Article:

Li, Degui orcid.org/0000-0001-6802-308X, Robinson, Peter M. and Shang, Hanlin (2023) Nonstationary Fractionally Integrated Functional Time Series. Bernoulli. pp. 1505-1526. ISSN 1350-7265

https://doi.org/10.3150/22-BEJ1508

Nonstationary Fractionally Integrated Functional Time Series

DEGUI LI[1], PETER M. ROBINSON[2] and HAN LIN SHANG[3]

[1] *Department of Mathematics, University of York, York, YO10 5DD, UK*
[2] *Department of Economics, London School of Economics, London WC2A 2AE, UK*
[3] *Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney NSW 2109, Australia*

We study a functional version of nonstationary fractionally integrated time series, covering the functional unit root as a special case. The time series taking values in an infinite-dimensional separable Hilbert space are projected onto a finite number of sub-spaces, the level of nonstationarity allowed to vary over them. Under regularity conditions, we derive a weak convergence result for the projection of the fractionally integrated functional process onto the asymptotically dominant sub-space, which retains most of the sample information carried by the original functional time series. Through the classic functional principal component analysis of the sample variance operator, we obtain the eigenvalues and eigenfunctions which span a sample version of the dominant sub-space. Furthermore, we introduce a simple ratio criterion to consistently estimate the dimension of the dominant sub-space, and use a semiparametric local Whittle method to estimate the memory parameter. Monte-Carlo simulation studies are given to examine the finite-sample performance of the developed techniques.

Keywords: fractional integration; functional principal component analysis; functional time series; local Whittle estimation; nonstationary process

1. Introduction

In recent years, there has been increasing interest in studying functional time series. Examples of functional time series include intraday stock price curves with each functional observation defined as a pricing function of time points within a trading day (Horváth, Kokoszka and Rice, 2014), and age-specific fertility rate curves with each functional observation defined as a function of different ages for a particular calendar year (Li, Robinson and Shang, 2020). Most of the existing literature assumes that the functional observations are either independent or stationary weakly dependent (e.g., Bosq, 2000, Ramsay and Silverman, 2005, Ferraty and Vieu, 2006, Bathia, Yao and Ziegelmann, 2010, Hörmann and Kokoszka, 2010, Horváth and Kokoszka, 2012). These authors introduce various approaches to reduce the infinite dimension to a finite dimension, and subsequently apply statistical tools developed for scalar or multivariate time series processes.

The stationarity assumption is often rejected when testing functional time series collected in economics and finance. For example, Horváth, Kokoszka and Rice (2014) reject it via a functional version of KPSS test (Kwiatkowski et al., 1992) for some intraday stock price curves; and Chang, Kim and Park (2016) find evidence of unit root or $I(1)$ structure for intra-month distribution of S&P 500 index returns. Therefore, it seems worthwhile to explore nonstationary pattern in functional time series analysis. Univariate and multivariate nonstationary time series have been extensively studied in the last few decades (e.g., Engle and Granger, 1987, Johansen, 1991, Phillips, 1995, Robinson and Marinucci, 2001, Robinson and Hualde, 2003, Chen and Hurvich, 2006). In contrast, infinite dimensional curve time series with nonstationarity have received much less attention.

Chang, Kim and Park (2016) study nonstationarity of the time series of state densities by dividing an infinite-dimensional Hilbert space into the $I(1)$ and stationarity subspaces, and use the *functional*

principal component analysis (FPCA) to estimate the finite-dimensional $I(1)$ subspace. Beare, Seo and Seo (2017) consider a cointegrated linear process in the Hilbert space and derive the Granger-Johansen representation theorem for infinite-dimensional $I(1)$ functional autoregressive processes. This result is extended by Beare and Seo (2020) to I(2) autoregressive processes in the Hilbert space and further extended by Franchi and Paruolo (2020) to more general functional $I(d)$ autoregressive processes with $d = 1, 2, \cdots$. Derivation of the main theoretical theorems in Beare and Seo (2020) and Franchi and Paruolo (2020) relies on the analytic Fredholm theorem (e.g., Gohberg, Goldberg and Kaashoek, 1990), which gives the inversion of the analytic operator functions. To facilitate estimation and inference, it is crucial to estimate the dimension of the dominant nonstationary subspace, which is often called as the attractor space (e.g., Beare, Seo and Seo, 2017, Beare and Seo, 2020, Franchi and Paruolo, 2020). Chang, Kim and Park (2016) use a generalised eigenvalue test to determine the dimension of the I(1) subspace, whereas Nielsen, Seo and Seong (2019) propose a variance ratio-type test to determine the dimension.

The aforementioned literature focuses on nonstationary functional time series with integration order being a positive integer. In the present paper, we introduce a more general nonstationary framework for functional time series, including functional $I(d)$ structure ($d = 1, 2, \cdots$) as a special case, namely a functional version of the nonstationary fractionally integrated time series. We allow the nonstationarity level to vary over different sub-spaces, as specified in Section 2 below. The proposed framework is a natural extension of the "Type II" multivariate fractionally integrated process, introduced by Marinucci and Robinson (2000) and Robinson and Marinucci (2001), and further studied in Phillips and Shimotsu (2004), Robinson (2005) and Shimotsu and Phillips (2005). We are particularly interested in a functional sub-space denoted by $\mathcal{H}_1$ on which the orthogonal projection of the nonstationary functional time series retains most of the sample information, and thus term it the *asymptotically dominant sub-space*, as in Li, Robinson and Shang (2020). Under regularity conditions, we derive a weak convergence result for projection of the fractionally integrated functional process onto the dominant sub-space, extending Theorem 1 in Marinucci and Robinson (2000) from the multivariate setting to the functional setting with multivariate Brownian motion replaced by Brownian motion on $\mathcal{H}_1$ and complementing weak convergence results developed for stationary functional process (e.g., Berkes, Horváth and Rice, 2013). Furthermore, in order to approximate the dominant nonstationary sub-space $\mathcal{H}_1$, we construct a sample variance operator, derive its asymptotic properties and then implement the classic FPCA. Consequently, we obtain the estimated eigenvalues and associated eigenfunctions whose span can be viewed as a sample version of $\mathcal{H}_1$.

To facilitate inference, we further estimate two key elements: the dimension of $\mathcal{H}_1$ and the memory parameter. As in Li, Robinson and Shang (2020), we use a simple ratio criterion to estimate the dimension of $\mathcal{H}_1$ and prove that the estimator is weakly consistent. Our method is easy to implement and substantially different from that in Chang, Kim and Park (2016) and Nielsen, Seo and Seong (2019). We apply the local Whittle method which is introduced by Künsch (1987) and systematically studied by Robinson (1995), Velasco (1999) and Phillips and Shimotsu (2004) to estimate the memory parameter. Noting that local Whittle estimation is biased when the true memory parameter exceeds one, we further introduce an algorithm involving appropriate integer-order differencing to obtain consistent estimate. The developed methodology and the relevant asymptotic property extend those in Li, Robinson and Shang (2020, 2021) from stationary long-range dependent functional processes to nonstationary fractionally integrated processes. Monte-Carlo simulation further shows that the developed methodologies have reliable performance in finite samples.

The rest of the paper is organised as follows. Section 2 introduces the model setting, gives some technical assumptions and the weak convergence for the projected functional process. Section 3 constructs the sample variance operator, implements FPCA and derives the relevant asymptotic results. Section 4 estimates the dimension and memory parameter of the dominant sub-space. Section 5 gives a Monte-Carlo simulation study. Section 6 concludes the paper. Proofs of the main results are given in Appendix

A. Some technical lemmas with their proofs, additional simulation studies and empirical applications are available in a supplement. Throughout the paper, we define the separable space $\mathcal{H}$ as the set of real measurable functions on a compact set $\mathbb{S}$ such that $\int_{\mathbb{S}} z^2(u)du < \infty$. The relevant inner product is $\langle z_1, z_2\rangle = \int_{\mathbb{S}} z_1(u)z_2(u)du$ and the norm is $\|z\| = \langle z, z\rangle^{1/2}$. Denote by $\mathcal{L}_{\mathcal{H}}$ the space of continuous linear operators from $\mathcal{H}$ to $\mathcal{H}$ equipped with the operator norm defined by $\|\boldsymbol{L}\| = \sup_{z\in\mathcal{H}}\{\|\boldsymbol{L}(z)\| : \|z\| \le 1\}$. Let $\boldsymbol{L}'$ be the adjoint of $\boldsymbol{L}$, which is defined via $\langle \boldsymbol{L}x, y\rangle = \langle x, \boldsymbol{L}'y\rangle$ for all $x, y \in \mathcal{H}$. For $\boldsymbol{L} \in \mathcal{L}_{\mathcal{H}}$, we define two associated linear subspaces: $\mathsf{ker}(\boldsymbol{L}) = \{x \in \mathcal{H} : \boldsymbol{L}(x) = 0\}$ and $\mathsf{ran}(\boldsymbol{L}) = \{\boldsymbol{L}(x) : x \in \mathcal{H}\}$, which are called as the kernel and range (or image) of $\boldsymbol{L}$, respectively. Let $z_1 \otimes z_2 = \langle z_1, \cdot\,\rangle z_2$ for all $z_1, z_2 \in \mathcal{H}$, and let $\mathcal{H}_1 \oplus \mathcal{H}_2$ denote a direct sum of two sub-spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ in the sense that for any $z \in \mathcal{H}_1 \oplus \mathcal{H}_2$, we can uniquely write $z = z_1 + z_2$ with $z_i \in \mathcal{H}_i$, $i = 1, 2$. Let $\xrightarrow{D}$, $\xrightarrow{P}$ and $\Rightarrow$ denote convergence in distribution, convergence in probability and weak convergence (in appropriate space), respectively.

2. Model and assumptions

Assume that X_t is a functional process in $\mathcal{H}$ defined by

$$X_t = \sum_{i=1}^{t} \boldsymbol{\Psi}_{t-i}\eta_i, \quad t = 1, 2, \cdots, \tag{1}$$

where $(\boldsymbol{\Psi}_i : i \ge 0)$ is a sequence of continuous linear operators in $\mathcal{L}_{\mathcal{H}}$ and $(\eta_i : \ i \ge 1)$ is a stationary linear process in $\mathcal{H}$, which satisfy Assumptions 2.1 and 2.2, respectively.

Assumption 2.1. *Let $\boldsymbol{\Psi}_0 = \boldsymbol{I}$, an identity operator from $\mathcal{H}$ to $\mathcal{H}$, and for $i \ge 1$,*

$$\boldsymbol{\Psi}_i = \sum_{j=1}^{s_0} i^{d_j-1} L_j(i)\boldsymbol{B}_j, \quad d_1 > d_2 > \cdots > d_{s_0}, \tag{2}$$

where $d_1 > 1/2$, s_0 is a fixed positive integer, $L_j(\cdot)$ is a slowly-varying function at infinity satisfying

$$\sup_{0\le\theta\le 1} \left|L_j(i+\theta) - L_j(i)\right| \le c_j L_j^{\diamond}(i)/i, \quad j = 1, \cdots, s_0$$

with c_j being a positive constant and $L_j^{\diamond}(i)$ being a positive slowly-varying function, and $\boldsymbol{B}_j$, $j = 1, \cdots, s_0$, are continuous linear operators in $\mathcal{L}_{\mathcal{H}}$ which are invariant over i.

