Inference on the dimension of the nonstationary subspace in functional time series*

Morten Ørregaard Nielsen†
Queen's University and CREATES
mon@econ.queensu.ca

Won-Ki Seo
University of Sydney
won-ki.seo@sydney.edu.au

Dakyung Seong
University of Sydney
dakyung.seong@sydney.edu.au

February 23, 2021

Abstract

We propose a statistical procedure to determine the dimension of the nonstationary subspace of cointegrated functional time series taking values in the Hilbert space of square-integrable functions defined on a compact interval. The procedure is based on sequential application of a proposed test for the dimension of the nonstationary subspace. To avoid estimation of the long-run covariance operator, our test is based on a variance ratio-type statistic. We derive the asymptotic null distribution and prove consistency of the test. Monte Carlo simulations show good performance of our test and provide evidence that it outperforms the existing testing procedure. We apply our methodology to three empirical examples: age-specific US employment rates, Australian temperature curves, and Ontario electricity demand.

JEL codes: C32.

Keywords: cointegration, functional data, nonstationarity, stochastic trends, variance ratio.

*We are grateful to Brendan Beare, Yoosoon Chang, Joon Park, Peter Phillips, Hanlin Shang, Yundong Tu, and seminar participants at Queen's University, UC Davis, Australian National University, the 2019 Canadian Economics Association Conference, the 2019 Time Series and Forecasting Symposium at University of Sydney, the 2020 ANZESG conference, the 2020 Econometric Society World Congress, the 2020 $(EC)^2$ Conference, and the CFE 2020 Conference for comments. An earlier version of this paper was circulated under the title "Variance ratio test for the number of stochastic trends in functional time series." Nielsen thanks the Canada Research Chairs program and the Social Sciences and Humanities Research Council of Canada for financial support. Seo thanks the Sir Edward Peacock Postdoctoral Fellowship at Queen's University for financial support. Data and R code to replicate the empirical results in Table 5 are available on the authors' websites.

†Corresponding author. Address: Department of Economics, 94 University Avenue, Queen's University, Kingston, Ontario K7L 3N6, Canada. Email: mon@econ.queensu.ca. Tel.: 613-533-2262. Fax: 613-533-6668.

1 Introduction

Much recent research in time series analysis has focused on so-called functional time series; that is, time series that take values in possibly infinite-dimensional Hilbert or Banach spaces rather than the usual finite-dimensional Euclidean space. Each observation of a functional time series may be, for example, a continuous function, a square-integrable function, or a probability density function (of course vector-valued time series are a special case). Recent monograph treatments include Bosq (2000), who considers stationary linear processes taking values in Hilbert and Banach spaces, and Horváth and Kokoszka (2012), who discuss statistical analysis of functional data and functional time series with many empirical examples.

The majority of recent developments in functional time series depend crucially on the assumption of stationarity. Despite its importance, this issue has received very limited attention in the literature. In the context of functional time series, there exists only a few articles that consider tests of the null hypothesis of stationarity. In particular, Horváth et al. (2014) and Kokoszka and Young (2016) develop a modified version of the univariate KPSS test (Kwiatkowski et al., 1992) of stationarity, and Aue and van Delft (2020) propose a test of stationarity in the frequency domain.

Of course, testing for stationarity is an important first step. However, when a time series is not stationary, an important problem is to determine the type and magnitude of departure from stationarity. The type of nonstationarity that we consider is the "unit root" or I(1) nonstationarity well-known from autoregressive processes. Thus, the issue is to determine the extent of nonstationarity, which we interpret as the dimension of the nonstationary subspace (to be made precise later). In this context, testing stationarity is the same as testing that the nonstationary subspace has dimension zero. To the best of our knowledge, the only article that considers this more general problem is the seminal contribution of Chang et al. (2016), though they denote it the "unit root dimension." Specifically, Chang et al. (2016) propose a test based on generalized eigenvalues associated with the covariance operator of the observations and the long-run covariance operator of the first-differenced observations.

In the analysis of vector-valued time series in finite-dimensional space, the dimension of the nonstationary subspace is the number of linearly independent linear combinations that are nonstationary and is called the number of common stochastic trends. Similarly, the dimension of the stationary subspace is the number of linearly independent linear combinations that are stationary and is called the cointegration rank. In finite-dimensional space, these numbers are both finite. In that context, Stock and Watson (1988) is an important early contribution, providing a statistical testing procedure for the dimension of the nonstationary subspace (number of stochastic trends) in cointegrated vector time series. Different from

most subsequent work, their procedure is interpreted as a way to find the number of stochastic trends rather than the cointegration rank. In finite-dimensional space, there is no meaningful difference between those interpretations since one determines the other. However, in infinite-dimensional spaces, and specifically in Chang et al.'s (2016) model and in our model of cointegrated linear processes that both have finite-dimensional nonstationary subspace, it is clear that the interpretation given in Stock and Watson (1988) is the most natural.

In the context of time series in finite-dimensional space, there has been a very large literature developing methods for determination of the cointegration rank or equivalently the number of common stochastic trends. These have mostly been based on canonical correlation analysis (e.g., Ahn and Reinsel, 1990; Bewley and Yang, 1995; Johansen, 1995) and eigenvalue analysis (e.g., Stock and Watson, 1988; Johansen, 1995; Zhang et al., 2019). In an important contribution to finite-dimensional time series, Müller (2008) demonstrates some desirable properties of variance ratio-type unit root test statistics that are not shared by other statistics that have to estimate the long-run covariance. In particular, tests based on variance ratio-type statistics are shown to be able to consistently discriminate between the unit root null and the stationary alternative. More practically, variance ratio statistics avoid estimation of the long-run covariance, which is known to be difficult in practice.

Inspired by the work of Müller (2008) and Chang et al. (2016), we consider a nonparametric variance ratio-type test statistic for the dimension of the nonstationary subspace. In the univariate special case our statistic reduces to the KPSS statistic with bandwidth zero and in the finite-dimensional case to the statistic considered by Breitung (2002); see also Taylor (2005), Nielsen (2009, 2010), and Pedroni et al. (2015). As in Chang et al. (2016), we assume that a cointegrated functional time series has finite-dimensional nonstationary subspace while the stationary subspace is infinite-dimensional (i.e., there are infinitely many cointegrating relations). We then discuss sequential application of our test to determine the dimension of the nonstationary subspace (the number of common stochastic trends). We derive the asymptotic null distribution and prove consistency of the proposed test.

Our procedure has several attractive features. First, our test is nonparametric; that is, we do not require the specification of a particular model. Second, it is easy to implement in practice. The statistic is given by the sum of generalized eigenvalues of sample covariance operators. Third, we do not need to estimate any long-run covariance operators. This is an important difference from existing methods in functional time series such as Horváth et al. (2014), Kokoszka and Young (2016), and Chang et al. (2016). Fourth, Chang et al.'s (2016) test requires projection onto a space that asymptotically includes the nonstationary subspace, and which is of the same dimension. On the other hand, our proposed method requires projection onto a space that asymptotically includes the nonstationary subspace,

but it may be of higher dimension than the latter. Clearly, this would appear to be easier in practice. Fifth, the asymptotic null distribution of the proposed test statistic does not depend on the choice of projection operator used to approximate the covariance operators, and is simply a functional of standard Brownian motion. In practice, therefore, our testing procedure can be easily implemented in a familiar finite-dimensional setting.

The Monte Carlo simulations in Section 5 suggest that our test has better finite-sample properties than the existing test of Chang et al. (2016). First, our test has much better size control across a wide range of simulation DGPs. Second, our test has superior finite-sample power (size-corrected power in cases where Chang et al.'s (2016) test is severely over-sized). Third, Chang et al.'s (2016) test is subject to a power reversal problem, which is not the case for our test. Fourth, and consequently, when implemented sequentially to determine the dimension of the nonstationary subspace, our test is very robust to the choice of the initial dimension, whereas Chang et al.'s (2016) test is very sensitive to this choice.

We present several empirical illustrations of our methodology, where we also compare with the test of Chang et al. (2016). In particular, we consider age-specific employment curves, Australian temperature curves, and Ontario electricity demand curves. Other applications of Chang et al. (2016) include global temperature distributions as in Chang et al. (2020). In our first empirical application, we also demonstrate how it is not feasible to consider a discrete approximation to the functional time series as a high-dimensional cointegrated vector autoregression, for example. In many cases, it is simply not computationally possible to perform the required eigenvalue analysis, while in others the high-dimensional cointegrated vector autoregression does not work well because of poor properties with high-dimensional time series; see Ho and Sørensen (1996), Onatski and Wang (2018), and Section 5.

The remainder of this paper is organized as follows. We review some essential mathematical preliminaries in Section 2. Our testing procedure and asymptotic theory is provided in Section 3. In Section 4 we present the results from our Monte Carlo simulations. We then apply our methodology to three empirical data sets in Section 5. All proofs are in the appendices.

2 Mathematical preliminaries

Let $\mathcal{H}$ denote the space of square-integrable functions defined on a compact interval $\mathcal{I}$ equipped with the inner product given by $\langle f, g\rangle = \int_{\mathcal{I}} f(u)g(u)du$ for $f, g \in \mathcal{H}$ and its induced norm $\|f\| = \langle f, f\rangle^{1/2}$ for $f \in \mathcal{H}$. Then $\mathcal{H}$ is a separable Hilbert space. Without loss of generality, we normalize and assume that $\mathcal{I} = [0, 1]$ and $\int f(u)du = \int_0^1 f(u)du$ throughout the paper. When there is no risk of confusion, we suppress the argument u and we use the terms vector and function interchangeably to denote an element of (the vector space) $\mathcal{H}$.

Given a subset $M \subset \mathcal{H}$, $M^\perp$ denotes the orthogonal complement of M and $\operatorname{cl} M$ denotes the closure of M. Given two subspaces $M_1, M_2 \subset \mathcal{H}$ with $M_1 \cap M_2 = \{0\}$, $\mathcal{H}$ is said to be a direct sum of M_1 and M_2, denoted by $\mathcal{H} = M_1 \oplus M_2$, if any element $x \in \mathcal{H}$ can be written as $x = x_{M_1} + x_{M_2}$ for some $x_{M_1} \in M_1$ and $x_{M_2} \in M_2$.

We let $\mathcal{L}_\mathcal{H}$ denote the space of bounded linear operators acting on $\mathcal{H}$ equipped with the uniform operator norm, $\|A\|_{\mathcal{L}_\mathcal{H}} = \sup_{\|x\|\leq 1} \|Ax\|$. For $A \in \mathcal{L}_\mathcal{H}$, we denote the kernel and range of A by $\ker A$ and $\operatorname{ran} A$, respectively. The dimension of $\operatorname{ran} A$ is called the rank of A.

The adjoint of an operator $A \in \mathcal{L}_\mathcal{H}$ is denoted by A^*. A linear operator $A \in \mathcal{L}_\mathcal{H}$ is said to be positive semi-definite (resp. positive definite) if $\langle Ax, x\rangle \geq 0$ (resp. $\langle Ax, x\rangle > 0$) for any $x \in \mathcal{H}$. In this paper, $f \otimes g$ denotes the operation $(f, g) \mapsto \langle f, \cdot\rangle g$ for $f, g \in \mathcal{H}$.

Sometimes we need to restrict the domain and the codomain of a bounded linear operator. Whenever this is required, we let $A|_{M_1 \to M_2}$ denote the operator $A \in \mathcal{L}_\mathcal{H}$ whose domain is $M_1 \subset \mathcal{H}$ and codomain is $M_2 \subset \mathcal{H}$.

2.1 $\mathcal{H}$-valued random variables

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the underlying probability space. An $\mathcal{H}$-valued random element Z is a measurable function from Ω to $\mathcal{H}$, where $\mathcal{H}$ is understood to be equipped with the Borel σ-field. The random element is said to be integrable if $\mathbb{E}\|Z\| < \infty$ and square-integrable if $\mathbb{E}\|Z\|^2 < \infty$. If the random element is integrable, there exists a unique element $\mu \in \mathcal{H}$ such that $\mathbb{E}\langle Z, f\rangle = \langle \mu, f\rangle$ for any $f \in \mathcal{H}$, and such μ is called the mean function of Z. If the random element is square-integrable, we define the covariance operator of Z as $C_z = \mathbb{E}[Z \otimes Z]$.

2.2 I(1) and cointegrated linear processes in $\mathcal{H}$

We require a notion of I(1) sequences and cointegration in our Hilbert space setting. To this end, we adopt the setting of Beare et al. (2017), who generalize the concept of a cointegrated I(1) sequence to an arbitrary complex Hilbert space and provide a rigorous mathematical treatment, although some of the concepts were introduced earlier by Chang et al. (2016).

A sequence $(X_t, t \geq 0)$ is said to be I(1) if its first differences $\Delta X_t = X_t - X_{t-1}$ satisfy

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

where $(\varepsilon_t, t \in \mathbb{Z})$ is a square-integrable i.i.d. sequence, and $(\Phi_j, j \geq 0)$ is a sequence in $\mathcal{L}_\mathcal{H}$ satisfying $\sum_{j=0}^{\infty} j\|\Phi_j\|_{\mathcal{L}_\mathcal{H}} < \infty$ and $\Phi(1) = \sum_{j=0}^{\infty} \Phi_j \neq 0$. We assume that the covariance operator C_ε of ε_t is positive definite and denote the long-run covariance operator of $(\Delta X_t, t \geq 1)$ by $\Lambda_{\Delta X} = \Phi(1) C_\varepsilon \Phi(1)^*$. For the sequence (2.1), the Beveridge-Nelson decomposition is

$$\Delta X_t = \Phi(1)\varepsilon_t + \nu_t - \nu_{t-1}, \quad t \geq 1, \tag{2.2}$$

where $\nu_t = \sum_{j=0}^{\infty} \tilde{\Phi}_j \varepsilon_{t-j}$ and $\tilde{\Phi}_j = -\sum_{k=j+1}^{\infty} \Phi_k$; see Phillips and Solo (1992).

The stationary subspace (cointegrating space) of X is the collection of all $h \in \mathcal{H}$ such that the scalar sequence $(\langle X_t, h\rangle, t \geq 0)$ is stationary for a suitable choice of X_0. Beare et al. (2017) showed that this space is given by $\ker \Lambda_{\Delta X}$. Since C_ε is positive definite, the stationary subspace is equal to $[\operatorname{ran} \Phi(1)]^\perp$. The nonstationary subspace (attractor space) is defined as the orthogonal complement of the stationary subspace, i.e. $\operatorname{cl} \operatorname{ran} \Phi(1)$. This is the subspace of $\mathcal{H}$ in which the I(1) stochastic trend in the Beveridge-Nelson decomposition (2.2) takes values. Throughout, $\mathfrak{A}$ denotes the nonstationary subspace and $\mathfrak{A}^\perp$ denotes the stationary subspace.

Given the direct sum decomposition $\mathcal{H} = \mathfrak{A} \oplus \mathfrak{A}^\perp$, any element $h \in \mathcal{H}$ can be uniquely decomposed as $h = h_{\mathfrak{A}} + h_{\mathfrak{A}^\perp}$ for some elements $h_{\mathfrak{A}} \in \mathfrak{A}$ and $h_{\mathfrak{A}^\perp} \in \mathfrak{A}^\perp$. If $h_{\mathfrak{A}} \neq 0$ then $\langle X_t, h\rangle$ is I(1). If in fact $h \in \mathfrak{A}$, then $\langle X_t, h\rangle$ may be called a stochastic trend, generalizing the notion from finite-dimensional space. Furthermore, when the nonstationary subspace, $\mathfrak{A}$, is finite-dimensional, its dimension can be called the number of common stochastic trends.

3 Inference on the dimension of the nonstationary subspace

In this section, we provide a statistical procedure to determine the dimension of the nonstationary subspace. For simplicity, we first focus on the case where $(X_t, t \geq 1)$ has zero mean. In Section 3.4, we consider I(1) sequences with deterministic components.

3.1 Model and hypotheses of interest

We apply the following assumption throughout.

Assumption 1. $(X_t, t \geq 1)$ satisfies the conditions stated in Section 2.2, and in particular (2.1), (2.2). The rank of $\Phi(1)$ is given by the integer s in $[0, \infty)$.

