsf{N}}_{d_L,s} + o_{\mathrm{P}}(1) \\ &\xrightarrow{\mathrm{d}} c_L \Lambda^{1/2}_{\Delta X}\left(\int \mathcal{W}_{d_L,\mathbb{d}_{\mathsf{N}}} \otimes \mathcal{W}_{d_L,\mathbb{d}_{\mathsf{N}}}\right)\Lambda^{1/2}_{\Delta X}, \end{aligned} \tag{B.6}$$

where $\widehat{\Gamma}^{\mathsf{N}}_{d,s} = \sum_{t=s+1}^{T} P_{\mathsf{N}}X_{d,t-s} \otimes P_{\mathsf{N}}X_{d,t}$ for $s \geq 0$ and $\widehat{\Gamma}^{\mathsf{N}}_{d,s} = \sum_{t=s+1}^{T} P_{\mathsf{N}}X_{d,t} \otimes P_{\mathsf{N}}X_{d,t-s}$ for $s < 0$. The first equality in (B.6) tells us that the limiting behavior of $h_L^{-1}T^{-2d_L}\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^*$ is the same as that of the finite-rank operator $h_L^{-1}T^{-2d_L}$ $P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}}$. The second equality follows by definition of $P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}}$. Due to the isomorphism between $\mathbb{R}^{\mathbb{d}_{\mathsf{N}}}$ and the $\mathbb{d}_{\mathsf{N}}$-dimensional subspace $\mathcal{H}_{\mathsf{N}} = \mathrm{ran}\,P_{\mathsf{N}}$, $P_{\mathsf{N}}X_t$ can be understood as the $\mathbb{d}_{\mathsf{N}}$-dimensional vector $x_{\mathbb{d}_{\mathsf{N}},t} = (\langle X_t, f_1\rangle, \ldots \langle X_t, f_{\mathbb{d}_{\mathsf{N}}}\rangle)'$, where $\{f_j\}_{j=1}^{\mathbb{d}_{\mathsf{N}}}$ is any orthonormal basis of $\mathcal{H}_{\mathsf{N}}$ (see Remark 3.2). Because $\mathbb{d}_{\mathsf{N}}$ is a fixed constant, we then deduce the convergence in distribution given in (B.6) from (A.1), Phillips (1991, unnumbered equation between (A.10) and (A.11)) and the continuous mapping theorem (applied to the integral functional on $[0, 1]$).

By similar arguments, using Assumption 2, $\widehat{\delta}_{\mathsf{K}} = o_{\mathrm{P}}(1)$, (B.5), Lemma B.1, and the fact that $h_R/T \to 0$ (Assumption 3), we can further show that

$$h_L^{-1}h_R^{1/2}T^{1/2-2d_L}\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* = h_L^{-1}h_R^{1/2}T^{1/2-2d_L}P_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}P^{\mathsf{K}}_{\mathsf{S}} + o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1), \tag{B.7}$$

$$h_L^{-1}h_R^{1/2}T^{1/2-2d_L}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* = h_L^{-1}h_R^{1/2}T^{1/2-2d_L}P^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}P_{\mathsf{N}} + o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1), \tag{B.8}$$

$$h_L^{-1}h_R T^{1-2d_L}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* = h_L^{-1}h_R T^{1-2d_L}P^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}P^{\mathsf{K}}_{\mathsf{S}} + o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1). \tag{B.9}$$

From the latter two equations together with $\widehat{\delta}_{\mathsf{K}} = o_{\mathrm{P}}(1)$, we also find that

$$h_L^{-1}T^{-2d_L}\widehat{\Upsilon}^{\mathsf{N}}_{d_L,L} = h_L^{-1}T^{-2d_L}(I - \widehat{\Pi})\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* = o_{\mathrm{P}}(h_R^{-1/2}T^{-1/2}), \tag{B.10}$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

$$h_L^{-1}h_R^{1/2}T^{1/2-2d_L}\widehat{\Upsilon}^{\mathsf{S}}_{d_L,L} = h_L^{-1}h_R^{1/2}T^{1/2-2d_L}(I-\widehat{\Pi})\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* = \mathcal{O}_{\mathrm{P}}(h_R^{-1/2}T^{-1/2}). \tag{B.11}$$

Second, $T^{-1}D_{h_RT}\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{1,R}\widehat{P}_{\mathrm{K}}D_{h_RT}$ can be understood as (see Appendix A)

$$\widehat{\mathfrak{D}}_R = \begin{pmatrix} h_R^{-1}T^{-2}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* + \widehat{\Upsilon}^{\mathsf{N}}_{1,R}) & h_R^{-1/2}T^{-3/2}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* + \widehat{\Upsilon}^{\mathsf{S}}_{1,R}) \\ h_R^{-1/2}T^{-3/2}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* & T^{-1}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* \end{pmatrix}.$$

As in (B.6)–(B.11), it holds that

$$h_R^{-1}T^{-2}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* + \widehat{\Upsilon}^{\mathsf{N}}_{1,R}) \xrightarrow{\mathrm{d}} c_R\Lambda_{\Delta X}^{1/2}\left(\int \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}}\otimes\mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}}\right)\Lambda_{\Delta X}^{1/2}, \tag{B.12}$$

$$h_R^{-1/2}T^{-3/2}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* + \widehat{\Upsilon}^{\mathsf{S}}_{1,R}) = \mathcal{O}_{\mathrm{P}}(1), \tag{B.13}$$

$$h_R^{-1/2}T^{-3/2}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{N}})^* = \mathcal{O}_{\mathrm{P}}(1). \tag{B.14}$$

We also find that

$$T^{-1}\widehat{\Pi}\widehat{P}^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathsf{K}}_{\mathsf{S}})^* = T^{-1}P^{\mathsf{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}P^{\mathsf{K}}_{\mathsf{S}} + \mathcal{O}_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} \Lambda^{\mathsf{K}}_{\mathsf{S},R}, \tag{B.15}$$

where $\Lambda^{\mathsf{K}}_{\mathsf{S},R} = P_{\mathrm{K}}\Lambda_{\mathsf{S},R}P_{\mathrm{K}}$ and the convergence follows from Theorem 2 of Horváth, Kokoszka, and Reeder (2013) (also see Theorem 5.3 of Kokoszka and Young, 2016) since $\{P_{\mathsf{S}}X_{1,t}\}_{t\geq 1}$ is an L^4-q-approximable sequence (Hörmann and Kokoszka, 2010, Prop. 2.1) with $q\sum_{j=q+1}^{\infty}\|\tilde{\Phi}_j\|_{\mathrm{op}} = \mathcal{O}(1)$ under the summability condition $\sum_{j=0}^{\infty}j\|\tilde{\Phi}_j\|_{\mathrm{op}} < \infty$ implied by (2.2).

