Supplementary Material for "Covariance Test for Discretely Observed Functional Data: When and How It Works?"

In Supplementary Material, we introduce some notations and auxiliary lemmas in Section S.1, which serve as the building blocks for the proof of main results in Section S.2. The proofs of lemmas are delegated to Section S.3. A discussion on two types of sampling scheme is provided in Section S.4. For practical implementation, we discuss the relations between the proposed test and the existing tests in Section S.5, while some additional simulations are presented in Section S.6. For brevity, we abuse the notation by removing the subscripts $\mathbf{Z}$ and $\mathbf{Z}'$ in the sequel when no confusion arises.

S.1 Notations and Auxiliary Lemmas

Recall that the Fourier coefficients $\xi_j = \langle X, e_j \rangle$ and $\eta_j = \langle Y, e_j \rangle$, where e_j's are the eigenfunctions of pooled covariance operator $\mathcal{C}_P$, and their second moments $\theta_{jk} := \mathbb{E}(\xi_j \xi_k)$ and $\zeta_{jk} := \mathbb{E}(\eta_j \eta_k)$, respectively. Recall the definitions (8)–(10) in the main paper that are collected below

$$\hat{\xi}_{ij} = \frac{1}{N} \sum_{p=1}^{N} X_{ip} \hat{e}_j(s_{ip}), \qquad \hat{\eta}_{ij} = \frac{1}{M} \sum_{q=1}^{M} Y_{iq} \hat{e}_j(t_{iq}).$$

$$\hat{\theta}_{jk,i} = \frac{1}{N(N-1)} \sum_{p \neq p'}^{N} X_{ip} X_{ip'} \hat{e}_j(s_{ip}) \hat{e}_k(s_{ip'}),$$

$$\hat{\zeta}_{jk,i} = \frac{1}{M(M-1)} \sum_{q \neq q'}^{M} Y_{iq} Y_{iq'} \hat{e}_j(t_{iq}) \hat{e}_k(t_{iq'}),$$

$$\hat{\theta}_{jk} = \bar{\hat{\theta}}_{jk} - \bar{\hat{\xi}}_j \bar{\hat{\xi}}_k, \qquad \hat{\zeta}_{jk} = \bar{\hat{\zeta}}_{jk} - \bar{\hat{\eta}}_j \bar{\hat{\eta}}_k,$$

where $\bar{\hat{\theta}}_{jk} = n^{-1} \sum_{i=1}^{n} \hat{\theta}_{jk,i}$ and $\bar{\hat{\xi}}_j = n^{-1} \sum_{i=1}^{n} \hat{\xi}_{ij}$ are the empirical averages, and analogously for $\bar{\hat{\zeta}}_{jk}$ and $\bar{\hat{\eta}}_j$. Furthermore, denote the Fourier coefficients of fully observed curves projected on the empirical eigenfunctions $\hat{e}_j$'s, and their corresponding empirical covariances by

$$\check{\xi}_{ij} = \langle X_i, \hat{e}_j \rangle, \qquad \bar{\check{\xi}}_j = \frac{1}{n}\sum_{i=1}^{n} \check{\xi}_{ij}, \qquad \check{\theta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \check{\xi}_{ij}\check{\xi}_{ik} - \bar{\check{\xi}}_j \bar{\check{\xi}}_k,$$

$$\check{\eta}_{ij} = \langle Y_i, \hat{e}_j \rangle, \qquad \bar{\check{\eta}}_j = \frac{1}{m}\sum_{i=1}^{m} \check{\eta}_{ij}, \qquad \check{\zeta}_{jk} = \frac{1}{m}\sum_{i=1}^{m} \check{\eta}_{ij}\check{\eta}_{ik} - \bar{\check{\eta}}_j \bar{\check{\eta}}_k.$$

The following lemma quantifies the errors caused by discretization.

Lemma S.1. Under Assumptions 1 and 5, we have for any $1 \leq j, k \leq K$, as $\min(n, m) \to \infty$,

$$\mathbb{E}\left[\left(\hat{\theta}_{jk} - \check{\theta}_{jk}\right)^2\right] = O\left(\frac{1}{nN} + \frac{1}{n^2}\right) \quad \text{and} \quad \mathbb{E}\left[\left(\hat{\zeta}_{jk} - \check{\zeta}_{jk}\right)^2\right] = O\left(\frac{1}{mM} + \frac{1}{m^2}\right).$$

Let $\mathcal{I}_K = \{(j,k) : 1 \leq j \leq k \leq K\}$ be the index set with cardinality $|\mathcal{I}_K|$. To derive the asymptotic null distribution of the statistic (24), we need the convergence rates of estimated variances $\hat{\rho}_{jk}$ uniformly for $(j,k) \in \mathcal{I}_K$ with diverging K, which is provided below.