Under Assumption 1, the nonstationary subspace $\mathfrak{A}$ is finite-dimensional and given by $\operatorname{ran} \Phi(1)$ since any finite-dimensional subspace is closed. Moreover, we have the direct sum decomposition $\mathcal{H} = \operatorname{ran} \Phi(1) \oplus [\operatorname{ran} \Phi(1)]^\perp = \mathfrak{A} \oplus \mathfrak{A}^\perp$ and the relation $\dim(\mathfrak{A}) = \operatorname{rank}(\Phi(1)) = s$.

Remark 1. For example, a sufficient condition for $\mathfrak{A}$ to be finite-dimensional is that $(X_t, t \geq 1)$ is autoregressive with compact autoregressive operators; see Beare and Seo (2020, Remark 3.6) or Franchi and Paruolo (2020). As another example, the functional observations could be generated by a factor model, where the (possibly nonstationary) factors could represent level, slope, curvature, etc., as in the Nelson and Siegel (1987) term structure model. Finally, our empirical examples in Section 5 as well as those in Chang et al. (2016) all suggest that $s < \infty$ is empirically reasonable. □

Let $(\phi_j, j \in \mathbb{N})$ be an orthonormal basis of $\mathcal{H}$ satisfying $\operatorname{span}(\phi_1, \ldots, \phi_s) = \operatorname{ran} \Phi(1)$. We may then understand X_t as the basis expansion

$$X_t = \sum_{j=1}^{\infty} \langle X_t, \phi_j \rangle \phi_j \quad \text{with} \quad \sum_{j=1}^{\infty} \langle X_t, \phi_j \rangle^2 < \infty \text{ almost surely.}$$

Let $\ell^2(\mathbb{N})$ denote the space of square-summable sequences equipped with the inner product $\langle\{x_j\},\{y_j\}\rangle_{\ell^2(\mathbb{N})} = \sum_{j=1}^{\infty} x_j y_j$ for $\{x_j\},\{y_j\} \in \ell^2(\mathbb{N})$. Then, under the isomorphism from $\mathcal{H}$ to $\ell^2(\mathbb{N})$, $(X_t, t \geq 0)$ may be viewed as the random infinite sequence

$$(\langle X_t, \phi_1\rangle, \ldots, \langle X_t, \phi_s\rangle, \langle X_t, \phi_{s+1}\rangle, \ldots), \quad t \geq 0. \tag{3.1}$$

Clearly, since $\operatorname{span}(\phi_1, \ldots, \phi_s) = \operatorname{ran}\Phi(1)$, the first s components of (3.1) are scalar-valued I(1) processes because

$$\Delta\langle X_t, \phi_j\rangle = \langle \Phi(1)\varepsilon_t, \phi_j\rangle + \langle \nu_t - \nu_{t-1}, \phi_j\rangle, \quad j = 1, \ldots, s,$$

are stationary with long-run covariances $\langle \Lambda_{\Delta X}\phi_j, \phi_j\rangle \neq 0$. On the other hand, the (infinitely many) remaining components of (3.1) are all stationary for a suitable choice of X_0 because

$$\langle X_t, \phi_j\rangle = \langle X_0 - \nu_0, \phi_j\rangle + \langle \nu_t, \phi_j\rangle, \quad j \geq s+1,$$

and $\langle \nu_t, \phi_j\rangle$ is stationary since $\langle \nu_t, \cdot\rangle$ is a measurable transformation of a stationary sequence. Moreover, $\langle X_t, \phi_j\rangle$ has nonzero long-run covariance if $\sum_{j=1}^{\infty} j\Phi_j \neq 0$, i.e. it is I(0) in this case.

It thus follows that the dimension of the nonstationary subspace, $\dim(\mathfrak{A}) = \operatorname{rank}(\Phi(1))$, can be interpreted as the number of stochastic trends embedded in the functional time series $(X_t, t \geq 0)$. We consider hypothesis testing on $\dim(\mathfrak{A})$ and provide a statistical procedure to determine $\dim(\mathfrak{A})$. In finite-dimensional Euclidean space, $\mathbb{C}^n$ or $\mathbb{R}^n$, this is closely related to cointegration rank (see references in the Introduction). If the cointegration rank is $r \leq n$ in $\mathbb{R}^n$, then there are $n-r$ stochastic trends in that setting. However, under Assumption 1 in our infinite-dimensional setting, only the dimension of the nonstationary subspace (number of stochastic trends) may be finitely identified while the cointegration rank is always ∞. Hence, it may not be proper to call our test a cointegration rank test, but it still may be viewed as a generalization of conventional cointegration rank tests from finite-dimensional space.

We apply the following assumption to obtain asymptotic results.

Assumption 2. In (2.1), (2.2) it holds that (i) $(\varepsilon_t, t \in \mathbb{Z})$ is an i.i.d. sequence with $\mathbb{E}\|\varepsilon_t\|^4 < \infty$; (ii) the covariance operator C_ν of $(\nu_t, t \geq 1)$ is positive definite on $\mathfrak{A}^\perp$, i.e., $\langle C_\nu x, x\rangle > 0$ for all $x \in \mathfrak{A}^\perp$.

The above assumption is convenient in our asymptotic analysis. Specifically, Assumption 2(i) is a standard condition to obtain weak convergence of linear processes in $\mathcal{H}$ as in Berkes et al. (2013). Assumption 2(ii), which does not seem restrictive in practice, ensures that a particular limiting quantity is positive definite; see (B.4) in the appendix.

We consider the following null and alternative hypotheses,

$$H_0 : \dim(\mathfrak{A}) = s_0 \quad vs \quad H_1 : \dim(\mathfrak{A}) \leq s_0 - 1. \tag{3.2}$$

The null hypothesis in (3.2) can be either a pre-specified hypothesis of interest, or (3.2) can

be applied sequentially to estimate s. We explore the latter possibility in Theorem 2.

3.2 Preliminary asymptotic analysis of covariance operators

We first fix notation for the subsequent discussion. Since $\operatorname{ran}\Lambda_{\Delta X} = \operatorname{ran}\Phi(1)$ under Assumption 1, only the first s eigenvalues of $\Lambda_{\Delta X}$ are nonzero. We let $((\alpha_j, \eta_j), j = 1, \ldots, s)$ denote the pairs of eigenvalues and eigenvectors of $\Lambda_{\Delta X}$ and assume $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_s > 0$ without loss of generality. Note that these can be used to define $\Lambda_{\Delta X}^{1/2}$ and $\Lambda_{\Delta X}^{-1/2}$ by the spectral decomposition. In addition, for convenience, we let $(\eta_j, j \geq s+1)$ denote an orthonormal basis of $[\operatorname{ran}\Phi(1)]^{\perp}$ so that $(\eta_j, j \in \mathbb{N})$ is an orthonormal basis of $\mathcal{H}$.

Let $(\mathcal{W}(r), r \in [0,1])$ denote a Brownian motion taking values in $\mathcal{H}$ with covariance operator $\sum_{j=1}^{s} \eta_j \otimes \eta_j$ and define $\mathcal{V}(r) = \int_0^r \mathcal{W}(w)dw$ for $r \in [0,1]$. Then

$$\begin{aligned} &\langle \mathcal{W}(r), \eta_j \rangle \overset{d}{=} \mathcal{W}_j(r) \quad \text{and} \quad \langle \mathcal{V}(r), \eta_j \rangle \overset{d}{=} \mathcal{V}_j(r), \quad j = 1, \ldots, s, \\ &\langle \mathcal{W}(r), \eta_j \rangle \overset{d}{=} 0 \qquad \text{and} \quad \langle \mathcal{V}(r), \eta_j \rangle \overset{d}{=} 0, \qquad j \geq s+1, \end{aligned} \tag{3.3}$$

where "$\overset{d}{=}$" denotes equality in distribution, $\mathcal{W}_j$ is a sequence of standard Brownian motions independent across j, and $\mathcal{V}_j(r) = \int_0^r \mathcal{W}_j(w)dw$. Under the isomorphism between s-dimensional real Hilbert space with element h and the Euclidean space $\mathbb{R}^s$ with element $(\langle h, \eta_1 \rangle, \langle h, \eta_2 \rangle, \cdots, \langle h, \eta_s \rangle)'$, we can consider $(\mathcal{W}(r), r \in [0,1])$ and $(\mathcal{V}(r), r \in [0,1])$ as the usual s-dimensional standard Brownian motion and integrated standard Brownian motion.

We define two random operators associated with $(\mathcal{W}(r), r \in [0,1])$ and $(\mathcal{V}(r), r \in [0,1])$ as

$$\widetilde{\mathcal{W}} = \int \mathcal{W}(r) \otimes \mathcal{W}(r)dr \quad \text{and} \quad \widetilde{\mathcal{V}} = \int \mathcal{V}(r) \otimes \mathcal{V}(r)dr. \tag{3.4}$$

Under the above-mentioned isomorphism, these operators may also be understood as random matrices taking values in $\mathbb{R}^{s \times s}$. Then we find from (3.3) and (3.4) that

$$\begin{aligned} &\langle \eta_i, \widetilde{\mathcal{W}}(\eta_j) \rangle \overset{d}{=} \int \mathcal{W}_i(r)\mathcal{W}_j(r)dr \quad \text{and} \quad \langle \eta_i, \widetilde{\mathcal{V}}(\eta_j) \rangle \overset{d}{=} \int \mathcal{V}_i(r)\mathcal{V}_j(r)dr, \quad 1 \leq i, j \leq s, \\ &\langle \eta_i, \widetilde{\mathcal{W}}(\eta_j) \rangle \overset{d}{=} 0 \qquad\qquad\qquad \text{and} \quad \langle \eta_i, \widetilde{\mathcal{V}}(\eta_j) \rangle \overset{d}{=} 0, \qquad\qquad \text{otherwise.} \end{aligned}$$

Given functional observations $(X_t, t = 1, \ldots, T)$, let $Y_t = \sum_{j=1}^{t} X_j$ for $t = 1, \ldots, T$. We define the unnormalized sample covariance operators

$$\hat{\mathcal{C}} = \sum_{t=1}^{T} X_t \otimes X_t \quad \text{and} \quad \hat{\mathcal{K}} = \sum_{t=1}^{T} Y_t \otimes Y_t. \tag{3.5}$$

Asymptotic properties of $\hat{\mathcal{C}}$ and $\hat{\mathcal{K}}$ play a crucial role in our analysis.

In particular, Lemma 1 in Appendix A shows that $T^{-2}\hat{\mathcal{C}}$ and $T^{-4}\hat{\mathcal{K}}$ converge to $\mathcal{C}$ and $\mathcal{K}$, respectively, where

$$\mathcal{C} \overset{d}{=} \Lambda_{\Delta X}^{1/2} \widetilde{\mathcal{W}} \Lambda_{\Delta X}^{1/2} \quad \text{and} \quad \mathcal{K} \overset{d}{=} \Lambda_{\Delta X}^{1/2} \widetilde{\mathcal{V}} \Lambda_{\Delta X}^{1/2}. \tag{3.6}$$

Given the definition of $\Lambda_{\Delta X}^{1/2}$, for any element of $h \in \mathcal{H}$ allowing the unique decomposition

$h = h_{\mathfrak{A}} + h_{\mathfrak{A}^\perp}$, we have

$$\mathcal{C}h = \mathcal{C}(h_{\mathfrak{A}} + h_{\mathfrak{A}^\perp}) = \mathcal{C}h_{\mathfrak{A}} \in \mathfrak{A} \quad \text{and} \quad \mathcal{K}h = \mathcal{K}(h_{\mathfrak{A}} + h_{\mathfrak{A}^\perp}) = \mathcal{K}h_{\mathfrak{A}} \in \mathfrak{A}. \tag{3.7}$$

That is, $\mathcal{C}$ and $\mathcal{K}$ eliminate any (infinite-dimensional) component in $\mathfrak{A}^\perp$ and leave only a (finite-dimensional) component in $\mathfrak{A}$. Moreover, the operators $\mathcal{C}$ and $\mathcal{K}$ are almost surely invertible on $\mathfrak{A}$, even if they are not invertible on $\mathcal{H}$. Note that, when $\mathcal{H} = \mathbb{R}^n$, this result specializes to $T^{-2}\sum_t X_t X_t' \xrightarrow{d} C$, where C satisfies $Ch = h'C = 0$ for $h \in \mathfrak{A}^\perp$ and $h'Ch$ is invertible almost surely for $h = [h_1, \ldots, h_s]$ with orthonormal vectors $h_1, \ldots, h_s \in \mathfrak{A}$.

3.3 Variance ratio test

Suppose first that we have a projection operator, denoted P_ℓ, whose range is an $\ell \geq s$ dimensional subspace that contains $\mathfrak{A}$. Since $((I - P_\ell)X_t, t \geq 1)$ is a stationary sequence, its nonstationary subspace is $\{0\}$. Thus, we may disregard this part of $(X_t, t \geq 1)$ and focus on the projected time series $(P_\ell X_t, t \geq 1)$. The latter is isomorphic to an ℓ-dimensional multivariate time series with $\ell - s$ cointegrating relationships and s (linearly independent) stochastic trends, i.e. with s-dimensional nonstationary subspace.

Of course, the assumption that P_ℓ is known is not reasonable in practice, so we need to replace P_ℓ with an estimate. We first apply the following high level condition.

Assumption 3. For some finite integer $\ell \geq s$ there exists $(\phi_1^T, \ldots, \phi_\ell^T)$ such that

$$P_\ell^T = \sum_{j=1}^{\ell} \phi_j^T \otimes \phi_j^T \quad \text{and} \quad \|P_\ell^T x - x\| = o_p(1) \quad \text{for any } x \in \mathfrak{A}. \tag{3.8}$$

Intuitively, Assumption 3 requires a finite collection of vectors whose span asymptotically includes the nonstationary subspace, $\mathfrak{A}$. We call ran P_ℓ^T an estimate of an asymptotic superspace of $\mathfrak{A}$. Of course, there are many possible empirical projection operators that could be applied in practice (e.g., based on eigenanalysis of various covariance, long-run covariance, or autocovariance operators), so we find it convenient to use Assumption 3 as a practically important guideline for what a candidate projection operator needs to satisfy. In Section 3.5 we discuss some practical, data-dependent choices of ℓ and P_ℓ^T that satisfy Assumption 3.

Some asymptotic implications of Assumption 3 are discussed in the following remark.

Remark 2. Without loss of generality, we may assume that $(\phi_1^T, \ldots, \phi_s^T)$ converges to some orthonormal basis $(\phi_1, \ldots, \phi_s)$ of $\mathfrak{A}$ under Assumption 3. The orthonormal set $(\phi_1, \ldots, \phi_s)$ may be random elements; we do not require that $(\phi_1^T, \ldots, \phi_s^T)$ to converges to a fixed orthonormal basis of $\mathfrak{A}$. Under Assumption 3, we also have

$$\|\phi_j^T - (I - P_{\mathfrak{A}})\phi_j^T\| = o_p(1), \quad s+1 \leq j \leq \ell.$$

That is, $(\phi_j^T, j = s+1, \ldots, \ell)$ is asymptotically included in $\mathfrak{A}^\perp$. It should be noted that we do not require any limiting behavior of ϕ_j^T for $j \geq s+1$ in our asymptotic analysis, i.e. it is

not required to converge to any (random or fixed) element of $\mathcal{H}$. □

On the subspace $\operatorname{ran} P_\ell^T$, consider the generalized eigenvalue problem

$$\tau_j^T P_\ell^T \hat{\mathcal{K}} P_\ell^T \xi_j^T = P_\ell^T \hat{\mathcal{C}} P_\ell^T \xi_j^T, \quad \xi_j^T \in \operatorname{ran} P_\ell^T. \tag{3.9}$$

The solution to the eigenvalue problem (3.9) can be characterized as

$$\tau_j^T = \frac{\langle P_\ell^T \hat{\mathcal{C}} P_\ell^T \xi_j^T, \xi_j^T \rangle}{\langle P_\ell^T \hat{\mathcal{K}} P_\ell^T \xi_j^T, \xi_j^T \rangle}. \tag{3.10}$$

The stochastic order and limiting behavior of the eigenvalue, τ_j^T, is different depending on whether or not the corresponding eigenvector, ξ_j^T, falls inside the nonstationary subspace, $\mathfrak{A}$. This is described in detail in Lemma 2 in Appendix A, where it is shown that $T^2\tau_j^T$ converges to a non-degenerate limit if $\xi_j^T \in \mathfrak{A}$ while $(T\tau_j^T)^{-1} = O_p(1)$ if $\xi_j^T \in \mathfrak{A}^\perp$, i.e. $T^2\tau_j^T$ diverges to infinity in the latter case. Thus, the s smallest eigenvalues have different stochastic order than the remaining eigenvalues and $T^2 \sum_{j=1}^s \tau_j^T$ has a non-degenerate limiting distribution.