From (B.6)–(B.15), $\widehat{\mathfrak{D}}_L$ and $\widehat{\mathfrak{D}}_R$ converge in distribution to random bounded linear operators acting on $\operatorname{ran}P_{\mathrm{K}}$. Specifically,

$$\widehat{\mathfrak{D}}_L \xrightarrow{\mathrm{d}} \mathfrak{D}_L = \begin{pmatrix} c_L\Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{d_L,\mathbb{d}_{\mathsf{N}}}\otimes\mathcal{W}_{d_L,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \text{ and} \tag{B.16}$$

$$\widehat{\mathfrak{D}}_R \xrightarrow{\mathrm{d}} \mathfrak{D}_R = \begin{pmatrix} c_R\Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}}\otimes\mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2} & 0 \\ 0 & \Lambda^{\mathsf{K}}_{\mathsf{S},R} \end{pmatrix}. \tag{B.17}$$

Using the isomorphism between $\mathbb{R}^{\mathrm{K}}$ and any K-dimensional Hilbert space, $\mathfrak{D}_L$ and $\mathfrak{D}_R$ may be understood as $\mathrm{K}\times\mathrm{K}$ matrices, say $[\mathfrak{D}_L]$ and $[\mathfrak{D}_R]$ (see Remark 3.2). Moreover, any orthonormal basis of $\operatorname{ran}P_{\mathrm{K}}$ can be used for these matrix representations since the eigenvalues that we are interested in are not affected by a change of basis. Hence, we may assume that $\{T^{2d_L-2}h_Lh_R^{-1}\mu_j\}_{j=1}^{\mathrm{K}}$ in (B.1) converge to the eigenvalues associated with $[\mathfrak{D}_L]$ and $[\mathfrak{D}_R]$, and these are represented with respect to some orthonormal basis $\{f_j\}_{j=1}^{\mathrm{K}}$ of $\operatorname{ran}P_{\mathrm{K}}$. Under Assumption 2, $[\Lambda^{\mathsf{K}}_{\mathsf{S},R}]$ has rank $\mathrm{K}-\mathbb{d}_{\mathsf{N}}$ and hence $[\mathfrak{D}_R]$ is invertible almost surely. We thus find that $\{(T^{2d_L-2}h_Lh_R^{-1}\mu_j)^{-1}\}_{j=1}^{\mathbb{d}_{\mathsf{N}}}$ converge to the $\mathbb{d}_{\mathsf{N}}$ largest eigenvalues of $[\mathfrak{D}_R]^{-1}[\mathfrak{D}_L]$ while $(T^{2d_L-2}h_Lh_R^{-1}\mu_j)^{-1}\to 0$ for $j\geq\mathbb{d}_{\mathsf{N}}+1$. This proves (3.8). From the expressions for $[\mathfrak{D}_L]$ and $[\mathfrak{D}_R]$, we further find that the $\mathbb{d}_{\mathsf{N}}$ largest eigenvalues of $[\mathfrak{D}_R]^{-1}[\mathfrak{D}_L]$ are distributionally identical to those of $(c_L/c_R)[\Lambda_{\Delta X}^{1/2}(\int\mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}}\otimes\mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2}]^{-1}[\Lambda_{\Delta X}^{1/2}(\int\mathcal{W}_{d_L,\mathbb{d}_{\mathsf{N}}}\otimes\mathcal{W}_{d_L,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2}]$. Then (3.7) follows from the continuous mapping theorem.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

B.2. Proof of Theorem 3.2

We consider only the case with $a_L > 0$. The proof can easily be modified to deal with $a_L = 0$.

As in the proof of Theorem 3.1, the eigenvalues μ_j from (3.4) with $d_L \geq 1$ and $d_R = 0$ are identical to those in

$$(h_L T^{2d_L-1}\mu_j)(T^{-2d_L}h_L^{-1}h_R^2 D_{h_R}^2 \widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}\widehat{P}_{\mathrm{K}}D_{h_R}^2)v_j = (T^{-1}h_R^2 D_{h_R}^2 \widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{0,R}\widehat{P}_{\mathrm{K}}D_{h_R}^2)v_j. \tag{B.18}$$

Note that $h_R^2 h_L^{-1}T^{-2d_L}D_{h_R}^2\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}\widehat{P}_{\mathrm{K}}D_{h_R}^2$ and $h_R^2T^{-1}D_{h_R}^2\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{0,R}\widehat{P}_{\mathrm{K}}D_{h_R}^2$ may, respectively, be understood as (see Appendix A)

$$\widehat{\mathfrak{D}}_L^{(1)} = \begin{pmatrix} h_L^{-1}T^{-2d_L}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{d_L,L}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* + \widehat{\Upsilon}_{d_L,L}^{\mathsf{N}}) & h_L^{-1}h_R T^{-2d_L}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{d_L,L}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* + \widehat{\Upsilon}_{d_L,L}^{\mathsf{S}}) \\ h_L^{-1}h_R T^{-2d_L}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{d_L,L}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* & h_L^{-1}h_R^2 T^{-2d_L}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{d_L,L}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* \end{pmatrix}, \tag{B.19}$$

$$\widehat{\mathfrak{D}}_R^{(1)} = \begin{pmatrix} T^{-1}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* + \widehat{\Upsilon}_{0,R}^{\mathsf{N}}) & h_R T^{-1}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* + \widehat{\Upsilon}_{0,R}^{\mathsf{S}}) \\ h_R T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* & h_R^2 T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* \end{pmatrix}, \tag{B.20}$$

where $\widehat{\Upsilon}_{d,m}^{\mathsf{N}}$ and $\widehat{\Upsilon}_{d,m}^{\mathsf{S}}$ are defined as in (B.4).

We first obtain the limiting behavior of the operator in (B.19). When $d_L \geq 2$ this is given in (B.16). When $d_L = 1$, we reverse the roles of L and R, use (B.12)–(B.15) and the facts that $h_R T^{-1/2} \to 0$ and $h_L^{-1}h_R^2T^{-1} \to 0$ (Assumption 3), and (B.16) still applies. Combining these results, $\widehat{\mathfrak{D}}_L^{(1)} \to_{\mathrm{d}} \mathfrak{D}_L$ for $d_L \geq 1$, where $\mathfrak{D}_L$ is given in (B.16).

We next establish the limiting behavior of the operator given in (B.20). Assumption 2, Lemma B.1, and the fact that $\{P_{\mathsf{N}}X_{0,t}\}_{t\geq 1}$ is L^4-q-approximable with $q\sum_{j=q+1}^{\infty}\|\Phi_j\|_{\mathrm{op}} = \mathcal{O}(1)$ imply that $T^{-1}\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* = T^{-1}P_{\mathsf{N}}\widehat{\Lambda}_{0,R}P_{\mathsf{N}} + \mathcal{O}_{\mathrm{P}}(1)$ and that $T^{-1}P_{\mathsf{N}}\widehat{\Lambda}_{0,R}P_{\mathsf{N}}$ converges to the long-run variance of $\{P_{\mathsf{N}}X_{0,t}\}_{t\geq 1}$, which is $\Lambda_{\Delta X}$ (Beare et al., 2017, Sect. 3.1). Thus,

$$T^{-1}\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* \xrightarrow{\mathrm{P}} \Lambda_{\Delta X}. \tag{B.21}$$

