

Attribution of Spurious Factors from High-Dimensional Functional Time Series

BY ADAM NIE

*Research School of Finance, Actuarial Studies and Statistics,
The Australian National University, Australia*

adam.nie@anu.edu.au

YANRONG YANG

*Research School of Finance, Actuarial Studies and Statistics,
The Australian National University, Australia*

yanrong.yang@anu.edu.au

HAN LIN SHANG

*Department of Actuarial Studies and Business Analytics,
Macquarie University, Australia*

hanlin.shang@mq.edu.au

SUMMARY

This article explores a general factor structure for high-dimensional nonstationary functional time series, encompassing a wide range of factor models studied in the existing literature. We investigate the asymptotic spectral behaviors of the sample covariance operator under this general data structure. A novel fundamental sufficient condition, formulated in terms of a new introduced effective rank tailored to this setup, is established under which empirical eigen-analysis yields spurious results, rendering sample eigenvalues and eigenvectors unreliable for accurately recovering the underlying factor structure. This generalizes the results of [Onatski & Wang \(2021\)](#) from typical high-dimensional time series (HDTs) to the more intricate functional framework. The newly defined effective rank is rigorously analyzed through a decomposition of the effects attributable to functional factor loadings and functional factors. Contrary to the findings in the HDTs setting, empirical eigen-analysis of models with only a small number of strong non-stationary factors may still produce spurious limits in the functional framework. Therefore, additional caution is warranted when applying covariance-based statistical methods to potentially nonstationary functional data. Simulation studies are performed to determine conditions under which spurious limits occur. Real data analysis on age-specific mortality rate data from multiple locations is conducted for evidence of spurious factors induced by empirical eigen-analysis.

Some key words: Factor Structure; High-dimensional Functional time series; Non-stationarity; Sample Covariance Operator; Spurious Analysis.

1. INTRODUCTION

Functional data analysis concerns random observations that take values in spaces of functions, typically square-integrable function spaces over compact intervals. Such data have gained substantial attention over the past few decades, both among theoreticians and researchers in applied fields (see, e.g., [Hsing & Eubank, 2015](#); [Wang et al., 2016, 2015](#); [Koner & Staicu, 2023](#);

Ramsay & Silverman, 1997). To address the inherent infinite-dimensional nature of functional data, dimension-reduction techniques—such as functional principal component analysis and functional factor models—are widely employed (Hall & Hosseini-Nasab, 2006; Benko et al., 2009; Shang, 2014; Hall et al., 2006).

Functional time series constitute a special class of functional data in which the observed random functions exhibit serial dependence. In the univariate functional time series setting, a substantial body of literature has extended classical time series methodologies to the functional framework (Hörmann & Kokoszka, 2012, 2010; Hyndman & Shang, 2009; Aue et al., 2015; Horváth et al., 2014, see, e.g., and references therein).

The analysis of multiple functional time series, particularly in high-dimensional regimes where the number of series is comparable to or exceeds the sample size, has received growing attention; see, for example, Gao et al. (2019); Jiménez-Varón et al. (2024); Hallin et al. (2023); Chang et al. (2024b,a); Guo et al. (2024); Leng et al. (2024). A central difficulty in this setting is the coexistence of multiple sources of dependence: within-curve dependence across the functional domain, serial dependence over time, and cross-sectional dependence across series. Each component may be high-dimensional. For instance, the Japanese sub-national mortality data analysed in Jiménez-Varón et al. (2024) comprise 46 functional time series observed over 50 years, with each annual observation recorded on a grid of 110 points.

Dimension-reduction methods for multivariate and high-dimensional functional time series include extensions of functional principal component analysis (see, e.g., Gao et al., 2019; Chiou et al., 2014; Happ & Greven, 2018; Di et al., 2009; Zapata et al., 2022) and functional factor models (see, e.g., Jiménez-Varón et al., 2024; Hallin et al., 2023; Chang et al., 2024b,a; Guo et al., 2024; Leng et al., 2024). These approaches approximate the data by low-rank representations and are effective when the underlying data-generating mechanism is itself approximately low-rank.

Many of the classical methods discussed above are primarily developed under the assumption of stationarity. In recent years, however, increasing attention has been devoted to the non-stationary setting (see, e.g., van Delft & Eichler, 2018; van Delft & Dette, 2021; Chang et al., 2016; Li et al., 2023; Horváth et al., 2014). For non-functional data, it is well established that directly applying principal component analysis (PCA) or factor analysis to high-dimensional non-stationary time series can lead to spurious conclusions. In such settings, temporal dependence may dominate cross-sectional dependence, even when strong common factors are present. As a result, the leading eigenvalues obtained from the standard PCA procedure primarily reflect serial dependence rather than genuine cross-sectional correlation. Consequently, the extracted components may fail to represent the underlying factor structure and instead capture persistent time dynamics.

This issue was rigorously studied by Onatski & Wang (2021), who showed that standard principal component analysis (PCA) may produce spurious factors when applied to high-dimensional integrated time series lacking a genuine low-rank structure. Even in the absence of true common factors, the leading sample principal components can explain a large fraction of the total variance, leading to erroneous inference. They further established that, under high-dimensional asymptotics, the sample eigenvalues and eigenvectors converge to functionals of a Wiener process, irrespective of the underlying covariance structure, indicating that strong temporal dependence can dominate the spectral behavior. Building on this framework, He & Zhang (2024) extended the analysis to models with more general forms of temporal dependence beyond the unit root case. They also derived the asymptotic distributions of the sample eigenvalues, which enable formal tests to distinguish genuine factors from spurious ones. To mitigate the occurrence of spurious factors in practice, Zhang et al. (2025) proposed a data-adaptive method to identify the underlying structure of high-dimensional time series before conducting eigen-analysis.

The present paper investigates the emergence of spurious phenomena in high-dimensional functional time series, with an extra functional dimension. We consider a general high-dimensional functional factor model with non-stationary factors that encompasses several existing frameworks in the literature (see, e.g., Hallin et al., 2023; Tavakoli et al., 2023; Guo et al., 2024; Leng

et al., 2024). In parallel with results from the high-dimensional literature, under mild rank-type conditions on the covariance operator of the process, the empirical eigenvalues and eigen-functions converge to a spurious limit that fails to represent the genuine covariance structure. As a result, methodologies relying on functional principal component analysis can produce inconsistent or misleading conclusions in the presence of non-stationarity.

We generalize the core results of Onatski & Wang (2021) to the setting of high-dimensional functional time series (see Theorem 1). The infinite-dimensional nature of functional observations introduces substantial technical challenges and gives rise to asymptotic behavior that is qualitatively different from that observed in finite-dimensional settings. In particular, the conclusions of our main theorems diverge significantly from those of Onatski & Wang (2021), highlighting structural features that are intrinsic to functional data. We elaborate on these distinctions below.

Onatski & Wang (2021) introduce a sufficient condition for the emergence of a spurious limit based on the notion of an “effective rank,” which characterizes the extent to which the model can be well approximated by a low-rank structure. They show that models with a large effective rank are prone to exhibiting spurious limits. In our setting, however, the concept of effective rank becomes more challenging due to an additional layer of dependence inherent in functional data. For high-dimensional (or multivariate) functional time series, the covariance operator encodes both cross-sectional dependence across coordinates and dependence along the functional domain. Disentangling these two sources of dependence requires additional technical work in order to derive results that admit a clear interpretation; see Theorem 2 and Corollary 1.

Furthermore, the existing theoretical framework does not accommodate models with a fixed number of strong factors. In contrast, this restriction is relaxed in our setting due to the intrinsic high dimensionality of functional data. Specifically, Theorem 2 establishes that even models driven by only a small number of strong non-stationary factors may converge to the spurious limits characterized in Theorem 1. This constitutes a substantial departure from the settings considered in Onatski & Wang (2021) and He & Zhang (2024), where a finite number of strong factors are treated as genuine and thus do not converge to spurious limits. Our findings indicate that, in the functional framework, the presence of an apparently well-defined factor structure does not safeguard against spurious eigenstructure in practice. Consequently, particular caution is warranted when applying covariance-based methods to non-stationary functional time series, even in the presence of strong factors.

The remainder of the paper is organized as follows. Section 2.1 introduces the model under consideration, and Section 2.2 presents several illustrative examples together with their connections to the existing literature. Section 2.3 defines the principal mathematical objects of interest, namely a class of sample covariance operators, and introduces the notation and preliminary results required for the development of our theory. Section 3.1 states and discusses the main assumptions underlying our theoretical analysis, while the primary results are presented in Section 3. Further discussion of the key sufficient condition, Assumption 2, is provided in Section 3.3, where we also establish additional theoretical results that clarify the circumstances under which spurious limits arise. Section 4 reports simulation results and examines their implications in light of the theory. In Section 5, we apply our methodology to empirical datasets. Finally, Section 6 summarizes the main findings and proposes a conjecture motivated by several notable simulation patterns for future investigation.

Finally, we will introduce some notations for the rest of the paper. Throughout the paper, for a separable Hilbert space $\mathcal{H}$, we write $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}}$ for its inner product and the associated norm. We will routinely drop the subscript $\mathcal{H}$ and simply write $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ when the context is clear. For elements f, g on $\mathcal{H}$, we write $f \otimes g$ for the bounded linear operator $h \mapsto f \langle h, g \rangle$. For a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we write $\|x\|_0 = |\{i \leq n, x_i \neq 0\}|$ for the number of non-zero coordinates of x and $\|x\|_p = (\sum_{i \leq n} x_i^p)^{1/p}$ for the usual ℓ^p norm of x . For $p \in \mathbb{N}$, we write $\mathcal{H}^p = \oplus_{k=1}^p \mathcal{H}$ for the external direct sum of p copies of $\mathcal{H}$ endowed with the usual inner product. Elements of $\mathcal{H}^p$ will be denoted as $f = (f_1, \dots, f_p)$ where $f_i \in \mathcal{H}$ for $1 \leq i \leq p$. For a bounded linear operator A

on $\mathcal{H}$, we write $\|A\|$ and A^* for the operator norm and the adjoint of A respectively. The rank, the trace, the nuclear norm and the Hilbert-Schmidt norm of A will be denoted as $\|A\|_0$, $\langle A \rangle$, $\|A\|_1$ and $\|A\|_2$ respectively, whenever they are finite. For Hilbert-Schmidt operators A, B on $\mathcal{H}$, we write $\langle A, B \rangle = \langle A^* B \rangle$ for the natural inner product in the Hilbert space of Hilbert-Schmidt operators on $\mathcal{H}$. For sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$, we write $a_n \lesssim b_n$ if $a_n \leq cb_n$ for all n for some positive constant c independent of n . We write $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. We write $a_n \ll b_n$ or $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$ and $a_n \gg b_n$ if $b_n \ll a_n$. Finally, we use O_p and o_p to denote the usual notions of stochastic compactness and convergence in probability.

2. SETUP: MODEL AND ASSUMPTIONS

2.1. A General Functional Factor Model

Let $\mathcal{H} = L^2(\mathcal{I})$ be the Hilbert space of square integrable functions on some compact interval $\mathcal{I}$ and $\mathcal{H}^p = \oplus_{k=1}^p \mathcal{H}$ be the external direct sum of p copies of $\mathcal{H}$ endowed with the usual inner product. We consider an integrated p -dimensional functional time series $X = (X_1, \dots, X_T)$ of length T taking values on $\mathcal{H}^p$. We assume X follows the functional factor model below

$$X_t(u) = (\Psi F_t)(u) + \zeta_t(u), \quad u \in \mathcal{I}, \quad (1)$$

where $(F_t)_t$ is an integrated functional time series on $\mathcal{H}^K$ for some $K < p$, the function Ψ is a bounded linear operator from $\mathcal{H}^K$ to $\mathcal{H}^p$, and $(\zeta_t)_t$ is a stationary functional time series on $\mathcal{H}^p$ with mean zero and covariance operator C_ζ . It is natural to view (1) as a functional factor model with non-stationary factors $(F_t)_t$ and factor loading operator Ψ . This functional factor model takes a general form in the sense of both factors and loadings being functions, which covers common factor styles in relevant literature [Hallin et al. \(2023\)](#); [Guo et al. \(2024\)](#) and [Leng et al. \(2024\)](#).

Analogous to [Onatski & Wang \(2021\)](#), the functional factor $(F_t)_t$ is assumed to be a random walk on $\mathcal{H}^K$ of the form $F_t = \sum_{s=1}^t \epsilon_s$ where $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{Kt})^\top$ and $(\epsilon_{kt})_{kt}$ is a collection of independent random functions in $\mathcal{H}$ with mean zero and covariance operator C_ϵ . The linear operator Ψ is assumed to be a matrix of integral operators, i.e., there exists a collection $(\Psi_{ik})_{i \leq p, k \leq K}$ of integral operators such that

$$(\Psi \epsilon_t)(u)_i = \sum_{k=1}^K \int_{\mathcal{I}} \Psi_{ik}(u, v) \epsilon_{kt}(v) dv, \quad u \in \mathcal{I},$$

where $\Psi_{ik}(\cdot, \cdot)$ denotes the integral kernel of the integral operator Ψ_{ik} . This structure is also adopted by [Leng et al. \(2024\)](#). We will write $\Psi(\cdot, \cdot)$ as the matrix of kernel functions $(\Psi_{ik}(\cdot, \cdot))_{ik}$ and write the above equation in the matrix form

$$(\Psi \epsilon)(u) = \int_{\mathcal{I}} \Psi(u, v) \epsilon(v) dv, \quad u \in \mathcal{I}.$$

Additional technical assumptions on the model are discussed in [Section 3](#).

2.2. Model Generality

Our framework is built upon a highly general functional factor structure in which both the factors F and the factor loadings Ψ are functional objects. This formulation is conceptually related to the dual functional factor model proposed by [Leng et al. \(2024\)](#); however, their analysis is restricted to stationary time series. The fully functional factor structure specified in (1) encompasses a broad class of existing models as special cases. We now illustrate this generality by presenting several concrete examples and discussing their connections to the literature.

Example 1. (Functional Loadings).

Suppose the factors F_{kt} takes the form of

$$F_{kt}(\cdot) = \tilde{F}_{kt}g(\cdot)$$

for some deterministic function $g \in \mathcal{H}$ and real random variables $(\tilde{F}_{kt})_{kt}$. Set

$$\tilde{\Psi}(\cdot) = \int \Psi(\cdot, v)g(v)dv,$$

then our factor structure in (1) reduces to the functional factor model of [Hallin et al. \(2023\)](#) and [Tavakoli et al. \(2023\)](#) with factors $(\tilde{F}_{kt})_{kt}$ and factor loading operator $\tilde{\Psi}$. In this case the factor loading is a functional object while the factors is given by a multivariate time series on $\mathbb{R}^K$.

Example 2. (Functional Factors).