In view of (3.10), we call $T^2 \sum_{j=1}^s \tau_j^T$ a variance ratio statistic. For univariate time series it reduces to the well-known KPSS statistic of Kwiatkowski et al. (1992) with zero bandwidth, and for finite-dimensional multivariate time series it reduces to the statistic considered by Breitung (2002). We prove the following theorem, which suggests a testing procedure to determine the dimension of the nonstationary subspace in our Hilbert space setting.

Theorem 1. *Suppose that Assumptions 1–3 hold, and let* $(\tau_1^T, \ldots, \tau_\ell^T)$ *with* $\tau_1^T \leq \ldots \leq \tau_\ell^T$ *and* $(\xi_1^T, \ldots, \xi_\ell^T) \subset \mathcal{H}$ *be the pairs of eigenvalues and eigenvectors satisfying (3.9). Then*

$$T^2 \sum_{j=1}^{s} \tau_j^T \xrightarrow{d} \operatorname{tr}\left((\widetilde{\mathcal{V}}|_{\mathfrak{A}\to\mathfrak{A}})^{-1} \widetilde{\mathcal{W}}|_{\mathfrak{A}\to\mathfrak{A}} \right), \tag{3.11}$$

$$T^2 \sum_{j=1}^{q} \tau_j^T \xrightarrow{p} \infty \quad \textit{for any } s+1 \leq q \leq \ell. \tag{3.12}$$

Remark 3. Suppose again that P_ℓ is known and replaces P_ℓ^T. In this case, additional intuition can be gained by considering the isomorphism between $\operatorname{ran} P_\ell$ and $\mathbb{R}^\ell$, where an element $x \in \operatorname{ran} P_\ell$, with the unique basis expansion $x = \sum_{j=1}^\ell \langle x, x_j \rangle x_j$ for an orthonormal basis $(x_1, \ldots, x_\ell)$ of $\operatorname{ran} P_\ell$, is identified as the vector $(\langle x, x_1 \rangle, \ldots, \langle x, x_\ell \rangle)'$. Under this isomorphism, $(P_\ell X_t, t \geq 1)$ may be viewed as a multivariate ℓ-dimensional time series and (3.9) may be viewed as a generalized eigenvalue problem in $\mathbb{R}^\ell$, where $P_\ell \hat{\mathcal{C}} P_\ell$ and $P_\ell \hat{\mathcal{K}} P_\ell$ appearing in (3.9)–(3.10) are isomorphic to the sample covariance matrices of this ℓ-dimensional time series and its cumulated sum, respectively. Then, as noted by Breitung (2002) for cointegrated systems in a Euclidean space setting, the stochastic order of τ_j^T depends on whether or not the corresponding eigenvector ξ_j^T falls inside the span of the cointegrating vectors (or equivalently the attractor space). □

Remark 4. Theorem 1 suggests a consistent test for the hypothesis in (3.2). Specifically, the statistic $T^2 \sum_{j=1}^{s_0} \tau_j^T$ has a well-defined limiting distribution under H_0, while it diverges to infinity under H_1. For a given significance level α, we therefore reject H_0 when $T^2 \sum_{j=1}^{s_0} \tau_j^T > c_\alpha$ for some c_α depending on α. Using the isomorphism between $\mathfrak{A}$ and $\mathbb{R}^s$, it follows that the asymptotic distribution in (3.11) satisfies

$$\operatorname{tr}\left((\widetilde{\mathcal{V}}|_{\mathfrak{A}\to\mathfrak{A}})^{-1}\widetilde{\mathcal{W}}|_{\mathfrak{A}\to\mathfrak{A}}\right) \stackrel{d}{=} \operatorname{tr}\left(\left(\int_0^1 V_s(r)V_s'(r)dr\right)^{-1}\int_0^1 W_s(r)W_s'(r)dr\right), \tag{3.13}$$

where W_s is s-dimensional standard Brownian motion and $V_s(r) = \int_0^r W_s(w)dw$. Hence, critical values c_α for the test statistic, $T^2 \sum_{j=1}^{s_0} \tau_j^T$, can be simulated from (3.13) with $s = s_0$ by standard methods. □

To estimate or determine the dimension s of $\mathfrak{A}$, we apply a top-down procedure, where we sequentially test (3.2) with $s_0 = s_{\max}, s_{\max} - 1, \ldots, 1$ for some reasonably chosen $s_{\max}$ and using nominal level α. The estimate of the dimension of the nonstationary subspace, $\hat{s}$, is then given by the first non-rejected null hypothesis. If H_0 is rejected for all values of s_0 considered, then we set $\hat{s} = 0$. The following result then follows from Theorem 1.

Theorem 2. *Suppose that the assumptions of Theorem 1 hold and that* $\mathbb{P}\{s_{\max} \geq s\} \to 1$. *Then, for a fixed nominal level* α, *the test described in Remark 4 is consistent and*

$$\mathbb{P}\{\hat{s} = s\} \to 1 - \alpha \quad \textit{and} \quad \mathbb{P}\{\hat{s} > s\} \to 0. \tag{3.14}$$

If the nominal level is chosen such that $\alpha \to 0$ *as* $T \to \infty$ *then* $\mathbb{P}\{\hat{s} = s\} \to 1$.

We note that Theorem 2 requires that $\mathbb{P}\{s_{\max} \geq s\} \to 1$. In theory, this could be guaranteed by letting $s_{\max} \to \infty$ as $T \to \infty$. In practice we would normally expect s to be quite small, so this would be guaranteed by letting $s_{\max}$ be some moderate number. As suggested by Chang et al. (2016, footnotes 4 and 13), $s_{\max}$ could be determined by graphical methods, like eigenvalue plots, or by the number of functional principal components that determine a large proportion of the total variance. The following remark considers the consequences of selecting $s_{\max} < s$.

Remark 5. From our proof of Theorem 1, it may easily be deduced that if $s_{\max} < s$, then

$$T^2 \sum_{j=1}^{s_{\max}} \tau_j^T \xrightarrow{d} \operatorname{tr}\left((\widetilde{\mathcal{V}}|_{\mathfrak{A}'\to\mathfrak{A}'})^{-1}\widetilde{\mathcal{W}}|_{\mathfrak{A}'\to\mathfrak{A}'}\right), \tag{3.15}$$

where $\mathfrak{A}'$ denotes some $s_{\max}$-dimensional subspace of $\mathfrak{A}$. As in Remark 4, the isomorphism between $\mathbb{R}^{s_{\max}}$ and any $s_{\max}$-dimensional subspace of $\mathcal{H}$ implies that the limit in (3.15) satisfies

$$\operatorname{tr}\left((\widetilde{\mathcal{V}}|_{\mathfrak{A}'\to\mathfrak{A}'})^{-1}\widetilde{\mathcal{W}}|_{\mathfrak{A}'\to\mathfrak{A}'}\right) \stackrel{d}{=} \operatorname{tr}\left(\left(\int_0^1 V_{s_{\max}}(r)V_{s_{\max}}'(r)dr\right)^{-1}\int_0^1 W_{s_{\max}}(r)W_{s_{\max}}'(r)dr\right).$$

Because $s_{\max} < s$, all the eigenvalues in the test statistic for the first hypothesis in the

sequential procedure, namely $H_0 : \dim(\mathfrak{A}) = s_{\max}$, are convergent and satisfy (3.15). It follows that, in this case, instead of (3.14) we have $\mathbb{P}\{\hat{s} = s_{\max}\} \to 1 - \alpha$. □

Remark 6. The consequence of Remark 5 is that, when $s_{\max} < s$, the sequential testing procedure will conclude that $\hat{s} = s_{\max}$ with probability converging to $1 - \alpha$ as $T \to \infty$. Therefore, in practice, if the sequential procedure results in $\hat{s} = s_{\max}$ it seems prudent to restart with a higher value of $s_{\max}$. Indeed, our test is very robust to choice of $s_{\max}$, even when $s_{\max}$ is much larger than s. For simulation evidence on this point, see Section 4.4 and Table 4, and for empirical evidence, see Section 5.4 and Figure 5. □

3.4 Deterministic components

Until now we assumed that $(X_t, t \geq 1)$ has mean zero. We now adapt the discussion to allow a deterministic component. A nonzero intercept function or a linear trend function seem most relevant in practice, so we focus on those, but an extension to more general deterministic components requires only a slight modification of the subsequent discussion.

Specifically, for some functions $\mu_1, \mu_2 \in \mathcal{H}$, we consider the unobserved components model

$$X_t = \mu_1 + \mu_2 t + U_t, \tag{3.16}$$

where $(U_t, t \geq 1)$ is an unobserved I(1) sequence with mean zero that is generated like $(X_t, t \geq 1)$ in the previous sections. The model (3.16) includes the intercept function, μ_1, and the linear trend function, μ_2. If only an intercept function is wanted, then we set $\mu_2 = 0$.

We define the functional residuals from least squares estimation,

$$U_t^{(1)} = X_t - \frac{1}{T}\sum_{t=1}^{T} X_t \quad \text{and} \quad U_t^{(2)} = U_t^{(1)} - \left(t - \frac{T+1}{2}\right) \frac{\sum_{t=1}^{T}\left(t - \frac{T+1}{2}\right) X_t}{\sum_{t=1}^{T}\left(t - \frac{T+1}{2}\right)^2};$$

see Kokoszka and Young (2016) for details. Here, the superscript $^{(1)}$ denotes residuals from the model with only an intercept function, while $^{(2)}$ denotes residuals from the model with both intercept and linear trend functions. Analogously to (3.5) we define, for $\varrho = 1, 2$,

$$\hat{\mathcal{C}}^{(\varrho)} = \sum_{t=1}^{T} U_t^{(\varrho)} \otimes U_t^{(\varrho)} \quad \text{and} \quad \hat{\mathcal{K}}^{(\varrho)} = \sum_{t=1}^{T} \left(\sum_{j=1}^{t} U_j^{(\varrho)} \otimes \sum_{j=1}^{t} U_j^{(\varrho)} \right). \tag{3.17}$$

We then consider the following generalized eigenvalue problem,

$$\tau_{j,(\varrho)}^T P_\ell^T \hat{\mathcal{K}}^{(\varrho)} P_\ell^T \xi_{j,(\varrho)}^T = P_\ell^T \hat{\mathcal{C}}^{(\varrho)} P_\ell^T \xi_{j,(\varrho)}^T, \quad \xi_{j,(\varrho)}^T \in \operatorname{ran} P_\ell^T. \tag{3.18}$$

To describe the asymptotic distributions, we let $(\mathcal{W}^{(\varrho)}(r), r \in [0,1])$ denote a demeaned (for $\varrho = 1$) or detrended (for $\varrho = 2$) Brownian motion taking values in $\mathcal{H}$ with covariance operator $\sum_{j=1}^{s} \eta_j \otimes \eta_j$ and define $\mathcal{V}^{(\varrho)}(r) = \int_0^r \mathcal{W}^{(\varrho)}(w)dw$ for $r \in [0,1]$ and $\varrho = 1, 2$. Based on these, we then define $\widetilde{\mathcal{W}}^{(\varrho)}$ and $\widetilde{\mathcal{V}}^{(\varrho)}$ as in (3.4).

Theorem 3. *Suppose that Assumptions 1–3 hold, and let* $(\tau^T_{1,(\varrho)}, \ldots, \tau^T_{\ell,(\varrho)})$ *with* $\tau^T_{1,(\varrho)} \leq \ldots \leq \tau^T_{\ell,(\varrho)}$ *and* $(\xi^T_{1,(\varrho)}, \ldots, \xi^T_{\ell,(\varrho)}) \subset \mathcal{H}$ *be the pairs of eigenvalues and eigenvectors satisfying (3.18). Then, for* $\varrho = 1, 2$,

$$T^2 \sum_{j=1}^{s} \tau^T_{j,(\varrho)} \xrightarrow{d} \operatorname{tr}\left((\widetilde{\mathcal{V}}^{(\varrho)}|_{\mathfrak{A}\to\mathfrak{A}})^{-1}\widetilde{\mathcal{W}}^{(\varrho)}|_{\mathfrak{A}\to\mathfrak{A}}\right), \tag{3.19}$$

$$T^2 \sum_{j=1}^{q} \tau^T_{j,(\varrho)} \xrightarrow{p} \infty \quad \text{for any } s+1 \leq q \leq \ell. \tag{3.20}$$

Remark 7. As in Remark 4, it can be shown that the limiting distribution in (3.19) satisfies

$$\operatorname{tr}\left((\widetilde{\mathcal{V}}^{(\varrho)}|_{\mathfrak{A}\to\mathfrak{A}})^{-1}\widetilde{\mathcal{W}}^{(\varrho)}|_{\mathfrak{A}\to\mathfrak{A}}\right) \stackrel{d}{=} \operatorname{tr}\left(\left(\int_0^1 V_s^{(\varrho)}(r)V_s^{(\varrho)\prime}(r)dr\right)^{-1}\int_0^1 W_s^{(\varrho)}(r)W_s^{(\varrho)\prime}(r)dr\right), \tag{3.21}$$

where $W_s^{(\varrho)}$ is s-dimensional demeaned (resp. detrended) Brownian motion when $\varrho = 1$ (resp. $\varrho = 2$), and $V_s^{(\varrho)}(r) = \int_0^r W_s^{(\varrho)}(w)dw$. As in Remark 4, quantiles of this distribution can be found by simulation. For $\varrho = 1, 2$ and $s_0 \leq 8$, these are tabulated in Breitung (2002). □

3.5 Practical choice of P^T_ℓ for feasible test

In this section we discuss practical, data-dependent choices of the projection operator, P^T_ℓ, that asymptotically span the nonstationary subspace, i.e. satisfy Assumption 3. To avoid basing the generalized eigenvalue problems in (3.9) and (3.18) on high-dimensional covariance matrices, which could lead to inaccuracy or even inconsistency of eigenvalues (Yao et al., 2012), we suggest projection operators constructed from a small set of orthonormal vectors. The next remark discusses one such projection, based on $\hat{\mathcal{C}}$ or $\hat{\mathcal{C}}^{(\varrho)}$, as applied by Chang et al. (2016). The subsequent theorem discusses another projection, based on our $\hat{\mathcal{K}}$ or $\hat{\mathcal{K}}^{(\varrho)}$

Remark 8. Let $\widetilde{\mathcal{C}}$ equal $\hat{\mathcal{C}}$ or $\hat{\mathcal{C}}^{(\varrho)}$ for $\varrho = 1, 2$, depending on the specification of deterministic components, and let $\{\gamma^T_j, \phi^T_j\}$ be the pairs of eigenvalues and eigenvectors satisfying $\gamma^T_j \phi^T_j = \widetilde{\mathcal{C}}\phi^T_j$ for $\gamma^T_1 \geq \gamma^T_2 \geq \ldots$. If Assumptions 1 and 2 hold, it follows from Theorem 3.3 of Chang et al. (2016) that $P^T_\ell = \sum_{j=1}^{\ell} \phi^T_j \otimes \phi^T_j$ satisfies Assumption 3 for any $\ell \geq s$. □

Theorem 4. *Suppose that Assumptions 1 and 2 hold. Let* $\widetilde{\mathcal{K}}$ *equal* $\hat{\mathcal{K}}$ *or* $\hat{\mathcal{K}}^{(\varrho)}$ *for* $\varrho = 1, 2$, *depending on the specification of deterministic components, and let* $\{\gamma^T_j, \phi^T_j\}$ *be the pairs of eigenvalues and eigenvectors satisfying*

$$\gamma^T_j \phi^T_j = \widetilde{\mathcal{K}}\phi^T_j \tag{3.22}$$

for $\gamma^T_1 \geq \gamma^T_2 \geq \ldots$. *Then* $P^T_\ell = \sum_{j=1}^{\ell} \phi^T_j \otimes \phi^T_j$ *satisfies Assumption 3 for any* $\ell \geq s$.