Similarly, from Assumption 2, Lemma B.1, and (B.5), we know that $h_R^2T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* = h_R^2T^{-1}P_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}P_{\mathsf{S}}^{\mathrm{K}} + \mathcal{O}_{\mathrm{P}}(1)$, where $P_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}P_{\mathsf{S}}^{\mathrm{K}}$ is the sample long-run variance computed from $\{P_{\mathsf{S}}^{\mathrm{K}}\Delta X_t\}_{t=1}^T$, which may be understood as a finite-dimensional vector-valued process (Remark 3.2). Then, from similar arguments used in the proof of Lemma 8.1(a) of Phillips (1995), we may deduce that $h_R^2T^{-1}P_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}P_{\mathsf{S}}^{\mathrm{K}} \to_{\mathrm{P}} -\mathrm{k}_R''(0)P_{\mathsf{S}}^{\mathrm{K}}\Lambda_{\mathsf{S},R}P_{\mathsf{S}}^{\mathrm{K}}$. Since $P_{\mathsf{S}}^{\mathrm{K}}\Lambda_{\mathsf{S},R}P_{\mathsf{S}}^{\mathrm{K}} = \Lambda_{\mathsf{S},R}^{\mathrm{K}}$, which is defined in our proof of Theorem 3.1:

$$h_R^2T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* \xrightarrow{\mathrm{P}} -\mathrm{k}_R''(0)\Lambda_{\mathsf{S},R}^{\mathrm{K}}. \tag{B.22}$$

In the same way, from Assumption 2, Lemma B.1, and (B.5), we find that $h_R T^{-1}\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* = h_R T^{-1}P_{\mathsf{N}}\widehat{\Lambda}_{0,R}P_{\mathsf{S}}^{\mathrm{K}} + \mathcal{O}_{\mathrm{P}}(1)$ and $h_R T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* = h_R T^{-1}P_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}P_{\mathsf{N}} + \mathcal{O}_{\mathrm{P}}(1)$, and by the same arguments used in the proof of Lemma 8.1(b) of Phillips (1995), both $P_{\mathsf{N}}\widehat{\Lambda}_{0,R}P_{\mathsf{S}}^{\mathrm{K}}$ and $P_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}P_{\mathsf{N}}$ are $\mathcal{O}_{\mathrm{P}}(h_R^{-1}) + \mathcal{O}_{\mathrm{P}}((h_R/T)^{1/2})$, from which we find that

$$h_R T^{-1}\widehat{P}_{\mathsf{N}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* = \mathcal{O}_{\mathrm{P}}(h_R^{-1}) + \mathcal{O}_{\mathrm{P}}((h_R/T)^{1/2}) + \mathcal{O}_{\mathrm{P}}(1) = \mathcal{O}_{\mathrm{P}}(1), \tag{B.23}$$

$$h_R T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{0,R}(\widehat{P}_{\mathsf{N}}^{\mathrm{K}})^* = \mathcal{O}_{\mathrm{P}}(h_R^{-1}) + \mathcal{O}_{\mathrm{P}}((h_R/T)^{1/2}) + \mathcal{O}_{\mathrm{P}}(1) = \mathcal{O}_{\mathrm{P}}(1). \tag{B.24}$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

From (B.22), (B.24), and the fact that $(I - \widehat{\Pi})\widehat{P}^{\mathrm{K}}_{\mathsf{S}} = \mathcal{O}_{\mathrm{P}}(1)$ (by (B.5) and $\widehat{\delta}_{\mathrm{K}} \to_{\mathrm{P}} 0$), we have

$$T^{-1}\widehat{\Upsilon}^{\mathsf{N}}_{0,R} = T^{-1}(I - \widehat{\Pi})\widehat{P}^{\mathrm{K}}_{\mathsf{S}}\widehat{\Lambda}_{0,R}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}})^* = \mathcal{O}_{\mathrm{P}}(h_R^{-1}), \tag{B.25}$$

$$h_R T^{-1}\widehat{\Upsilon}^{\mathsf{S}}_{0,R} = h_R T^{-1}(I - \widehat{\Pi})\widehat{P}^{\mathrm{K}}_{\mathsf{S}}\widehat{\Lambda}_{0,R}(\widehat{P}^{\mathrm{K}}_{\mathsf{S}})^* = \mathcal{O}_{\mathrm{P}}(h_R^{-1}). \tag{B.26}$$

Combining (B.21)–(B.26), we find that

$$\widehat{\mathfrak{D}}^{(1)}_R \xrightarrow{\mathrm{P}} \mathfrak{D}^{(1)}_R = \begin{pmatrix} \Lambda_{\Delta X} & 0 \\ 0 & -\mathrm{k}''_R(0)\Lambda^{\mathrm{K}}_{\mathsf{S},R} \end{pmatrix}. \tag{B.27}$$

As in our proof of Theorem 3.1, we may assume that $\{h_L T^{2d_L-1}\mu_j\}_{j=1}^{\mathrm{K}}$ converge in distribution to the eigenvalues associated with $[\mathfrak{D}^{(1)}_L]$ and $[\mathfrak{D}^{(1)}_R]$, which are matrix representations of $\mathfrak{D}^{(1)}_L$ and $\mathfrak{D}^{(1)}_R$ with respect to some orthonormal basis of $\operatorname{ran} P_{\mathrm{K}}$. Because $[\mathfrak{D}^{(1)}_R]$ is invertible under Assumption 2, the desired results follow from nearly identical arguments used to conclude our proof of Theorem 3.1.

B.3. Proof of Theorem 3.3

As in our proof of Theorem 3.1, we only consider the case with $a_L > 0$. We first consider the case with $\mathrm{K} > \mathrm{d}_{\mathsf{N}}$ and rewrite the VR(d_L,1) eigenvalue problem with $d_L \geq 2$ as

$$h_L^{-1}T^{2-2d_L}D_T^2\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}\widehat{P}_{\mathrm{K}}D_T^2 v_j = (T^{2d_L-3}h_L\mu_j)^{-1}T^{-1}D_T^2\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{1,R}\widehat{P}_{\mathrm{K}}D_T^2 v_j, \tag{B.28}$$

where $h_L^{-1}T^{2-2d_L}D_T^2\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}\widehat{P}_{\mathrm{K}}D_T^2$ and $T^{-1}D_T^2\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{1,R}\widehat{P}_{\mathrm{K}}D_T^2$ are, respectively, understood as

$$\widehat{\mathfrak{D}}^{(2)}_L = \begin{pmatrix} T^{-2d_L}h_L^{-1}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}})^* + \widehat{\Upsilon}^{\mathsf{N}}_{d_L,L}) & T^{1-2d_L}h_L^{-1}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathrm{K}}_{\mathsf{S}})^* + \widehat{\Upsilon}^{\mathsf{S}}_{d_L,L}) \\ T^{1-2d_L}h_L^{-1}\widehat{\Pi}\widehat{P}^{\mathrm{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}})^* & T^{2-2d_L}h_L^{-1}\widehat{\Pi}\widehat{P}^{\mathrm{K}}_{\mathsf{S}}\widehat{\Lambda}_{d_L,L}(\widehat{P}^{\mathrm{K}}_{\mathsf{S}})^* \end{pmatrix}, \tag{B.29}$$

$$\widehat{\mathfrak{D}}^{(2)}_R = \begin{pmatrix} T^{-3}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}})^* + \widehat{\Upsilon}^{\mathsf{N}}_{1,R}) & T^{-2}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathrm{K}}_{\mathsf{S}})^* + \widehat{\Upsilon}^{\mathsf{S}}_{1,R}) \\ T^{-2}\widehat{\Pi}\widehat{P}^{\mathrm{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathrm{K}}_{\mathsf{N}})^* & T^{-1}\widehat{\Pi}\widehat{P}^{\mathrm{K}}_{\mathsf{S}}\widehat{\Lambda}_{1,R}(\widehat{P}^{\mathrm{K}}_{\mathsf{S}})^* \end{pmatrix} \tag{B.30}$$