Assumption 2.2. *The stationary functional process η_t is generated by*

$$\eta_t = \sum_{i=0}^{\infty} \boldsymbol{A}_i \varepsilon_{t-i}, \quad t = 1, 2, \cdots, \tag{3}$$

where $(\boldsymbol{A}_i : i \ge 0)$ is a sequence of continuous linear operators in $\mathcal{L}_{\mathcal{H}}$ with the operator norm satisfying $\sum_{i=1}^{\infty} i\|\boldsymbol{A}_i\| < \infty$, and $(\varepsilon_t : t \in \mathbb{Z})$ is a sequence of independent and identically distributed (i.i.d.) random elements in $\mathcal{H}$ with mean zero, covariance function $\boldsymbol{\Omega}(u, v)$, $u, v \in \mathbb{C}$ and $\mathsf{E}\left[\|\varepsilon_i\|^p\right] < \infty$, $p > \max\{2, 2/(2d_1 - 1)\}$.

Assumption 2.1 indicates that the sequence $\left(i^{d_j-1}L_j(i) : i \geq 1\right)$ is quasi-monotonically decaying to zero, which is commonly assumed in the literature on scalar strongly dependent process (e.g., Marinucci and Robinson, 2000, Robinson and Marinucci, 2001). A key assumption to our main asymptotic theory such as Proposition 2.1 and Theorem 3.1 is $d_1 > 1/2$, whereas the remaining parameters $d_2, \cdots, d_{s_0}$ can be either larger than $1/2$ (i.e., nonstationary region) or smaller than $1/2$ (i.e., stationary region). As $d_1 > 1/2$, the operator norm of $\boldsymbol{\Psi}_i$ defined in (2) is not square-summable and the functional process X_t defined in (1) is not asymptotically stationary. The conditions in Assumption 2.2 ensure that η_t is a stationary and short-range dependent functional process. As in Johansen and Nielsen (2012), the moment condition on ε_i becomes very strong when d_1 approaches the boundary value $1/2$.

In this paper, we call model (1) a fractionally integrated functional process, which can be viewed as a functional generalisation of the "Type II" fractionally integrated multivariate process studied in Marinucci and Robinson (2000), Robinson and Marinucci (2001), Phillips and Shimotsu (2004) and Shimotsu and Phillips (2005). For the very special case $\boldsymbol{\Psi}_i \equiv \boldsymbol{I}$ for $i \geq 0$, (1) reduces to the $I(1)$ functional process $X_t = \sum_{i=1}^{t} \eta_i$ which is considered by Chang, Kim and Park (2016) and Beare, Seo and Seo (2017). Noting that $\boldsymbol{\Psi}_i = i^{d_1-1}L_1(i)\boldsymbol{B}_1(1+o(1))$ for i sufficiently large, we may call (1) a functional $I(d_1)$ process. An alternative nonstationary fractionally integrated functional process is

$$X_t = X_{t-1} + U_t = X_0 + \sum_{k=1}^{t} U_t, \tag{4}$$

where U_t is an $I(d_1-1)$ functional process with $1/2 \leq d_1 < 3/2$, which may be generated by a functional version of stationary fractionally integrated autoregressive moving averages considered as in Section 4 of Li, Robinson and Shang (2020). Model (4) is a natural functional extension of the "Type I" multivariate fractionally integrated process introduced by Velasco (1999), and can be extended to cover $d_1 \geq 3/2$ by repeated use of partial summation. Robinson (2005) discusses the difference between the "Type I" and "Type II" processes in the univariate time series setting. The present paper concentrates on "Type II" structure (1) since it directly provides a valid representation for all $d_1 > 1/2$.

We next provide a decomposition of X_t using the Beveridge-Nelson (B-N) decomposition (Phillips and Solo, 1992) and Abel summation by parts. This decomposition motivates the introduction of Assumption 2.3 below which is crucial to achieve dimension reduction. Letting $S_k(\eta) = \sum_{j=1}^{k} \eta_j$, by the B-N decomposition, we can show that

$$S_k(\eta) = \boldsymbol{A} \sum_{j=1}^{k} \varepsilon_j + \widetilde{\eta}_0 - \widetilde{\eta}_k, \tag{5}$$

where $\boldsymbol{A} = \sum_{i=0}^{\infty} \boldsymbol{A}_i$ and $\widetilde{\eta}_t = \sum_{i=0}^{\infty} \widetilde{\boldsymbol{A}}_i \varepsilon_{t-i}$ with $\widetilde{\boldsymbol{A}}_i = \sum_{j=i+1}^{\infty} \boldsymbol{A}_j$. From (2), we may write $\boldsymbol{\Psi}_i = \sum_{j=1}^{s_0} \psi_{i,j}\boldsymbol{B}_j$ with $\psi_{i,j} = i^{d_j-1}L_j(i)$, which, together with (1), leads to

$$X_t = \sum_{j=1}^{s_0} \sum_{i=1}^{t} \psi_{t-i,j}\boldsymbol{B}_j \eta_i =: \sum_{j=1}^{s_0} X_{t,j}. \tag{6}$$

Since $d_1 > d_2 > \cdots > d_{s_0}$, we may show that $X_{t,1}$ is the leading term on the right side of (6) when t is sufficiently large. Letting $\boldsymbol{\Psi}_{i,1} = \psi_{i,1}\boldsymbol{B}_1 = i^{d_1-1}L_1(i)\boldsymbol{B}_1$ for $i \geq 1$ and $\boldsymbol{\Psi}_{0,1} = \boldsymbol{I}$, and noting that

$S_0(\eta) = 0$, by Abel summation by parts and the B-N decomposition in (5), for $t \geq 2$,

$$\begin{aligned} X_{t,1} &= \sum_{i=1}^{t} \mathbf{\Psi}_{t-i,1}\eta_i = \sum_{i=1}^{t} \mathbf{\Psi}_{t-i,1}\left(S_i(\eta) - S_{i-1}(\eta)\right) \\ &= \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) S_i(\eta) + S_t(\eta) \\ &= \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) \overline{S}_i(\eta) + \overline{S}_t(\eta) + \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) \widetilde{\eta}_0 - \\ &\quad \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) \widetilde{\eta}_i + \widetilde{\eta}_0 - \widetilde{\eta}_t, \\ &=: X^*_{t,1} + X^\dagger_{t,1}, \end{aligned} \tag{7}$$

where $\overline{S}_t(\eta) = \boldsymbol{A} \sum_{k=1}^{t} \varepsilon_k$,

$$X^*_{t,1} = \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) \overline{S}_i(\eta) + \overline{S}_t(\eta),$$

$$X^\dagger_{t,1} = \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) \widetilde{\eta}_0 - \sum_{i=1}^{t-1} \left(\mathbf{\Psi}_{t-i,1} - \mathbf{\Psi}_{t-i-1,1}\right) \widetilde{\eta}_i + \widetilde{\eta}_0 - \widetilde{\eta}_t.$$

Combining (6) and (7), we have

$$X_t = X^*_{t,1} + X^\dagger_{t,1} + X_{t,-1} \quad \text{with} \quad X_{t,-1} = \sum_{j=2}^{s_0} X_{t,j}. \tag{8}$$

We may show that $X^\dagger_{t,1}$ and $X_{t,-1}$ are asymptotically dominated by $X^*_{t,1}$ when t is sufficiently large (see the proof of Proposition 2.1 in Appendix A). Furthermore, by (2) and the definition of $\overline{S}_i(\eta)$, we write

$$X^*_{t,1} = \sum_{i=1}^{t} \mathbf{\Psi}_{t-i,1}\boldsymbol{A}\varepsilon_i = \sum_{i=1}^{t-1} \psi_{t-i,1}\boldsymbol{B}_1\boldsymbol{A}\varepsilon_i + \boldsymbol{A}\varepsilon_t. \tag{9}$$

Assumption 2.3. *The operator $\boldsymbol{B}_{1,A} := \boldsymbol{B}_1\boldsymbol{A}$ has finite positive rank q_1.*

Assumption 2.3 is key to reduct infinite dimension to finite dimension in our main asymptotic analysis. Motivated by the representation in (8) and (9), we define $\mathcal{H}_1 = \mathsf{ran}(\boldsymbol{B}_{1,A})$, which is a closed linear subspace of $\mathcal{H}$ with finite dimension q_1. For any non-zero $v \in \mathcal{H}_1$, the coordinate process $\langle v, X_t \rangle$ is univariate nonstationary $I(d_1)$. Hence, we call $\mathcal{H}_1$ as the nonstationary $I(d_1)$ sub-space or the attractor space as in Beare, Seo and Seo (2017) and Beare and Seo (2020). As the orthogonal projection of X_t onto $\mathcal{H}_1$ results in the strongest nonstationary signal and retains most of the sample information, we may also call $\mathcal{H}_1$ the dominant sub-space as in Li, Robinson and Shang (2020). Let $\mathcal{H}_1^\perp$ be the orthogonal complement of $\mathcal{H}_1$ such that $\mathcal{H}_1 \oplus \mathcal{H}_1^\perp = \mathcal{H}$. By the definition of $\mathcal{H}_1$, we have

$\mathcal{H}_1^\perp = \left[\mathsf{ran}(\boldsymbol{B}_{1,A})\right]^\perp = \mathsf{ker}(\boldsymbol{B}'_{1,A})$. In fact, for any $v \in \mathcal{H}_1^\perp$, we may show that the order of integration for the coordinate process $\langle v, X_t\rangle$ is strictly smaller than d_1, and thus term $\mathcal{H}_1^\perp$ as the cointegrating space. The definitions of $\mathcal{H}_1$ and $\mathcal{H}_1^\perp$ are similar to those in Beare, Seo and Seo (2017), Beare and Seo (2020) and Franchi and Paruolo (2020), which derive the Granger-Johansen representation theorems for the autoregressive processes in the Hilbert space and then construct the $I(d)$ subspace ($d = 1, 2, \cdots$) and the cointegrating space using the autoregressive coefficient operators.

Let $\boldsymbol{P}_1$ and $\boldsymbol{P}_1^\perp$ be the orthogonal projections on $\mathcal{H}_1$ and $\mathcal{H}_1^\perp$, respectively, and $\boldsymbol{P}_1 + \boldsymbol{P}_1^\perp = \boldsymbol{I}$. Let $\mathcal{D}([0,1],\mathcal{R})$ be the space of real-valued functions on $[0,1]$ which are continuous on the right and have finite left limit. The space $\mathcal{D}([0,1],\mathcal{H}_1)$ is defined similarly to $\mathcal{D}([0,1],\mathcal{R})$ but with $\mathcal{H}_1$-valued functions replacing real-valued ones. For a random sequence $(Z_t : t \geq 1)$ on $\mathcal{H}_1$, we write $Z_{\lfloor nr\rfloor} \Rightarrow Z(r)$ in the space $\mathcal{D}([0,1],\mathcal{H}_1)$, if $\langle v, Z_{\lfloor nr\rfloor}\rangle \Rightarrow \langle v, Z(r)\rangle$ in the space $\mathcal{D}([0,1],\mathcal{R})$ for any $v \in \mathcal{H}_1$, where $\lfloor\cdot\rfloor$ denotes the floor function and $Z(\cdot)$ is a limiting stochastic process on $\mathcal{H}_1$. The following proposition gives the weak convergence of X_t in the direction determined by $\boldsymbol{P}_1$, which plays a key role in proving the main asymptotic theorems to be given in Section 3.