Suppose the factor loadings are given by

$$\Psi_{ik}(u, v) = \delta(u - v)\tilde{\Psi}_{ik}$$

where $\tilde{\Psi} = (\tilde{\Psi}_{ik})_{ik}$ is a $p \times K$ real matrix and δ is the Dirac δ function at zero. Then we have

$$(\Psi F_t)(u) = \int_{\mathcal{I}} \delta(u - v)\tilde{\Psi}F_t(v)dv = \tilde{\Psi}F_t(u)$$

and our factor structure in (1) reduces to the functional factor model recently proposed by [Guo et al. \(2024\)](#). In this case the factors are functional objects while the factor loadings are given by a $p \times K$ real matrix.

We note that, under the factor structure in Example 1, both the cross-sectional dependence and the dependence along the functional domain are entirely induced by the factor loadings. In contrast, under the structure in Example 2, the cross-sectional dependence is determined by the factor loading matrix, whereas the dependence along the functional domain is jointly driven by the latent factors and the loading matrix. In our general framework, both types of dependence emerge from the interaction between the factors and the factor loadings, as will become evident from the main theoretical results.

The fully functional factor structure is also closely related to the study of non-stationarity and cointegration in functional time series, topics that have attracted increasing attention in the recent literature (see, e.g., [Beare et al., 2017](#); [Chang et al., 2016](#), and references therein). We present a simplified example below to illustrate these connections.

Example 3. (Cointegrated Functional Time Series).

Suppose $p = K = 1$ and Ψ is a bounded linear operator on $\mathcal{H}$ of finite rank. Writing

$$\Delta X_t = \Psi \epsilon_t + \Delta \zeta_t,$$

our model in (1) becomes the Beveridge-Nelson decomposition of the cointegrated functional time series studied in [Beare et al. \(2017\)](#); [Chang et al. \(2016\)](#). The rank of Ψ here represents the dimension of the attractor space, i.e. the vectors h in the kernel of Ψ are the directions such that the projection $\langle h, X_t \rangle$ is of $I(0)$. Our model can be extended to be compatible with recent results that study integrated time series of order $I(d)$ where d can be fractional (see [Li et al., 2023](#); [Beare & Seo, 2020](#)). We do not pursue such extensions in our current work.

Finally, although our model is formulated with the factors F evolving as a random walk on $\mathcal{H}$, the framework in fact accommodates a broad class of temporal dependence structures, in the spirit of [Onatski & Wang \(2021\)](#). In particular, the specification in (1) can be interpreted as the Beveridge-Nelson decomposition of $I(1)$ linear process on $\mathcal{H}$ exhibiting more general short-run dynamics.

As demonstrated by our main results, the random-walk component of the factors asymptotically dominates the behavior of the sample covariance operator, while mild deviations from pure

random-walk dynamics have asymptotically negligible effects. We conclude this section with an additional example illustrating the range of temporal dependence structures encompassed by our framework.

Example 4. (Multivariate Functional Time Series).

Suppose the factors $(F_1, \dots, F_T) \subseteq \mathcal{H}^K$ follow the model

$$F_t(u) = \sum_{n=1}^N Y_t^n \phi_n(u),$$

where (ϕ_n) is a complete orthonormal set of functions on $\mathcal{H}$ and for each n , the vector-valued coefficients $(Y_t^n, t = 1, \dots, T)$ follow a K -dimensional vector ARIMA model of the form

$$\Phi_n(L)(1-L)(Y_t^n - Y_0^n) = \Theta_n(L)\epsilon_t^n$$

where Φ_n and Θ_n are the matrix valued autoregressive and moving average polynomials. Assuming causality of the ARIMA processes and writing $\Psi_n(L) := \Phi_n(L)^{-1}\Theta_n(L)$, the multivariate Beveridge-Nelson decomposition of Y_t^n is given by

$$Y_t^n = \Psi_n(1) \sum_{s \leq t} \epsilon_s^n + \zeta_t^n,$$

where $(\zeta_t^n)_t$ is a stationary process given by $\zeta_t^n = -\sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} \Psi_i \epsilon_t^n$. Now, let $\Psi : \mathcal{H}^K \rightarrow \mathcal{H}^K$ be a linear operator given by $\Psi(a\phi_n) := \Psi_n(1)a\phi_n$ where a is an arbitrary vector in $\mathbb{R}^K$. The factor process can then be written as

$$F_t(u) = \Psi \sum_{s \leq t} \sum_{n=1}^N \epsilon_s^n \phi_n(u) + \sum_{n=1}^N \zeta_t^n \phi_n(u).$$

This representation shows that, the multivariate functional time series $F_t(u)$ is also a special case of the functional factor model in (1).

In light of Examples 1–4, the proposed functional factor model accommodates a rich variety of cross-sectional and temporal dependence structures.

2.3. The Sample Covariance Operator Matrix

The analysis of the model (1) and the estimation of factors and factor loadings typically rely on the eigen-analysis of the sample covariance operator $\frac{1}{T} \sum_{t=1}^T [X_t \otimes X_t]$ on $\mathcal{H}^p$. In Euclidean space, it is common practice to leverage on the duality between the column space and the row space of a matrix and consider the $T \times T$ Gram matrix instead. In the context of functional data, this estimation approach has also been taken in existing work (Benko et al., 2009; Leng et al., 2024; Tavakoli et al., 2023) when estimating factors and factor loadings. We provide a formal justification of this duality in Lemma A5 in Appendix.

Let $\bar{X}(u) := \frac{1}{T} \sum_{t=1}^T X_t(u)$ be the sample mean. By Lemma A5, the sample covariance operator of X has the same non-zero spectrum as the Gram matrix $\hat{S} = (\hat{S}_{st})$ given by

$$\hat{S}_{st} := \frac{1}{p} \sum_{i=1}^p \langle X_{is} - \bar{X}_i, X_{it} - \bar{X}_i \rangle \quad (2)$$

Let M denote the orthogonal projection onto the orthogonal complement of the T -dimensional vector of ones, i.e. $M := I_T - T^{-1}1_T 1_T'$. Then we may write

$$\hat{S} = \frac{1}{p} \int_{\mathcal{U}} (X(u) - \bar{X}(u))' (X(u) - \bar{X}(u)) du = \frac{1}{p} \int_{\mathcal{U}} M X(u)' X(u) M du$$

As will be shown, in cases where spurious behaviors arise, this Gram matrix is dominated by the Gram matrix of the non-stationary part of the original data, which we will denote as

$$\tilde{S} := \frac{1}{p} \int_{\mathcal{U}} M(\Psi F)(u)' (\Psi F)(u) M du. \quad (3)$$

Write Θ for the $T \times T$ upper triangular matrix with entries equal to 1 so that, based on the random walk structure on $F(u)$,

$$(\Psi F)(u) = (\Psi \epsilon)(u) \Theta = \int_{\mathcal{I}} \Psi(u, v) \epsilon(v) \Theta dv.$$

By Fubini's theorem, the Gram matrix $\tilde{S}$ can be written as

$$\begin{aligned} \tilde{S} &= \frac{1}{p} M \Theta' \int_{\mathcal{I}} \left(\int_{\mathcal{I}} \epsilon(v)' \Psi(u, v)' dv \int_{\mathcal{I}} \Psi(u, w) \epsilon(w) dw \right) du \Theta M \\ &= \frac{1}{p} M \Theta' \iint_{\mathcal{I}^2} \epsilon(v)' \left(\int_{\mathcal{I}} \Psi(u, v)' \Psi(u, w) du \right) \epsilon(w) dv dw \Theta M =: \frac{1}{p} M \Theta' W \Theta M, \end{aligned}$$

where W denotes the $T \times T$ random matrix given by

$$W := \iint_{\mathcal{I}^2} \epsilon(v)' \Omega(v, w) \epsilon(w) dv dw, \quad (4)$$

and $\Omega(v, w)$ is the $K \times K$ matrix given by

$$\Omega(v, w) := \int_{\mathcal{I}} \Psi(u, v)' \Psi(u, w) du, \quad v, w \in \mathcal{I}. \quad (5)$$

Note that the transpose of Ω is given by $\Omega(v, w)' = \int_{\mathcal{I}} \Psi(u, w)' \Psi(u, v) du = \Omega(w, v)$, i.e. for any i, j we have $\Omega_{ji} = \Omega_{ij}^*$. Proposition 1 below shows that Ω is a self-adjoint operator on $\mathcal{H}^K$.

In the analysis of the Gram operator $\tilde{S}$, we will often encounter certain compositions between matrices of operators. For convenience, we provide a new definition on the operator product and explore its basic properties.

DEFINITION 1. (*A New Operator-Matrices Product*)

Let C_1 and C_2 be bounded linear operators on $\mathcal{H}$ and $\Omega = (\Omega_{ij})$ be a $K \times K$ matrix of bounded linear operators on $\mathcal{H}$ for some fixed K . We define

$$C_1 \Omega C_2 := (C_1 \Omega_{ij} C_2)_{ij}$$

to be the $K \times K$ matrix of operators where $C_1 \Omega_{ij} C_2$ is the common composition between operators. The operator $C_1 \Omega C_2$ acts on $\mathcal{H}^K$ by

$$(C_1 \Omega C_2 f)_i := \sum_{j=1}^K C_1 \Omega_{ij} C_2 f_j, \quad (f_j)_{j=1}^K \in \mathcal{H}^K.$$

We develop some basic properties for such matrices of operators. The proofs can be found in Section A.

PROPOSITION 1. Let C_1 and C_2 be given as in Definition 1, and Ω is defined in (5). Then

1. The adjoint of Ω is given by $(\Omega^*)_{ij} = \Omega_{ji}^*$, hence Ω is self-adjoint if and only if $\Omega_{ij}^* = \Omega_{ji}$. Furthermore, the adjoint of $C_1 \Omega C_2$ is given by

$$(C_1 \Omega C_2)^* = C_2^* \Omega^* C_1^*.$$

2. Suppose $C_1\Omega_{ii}C_2$ is a trace-class operator on $\mathcal{H}$ for all $i = 1, \dots, K$. Then $C_1\Omega C_2$ is a trace class operator on $\mathcal{H}^K$ and its trace, denoted as $\langle C_1\Omega C_2 \rangle$, is given by

$$\langle C_1\Omega C_2 \rangle = \sum_{i=1}^K \langle C_1\Omega_{ii}C_2 \rangle.$$

3. Suppose the operator $C_1\Omega_{ij}C_2$ is Hilbert-Schmidt on $\mathcal{H}$ for all $i, j = 1, \dots, K$. Then the operator $C_1\Omega C_2$ is Hilbert-Schmidt on $\mathcal{H}^K$ and

$$\|C_1\Omega C_2\|_2^2 := \langle C_1\Omega C_2 C_2^* \Omega^* C_1^* \rangle = \sum_{ij} \|C_1\Omega_{ij}C_2\|_2^2.$$

3. ASYMPTOTIC THEORY FOR SPURIOUS PHENOMENON

3.1. Assumptions

The following set of assumptions are made throughout the paper.

Assumption 1 (Moments Condition). Assume $p \rightarrow \infty$ as $T \rightarrow \infty$. Furthermore, suppose that

1. $\epsilon := (\epsilon_{it})_{it}$ is a $K \times T$ matrix of independent random elements on $\mathcal{H}$ with $\mathbb{E}[\epsilon_{it}] = 0$ and covariance operator $C_\epsilon := \mathbb{E}[\epsilon_{it} \otimes \epsilon_{it}]$ for all i, t .
2. The fourth moment of ϵ_{it} is uniformly bounded in the sense that

$$\kappa_4^* := \sup_{i,t} \sup_{u,v,\alpha,\beta} \frac{\mathbb{E}[\epsilon_{it}(u)\epsilon_{it}(v)\epsilon_{it}(\alpha)\epsilon_{it}(\beta)]}{C_\epsilon(u,v)C_\epsilon(\alpha,\beta)} < \infty.$$

Remark 1. The condition $p \rightarrow \infty$ places the analysis in a high-dimensional regime, while no restriction is imposed on the relative growth rates of p and T .

Assumption 1.1 ensures that the latent factors ϵ follow a covariance stationary process with K independent components. The requirement that the elements of ϵ be independent across $t = 1, 2, \dots, T$ may appear restrictive. However, this assumption can be readily relaxed to accommodate common time-series structures such as ARMA models; see Example 4 in Section 2.2. The present formulation of the factor structure is adopted for convenience, as it clearly distinguishes the $I(1)$ and $I(0)$ components of the model, which is the main focus of this article.

Assumption 1.2 simply ensures that the latent factors have uniformly bounded 4-th moment. The appearance of the covariance function C_ϵ in the denominator takes into account the fact that ϵ_{it} is an element of a Hilbert space and has a nontrivial covariance structure.

The following assumption on the covariance operators C_ϵ and Ω outlines the key condition for the emergence of spurious behaviours in sample covariance-based eigen-analysis.

Assumption 2 (Effective Rank). As $p, T \rightarrow \infty$, the effective rank, based on the operators C_ϵ and Ω , satisfies

$$\mathcal{R} := \langle C_\epsilon^{1/2} \Omega C_\epsilon^{1/2} \rangle / \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\| \rightarrow \infty. \quad (6)$$

Remark 2. The ratio $\mathcal{R}$ is the effective rank of the operator $C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}$, a quantity widely used in the high dimensional probability literature (Koltchinskii & Lounici, 2016; Rudelson & Vershynin, 2010; Wainwright, 2019). A large effective rank indicates that the operator cannot be well approximated by low-rank operators in the operator norm.

It is straightforward to see that the operator $\mathcal{C} := C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}$, as defined in Definition 1, is a self-adjoint, non-negative definite operator on $\mathcal{H}^K$. This implies $\|\mathcal{C}\|^2 \leq \|\mathcal{C}\|_2^2 \leq \langle \mathcal{C} \rangle \|\mathcal{C}\|$, from which we obtain the estimates $\langle \mathcal{C} \rangle / \|\mathcal{C}\|_2 \leq \langle \mathcal{C} \rangle^2 / \|\mathcal{C}\|_2^2 \leq \langle \mathcal{C} \rangle^2 / \|\mathcal{C}\|^2$. Therefore, Assumption 2 is equivalent to the condition

$$\|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_2 \ll \langle C_\epsilon^{1/2} \Omega C_\epsilon^{1/2} \rangle.$$

We now offer several remarks on the interpretation of Assumption 2 and its relation to the existing literature.

Remark 3. In the non-functional settings considered by Onatski & Wang (2021) and He & Zhang (2023), the effective rank $\mathcal{R}$ is bounded below by the number of non-stationary factors K . Consequently, condition (6) is ensured under the stronger but interpretable assumption that $K \rightarrow \infty$. This is also the framework within which Onatski & Wang (2021) and He & Zhang (2024) develop the asymptotic theory for spurious phenomena.

We work directly under the intrinsic condition (6) in the functional setting. One reason is that the lower bound $\mathcal{R} \gtrsim K$ no longer holds due to the presence of the operator C_ϵ . In fact, condition (6) reflects both the cross-sectional and the functional dependence structures in the data; see Section 3.3 for further discussion. This represents a key distinction between the present paper and the existing literature. We show that models with $K \sim p$ factors may exhibit no spurious behaviour, while models with only a few factors may display pronounced spurious effects. Consequently, even models with a small number of genuine non-stationary factors, eigen-analysis based on the sample covariance can produce spurious results in the functional setting. Extra caution is therefore required when applying eigen-analysis to potentially non-stationary functional time series.