(see Appendix A). By Assumption 2, Lemma B.1, (B.5), and the conditions on h_L, h_R in Assumption 3, we may replace $\widehat{P}^{\mathrm{K}}_{\mathsf{N}}$ (resp. $\widehat{P}^{\mathrm{K}}_{\mathsf{S}}$ and $\widehat{\Pi}\widehat{P}^{\mathrm{K}}_{\mathsf{S}}$) with P_{N} (resp. $P^{\mathrm{K}}_{\mathsf{S}}$) in (B.29) and (B.30) and add a $\mathcal{O}_{\mathrm{P}}(1)$ component. Moreover, from (B.10)–(B.13), we find that $T^{-2d_L}h_L^{-1}\widehat{\Upsilon}^{\mathsf{N}}_{d_L,L} = \mathcal{O}_{\mathrm{P}}(1)$, $T^{1-2d_L}h_L^{-1}\widehat{\Upsilon}^{\mathsf{S}}_{d_L,L} = \mathcal{O}_{\mathrm{P}}(1)$, $T^{-3}\widehat{\Upsilon}^{\mathsf{N}}_{1,R} = \mathcal{O}_{\mathrm{P}}(1)$, and $T^{-2}\widehat{\Upsilon}^{\mathsf{S}}_{1,R} = \mathcal{O}_{\mathrm{P}}(1)$. Let $\{g_j\}_{j=1}^{\mathrm{K}-\mathrm{d}_{\mathsf{N}}}$ be eigenvectors of $(\Lambda^{\mathrm{K}}_{\mathsf{S},R})^{1/2}$ (where $\Lambda^{\mathrm{K}}_{\mathsf{S},R} = P_{\mathrm{K}}\Lambda_{\mathsf{S},R}P_{\mathrm{K}}$ as in our proof of Theorem 3.1, and $(\Lambda^{\mathrm{K}}_{\mathsf{S},R})^{1/2}$ is well defined), and let $\mathcal{B}_{\mathrm{K}-\mathrm{d}_{\mathsf{N}}}(r) = \sum_{j=1}^{\mathrm{K}-\mathrm{d}_{\mathsf{N}}} W_j(r)g_j$ be an $\mathcal{H}$-valued random element such that $(\langle\mathcal{B}_{\mathrm{K}-\mathrm{d}_{\mathsf{N}}}(r), g_1\rangle, \ldots, \langle\mathcal{B}_{\mathrm{K}-\mathrm{d}_{\mathsf{N}}}(r), g_{\mathrm{K}-\mathrm{d}_{\mathsf{N}}}\rangle)'$ is a $(\mathrm{K} - \mathrm{d}_{\mathsf{N}})$-dimensional vector-valued standard Brownian motion. From the functional central limit theorem and continuous mapping theorem, we find that $\sum_{s=1}^{t} P^{\mathrm{K}}_{\mathsf{S}}X_s$ and $\sum_{s=1}^{t} P_{\mathsf{N}}X_s$, viewed as partial sum processes, jointly converge weakly to $(\Lambda^{\mathrm{K}}_{\mathsf{S},R})^{1/2}\mathcal{B}_{\mathrm{K}-\mathrm{d}_{\mathsf{N}}}$ and $\Lambda^{1/2}_{\Delta X}\mathcal{W}_{2,\mathrm{d}_{\mathsf{N}}}$, respectively. Then, from similar arguments used in Phillips (1991, proof of Theorem 3.1(a) and

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

unnumbered equation between (A.10) and (A.11)), we deduce that

$$\widehat{\mathfrak{D}}_L^{(2)} \xrightarrow{\mathrm{d}} \mathfrak{D}_L^{(2)} = c_L \begin{pmatrix} \Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{d_L,\mathbb{d}_\mathsf{N}} \otimes \mathcal{W}_{d_L,\mathbb{d}_\mathsf{N}})\Lambda_{\Delta X}^{1/2} & \Lambda_{\Delta X}^{1/2}(\int \mathcal{B}_{\mathrm{K}-\mathbb{d}_\mathsf{N}} \otimes \mathcal{W}_{d_L,\mathbb{d}_\mathsf{N}})(\Lambda_{\mathsf{S},R}^{\mathrm{K}})^{1/2} \\ (\Lambda_{\mathsf{S},R}^{\mathrm{K}})^{1/2}(\int \mathcal{W}_{d_L,\mathbb{d}_\mathsf{N}} \otimes \mathcal{B}_{\mathrm{K}-\mathbb{d}_\mathsf{N}})\Lambda_{\Delta X}^{1/2} & (\Lambda_{\mathsf{S},R}^{\mathrm{K}})^{1/2}(\int \mathcal{B}_{\mathrm{K}-\mathbb{d}_\mathsf{N}} \otimes \mathcal{B}_{\mathrm{K}-\mathbb{d}_\mathsf{N}})(\Lambda_{\mathsf{S},R}^{\mathrm{K}})^{1/2} \end{pmatrix}. \tag{B.31}$$

We also find that

$$\widehat{\mathfrak{D}}_R^{(2)} \xrightarrow{\mathrm{P}} \mathfrak{D}_R^{(2)} = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_{\mathsf{S},R}^{\mathrm{K}} \end{pmatrix}. \tag{B.32}$$

As in our proof of Theorem 3.1, we let $[\mathfrak{D}_L^{(2)}]$ and $[\mathfrak{D}_R^{(2)}]$ be matrix representations of $\mathfrak{D}_L^{(2)}$ and $\mathfrak{D}_R^{(2)}$ for some orthonormal basis of $\operatorname{ran} P_{\mathrm{K}}$. It then follows that $\{T^{2d_L-3}h_L c_L \mu_j\}_{j=\mathbb{d}_\mathsf{N}+1}^{\mathrm{K}}$ converges in distribution to the $\mathrm{K}-\mathbb{d}_\mathsf{N}$ largest eigenvalues of $c_L[\mathfrak{D}_L^{(2)}]^{-1}[\mathfrak{D}_R^{(2)}]$. The limit of the latter matrix simplifies to $(\int B_{\mathrm{K}-\mathbb{d}_\mathsf{N}} B'_{\mathrm{K}-\mathbb{d}_\mathsf{N}} - \int B_{\mathrm{K}-\mathbb{d}_\mathsf{N}} W'_{d_L,\mathbb{d}_\mathsf{N}}(\int W_{d_L,\mathbb{d}_\mathsf{N}} W'_{d_L,\mathbb{d}_\mathsf{N}})^{-1} \int W_{d_L,\mathbb{d}_\mathsf{N}} B'_{\mathrm{K}-\mathbb{d}_\mathsf{N}})^{-1}$ without affecting the distributional properties of the eigenvalues (see the proof of Theorem 3.1), where $B_{\mathrm{K}-\mathbb{d}_\mathsf{N}}$ is a $(\mathrm{K}-\mathbb{d}_\mathsf{N})$-dimensional vector-valued standard Brownian motion independent of $W_{d_L,\mathbb{d}_\mathsf{N}}$. Then (3.14) follows immediately. Moreover, from Theorem 3.1, we know $T^{2d_L-3}h_L\mu_j \to_{\mathrm{P}} 0$ holds for $j \le \mathbb{d}_\mathsf{N}$, which implies that $(T^{2d_L-3}c_L h_L \mu_j)^{-1} \to_{\mathrm{P}} \infty$ for such j.