Proposition 2.1. *Suppose that Assumptions 2.1– 2.3 are satisfied and let $\boldsymbol{\Omega}_1 = \boldsymbol{P}_1\boldsymbol{B}_{1,A}\boldsymbol{\Omega}\boldsymbol{B}'_{1,A}\boldsymbol{P}_1$ be positive definite on $\mathcal{H}_1$. For $0 \leq r \leq 1$,*

$$\frac{1}{n^{d_1-1/2}L_1(n)}\boldsymbol{P}_1 X_{\lfloor nr\rfloor} \Rightarrow W_{d_1,\boldsymbol{\Omega}_1}(r) \tag{10}$$

in the space $\mathcal{D}([0,1],\mathcal{H}_1)$, where the limiting process $W_{d_1,\boldsymbol{\Omega}_1}(\cdot)$ takes value on $\mathcal{H}_1$ with $W_{d_1,\boldsymbol{\Omega}_1}(0) = 0$ almost surely (a.s.) and

$$W_{d_1,\boldsymbol{\Omega}_1}(r) = \int_0^r (r-s)^{d_1-1} dB_{\boldsymbol{\Omega}_1}(s), \tag{11}$$

$B_{\boldsymbol{\Omega}_1}(\cdot)$ is a Brownian motion on $\mathcal{H}_1$ with variance operator $\boldsymbol{\Omega}_1$.

The above proposition is an extension of Theorem 1 in Marinucci and Robinson (2000) from the multivariate setting to the functional setting with the multivariate Brownian motion in Marinucci and Robinson (2000) replaced by a Brownian motion on the sub-space $\mathcal{H}_1$. For the special case $d_1 = 1$ and $L_1(\cdot) \equiv c_0$, a positive constant, $W_{d_1,\boldsymbol{\Omega}_1}(\cdot)$ reduces to the Brownian motion $B_{\boldsymbol{\Omega}_1}(\cdot)$, and Proposition 2.1 becomes the functional weak convergence result derived by Chang, Kim and Park (2016).

We can make a further decomposition on the cointegrating space $\mathcal{H}_1^\perp$. Suppose that $d_2 > 1/2$, $0 < d_1 - d_2 < 1/2$, and the operator $\boldsymbol{P}_1^\perp\boldsymbol{B}_{2,A}$ with $\boldsymbol{B}_{2,A} = \boldsymbol{B}_2\boldsymbol{A}$ has finite positive rank q_2. Let $\mathcal{H}_2 = \mathsf{ran}(\boldsymbol{P}_1^\perp\boldsymbol{B}_{2,A})$, which is closed and finite dimensional. For any non-zero $v \in \mathcal{H}_2$, we may show that $\langle v, X_t\rangle$ is a univariate nonstationary $I(d_2)$ process. Define the sub-space $\mathcal{H}_2^\perp$ via $\mathcal{H}_2 \oplus \mathcal{H}_2^\perp = \mathcal{H}_1^\perp$. Let $\boldsymbol{P}_2$ and $\boldsymbol{P}_2^\perp$ be the orthogonal projections on $\mathcal{H}_2$ and $\mathcal{H}_2^\perp$, respectively, and $\boldsymbol{P}_2 + \boldsymbol{P}_2^\perp = \boldsymbol{P}_1^\perp$. Following the proof of Proposition 2.1 in Appendix A and strengthening the moment condition in Assumption 2.2, we may show that

$$\frac{1}{n^{d_2-1/2}L_2(n)}\boldsymbol{P}_2 X_{\lfloor nr\rfloor} \Rightarrow W_{d_2,\boldsymbol{\Omega}_2}(r) \tag{12}$$

in the space $\mathcal{D}([0,1],\mathcal{H}_2)$, where $W_{d_2,\boldsymbol{\Omega}_2}(\cdot)$ takes value on $\mathcal{H}_2$ with $W_{d_2,\boldsymbol{\Omega}_2}(0) = 0$ a.s. and $W_{d_2,\boldsymbol{\Omega}_2}(r) = \int_0^r (r-s)^{d_2-1} dB_{\boldsymbol{\Omega}_2}(s)$, $B_{\boldsymbol{\Omega}_2}(\cdot)$ is a Brownian motion on $\mathcal{H}_2$ with variance operator $\boldsymbol{\Omega}_2 = \boldsymbol{P}_2\boldsymbol{B}_{2,A}\boldsymbol{\Omega}\boldsymbol{B}'_{2,A}\boldsymbol{P}_2$ which is assumed to be positive definite on $\mathcal{H}_2$. Note that $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_1^\perp = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_2^\perp$, which is similar to the subspace decomposition in Remark 4.5 of Beare and Seo (2020) and the POLE(2) condition in Franchi and Paruolo (2020).

3. Sample variance operator and FPCA

In this section, we consider approximating the sub-space $\mathcal{H}_1$ and the associated orthogonal projection $\boldsymbol{P}_1$ by the corresponding sample version. To achieve this, we have to construct a sample variance operator, derive its asymptotic property and then implement the FPCA. Define an unnormalised variance operator by

$$\boldsymbol{V}_n = \sum_{t=1}^{n} X_t \otimes X_t, \quad \text{or equivalently,} \quad \boldsymbol{V}_n(z) = \sum_{t=1}^{n} \langle X_t, z \rangle X_t \tag{13}$$

for $z \in \mathcal{H}$, where X_t is defined in (1). As $X_t = \boldsymbol{P}_1 X_t + \boldsymbol{P}_1^{\perp} X_t$, we write

$$\begin{aligned} \boldsymbol{V}_n &= \sum_{t=1}^{n} \left(\boldsymbol{P}_1 + \boldsymbol{P}_1^{\perp}\right) X_t \otimes X_t \left(\boldsymbol{P}_1 + \boldsymbol{P}_1^{\perp}\right) \\ &= \sum_{t=1}^{n} \boldsymbol{P}_1 X_t \otimes X_t \boldsymbol{P}_1 + \sum_{t=1}^{n} \boldsymbol{P}_1 X_t \otimes X_t \boldsymbol{P}_1^{\perp} + \sum_{t=1}^{n} \boldsymbol{P}_1^{\perp} X_t \otimes X_t \boldsymbol{P}_1 + \sum_{t=1}^{n} \boldsymbol{P}_1^{\perp} X_t \otimes X_t \boldsymbol{P}_1^{\perp} \\ &=: \boldsymbol{V}_{n1} + \boldsymbol{V}_{n2} + \boldsymbol{V}_{n3} + \boldsymbol{V}_{n4}. \end{aligned} \tag{14}$$

Using Proposition 2.1 in Section 2, we can prove the following theorem which describes the asymptotic behavior of $\boldsymbol{V}_n$. For a sequence of operators $(\boldsymbol{Z}_t : t \geq 1)$ on $\mathcal{H}_1$, we write $\boldsymbol{Z}_n \xrightarrow{D} \boldsymbol{Z}$, if $\langle v_1, \boldsymbol{Z}_n(v_2) \rangle \xrightarrow{D} \langle v_1, \boldsymbol{Z}(v_2) \rangle$ for any $v_1, v_2 \in \mathcal{H}_1$.

Theorem 3.1. *Suppose that the assumptions of Proposition 2.1 are satisfied. Then we have*

$$\frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_{n1} \xrightarrow{D} \boldsymbol{V} := \int_0^1 W_{d_1, \boldsymbol{\Omega}_1}(r) \otimes W_{d_1, \boldsymbol{\Omega}_1}(r) dr, \tag{15}$$

where $W_{d_1, \boldsymbol{\Omega}_1}(\cdot)$ is defined in Proposition 2.1. For $k = 2, 3, 4$, $\|\boldsymbol{V}_{nk}\| = o_P\left(n^{2d_1} L_1^2(n)\right)$.

When $d_1 = 1$, $W_{d_1, \boldsymbol{\Omega}_1}(\cdot)$ in (15) would be replaced by the Brownian motion $B_{\boldsymbol{\Omega}_1}(\cdot)$ and we can obtain convergence results similar to those in Lemma 3.1 in Chang, Kim and Park (2016). If $s_0 \geq 2$ and $d_2 > 1/2$, we may derive more explicit rates for $\|\boldsymbol{V}_{nk}\|$, $k = 2, 3, 4$. Specifically, for $k = 2$ and 3,

$$\|\boldsymbol{V}_{nk}\| = \begin{cases} O_P\left(\left[n^{d_1+d_2} L_*^2(n)\right] \vee \left[n^{2d_1 - 1/2} L_*^2(n)\right]\right), & d_1 > 1, \\ O_P\left(n^{d_1+d_2} L_*^2(n)\right), & 1/2 < d_1 \leq 1, \end{cases} \tag{16}$$

and

$$\|\boldsymbol{V}_{n4}\| = \begin{cases} O_P\left(\left[n^{2d_2} L_*^2(n)\right] \vee \left[n^{2d_1 - 1} L_*^2(n)\right]\right), & d_1 > 1, \\ O_P\left(n^{2d_2} L_*^2(n)\right), & 1/2 < d_1 \leq 1, \end{cases} \tag{17}$$

where $L_*(\cdot) = \max\left\{|L_1(\cdot)|, |L_1^{\diamond}(\cdot)|, |L_2(\cdot)|\right\}$. Theorem 3.1 indicates that $\boldsymbol{V}_{n1}$ is the asymptotic leading term of $\boldsymbol{V}_n$, i.e.,

$$\frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_n = \frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_{n1} + o_P(1). \tag{18}$$

We next implement FPCA. Let $(\lambda_k(\boldsymbol{V}_n),\ \nu_k(\boldsymbol{V}_n))$, $k=1,\cdots,q_1$, be pairs of eigenvalues and eigenfunctions of the unnormalised sample variance operator $\boldsymbol{V}_n$. If the dimension q_1 is known a priori, we may construct the sample nonstationary $I(d_1)$ (or dominant) subspace via

$$\mathcal{H}_1(\boldsymbol{V}_n)=\mathcal{S}\left(\nu_1(\boldsymbol{V}_n),\cdots,\nu_{q_1}(\boldsymbol{V}_n)\right) \tag{19}$$

which is the span of the eigenfunctions $\nu_1(\boldsymbol{V}_n),\cdots,\nu_{q_1}(\boldsymbol{V}_n)$. Let $\boldsymbol{P}_1(\boldsymbol{V}_n)$ be the orthogonal projection on $\mathcal{H}_1(\boldsymbol{V}_n)$ and $\boldsymbol{P}_1^{\perp}(\boldsymbol{V}_n)=\boldsymbol{I}-\boldsymbol{P}_1(\boldsymbol{V}_n)$. Using Theorem 3.1, we readily have the following asymptotic result.