In Theorem 4, we suggest an estimate P^T_ℓ based on eigenanalysis of $\widetilde{\mathcal{K}}$ following the idea of Chang et al. (2016) who estimate P^T_ℓ based on eigenanalysis of $\widetilde{\mathcal{C}}$; see Remark 8. In either case, the space spanned by the first s eigenvectors converges to the nonstationary subspace, such that Assumption 3 is satisfied for any $\ell \geq s$. We use these results to form a test statistic

associated with the null hypothesis in (3.2) following Remark 4. We let the required estimate P_ℓ^T be based on $\widetilde{\mathcal{K}}$ and Theorem 4, and define the test statistic

$$\mathcal{T}_{\mathcal{K}} = T^2 \sum_{j=1}^{s_0} \tau_j^T. \tag{3.23}$$

Remark 9. Of course, we could also form a test statistic based on Remark 8 and $\widetilde{\mathcal{C}}$ instead of $\widetilde{\mathcal{K}}$. It follows from Theorem 4 and Remark 8 that all our previous results and remarks apply to both versions of the test. From our experience in simulations, the statistic $\mathcal{T}_{\mathcal{K}}$ in (3.23) is preferred in terms of finite-sample performance. □

The statistic $\mathcal{T}_{\mathcal{K}}$ depends on the tuning parameter ℓ, the choice of which is discussed next. However, we note that the limiting distributions of $\mathcal{T}_{\mathcal{K}}$ in Theorems 1 and 3 do not depend on any nuisance parameters. Furthermore, the computation of the statistic does not require an estimate of the long-run covariance operator $\Lambda_{\Delta X}$. In contrast, Chang et al.'s (2016) statistic requires computation of the long-run covariance operator $P_\ell^T \Lambda_{\Delta X} P_\ell^T$ for their choice of P_ℓ^T, and the functional KPSS statistics of Horváth et al. (2014) and Kokoszka and Young (2016) require estimation of the long-run covariance operator of the sequence $(X_t, t \geq 1)$.

Remark 10. Note that s in Theorem 4 and Remark 8 is the true dimension of the nonstationary subspace, which of course is unknown in general. However, this is not a problem in practice because we start with $s_0 = s_{\max}$, where $s_{\max} \geq s$, and test down; see Theorem 2. Then we can apply Theorem 4 and Remark 8 with $\ell \geq s_0$, which is feasible. See also Remarks 5 and 6 regarding the choice of $s_{\max}$. □

Remark 11. For the choice of ℓ, we suggest either a fixed value, for example $\ell = s_{\max} + k$ for some integer $k \geq 0$, or alternatively a value that depends on the null hypothesis, $\ell = s_0 + k$ for some integer $k \geq 0$. Note that both these choices are feasible in practical application since both $s_{\max}$ and s_0 are known to the practitioner. Clearly, the choice of ℓ may be an important issue for finite sample properties of our tests. We experimented with both $\ell = s_0$ and $\ell = s_0 + 2$ in unreported Monte Carlo simulations, and found that $\ell = s_0 + 2$ (which of course is feasible in practice) provides a good compromise between size and power. □

Remark 12. In Chang et al.'s (2016) test of (3.2), exactly s_0 orthonormal vectors that asymptotically span the nonstationary subspace, $\mathfrak{A}$, are required for consistency. An interesting feature of our Assumption 3, which is supported in practice by Theorem 4, is that our testing procedure does not require exactly s_0 orthonormal vectors that asymptotically span $\mathfrak{A}$. Instead, our testing procedure allows ℓ to be larger than s_0 in the estimate of the asymptotic superspace, ran P_ℓ^T, whose span asymptotically includes $\mathfrak{A}$. Intuitively, it seems clear that estimation of $\mathfrak{A}$ is much more difficult than estimation of any space that is asymptotically a superspace of $\mathfrak{A}$, and this may cause problems for Chang et al.'s (2016) test in finite

samples. This is, to some extent, confirmed in unreported simulations, where the test with $\ell = s_0 + 2$ outperforms that with $\ell = s_0$. □

Remark 13. Sometimes, estimation of $\mathfrak{A}$ is of independent interest. Theorem 4 shows that the first s eigenvectors of $\widetilde{\mathcal{K}}$ converge to an orthonormal basis of $\mathfrak{A}$. Therefore, estimation of $\mathfrak{A}$ reduces to estimation of s, which can be determined by our testing procedure. See also Theorem 3.3 of Chang et al. (2016), where it is shown that the first s eigenvectors of $\widetilde{\mathcal{C}}$ converge to orthonormal basis of $\mathfrak{A}$. □

4 Monte Carlo simulations

In this section, we investigate the finite sample performance of our test by Monte Carlo simulation. For all simulation experiments, the sample sizes are $T = 200$ and $T = 500$, the number of replications is 10,000, the nominal size is 5%, and critical values for the variance ratio tests are from Table 6 in Breitung (2002) (see Remarks 4 and 7). Note that s is the true value of the dimension of $\mathfrak{A}$ in the DGP and s_0 is the value under the null hypothesis. Thus, we simulate size when $s_0 = s$ and power when $s_0 \geq s + 1$. We report results for two statistics: CKP is the statistic of Chang et al. (2016) and $\mathcal{T}_{\mathcal{K}}$ is our preferred variance ratio statistic using $\ell = s_0 + 2$ eigenvalues of the sample covariance operator $\hat{\mathcal{K}}^{(1)}$ to construct P_ℓ^T; see (3.23) and Remark 9. The performance of the CKP statistic is sensitive to the choice of bandwidth parameter used in the estimation of the long-run covariance operator. We follow Chang et al. (2016) and use the Parzen kernel with the automatic data-dependent bandwidth rule of Andrews (1991). Finally, all statistics include correction for a non-zero intercept function, but no linear trend function.

4.1 Experiment 1: densities of individual earnings

This simulation experiment is based on the time series of cross-sectional densities of individual earnings that is analyzed in Chang et al. (2016). The observations of individual weekly earnings are obtained from the Current Population Survey (CPS) and deflated using inflation-adjustment factors suggested by CPS with base year 2005; see `https://cps.ipums.org/cps/cpi99.shtml`. Moreover, as in Chang et al. (2016), we drop top-coded earnings as well as zero earnings. As a result, our data set provides cross-sectional observations of individual earnings for 247 months from January 1994 to July 2014, and the number of cross-sectional observations for each month ranges from 12,180 in April 1996 to 15,826 in October 2001. For convenience, we divide each observation by 3500, which is strictly larger than the historically maximal observation, 3394.81, so that all observations are in $[0, 1]$. Clearly, this normalization does not cause any numerical differences compared with the results without normalization. As in Chang et al. (2016), we estimate monthly densities of individual earnings by kernel density estimation with the Epanechnikov kernel and bandwidth

given by $2.3449\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the standard deviation of cross-sectional observations and n is the cross-sectional sample size.

The basic data-generating process (DGP) for the simulation experiment is constructed in the same way as in Chang et al. (2016) with the only difference that we use 247 B-spline basis functions for the representation of $L^2[0,1]$-functions, and obtain the eigenvectors $(\hat{v}_1, \ldots, \hat{v}_{247})$ of the covariance operator $\hat{\mathcal{C}}$. We thus let

$$X_t - \bar{X}_T = \sum_{j=1}^{247} a_{j,t}\hat{v}_j \tag{4.1}$$

and

$$\begin{aligned} \Delta a_{j,t} &= \beta_j \Delta a_{j,t-1} + \sigma_j \eta_{j,t}, \qquad j = 1, 2(= s), \\ a_{j,t} &= \beta_j a_{j,t-1} + \sigma_j \eta_{j,t}, \qquad j \geq 3, \end{aligned} \tag{4.2}$$

where $\eta_{j,t}$ are i.i.d. N(0,1) across j and t, and β_j and σ_j are replaced by the estimates from the observations $\hat{a}_{j,t} = \langle X_t - \bar{X}_T, \hat{v}_j \rangle$ for $t = 1, \ldots, T$. Note that this DGP is a special case of the functional AR(1) processes in Section 4.3.

In the basic DGP in (4.1)–(4.2) we have $\sum_{j=1}^{2} \sigma_j^2 / \sum_{j=1}^{247} \sigma_j^2 \simeq 0.65$. This implies that 65% of the random functional variation at time $t+1$, given all the information up to time t, occurs in the nonstationary subspace. In our empirical examples to age-specific employment rates and to Ontario electricity demand in Sections 5.1 and 5.3, this number is 14% and 1.6%, respectively. Thus, the value 65% in the DGP (4.1)–(4.2) may seem like a very high value. We therefore replace σ_j in (4.2) with $\tilde{\sigma}_j$ defined as

$$\begin{aligned} \tilde{\sigma}_j^2 &= (1/q_1)\sigma_j^2, \qquad j = 1, 2(= s), \\ \tilde{\sigma}_j^2 &= q_1 \sigma_j^2, \qquad j \geq 3, \end{aligned} \tag{4.3}$$

for $q_1 \in \{1.0, 1.5, \ldots, 3.5\}$. Now $\sum_{j=1}^{2} \tilde{\sigma}_j^2 / \sum_{j=1}^{247} \tilde{\sigma}_j^2$ varies from approximately 65% to 13% as q_1 varies from 1.0 to 3.5.

In Table 1 we report the results from the DGP in (4.1)–(4.3). In the rows with $s_0 = s = 2$ we report the simulated size of the tests. It is clear that the CKP test is very sensitive to the value of q_1 with severe over-sizing when $q_1 \geq 2$ for $T = 200$ and when $q_1 \geq 2.5$ when $T = 500$. On the other hand, the $\mathcal{T}_{\mathcal{K}}$ variance ratio test is very robust to q_1, though it is somewhat under-sized when $T = 200$ and slightly under-sized when $T = 500$.

Recalling that the true value is $s = 2$, the rows with $s_0 \geq 3$ report simulated power. To make the comparison meaningful, these are size-corrected. To this end, we need to modify the DGP in (4.2) such that data can be generated with $s \geq 3$ nonstationary components. To do this, we randomly choose $s-2$ values from $j = 3, \ldots, 8$ and set their values of β_j equal to one. From the results in Table 1 we see that the large size distortions of the CKP test imply that its power is nearly zero in many cases. It is only with small values of q_1 that the CKP test

Table 1: Simulation results for DGP (4.1)–(4.3)

		$T = 200$						$T = 500$					
Test	$s_0 \backslash q_1$	1.0	1.5	2.0	2.5	3.0	3.5	1.0	1.5	2.0	2.5	3.0	3.5
$\mathcal{T}_{\mathcal{K}}$	2	0.011	0.010	0.009	0.009	0.011	0.010	0.037	0.036	0.036	0.036	0.036	0.036
	3	0.463	0.457	0.473	0.456	0.468	0.431	0.882	0.887	0.884	0.888	0.884	0.886
	4	0.870	0.869	0.879	0.874	0.871	0.845	0.999	0.999	0.999	0.999	0.999	0.999
	5	0.986	0.987	0.989	0.989	0.987	0.985	1.000	1.000	1.000	1.000	1.000	1.000
CKP	2	0.024	0.044	0.182	0.393	0.590	0.756	0.040	0.041	0.052	0.111	0.230	0.383
	3	0.992	0.658	0.158	0.081	0.071	0.073	1.000	1.000	1.000	0.333	0.083	0.048
	4	0.991	0.388	0.106	0.065	0.052	0.053	1.000	1.000	0.627	0.126	0.040	0.020
	5	0.947	0.251	0.070	0.040	0.020	0.017	1.000	0.993	0.260	0.032	0.008	0.004

Notes: Based on 10,000 Monte Carlo replications. The DGP true value is $s = 2$ and the H_0 value is s_0. The nominal size is 5%.

has meaningful size and power. Furthermore, in some cases where the CKP test has accurate size it suffers from a power-reversal problem, in the sense that power declines as s_0 increases and is further away from the true s (e.g., when $T = 500$ and $q_1 = 2.0$). This phenomenon was also observed by Chang et al. (2016) in their Table 7, where the true number of stochastic trends is one, and columns 1 and 2 are reasonable, but columns 3–5 correspond to alternatives that are farther away from the null hypothesis with declining power. On the other hand, the power of the $\mathcal{T}_{\mathcal{K}}$ variance ratio test is unaffected by the value of q_1, as was the size.

4.2 Experiment 2: densities of individual earnings with measurement error

We next consider a modification of the basic DGP in (4.1)–(4.2), where we include estimation/measurement error in the DGP. First, the time series in Chang et al.'s (2016) DGP in (4.1)–(4.2) consist of estimated densities, so we should not disregard estimation error. Second, the first step of functional time series analysis typically includes smoothing of discrete observations to obtain functional observations. This data pre-processing therefore entails estimation error for each X_t, and we make the DGP a little more realistic by adding an i.i.d. noise to each realization. That is, in addition to (4.1)–(4.2), we consider

$$\widetilde{X}_t = X_t + q_2 u_t, \quad u_t = P_{F10} B_t, \tag{4.4}$$

where B_t is a sequence of i.i.d. standard Brownian bridges and P_{F10} denote the projection operator onto the span of the first 10 Fourier basis functions (without a constant function). The projection P_{F10} is not essential to this experiment and is applied to make $\widetilde{X}_t$ a smooth function for each t. The inverse signal-to-noise ratio is $q_2 \in \{0.00, 0.05, \ldots, 0.25\}$.

The results for the DGP in (4.1), (4.2), and (4.4) are reported in Table 2. Again, the CKP test has very poor size control for this DGP. Even for the larger sample with $T = 500$,

Table 2: Simulation results for DGP (4.1), (4.2), (4.4)

		$T = 200$						$T = 500$					
Test	$s_0 \backslash q_2$	0.00	0.05	0.10	0.15	0.20	0.25	0.00	0.05	0.10	0.15	0.20	0.25
$\mathcal{T}_{\mathcal{K}}$	2	0.011	0.017	0.031	0.061	0.094	0.137	0.037	0.041	0.050	0.064	0.083	0.107
	3	0.456	0.499	0.525	0.564	0.558	0.556	0.875	0.905	0.931	0.953	0.960	0.966
	4	0.871	0.894	0.920	0.921	0.917	0.902	0.999	1.000	1.000	1.000	1.000	1.000
	5	0.984	0.992	0.995	0.996	0.992	0.987	1.000	1.000	1.000	1.000	1.000	1.000
CKP	2	0.024	0.035	0.082	0.232	0.466	0.677	0.040	0.054	0.108	0.230	0.413	0.601
	3	0.993	0.992	0.860	0.306	0.155	0.097	1.000	1.000	1.000	1.000	0.894	0.448
	4	0.999	0.995	0.797	0.397	0.268	0.192	1.000	1.000	1.000	1.000	0.914	0.649
	5	1.000	0.992	0.821	0.510	0.378	0.280	1.000	1.000	1.000	1.000	0.963	0.796

Notes: Based on 10,000 Monte Carlo replications. The DGP true value is $s = 2$ and the H_0 value is s_0. The nominal size is 5%.

the size of the CKP test is 10.8% with $q_2 = 0.10$ and 60.1% with $q_2 = 0.25$. In contrast, the $\mathcal{T}_{\mathcal{K}}$ test has only slight size distortions for the largest values of q_2 considered.

We again consider size-corrected power due to the large size distortions for the CKP test in particular. In cases where the size of the CKP test is reasonable, it appears to have good power, and for the smallest values of q_2 the CKP test has higher power than the $\mathcal{T}_{\mathcal{K}}$ test. On balance, though, the $\mathcal{T}_{\mathcal{K}}$ test has higher power than the CKP test in most cases.

Overall, the CKP test is clearly very sensitive to the specifications of q_1 and q_2 in our Experiments 1 and 2, in terms of both simulated size and power. In constrast, the $\mathcal{T}_{\mathcal{K}}$ variance ratio test is very robust to all specifications.