If $\mathrm{K} \le \mathbb{d}_\mathsf{N}$, and thus $\|\widehat{P}_{\mathrm{K}} P_\mathsf{S}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1)$, it holds that $\|\widehat{P}_{\mathrm{K}} - P_\mathsf{N}^{\mathrm{K}}\|_{\mathrm{op}} \to_{\mathrm{P}} 0$ for some (possibly random) orthogonal projection $P_\mathsf{N}^{\mathrm{K}}$ onto a subspace of $\mathcal{H}_\mathsf{N}$. We deduce from our previous discussion that both $T^{-2d_L}h_L^{-1}\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{d_L,L}(\widehat{P}_{\mathrm{K}})^*$ and $h_R^{-1}T^{-2}\widehat{P}_\mathsf{N}^{\mathrm{K}}\widehat{\Lambda}_{1,R}(\widehat{P}_\mathsf{N}^{\mathrm{K}})^*$ converge to random bounded linear operators allowing K (almost surely) positive eigenvalues. From these results and because $T/h_R \to \infty$ by Assumption 3, $(T^{2d_L-3}c_L h_L \mu_j)^{-1} \to_{\mathrm{P}} \infty$ follows for all $j \le \mathrm{K}$.

B.4. Proof of Theorem 3.4

In this proof, we let $\widehat{P}_\mathsf{N} = \sum_{j=1}^{\mathbb{d}_\mathsf{N}} \hat{f}_j \otimes \hat{f}_j$ and $\widehat{P}_\mathsf{S} = I - \widehat{P}_\mathsf{N}$. We first consider the case $a_R > 0$ (and hence $h_P > 0$). Note that, under Assumption 3, $\mathrm{k}_P(\cdot)$ has bounded support $[-\ell, \ell]$ for some $\ell < \infty$ and $|\mathrm{k}_P(\cdot)| \le 1$. Without loss of generality, we assume that $\ell = 1$, and then $\mathrm{k}_P(s/h_P) = 0$ holds if and only if $s \notin [-h_P, h_P]$.

First, consider $\mathrm{K} > \mathbb{d}_\mathsf{N}$. We then argue that

$$\|h_P^{-1}T^{-2}\widehat{\Lambda}_{1,P} - h_P^{-1}T^{-2}P_\mathsf{N}\widehat{\Lambda}_{1,P}P_\mathsf{N}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1). \tag{B.33}$$

This holds because

$$\frac{1}{h_P T^2}\sum_{s=0}^{h_P}\sum_{t=1}^{T}(\|P_\mathsf{N}X_{t-s}\|\|P_\mathsf{S}X_t\| + \|P_\mathsf{S}X_{t-s}\|\|P_\mathsf{N}X_t\| + \|P_\mathsf{S}X_{t-s}\|\|P_\mathsf{S}X_t\|) = \mathcal{O}_{\mathrm{P}}(1),$$

which follows from the facts that $\sup_{1\le t\le T}\|P_\mathsf{N}X_t/\sqrt{T}\| = \mathcal{O}_{\mathrm{P}}(1)$ and $T^{-1}\sum_{t=1}^{T}\mathbb{E}\|P_\mathsf{S}X_t\|^2 = \mathcal{O}(1)$, which both hold under our assumptions, along with the Cauchy–Schwarz inequality. By (B.33), we then find, from the same arguments used in the proof of Proposition 3.2 of Chang et al. (2016), that $\widehat{P}_\mathsf{N} - P_\mathsf{N} = \mathcal{O}_{\mathrm{P}}(T^{-1})$, $\widehat{P}_\mathsf{S} - P_\mathsf{S} = \mathcal{O}_{\mathrm{P}}(T^{-1})$, and $\widehat{P}_\mathsf{S}P_\mathsf{N} = \mathcal{O}_{\mathrm{P}}(T^{-1})$. These results combined with Lemma B.1 imply that

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

$T^{-1}\widehat{P}_{\mathsf{S}}P_{\mathsf{N}}\widehat{\Lambda}_{1,P}P_{\mathsf{N}}\widehat{P}_{\mathsf{S}}$, $T^{-1}\widehat{P}_{\mathsf{S}}P_{\mathsf{N}}\widehat{\Lambda}_{1,P}P_{\mathsf{S}}\widehat{P}_{\mathsf{S}}$, and $T^{-1}\widehat{P}_{\mathsf{S}}P_{\mathsf{N}}\widehat{\Lambda}_{1,P}P_{\mathsf{S}}\widehat{P}_{\mathsf{S}}$ are all $\mathcal{O}_{\mathrm{P}}(1)$, and, in addition, that

$$T^{-1}\widehat{P}_{\mathsf{S}}P_{\mathsf{S}}\widehat{\Lambda}_{1,P}P_{\mathsf{S}}\widehat{P}_{\mathsf{S}} \xrightarrow{\mathrm{P}} \Lambda_{\mathsf{S},R}, \tag{B.34}$$

where the convergence follows from Horváth et al. (2013, Theorem 2) (see also our proof of Theorem 3.1). We may then deduce from Bosq (2000, Lemma 4.2) that $\sup_{j\geq 1}|\widehat{\tau}_{j+\mathbb{d}_{\mathsf{N}}} - \tau_j| \to_{\mathrm{P}} 0$. Moreover, from Bosq (2000, Lemmas 4.3 and 4.4), we find that (B.34) implies that

$$\Big\| \sum_{j=\mathbb{d}_{\mathsf{N}}+1}^{\mathrm{K}} \hat{f}_j \otimes \hat{f}_j - P_{\mathsf{S}}^{\mathrm{K}} \Big\|_{\mathrm{op}} \xrightarrow{\mathrm{P}} 0, \tag{B.35}$$

where $P_{\mathsf{S}}^{\mathrm{K}}$ is the projection onto the span of $\{f_j\}_{j=\mathbb{d}_{\mathsf{N}}+1}^{\mathrm{K}}$, and we know that $P_{\mathsf{S}}^{\mathrm{K}}$ is a nonrandom projection since $\tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}} \neq \tau_{\mathrm{K}-\mathbb{d}_{\mathsf{N}}+1}$. Therefore, $\widehat{P}_{\mathrm{K}}$ satisfies Assumption 2.

Next, if $\mathrm{K} = \mathbb{d}_{\mathsf{N}}$ then Assumption 2 holds for $P_{\mathsf{S}}^{\mathrm{K}} = 0$ obviously.

Finally, consider $\mathrm{K} < \mathbb{d}_{\mathsf{N}}$. From (B.12)–(B.15), we find that $\|h_P^{-1}T^{-2}\widehat{\Lambda}_{1,P} - c_P\Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}} \otimes \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1)$ (where $c_P = 2\int_0^1 \mathrm{k}_P(u)du$). Since the nonzero eigenvalues of $\Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}} \otimes \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2}$ are almost surely distinct and $\operatorname{ran}\Lambda_{\Delta X}^{1/2} = \mathcal{H}_{\mathsf{N}}$, we deduce from Bosq (2000, Lemma 4.3) that

$$\sup_{1\leq j\leq \mathrm{K}} \|\hat{f}_j - \operatorname{sgn}(\langle \hat{f}_j, f_j\rangle)f_j\| \xrightarrow{\mathrm{P}} 0, \tag{B.36}$$

where $\{f_j\}_{j=1}^{\mathrm{K}}$ is an orthonormal set in $\mathcal{H}_{\mathsf{N}}$. Thus, we find that $\|(\sum_{j=1}^{\mathrm{K}}\hat{f}_j \otimes \hat{f}_j)P_{\mathsf{S}}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1)$.