It is worth noting that, in Section 3.3, we derive more interpretable results by establishing upper and lower bounds for $\mathcal{R}$ under common time series models. These results provide clearer insight into when spurious behaviour arises and how it relates to K .

Finally, the following assumption gives a lenient bound on the size of the error time series ζ .

Assumption 3 (Error Components). Assume that the error term $(\zeta_t, t = 1, \dots, T)$ is a weakly stationary time series on $\mathcal{H}^p$ with mean zero and covariance operator $C_\zeta = \mathbb{E}[\zeta_t \otimes \zeta_t] \in \mathcal{H}^p$ satisfying

$$\langle C_\zeta \rangle \ll T \langle C_\epsilon \Omega \rangle, \quad T \rightarrow \infty. \quad (7)$$

Condition (3) ensures ζ is negligible asymptotically compared to the non-stationary component of the model. To illustrate the scope of this condition, suppose the components $\zeta_{1t}, \dots, \zeta_{pt}$ of ζ_t have uniformly bounded covariance operators, in which case condition (7) is equivalent to $\langle C_\epsilon \Omega \rangle \gg pT^{-1}$. Drawing a parallel with the high dimensional factor modelling literature, it is natural to assume that the strength of the factors increases with p , i.e. $\|\Omega\| \sim p^{1-\delta}$ for some $\delta \in [0, 1]$ (see Lam & Yao, 2012; Fan et al., 2021). Since C_ϵ is a trace class operator, condition (7) reduces to $p^\delta T^{-1} = o(1)$, which clearly holds for a wide range of high dimensional models. In particular, condition (7) is trivially satisfied when the factors are “strong”, i.e. when $\delta = 0$.

3.2. Main result

We are now ready to state the main result of our work.

THEOREM 1. Let $(\hat{\lambda}_t)_{t=1}^T$ be the eigenvalues of the sample covariance matrix $\hat{S}$ arranged in non-ascending order and $(\hat{u}_n)$ be the corresponding eigenvectors. Suppose Assumptions 1-3 hold, then for any fixed natural number k , as $T \rightarrow \infty$, we have

(i) the sample eigenvector $\hat{u}_k$ satisfies

$$|\langle \hat{u}_k, d_k \rangle| = 1 + o_p(1)$$

where $d_k \in \mathbb{R}^T$ with the t -th coordinate equal to $d_{kt} = \sqrt{2/T} \cos(\pi kt/T)$,

(ii) the sample eigenvalue $\hat{\lambda}_k$ satisfies

$$\hat{\lambda}_k = \frac{T^2}{k^2 \pi^2 p} \langle C_\epsilon \Omega \rangle (1 + o_p(1)),$$

(iii) the percentage of variance explained by $\hat{\lambda}_k$ satisfies

$$\frac{\hat{\lambda}_k}{\sum_j \hat{\lambda}_j} = \frac{6}{(k\pi)^2} + o_p(1).$$

Theorem 1 gives the asymptotic limits of sample eigenvalues and eigenvectors as T tends to infinity. Part (ii) states that the sample eigenvalues $\hat{\lambda}_k$ tends to deterministic limits. There are several observations to be made. First, the true covariance structure of the model, captured by C_ϵ and Ω , appears in the limit of the eigenvalue $\hat{\lambda}_k$ only in the form of a tracial quantity. In particular, the limit of $\hat{\lambda}_k$ depends on the index k only through a factor of k^{-2} and is not related to the k -th eigenvalue of C_ϵ and Ω . Thus, the sample eigenvalues $\hat{\lambda}_k$ do not capture any useful information about the cross-sectional dependence structure of the model.

Furthermore, the limit of $\hat{\lambda}_k$ is of order $T^2 p^{-1} \langle C_\epsilon \Omega \rangle$ for any fixed k , regardless of how many factors are actually present in the model. In particular, even though the sample eigenvalues appear to be decaying rapidly at the speed of k^{-2} , any inference on the number of factors based on sample eigenvalues is guaranteed to be misguided.

Lastly, we recall that in the factor modelling literature, the term ‘strong factors’ is typically used to refer to factor loadings with $\|C_\epsilon \Omega\| \sim p$. However, we see that even under the assumption of strong factors, the leading sample eigenvalues $\hat{\lambda}_k$ are still of order T^2 , which is much larger than the strength of factors. This suggests that the non-stationarity in the model dominates the covariance structure even in the presence of strong factors.

Part (i) of the theorem states that the estimated eigenvectors $\hat{u}_k$, which are used as estimates for ‘factors’ in certain models (see Tavakoli et al., 2023; Leng et al., 2024) converge in probability to a deterministic trigonometric function d_k . It is important to note that this limit only depends on the index k and is independent of the actual covariance structure of the model. Factors estimated this way are hence labeled as spurious in nature since they do not relate to the true factors ϵ and the factor loadings Ψ in any way at all.

The appearance of trigonometric functions in the limit is not surprising in this context. These functions are precisely the eigenfunctions appearing in the Karhunen-Loève expansion of a demeaned standard Wiener process, which arises as the limit of a scaled random walk. As explained above, the covariance structure in the spurious case is dominated by the non-stationarity in the model instead of the cross-sectional dependence, and the factor estimates are essentially estimating the eigenfunctions of the Wiener process.

Lastly, part (iii) of the theorem gives limits of the proportion of variance explained by each sample eigenvalue. It can be seen that the limit is deterministic and does not contain any information about the true factor structure C_ϵ and Ω . As a consequence, any inference about the factor strength based on the scree plot would be inaccurate.

3.3. Discussion on Assumption 2

Assumption 2 gives a sufficient condition for the emergence of the spurious limit similar to the preceding works Onatski & Wang (2021) and He & Zhang (2024). The key quantity $\langle \mathcal{C} \rangle / \|\mathcal{C}\|_2$, the ratio between the trace and the Hilbert-Schmidt norm of the operator $\mathcal{C} = C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}$, depends on both the cross-sectional structure and the functional structure of the model. In certain special cases, we can separate the two effects and give concrete bounds on this ratio, which allows us to formulate sufficient conditions for Theorem 1 in terms of more basic model parameters.

By Mercer’s theorem, the operator C_ϵ has a spectral decomposition given by

$$C_\epsilon = \sum_{n=1}^{\infty} c_n \phi_n \otimes \phi_n \tag{8}$$

where $(\phi_k)_k$ is a complete orthonormal basis for $\mathcal{H}$ and $(c_n)_n$ is a non-negative sequence in ℓ^1 . Suppose that the factor loading operators (Ψ_{ik}) are given by

$$\Psi_{ik} = \sum_{n=1}^q a_{nik} \phi_n \otimes \phi_n \quad (9)$$

for some collection of real numbers $(a_{nik})_{nik}$ satisfying $\sup_n \sum_{i,k} a_{nik}^2 < \infty$. For $n \in \mathbb{N}$, define the matrices $A_n = (a_{nik})_{ik}$ and $B_n = A_n A_n'$. Essentially, $c_n B_n$ is the covariance matrix of the factor part of the model $\Psi \epsilon_t$ projected component-wise to the one dimensional closed linear subspace of $\mathcal{H}$ spanned by the function ϕ_n .

Under this specific structure for the factor loadings, we can formulate the following two-sided estimates on the ratio $\langle \mathcal{C} \rangle / \|\mathcal{C}\|_2$.

THEOREM 2. *Let C_ϵ and Ψ be given as in (8) and (9). Suppose that the matrices $(B_n, n \in \mathbb{N})$ satisfy $\|B_n\|_2 \lesssim \|B_m\|_2$ uniformly in $n, m \in \mathbb{N}$. Then we have the upper bound*

$$\frac{\langle \mathcal{C} \rangle}{\|\mathcal{C}\|_2} \lesssim \frac{1}{\|C_\epsilon\|_2} \sup_{n \in \mathbb{N}} \frac{\langle B_n \rangle}{\|B_n\|_2} \leq \frac{1}{\|C_\epsilon\|_2} \sup_{n \in \mathbb{N}} \|B_n\|_0^{1/2}, \quad (10)$$

and the lower bound

$$\frac{\langle \mathcal{C} \rangle}{\|\mathcal{C}\|_2} \gtrsim \frac{1}{\|C_\epsilon\|_2} \left(1 + \sup_n c_n \frac{\langle B_n \rangle}{\|B_n\|_2} \right) \gtrsim \frac{1}{\|C_\epsilon\|_2} \left(1 + \sup_n c_n \alpha(B_n)^{1/2} \|B_n\|_0^{1/2} \right), \quad (11)$$

where $\alpha(B_n)$ denotes the ratio between the least non-zero eigenvalues and the largest eigenvalue of B_n . The omitted constants in the above estimates are all uniform in T .

Theorem 2 allows us to separate condition (6) into a cross-sectional component captured by the matrices $(B_n, n \in \mathbb{N})$ and a functional component captured by the operator C_ϵ . The term $\langle B_n \rangle^2 / \|B_n\|_2^2$ is analogous to the ratio in Assumption A3 of Onatski & Wang (2021). It is bounded from the above by the rank of B_n and from the below by the rank of B_n scaled by $\alpha(B_n)$ appearing in the Theorem. Intuitively, this term measures the effective rank of the model projected in the direction of ϕ_n . The term $\|C_\epsilon\|_2^{-1}$ serves as an effective rank of C_ϵ which measures the decay of the eigenvalues c_n . When the decay is fast, the ratio is close to 1 and when the decay is slow, the ratio is close to the actual rank of C_ϵ , which can be infinite. For instance, if $c_n = q^{-1} 1_{n \leq q}$ for some fixed $q > 0$, then $\|C_\epsilon\|_2^{-1} = \sqrt{q}$ which is equal to the square root of the rank of C_ϵ . On the other hand, suppose $c_n = 2^{-n}$ for all n , then $\|C_\epsilon\|_2^{-1} = \sqrt{3}$ even though the rank of C_ϵ is infinite.

To facilitate the discussion, it is helpful to define the following two different regimes for C_ϵ :

DEFINITION 2. *The operator C_ϵ is said to be “localized” if $\|C_\epsilon\|_2$ is uniformly bounded away from zero as $T \rightarrow \infty$, and “delocalized” if $\|C_\epsilon\|_2 \rightarrow 0$ as $T \rightarrow \infty$.*

The delocalized regime is when the factors are generated by an increasing number of basis functions with comparable weights as $T \rightarrow \infty$. A simple yet very general example of this can be seen by setting $c_n := (aT)^{-\delta} 1_{n \leq (aT)^\delta}$ for some fixed $a > 0$ and $\delta \in (0, 1]$ in the spectral decomposition of C_ϵ . The localized regime refers to when the factors are generated by an arbitrary number of basis functions, but the weights (c_n) rapidly decay to zero. In this setting $\|C_\epsilon\|_2^{-1}$ is bounded even though the rank of C_ϵ can be infinite.

As previously discussed in Section 3.1, the sufficient condition for the main theorem of Onatski & Wang (2021) implies that the number of stochastic trends K in the data must diverge as $T \rightarrow \infty$. As a consequence, their main theorem necessarily exclude models with a finite number of strong non-stationary factors. This observation is analogous to the upper bounds in (10) in our case. In particular, under the localized regime defined above, condition (6) still implies $K \rightarrow \infty$. However, this is no longer the case under the delocalized regime. Condition (6) can in fact be satisfied by models with a finite number of strong factors, as long as those factors are generated by an increasing number of basis functions with comparable weights as $T \rightarrow \infty$. This is an important

distinction since it implies that functional time series with few strong non-stationary factors can exhibit spurious behaviour as well. We will illustrate this through simulations in Section 4.

In addition to this, the lower bounds in Theorem 2 allows us to formulate sufficient conditions to Theorem 1 in terms of basic model parameters. In fact, from (11) we can conclude that condition (6) holds in the delocalized regime when all the B_n 's are comparable in $\|\cdot\|_2$, regardless of their actual rank. This is a significant departure from the conclusions in Onatski & Wang (2021) for the non-functional case, since it shows concretely that a model with single ($K = 1$) integrated functional factor can indeed tend to the spurious limit.

Finally, in the localized regime, a sufficient condition for (6) is the divergence of the quantity $c_n \langle B_n \rangle / \|B_n\|$ for some n . By the second lower bound in (11), this is satisfied whenever the rank of B_n diverges as $T \rightarrow \infty$, and the quantities c_n and $\alpha(B_n)$ do not decay too fast. This concretely shows that a model with large effective rank in some direction ϕ_n with non-diminishing weights has to tend to the spurious limit, which is in line with the findings of Onatski & Wang (2021).

For ease of referencing we summarize the above discussion into the following corollary.

COROLLARY 1. *Suppose that the conditions of Theorem 2 are satisfied. Then condition 6 is satisfied when C_ϵ is delocalized, or when $c_n \langle B_n \rangle / \|B_n\|_2$ diverges for some n .*

4. SIMULATION STUDY

4.1. Setup

We start with a description of the data generating process. Recall from (1) the model

$$X_{it}(u) = \sum_{s=1}^t \sum_{k=1}^K (\Psi_{ik} \epsilon_{ks})(u) + \zeta_{it}(u), \quad (i = 1, \dots, p, \ t = 1, \dots, T),$$

where $(\epsilon_{kt})_{kt}$ are independent random functions on $\mathcal{H}$ with mean zero and covariance operator C_ϵ and $(\zeta_{it})_{it}$ is a weakly stationary functional times series on $\mathcal{H}^p$ with zero mean and covariance C_ζ . For simplicity, we focus on the settings discussed in Section 3.3 and Theorem 2.

Fix $q > 0$ and let $(\phi_n)_{n=1}^q$ be a set of Fourier basis functions on $\mathcal{H}$. We set

$$C_\epsilon = \sum_{n=1}^{\infty} c_n \phi_n \otimes \phi_n, \quad (12)$$

where c_n is a sequence of non-negative real numbers summing to one, to be specified below. Similar to Leng et al. (2024) and Guo et al. (2024), we generate the factor process ϵ by setting

$$\epsilon_{kt}(u) := \sum_{n=1}^q Z_{kt}^n \phi_n(u),$$

where (Z_{kt}^n) is a collection of i.i.d. Gaussian variables with mean zero and variance equal to $\text{var}(Z_{kt}^n) = c_n$ for all k, t . For the noise time series ζ , we simply set

$$\zeta_{it}(u) := \sum_{n=1}^q W_{it}^n \phi_n(u),$$

where (W_{it}^n) is a collection of i.i.d. Gaussian variables with mean zero and variance equal to $\text{var}(W_{it}^n) = 2^{-n}$ for all i, t . Following (9), we generate the factor loadings $\{\Psi_{ik}\}$ by setting

$$\Psi_{ik}(u, v) = \sum_{n=1}^q a_{ik}^n \phi_n(u) \phi_n(v), \quad (13)$$

where for each $n = 1, \dots, q$, the $p \times K$ matrix $A_n := \{a_{ik}^n\}_{i,k}$ is to be specified below.