4.3 Experiment 3: functional AR(1) process

For our next simulation experiment we consider the commonly applied functional AR(1) model. In particular, our setup follows that of Beare et al. (2017) and Aue et al. (2017). Let $(\zeta_j, j = 1, \ldots, 21)$ be the first 21 orthonormal polynomial basis functions defined on $[0, 1]$, and let $(\zeta_{(j)}, j = 1, \ldots, 21)$ be the same collection in a different order obtained by randomly permuting $(\zeta_j, j = 1, \ldots, 8)$ and $(\zeta_j, j = 9, \ldots, 21)$, separately. We generate the functional time series as

$$X_t = \mu + \sum_{j=1}^{21} \theta_j \langle \zeta_{(j)}, X_{t-1} \rangle \zeta_{(j)} + B_t, \tag{4.5}$$

where $(B_t, t = 1, \ldots, T)$ is a sequence of i.i.d. standard Brownian bridges and

$$\theta_j = \begin{cases} 1 & \text{for } j \leq s, \\ \theta^{(j-s)} & \text{for } j \geq s + 1, \end{cases}$$

Table 3: Simulation results for functional AR(1)

			$T = 200$				$T = 500$			
θ	Test	s_0	$s = 0$	$s = 1$	$s = 2$	$s = 3$	$s = 0$	$s = 1$	$s = 2$	$s = 3$
0.0	$\mathcal{T}_{\mathcal{K}}$	s		0.049	0.045	0.058		0.050	0.050	0.047
		$s+1$	0.999	0.966	0.954	0.955	1.000	1.000	1.000	0.999
		$s+2$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$s+3$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	CKP	s		0.372	0.644	0.567		0.173	0.424	0.577
		$s+1$	0.187	0.134	0.036	0.020	0.496	0.209	0.161	0.038
		$s+2$	0.124	0.001	0.001	0.002	0.372	0.091	0.000	0.001
		$s+3$	0.000	0.000	0.000	0.000	0.199	0.000	0.000	0.000
0.5	$\mathcal{T}_{\mathcal{K}}$	s		0.047	0.043	0.041		0.050	0.048	0.052
		$s+1$	0.981	0.907	0.897	0.859	1.000	0.996	0.994	0.996
		$s+2$	1.000	0.999	0.999	0.997	1.000	1.000	1.000	1.000
		$s+3$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	CKP	s		0.390	0.688	0.799		0.195	0.455	0.635
		$s+1$	0.149	0.067	0.044	0.032	0.342	0.117	0.060	0.032
		$s+2$	0.000	0.009	0.010	0.012	0.000	0.002	0.003	0.003
		$s+3$	0.000	0.001	0.002	0.004	0.000	0.000	0.000	0.000
0.8	$\mathcal{T}_{\mathcal{K}}$	s		0.027	0.023	0.019		0.045	0.042	0.041
		$s+1$	0.785	0.618	0.537	0.437	0.976	0.915	0.901	0.860
		$s+2$	0.982	0.959	0.940	0.917	1.000	1.000	1.000	0.999
		$s+3$	1.000	0.999	0.999	0.997	1.000	1.000	1.000	1.000
	CKP	s		0.459	0.744	0.823		0.271	0.569	0.729
		$s+1$	0.143	0.064	0.070	0.072	0.315	0.104	0.069	0.077
		$s+2$	0.019	0.056	0.080	0.113	0.005	0.020	0.046	0.091
		$s+3$	0.012	0.042	0.092	0.317	0.000	0.007	0.031	0.231

Notes: Based on 10,000 Monte Carlo replications. The DGP true value is s and the H_0 value is s_0. Nominal size is 5%.

for $0 \leq \theta < 1$. Following, e.g., Aue et al. (2017), we permute (ζ_j) as described above to avoid any effects caused by the particular shape and ordering of the basis functions, and hence the shapes of the stationary and nonstationary subspaces. Intuitively, when the nonstationary subspace is s-dimensional, s elements are randomly drawn from the first eight polynomials. Similarly, to avoid any effects caused by the particular shape of the mean function μ in (4.5), it is generated by $\sum_{j=1}^{21} g_j \zeta_j$, where g_j are i.i.d. standard normal random variables. Finally, the functional observations are constructed by smoothing $(X_t, t = 1, \ldots, T)$ in (4.5) using 41 Fourier basis functions (the choice of basis functions has minimal effect in this setting).

Table 3 presents simulation results for $\theta \in \{0.0, 0.5, 0.8\}$. We first note that the CKP

test has very poor size control for all θ and all $s = \dim(\mathfrak{A})$. Even though it improves as the sample size increases, there is still severe over-rejection for $T = 500$. The size-corrected simulated power of the CKP test is very low in all cases, presumably due to the large size distortions. On the other hand, the $\mathcal{T}_{\mathcal{K}}$ variance ratio test has excellent size and power for all θ and s considered in this simulation setup.

In unreported simulations, we considered model (4.5) with the nonstationary subspace spanned by the first s polynomial basis functions, i.e. without permutation of $(\zeta_j, j = 1, \ldots, 8)$. Because lower-order polynomial basis functions are much more smooth than higher-order polynomials, all tests have better finite-sample properties in this case. However, the CKP test still substantially over-rejects the null hypothesis when either s or θ is relatively large, while the $\mathcal{T}_{\mathcal{K}}$ test performs very well overall.

4.4 Experiment 4: sequential application of tests

In practice, rather than testing a specific hypothesis of interest, we expect that the most common application of our proposed test, and of the CKP test, is to estimate the dimension of the nonstationary subspace as in Theorem 2. To this end, both procedures require a priori setting an upper bound, denoted $s_{\max}$. The hypotheses $H_0 : \dim(\mathfrak{A}) = s_0$ for $s_0 = s_{\max}, s_{\max} - 1, \ldots, 1$ are then tested sequentially, and the estimate $\hat{s}$ is the first non-rejected null hypothesis. It is preferable that this $\hat{s}$ is robust to the choice of $s_{\max}$, i.e. that it does not depend on the choice of $s_{\max}$ (as long as $s_{\max} \geq s$). In our last simulation experiment we consider this sequential application of the tests to obtain $\hat{s}$. The data is generated by the functional AR(1) process (4.5) with $\theta = 0.5$.

Table 4 reports the results as relative frequencies of $\hat{s}$ divided into the four "bins", $\hat{s} < s$, $\hat{s} = s$, $\hat{s} = s+1$, and $\hat{s} > s+1$. For each of four true values, $s \in \{0, 1, 5, 8\}$, we consider three different choices for the initial hypothesis, $s_{\max} \in \{s + 1, s + 3, 20\}$. The first two choices are a simple way to simulate careful selection of $s_{\max}$ based on graphical or other measures as discussed in Chang et al. (2016, Section 5), with the second choice being slightly more liberal in view of the requirement that $s_{\max} \geq s$ (and s is unknown in practice).

The results in Table 4 are clearly favorable to the $\mathcal{T}_K$ test. In particular, the CKP test tends to find $\hat{s} > s$ when either $s_{\max}$ or s increases. This issue with the CKP test is likely due to the power reversal problem of the test as mentioned in Section 4.1 and also observed by Chang et al. (2016) in their Table 7. On the other hand, our $\mathcal{T}_K$ test finds $\hat{s} = s$ with large probability that is nearly unaffected by the choice of $s_{\max}$. Although this probability decreases somewhat when s increases, it also increases substantially with the sample size.

Unreported Monte Carlo simulations have shown that the variance ratio tests based on $\mathcal{C}$ and/or with $\ell = s_0$ are somewhat over-sized, though not nearly as much as the CKP

Table 4: Relative frequencies of $\hat{s}$ for functional AR(1)

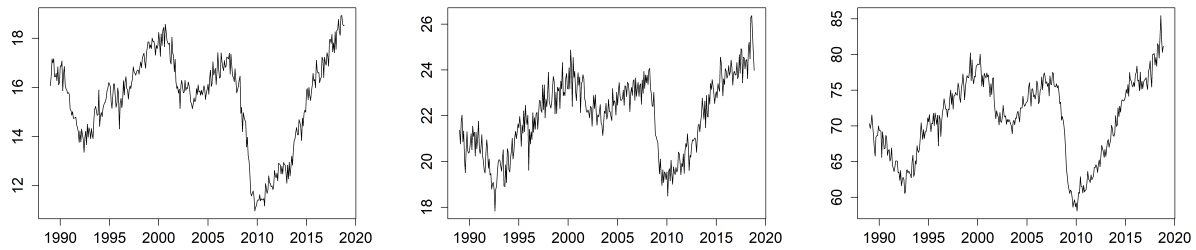
			$T = 200$				$T = 500$			
$s_{\max}$	Test	s	$\hat{s} < s$	$\hat{s} = s$	$\hat{s} = s{+}1$	$\hat{s} > s{+}1$	$\hat{s} < s$	$\hat{s} = s$	$\hat{s} = s{+}1$	$\hat{s} > s{+}1$
$s{+}1$	$\mathcal{T}_{\mathcal{K}}$	0		0.979	0.021			1.000	0.000	
		1	0.042	0.853	0.105		0.051	0.944	0.005	
		5	0.088	0.734	0.178		0.077	0.919	0.004	
		8	0.334	0.624	0.042		0.259	0.741	0.000	
	CKP	0		1.000	0.000	0.000		1.000	0.000	0.000
		1	0.396	0.603	0.001	0.000	0.199	0.801	0.000	0.000
		5	0.114	0.054	0.831	0.000	0.762	0.238	0.000	0.000
		8	0.000	0.000	1.000	0.000	0.063	0.027	0.909	0.000
$s{+}3$	$\mathcal{T}_{\mathcal{K}}$	0		0.979	0.021	0.000		1.000	0.000	0.000
		1	0.041	0.853	0.105	0.001	0.051	0.944	0.005	0.000
		5	0.088	0.734	0.176	0.002	0.077	0.919	0.004	0.000
		8	0.334	0.624	0.041	0.000	0.259	0.741	0.000	0.000
	CKP	0		1.000	0.000	0.000		1.000	0.000	0.000
		1	0.354	0.528	0.000	0.117	0.199	0.801	0.000	0.000
		5	0.000	0.000	0.000	1.000	0.757	0.235	0.000	0.008
		8	0.000	0.000	0.000	1.000	0.001	0.000	0.000	0.999
20	$\mathcal{T}_{\mathcal{K}}$	0		0.979	0.021	0.000		1.000	0.000	0.000
		1	0.041	0.853	0.105	0.001	0.051	0.944	0.005	0.000
		5	0.088	0.734	0.176	0.002	0.077	0.919	0.004	0.000
		8	0.334	0.624	0.041	0.000	0.259	0.741	0.000	0.000
	CKP	0		0.000	0.000	1.000		0.000	0.000	1.000
		1	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
		5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
		8	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Notes: Based on 10,000 Monte Carlo replications. The DGP true value is s and $\theta = 0.5$. Nominal size is 5%.

test. This suggests that estimation of $\mathfrak{A}$ is more difficult than estimation of an asymptotic superspace of $\mathfrak{A}$, and we conjecture that this may be the main reason for the size distortion of the CKP test; see also Remarks 11 and 12.

Overall, our Monte Carlo simulations strongly support the use of the $\mathcal{T}_{\mathcal{K}}$ variance ratio test. It is very robust to the DGP specifications with excellent size control throughout. Further evidence on the robustness based on empirical applications is presented in Section 5.4.

Figure 1: Group characteristics



(a) younger age group (b) older age group (c) all ages

5 Empirical applications

5.1 Logit transformed age-specific employment rates

We first apply our methodology to the time series of age-specific employment rates in the US observed monthly from January 1989 to November 2018. The data is available from the CPS at `https://ipums.org`; see Flood et al. (2018). We only consider individuals in the working age (15–64) population. For age a, the age-specific employment rate at time t is computed as

$$X_{a,t} = \frac{\sum_{i=1}^{n_t} w_{i,t} Z_{i,t} 1\{a_{i,t} = a\}}{\sum_{i=1}^{n_t} w_{i,t} 1\{a_{i,t} = a\}},$$

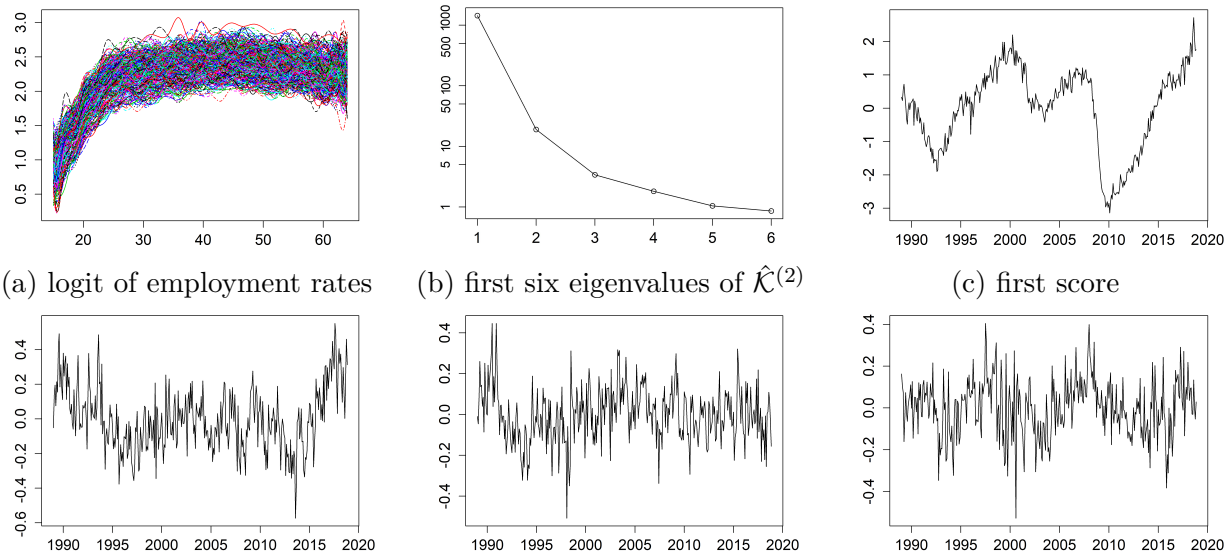
where $1\{\cdot\}$ denotes the indicator function, n_t is the number of individuals observed at time t, and $w_{i,t}$, $a_{i,t}$, and $Z_{i,t}$ denote the weight (WTFINL in CPS), age, and employment status dummy of individual i at time t, respectively. The employment rate specific to each age, $X_{a,t}$, is then seasonally adjusted using the software package provided by the US Census Bureau. The age-specific employment rate takes values between 0 to 1 by construction, so as is common in the literature, we hereafter consider the logit transformation, $\psi(X_{a,t})$, instead of $X_{a,t}$. Finally, the functional observations $X_t(u)$ for $u \in [15, 64]$ and $t = 1, \ldots, T = 359$ are obtained by smoothing $\psi(X_{a,t})$ over a using 31 B-spline basis functions.

In Figure 1 we plot three real-valued sequences $(\langle X_t, v\rangle, t \geq 1)$ to explore characteristics of the functional time series. Specifically, we consider $v = v_y$, v_o, and v_a, where

$$v_y(u) = 1\{u \leq 25\}, \quad v_o(u) = 1\{u \geq 54\}, \quad v_a(u) = 1, \quad u \in [15, 64].$$

Clearly, $\langle X_t, v_y\rangle$ and $\langle X_t, v_o\rangle$) are the average employment rates for the younger and older age groups, respectively, and $\langle X_t, v_a\rangle$ is the overall average employment rate. Firstly, Figure 1 suggests that the functional time series of age-specific employment rates is nonstationary, because if it were stationary, then $(\langle X_t, v\rangle, t \geq 1)$ would be stationary for any choice of $v \in \mathcal{H}$. Secondly, it seems that the series may have a linear time trend. Thirdly, the three series clearly have some degree of co-movement, but they also have their own characteristics. For example, in 2009 employment rates decline sharply in both age groups, but the decline seems

Figure 2: Monthly age-specific employment rates January 1989 to November 2018



(a) logit of employment rates (b) first six eigenvalues of $\hat{\mathcal{K}}^{(2)}$ (c) first score

(d) second score (e) third score (f) fourth score

more severe in the younger age group than in the older age group. Note that, if we only focus on the employment rate that is aggregated over ages as in Figure 1(c), this information is lost.

Panel (a) of Figure 2 displays the functional observations. In Figure 2(b) we display the first six (largest) eigenvalues of $\hat{\mathcal{K}}^{(2)}$ from (3.22) on a logarithmic scale. Let the inner product of the jth eigenvector of (3.22) with $(U_t^{(\varrho)}, 1 \leq t \leq T)$ denote jth score process. Figures 2(c)–(f) display the first four score processes. Because Figure 1 suggested the possible presence of a linear time trend, this has been applied in the specification of the deterministic component for the plots. Note that the horizontal axis in Figure 2(a) is age and not time, so a "trend" in this plot is related to the shape of the intercept function and not the time trend.