Theorem 3.2. *Suppose that the assumptions of Theorem 3.1 are satisfied. Then*

$$\boldsymbol{P}_1(\boldsymbol{V}_n)=\boldsymbol{P}_1+o_P(1),\quad \boldsymbol{P}_1^{\perp}(\boldsymbol{V}_n)=\boldsymbol{P}_1^{\perp}+o_P(1), \tag{20}$$

and for $k=1,\cdots,q_1$,

$$\left(\frac{1}{n^{2d_1}L_1^2(n)}\lambda_k(\boldsymbol{V}_n),\ \nu_k(\boldsymbol{V}_n)\right)\xrightarrow{P}\left(\lambda_k(\boldsymbol{V}),\ \nu_k(\boldsymbol{V})\right), \tag{21}$$

where $(\lambda_k(\boldsymbol{V}),\ \nu_k(\boldsymbol{V}))$, $k=1,\cdots,q_1$, are non-zero eigenvalues (arranged in the decreasing order and distinct with probability one) of $\boldsymbol{V}$ defined in (15) and the corresponding eigenfunctions.

Theorem 3.2 can be seen as an extension of Proposition 3.2 and Theorem 3.3 in Chang, Kim and Park (2016) from functional $I(1)$ processes to more general functional $I(d_1)$ processes. Without loss of generality, we define $\nu_k(\boldsymbol{V})=\mathsf{sign}(\langle\nu_k(\boldsymbol{V}_n),\nu_k(\boldsymbol{V})\rangle)\nu_k(\boldsymbol{V})$ so that $\nu_k(\boldsymbol{V}_n)$ and $\nu_k(\boldsymbol{V})$ have the same sign. The dominant subspace $\mathcal{H}_1$ may be generated as the span of the eigenfunctions $\nu_1(\boldsymbol{V}),\cdots,\nu_{q_1}(\boldsymbol{V})$. Hence $\mathcal{H}_1(\boldsymbol{V}_n)$ defined in (19) is a sensible approximation of $\mathcal{H}_1$. Note that although $\nu_k(\boldsymbol{V})$, $k=1,\cdots,q_1$, are *random* as functionals of $W_{d_1,\boldsymbol{\Omega}_1}(\cdot)$, the space spanned by them is non-random and uniquely determined as pointed out by Chang, Kim and Park (2016).

Motivated by Theorem 3.1 above, we may also define the normalised sample variance operator:

$$\overline{\boldsymbol{V}}_n=\frac{1}{n^{2d_1}L_1^2(n)}\sum_{t=1}^{n}X_t\otimes X_t,\quad\text{or}\quad \overline{\boldsymbol{V}}_n(z)=\frac{1}{n^{2d_1}L_1^2(n)}\sum_{t=1}^{n}\langle X_t,z\rangle X_t. \tag{22}$$

In practice, we can obtain the following sample variance (or covariance) function:

$$\overline{V}_n(u,v)=\frac{1}{n^{2d_1}L_1^2(n)}\sum_{t=1}^{n}X_t(u)X_t(v),\quad u,v\in\mathbb{S}, \tag{23}$$

which is connected to the normalised sample variance operator via $\overline{\mathbf{V}}_n(z)(u)=\int_{\mathbb{S}}\overline{V}_n(u,v)z(v)dv$, $z\in\mathcal{H}$. Note that $\overline{\boldsymbol{V}}_n$ is proportional to $\boldsymbol{V}_n$, indicating that the eigenfunctions via FPCA of $\overline{\boldsymbol{V}}_n$ are the same as those of $\boldsymbol{V}_n$. However, the latter does not contain any unknown quantity (say, d_1) so is preferred in practical implementation.

4. Estimation of q_1 and d_1

In this section, we introduce methods to estimate the dimension q_1 and the parameter d_1, which are important when the main interest lies in statistical inference.

4.1. Estimation of q_1

In practice, the dimension q_1 of the dominant subspace $\mathcal{H}_1$ is unknown, and needs to be estimated before implementing the FPCA technique introduced in Section 3. We estimate q_1 via a simple ratio criterion and then show its consistency. Recall that $\lambda_k(\boldsymbol{V}_n)$ is the k-th largest eigenvalue of $\boldsymbol{V}_n$. We estimate q_1 by

$$\widehat{q}_1 = \arg\min_{1\le k\le \overline{K}} \left|\frac{\lambda_{k+1}(\boldsymbol{V}_n)}{\lambda_k(\boldsymbol{V}_n)}\right|, \tag{24}$$

where $\overline{K}$ is a pre-specified positive integer and $0/0 = 1$. The above ratio criterion is easy to implement and a similar idea has been commonly used to estimate the number of latent factors in factor models (Lam and Yao, 2012, Ahn and Horenstein, 2013), to identify cointegrated components of nonstationary time series (Zhang, Robinson and Yao, 2019), and to determine the dimension of the dominant subspace for stationary long memory functional time series (Li, Robinson and Shang, 2020). Other dimension selection methods proposed in the literature on functional data analysis include the cumulative percentage of total variation (Horváth and Kokoszka, 2012), bootstrap (Hall and Vial, 2006) and the Akaike information criterion (Li, Wang and Carroll, 2013). In practical implementation of the ratio criterion, we may set $\lambda_k(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)$ as 0 if $|\lambda_k(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)| < \delta$ with δ being a pre-specified small positive number. Consequently, we can show that

$$\left|\frac{\lambda_{k+1}(\boldsymbol{V}_n)}{\lambda_k(\boldsymbol{V}_n)}\right| = \frac{|\lambda_{k+1}(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)|}{|\lambda_k(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)|} = 0/0 = 1, \tag{25}$$

if both $|\lambda_{k+1}(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)|$ and $|\lambda_k(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)|$ are smaller than δ. The following theorem shows that $\widehat{q}_1$ is weakly consistent, extending Proposition 4 in Li, Robinson and Shang (2020) from the stationary long-range dependence setting to nonstationary fractional integration.

Theorem 4.1. *Suppose that the assumptions of Theorem 3.2 are satisfied. Then we have* $\mathsf{P}(\widehat{q}_1 = q_1) \to 1$.

4.2. Local Whittle estimation of d_1

Semiparametric estimation of the memory parameter d_1 in fractionally integrated time series has received much attention. In this section, we extend the most commonly-used semiparametric local Whittle method to functional processes. The local Whittle estimation method is introduced by Künsch (1987), and its asymptotic properties including consistency and asymptotic normality are established by Robinson (1995, 2008) for stationary univariate and multivariate time series with memory parameter between $-1/2$ and $1/2$. These results are extended by Velasco (1999) to "Type I" fractionally integrated processes with memory parameter between $-1/2$ and 1, and further extended by Phillips and Shimotsu (2004) to "Type II" fractionally integrated processes with memory parameter exceeding $1/2$ and possibly larger than 1. Li, Robinson and Shang (2020) apply this method to estimate the memory parameter for stationary and long-range dependent functional time series and Li, Robinson and Shang (2021) further derive its asymptotic theory including consistency and asymptotic normality.

We now limit our attention to the following $I(d_1)$ functional time series model:

$$\nabla^{d_1} X_t = \eta_t I\{t \ge 1\} \quad \text{with} \quad \nabla = 1 - B, \tag{26}$$

where B denotes the backshift operator, $d_1 > 1/2$, $I\{\cdot\}$ is an indicator function and $(\eta_t : t \ge 1)$ is a stationary and short-range dependent functional process satisfying Assumption 2.2 in Section 2. Model

(26) extends the functional FARIMA model in Li, Robinson and Shang (2020) from the stationary region to the nonstationary one, and generalises the model in Phillips and Shimotsu (2004) and Shimotsu and Phillips (2005) from the classic univariate setting to the functional setting. Note that

$$X_t = \nabla^{-d_1}\eta_t I\{t \geq 1\} = \sum_{i=0}^{t-1}\beta_{i,d_1}B^i\eta_t = \sum_{i=0}^{t-1}\beta_{i,d_1}\eta_{t-i} = \sum_{i=1}^{t}\beta_{t-i,d_1}\eta_i, \tag{27}$$

where, by the Stirling's formula,

$$\beta_{i,d_1} = \frac{\Gamma(d_1+i)}{\Gamma(d_1)i!} = \frac{1}{\Gamma(d_1)}i^{d_1-1} + O(i^{d_1-2}) \tag{28}$$

with $\Gamma(\cdot)$ being the gamma function. Combining (27) and (28), it is easy to show that (26) falls within the model framework (1), so the theory and methodology developed in Sections 2 and 3 are applicable to X_t defined in (26). For notational simplicity, we let $\mathcal{H}_{1,n} := \mathcal{H}_1(\boldsymbol{V}_n)$ defined in (19), and $\boldsymbol{P}_{1,n} := \boldsymbol{P}_1(\boldsymbol{V}_n)$, the orthogonal projection on $\mathcal{H}_{1,n}$. For a real function $\nu \in \mathcal{H}$ such that $\|\boldsymbol{P}_1\nu\| > 0$, by (20) in Theorem 3.2, we readily have that

$$\|\boldsymbol{P}_{1,n}\nu - \boldsymbol{P}_1\nu\| = o_P(1). \tag{29}$$

Define the sequence of inner products:

$$x_t^\nu = \langle \boldsymbol{P}_1\nu, X_t\rangle = \sum_{i=1}^{t}\beta_{t-i,d_1}\langle \boldsymbol{P}_1\nu, \eta_i\rangle = \sum_{i=1}^{t}\beta_{t-i,d_1}\eta_i^\nu, \tag{30}$$

which is univariate nonstationary $I(d_1)$ with $d_1 > 1/2$. Furthermore, by (27) and Proposition 2.1, we readily obtain

$$\begin{aligned}\frac{1}{n^{d_1-1/2}}x_{\lfloor nr\rfloor}^\nu &= \frac{1}{n^{d_1-1/2}}\left\langle \boldsymbol{P}_1\nu, \boldsymbol{P}_1X_{\lfloor nr\rfloor} + \boldsymbol{P}_1^\perp X_{\lfloor nr\rfloor}\right\rangle \\ &= \frac{1}{n^{d_1-1/2}}\left\langle \boldsymbol{P}_1\nu, \boldsymbol{P}_1X_{\lfloor nr\rfloor}\right\rangle \\ &\Rightarrow w^\nu(r) := \frac{1}{\Gamma(d_1)}\left\langle \boldsymbol{P}_1\nu, W_{d_1,\boldsymbol{\Omega}_\star}(r)\right\rangle \end{aligned} \tag{31}$$

in $\mathcal{D}([0,1],\mathcal{R})$, where $W_{d_1,\boldsymbol{\Omega}_\star}(\cdot)$ is defined similarly to $W_{d_1,\boldsymbol{\Omega}_1}$ but with $\boldsymbol{\Omega}_1$ replaced by $\boldsymbol{\Omega}_\star = \boldsymbol{P}_1\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\boldsymbol{P}_1$.