Lemma S.2. Under the null hypothesis and Assumptions 1–5, it holds that, for $1 \leq j, k \leq K$,

$$\mathbb{E}\left(|\hat{\rho}_{jk} - \rho_{jk}|^2\right) = O\left(\frac{1}{nN} + \frac{1}{N^2} + \frac{\lambda_j^2 \lambda_k^2}{n} + \lambda_j \lambda_k \omega_{n,N,K,h}\right).$$

as $n \to \infty$, where $\omega_{n,N,j,h}$ denotes the convergence rates of $\mathbb{E}\left(\|\hat{e}_j - e_j\|^2\right)$ in Proposition 1.

To make the proofs comprehensive and easy for readers to follow, we introduce some notations. Define that

$$\dot{\theta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \xi_{ij}\xi_{ik} - \bar{\xi}_j \bar{\xi}_k, \qquad \dot{\zeta}_{jk} = \frac{1}{m}\sum_{i=1}^{m} \eta_{ij}\eta_{ik} - \bar{\eta}_j \bar{\eta}_k,$$

where $\bar{\xi}_j = n^{-1}\sum_{i=1}^{n} \xi_{ij}$ and $\bar{\eta}_j = m^{-1}\sum_{i=1}^{m} \eta_{ij}$ are the empirical averages of ξ_{ij} and η_{ij}, respectively. We can rewrite

$$\begin{aligned}
\dot{\theta}_{jk} &= \frac{1}{n}\sum_{i=1}^{n} \langle X_i, e_j \rangle \langle X_i, e_k \rangle - \langle \bar{X}, e_j \rangle \langle \bar{X}, e_k \rangle = \langle \widetilde{\mathcal{C}}_X e_j, e_k \rangle, \\
\dot{\zeta}_{jk} &= \frac{1}{n}\sum_{i=1}^{n} \langle Y_i, e_j \rangle \langle Y_i, e_k \rangle - \langle \bar{Y}, e_j \rangle \langle \bar{Y}, e_k \rangle = \langle \widetilde{\mathcal{C}}_Y e_j, e_k \rangle,
\end{aligned} \tag{S.1}$$

where $\widetilde{\mathcal{C}}_X, \widetilde{\mathcal{C}}_Y$ are the sample covariances defined in Section 2.1. Furthermore, for each (j,k) in the index set $\mathcal{I}_K$, define that

$$\begin{aligned} \widehat{S}_{jk} &= \sqrt{nm/(n+m)}(\hat{\theta}_{jk} - \hat{\zeta}_{jk}), \\ \widetilde{S}_{jk} &= \sqrt{nm/(n+m)}\langle(\widetilde{\mathcal{C}}_X - \widetilde{\mathcal{C}}_Y)\tilde{e}_j, \tilde{e}_k\rangle = \langle\mathcal{L}_{nm}\tilde{e}_j, \tilde{e}_k\rangle, \\ S_{jk} &= \sqrt{nm/(n+m)}\langle(\widetilde{\mathcal{C}}_X - \widetilde{\mathcal{C}}_Y)e_j, e_k\rangle = \langle\mathcal{L}_{nm}e_j, e_k\rangle, \end{aligned} \tag{S.2}$$

where

$$\mathcal{L}_{nm} = \sqrt{nm/(n+m)}(\widetilde{\mathcal{C}}_X - \widetilde{\mathcal{C}}_Y). \tag{S.3}$$

In the sequel, we use $A(\mathcal{I}_K)$ to denote the vector by stacking $\{A_{jk} : (j,k) \in \mathcal{I}_K\}$. Then based on (S.2), we can define vectors $\widehat{S}(\mathcal{I}_K)$, $\widetilde{S}(\mathcal{I}_K)$ and $S(\mathcal{I}_K)$. With these notations and lemmas, in the next section we first prove Theorem 1 and Theorem 2 for non-standardized statistics on fully and discretely observed functional data, respectively. Then we proceed the proof of Theorem 3 for the normalized statistic on discretely observed functional data.