The order of magnitude and rate of decay of the eigenvalues can be suggestive of the dimension of the nonstationary subspace, as discussed in Chang et al. (2016) for the eigenvalues of $\hat{\mathcal{C}}$. In particular, the number of "large" eigenvalues should correspond to the dimension of the nonstationary subspace. Furthermore, based on Theorem 4 we expect that the first s score processes behave as unit root processes. From the plots of both the eigenvalues and the score processes, there seems to be quite strong graphical evidence in favor of the dimension being at least one or two.

Table 5 summarizes the test results under two different specifications of the deterministic component, nonzero intercept and linear trend. Even though Chang et al. (2016) do not explicitly consider the case where the DGP includes a linear trend function, we may apply their test to the functional residuals $U_t^{(2)}$ in which case the asymptotic distribution of their test statistic under the null is given by the minimum eigenvalue of $\int_0^1 W_{s_0}^{(2)}(r) W_{s_0}^{(2)\prime}(r) dr$.

Table 5: Test results for logit of age-specific employment rates

Test	$s_0 = 1$	$s_0 = 2$	$s_0 = 3$	$s_0 = 4$	$s_0 = 5$
Intercept only					
$\mathcal{T}_{\mathcal{K}}$	17.67	156.38	579.59	1623.55**	4411.19***
CKP	0.1134	0.0325	0.0077***	0.0075***	0.0068***
Linear trend & intercept					
$\mathcal{T}_{\mathcal{K}}$	130.43	426.18	1401.94**	4381.61***	7107.46***
CKP	0.1110	0.0093***	0.0074***	0.0073***	0.0066***

Notes: The functional data are smoothed with 31 B-spline functions, and the number of observations is $T = 359$. We use *, **, and *** to denote rejection at 10%, 5%, and 1% significance level, respectively. Critical values for the CKP test in the linear trend case are calculated from the functional residuals $U_t^{(2)}$ and 100,000 approximate realizations from the asymptotic distribution. Data and R code to replicate this table are available on the authors' websites.

Table 6: Johansen trace test results for logit of age-specific employment rates

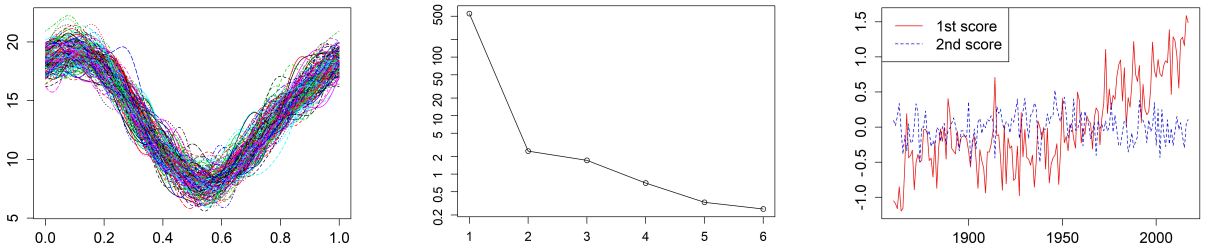
Lag length, k	1	2	3	4	5
$\hat{s}$	1	2	23	20	13

Notes: The data are the 50-dimensional vector-valued series $\psi(X_{a,t}), a = 1, \ldots, 50$. The estimated number of stochastic trends is based on Johansen's trace test with lag length k, a restricted trend, and significance level 5%.

For both specifications of deterministic terms in Table 5, the CKP test rejects more than the $\mathcal{T}_{\mathcal{K}}$ variance ratio test. In particular, with only an intercept, the CKP test suggests that the dimension of the nonstationary subspace is $\hat{s} = 2$ and the $\mathcal{T}_{\mathcal{K}}$ test suggests $\hat{s} = 3$. Allowing for a linear trend, which based on Figure 1 seems prudent, the CKP and $\mathcal{T}_{\mathcal{K}}$ tests suggest $\hat{s} = 1$ and $\hat{s} = 2$, respectively. This pattern is, at least to some extent, expected from the simulation evidence, where the CKP test was often found to be over-sized. For that reason, we would be most inclined to conclude that the logit transformed age-specific employment curves have a two-dimensional nonstationary subspace (i.e., are driven by two stochastic trends).

Before moving on, we entertain the following idea. We consider the series $(\psi(X_{a,t}), t = 1, \ldots, 359)$ for $a = 1, \ldots, 50$ as a 50-dimensional vector-valued time series and apply standard methods from finite-dimensional Euclidean space (note that in the next empirical examples the dimension is too high for this idea to be feasible). In particular, we apply Johansen's (1995) trace test for cointegration rank (using the obvious correspondence between cointegration rank and number of stochastic trends in the finite-dimensional case) in a vector error-correction model with k lags and an unrestricted constant and a restricted trend. The results are presented in Table 6 for $k = 1, \ldots, 5$. While more lags may be preferable with monthly data, e.g. $k = 12$, the computer programs fail for $k > 5$; see Yao et al. (2012) for theoretical re-

Figure 3: Annual minimum temperature curves at Sydney station 1860–2018



(a) minimum temp. curves (b) first six eigenvalues of $\hat{\mathcal{K}}^{(1)}$ (c) first and second scores

sults on failure of eigenvalue calculations with large covariance matrices. It is well known that Johansen's (1995) test has low power in high-dimensional systems (Ho and Sørensen, 1996; Onatski and Wang, 2018), and of course it can be sensitive to the parametric specification of the lag-order. This is also what we find in Table 6, at least for $k \geq 3$, whereas for $k = 1, 2$ the results could be a consequence of size-distortion due to under-specification of the lag-order.

5.2 Minimum temperatures in Australia

The next empirical example is an application to yearly minimum temperature curves in Australia. This example is also considered in Aue et al. (2017), who reject the null of stationarity against the alternative of structural change in the mean function. However, their finding could also be a consequence of a nontrivial nonstationary subspace.

The raw data is obtained from the Australian Bureau of Meteorology at `http://www.bom.gov.au` and consists of daily minimum temperature observations. For each year, the observations are smoothed using 23 Fourier basis functions to obtain a curve of minimum temperatures through the year. We consider six weather stations that have relatively large samples, and we allow for a non-zero intercept function in the processes.

Figure 3 shows a graphical summary of the Sydney data set. In particular, Figure 3(b) shows the six largest eigenvalues of $\hat{\mathcal{K}}^{(1)}$ on a logarithmic scale. The first eigenvalue is clearly very different from the remaining eigenvalues, suggesting a one-dimensional nonstationary subspace (one stochastic trend). This is also suggested from the plots of the first and second scores in Figure 3(c), of which only the first seems nonstationary.

Table 7 reports the test results for the six temperature series. The findings for the CKP test are very mixed. In all of the data sets, $s_0 = 5$ is not rejected, although smaller values are rejected, and smaller yet are not. The CKP findings are thus strongly dependent on the starting point, $s_{\max}$, of the procedure. On the other hand, our $\mathcal{T}_{\mathcal{K}}$ test detects one stochastic trend for all temperature curves with at least 5% significance ($s_0 = 1$ is in fact rejected for the Gunnedah Pool series, but only at the 10% level).

Clearly, this strong dependence of the CKP procedure on the starting value, $s_{\max}$, is

Table 7: Test results for Australian minimum temperatures

Test	$s_0 = 1$	$s_0 = 2$	$s_0 = 3$	$s_0 = 4$	$s_0 = 5$
Sydney, 1860–2018					
$\mathcal{T}_{\mathcal{K}}$	16.27	440.81**	1508.09***	3348.58***	6728.21***
CKP	0.0523	0.0199**	0.0184	0.0149	0.0148
Melbourne, 1856–2014					
$\mathcal{T}_{\mathcal{K}}$	14.65	397.21**	1331.09***	4421.53***	7933.17***
CKP	0.0846	0.0194**	0.0174*	0.0138	0.0137
Gunnedah Pool, 1877–2011					
$\mathcal{T}_{\mathcal{K}}$	68.92*	393.15**	1369.26***	2809.60***	4820.76***
CKP	0.0202***	0.0166**	0.0141**	0.0142	0.0134
Cape Otway, 1864–2018					
$\mathcal{T}_{\mathcal{K}}$	55.64	980.81***	2197.37***	4611.09***	8595.01***
CKP	0.0187***	0.0140***	0.0143**	0.0138	0.0131
Boulia Airport, 1888–2018					
$\mathcal{T}_{\mathcal{K}}$	31.56	226.93	1376.03***	3510.24***	6449.36***
CKP	0.0176***	0.0160***	0.0129**	0.0130*	0.0140
Gaydah Post Office, 1894–2008					
$\mathcal{T}_{\mathcal{K}}$	16.72	429.14**	1109.67**	2574.92***	4469.61***
CKP	0.0517	0.0208**	0.0184	0.0182	0.0154

Notes: The functional data are smoothed with 23 Fourier basis functions. In the order of stations reported in the table, the numbers of observations are 160, 161, 133, 155, 126, and 117, respectively. We use *, **, and *** to denote rejection at 10%, 5%, and 1% significance level, respectively.

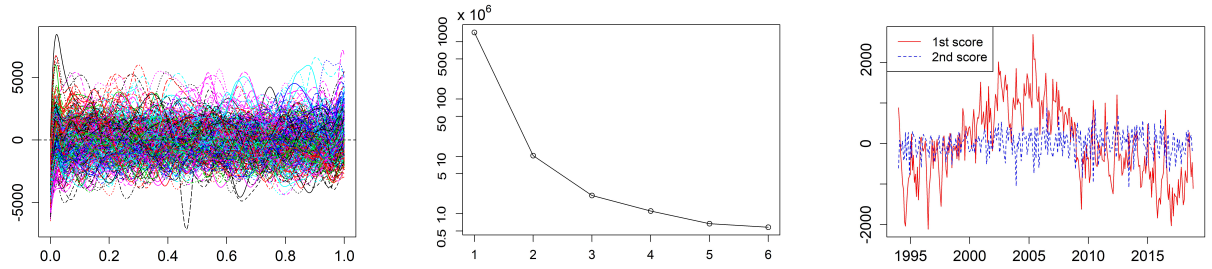
undesirable in practice. At the same time, our $\mathcal{T}_{\mathcal{K}}$ variance ratio test seems to be much less subject to this problem, and this should be a very appealing feature of our test for applied researchers. The robustness (or lack thereof) of the CKP procedure and the $\mathcal{T}_{\mathcal{K}}$ variance ratio test to the starting value, $s_{\max}$, is further explored in Section 5.4.

5.3 Ontario monthly electricity demand

In our final empirical example, we examine the existence of nonstationarity in Ontario electricity demand. The raw data is observed every hour from January 1994 to November 2018, and is available at http://www.ieso.ca. We obtain $T = 299$ monthly electricity demand curves using around 700 hourly data points for each month smoothed with 31 B-spline basis functions. The monthly curves are seasonally adjusted by functional regression on a set of 12 seasonal dummies (of course, this implies that the series have zero mean, but in the asymptotic theory it corresponds to inclusion of an intercept function).

Figure 4 shows a graphical summary of the time series of monthly electricity demand

Figure 4: Monthly electricity demand in Ontario January 1994 to November 2018



(a) electricity demand curves (b) first six eigenvalues of $\hat{\mathcal{K}}$ (c) first and second scores

Table 8: Test results for Ontario monthly electricity demand

Test	$s_0 = 1$	$s_0 = 2$	$s_0 = 3$	$s_0 = 4$	$s_0 = 5$
$\mathcal{T}_{\mathcal{K}}$	49.26	446.95**	2533.74***	5163.29***	8154.96***
CKP	0.0188***	0.0187**	0.0190	0.0102**	0.0075***

Notes: The seasonally adjusted functional data are smoothed with 31 B-spline basis functions. The number of observations is $T = 299$. We use *, **, and *** to denote rejection at 10%, 5%, and 1% significance level, resp.

curves. As in the previous example, the first eigenvalue is orders of magnitude larger than the remaining eigenvalues, and the first score process looks nonstationary while the second looks stationary. Again, this is suggestive of a one-dimensional nonstationary subspace.

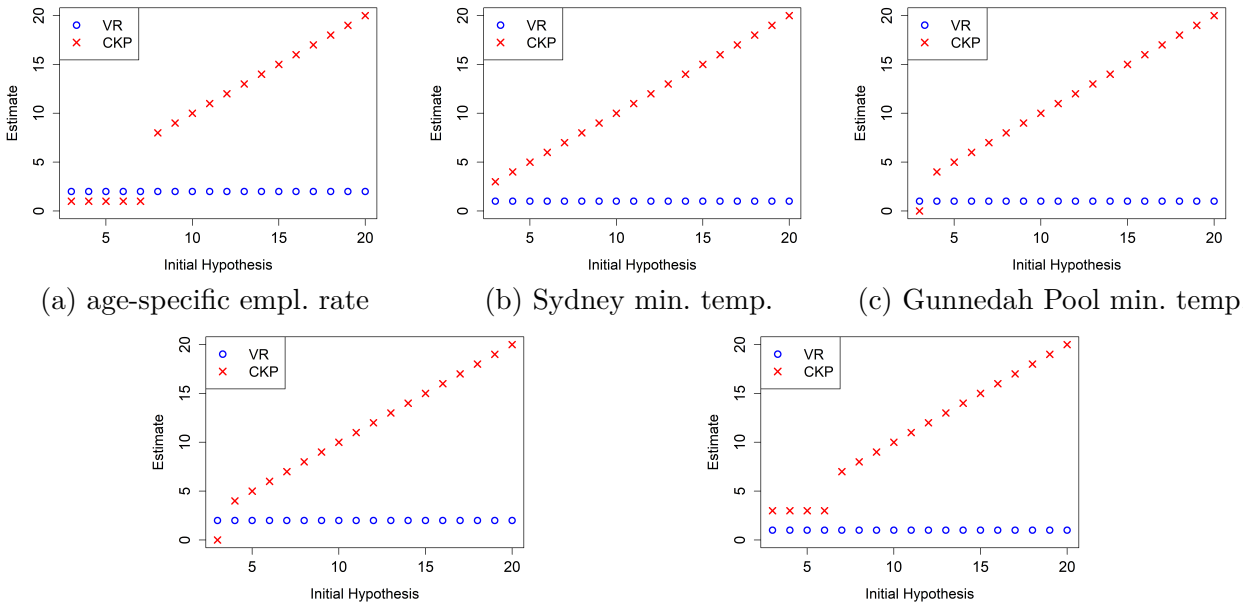
Table 8 presents test results for the electricity demand data set. The $\mathcal{T}_{\mathcal{K}}$ test concludes that the nonstationary subspace is one-dimensional at the 5% significance level. The CKP test, on the other hand, concludes that the process is either stationary or has a three-dimensional nonstationary subspace at the 1% significance level ($s_0 = 3$ is the only non-rejected hypothesis). In view of the Monte Carlo simulation results, this is not too surprising.

5.4 Robustness to choice of $s_{\max}$

In practice, rather than testing a specific hypothesis of interest, we expect that the most common application of our proposed test, and of the CKP test, is to estimate the dimension of the nonstationary subspace as in Theorem 2. To this end, both procedures require a priori setting an upper bound, denoted $s_{\max}$. The hypotheses $H_0 : \dim(\mathfrak{A}) = s_0$ for $s_0 = s_{\max}, s_{\max} - 1, \ldots, 1$ are then tested sequentially. It is preferable that the result of this procedure is robust to the choice of $s_{\max}$, i.e. that it does not depend on the choice of $s_{\max}$ (as long as $s_{\max} \geq s$).

We know from the simulation evidence in Section 4.4 and Table 4, as well as the empirical examples discussed above, that the CKP results depend heavily on the choice of $s_{\max}$, while the $\mathcal{T}_{\mathcal{K}}$ variance ratio test is more robust to this choice. We now investigate this issue further in Figure 5, where we report the estimated dimension for each data set using the sequential testing procedure in Theorem 2 with 5% significance level for $s_{\max} = 3, \ldots, 20$. It is obvious

Figure 5: Estimates from the $\mathcal{T}_{\mathcal{K}}$ and CKP procedures for the data sets



(a) age-specific empl. rate (b) Sydney min. temp. (c) Gunnedah Pool min. temp.