Given that $(x_t^\nu : t \geq 1)$ is univariate $I(d_1)$ with $d_1 > 1/2$, we may estimate d_1 by local Whittle. Assume that $\eta_i^\nu = \langle \boldsymbol{P}_1\nu, \eta_i\rangle$ has a spectral density $f_\nu(\cdot)$ satisfying $f_\nu(\lambda) \sim G_\nu$ as $\lambda \to 0+$, where G_ν is a positive constant relying on choice of ν. This is a sensible condition as $(\eta_t : t \geq 1)$ is stationary and short-range dependent (see Assumption 2.2). Define the discrete Fourier transform and the periodogram of x_t^ν at the frequency $\lambda_j = 2\pi j/n$ as

$$w_x^\nu(\lambda_j) = \frac{1}{(2\pi n)^{1/2}}\sum_{t=1}^{n}x_t^\nu e^{it\lambda_j} \quad\text{and}\quad I_x^\nu(\lambda_j) = \left|w_x^\nu(\lambda_j)\right|^2. \tag{32}$$

Consider the following Gaussian objective function:

$$\mathrm{Q}(G,d) = \frac{1}{m}\sum_{j=1}^{m}\left[\log(G\lambda_j^{-2d}) + G^{-1}\lambda_j^{2d}I_x^\nu(\lambda_j)\right], \tag{33}$$

where m is a tuning parameter satisfying that $m \to \infty$ but $m = o(n)$. As in Robinson (1995), we may concentrate $\mathsf{Q}(G, d)$ in (33) with respect to G, and obtain the local Whittle estimate:

$$\widetilde{d}_1 = \arg\min_{d\in\Theta} R(d) = \arg\min_{d\in\Theta} \left\{ \log \widetilde{G}(d) - \frac{2d}{m} \sum_{j=1}^{m} \log \lambda_j \right\} \tag{34}$$

with

$$\widetilde{G}(d) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_x^{\nu}(\lambda_j),$$

where $\Theta = [\Delta_1, \Delta_2]$ with Δ_1 and Δ_2 chosen such that $-1/2 < \Delta_1 < \Delta_2 < \infty$. Unfortunately, the local Whittle estimation defined in (34) is practically infeasible as the orthogonal projection $\boldsymbol{P}_1$ (involved in the definition of x_t^{ν}) is unobservable. By (20) in Theorem 3.2, it is sensible to approximate $\boldsymbol{P}_1$ by its sample version $\boldsymbol{P}_{1,n}$, and thus we may expect that x_t^{ν} is reasonably close to $\overline{x}_t^{\nu} = \langle \boldsymbol{P}_{1,n}\nu, X_t \rangle$, which is an approximate $I(d_1)$ process. Following the local Whittle estimation procedure with x_t^{ν} replaced by $\overline{x}_t^{\nu}$, we can obtain a feasible estimator of d_1, denoted by $\overline{d}_1$. Similar to Theorems 3.1 and 3.2 in Phillips and Shimotsu (2004), the following theorem shows that the feasible local Whittle estimator is only consistent when $1/2 < d_1 \le 1$, and converges to unity when d_1 exceeds 1.

Theorem 4.2. *Suppose that the spectral density of η_i^{ν} satisfies that $f_{\nu}(\lambda) \sim G_{\nu}$ as $\lambda \to 0+$, it is differentiable in a small neighbourhood of the origin, and*

$$\frac{d}{d\lambda} \log f_{\nu}(\lambda) = O(1/\lambda) \quad \text{as} \quad \lambda \to 0+.$$

In addition, the tuning parameter m satisfies that $m \to \infty$ and $m = o(n)$.

(i) *For $1/2 < d_1 \le 1$, $\overline{d}_1 \xrightarrow{P} d_1$.*

(ii) *For $1 < d_1 \le \overline{c}$ with $1 < \overline{c} < \infty$, $\overline{d}_1 \xrightarrow{P} 1$.*

Theorem 4.2, together with Theorem 1(i) in Li, Robinson and Shang (2021), indicates that, to obtain a consistent estimated value of d_1, an integer-order differencing is needed to process the approximate coordinate process $(\overline{x}_t^{\nu} : t \ge 1)$ when $\overline{d}_1$ is close to one. Hence, we propose the following algorithm.

Algorithm 1. Local Whittle with integer-order differencing

1. Apply local Whittle using $\overline{x}_t^{\nu}$ to obtain the estimate $\overline{d}_1$.

2. If $\overline{d}_1 < 1 - \gamma$ with γ being a pre-determined small tuning parameter, we stop the algorithm and use $\overline{d}_1$ as the final estimate. Otherwise, take point-wise differences of $\overline{x}_t^{\nu}$ and obtain the local Whittle estimate $\overline{d}_1(1)$.

3. Repeat Step 2 k_0 times. Terminate the algorithm when $\overline{d}_1(k_0) < 1 - \gamma$ and obtain the final estimate: $\overline{d}_1(k_0) + k_0$.

It would be interesting to further study the asymptotic distribution theory of the developed local Whittle estimate. Velasco (1999) and Phillips and Shimotsu (2004) derive the estimation distribution

properties for the "Type I" and "Type II" univariate fractionally integrated processes, respectively. In fact, following Theorem 4.1 in Phillips and Shimotsu (2004), we may show the following asymptotic distributions for the infeasible local Whittle estimate $\widetilde{d}_1$:

$$m^{1/2}\left(\widetilde{d}_1 - d_1\right) \xrightarrow{D} \frac{1}{2}\mathsf{N}_1 \text{ for } 1/2 < d_1 < 3/4,$$

$$m^{1/2}\left(\widetilde{d}_1 - d_1\right) \xrightarrow{D} \frac{1}{2}\mathsf{N}_1 + C(d_1)\mathsf{N}_2^2 \text{ for } d_1 = 3/4,$$

$$m^{2-2d_1}\left(\widetilde{d}_1 - d_1\right) \xrightarrow{D} C(d_1)\mathsf{N}_2^2 \text{ for } 3/4 < d_1 < 1,$$

where N_1 and N_2 are two independent standard normal random variables and

$$C(d) = (2\pi)^{2d-2}[\Gamma(d)]^{-2}(2d-1)^{-3}(1-d).$$

When $d_1 = 1$, using Theorem 4.2 in Phillips and Shimotsu (2004), we may show that $m^{1/2}\left(\widetilde{d}_1 - d_1\right)$ converges in distribution to a mixed normal distribution. We conjecture this distribution property may also hold for the feasible local Whittle estimate $\overline{d}_1$. However, more technical assumptions would be required and the mathematical proofs would be quite involved. We will leave it in our future studies.

5. Monte-Carlo simulation

We next present a Monte-Carlo simulation study to assess numerical performance of the simple ratio criterion and local Whittle estimation in finite samples. The supplemental document contains extra simulation studies and two empirical applications.

Example 5.1. Following the simulation setting in Aue, Rice and Sönmez (2018) and Nielsen, Seo and Seong (2019), we let $v_\star^1, v_\star^2, \cdots, v_\star^{21}$ be the first 21 polynomial basis functions defined on $\mathbb{S} = [0,1]$, and generate functional time series via

$$X_t = \sum_{j=1}^{21} x_t^j v_\star^j = \sum_{j=1}^{3} x_t^j v_\star^j + \sum_{j=4}^{21} x_t^j v_\star^j = \boldsymbol{x}_t^N \boldsymbol{v}^N + \boldsymbol{x}_t^S \boldsymbol{v}^S, \tag{35}$$

where $\boldsymbol{v}^N = \left(v_\star^1, v_\star^2, v_\star^3\right)^\top$, $\boldsymbol{v}^S = \left(v_\star^4, v_\star^5, \ldots, v_\star^{21}\right)^\top$, $\boldsymbol{x}_t^N = \left(x_t^1, x_t^2, x_t^3\right)^\top$ is generated from a three-dimensional "Type II" fractionally integrated process:

$$\nabla^{d_1}\boldsymbol{x}_t^N = \boldsymbol{\eta}_t I\{t \geq 1\}, \tag{36}$$

where the memory parameter d_1 is set as 0.75, 1.0 or 1.25, and $(\boldsymbol{\eta}_t : t \geq 1)$ is independently generated by a three-dimensional normal distribution with mean zero and covariance matrix

$$\boldsymbol{\Omega}_3 = \begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix},$$

and $\boldsymbol{x}_t^S = \left(x_t^4, x_t^5, \cdots, x_t^{21}\right)^\top$ is generated from a stationary VAR(1) with dimension 18:

$$\boldsymbol{x}_t^S = \boldsymbol{A}_S \boldsymbol{x}_{t-1}^S + \boldsymbol{\xi}_t, \tag{37}$$

where $\boldsymbol{A}_S = (a_{ij})_{18\times 18}$ is a banded autoregressive matrix with a_{ij} independently generated from $U(-0.3, 0.3)$ when $|i-j| \leq 3$ and $a_{ij} = 0$ when $|i-j| > 3$ or $\boldsymbol{A}_S$ is a diagonal matrix with diagonal elements drawn from a $U(-0.5, 0.5)$, and $(\boldsymbol{\xi}_t : t \in \mathbb{Z})$ is independently generated by an 18-dimensional normal distribution with mean zero and identity covariance matrix.

Models (35)–(37) above show that the dimension q_1 of the nonstationary dominant sub-space is 3. As in Aue, Rice and Sönmez (2018) and Nielsen, Seo and Seong (2019), we permute the selected 21 basis functions so that the shape of the nonstationary sub-space would not be affected by particular shape and ordering of the polynomial basis functions. In Table 1 below, we report the number of retained eigenfunctions for 1000 replications with the ratio criterion proposed in Section 4.1. In nearly all the settings, there are more than 70% of times that the ratio criterion could correctly select the true dimension $q_1 = 3$. The estimation accuracy improves as the sample size n increases.