For the model parameters, we set $T = 200$, $p = 100$ and $q = 20$. Recall from Assumption 2 and Theorem 1 that the sample covariance structure tends to a spurious limit whenever the ratio

$$\mathcal{R} = \frac{\langle C_\epsilon^{1/2} \Omega C_\epsilon^{1/2} \rangle^2}{\|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_2^2}.$$

diverges as $T \rightarrow \infty$. From Theorem 2, it is clear that whether $\mathcal{R}$ diverges is determined by the parameters K , the eigenvalues $(c_n)_n$ in (12) and the matrices $(A_n)_n$ in (13). We will perform simulations with various choices of these parameters described below.

For the number of factors we set $K \in \{50, 10, 2\}$. In He & Zhang (2023) and Onatski & Wang (2021), the $K = 2$ setting represents as a model with genuine factors and is shown to be far from the spurious limit, while the $K = 50$ setting represents a model with no genuine factors for which spurious behaviours occur. We will see that this is not necessarily the case with functional data.

For the loading matrices $(A_n, n \in \mathbb{N})$, we will consider the high rank setting where

$$A_n \sim G_{p,K} \quad (*)$$

independently for all $n \in \mathbb{N}$, where $G_{p,K}$ denote a $p \times K$ matrix with i.i.d. standard Gaussian random variables, and the low effective rank setting where

$$A_n \sim p^{1/2} U_p \text{diag}(\{2^{-n/2}\}_{n=1}^K, 0, \dots, 0) U_K' \quad (**)$$

independently for all $n \in \mathbb{N}$, where U_N denotes an $N \times N$ random orthogonal matrix uniformly sampled from $\mathbb{S}^{N-1}$ for $N \in \{p, K\}$. In setting (*), each A_n has full column rank with high probability, while in setting (**), each A_n has low effective rank with

$$\frac{\langle A_n A_n' \rangle^2}{\|A_n A_n'\|_2^2} = 3 \frac{(1 - 2^{-K})^2}{1 - 4^{-K}} \ll K.$$

For the covariance operator C_ϵ , we consider the following three scenarios by setting

$$c_n = q^{-1} 1_{n \leq q}, \quad (\text{A})$$

$$c_n = 2^{-n} 1_{n \leq q}, \quad (\text{B})$$

$$c_n = 2^{-1} 1_{n \leq 2}. \quad (\text{C})$$

In Setting (A), each basis function ϕ_n has equal weight and we have $\|C_\epsilon\|_2^2 = q^{-1} = 0.02$ which is very close to zero. This represent the scenario where the operator C_ϵ is delocalized as defined in Definition 2. In Setting (B), the eigenvalues rapidly decay to zero and we have $\|C_\epsilon\|_2^2 = 1/3$. This represents the situation where C_ϵ is localized. Finally, a more extreme version of this is found in Setting (C), where C_ϵ is completely localized on two basis functions and we have $\|C_\epsilon\|_2^2 = 0.68$. We summarize the different combination of choices for A_n and C_ϵ in Table 1.

	Setting I	Setting II	Setting III	Setting IV	Setting V	Setting VI
C_ϵ	(A)	(B)	(C)	(A)	(B)	(C)
A_n	(*)	(**)	(*)	(**)	(*)	(**)

Table 1: Summary of the settings considered

For each setting in Table 1, we simulate X from the above data generating process with $K \in \{50, 10, 2\}$ and compute the leading eigenvalues and eigenvectors of the sample Gram matrix defined in (2). We plot the top five sample eigenvectors for $K \in \{50, 10, 2\}$ (red, green and blue lines respectively) against the spurious limit (black line) as shown in Theorem 1. We also plot the proportion of variance explained by each eigenvalue with the same color coding.

4.2. Results

We start with settings that shows obvious spurious behaviour. Figure 1 plots the sample eigenvectors of Setting I where C_ϵ is in the delocalized regime and A_n is of high rank. It can be seen that all sample eigenvectors as well as the sample eigenvalues closely resemble the spurious limits.

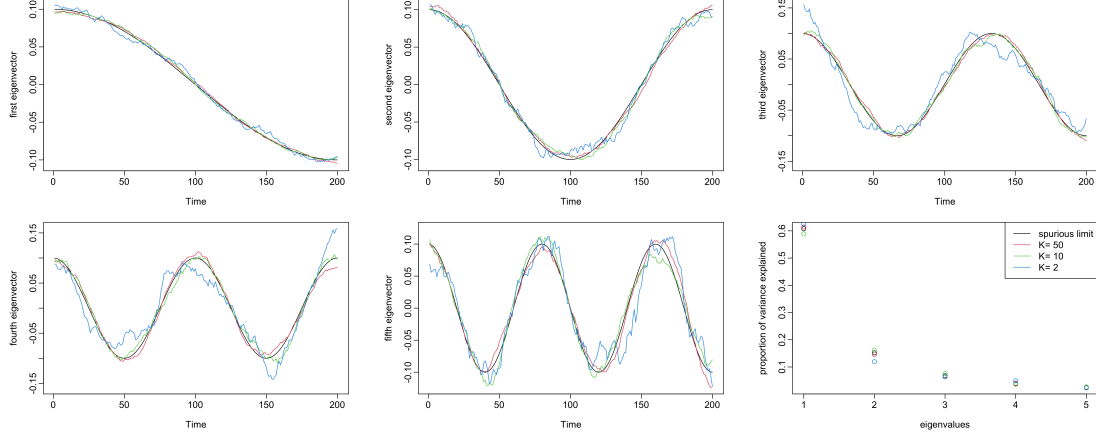


Fig. 1: Setting I - C_ϵ delocalized, loading matrices are of full column rank

This behaviour is in agreement with our theoretical results, since Corollary 1 states that spurious limits must appear whenever C_ϵ is delocalized. The results of Setting II (C_ϵ delocalized with A_n having low effective rank) are visually very similar to Figure 1 and are hence omitted. Most importantly, these limits arise even in the case with only $K = 2$ factors, and in the case where A_n has low effective rank, which marks a clear departure from previous results in He & Zhang (2024) and Onatski & Wang (2021). This illustrates a crucial differences between functional data and high dimensional data, where a large number of basis functions generating the data can act as an additional source of dimensionality.

Figure 2 plots the results for Setting III where the eigenvalues of C_ϵ is fast decaying and the rank of each A_n is equal to K . As can be seen from Figure 2, the sample eigenstructure in Setting III closely resembles the spurious limit for $K = 50$, however, as K decreases, visible deviations from the limit can be observed. This is to be expected, since for Setting III, the first upper bound in (10) is bounded from above by a multiple of K . By Theorem 2 this implies that condition (6) is not satisfied when K is finite. The results for this setting is in line with the findings of Onatski & Wang (2021) and He & Zhang (2024), where a small value of K represents the presence of genuine factors.

Figure 3 plots the results for Setting IV where C_ϵ has decaying eigenvalues and A_n has low effective rank. It can be seen that the sample eigenstructure begins to visibly deviate from the limit, regardless of the value of K . Similar to Setting III, this can be explained by the upper bounds in (10) and Theorem 2. Note that in particular, having K diverge do not seem to guarantee the presence of the spurious limit.

Figures 4 and 5 show the results for Settings V and VI, where C_ϵ is completely localized on two basis functions, and A_n has full rank and low effective rank respectively.

In Figures 4, the case $K = 2$ significantly deviates from the spurious limit. This effect is similar to Setting III but is visibly more pronounced, since the effective rank of C_ϵ , which is an important quantity due to Theorem 2, is even smaller in Setting V. In Setting VI, the sample eigenstructures all deviates from the spurious limit regardless of the value K , similar to Setting IV. The deviations are more visibly pronounced compared to Setting IV due to the decrease in the rank of A_n .

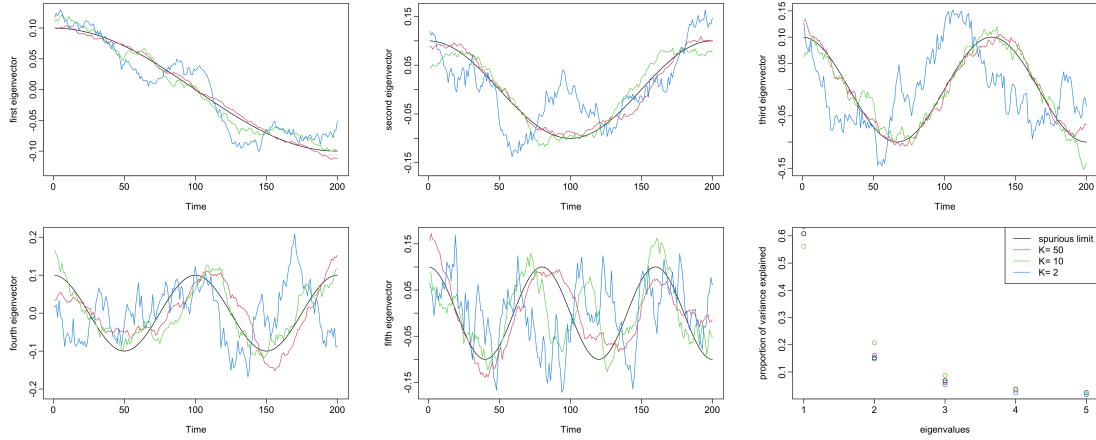


Fig. 2: Setting III - C_ϵ localized, loading matrices are of full column rank

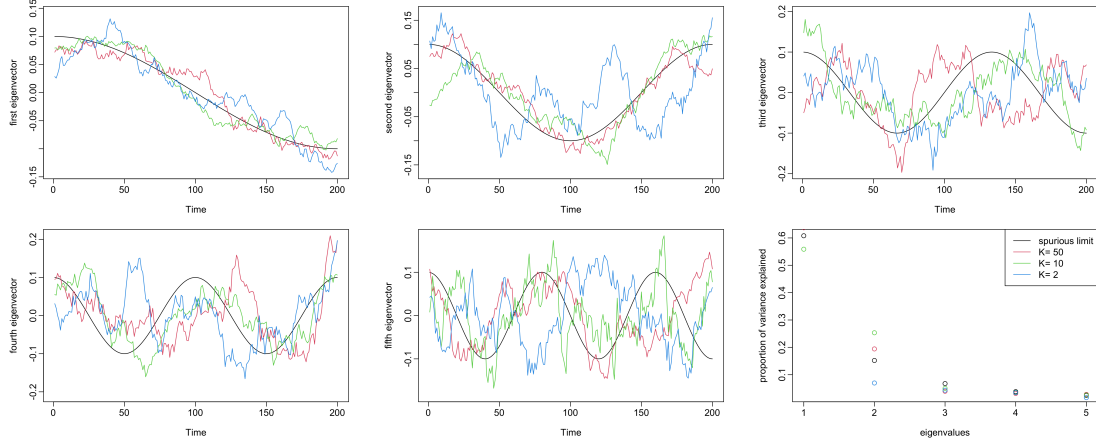


Fig. 3: Setting IV - C_ϵ localized, loading matrices are of low effective rank

5. EMPIRICAL APPLICATIONS

We consider two age-specific mortality rate datasets as an empirical illustration of our results. Our first data set contains age-specific and gender-specific mortality rates obtained from the Human Mortality Database. The dataset consists of mortality rates for $p = 32$ countries observed from 1960 to 2013 ($n = 54$). Since exposures and death counts are sparse at high ages, we aggregate data with age over 100 into a single observation. The raw data is transformed into functional data via smoothed using the `demography` package in R and the logarithm of the mortality rates are computed. This dataset was considered in [Tang et al. \(2022\)](#) where a novel clustering algorithm is proposed based on functional principal component analysis.

From the smoothed log mortality rates, we computed the covariance matrix as defined in (2) and computed its eigenvalues and eigenvectors. Figure 6 plots the first four eigenvectors against the spurious limits as found in Theorem 1. As can be seen, the sample eigenvectors closely resemble the trigonometric curves in the spurious limit. This suggest that the mortality rates is likely non-stationary and the number of non-stationary components in the data, as measured by the effective rank of the model, may be large. In this case, as seen from Theorem 1, the covariance structure is largely driven by the non-stationarity of the data and not the crosssectional dependence

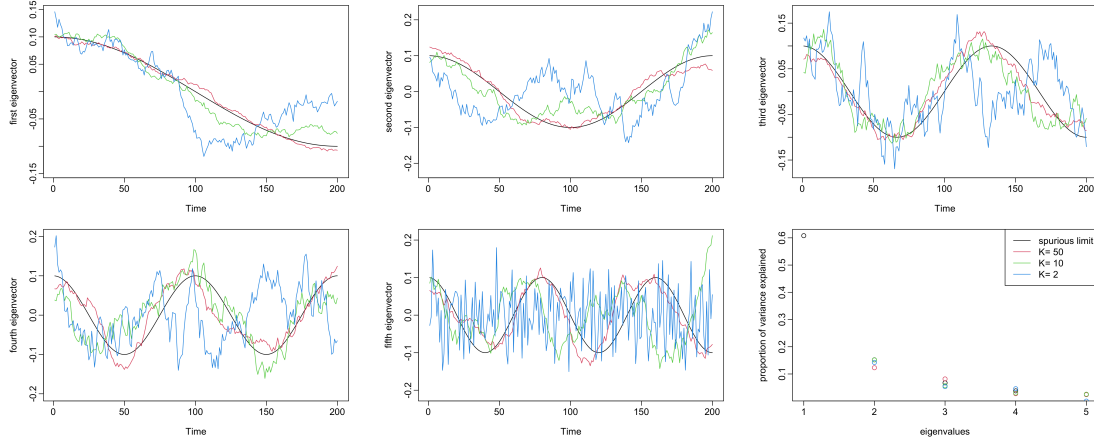


Fig. 4: Setting V - C_ϵ is of rank 2, loading matrices are of full column rank

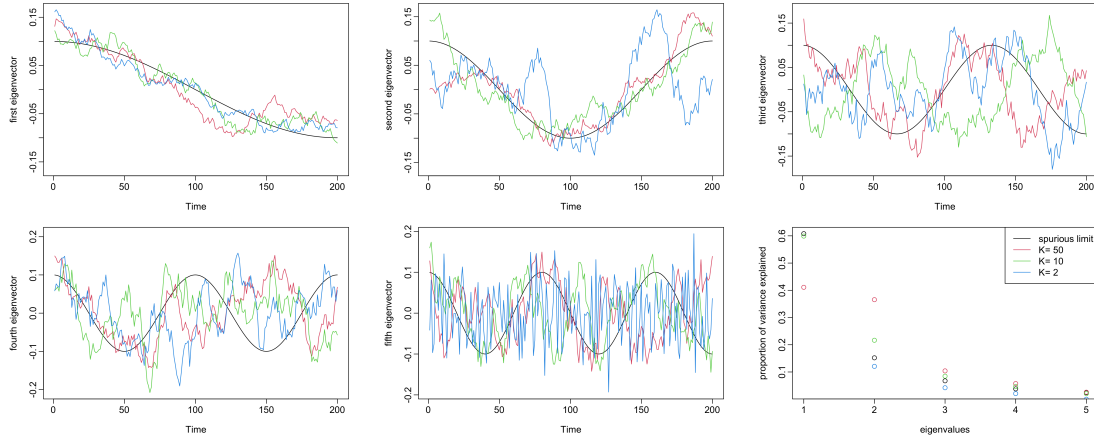


Fig. 5: Setting VI - C_ϵ is of rank 2, loading matrices are of low effective rank

in the data. This raises concerns about the validity of principal component analysis based methods when analysing this dataset without differencing.