(d) Boulia Airport min. temp. (e) Ontario electricity demand

Notes: For age-specific employment rates a linear trend is included, for the minimum temperature series an intercept but no trend is included, and for the electricity demand seasonal dummies are included. In all cases the significance level of the tests is 5%. The results for the minimum temperatures in Gaydah Post Office, Melbourne, and Cape Otway are identical to those in Panels (b), (b), and (d), respectively.

from Figure 5 that the CKP procedure is very sensitive to the initial hypothesis, $s_{\max}$, while the variance ratio test is very robust.

In particular, the CKP test tends not to reject the initial hypothesis, $H_0 : \dim(\mathfrak{A}) = s_{\max}$, when $s_{\max}$ gets bigger. Consequently, sequential application of the CKP test would either lead to the conclusion that the dimension of the nonstationary subspace is $s_{\max}$, or $s_{\max}$ would be increased and the sequential test repeated (as discussed in Remark 6) thus exacerbating the problem. In some cases, this dependence of the CKP test on the choice of $s_{\max}$ is a relatively minor issue, and can be avoided by careful choice of $s_{\max}$. For example, for the age-specific employment and electricity demand applications, the CKP procedure gives the same estimate for $s_{\max} \leq 7$ and $s_{\max} \leq 6$, respectively. However, in all the other applications, the CKP procedure finds that $\hat{s} = s_{\max}$ for all choices of $s_{\max} \geq 4$ (or even less). This is problematic, not only in view of Remark 6, but also since two researchers with different choices of $s_{\max}$ would frequently find different estimates from the CKP procedure, even though they retain the same significance level.

On the other hand, our $\mathcal{T}_{\mathcal{K}}$ variance ratio procedure is very robust to the choice of $s_{\max}$. For all the data sets, it gives the same estimate for all $s_{\max}$ considered. For practical

application, this is a substantial advantage of our procedure.

6 Conclusion

We have proposed a testing procedure to determine the dimension of the nonstationary subspace (number of stochastic trends) in functional time series taking values in a Hilbert space. Our test statistic is of the variance ratio type, and in the univariate special case it reduces to the well-known KPSS statistic of Kwiatkowski et al. (1992) with bandwidth zero. The test is based on a projection onto a subspace of Hilbert space that is a superspace of the true nonstationary subspace with probability converging to one. We provided an easily implemented candidate for this required projection operator using empirical eigenvectors of covariance operators. We have derived the asymptotic distribution of the test statistic under the null hypothesis, which is a functional of standard Brownian motion. It does not depend on the choice of projection operator nor on the number of eigenvectors used to construct the projection operator. Monte Carlo simulation results were reported which provide evidence that our test has good finite sample properties and is preferred to the existing test of Chang et al. (2016). Finally, we applied our methodology to three empirical data sets, age-specific US employment curves, Australian temperature curves, and Ontario electricity demand curves, and in all cases found evidence of nontrivial nonstationary subspaces.

Appendix A: Preliminary lemmas

The first lemma shows convergence of the sample covariance operators.

Lemma 1. *Suppose that Assumptions 1 and 2 are satisfied. With the notation in Section 3.2 it holds that*

$$\|T^{-2}\hat{\mathcal{C}} - \mathcal{C}\|_{\mathcal{L}_{\mathcal{H}}} = o_p(1), \tag{A.1}$$

$$\|T^{-4}\hat{\mathcal{K}} - \mathcal{K}\|_{\mathcal{L}_{\mathcal{H}}} = o_p(1). \tag{A.2}$$

The next lemma shows the different behavior of the sample covariance operators in different directions of the parameter space.

Lemma 2. *Suppose that Assumptions 1, 2 and 3 are satisfied. With the notation in Section 3.3 the following holds.*

(i) *For any $v^T \in \operatorname{ran} P_\ell^T$ satisfying $\sup_T \|v^T\| < \infty$,*

$$|\langle T^{-2} P_\ell^T \hat{\mathcal{C}} P_\ell^T v^T, v^T\rangle - \langle \mathcal{C} v^T, v^T\rangle| = o_p(1),$$

$$|\langle T^{-4} P_\ell^T \hat{\mathcal{K}} P_\ell^T v^T, v^T\rangle - \langle \mathcal{K} v^T, v^T\rangle| = o_p(1).$$

(ii) *For any* $v^T \in \mathfrak{A}^\perp \cap \operatorname{ran} P_\ell^T$ *satisfying* $\sup_T \|v^T\| < \infty$,

$$\langle \mathcal{C} v^T, v^T \rangle = 0 \text{ and } \langle \mathcal{K} v^T, v^T \rangle = 0,$$
$$|\langle T^{-1} P_\ell^T \hat{\mathcal{C}} P_\ell^T v^T, v^T \rangle - \langle (I - P_{\mathfrak{A}}) C_\nu (I - P_{\mathfrak{A}}) v^T, v^T \rangle| = o_p(1),$$
$$\langle T^{-2} P_\ell^T \hat{\mathcal{K}} P_\ell^T v^T, v^T \rangle = O_p(1).$$

The results of Lemmas 1 and 2 will be important in the derivation of the limiting distribution of our test statistic. In the next two lemmas, these results are extended to accommodate deterministic terms.

Lemma 3. *Suppose that Assumptions 1 and 2 are satisfied. With the notation in Section 3.4 it holds that, for* $\varrho = 1, 2$,

$$\|T^{-2} \hat{\mathcal{C}}^{(\varrho)} - \mathcal{C}^{(\varrho)}\|_{\mathcal{L}_{\mathcal{H}}} = o_p(1), \tag{A.3}$$
$$\|T^{-4} \hat{\mathcal{K}}^{(\varrho)} - \mathcal{K}^{(\varrho)}\|_{\mathcal{L}_{\mathcal{H}}} = o_p(1), \tag{A.4}$$

where $\mathcal{C}^{(\varrho)} \overset{d}{=} \Lambda_{\Delta X}^{1/2} \widetilde{\mathcal{W}}^{(\varrho)} \Lambda_{\Delta X}^{1/2}$ *and* $\mathcal{K}^{(\varrho)} \overset{d}{=} \Lambda_{\Delta X}^{1/2} \widetilde{\mathcal{V}}^{(\varrho)} \Lambda_{\Delta X}^{1/2}$.

Lemma 4. *Suppose that Assumptions 1, 2, and 3 are satisfied. With the notation in Section 3.4 the following holds for* $\varrho = 1, 2$.

(i) *For any* $v^T \in \operatorname{ran} P_\ell^T$ *satisfying* $\sup_T \|v^T\| < \infty$,

$$|\langle T^{-2} P_\ell^T \hat{\mathcal{C}}^{(\varrho)} P_\ell^T v^T, v^T \rangle - \langle \mathcal{C}^{(\varrho)} v^T, v^T \rangle| = o_p(1),$$
$$|\langle T^{-4} P_\ell^T \hat{\mathcal{K}}^{(\varrho)} P_\ell^T v^T, v^T \rangle - \langle \mathcal{K}^{(\varrho)} v^T, v^T \rangle| = o_p(1).$$

(ii) *For any* $v^T \in \mathfrak{A}^\perp \cap \operatorname{ran} P_\ell^T$ *satisfying* $\sup_T \|v^T\| < \infty$,

$$\langle \mathcal{C}^{(\varrho)} v^T, v^T \rangle = 0 \text{ and } \langle \mathcal{K}^{(\varrho)} v^T, v^T \rangle = 0,$$
$$|\langle T^{-1} P_\ell^T \hat{\mathcal{C}}^{(\varrho)} P_\ell^T v^T, v^T \rangle - \langle (I - P_{\mathfrak{A}}) C_\nu (I - P_{\mathfrak{A}}) v^T, v^T \rangle| = o_p(1),$$
$$\langle T^{-2} P_\ell^T \hat{\mathcal{K}}^{(\varrho)} P_\ell^T v^T, v^T \rangle = O_p(1).$$

Appendix B: Proofs of theorems

B.1 Proof of Theorem 1

We consider the decomposition $P_\ell^T = P_\ell^T P_{\mathfrak{A}} + P_\ell^T (I - P_{\mathfrak{A}}) = P_{\ell,\mathfrak{A}}^T + P_{\ell,\mathfrak{A}^\perp}^T$, for $P_{\ell,\mathfrak{A}}^T = P_\ell^T P_{\mathfrak{A}}$ and $P_{\ell,\mathfrak{A}^\perp}^T = P_\ell^T (I - P_{\mathfrak{A}})$, and the following operator matrices,

$$\hat{\mathcal{K}}_\ell = P_\ell^T \hat{\mathcal{K}} P_\ell^T|_{\operatorname{ran} P_\ell^T \to \operatorname{ran} P_\ell^T} = \begin{pmatrix} (P_{\ell,\mathfrak{A}}^T)^* \hat{\mathcal{K}} P_{\ell,\mathfrak{A}}^T & (P_{\ell,\mathfrak{A}^\perp}^T)^* \hat{\mathcal{K}} P_{\ell,\mathfrak{A}}^T \\ (P_{\ell,\mathfrak{A}}^T)^* \hat{\mathcal{K}} P_{\ell,\mathfrak{A}^\perp}^T & (P_{\ell,\mathfrak{A}^\perp}^T)^* \hat{\mathcal{K}} P_{\ell,\mathfrak{A}^\perp}^T \end{pmatrix},$$
$$\hat{\mathcal{C}}_\ell = P_\ell^T \hat{\mathcal{C}} P_\ell^T|_{\operatorname{ran} P_\ell^T \to \operatorname{ran} P_\ell^T} = \begin{pmatrix} (P_{\ell,\mathfrak{A}}^T)^* \hat{\mathcal{C}} P_{\ell,\mathfrak{A}}^T & (P_{\ell,\mathfrak{A}^\perp}^T)^* \hat{\mathcal{C}} P_{\ell,\mathfrak{A}}^T \\ (P_{\ell,\mathfrak{A}}^T)^* \hat{\mathcal{C}} P_{\ell,\mathfrak{A}^\perp}^T & (P_{\ell,\mathfrak{A}^\perp}^T)^* \hat{\mathcal{C}} P_{\ell,\mathfrak{A}^\perp}^T \end{pmatrix}.$$

Let D_T denote the normalization operator matrix

$$D_T = \begin{pmatrix} T^{-1/2} I_1 & 0 \\ 0 & I_2 \end{pmatrix},$$

where I_1 and I_2 are properly defined identity operators. Then the generalized eigenvalue problem (3.9) can be rewritten as

$$(T^2\tau_j^T)(T^{-3}D_T\hat{\mathcal{K}}_\ell D_T)\xi_j^T = (T^{-1}D_T\hat{\mathcal{C}}_\ell D_T)\xi_j^T, \quad \xi_j^T \in \operatorname{ran} P_\ell^T. \tag{B.1}$$

By the isomorphism between $\mathbb{R}^\ell$ and any ℓ-dimensional subspace of $\mathcal{H}$, the generalized eigenvalue problem (B.1) may be understood as a standard eigenvalue problem in $\mathbb{R}^\ell$. Let $[T^{-1}D_T\hat{\mathcal{C}}_\ell D_T]$ (resp. $[T^{-3}D_T\hat{\mathcal{K}}_\ell D_T]$) be the matrix representation of $T^{-1}D_T\hat{\mathcal{C}}_\ell D_T$ (resp. $T^{-3}D_T\hat{\mathcal{K}}_\ell D_T$) with respect to the orthonormal basis $(\phi_1^T, \ldots, \phi_\ell^T)$ of $\operatorname{ran} P_\ell^T$, as given by

$$[T^{-1}D_T\hat{\mathcal{C}}_\ell D_T]_{ij} = \langle T^{-1}D_T\hat{\mathcal{C}}_\ell D_T\phi_j^T, \phi_i^T\rangle, \quad 1 \le i,j \le \ell,$$
$$[T^{-3}D_T\hat{\mathcal{K}}_\ell D_T]_{ij} = \langle T^{-3}D_T\hat{\mathcal{K}}_\ell D_T\phi_j^T, \phi_i^T\rangle, \quad 1 \le i,j \le \ell.$$

From Assumption 3, Remark 2, and the results in Lemma 2, it follows that

$$[T^{-3}D_T\hat{\mathcal{K}}_\ell D_T] - \begin{pmatrix} [\mathcal{K}] & 0 \\ 0 & 0 \end{pmatrix} \xrightarrow{p} 0, \qquad [\mathcal{K}]_{ij} = \langle \mathcal{K}\phi_j, \phi_i\rangle, \qquad 1 \le i,j \le s, \tag{B.2}$$

$$[T^{-1}D_T\hat{\mathcal{C}}_\ell D_T] - \begin{pmatrix} [\mathcal{C}] & 0 \\ 0 & [\widetilde{C}_\nu]^T \end{pmatrix} \xrightarrow{p} 0, \qquad [\mathcal{C}]_{ij} = \langle \mathcal{C}\phi_j, \phi_i\rangle, \qquad 1 \le i,j \le s, \tag{B.3}$$

$$[\widetilde{C}_\nu]_{ij}^T = \langle \widetilde{C}_\nu\phi_j^T, \phi_i^T\rangle, \qquad s+1 \le i,j \le \ell, \tag{B.4}$$

where $\widetilde{C}_\nu = (I - P_{\mathfrak{A}})C_\nu(I - P_{\mathfrak{A}})$. Moreover, by our Assumption 2(ii) and Remark 2, the matrix $[\widetilde{C}_\nu]^T$ is positive definite regardless of the limiting behavior of $(\phi_{s+1}^T, \ldots, \phi_\ell^T)$.

Given the results in (B.2)–(B.4), only the first s eigenvalues of (B.1) are finite in the limit when normalized by T^2. Specifically, it follows that

$$T^2 \sum_{j=1}^{s} \tau_j^T \xrightarrow{p} \sum_{j=1}^{s} \tau_j,$$
$$(T^2\tau_j^T)^{-1} \xrightarrow{p} 0, \quad j = s+1, \ldots, \ell,$$

which proves (3.12). The limiting eigenvalues $(\tau_1, \ldots, \tau_s)$ and corresponding eigenvectors $(\xi_1, \ldots, \xi_s)$ are defined by the limiting eigenvalue problem

$$\tau_j[\mathcal{K}]\xi_j = [\mathcal{C}]\xi_j, \quad \xi_j \in \mathbb{R}^s. \tag{B.5}$$

Thus, we may deduce from (B.5) and Remark 4 that

$$\sum_{j=1}^{s} \tau_j = \operatorname{tr}\left([\mathcal{K}]^{-1}[\mathcal{C}]\right) = \operatorname{tr}\left((\Lambda_{\Delta X}^{1/2}\widetilde{\mathcal{V}}\Lambda_{\Delta X}^{1/2}|_{\mathfrak{A}\to\mathfrak{A}})^{-1}\Lambda_{\Delta X}^{1/2}\widetilde{\mathcal{W}}\Lambda_{\Delta X}^{1/2}\right).$$

Finally, recall that $\Lambda_{\Delta X}^{1/2} : \mathfrak{A} \to \mathfrak{A}$ and $\widetilde{\mathcal{V}} : \mathfrak{A} \to \mathfrak{A}$ are both invertible operators (almost surely for $\widetilde{\mathcal{V}}$). By the properties of the trace, it then follows that

$$\sum_{j=1}^{s} \tau_j = \operatorname{tr}\left(((\Lambda_{\Delta X}^{1/2})|_{\mathfrak{A}\to\mathfrak{A}})^{-1}(\widetilde{\mathcal{V}}|_{\mathfrak{A}\to\mathfrak{A}})^{-1}\widetilde{\mathcal{W}}|_{\mathfrak{A}\to\mathfrak{A}}(\Lambda_{\Delta X}^{1/2}|_{\mathfrak{A}\to\mathfrak{A}})\right) = \operatorname{tr}\left((\widetilde{\mathcal{V}}|_{\mathfrak{A}\to\mathfrak{A}})^{-1}\widetilde{\mathcal{W}}|_{\mathfrak{A}\to\mathfrak{A}}\right),$$

which proves (3.11).