		$\boldsymbol{A}_S$="diag"				$\boldsymbol{A}_S$="band"		
n	d_1	$\widehat{q}_1 = 2$	$\widehat{q}_1 = 3$	$\widehat{q}_1 = 4$	$\widehat{q}_1 = 6$	$\widehat{q}_1 = 2$	$\widehat{q}_1 = 3$	$\widehat{q}_1 = 4$
500	0.75	451	547	1	1	456	544	
	1.00	237	760	3		239	759	2
	1.25	286	703	11		270	719	11
1000	0.75	264	736			280	720	
	1.00	152	846	2		158	839	3
	1.25	246	739	15		244	750	6
2000	0.75	114	886			128	872	
	1.00	111	886	3		111	882	7
	1.25	220	771	9		224	759	17

Table 1.. The number of replications for $\widehat{q}_1$ estimated by the ratio criterion for Example 5.1 (with $q_1 = 3$ as the true dimension)

To assess the sensitivity of the proposed local Whittle estimation to different choices of the function ν, we consider the following three ways to construct the coordinate process: (i) $\nu = \nu_1(\boldsymbol{V}_n)$ defined in Section 3 and consequently $\overline{x}_t^{\nu} = \langle \boldsymbol{P}_{1,n}\nu, X_t\rangle = \langle \nu_1(\boldsymbol{V}_n), X_t\rangle$ which becomes the approximation of the first score; (ii) $\nu = \nu_\star^1$, the first-order polynomial basis function, and $\overline{x}_t^{\nu} = \langle \boldsymbol{P}_{1,n}\nu_\star^1, X_t\rangle$; and (iii) use the coordinate process $\langle \nu_\star^1, X_t\rangle$ in Algorithm 1. In Tables 2–4, we present mean, median, bias, variance and MSE of the local Whittle estimates over 1000 replications (an integer-order differencing may be needed in the estimation algorithm) for the three cases, respectively. In general, the bias (in absolute value), variance and MSE of the estimates decrease as the sample size n increases. The estimation bias tends to be negative when d_1 is either 1 or 1.25. Furthermore, the local Whittle estimation performance does not differ significantly over Cases (i)–(iii). Meanwhile, the local Whittle estimation and the associated algorithm proposed in Section 4.2 rely on selection of two tuning parameters: m defined as in (33) and γ that determines whether an integer-order differencing is needed. For simplicity, we consider the following ad-hoc tuning parameter selection: $m = \lfloor n^{\iota} \rfloor + 1$ with $\iota = 0.65$ satisfying the restriction in Theorem 4.2, and $\gamma = c/\log(m)$ with $c = 0.4$, which slowly tends to zero as the sample size increases. In Appendix C of the supplement, we further perform the estimation sensitivity analysis for different values of ι and c.

A_S	d_1	n	Mean	Median	Bias	Variance	MSE
diag	0.75	500	0.7907	0.7965	0.0407	0.0049	0.0066
		1000	0.7948	0.7957	0.0448	0.0032	0.0052
		2000	0.7924	0.7931	0.0424	0.0022	0.0040
	1.0	500	0.9584	0.9584	-0.0416	0.0055	0.0073
		1000	0.9703	0.9726	-0.0297	0.0036	0.0044
		2000	0.9816	0.9838	-0.0184	0.0020	0.0023
	1.25	500	1.1695	1.1702	-0.0805	0.0058	0.0123
		1000	1.1902	1.1920	-0.0598	0.0035	0.0071
		2000	1.2079	1.2095	-0.0421	0.0019	0.0037
band	0.75	500	0.7886	0.7917	0.0386	0.0046	0.0061
		1000	0.7937	0.7962	0.0437	0.0033	0.0052
		2000	0.7924	0.7929	0.0424	0.0023	0.0040
	1.0	500	0.9555	0.9538	-0.0445	0.0055	0.0075
		1000	0.9697	0.9721	-0.0303	0.0035	0.0044
		2000	0.9817	0.9849	-0.0183	0.0021	0.0024
	1.25	500	1.1671	1.1688	-0.0829	0.0057	0.0126
		1000	1.1892	1.1896	-0.0608	0.0035	0.0072
		2000	1.2081	1.2093	-0.0419	0.0019	0.0036

Table 2.. Mean, median, bias, variance and MSE of the local Whittle estimates with $\bar{x}_t^{\nu} = \langle \nu_1(\boldsymbol{V}_n), X_t \rangle$ over 1000 replications

A_S	d_1	n	Mean	Median	Bias	Variance	MSE
diag	0.75	500	0.7776	0.7822	0.0276	0.0054	0.0061
		1000	0.7841	0.7854	0.0341	0.0037	0.0048
		2000	0.7852	0.7866	0.0352	0.0024	0.0036
	1.0	500	0.9518	0.9555	-0.0482	0.0059	0.0082
		1000	0.9653	0.9678	-0.0347	0.0036	0.0048
		2000	0.9788	0.9812	-0.0212	0.0021	0.0026
	1.25	500	1.1643	1.1690	-0.0857	0.0064	0.0138
		1000	1.1873	1.1885	-0.0627	0.0037	0.0076
		2000	1.2058	1.2066	-0.0442	0.0020	0.0040
band	0.75	500	0.7735	0.7755	0.0235	0.0050	0.0056
		1000	0.7844	0.7882	0.0344	0.0036	0.0048
		2000	0.7851	0.7860	0.0351	0.0024	0.0036
	1.0	500	0.9475	0.9443	-0.0525	0.0053	0.0081
		1000	0.9652	0.9674	-0.0348	0.0038	0.0050

Continued on next page

$\boldsymbol{A}_S$	d_1	n	Mean	Median	Bias	Variance	MSE
		2000	0.9779	0.9814	-0.0221	0.0021	0.0026
	1.25	500	1.1620	1.1641	-0.0880	0.0055	0.0132
		1000	1.1867	1.1882	-0.0633	0.0036	0.0076
		2000	1.2052	1.2082	-0.0448	0.0020	0.0040

Table 3.: Mean, median, bias, variance and MSE of the local Whittle estimates with $\overline{x}_t^{\gamma} = \langle \boldsymbol{P}_{1,n} v_{\star}^1, X_t \rangle$ over 1000 replications

$\boldsymbol{A}_S$	d_1	n	Mean	Median	Bias	Variance	MSE
diag	0.75	500	0.7783	0.7827	0.0283	0.0054	0.0062
		1000	0.7841	0.7853	0.0341	0.0036	0.0048
		2000	0.7853	0.7861	0.0353	0.0024	0.0036
	1.0	500	0.9518	0.9552	-0.0482	0.0059	0.0082
		1000	0.9653	0.9677	-0.0347	0.0036	0.0048
		2000	0.9788	0.9812	-0.0212	0.0021	0.0026
	1.25	500	1.1644	1.1690	-0.0856	0.0064	0.0138
		1000	1.1871	1.1884	-0.0629	0.0038	0.0077
		2000	1.2058	1.2066	-0.0442	0.0020	0.0040
band	0.75	500	0.7740	0.7758	0.0240	0.0048	0.0054
		1000	0.7845	0.7878	0.0345	0.0036	0.0048
		2000	0.7850	0.7856	0.0350	0.0024	0.0036
	1.0	500	0.9476	0.9439	-0.0524	0.0053	0.0081
		1000	0.9652	0.9674	-0.0348	0.0038	0.0050
		2000	0.9779	0.9814	-0.0221	0.0021	0.0026
	1.25	500	1.1619	1.1633	-0.0881	0.0055	0.0132
		1000	1.1866	1.1879	-0.0634	0.0036	0.0076
		2000	1.2051	1.2081	-0.0449	0.0020	0.0040

Table 4.: Mean, median, bias, variance and MSE of the local Whittle estimates using the coordinate process $\langle v_{\star}^1, X_t \rangle$ over 1000 replications

6. Conclusions

In this paper, we have extended the fractionally integrated time series model from the classic univariate or finite-dimensional setting to the infinite-dimensional functional setting. The proposed nonstationary functional time series model framework is general, covering the functional unit root as a special case. We project the functional process onto a finite number of sub-spaces with varying strength of nonstationarity. The primary interest lies in the nonstationary dominant sub-space $\mathcal{H}_1$ on which the projection of the original nonstationary functional time series retains most of the sample information and results in the strongest signal. Under some technical conditions, we derive a weak convergence result for the projection of the fractionally integrated functional process onto the sub-space $\mathcal{H}_1$, extending some existing results developed by Marinucci and Robinson (2000) and Berkes, Horváth and Rice (2013). Applying the classic FPCA to the sample variance operator, we obtain the eigenvalues and eigenfunctions which span a sample version of the dominant sub-space $\mathcal{H}_1$. The dimension of $\mathcal{H}_1$ is consistently estimated by an easy-to-implement ratio criterion and the relevant memory parameter is estimated by a semiparametric local Whittle method. The Monte-Carlo simulation study shows that the proposed estimation techniques have satisfactory finite-sample performance.

Appendix A: Proofs of the asymptotic results

In this appendix, we give the detailed proofs of Proposition 2.1, Theorems 3.1, 3.2 and 4.1, and sketch the proof of Theorem 4.2. Throughout the proof, we let C denote a generic positive constant whose value may change from line to line.

Proof of Proposition 2.1. For $t = \lfloor nr \rfloor$ with $0 \le r \le 1$, we note that

$$\boldsymbol{P}_1 X_{\lfloor nr \rfloor} = \sum_{j=1}^{s_0} \boldsymbol{P}_1 X_{\lfloor nr \rfloor, j} = \boldsymbol{P}_1 X^*_{\lfloor nr \rfloor, 1} + \boldsymbol{P}_1 X^\dagger_{\lfloor nr \rfloor, 1} + \boldsymbol{P}_1 X_{\lfloor nr \rfloor, -1}, \tag{A.1}$$

where $X_{t,j}$ is defined in (6), $X^*_{t,1}$ and $X^\dagger_{t,1}$ are defined as in (7) and $X_{t,-1}$ is defined in (8). Let

$$\overline{X}^*_{\lfloor nr \rfloor, 1} = \frac{1}{n^{d_1 - 1/2} L_1(n)} X^*_{\lfloor nr \rfloor, 1}.$$

In order to prove (10), we only need to show that as $n \to \infty$,

$$\left\langle v, \boldsymbol{P}_1 \overline{X}^*_{\lfloor nr \rfloor, 1} \right\rangle \Rightarrow \left\langle v, W_{d_1, \boldsymbol{\Omega}_1}(r) \right\rangle \quad \text{in } \mathcal{D}([0,1], \mathcal{R}), \tag{A.2}$$

$$\sup_{0 \le r \le 1} \left| \left\langle v, \boldsymbol{P}_1 X^\dagger_{\lfloor nr \rfloor, 1} \right\rangle \right| = o_P\left(n^{d_1 - 1/2} L_1(n) \right), \tag{A.3}$$

$$\sup_{0 \le r \le 1} \left| \left\langle v, \boldsymbol{P}_1 X_{\lfloor nr \rfloor, -1} \right\rangle \right| = o_P\left(n^{d_1 - 1/2} L_1(n) \right), \tag{A.4}$$

where $v \in \mathcal{H}_1$ is arbitrarily chosen.