For comparison, we also consider the Japanese sub-national mortality rates observed for $p = 47$ prefectures over 1975-2022 ($T = 48$). Since the mortality rates are recorded from different prefectures in the same country, it is reasonable to expect a higher level of homogeneity within the across the 47 covariates. If the crosssectional dependence (factor strength) is strong enough, i.e. if the model has low effective rank in the sense of condition 6, it is possible to observe genuine factors present in the model instead of the spurious limit.

Similar to the multi-country data set mentioned above, we aggregate ages above 100 before smoothing and taking the logarithm. Figure 7 shows the sample eigenvectors of the covariance matrix defined in (2) plotted against the spurious limits of Theorem 1. The eigenvectors still resemble the spurious limits, but the effects are not as visually pronounced as in the multi-country dataset. The second and third sample eigenvector show significant deviation from the trigonometric limits at the year 2011. A possible explanation is the Tohoku Earthquake and Tsunami that took place in March of 2011 which results in nearly 20,000 deaths (Nakahara & Ichikawa, 2013). In this case, the eigenvectors could be thought of as genuine factors that contains relevant information

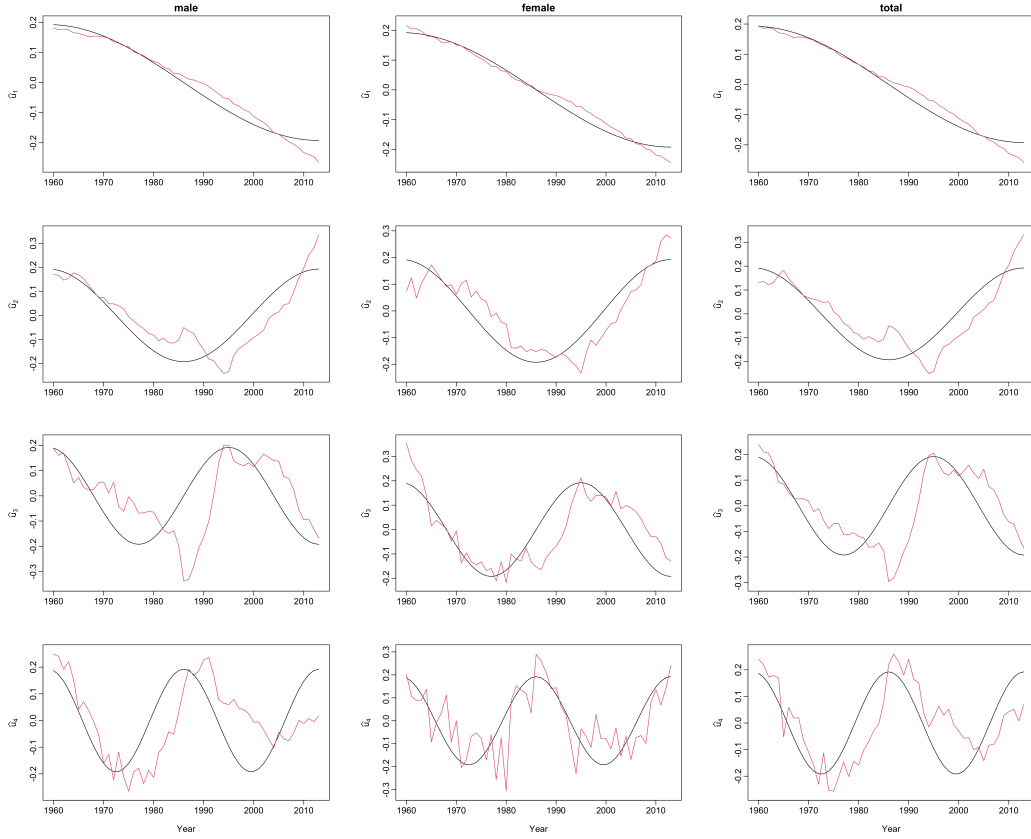


Fig. 6: Sample eigenvectors of the covariance matrix of the male, female and total age-specific mortality rates of 32 countries from 1960 to 2013

on the dataset and instead of being labeled as spurious. Nevertheless, the overall resemblance between the sample eigenvectors and the trigonometric limits suggests that additional care should be taken when dealing with the non-stationary nature of the dataset.

6. CONCLUSION AND FUTURE WORK

We study the asymptotic spectral behavior of high-dimensional functional factor models with non-stationary factors. Under a rank-type condition on the covariance operator, we show that the conventional functional principal component analysis may yield spurious results, thereby extending the findings of [Onatski & Wang \(2021\)](#) to the functional setting. In particular, the sample eigenvalues and eigenfunctions are shown to converge to limits that are unrelated to the true covariance structure. As a consequence, the leading principal components may explain a substantial proportion of the total variation even when the underlying model contains no genuine factors. Moreover, owing to the intrinsic high dimensionality of functional data, we demonstrate that even a model driven by a single non-stationary factor can exhibit spurious behavior. This represents a notable departure from the conclusions of [Onatski & Wang \(2021\)](#) and [He & Zhang \(2024\)](#). Simulation studies are conducted to illustrate the theoretical results and to explore scenarios in which spurious behavior arises. As an empirical application, we analyze two multivariate age-specific mortality datasets and show that they exhibit spurious behavior under the standard functional principal component analysis procedure.

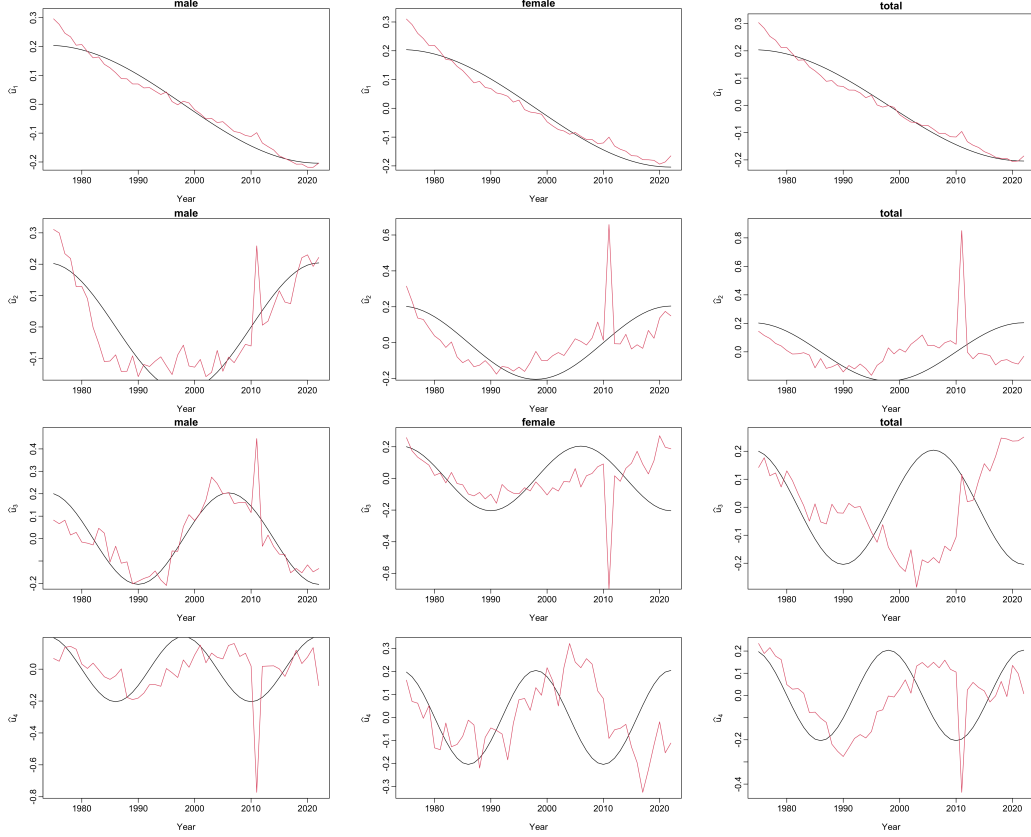


Fig. 7: Sample eigenvectors of the covariance matrix of the male, female and total mortality rates of 47 Japanese prefectures from 1975 to 2022

Our simulation results reveal several intriguing asymptotic features of the sample eigenvectors that are not explained by the existing theory. We intend to formalize these observations and rigorously establish the associated conjecture in future work. Such developments would provide practical tools for detecting spurious behavior in empirical datasets and for identifying key characteristics of common factors, including their effective dimensionality. Another promising direction is to investigate the asymptotic distribution of the sample eigenvalues, in the spirit of [He & Zhang \(2024\)](#). Establishing such results would enable formal inference procedures for testing the presence of high-rank components and for assessing the emergence of spurious behavior in functional factor models.

Lastly, we end the discussion with some interesting observations and a conjecture, which forms the basis of some future research. We observe that for the case $K = 2$ in Setting V (Figure 4), the fifth sample eigenvector seemingly behaves in a significantly different way to the first four. More specifically, the first four sample eigenvectors vaguely resembles the spurious limit and are relatively smooth, while the fifth eigenvector resembles a white noise process which has rough sample paths.

The model in this setting is generated by two basis functions with equal weight, and along the direction of each basis function, the non-stationary factor is loaded by a matrix A_n of rank two. Heuristically, the non-stationary component of the model spans a subspace of dimension four, while the orthogonal complement of this subspace is spanned by the stationary components of the model. This is similar to the concept of the attractor and cointegrating space that has recently

garnered much attention in the functional time series literature (see [Chang et al., 2016](#); [Beare et al., 2017](#); [Li et al., 2023](#)). In this line of work, the dimension of the non-stationary space is determined by model parameters, and the asymptotic behaviour of the eigenstructure differs on the stationary and non-stationary subspaces.

It is therefore natural to conjecture that for models with C_ϵ and A_n being of finite rank, only the first $\|C_\epsilon\|_0\|A_n\|_0$ eigenvectors will tend to the spurious limit, while the rest behave like white noise. This is because the limits in Theorem 1 arise as eigenfunctions of Wiener processes, which are the scaling limits of the non-stationary part of the model. This is confirmed empirically by simulations under a wide range of different settings, we include a simple example to illustrate what typically occurs. The top eight subfigures of Figure 8 plot the sample eigenvectors of a

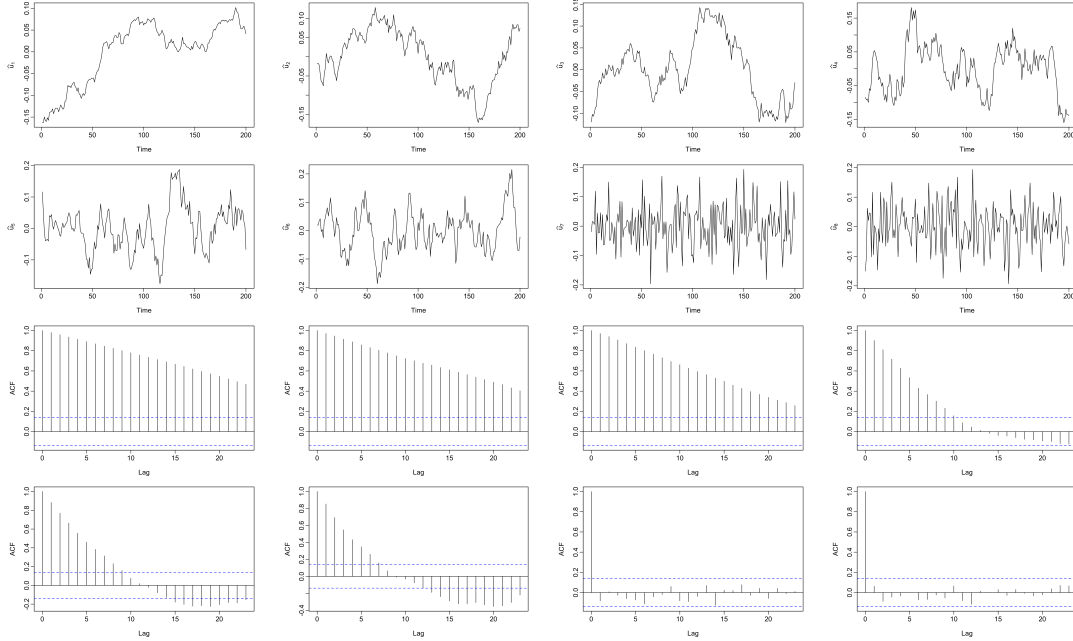


Fig. 8: An eample where C_ϵ is of rank 2 and loading matrices are of rank 3

model with $K = 10$, C_ϵ following Setting (C) and $A_n = G_{p \times 3} \times G_{3 \times K}$, where $G_{m,n}$ denotes a $m \times n$ matrix of i.i.d. standard Gaussians. Clearly the first six eigenvectors behave differently than the last two, which is in line with our conjecture since here the “effective rank” of the model is equal to $\|C_\epsilon\|_0\|A_n\|_0 = 6$. The bottom eight figures plot the sample ACF of each of the sample eigenvectors as a time series of length 200. Clearly the first six eigenvectors exhibit persistent dependence which is reflected in their smoother sample paths, while the last two do not.

A rigorous justification of these observations and conjectures will serve as a valuable tool to identify non-stationary factors in a functional time series model, as well as distinguish spurious behaviour from genuine factors in the model. A much more intricate analysis of the eigenstructure of the model is needed to derive such results, and we will pursue this direction in a standalone work in the future.

SUPPLEMENTARY MATERIAL

A. PROOFS

A.1. Preliminary results

We first give the proofs to Proposition 1.

Proof of Proposition 1. For $f = (f_i)$ and $g = (g_i)$ in $\mathcal{H}^K$, we have

$$\langle \Omega f, g \rangle = \sum_i \langle \sum_j \Omega_{ij} f_j, g_i \rangle = \sum_j \langle f_j, \sum_i \Omega_{ij}^* g_i \rangle = \langle f, (\Omega_{ji}^*)_{ij} g \rangle.$$

Similarly, we have

$$\langle C_1 \Omega C_2 f, g \rangle = \sum_i \langle C_1 \sum_j \Omega_{ij} C_2 f_j, g_i \rangle = \sum_j \langle f_j, \sum_i C_2^* \Omega_{ij}^* C_1^* g_i \rangle$$

which shows $(C_1 \Omega C_2)^* = C_2^* \Omega^* C_1^*$.