In the case $a_R = 0$ and $\mathrm{K} \geq \mathbb{d}_{\mathsf{N}}$, the result (B.33) reduces to $\|T^{-2}\widehat{\Lambda}_{1,P} - T^{-2}P_{\mathsf{N}}\widehat{\Lambda}_{1,P}P_{\mathsf{N}}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1)$ and (B.34) still holds with $a_R = 0$ (whereby $\Lambda_{\mathsf{S},R}$ becomes $\mathbb{E}[P_{\mathsf{S}}X_t \otimes P_{\mathsf{S}}X_t]$). Using nearly identical arguments, we then find that (B.35) is established and $\widehat{P}_{\mathrm{K}}$ satisfies the requirements in Assumption 2. In the case $\mathrm{K} < \mathbb{d}_{\mathsf{N}}$, we similarly deduce from (B.12)–(B.15) that $\|T^{-2}\widehat{\Lambda}_{1,P} - \Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}} \otimes \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}(1)$, from which the desired result follows.

The above proof does not require any significant changes even if we allow for a nonzero intercept or a linear trend as in Section 3.5. Provided that $\widehat{\Lambda}_{1,P}$ is computed from the relevant residuals, that is, $U_t^{(1)}$ or $U_t^{(2)}$ given in (3.22), we may show that (B.33) holds and that $\widehat{P}_{\mathsf{N}} - P_{\mathsf{N}} = \mathcal{O}_{\mathrm{P}}(T^{-1})$ and $\widehat{P}_{\mathsf{S}} - P_{\mathsf{S}} = \mathcal{O}_{\mathrm{P}}(T^{-1})$. Then $T^{-1}\widehat{P}_{\mathsf{S}}P_{\mathsf{S}}\widehat{\Lambda}_{1,R}P_{\mathsf{S}}\widehat{P}_{\mathsf{S}} \to_{\mathrm{P}} \Lambda_{\mathsf{S},R}$ follows from Kokoszka and Young (2016, Theorem 5.3); cf. (B.34). We thus find from Bosq (2000, Lemmas 4.3 and 4.4) that (B.35) holds if $\hat{f}_j$ is an eigenvector of $\widehat{\Lambda}_{1,P}$ computed from the relevant residuals.

B.5. Proof of Theorem 3.5

We first consider the case $a_R > 0$ and hence $h_P > 0$. We note that

$$T^{-1}\widehat{\Lambda}_{1,P} = T^{-1}P_{\mathsf{N}}\widehat{\Lambda}_{1,P}P_{\mathsf{N}} + T^{-1}(P_{\mathsf{N}}\widehat{\Lambda}_{1,P}P_{\mathsf{S}} + P_{\mathsf{S}}\widehat{\Lambda}_{1,P}P_{\mathsf{N}}) + T^{-1}P_{\mathsf{S}}\widehat{\Lambda}_{1,P}P_{\mathsf{S}},$$

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

where the first term is $\mathcal{O}_{\mathrm{P}}(h_P T)$ by Lemma B.1. The second term is $\mathcal{O}_{\mathrm{P}}(h_P)$, and this is deduced from similar arguments used in our proof of Lemma B.1 (see also Lemma 3.1 of Chang et al., 2016). The third term is $\mathcal{O}_{\mathrm{P}}(1)$, which results from Horváth et al. (2013, Theorem 2). Since $\widehat{P}_{\mathsf{N}} - P_{\mathsf{N}} = \mathcal{O}_{\mathrm{P}}(T^{-1})$ and $\widehat{P}_{\mathsf{S}} - P_{\mathsf{S}} = \mathcal{O}_{\mathrm{P}}(T^{-1})$, $T^{-1}\widehat{P}_{\mathsf{S}} P_{\mathsf{N}} \widehat{\Lambda}_{1,P} P_{\mathsf{N}} \widehat{P}_{\mathsf{S}}$, $T^{-1}\widehat{P}_{\mathsf{S}} P_{\mathsf{N}} \widehat{\Lambda}_{1,P} P_{\mathsf{S}} \widehat{P}_{\mathsf{S}}$, and $T^{-1}\widehat{P}_{\mathsf{S}} P_{\mathsf{S}} \widehat{\Lambda}_{1,P} P_{\mathsf{N}} \widehat{P}_{\mathsf{S}}$ are all $\mathcal{O}_{\mathrm{P}}(h_P/T)$, and, in addition, (B.34) holds. Combining all these results,

$$T^{-1}\widehat{P}_{\mathsf{S}} \widehat{\Lambda}_{1,P} \widehat{P}_{\mathsf{S}} - \Lambda_{\mathsf{S},R} = T^{-1}\widehat{P}_{\mathsf{S}}(P_{\mathsf{S}} \widehat{\Lambda}_{1,P} P_{\mathsf{S}} - \Lambda_{\mathsf{S},R})\widehat{P}_{\mathsf{S}} + \mathcal{O}_{\mathrm{P}}(h_P/T) \xrightarrow{\mathrm{P}} 0, \tag{B.37}$$

where the first equality follows from the facts that $\widehat{P}_{\mathsf{N}} \Lambda_{\mathsf{S},R}$ and $\Lambda_{\mathsf{S},R} \widehat{P}_{\mathsf{N}}$ are both $\mathcal{O}_{\mathrm{P}}(T^{-1})$. Under the conditions in Assumption 5, we may deduce from Berkes et al. (2016, Theorems 2.2 and 2.3, and their proofs) that $T^{-1} P_{\mathsf{S}} \widehat{\Lambda}_{1,R} P_{\mathsf{S}} - \Lambda_{\mathsf{S},R} = \mathcal{O}_{\mathrm{P}}((T/h_P)^{-1/2}) + \mathcal{O}_{\mathrm{P}}(T^{-1}) + \mathcal{O}_{\mathrm{P}}(h_P^{-\varphi})$. This result, combined with (B.37), implies that $(T/h_P)^{1/2}(T^{-1}\widehat{P}_{\mathsf{S}} \widehat{\Lambda}_{1,P} \widehat{P}_{\mathsf{S}} - \Lambda_{\mathsf{S},R}) = \mathcal{O}_{\mathrm{P}}(1) + \mathcal{O}_{\mathrm{P}}((T/h_P)^{1/2} h_P^{-\varphi})$, which is $\mathcal{O}_{\mathrm{P}}(1)$ under Assumption 5(iv). We thus find that $T^{-1}\widehat{P}_{\mathsf{S}} \widehat{\Lambda}_{1,P} \widehat{P}_{\mathsf{S}} = \Lambda_{\mathsf{S},R} + \mathcal{O}_{\mathrm{P}}((h_P/T)^{1/2})$. We then deduce from Lemma 4.2 of Bosq (2000) that the eigenvalues $\{\widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j}\}_{j\geq 1}$ of $T^{-1}\widehat{P}_{\mathsf{S}} \widehat{\Lambda}_{1,P} \widehat{P}_{\mathsf{S}}$ satisfy

$$\sup_{j\geq 1} |\widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j} - \tau_j| \leq \|T^{-1}\widehat{P}_{\mathsf{S}} \widehat{\Lambda}_{1,P} \widehat{P}_{\mathsf{S}} - \Lambda_{\mathsf{S},R}\|_{\mathrm{op}} = \mathcal{O}_{\mathrm{P}}((h_P/T)^{1/2}). \tag{B.38}$$

From (B.38), we find that

$$\begin{aligned} \widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j} - \widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j+1} &= (\widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j} - \tau_j) + (\tau_j - \tau_{j+1}) + (\tau_{j+1} - \widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j+1}) \\ &= \tau_j - \tau_{j+1} + \mathcal{O}_{\mathrm{P}}((h_P/T)^{1/2}), \end{aligned} \tag{B.39}$$

which is the desired result for $a_R > 0$. In the case $a_R = 0$, and hence $P_{\mathsf{S}} \widehat{\Lambda}_{1,P} P_{\mathsf{S}}$ becomes the sample variance operator of $P_{\mathsf{S}} X_t$, we deduce from Theorems 3.1 and 3.2 of Hörmann and Kokoszka (2010) that $\sup_{j\geq 1} |\widehat{\tau}_{\mathrm{d}_{\mathsf{N}}+j} - \tau_j| = \mathcal{O}_{\mathrm{P}}(T^{-1/2})$.