Without loss of generality, we only consider the case of $\lfloor nr \rfloor \ge 2$ in the remaining proof. For $t \ge 2$, by (9), we readily have that

$$\begin{aligned} \left\langle v, \boldsymbol{P}_1 \overline{X}^*_{t,1} \right\rangle &= \frac{1}{n^{d_1 - 1/2} L_1(n)} \left(\sum_{i=1}^{t-1} \psi_{t-i,1} \left\langle v, \boldsymbol{P}_1 \boldsymbol{B}_{1,A} \varepsilon_i \right\rangle + \left\langle v, \boldsymbol{P}_1 \boldsymbol{A} \varepsilon_t \right\rangle \right) \\ &=: \frac{1}{n^{d_1 - 1/2} L_1(n)} \sum_{i=1}^{t-1} \psi_{t-i,1} \xi^*_{i,v} + \frac{1}{n^{d_1 - 1/2} L_1(n)} \xi^\dagger_{t,v} \end{aligned}$$

for $v \in \mathcal{H}_1$. By Assumptions 2.2 and 2.3, $(\xi^*_{i,v} : i \geq 1)$ is a sequence of i.i.d. random variables with mean zero, positive variance and $\mathsf{E}|\xi^*_{t,v}|^p < \infty$, which together with Assumption 2.1, implies that Assumptions A and B in Marinucci and Robinson (2000) are satisfied. Then, using Theorem 1 in Marinucci and Robinson (2000), (A.2) can be proved if we can show that

$$\max_{1\leq t\leq n} \left|\xi^\dagger_{t,v}\right| = o_P\left(n^{d_1-1/2}L_1(n)\right). \tag{A.5}$$

By Bonferroni's and Markov's inequalities, we readily have for any $\epsilon > 0$

$$\begin{aligned}
\mathsf{P}\left(\max_{1\leq t\leq n}\left|\xi^\dagger_{t,v}\right| \geq \epsilon n^{d_1-1/2}L_1(n)\right) &\leq \sum_{t=1}^{n}\mathsf{P}\left(\left|\xi^\dagger_{t,v}\right| \geq \epsilon n^{d_1-1/2}L_1(n)\right) \\
&\leq \epsilon^{-p}n^{-p(d_1-1/2)}L_1^{-p}(n)\sum_{t=1}^{n}\mathsf{E}\left|\xi^\dagger_{t,v}\right|^p \\
&= O\left(n^{1-p(d_1-1/2)}L_1^{-p}(n)\right) = o(1),
\end{aligned} \tag{A.6}$$

as $1 - p(d_1 - 1/2) < 0$ by noting that $p > 2/(2d_1 - 1)$ in Assumption 2.2. So (A.5) is proved and then (A.2) is proved.

For any t, by Rosenthal's inequalities and Assumption 2.2, we may show that

$$\begin{aligned}
\mathsf{E}\,|\langle v, \widetilde{\eta}_t\rangle|^p &\leq C\left\{\left(\sum_{i=0}^{\infty}\mathsf{E}\langle v, \widetilde{\boldsymbol{A}}_i\varepsilon_{t-i}\rangle^2\right)^{p/2} + \sum_{i=0}^{\infty}\mathsf{E}\left|\langle v, \widetilde{\boldsymbol{A}}_i\varepsilon_{t-i}\rangle\right|^p\right\} \\
&\leq C\left\{\left[\sum_{i=0}^{\infty}\left\|\widetilde{\boldsymbol{A}}_i\right\|^2\left(\mathsf{E}\|\varepsilon_{t-i}\|^2\right)\right]^{p/2} + \sum_{i=0}^{\infty}\left\|\widetilde{\boldsymbol{A}}_i\right\|^p\left(\mathsf{E}\|\varepsilon_{t-i}\|^p\right)\right\} \\
&\leq C\left[\left(\sum_{i=0}^{\infty}\left\|\widetilde{\boldsymbol{A}}_i\right\|^2\right)^{p/2} + \sum_{i=0}^{\infty}\left\|\widetilde{\boldsymbol{A}}_i\right\|^p\right]\left(\mathsf{E}\|\varepsilon_0\|^p\right) < \infty,
\end{aligned} \tag{A.7}$$

noting that

$$\sum_{i=0}^{\infty}\left\|\widetilde{\boldsymbol{A}}_i\right\|^2 \leq \sum_{i=0}^{\infty}\left(\sum_{j=i+1}^{\infty}\|\boldsymbol{A}_j\|\right)^2 \leq C\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\|\boldsymbol{A}_j\| \leq C\sum_{i=0}^{\infty} i\|\boldsymbol{A}_i\| < \infty$$

and thus $\sum_{i=0}^{\infty}\left\|\widetilde{\boldsymbol{A}}_i\right\|^p < \infty$ for $p > 2$. Then, using (A.7) and the condition $p > \max\{2, 2/(2d_1 - 1)\}$, as in (A.6), we readily have

$$\max_{1\leq t\leq n}|\langle v, \widetilde{\eta}_t\rangle| = O_P\left(n^{1/p}\right) = o_P\left(n^{d_1-1/2}L_1(n)\right), \tag{A.8}$$

and furthermore, by Assumption 2.1,

$$\begin{aligned}
\max_{2\leq t\leq n}\left|\sum_{i=1}^{t-1}\left(\psi_{t-i,1} - \psi_{t-i-1,1}\right)\langle v, \boldsymbol{B}_1\widetilde{\eta}_i\rangle\right| &\leq O_P\left(n^{1/p}\right)\max_{2\leq t\leq n}\left(\sum_{i=1}^{t-1}\left|\psi_{t-i,1} - \psi_{t-i-1,1}\right|\right) \\
&= O_P\left(n^{1/p}\sum_{i=1}^{n} i^{d_1-2}\left[L_1^\diamond(i) + L_1(i)\right]\right) \\
&= o_P\left(n^{d_1-1/2}L_1(n)\right).
\end{aligned} \tag{A.9}$$

Meanwhile, it is obvious that

$$|\langle v, \widetilde{\eta}_0 \rangle| = O_P(1) = o_P\left(n^{d_1-1/2} L_1(n)\right), \tag{A.10}$$

and as in (A.9),

$$\begin{aligned}\max_{2\le t\le n}\left|\sum_{i=1}^{t-1}\left(\psi_{t-i,1}-\psi_{t-i-1,1}\right)\langle v, \boldsymbol{B}_1\widetilde{\eta}_0\rangle\right| &= O_P(1)\max_{2\le t\le n}\left(\sum_{i=1}^{t-1}\left|\psi_{t-i,1}-\psi_{t-i-1,1}\right|\right)\\ &= O_P\left(\sum_{i=1}^{n} i^{d_1-2}\left[L_1^{\diamond}(i)+L_1(i)\right]\right)\\ &= o_P\left(n^{d_1-1/2}L_1(n)\right). \end{aligned}\tag{A.11}$$

With (A.8)–(A.11), we can prove (A.3) uniformly for $0 \le r \le 1$.

Finally, combining the arguments in the proofs of (A.2) and (A.3), we can similarly show that

$$\max_{1\le t\le n}\left|\left\langle v, \boldsymbol{P}_1 X_{t,j}\right\rangle\right| = o_P\left(n^{d_1-1/2}L_1(n)\right), \quad j = 2, \cdots, s_0, \tag{A.12}$$

as $d_1 > d_2 > \cdots > d_{s_0}$, indicating that (A.4) holds. We have completed the proof of Proposition 2.1. □

Proof of Theorem 3.1. By Proposition 2.1 and the continuous mapping theorem in Billingsley (1968), we readily have (15). We next give the proofs of $\|\boldsymbol{V}_{nk}\| = o_P\left(n^{2d_1}L_1^2(n)\right)$ for $k = 2, 3, 4$.

We first prove $\|\boldsymbol{V}_{n2}\| = o_P\left(n^{2d_1}L_1^2(n)\right)$. By the definition of $\boldsymbol{V}_{n2}$ and Proposition 2.1, we have

$$\|\boldsymbol{V}_{n2}\| \le \sum_{t=1}^{n}\|\boldsymbol{P}_1 X_t\|\left\|\boldsymbol{P}_1^{\perp}X_t\right\| = O_P\left(n^{d_1-1/2}L_1(n)\right)\cdot\sum_{t=1}^{n}\left\|\boldsymbol{P}_1^{\perp}X_t\right\|. \tag{A.13}$$

By (7) and (8) and noting that $\boldsymbol{P}_1^{\perp}\boldsymbol{B}_{1,A} = 0$, we have

$$\boldsymbol{P}_1^{\perp}X_t = \sum_{i=1}^{t-1}\boldsymbol{P}_1^{\perp}\left(\boldsymbol{\Psi}_{t-i,1}-\boldsymbol{\Psi}_{t-i-1,1}\right)\left(\widetilde{\eta}_0-\widetilde{\eta}_i\right)+\boldsymbol{P}_1^{\perp}\left(\widetilde{\eta}_0-\widetilde{\eta}_t\right)+\boldsymbol{P}_1^{\perp}\boldsymbol{A}\varepsilon_t+\boldsymbol{P}_1^{\perp}X_{t,-1}. \tag{A.14}$$

By Assumption 2.2, similar to (A.7), we may show that $\mathsf{E}\|\widetilde{\eta}_t\|^2 = O(1)$. This indicates that

$$\begin{aligned}&\sum_{t=2}^{n}\mathsf{E}\left(\left\|\sum_{i=1}^{t-1}\boldsymbol{P}_1^{\perp}\left(\boldsymbol{\Psi}_{t-i,1}-\boldsymbol{\Psi}_{t-i-1,1}\right)\left(\widetilde{\eta}_0-\widetilde{\eta}_i\right)+\boldsymbol{P}_1^{\perp}\left(\widetilde{\eta}_0-\widetilde{\eta}_t\right)\right\|\right)\\ &= O\left(\sum_{t=2}^{n}\sum_{i=1}^{t-1}(t-i)^{d_1-2}\left[L_1^{\diamond}(t-i)+L_1(t-i)\right]\right)+O(n)\\ &= \begin{cases} O(n), & 1/2 < d_1 < 1,\\ o(n^{1+\zeta_1}), & d_1 = 1,\\ O\left(n^{d_1}L_{\star}(n)\right), & d_1 > 1,\end{cases}\end{aligned}\tag{A.15}$$

where $0 < \zeta_1 < d_1 - 1/2$ and $L_{\star}(\cdot) = \max\left\{|L_1(\cdot)|, |L_1^{\diamond}(\cdot)|\right\}$. It is straightforward to prove that

$$\sum_{t=1}^{n}\mathsf{E}\left(\left\|\boldsymbol{P}_1^{\perp}\boldsymbol{A}\varepsilon_t\right\|\right) = O(n). \tag{A.16}$$

By (A.12) and the assumption $d_1 > d_2 > \cdots > d_{s_0}$, we have

$$\sum_{t=1}^{n} \left\| \boldsymbol{P}_1^{\perp} X_{t,-1} \right\| = o_P\left(n^{d_1+1/2} L_1(n)\right). \tag{A.17}$$

With (A.13)–(A.17), we complete the proof of $\|\boldsymbol{V}_{n2}\| = o_P\left(n^{2d_1} L_1^2(n)\right)$. The proof of $\|\boldsymbol{V}_{n3}\| = o_P\left(n^{2d_1} L_1^2(n)\right)$ can be done in exactly the same way.