Let $(e_n)_n$ be an orthonormal basis for $\mathcal{H}$. Let $(\tilde{e}_n)_n$ be a set of vectors in $\mathcal{H}^K$ given by $\tilde{e}_{(n-1)K+m} := (0, \dots, e_n, \dots, 0)$ where $n = 1, \dots, m = 1, \dots, K$ and e_n appears in the m -th co-ordinate. Clearly $(\tilde{e}_n)$ is an orthonormal basis of $\mathcal{H}^K$. Then we have

$$\sum_n \langle C_1 \Omega C_2 \tilde{e}_n, \tilde{e}_n \rangle = \sum_n \sum_{ij} \langle C_1 \Omega_{ij} C_2 \tilde{e}_n^j, \tilde{e}_n^i \rangle,$$

where $\tilde{e}_n^i$ denotes the i -th co-ordinate of $\tilde{e}_n$. By definition, at least one of $\tilde{e}_n^i$ and $\tilde{e}_n^j$ is 0 whenever $i \neq j$. Therefore

$$\sum_n \langle C_1 \Omega C_2 \tilde{e}_n, \tilde{e}_n \rangle = \sum_i \sum_n \langle C_1 \Omega_{ii} C_2 e_n, e_n \rangle = \sum_i \langle C_1 \Omega_{ii} C_2 \rangle,$$

which gives the trace of $C\Omega$. Finally, we have

$$\|C_1 \Omega C_2\|_F^2 = \langle C_1 \Omega C_2 C_2^* \Omega^* C_1^* \rangle = \sum_i \langle C_1 (\Omega C_2 C_2^* \Omega^*)_{ii} C_1^* \rangle,$$

where the last equality follows from the second claim of the Lemma. Since $(\Omega^*)_{ji} = \Omega_{ij}^*$ by part 1 of the Lemma, we have

$$\|C_1 \Omega C_2\|_F^2 = \sum_{ij} \langle C_1 \Omega_{ij} C_2 C_2^* (\Omega^*)_{ji} C_1^* \rangle = \sum_{ij} \langle C_1 \Omega_{ij} C_2 C_2^* \Omega_{ij}^* C_1^* \rangle = \sum_{ij} \|C_1 \Omega_{ij} C_2\|_F^2$$

where $\|C_1 \Omega_{ij} C_2\|_F$ denotes the Hilbert-Schmidt norm of $C_1 \Omega_{ij} C_2$ on $\mathcal{H}$. $\square$

The following lemma is the key technical result underpinning Theorem 1. It computes moments of certain bilinear forms associated with the random matrix W appearing in (4).

LEMMA A1. *Let W be the $T \times T$ random matrix defined in (4). Under Assumption 1, for any deterministic vectors $a, b \in \mathbb{R}^T$, we have*

$$\mathbb{E}[a' W b] = \langle a, b \rangle \langle C_\epsilon \Omega \rangle, \quad (\text{A1})$$

Furthermore, for any deterministic vectors $a, b, c, d \in \mathbb{R}^T$, we have

$$\begin{aligned} & \left| \text{cov}(a' W b, c' W d) - \langle a, c \rangle \langle b, d \rangle \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2 - \langle a, d \rangle \langle b, c \rangle \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2 \right| \\ & \leq 2\kappa_4^* \sum_t a_t b_t c_t d_t \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2, \end{aligned}$$

where a_t, b_t, c_t and d_t are the t -th coordinate of a, b, c and d respectively.

Proof. From the definition of W we may write

$$a' W b = \sum_{i,j=1}^K \iint \langle a, \epsilon_i(v) \rangle \langle b, \epsilon_j(v) \rangle \Omega_{ij}(v, w) dv dw,$$

where $\epsilon_{i\cdot}(v) \in \mathbb{R}^T$ i -th column of the matrix $\epsilon'(v)$. Since $(\epsilon_{it})_{it}$ is a matrix of uncorrelated elements by assumptions, we have

$$\begin{aligned}\mathbb{E}[a'Wb] &= \sum_{i,j=1}^K \iint \sum_{s,t=1}^T \mathbb{E}[a_s \epsilon_{is}(v) b_t \epsilon_{jt}(w)] \Omega_{ij}(v, w) dv dw \\ &= \sum_{t=1}^T a_t b_t \sum_{i=1}^K \iint \mathbb{E}[\epsilon_{it}(v) \epsilon_{it}(w)] \Omega_{ii}(v, w) dv dw \\ &= \langle a, b \rangle \sum_{i=1}^K \iint C_\epsilon(v, w) \Omega_{ii}(v, w) dv dw.\end{aligned}\tag{A2}$$

By Proposition 1 we obtain

$$\mathbb{E}[a'Wb] = \sum_{i=1}^K \langle C_\epsilon \Omega_{ii} \rangle = \langle a, b \rangle \langle C_\epsilon \Omega \rangle.$$

This gives the equality in (A1). For second moments, we write

$$\mathbb{E}[a'Wb, c'Wd] = \sum_{i,j,k,l=1}^K \iiint \iint A_{ijkl}(v, w, \alpha, \beta) \Omega_{ij}(v, w) \Omega_{kl}(\alpha, \beta) dv dw d\alpha d\beta.$$

where we have defined

$$A_{ijkl}(v, w, \alpha, \beta) := \mathbb{E}[\langle a, \epsilon_{i\cdot}(v) \rangle \langle b, \epsilon_{j\cdot}(w) \rangle \langle c, \epsilon_{k\cdot}(\alpha) \rangle \langle d, \epsilon_{l\cdot}(\beta) \rangle].$$

We will first compute A_{ijkl} and $\sum_{ijkl} \Omega_{ij}(v, w) \Omega_{kl}(\alpha, \beta)$ for all the possible partitions of index set $\{i, j, k, l\}$. For ease of notation we will write

$$\tilde{\Omega} := \sum_i \Omega_{ii}(v, w) \Omega_{ii}(\alpha, \beta).$$

On the set $\{i = j \neq k = l\}$, we have

$$\begin{aligned}A_{iikk}(v, w, \alpha, \beta) &= \mathbb{E}[\langle a, \epsilon_{i\cdot}(v) \rangle \langle b, \epsilon_{i\cdot}(w) \rangle] \mathbb{E}[\langle c, \epsilon_{k\cdot}(\alpha) \rangle \langle d, \epsilon_{k\cdot}(\beta) \rangle] \\ &= \langle a, b \rangle \langle c, d \rangle C_\epsilon(v, w) C_\epsilon(\alpha, \beta),\end{aligned}$$

where the last line follows from similar computations as in (A2), and

$$\begin{aligned}\sum_{i=j \neq k=l} \Omega_{ij}(v, w) \Omega_{kl}(\alpha, \beta) &= \sum_{i \neq k} \Omega_{ii}(v, w) \Omega_{kk}(\alpha, \beta) \\ &= \text{tr } \Omega(v, w) \text{tr } \Omega(\alpha, \beta) - \tilde{\Omega}.\end{aligned}$$

Similarly, on the set $\{i = k \neq j = l\}$, we have

$$\begin{aligned}A_{ijij}(v, w, \alpha, \beta) &= \mathbb{E}[\langle a, \epsilon_{i\cdot}(v) \rangle \langle c, \epsilon_{i\cdot}(\alpha) \rangle] \mathbb{E}[\langle b, \epsilon_{j\cdot}(w) \rangle \langle d, \epsilon_{j\cdot}(\beta) \rangle] \\ &= \langle a, c \rangle \langle b, d \rangle C_\epsilon(v, \alpha) C_\epsilon(w, \beta),\end{aligned}$$

and

$$\begin{aligned}\sum_{i=k \neq j=l} \Omega_{ij}(v, w) \Omega_{kl}(\alpha, \beta) &= \sum_{i \neq j} \Omega_{ij}(v, w) \Omega_{ij}(\alpha, \beta) \\ &= \text{tr } \Omega(v, w)' \Omega(\alpha, \beta) - \tilde{\Omega}.\end{aligned}$$

On the set $\{i = l \neq j = k\}$, we have

$$\begin{aligned} A_{ijji}(v, w, \alpha, \beta) &= \mathbb{E}[\langle a, \epsilon_i(v) \rangle \langle d, \epsilon_i(\beta) \rangle] \mathbb{E}[\langle b, \epsilon_j(w) \rangle \langle c, \epsilon_j(\alpha) \rangle] \\ &= \langle a, d \rangle \langle b, c \rangle C_\epsilon(v, \beta) C_\epsilon(w, \alpha), \end{aligned}$$

and

$$\begin{aligned} \sum_{i=l \neq j=k} \Omega_{ij}(v, w) \Omega_{kl}(\alpha, \beta) &= \sum_{i \neq j} \Omega_{ij}(v, w) \Omega_{ji}(\alpha, \beta) \\ &= \text{tr } \Omega(v, w) \Omega(\alpha, \beta) - \tilde{\Omega}. \end{aligned}$$

In cases where exactly three of the four indices coincide we have $A_{i,j,k,l=0}$ since $\mathbb{E}[\epsilon = 0]$. Finally on the diagonal $\{i = j = k = l\}$, we have

$$A_{iiii}(v, w, \alpha, \beta) = \mathbb{E}[\langle a, \epsilon_i(v) \rangle \langle b, \epsilon_i(w) \rangle \langle c, \epsilon_i(\alpha) \rangle \langle d, \epsilon_i(\beta) \rangle]$$

and

$$\sum_{i=j=k=l} \Omega_{ij}(v, w) \Omega_{kl}(\alpha, \beta) = \sum_i \Omega_{ii}(v, w) \Omega_{ii}(\alpha, \beta) = \tilde{\Omega}.$$

Expanding the expression for A_{iiii} and partitioning the index set in a similar way, we get

$$\begin{aligned} A_{iiii}(v, w, \alpha, \beta) &= \sum_{s,t,m,n} \mathbb{E}[a_s b_t c_m d_n \epsilon_{is}(v) \epsilon_{it}(w) \epsilon_{im}(\alpha) \epsilon_{in}(\beta)] \\ &= \left(\sum_{s=t \neq m=n} + \sum_{s=m \neq t=n} + \sum_{s=n \neq t=m} + \sum_{s=t=m=n} \right) \mathbb{E}[a_s b_t c_m d_n \epsilon_{is}(v) \epsilon_{it}(w) \epsilon_{im}(\alpha) \epsilon_{in}(\beta)], \end{aligned}$$

since $\mathbb{E}[\epsilon_{is} \epsilon_{it} \epsilon_{im} \epsilon_{in}] = 0$ on sets of exactly three coinciding indices. Therefore

$$\begin{aligned} A_{iiii}(v, w, \alpha, \beta) &= \langle a, b \rangle \langle c, d \rangle C_\epsilon(v, w) C_\epsilon(\alpha, \beta) \\ &\quad + \langle a, c \rangle \langle b, d \rangle C_\epsilon(v, \alpha) C_\epsilon(w, \beta) \\ &\quad + \langle a, d \rangle \langle b, c \rangle C_\epsilon(v, \beta) C_\epsilon(w, \alpha) \\ &\quad + \sum_t \kappa_{it}(v, w, \alpha, \beta) a_t b_t c_t d_t \end{aligned}$$

where we have defined

$$\kappa_{it}(v, w, \alpha, \beta) := \mathbb{E}[\epsilon_{it}(v) \epsilon_{it}(w) \epsilon_{it}(\alpha) \epsilon_{it}(\beta) - 3].$$

Recall from (A1) that

$$\mathbb{E}[a' W b] = \langle a, b \rangle \langle C_\epsilon \Omega \rangle, \quad \mathbb{E}[c' W d] = \langle c, d \rangle \langle C_\epsilon \Omega \rangle.$$

After collecting the above terms and simplifying, we note that all terms containing the quantity $\tilde{\Omega}$ cancel out and we obtain

$$\begin{aligned} \text{cov}(a' W b, c' W d) &= \langle a, c \rangle \langle b, d \rangle \int_{\mathcal{I}^4} C_\epsilon(v, \alpha) C_\epsilon(w, \beta) \text{tr } \Omega(v, w)' \Omega(\alpha, \beta) dv dw d\alpha d\beta \\ &\quad + \langle a, d \rangle \langle b, c \rangle \int_{\mathcal{I}^4} C_\epsilon(v, \beta) C_\epsilon(w, \alpha) \text{tr } \Omega(v, w) \Omega(\alpha, \beta) dv dw d\alpha d\beta \\ &\quad + \sum_t a_t b_t c_t d_t \sum_i \int_{\mathcal{I}^4} \kappa_{it}(v, w, \alpha, \beta) \Omega_{ii}(v, w) \Omega_{ii}(\alpha, \beta) dv dw d\alpha d\beta. \end{aligned}$$

Recall that C_ϵ is self-adjoint. By definition of Ω and Proposition 1, we have

$$\Omega_{ij}(u, v) = \Omega_{ji}^*(u, v) = \Omega_{ji}(v, u). \quad (\text{A3})$$

By linearity, the first term in the above expression simplifies to

$$\begin{aligned}
& \int C_\epsilon(v, \alpha) C_\epsilon(w, \beta) \text{tr } \Omega(v, w)' \Omega(\alpha, \beta) dv dw d\alpha d\beta \\
&= \sum_{i,j} \int_{\mathcal{I}^4} C_\epsilon(\beta, w) \Omega_{ij}(w, v) C_\epsilon(v, \alpha) \Omega_{ji}(\alpha, \beta) dv dw d\alpha d\beta \\
&= \sum_{i,j} \langle C_\epsilon \Omega_{ij} C_\epsilon \Omega_{ji} \rangle
\end{aligned}$$

Using (A3) and the self-adjointness of C_ϵ again, we obtain

$$\begin{aligned}
\sum_{i,j} \langle C_\epsilon \Omega_{ij} C_\epsilon \Omega_{ji} \rangle &= \sum_{i,j} \langle C_\epsilon^{1/2} \Omega_{ij} C_\epsilon^{1/2} C_\epsilon^{1/2} \Omega_{ij}^* C_\epsilon^{1/2} \rangle \\
&= \sum_{i,j} \|C_\epsilon^{1/2} \Omega_{ij} C_\epsilon^{1/2}\|_F^2 = \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2,
\end{aligned}$$

Similarly we have

$$\int_{\mathcal{I}^4} C_\epsilon(v, \beta) C_\epsilon(w, \alpha) \text{tr } \Omega(v, w) \Omega(\alpha, \beta) dv dw d\alpha d\beta = \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2.$$

Therefore we have

$$\begin{aligned}
& \text{cov}(a'Wb, c'Wd) - (\langle a, c \rangle \langle b, d \rangle + \langle a, d \rangle \langle b, c \rangle) \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2 \\
&= \sum_t a_t b_t c_t d_t \sum_i \int_{\mathcal{I}^4} \kappa_{it}(v, w, \alpha, \beta) \Omega_{ii}(v, w) \Omega_{ii}(\alpha, \beta) dv dw d\alpha d\beta
\end{aligned}$$

and it remains to bound the last term. By Assumption 1.2, we have

$$\begin{aligned}
& \left| \sum_i \int_{\mathcal{I}^4} \kappa_{it}(v, w, \alpha, \beta) \Omega_{ii}(v, w) \Omega_{ii}(\alpha, \beta) dv dw d\alpha d\beta \right| \\
&\leq \sum_i \int_{\mathcal{I}^4} C_\epsilon(\beta, v) \Omega_{ii}(v, w) C_\epsilon(w, \alpha) \Omega_{ii}(\alpha, \beta) dv dw d\alpha d\beta \\
&= \sum_i \langle C_\epsilon \Omega_{ii} C_\epsilon \Omega_{ii} \rangle.
\end{aligned}$$

By Proposition 1.2, we have

$$\sum_i \langle C_\epsilon \Omega_{ii} C_\epsilon \Omega_{ii} \rangle = \sum_i \|C_\epsilon^{1/2} \Omega_{ii} C_\epsilon^{1/2}\|_F^2 \leq \sum_{i,j} \|C_\epsilon^{1/2} \Omega_{ij} C_\epsilon^{1/2}\|_F^2 = \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2,$$

where the last equality follows from Proposition 1.3. $\square$

The following lemma gives the spectral decomposition of the matrix $M\Theta'$.