The above proof does not require any significant changes if $\widehat{\Lambda}_{1,P}$ is computed from $U_t^{(1)}$ or $U_t^{(2)}$ given in (3.22) to accommodate a nonzero intercept or linear trend. In those cases, (B.37) still holds and $(T/h_P)^{1/2}(T^{-1} P_{\mathsf{S}} \widehat{\Lambda}_{1,P} P_{\mathsf{S}} - \Lambda_{\mathsf{S},R}) = \mathcal{O}_{\mathrm{P}}(1)$ (see Theorem 5.3 of Kokoszka and Young, 2016 and Theorem 2.1 of Berkes et al., 2016). Then the rest of the proof is almost identical and hence omitted.

B.6. Proof of Theorem 4.2

If $\{\mu_j\}_{j=1}^{\mathrm{d}_{\max}+1}$ are the eigenvalues of the VR(d_L,1) problem with $d_L \geq 2$, then the consistency result follows immediately from Theorems 3.1 and 3.3.

For the VR(d_L,0) problem with $d_L \geq 2$, we consider only $a_L > 0$ and rewrite (3.4) as

$$h_L^{-1} T^{2-2d_L} D_T^2 \widehat{P}_{\mathrm{K}} \widehat{\Lambda}_{d_L,L} \widehat{P}_{\mathrm{K}} D_T^2 v_j = (T^{2d_L-3} h_L h_R^2 \mu_j)^{-1} T^{-1} h_R^2 D_T^2 \widehat{P}_{\mathrm{K}} \widehat{\Lambda}_{0,R} \widehat{P}_{\mathrm{K}} D_T^2 v_j. \tag{B.40}$$

The operator matrix corresponding to the left-hand side of (B.40) is $\widehat{\mathfrak{D}}_L^{(2)}$ from our proof of Theorem 3.3. From (B.31), we know that $\widehat{\mathfrak{D}}_L^{(2)} \to_{\mathrm{d}} \mathfrak{D}_L^{(2)}$, where $\mathfrak{D}_L^{(2)}$ is given in (B.31). We let $\widehat{\mathfrak{D}}_R^{(3)}$ be the operator matrix corresponding to $T^{-1} h_R^2 D_T^2 \widehat{P}_{\mathrm{K}} \widehat{\Lambda}_{0,R} \widehat{P}_{\mathrm{K}} D_T^2$, which can

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

be constructed similarly to $\widehat{\mathfrak{D}}_R^{(2)}$ in the proof of Theorem 3.3. From (B.21)–(B.26) and the fact that $h_R T^{-1} \to 0$, we deduce that

$$\widehat{\mathfrak{D}}_R^{(3)} \xrightarrow{\mathrm{P}} -\mathrm{k}_R''(0)\mathfrak{D}_R^{(2)}, \tag{B.41}$$

where $\mathfrak{D}_R^{(2)}$ is given in (B.32). We thus find that $T^{2d_L-3}h_L h_R^2 c_L \mu_j / \mathrm{k}_R''(0)$ converges to well-defined nonzero nonrandom eigenvalues for $j = \mathbb{d}_{\mathsf{N}} + 1, \ldots, \mathrm{K}$ (jointly), but converges to zero for $j \le \mathbb{d}_{\mathsf{N}}$. From this result and Theorem 3.2, the desired consistency result follows.

Next, suppose $\{\mu_j\}_{j=1}^{\mathbb{d}_{\max}+1}$ are the eigenvalues of the VR(1,0) problem. As before, we only consider $a_L > 0$ and rewrite (3.4) as

$$T^{-1}D_{h_L T}\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{1,L}\widehat{P}_{\mathrm{K}}D_{h_L T}v_j = (h_R^2\mu_j)^{-1}T^{-1}h_R^2 D_{h_L T}\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{0,R}\widehat{P}_{\mathrm{K}}D_{h_L T}v_j, \tag{B.42}$$

and let $\widehat{\mathfrak{D}}_L^{(4)}$ (resp. $\widehat{\mathfrak{D}}_R^{(4)}$) be the operator matrix corresponding to $T^{-1}D_{h_L T}\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{1,L}\widehat{P}_{\mathrm{K}}D_{h_L T}$ (resp. $T^{-1}h_R^2 D_{h_L T}\widehat{P}_{\mathrm{K}}\widehat{\Lambda}_{0,R}\widehat{P}_{\mathrm{K}}D_{h_L T}$). Let $\Lambda_{\mathsf{S},L}$ be defined as in (3.2) but with a_R replaced by a_L. From similar arguments used to obtain (B.16) and the facts that $h_L^{-1/2}h_R T^{-1/2} \to 0$ and $T^{-1}\widehat{\Pi}\widehat{P}_{\mathsf{S}}^{\mathrm{K}}\widehat{\Lambda}_{1,L}(\widehat{P}_{\mathsf{S}}^{\mathrm{K}})^* \to_{\mathrm{P}} \Lambda_{\mathsf{S},L}^{\mathrm{K}} = P_{\mathrm{K}}\Lambda_{\mathsf{S},L}P_{\mathrm{K}}$, see (B.15), we find that

$$\widehat{\mathfrak{D}}_L^{(4)} \xrightarrow{\mathrm{d}} \mathfrak{D}_L^{(4)} = \begin{pmatrix} c_L \Lambda_{\Delta X}^{1/2}(\int \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}} \otimes \mathcal{W}_{1,\mathbb{d}_{\mathsf{N}}})\Lambda_{\Delta X}^{1/2} & 0 \\ 0 & \Lambda_{\mathsf{S},L}^{\mathrm{K}} \end{pmatrix}. \tag{B.43}$$

From (B.21)–(B.26) and the fact that $h_L^{-1/2}h_R T^{-1/2} \to 0$, we find, as in (B.41), that

$$\widehat{\mathfrak{D}}_R^{(4)} \xrightarrow{\mathrm{P}} -\mathrm{k}_R''(0)\mathfrak{D}_R^{(2)}, \tag{B.44}$$

where $\mathfrak{D}_R^{(2)}$ is given in (B.32). From (B.42)–(B.44), the conclusions follow by the same arguments as for the VR(d_L,0) problem with $d_L \ge 2$.