We finally turn to the proof of $\|\boldsymbol{V}_{n4}\| = o_P\left(n^{2d_1} L_1^2(n)\right)$. By the definition of $\boldsymbol{V}_{n4}$, we have

$$\|\boldsymbol{V}_{n4}\| \le \sum_{t=1}^{n} \left\| \boldsymbol{P}_1^{\perp} X_t \right\|^2. \tag{A.18}$$

Similar to (A.15), we can show that

$$\begin{aligned}
&\sum_{t=2}^{n} \mathsf{E}\left(\left\| \sum_{i=1}^{t-1} \boldsymbol{P}_1^{\perp} \left(\boldsymbol{\Psi}_{t-i,1} - \boldsymbol{\Psi}_{t-i-1,1} \right) \left(\widetilde{\eta}_0 - \widetilde{\eta}_i \right) + \boldsymbol{P}_1^{\perp} \left(\widetilde{\eta}_0 - \widetilde{\eta}_t \right) + \boldsymbol{P}_1^{\perp} \mathbf{A} \varepsilon_t \right\|^2 \right) \\
&\le 2 \sum_{t=2}^{n} \mathsf{E} \left\| \sum_{i=1}^{t-1} \boldsymbol{P}_1^{\perp} \left(\boldsymbol{\Psi}_{t-i,1} - \boldsymbol{\Psi}_{t-i-1,1} \right) \left(\widetilde{\eta}_0 - \widetilde{\eta}_i \right) \right\|^2 \\
&\quad + 2 \sum_{t=2}^{n} \mathsf{E} \left\| \boldsymbol{P}_1^{\perp} \left(\widetilde{\eta}_0 - \widetilde{\eta}_t \right) + \boldsymbol{P}_1^{\perp} \boldsymbol{A} \varepsilon_t \right\|^2 \\
&= O\left(\sum_{t=2}^{n} \left\{ \sum_{i=1}^{t-1} (t-i)^{d_1-2} \left[L_1(t-i) + L_1^{\diamond}(t-i) \right] \right\}^2 \right) + O(n) \\
&= \begin{cases} O(n), & 1/2 < d_1 < 1, \\ o(n^{1+\zeta_2}), & d_1 = 1, \\ O\left(n^{2d_1-1} L_*^2(n)\right), & d_1 > 1, \end{cases}
\end{aligned} \tag{A.19}$$

where $0 < \zeta_2 < 2d_1 - 1$. On the other hand, as in (A.17), we have

$$\sum_{t=1}^{n} \left\| \boldsymbol{P}_1^{\perp} X_{t,-1} \right\|^2 = o_P\left(n^{2d_1} L_1^2(n)\right). \tag{A.20}$$

Using (A.14) and (A.18)–(A.20), we can complete the proof of $\|\boldsymbol{V}_{n4}\| = o_P\left(n^{2d_1} L_1^2(n)\right)$. □

Proof of Theorem 3.2. Let

$$\boldsymbol{V}_{n1}^{\diamond} = \sum_{t=1}^{n} \boldsymbol{P}_1 X_{t,1}^{\diamond} \otimes X_{t,1}^{\diamond} \boldsymbol{P}_1 \quad \text{with} \quad X_{t,1}^{\diamond} = \sum_{i=1}^{t-1} \psi_{t-i,1} \boldsymbol{B}_{1,A} \varepsilon_i.$$

From (18) and the arguments in the proofs of Proposition 2.1 and Theorem 3.1, we readily have that

$$\frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_n = \frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_{n1} + o_P(1) = \frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_{n1}^{\diamond} + o_P(1). \tag{A.21}$$

Note that the dominant $I(d_1)$ subspace $\mathcal{H}_1$ is spanned by the q_1 eigenvectors corresponding to the non-zero eigenvalues of $\frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_{n1}^{\diamond}$. By the definition of $\boldsymbol{P}_1(\boldsymbol{V}_n)$ and (A.21), we prove the first assertion in (20), which then together with the definitions $\boldsymbol{P}_1^{\perp}(\boldsymbol{V}_n) = \boldsymbol{I} - \boldsymbol{P}_1(\boldsymbol{V}_n)$ and $\boldsymbol{P}_1^{\perp} = \boldsymbol{I} - \boldsymbol{P}_1$, leads to the second assertion in (20).

We next turn to the proof of (21). By (A.21), invoking the classic Skorokhod representation theorem as in Chang, Kim and Park (2016), we may show that

$$\left\| \frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_n - \boldsymbol{V} \right\| = o_P(1),$$

which together with Lemma B.1 in Appendix B of the supplement, implies that (21) holds. □

Proof of Theorem 4.1. By (A.21), Assumption 2.3 and Lemma B.1 in the supplement, we may show that there exists a positive constant χ_0 such that

$$\mathsf{P}\left(\left| \frac{\lambda_{k+1}(\boldsymbol{V}_n)}{\lambda_k(\boldsymbol{V}_n)} \right| = \left| \frac{\lambda_{k+1}(\boldsymbol{V}_n)/\left[n^{2d_1} L_1^2(n)\right]}{\lambda_k(\boldsymbol{V}_n)/\left[n^{2d_1} L_1^2(n)\right]} \right| > \chi_0 \right) \to 1, \quad k = 1, \cdots, q_1 - 1. \tag{A.22}$$

On the other hand, for $k = q_1 + 1, \cdots, \overline{K}$, the k-th largest eigenvalue of $\frac{1}{n^{2d_1} L_1^2(n)} \boldsymbol{V}_{n1}^{\diamond}$ (defined in the proof of Theorem 3.2) is zero by Assumption 2.3. Furthermore, by (A.21) and Lemma B.1, we can prove that

$$\lambda_k(\boldsymbol{V}_n) = o_P\left(n^{2d_1} L_1^2(n)\right), \quad k = q_1 + 1, \cdots, \overline{K},$$

which together with (21), indicates that

$$\mathsf{P}\left(\left| \frac{\lambda_{k+1}(\boldsymbol{V}_n)}{\lambda_1(\boldsymbol{V}_n)} \right| \geq \delta \right) \to 0, \tag{A.23}$$

where $\delta > 0$ is a very small constant. By (25) and (A.23), we may show that

$$\mathsf{P}\left(\left| \frac{\lambda_{k+1}(\boldsymbol{V}_n)}{\lambda_k(\boldsymbol{V}_n)} \right| = \left| \frac{\lambda_{k+1}(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)}{\lambda_k(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)} \right| = \frac{0}{0} = 1 \right) \to 1 \tag{A.24}$$

for $k = q_1 + 1, \cdots, \overline{K}$, and

$$\left| \frac{\lambda_{q_1+1}(\boldsymbol{V}_n)}{\lambda_{q_1}(\boldsymbol{V}_n)} \right| = \left| \frac{\lambda_{q_1+1}(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)}{\lambda_{q_1}(\boldsymbol{V}_n)/\lambda_1(\boldsymbol{V}_n)} \right| = o_P(1). \tag{A.25}$$

We then prove Theorem 4.1 combining (A.22), (A.24) and (A.25). □

Proof of Theorem 4.2. The main idea used in this proof is similar to that in Robinson (1995) and Phillips and Shimotsu (2004). In fact, using Theorems 3.1 and 3.2 in Phillips and Shimotsu (2004), we may show that the infeasible local Whittle estimator $\widetilde{d}_1$ is consistent when $1/2 < d_1 \leq 1$, but is biased and convergent to unity when d_1 exceeds 1. Hence, the main step in the proof is to show that replacement of x_t^{ν} by $\overline{x}_t^{\nu}$ in the local Whittle estimation has an asymptotically negligible effect in proving the consistency property. As the detailed proof is tedious, we provide it in Appendix B of the supplement. □

Acknowledgments

The authors would like to thank two reviewers for the comments, which helped to improve the paper.

Funding

The first author is partially supported by the National Natural Science Foundation of China (No. 72033002).

Supplementary Material

The supplemental document contains some technical lemmas together with their proofs, the detailed proof of Theorem 4.2, additional simulation results and empirical applications.

References

AHN, S. and HORENSTEIN, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81** 1203-1227.

AUE, A., RICE, G. and SÖNMEZ, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society B* **80** 509-529.

BATHIA, N., YAO, Q. and ZIEGELMANN, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics* **38** 3352-3386.

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

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

BERKES, I., HORVÁTH, L. and RICE, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and Their Applications* **123** 385-403.

BILLINGSLEY, P. (1968). *Convergence of Probability Measure*. Wiley, New York.

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

CHANG, Y., KIM, C. S. and PARK, J. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics* **192** 152-167.

CHEN, W. W. and HURVICH, C. (2006). Semiparametric estimation of fractional cointegrating subspaces. *The Annals of Statistics* **34** 2939-2979.

ENGLE, R. and GRANGER, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* **55** 251-276.

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.

FRANCHI, M. and PARUOLO, P. (2020). Cointegration in functional autoregressive processes. *Econometric Theory* **36** 803-839.

GOHBERG, I., GOLDBERG, S. and KAASHOEK, M. (1990). *Classes of Linear Operators* **Vol. I**. Birkhäuser, Basel; Boston.

HALL, P. and VIAL, C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society Series B* **68** 689-705.

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38** 1845-1884.

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.

HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics* **179** 66-82.

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

JOHANSEN, S. and NIELSEN, M. O. (2012). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory* **28** 671-679.

KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings First World Congress of the Bernoulli Society* (Y. A. PROHOROV and V. V. SAZONOV, eds.) 67-74. VNU Science Press, Utrecht.

KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P. and SHIN, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **54** 159-178.

LAM, C. and YAO, Q. (2012). Factor modelling for high-dimensional time series: Inference for the number of factor. *The Annals of Statistics* **40** 694-726.

LI, D., ROBINSON, P. M. and SHANG, H. L. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association: Theory and Methods* **115** 957-971.

LI, D., ROBINSON, P. M. and SHANG, H. L. (2021). Local Whittle estimation of long range dependence for functional time series. *Journal of Time Series Analysis* **in press**.

LI, Y., WANG, N. and CARROLL, R. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association: Theory and Methods* **108** 1284-1294.

MARINUCCI, D. and ROBINSON, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* **86** 103-120.

NIELSEN, M., SEO, W. and SEONG, D. (2019). Inference on the dimension of the nonstationary subspace in functional time series Technical Report, Queen's Economics Department Working Paper No. 1420.

PHILLIPS, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica* **63** 1023-1078.

PHILLIPS, P. C. B. and SHIMOTSU, K. (2004). Local Whittle estimation in nonstationary and unit root cases. *The Annals of Statistics* **32** 656-692.

PHILLIPS, P. C. B. and SOLO, V. (1992). Asymptotics for linear processes. *The Annals of Statistics* **20** 971-1001.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York.

ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23** 1630-1661.

ROBINSON, P. M. (2005). The distance between rival nonstationary fractional processes. *Journal of Econometrics* **128** 283-300.

ROBINSON, P. M. (2008). Multiple local whittle estimation in stationary systems. *The Annals of Statistics* **36** 2508-2530.

ROBINSON, P. M. and HUALDE, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71** 1727-1766.

ROBINSON, P. M. and MARINUCCI, D. (2001). Narrow-banded analysis of nonstationary processes. *The Annals of Statistics* **29** 947-986.

SHIMOTSU, K. and PHILLIPS, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics* **33** 1890-1933.

VELASCO, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20** 87-127.

ZHANG, R., ROBINSON, P. M. and YAO, Q. (2019). Identifying cointegration by eigenanalysis. *Journal of American Statistical Association: Theory and Methods* **114** 916-927.