B.2 Proof of Theorem 2

The result that $\mathbb{P}\{\hat{s} > s\} \to 0$ is a direct consequence of the consistency of the test, i.e. of (3.12) in Theorem 1, and the assumption that $\mathbb{P}\{s_{\max} \geq s\} \to 1$. Thus, because $\mathbb{P}\{\hat{s} > s\} \to 0$, the sequential test procedure will reach the test of the null hypothesis that $s_0 = s$ with probability converging to one. This is a test of a true null, so we find from (3.11) in Theorem 1 that $\mathbb{P}\{\hat{s} = s\} \to 1 - \alpha$, which proves the required result.

B.3 Proof of Theorem 3

The proof is nearly identical to that of Theorem 1, but using Lemmas 3 and 4 instead of Lemmas 1 and 2, and hence is omitted.

B.4 Proof of Theorem 4

Let $\widetilde{\mathcal{K}}_\infty$ denote the limit $\mathcal{K}$ or $\mathcal{K}^{(\varrho)}$ depending on the specification of the deterministic component. Then it follows from Lemmas 1 and 3 that $\|T^{-4}\widetilde{\mathcal{K}} - \widetilde{\mathcal{K}}_\infty\|_{\mathcal{L}_\mathcal{H}} = o_p(1)$.

Let ϕ_j^T and ϕ_j denote the eigenvectors corresponding to the ordered eigenvalues of $\widetilde{\mathcal{K}}$ and $\widetilde{\mathcal{K}}_\infty$, respectively. Using the fact that the first s eigenvalues of $\widetilde{\mathcal{K}}_\infty$ are almost surely distinct, it follows from Lemma 3.2 of Hörmann and Kokoszka (2010) (as a generalization of Lemma 4.3 of Bosq (2000)) that

$$\|\phi_j^T - \operatorname{sgn}(\langle\phi_j^T, \phi_j\rangle)\phi_j\| = o_p(1), \quad j = 1, \ldots, s. \tag{B.6}$$

Note that $(\phi_1, \ldots, \phi_s)$ is a random orthonormal set, but the span is nonrandomly given by $\operatorname{span}(\phi_1, \ldots, \phi_s) = \mathfrak{A}$. Therefore, (B.6) implies that the set of eigenvectors $(\phi_1^T, \ldots, \phi_s^T)$ asymptotically spans $\mathfrak{A}$. Specifically, for any $x \in \mathfrak{A}$,

$$\begin{aligned} \|\langle\phi_j^T, x\rangle\phi_j^T - \langle\phi_j, x\rangle\phi_j\| &\xrightarrow{p} 0, \qquad j = 1, \ldots, s, \\ \|\langle\phi_j^T, x\rangle\phi_j^T\| &\xrightarrow{p} 0, \qquad j = s+1, \ldots, \ell. \end{aligned} \tag{B.7}$$

From (B.7), we may easily deduce that Assumption 3 is satisfied.

Appendix C: Proofs of lemmas

C.1 Proof of Lemma 1

Recalling that $X_t(u)$ is a random function of the argument $u \in [0, 1]$, we define the double-indexed function $Z_T(r, u) = T^{-1/2}X_{\lfloor Tr\rfloor}(u) = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta X_t(u)$ for $r \in [0, 1]$ and at the

functional value $u \in [0,1]$. When there is no risk of confusion, we also use the notation $Z_T(r) = T^{-1/2}X_{\lfloor Tr \rfloor}$ to denote the entire function of u. Under the summability condition $\sum_{j=0}^{\infty} j\|\Phi_j\|_{\mathcal{L}_\mathcal{H}} < \infty$ and Assumption 2, the sequence ΔX_t in (2.1) is a so-called L^4-m-approximable sequence; see Proposition 2.1 in Hörmann and Kokoszka (2010). Then, from Theorem 1.1 in Berkes et al. (2013) and the Skorokhod representation, it follows that

$$\sup_{0\le r\le 1} \left\| Z_T(r) - \overline{W}(r) \right\| \xrightarrow{p} 0, \tag{C.1}$$

where $\overline{W} \overset{d}{=} \Lambda_{\Delta X}^{1/2}\mathcal{W}$. Let $\mathcal{C} = \int \overline{W}(r) \otimes \overline{W}(r)dr$, then clearly $\mathcal{C} \overset{d}{=} \Lambda_{\Delta X}^{1/2}\widetilde{\mathcal{W}}\Lambda_{\Delta X}^{1/2}$.

Similarly, for all $r \in [0,1]$, we let $\overline{W}(r,u)$ denote the function value at $u \in [0,1]$ and use $\overline{W}(r)$ to denote the random (square-integrable) function of u. Both $\hat{\mathcal{C}}$ and $\mathcal{C}$ are integral operators, so we let $\hat{c}(u,w)$ (resp. $c(u,w)$) denote the kernel function of $T^{-2}\hat{\mathcal{C}}$ (resp. $\mathcal{C}$). These are given as follows,

$$\hat{c}(u,w) = \frac{1}{T^2}\sum_{t=1}^{T} X_t(u)X_t(w) = \int Z_T(r,u)Z_T(r,w)dr,$$
$$c(u,w) = \int \overline{W}(r,u)\overline{W}(r,w)dr.$$

To prove (A.1) we show a stronger result. An operator A is a compact operator if there exists two orthonormal bases, $(f_j, j \in \mathbb{N})$ and $(g_j, j \in \mathbb{N})$, and a real-valued sequence $(\gamma_j, j \in \mathbb{N})$ tending to zero, such that

$$Ax = \sum_{j=1}^{\infty} \gamma_j f_j \otimes g_j(x).$$

A compact operator A is said to be a Hilbert-Schmidt operator if $\sum_{j=1}^{\infty}\gamma_j^2 < \infty$. The so-called Hilbert-Schmidt norm of A is then given by

$$\|A\|_{\text{HS}} = \left(\sum_{j=1}^{\infty}\|Ag_j\|^2\right)^{1/2}$$

for any arbitrary orthonormal basis $(g_j, j \in \mathbb{N})$. The following norm inequality is well known,

$$\|\cdot\|_{\mathcal{L}_\mathcal{H}} \le \|\cdot\|_{\text{HS}}. \tag{C.2}$$

It thus suffices to show that $\|T^{-2}\hat{\mathcal{C}} - \mathcal{C}\|_{\text{HS}} = o_p(1)$.

Define the norm $\|g\|_{L\times L} = (\int\int g(u,w)^2 dudw)^{1/2}$ for a kernel function $g : [0,1]\times[0,1] \to \mathbb{R}$. Then note that

$$\|T^{-2}\hat{\mathcal{C}} - \mathcal{C}\|_{\text{HS}} = \|\hat{c} - c\|_{L\times L} \le \left(\int\int\int \left(Z_T(r,u)Z_T(r,w) - \overline{W}(r,u)\overline{W}(r,w)\right)^2 dudwdr\right)^{1/2}, \tag{C.3}$$

where the equality is because both $T^{-2}\hat{\mathcal{C}}$ and $\mathcal{C}$ are integral operators and the inequality is

the Cauchy-Schwarz inequality. The integrand in (C.3) is equal to the square of

$$(Z_T(r,u)-\overline{W}(r,u))(Z_T(r,w)-\overline{W}(r,w)) \\ +\overline{W}(r,u)(Z_T(r,w)-\overline{W}(r,w))+(Z_T(r,u)-\overline{W}(r,u))\overline{W}(r,w). \tag{C.4}$$

Using (C.4) and Minkowski's inequality, we deduce that (C.3) is bounded from above by

$$\begin{aligned}
&\left(\int\int\int \left(Z_T(r,u)-\overline{W}(r,u)\right)^2\left(Z_T(r,w)-\overline{W}(r,w)\right)^2 dudwdr\right)^{1/2} \\
&\quad +2\left(\int\int\int \overline{W}(r,u)^2\left(Z_T(r,w)-\overline{W}(r,w)\right)^2 dudwdr\right)^{1/2} \\
&=\left(\int\left(\int\left(Z_T(r,u)-\overline{W}(r,u)\right)^2 du\right)^2 dr\right)^{1/2} \\
&\quad +2\left(\int\left(\int \overline{W}(r,u)^2 du\right)\left(\int\left(Z_T(r,w)-\overline{W}(r,w)\right)^2 dw\right)dr\right)^{1/2} \\
&\le \sup_{0\le r\le 1}\left(\int\left(Z_T(r,u)-\overline{W}(r,u)\right)^2 du\right) \\
&\quad +2\sup_{0\le r\le 1}\left(\int \overline{W}(r,u)^2 du\right)^{1/2}\sup_{0\le r\le 1}\left(\int\left(Z_T(r,w)-\overline{W}(r,w)\right)^2 dw\right)^{1/2}=o_p(1),
\end{aligned}$$

where the last equality is from (C.1) and the fact that

$$\sup_{0\le r\le 1}\left\|\overline{W}(r)\right\|<\infty \quad \text{almost surely}$$

because $\overline{W}(r)$ is almost surely continuous on a bounded interval. Thus, (A.1) is established.

To prove (A.2), we note that instead of (C.1) we now have

$$\sup_{0\le r\le 1}\left\|T^{-3/2}\sum_{t=1}^{[Tr]}X_t-\int_0^r \overline{W}(u)du\right\|=\sup_{0\le r\le 1}\left\|\int_0^r Z_T(u)du-\int_0^r \overline{W}(u)du\right\|=o_p(1), \tag{C.5}$$

which follows from Lemma B.3 in Horváth et al. (2014) and the Skorokhod representation. The remainder of the proof is almost identical to that of (A.1), and is therefore omitted.

C.2 Proof of Lemma 2

First note that $v^T\in \operatorname{ran} P_\ell^T$ implies $P_\ell^T v^T=v^T$, so that

$$|\langle T^{-2}P_\ell^T\hat{\mathcal{C}}P_\ell^T v^T,v^T\rangle-\langle \mathcal{C}v^T,v^T\rangle|=|\langle (T^{-2}\hat{\mathcal{C}}-\mathcal{C})v^T,v^T\rangle|\le \|T^{-2}\hat{\mathcal{C}}-\mathcal{C}\|_{\mathcal{L}_{\mathcal{H}}}\sup_T\|v^T\|^2=o_p(1).$$

The inequality follows from Cauchy-Schwarz and properties of the operator norm and the final equality follows from Lemma 1. The proof of the second statement of part (i) is identical.

Next, we prove part (ii). The first statement is a direct consequence of the definitions of $\mathcal{C}$ and $\mathcal{K}$; see (3.7). Because $v^T\in\mathfrak{A}^\perp\cap\operatorname{ran}P_\ell^T$ we have $P_\ell^T(I-P_{\mathfrak{A}})v^T=v^T$, so that

$$\langle P_\ell^T\hat{\mathcal{C}}P_\ell^T v^T,v^T\rangle=\langle (I-P_{\mathfrak{A}})\hat{\mathcal{C}}(I-P_{\mathfrak{A}})v^T,v^T\rangle. \tag{C.6}$$

We then find that

$$\begin{aligned}
&|\langle T^{-1}P_\ell^T\hat{\mathcal{C}}P_\ell^T v^T, v^T\rangle - \langle (I-P_{\mathfrak{A}})C_\nu(I-P_{\mathfrak{A}})v^T, v^T\rangle| \\
&\quad = |\langle T^{-1}(I-P_{\mathfrak{A}})\hat{\mathcal{C}}(I-P_{\mathfrak{A}})v^T, v^T\rangle - \langle (I-P_{\mathfrak{A}})C_\nu(I-P_{\mathfrak{A}})v^T, v^T\rangle| \\
&\quad \leq \|T^{-1}(I-P_{\mathfrak{A}})\hat{\mathcal{C}}(I-P_{\mathfrak{A}})v^T, v^T\rangle - \langle (I-P_{\mathfrak{A}})C_\nu(I-P_{\mathfrak{A}})\|_{\mathcal{L}_{\mathcal{H}}} \sup_T \|v^T\|^2,
\end{aligned}$$

where the inequality follows from Cauchy-Schwarz and properties of the operator norm. We note that $T^{-1}(I-P_{\mathfrak{A}})\hat{\mathcal{C}}(I-P_{\mathfrak{A}})$ may be viewed as the sample covariance operator of $((I-P_{\mathfrak{A}})\nu_t, t \geq 1)$, and it follows from Assumption 2 and Mas (2002) that

$$\|T^{-1}(I-P_{\mathfrak{A}})\hat{\mathcal{C}}(I-P_{\mathfrak{A}}) - (I-P_{\mathfrak{A}})C_\nu(I-P_{\mathfrak{A}})\|_{\mathcal{L}_{\mathcal{H}}} = o_p(1), \tag{C.7}$$

which shows the second statement of part (ii). The proof of the third statement is identical.

C.3 Proof of Lemma 3

We need the result corresponding to (C.1) for the case with residuals. Similarly to the proof of Lemma 1, we define $Z_T^{(\varrho)}(r) = T^{-1/2}U_{\lfloor Tr\rfloor}^{(\varrho)}$ and show that

$$\sup_{0\leq r\leq 1}\left\|Z_T^{(\varrho)}(r) - \overline{W}^{(\varrho)}(r)\right\| = o_p(1), \tag{C.8}$$

where $\overline{W}^{(\varrho)}(r) \stackrel{d}{=} \Lambda_{\Delta X}^{1/2}\mathcal{W}^{(\varrho)}$. After showing (C.8), the remainder of the proof is identical to that of Lemma 1 and is therefore omitted.

The proof of (C.8) applies well-known techniques combined with the convergence results in (C.1) and (C.5). For example, for $\varrho = 1$, we have $U_t^{(1)} = X_t - T^{-1}\sum_{t=1}^T X_t$, so that $Z_T^{(1)}(r) = Z_T(r) - \int Z_T(w)dw$. Since $\overline{W}^{(1)}(r) = \overline{W}(r) - \int \overline{W}(w)dw$, the left-hand side of (C.8) for $\varrho = 1$ is bounded by

$$\sup_{0\leq r\leq 1}\left\|Z_T(r) - \overline{W}(r)\right\| + \sup_{0\leq r\leq 1}\left\|\int Z_T(w)dw - \int \overline{W}(w)dw\right\| = o_p(1), \tag{C.9}$$

where the convergence is from (C.1) and (C.5).

C.4 Proof of Lemma 4

This follows by nearly identical arguments to the proof of Lemma 2, using the results in Lemma 3, and is therefore omitted.

References

Ahn, S. K. and G. C. Reinsel (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.

Aue, A., G. Rice, and O. Sönmez (2017). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society B 80*, 509–529.

Aue, A. and A. van Delft (2020). Testing for stationarity of functional time series in the frequency domain. *Annals of Statistics 48*, 2505–2547.

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

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

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

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

Bosq, D. (2000). *Linear Processes in Function Spaces*. New York: Springer.

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

Chang, Y., R. K. Kaufmann, C. S. Kim, I. J. Miller, J. Y. Park, and S. Park (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics 214*, 274–294.

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

Flood, S., M. King, R. Rodgers, S. Ruggles, and J. R. Warren (2018). *Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]*. Minneapolis, MN: IPUMS.

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

Ho, M. S. and B. E. Sørensen (1996). Finding cointegration rank in high dimensional systems using the Johansen test: an illustration using data based Monte Carlo simulations. *Review of Economics and Statistics 78*, 726–732.

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

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. New York: Springer.

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

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

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

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

Mas, A. (2002). Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Processes and their Applications 99*, 117–135.

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

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

Nielsen, M. Ø. (2009). A powerful test of the autoregressive unit root hypothesis based on a tuning parameter free statistic. *Econometric Theory 25*, 1515–1544.

Nielsen, M. Ø. (2010). Nonparametric cointegration analysis of fractional systems with unknown integration orders. *Journal of Econometrics 155*, 170–187.

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

Pedroni, P. L., T. J. Vogelsang, M. Wagner, and J. Westerlund (2015). Nonparametric rank tests for non-stationary panels. *Journal of Econometrics 185*, 378–391.

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

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

Taylor, A. M. R. (2005). Variance ratio tests of the seasonal unit root hypothesis. *Journal of Econometrics 124*, 33–54.

Yao, J., A. Kammoun, and J. Najim (2012). Eigenvalue estimation of parameterized covariance matrices of large dimensional data. *IEEE Transactions on Signal Processing 60*, 5893–5905.

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