LEMMA A2. *The matrix $M\Theta'$ admits a singular value decomposition*

$$M\Theta' = \sum_{t=1}^T \sigma_t w_t v_t',$$

where $\sigma_t = (2 \sin(\frac{t\pi}{2T}))^{-1}$ and

$$w_{tn} = -\sqrt{\frac{2}{T}} \cos\left(\frac{(n-1/2)\pi t}{T}\right), \quad v_{tn} = \sqrt{\frac{2}{T}} \sin\left(\frac{(n-1)\pi t}{T}\right)$$

for $t < T$, and $\sigma_T = 0$, $w_{Tn} = 1/\sqrt{T}$ and $v_{Tn} = 1_{n=1}$. Furthermore, we have $\sigma_n \sim T$ for any fixed n as $T \rightarrow \infty$.

Proof. The spectral decomposition follows from Lemma 5 of [Onatski & Wang \(2021\)](#). For the last estimate, note that $\sin(\frac{n\pi}{2T}) \approx \frac{n\pi}{2T}$ as $T \rightarrow \infty$, so $\sigma_n \sim T$. $\square$

Using Lemma [A1](#) and Lemma [A2](#), we may derive the following concentration bounds for certain quadratic forms of W .

LEMMA A3. *Let $(\sigma_i)_i, (v_i)_i$ be as defined in Lemma [A2](#). Under Assumptions [1-3](#), for any $i \leq j \leq T$, we have the following estimate on the quadratic form of W :*

$$v_i' W v_j = \delta_{ij} \langle C_\epsilon \Omega \rangle + \langle C_\epsilon \Omega \rangle_{op}(1).$$

Furthermore, we have

$$\sum_{k=i}^j \sigma_k^2 v_k' W v_k = \langle C_\epsilon \Omega \rangle \sum_{k=i}^j \sigma_k^2 + \langle C_\epsilon \Omega \rangle_{op}(T^2).$$

Proof. Since $\langle v_i, v_j \rangle = \delta_{ij}$, by Lemma [A1](#) we have $\mathbb{E}[v_i' W v_j] = \delta_{ij} \langle C_\epsilon \Omega \rangle$. Furthermore,

$$\begin{aligned} \text{var}(v_i' W v_j) &= (1 + \delta_{ij}) \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2 + O(2\kappa_4^* \sum_t v_{it}^2 v_{jt}^2 \|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2) \\ &= O(\|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F^2), \end{aligned}$$

where the last estimate holds since $\sum_t v_{it}^2 v_{jt}^2 \leq \sum_{s,t} v_{it}^2 v_{js}^2 = 1$. By Chebyshev's inequality

$$v_i' W v_j = \delta_{ij} \langle C_\epsilon \Omega \rangle + O_p(\|C_\epsilon^{1/2} \Omega C_\epsilon^{1/2}\|_F)$$

and the first claim follows from Assumption [2](#). For the second estimate, recall from Lemma [A2](#) that $\sigma_k^{-1} \sim \sin(k/T)$ as $T \rightarrow \infty$. Clearly this implies $\sigma_k \sim T$. Since i, j are fixed, this implies $\sum_{k=i}^j \sigma_k^2 = O(T^2)$ and the second claim follows. $\square$

The following algebraic identity from Lemma 3.3 of [Yin et al. \(2014\)](#) gives a geometric interpretation on the ratio between ℓ^1 and ℓ^2 norms of a vector.

LEMMA A4. *Let $x \in \mathbb{R}^n$. Then*

$$\frac{2\alpha(x)^{1/2}}{1 + \alpha(x)} \|x\|_0^{1/2} \leq \frac{\|x\|_1}{\|x\|_2} \leq \|x\|_0^{1/2}$$

where $\|x\|_0 := |\{i \leq n, x_i \neq 0\}|$ and $\alpha(x)$ is given by

$$\alpha(x) := \frac{\inf_{i \leq n, x_i \neq 0} x_i}{\sup_{i \leq n, x_i \neq 0} x_i}.$$

Proof. This is Lemma 3.3 of [Yin et al. \(2014\)](#). $\square$

Finally, Lemma [A5](#) is a linear algebraic result used to compute the spectrums of certain operators.

LEMMA A5. *Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and p, q, T be positive integers. Suppose $f_1, \dots, f_T$ and $g_1, \dots, g_T$ are elements of $\mathcal{H}^p$ with components $f_t = (f_{1t}, \dots, f_{pt})$ and $g_t = (g_{1t}, \dots, g_{pt})$ respectively. Let A be the $T \times T$ matrix given by*

$$A_{st} = \sum_{i=1}^p \langle f_{is}, g_{it} \rangle, \quad s, t = 1, \dots, T$$

and $\mathcal{A}$ be the compact operator on $\mathcal{H}^p$ given by

$$\mathcal{A} = \sum_{t=1}^T f_t \otimes g_t.$$

Then A and $\mathcal{A}$ share the same non-zero singular values. Furthermore, we have

$$\|A\|^2 \leq \|A_f\| \|A_g\|,$$

where A_f and A_g are the matrices whose entries are given by

$$(A_f)_{st} = \sum_{i=1}^p \langle f_{is}, f_{it} \rangle, \quad (A_g)_{st} = \sum_{i=1}^p \langle g_{is}, g_{it} \rangle.$$

Proof. Let $F : \mathbb{R}^T \rightarrow \mathcal{H}^p$ be the linear operator given by

$$Fu = \sum_{t=1}^T u_t f_t, \quad u = (u_1, \dots, u_T)' \in \mathbb{R}^T.$$

For $h = (h_1, \dots, h_p) \in \mathcal{H}^p$, since

$$\langle h, Fu \rangle = \sum_{i=1}^p \sum_{t=1}^T \langle h_i, u_t f_{it} \rangle = \sum_{t=1}^T u_t \langle h, f_t \rangle,$$

the adjoint of F is given by $F^* : \mathcal{H}^p \rightarrow \mathbb{R}^T$

$$F^*h = (\langle h, f_1 \rangle, \dots, \langle h, f_T \rangle)', \quad h \in \mathcal{H}^p.$$

Similarly, define $G : \mathcal{H}^p \rightarrow \mathbb{R}^T$ by

$$Gu = \sum_{t=1}^T u_t g_t, \quad u \in \mathbb{R}^T$$

whose adjoint is given by

$$G^*h = (\langle h, g_1 \rangle, \dots, \langle h, g_T \rangle)', \quad h \in \mathcal{H}^p.$$

Let $(e_t)_{t=1, \dots, T}$ be the standard basis of $\mathbb{R}^T$, then we have

$$\langle e_s, F^*G e_t \rangle = \langle F e_s, G e_t \rangle = \langle f_s, g_t \rangle = A_{st}$$

the matrix A can be written as $A = F^*G$. Similarly, we have $\mathcal{A} = FG^*$ since

$$\langle h_1, FG^*h_2 \rangle = \langle F^*h_1, G^*h_2 \rangle = \sum_{t=1}^T \langle h_1, f_t \rangle \langle h_2, g_t \rangle = \langle h_1, \mathcal{A}h_2 \rangle$$

for $h_1, h_2 \in \mathcal{H}^p$. It remains to observe that $A = F^*G$ shares the same non-zero singular values as $A^* = G^*F$ which shares the same non-zero singular values as $\mathcal{A} = FG^*$. This gives the first claim. The second claim follows from the inequality

$$\|A\|^2 \leq \|F\|^2 \|G\|^2 = \|FF^*\| \|GG^*\|$$

and the fact that $A_f = FF^*$ and $A_g = GG^*$. $\square$

A.2. Proof of the main results

Proof of Theorem 1. First we show that under Assumption 1-3, Theorem 1 holds for the matrix $\hat{S}$ whenever it holds for the covariance matrix of the non-stationary part of the model $\tilde{S}$, i.e. the perturbation of the stationary error term ζ is asymptotically negligible.

We will show below that the spectral gaps of $\tilde{S}$ are asymptotically of size

$$|\lambda_k(\tilde{S}) - \lambda_{k+1}(\tilde{S})| \sim \|\tilde{S}\| \sim T^2 p^{-1} \langle C_\epsilon \Omega \rangle. \quad (\text{A4})$$

By standard perturbation theory (Hsing & Eubank (2015); Kato (2013)), to prove that the perturbation of ζ is asymptotically negligible, it suffices to show that $\|\hat{S} - \tilde{S}\| = o_p(T^2 p^{-1} \langle C_\epsilon \Omega \rangle)$. This is what we show now. Since $X(u) := (\Psi F)(u) + \zeta(u)$, by the triangle inequality we have

$$\|\hat{S} - \tilde{S}\| \lesssim \left\| \frac{1}{p} \int_{\mathcal{I}} M \zeta(u)' \zeta(u) M du \right\| + \left\| \frac{1}{p} \int_{\mathcal{I}} M (\Psi F)(u)' \zeta(u) M du \right\|. \quad (\text{A5})$$

where M denotes the orthogonal projection onto the orthogonal complement of the T -dimensional vector of ones, i.e. $M = I_T - T^{-1} 1_T 1_T'$. We will use the fact that $\|M\| = 1$ throughout the rest of the proof without mention.

For the first term in (A5), by Lemma A5, we have

$$\left\| \frac{1}{p} \int_{\mathcal{I}} M \zeta(u)' \zeta(u) M du \right\| \leq \frac{1}{p} \left\| \sum_{t=1}^T \zeta_t \otimes \zeta_t \right\| \leq \frac{1}{p} \sum_{t=1}^T \|\zeta_t\|_{\mathcal{H}^p}^2.$$

Since $\mathbb{E}[\|\zeta_t\|^2] = \langle C_\zeta \rangle$, by Markov's inequality we have

$$\left\| \frac{1}{p} \int_{\mathcal{I}} M \zeta(u)' \zeta(u) M du \right\| = O_p(T p^{-1} \langle C_\zeta \rangle) = o_p(T^2 p^{-1} \langle C_\epsilon \Omega \rangle), \quad (\text{A6})$$

where the last estimate follows from Assumption 3. Similarly, by the second claim of Lemma A5, the second term in (A5) is bounded by

$$\begin{aligned} \left\| \frac{1}{p} \int_{\mathcal{I}} M (\Psi F)(u)' \zeta(u) M du \right\| &\leq \left\| \frac{1}{p} \int_{\mathcal{I}} M (\Psi F)(u)' (\Psi F)(u) M du \right\|^{1/2} \left\| \frac{1}{p} \int_{\mathcal{I}} M \zeta(u)' \zeta(u) M du \right\|^{1/2} \\ &= \|\tilde{S}\|^{1/2} \left\| \frac{1}{p} \int_{\mathcal{I}} M \zeta(u)' \zeta(u) M du \right\|^{1/2} \\ &= o_p(T^2 p^{-1} \langle C_\epsilon \Omega \rangle), \end{aligned}$$

where the last estimate follows from (A6) and the bound

$$\|\tilde{S}\| = O_p(T^2 p^{-1} \langle C_\epsilon \Omega \rangle) \quad (\text{A7})$$

which will be shown below. From this we conclude that $\|\tilde{S} - \hat{S}\| = o_p(T^2 p^{-1} \langle C_\epsilon \Omega \rangle)$ which shows that the perturbation of ζ is asymptotically negligible.

It remains to show that Theorem 1 holds for the covariance operator $\tilde{S}$, which also proves the two estimated (A4) and (A7) we used above. This proof is adapted from the arguments in Section A.3.1 of Onatski & Wang (2021). We sketch the key components of the argument adapted to the functional setting of our model.

Let $\tilde{\lambda}_n, \tilde{u}_n, n = 1, \dots, T$ be the eigenvalues of $\tilde{S}$ arranged in non-ascending order and the corresponding eigenvectors. We will focus on the largest eigenvalues $\tilde{\lambda}_1$ and the corresponding eigenvector $\tilde{u}_1$, the general case can be obtained via mathematical induction. We omit the details of the induction argument since it can be found on p.610-611 of Onatski & Wang (2021).

Recall from Lemma A2 that the set $\{w_n, n = 1, \dots, T\}$ forms an orthonormal basis of $\mathbb{R}^T$ and w_T is proportional to a vector of ones. Since $\tilde{u}_1$ is orthogonal to w_T by construction, we have the representation

$$\tilde{u}_1 = \sum_{t=1}^{T-1} \alpha_t w_t$$

for some set of weights (α_t) . Note that since (w_t) is an orthonormal set, we have

$$1 = \|\tilde{u}_1\|^2 = \left\| \sum_{t=1}^{T-1} \alpha_t w_t \right\|^2 = \sum_{t=1}^{T-1} \alpha_t^2 \|w_t\|^2 = \sum_{t=1}^{T-1} \alpha_t^2.$$

By definition of $\tilde{\lambda}_1, \tilde{u}_1, \tilde{S}$ we have

$$\tilde{\lambda}_1 = \tilde{u}_1' \tilde{S} \tilde{u}_1 = \sum_{s,t=1}^{T-1} \alpha_s \alpha_t w_s' \tilde{S} w_t.$$

Since $\tilde{S} = p^{-1} M \Theta' W \Theta M$ where $M \Theta' = \sum_t \sigma_t w_t v_t'$ by Lemma A2, we have

$$p \tilde{\lambda}_1 = \sum_{s,t=1}^{T-1} \alpha_s \alpha_t \tilde{\sigma}_s \tilde{\sigma}_t v_s' W v_t.$$

The quadratic form $v_s' W v_t$ can be estimated using Lemma A3, however, the sum of $T - 1$ such terms has to be dealt with additional care since $T \rightarrow \infty$. Using a truncation argument similar to the one found in p.607-p.609 of Onatski & Wang (2021), we may show that

$$p \hat{\lambda}_1 \leq \langle C_\epsilon \Omega \rangle \left(\sum_{t=1}^{T-1} \alpha_t^2 \tilde{\sigma}_t^2 + o_p(T^2) \right).$$

Since $\tilde{\sigma}_1$ is the leading eigenvalue, and $\sum_{t=1}^{T-1} \alpha_t^2 = 1$, we have

$$\sum_{t=1}^{T-1} \alpha_t^2 \tilde{\sigma}_t^2 = \alpha_1^2 \tilde{\sigma}_1^2 + \sum_{t=2}^{T-1} \alpha_t^2 \tilde{\sigma}_t^2 \leq \alpha_1^2 \tilde{\sigma}_1^2 + (1 - \alpha_1^2) \tilde{\sigma}_2^2,$$

which gives

$$p \hat{\lambda}_1 \leq \alpha_1^2 \tilde{\sigma}_1^2 \langle C_\epsilon \Omega \rangle + (1 - \alpha_1^2) \tilde{\sigma}_2^2 \langle C_\epsilon \Omega \rangle + o_p(T^2) \langle C_\epsilon \Omega \rangle.$$

On the other hand, since $\tilde{\lambda}_1$ is the leading eigenvalues, we have

$$p \tilde{\lambda}_1 \geq p w_1' \tilde{S} w_1 = \sigma_1 v_1' W v_1 = \sigma_1^2 \langle C_\epsilon \Omega \rangle + o_p(T^2) \langle C_\epsilon \Omega \rangle.$$

Combining the upper and lower bounds and simplifying we obtain

$$(\sigma_1^2 - \sigma_2^2)(1 - \alpha_1^2) \langle C_\epsilon \Omega \rangle \leq o_p(T^2) \langle C_\epsilon \Omega \rangle.$$

Using the fact that $\sigma_1^2 - \sigma_2^2 \sim T^2$, we obtain $1 - \alpha_1^2 = o_p(1)$, which implies $\alpha_1^2 = (\tilde{u}_1' w_1)^2 \rightarrow 1$ in probability as $T \rightarrow \infty$. This gives the first claim.