B.7. Proof of Theorem 5.1

Consider first the hypotheses in (5.1). Under H_0, the dimension of the nonstationary subspace of the residual time series $\{(I - P_{\mathcal{H}_0})X_t\}$ is $\mathbb{d}_{\mathsf{N}} - p_0$, while it is greater than $\mathbb{d}_{\mathsf{N}} - p_0$ under H_1. We thus know from Corollary 3.3 that under H_0 the statistics $\mathcal{F}_{\max}(\{\tilde{\mu}_j^{-1}\}_{j=\mathbb{d}_{\mathsf{N}}-p_0}^{\mathrm{K}})$ and $\mathcal{F}_{\mathrm{trace}}(\{\tilde{\mu}_j^{-1}\}_{j=\mathbb{d}_{\mathsf{N}}-p_0}^{\mathrm{K}})$ converge to the maximum eigenvalue and trace, respectively, of $\mathcal{A}$ with $\mathbb{d}_0 = \mathbb{d}_{\mathsf{N}} - p_0$, where $\mathcal{A}$ is defined in Corollary 3.3. On the other hand, also by Corollary 3.3, the statistics both diverge to infinity under H_1.

Next, consider the hypotheses in (5.2). Under H_0, the residual time series $\{(I - P_{\mathcal{H}_0})X_t\}$ is stationary, while it is nonstationary under H_1. In this case, we know from Corollary 3.3 that under H_0 the statistics $\mathcal{F}_{\max}(\{\tilde{\mu}_j^{-1}\}_{j=1}^{\mathrm{K}})$ and $\mathcal{F}_{\mathrm{trace}}(\{\tilde{\mu}_j^{-1}\}_{j=1}^{\mathrm{K}})$ converge to the maximum eigenvalue and trace, respectively, of $\int B_{\mathrm{K}}B_{\mathrm{K}}'$, while they diverge to infinity under H_1.

The arguments for the hypotheses in (5.3) are similar and omitted for brevity.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466626100462.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

FUNDING STATEMENT

M.Ø.N. is grateful for financial support from the Aarhus Center for Econometrics (ACE) funded by the Danish National Research Foundation (Grant No. DNRF186) and from the DNRF Chair Program (Grant No. DNRF154).

COMPETING INTEREST

The authors declare that no competing interests exist.

REFERENCES

Ahn, S. K. (1997). Inference of vector autoregressive models with cointegration and scalar components. *Journal of the American Statistical Association*, 92, 350–356.

Ahn, S. K., & Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85, 813–823.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817–858.

Barigozzi, M., & Trapani, L. (2022). Testing for common trends in nonstationary large datasets. *Journal of Business & Economic Statistics*, 40, 1107–1122.

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

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

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

Berkes, I., Horváth, L., & Rice, G. (2016). On the asymptotic normality of kernel estimators of the long run covariance of functional time series. *Journal of Multivariate Analysis*, 144, 150–175.

Bewley, R., & Yang, M. (1995). Tests for cointegration based on canonical correlation analysis. *Journal of the American Statistical Association*, 90, 990–996.

Bewley, R., & Yang, M. (1998). On the size and power of system tests for cointegration. *Review of Economics and Statistics*, 80, 675–679.

Bosq, D. (2000). *Linear processes in function spaces*. Springer.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of Econometrics*, 108, 343–363.

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

Bykhovskaya, A., & Gorin, V. (2024). Asymptotics of cointegration tests for high-dimensional VAR(k). *Review of Economics and Statistics*, Forthcoming.

Chang, J., Cheng, G., & Yao, Q. (2021). Testing for unit roots based on sample autocovariances. *Biometrika*, 109, 543–550.

Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, 192, 152–167.

Chen, X., & White, H. (1998). Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory*, 14, 260–284.

Conway, J. B. (1990). *A course in functional analysis*. (2nd ed.) Springer.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130, 337–364.

Franchi, M., Georgiev, I., & Paruolo, P. (2026). Estimating the number of common trends in large T and N factor models via canonical correlations analysis. *Econometrics and Statistics*, 39, 81–95.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

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

Franchi, M., & Paruolo, P. (2026). Inference on the attractor space via functional approximation. *Oxford Bulletin of Economics and Statistics*, Forthcoming.

Haug, A. A. (1996). Tests for cointegration: A Monte Carlo comparison. *Journal of Econometrics*, 71, 89–115.

Hörmann, S., & Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics*, 38, 1845–1884.

Horváth, L., Kokoszka, P., & Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society B*, 75, 103–122.

Horváth, L., Kokoszka, P., & Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.

Kokoszka, P., & Young, G. (2016). KPSS test for functional time series. *Statistics*, 50, 957–973.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54, 159–178.

Li, D., Li, Y., & Phillips, P. C. B. (2025). *Large-scale curve time series with common stochastic trends*. Preprint, arXiv:2509.11060.

Li, D., Robinson, P. M., & Shang, H. L. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association*, 115, 957–971.

Li, D., Robinson, P. M., & Shang, H. L. (2023). Nonstationary fractionally integrated functional time series. *Bernoulli* 29, 1505–1526. Earlier working paper version. https://www.researchgate.net/publication/340503853

Martínez-Hernández, I., Gonzalo, J., & González-Farías, G. (2022). Nonparametric estimation of functional dynamic factor model. *Journal of Nonparametric Statistics*, 34, 895–916.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34, 574–589.

Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, 141, 1331–1352.

Müller, U. K. (2008). The impossibility of consistent discrimination between I(0) and I(1) processes. *Econometric Theory*, 24, 616–630.

Nam, K., & Seo, W.-K. (2025). *Functional regression with nonstationarity and error contamination: Application to the economic impact of climate change*. Preprint, arXiv:2509.08591.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60, 473–489.

Nielsen, M. Ø., Seo, W., & Seong, D. (2023). Inference on the dimension of the nonstationary subspace in functional time series. *Econometric Theory*, 39, 443–480.

Nyblom, J., & Harvey, A. (2000). Tests of common stochastic trends. *Econometric Theory*, 16, 176–199.

Onatski, A., & Wang, C. (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica*, 86, 1465–1478.

Peña, D., & Poncela, P. (2004). Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, 119, 291–321.

Peña, D., & Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136, 1237–1257.

Phillips, P. C. B. (1991). Spectral regression for cointegrated time series. In W. A. Barnett, J. Powell, & G. E. Tauchen (Eds.), *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the fifth international symposium in economic theory and econometrics* (pp. 413–435). Cambridge University Press.

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

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press

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

Rice, G., & Shang, H. L. (2017). A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. *Journal of Time Series Analysis*, 38, 591–609.

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

Seo, W.-K. (2024). Functional principal component analysis for cointegrated functional time series. *Journal of Time Series Analysis*, 45, 320–330.

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

Shintani, M. (2001). A simple cointegrating rank test without vector autoregression. *Journal of Econometrics*, 105, 337–362.

Stock, J. H., & Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83, 1097–1107.

Tang, C., Shang, H. L., Yang, Y., & Yang, Y. (2025). Forecasting high-dimensional functional time series with dual-factor structures. *Journal of the Royal Statistical Society Series A*. Forthcoming.

Toda, H. Y. (1995). Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions. *Econometric Theory*, 11, 1015–1032.

Zhang, R., & Chan, N. H. (2018). Portmanteau-type tests for unit-root and cointegration. *Journal of Econometrics*, 207, 307–324.

Zhang, R., Robinson, P., & Yao, Q. (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association*, 114, 916–927.

https://doi.org/10.1017/S0266466626100462 Published online by Cambridge University Press