Using the upper and lower bound again, we may obtain

$$|p \tilde{\lambda}_1 - \sigma^2 \langle C_\epsilon \Omega \rangle| \leq |1 - \alpha_1^2| (\sigma_1^2 + \sigma_2^2) \langle C_\epsilon \Omega \rangle + o_p(T^2) \langle C_\epsilon \Omega \rangle = o_p(T^2) \langle C_\epsilon \Omega \rangle,$$

since $\sigma_1^2 \sim T^2/\pi^2$ and $1 - \alpha_1^2 = o_p(1)$. This gives

$$\tilde{\lambda}_1 = \frac{T^2}{\pi^2 p} \langle C_\epsilon \Omega \rangle (1 + o_p(1)),$$

which the asymptotic limit of the eigenvalues. This also implies the bounds (A4) and (A7). Finally, the last claim follows from the same arguments as p.609-610 of Onatski & Wang (2021) and we omit the details. $\square$

Proof of Theorem 2. Using (5) and (9), the operator Ω can be written as

$$\begin{aligned}\Omega_{kl}(v, w) &= \sum_{i=1}^p \int_{\mathcal{I}} \Psi_{ik}(u, v) \Psi_{il}(u, w) du \\ &= \sum_{i=1}^p \sum_{n=1}^{\infty} a_{nik} a_{nil} \phi_n(v) \phi_n(w) \\ &= \sum_{n=1}^{\infty} (A'_n A_n)_{kl} \phi_n(v) \phi_n(w).\end{aligned}$$

Since $C_\epsilon^{1/2} = \sum_n c_n^{1/2} \phi_n \otimes \phi_n$, we have

$$\begin{aligned}(C_\epsilon^{1/2} \Omega_{kl} C_\epsilon^{1/2})(u, v) &= \int_{\mathcal{I}} C_\epsilon^{1/2}(u, \alpha) \Omega_{kl}(\alpha, \beta) C_\epsilon^{1/2}(\beta, v) d\alpha d\beta \\ &= \sum_{n=1}^{\infty} c_n (A'_n A_n)_{kl} \phi_n(u) \phi_n(v).\end{aligned}$$

By Proposition 1, the trace of $\mathcal{C}$ is given by

$$\langle \mathcal{C} \rangle = \sum_{k=1}^K \langle C_\epsilon \Omega_{kk} \rangle = \sum_{k=1}^K \sum_{n=1}^{\infty} c_n (A'_n A_n)_{kk} = \sum_{n=1}^{\infty} c_n \langle B_n \rangle.$$

Similarly, the Hilbert-Schmidt norm of $\mathcal{C}$ is given by

$$\|\mathcal{C}\|_2^2 = \sum_{kl} \|C_\epsilon^{1/2} \Omega_{kl} C_\epsilon^{1/2}\|^2 = \sum_{n=1}^{\infty} \sum_{kl} c_n^2 (A'_n A_n)_{kl}^2 = \sum_{n=1}^{\infty} c_n^2 \|B_n\|_2^2.$$

Since $1 = \langle C_\epsilon \rangle = \sum_n c_n$ and $\|C_\epsilon\|_2^2 = \sum_n c_n^2$, we have the following upperbound

$$\begin{aligned}\frac{\langle \mathcal{C} \rangle^2}{\|\mathcal{C}\|_2^2} &= \limsup_{q \rightarrow \infty} \frac{(\sum_{n=1}^q c_n \langle B_n \rangle)^2}{\sum_{n=1}^q c_n^2 \|B_n\|_2^2} \\ &\leq \inf_{Q \in \mathbb{N}} \sup_{q > Q} \frac{\max_{n \leq q} \langle B_n \rangle^2}{\min_{n \leq q} \|B_n\|_2^2} \frac{(\sum_{n=1}^q c_n)^2}{\sum_{n=1}^q c_n^2} \\ &\leq \frac{1}{\|C_\epsilon\|_2^2} \inf_{Q \in \mathbb{N}} \sup_{q > Q} \frac{\max_{n \leq q} \langle B_n \rangle^2}{\min_{n \leq q} \|B_n\|_2^2}.\end{aligned}$$

Let k_q be the index such that the maximum in the numerator is attained, i.e.

$$k_q := \arg \max_{n \leq q} \langle B_n \rangle.$$

Since $\|B_n\|_2 \gtrsim \|B_{k_q}\|_2$ uniformly in n and q by assumption, we have

$$\frac{\max_{n \leq q} \langle B_n \rangle}{\min_{n \leq q} \|B_n\|_2} \lesssim \frac{\langle B_{k_q} \rangle}{\|B_{k_q}\|_2} \leq \max_{n \leq q} \frac{\langle B_n \rangle}{\|B_n\|_2},$$

which gives the first upper bound

$$\frac{\langle \mathcal{C} \rangle}{\|\mathcal{C}\|_2} \lesssim \frac{1}{\|C_\epsilon\|_2} \inf_{Q \in \mathbb{N}} \sup_{q > Q} \max_{n \leq q} \frac{\langle B_n \rangle}{\|B_n\|_2} \leq \frac{1}{\|C_\epsilon\|_2} \sup_n \frac{\langle B_n \rangle}{\|B_n\|_2}.$$

By the Cauchy-Schwarz inequality we have $\langle B_n \rangle \leq \|B_n\|_0^{1/2} \|B_n\|_2$ and the second upper bound follows from the first upper bound.

Similar to the above computations, using $\|B_n\|_2 \sim \|B_m\|_2$ we obtain a lower bound of the form

$$\frac{\langle \mathcal{C} \rangle}{\|\mathcal{C}\|_2} \geq \frac{1}{\|C_\epsilon\|_2} \sup_{Q \in \mathbb{N}} \inf_{q > Q} \frac{\min_{n \leq q} \langle B_n \rangle}{\max_{n \leq q} \|B_n\|_2} \gtrsim \frac{1}{\|C_\epsilon\|_2} \sup_{Q \in \mathbb{N}} \inf_{q > Q} \min_{n \leq q} \frac{\langle B_n \rangle}{\|B_n\|_2}.$$

Since $\langle B_n \rangle \geq \|B_n\|_2$, we obtain the first half of the first lower bound

$$\frac{\langle \mathcal{C} \rangle}{\|\mathcal{C}\|_2} \gtrsim \frac{1}{\|C_\epsilon\|_2}.$$

For the second half, using the non-negativity of the summand, we have

$$\begin{aligned} \frac{(\sum_{n=1}^q c_n \langle B_n \rangle)^2}{\sum_{n=1}^q c_n^2 \|B_n\|_2^2} &\geq \frac{\sum_{n=1}^q c_n^2 \langle B_n \rangle^2}{\sum_{n=1}^q c_n^2 \|B_n\|_2^2} \\ &= \frac{c_m^2 \langle B_m \rangle^2 + \sum_{n \neq m}^q c_n^2 \langle B_n \rangle^2}{\sum_{n=1}^q c_n^2 \|B_n\|_2^2}, \end{aligned}$$

where $m \leq q$ is arbitrarily chosen. Since $\|B_n\|_2 \sim \|B_m\|_2$ uniformly for all n , we have

$$\frac{(\sum_{n=1}^q c_n \langle B_n \rangle)^2}{\sum_{n=1}^q c_n^2 \|B_n\|_2^2} \gtrsim \frac{c_m^2 \langle B_m \rangle^2 + \sum_{n \neq m}^q c_n^2 \langle B_n \rangle^2}{\|C_\epsilon\|_2^2 \|B_m\|_2^2} \geq \frac{1}{\|C_\epsilon\|_2^2} c_m^2 \frac{\langle B_m \rangle^2}{\|B_m\|_2^2}.$$

Taking the limit in q yields the first lower bound. Finally, using Lemma A4, we get

$$\frac{\langle B_m \rangle}{\|B_m\|_2} \gtrsim \|B_m\|_0^{1/2} \alpha(B_m)^{1/2}$$

which completes the proof. $\square$

REFERENCES

- AUE, A., NORINHO, D. D. & HÖRMANN, S. (2015). On the Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association* **110**, 378–392.
- BEARE, B. K., SEO, J. & SEO, W.-K. (2017). Cointegrated Linear Processes in Hilbert Space. *Journal of Time Series Analysis* **38**, 1010–1027.
- BEARE, B. K. & SEO, W.-K. (2020). Representation of I(1) and I(2) autoregressive Hilbertian processes. *Econom. Theory* **36**, 773–802.
- BENKO, M., HÄRDLE, W. & KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37**.
- CHANG, J., CHEN, C., QIAO, X. & YAO, Q. (2024a). An autocovariance-based learning framework for high-dimensional functional time series. *Journal of Econometrics* **239**, 105385.
- CHANG, J., FANG, Q., QIAO, X. & YAO, Q. (2024b). On the modelling and prediction of high-dimensional functional time series.
- CHANG, Y., KIM, C. S. & PARK, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics* **192**, 152–167.
- CHIOU, J.-M., CHEN, Y.-T. & YANG, Y.-F. (2014). Multivariate Functional Principal Component Analysis: A Normalization Approach. *Statistica Sinica* **24**, 1571–1596.
- DI, C.-Z., CRAINICEANU, C. M., CAFFO, B. S. & PUNJABI, N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics* **3**, 458–488.
- FAN, J., LI, K. & LIAO, Y. (2021). Recent Developments on Factor Models and its Applications in Econometric Learning. *Annu. Rev. Financ. Econ.* **13**, 401–430.
- GAO, Y., SHANG, H. L. & YANG, Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis* **170**, 232–243.
- GUO, S., QIAO, X., WANG, Q. & WANG, Z. (2024). Factor modelling for high-dimensional functional time series.
- HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **68**, 109–126.
- HALL, P., MÜLLER, H.-G. & WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**.

- HALLIN, M., NISOL, G. & TAVAKOLI, S. (2023). Factor models for high-dimensional functional time series I: Representation results. *Journal Time Series Analysis* **44**, 578–600.
- HAPP, C. & GREVEN, S. (2018). Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association* **113**, 649–659.
- HE, Y. & ZHANG, B. (2023). Detecting Spurious Factor Models. *Available at SSRN 4615130*.
- HE, Y. & ZHANG, B. (2024). Testing for Spurious Factor Analysis on High Dimensional Nonstationary Time Series.
- HÖRMANN, S. & KOKOSZKA, P. (2010). Weakly dependent functional data.
- HÖRMANN, S. & KOKOSZKA, P. (2012). Functional time series. In *Handbook of Statistics*, vol. 30. Elsevier, pp. 157–186.
- HORVÁTH, L., KOKOSZKA, P. & RICE, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics* **179**, 66–82.
- HSING, T. & EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, vol. 997. John Wiley & Sons.
- HYNDMAN, R. J. & SHANG, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society* **38**, 199–211.
- JIMÉNEZ-VARÓN, C. F., SUN, Y. & SHANG, H. L. (2024). Forecasting High-Dimensional Functional Time Series: Application to Sub-National Age-Specific Mortality. *Journal of Computational and Graphical Statistics*, 1–15.
- KATO, T. (2013). *Perturbation Theory for Linear Operators*, vol. 132. Springer Science & Business Media.
- KOLTCHINSKII, V. & LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance.
- KONER, S. & STAICU, A.-M. (2023). Second-Generation Functional Data. *Annu. Rev. Stat. Appl.* **10**, 547–572.
- LAM, C. & YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 694–726.
- LENG, C., LI, D., SHANG, H. & XIA, Y. (2024). Covariance Function Estimation for High-Dimensional Functional Time Series with Dual Factor Structures.
- LI, D., ROBINSON, P. M. & SHANG, H. L. (2023). Nonstationary fractionally integrated functional time series. *Bernoulli* **29**, 1505–1526.
- NAKAHARA, S. & ICHIKAWA, M. (2013). Mortality in the 2011 Tsunami in Japan. *Journal of Epidemiology* **23**, 70–73.
- ONATSKI, A. & WANG, C. (2021). Spurious Factor Analysis. *Econometrica* **89**, 591–614.
- RAMSAY, J. & SILVERMAN, B. W. (1997). *Functional data analysis (springer series in statistics)*.
- RUDELSON, M. & VERSHYNIN, R. (2010). Non-asymptotic theory of random matrices: Extreme singular values. *arXiv:1003.2990 [math]*.
- SHANG, H. L. (2014). A survey of functional principal component analysis. *ASTA Adv Stat Anal* **98**, 121–142.
- TANG, C., SHANG, H. L. & YANG, Y. (2022). Clustering and Forecasting Multiple Functional Time Series.
- TAVAKOLI, S., NISOL, G. & HALLIN, M. (2023). Factor models for high-dimensional functional time series II: Estimation and forecasting. *Journal Time Series Analysis* **44**, 601–621.
- VAN DELFT, A. & DETTE, H. (2021). A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing.
- VAN DELFT, A. & EICHLER, M. (2018). Locally stationary functional time series.
- WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, vol. 48. Cambridge university press.
- WANG, J.-L., CHIOU, J.-M. & MUELLER, H.-G. (2015). Review of Functional Data Analysis.
- WANG, J.-L., CHIOU, J.-M. & MÜLLER, H.-G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application* **3**, 257–295.
- YIN, P., ESSER, E. & XIN, J. (2014). Ratio and difference of ℓ_1 and ℓ_2 norms and sparse representation with coherent dictionaries. *Communications in Information and Systems* **14**, 87–109.
- ZAPATA, J., OH, S.-Y. & PETERSEN, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika* **109**, 665–681.
- ZHANG, B., GAO, J., PAN, G. & YANG, Y. (2025). Identifying the structure of high-dimensional time series via eigen-analysis. *Journal of the American Statistical Association* **120**, 2822–2833.