S.2 Proofs of Main Results

Proof of Theorem 1. Recall that $\widetilde{\Delta}_X = \widetilde{\mathcal{C}}_X - \mathcal{C}_X$, where $\widetilde{\mathcal{C}}_X = n^{-1}\sum_{i=1}^n (X_i \otimes X_i) - \bar{X} \otimes \bar{X}$. Assumption 1 implies that $\mathbb{E}(\|X_1 \otimes X_1\|_{HS}^2) = \mathbb{E}(\|X_1\|^4) < \infty$. Then by the independence of X_i's and $\mathbb{E}(X_1) = 0$, it holds that $\mathbb{E}(\|\bar{X} \otimes \bar{X}\|_{HS}^2) = O(n^{-2})$. Furthermore, by direct expansions we have

$$\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^n (X_i \otimes X_i - \mathcal{C}_X)\right\|_{HS}^2\right) = O\left(\frac{1}{n}\right). \tag{S.4}$$

This implies that $\mathbb{E}(\|\widetilde{\Delta}_X\|_{HS}^2) = O(n^{-1})$ and hence $\|\widetilde{\Delta}_X\|_{HS} = O_p(n^{-1/2})$. Similar result holds for $\|\widetilde{\Delta}_Y\|_{HS}$. Under Assumption 3, it holds that the eigen-gap $\varrho_{X,l} = \min_{k\neq l}|\lambda_k - \lambda_l| \asymp l^{-\alpha-1}$. Together with above, if the condition $K = o(n^{1/(2\alpha+2)})$ holds, it implies that $\varrho_{X,l}^{-1}\|\widetilde{\Delta}_X\|_{HS} = o_p(1)$ uniformly for all $1 \le l \le K$. This shows that $\mathbb{P}(\mathcal{E}_K) \to 1$ as $\min\{n,m\} \to \infty$.

For the second assertion, the proof is divided into three parts:

Part (1): $\widetilde{T}_K = \|\widetilde{S}(\mathcal{I}_K)\|^2 \xrightarrow{d} \|S(\mathcal{I}_K)\|^2$ **uniformly for** K**.**

Note that the statistic $\widetilde{T}_K = \|\widetilde{S}(\mathcal{I}_K)\|^2 = \sum_{1\le j\le k\le K} |\widetilde{S}_{jk}|^2$. We first show that $\|\widetilde{S}(\mathcal{I}_K)\|^2 \xrightarrow{d} \|S(\mathcal{I}_K)\|^2$ uniformly for K. Recall the definition of $\mathcal{L}_{nm}$ in (S.3), we have

$$
\begin{aligned}
\mathbb{E}\left(|\widetilde{S}_{jk} - S_{jk}|^2\right) &= \mathbb{E}\left(|\langle \mathcal{L}_{nm}\tilde{e}_j, \tilde{e}_k\rangle - \langle \mathcal{L}_{nm}e_j, e_k\rangle|^2\right) \\
&= \mathbb{E}\left(|\langle \mathcal{L}_{nm}(\tilde{e}_j - e_j), \tilde{e}_k\rangle + \langle \mathcal{L}_{nm}e_j, \tilde{e}_k - e_k\rangle|^2\right) \\
&\le 2\mathbb{E}\left(\langle \mathcal{L}_{nm}(\tilde{e}_j - e_j), \tilde{e}_k\rangle^2\right) + 2\mathbb{E}\left(\langle \mathcal{L}_{nm}e_j, \tilde{e}_k - e_k\rangle^2\right) \\
&\le 2\mathbb{E}^{1/2}\left(\|\mathcal{L}_{nm}\|_{HS}^4\right)\mathbb{E}^{1/2}\left(\|\tilde{e}_j - e_j\|^4\right) + 2\mathbb{E}^{1/2}\left(\|\mathcal{L}_{nm}\|_{HS}^4\right)\mathbb{E}^{1/2}\left(\|\tilde{e}_k - e_k\|^4\right),
\end{aligned}
$$

where the last inequality holds by the Cauchy-Schwarz inequality and noting that $\|\tilde{e}_j\| = \|e_j\| = 1$. Since $\mathcal{L}_{nm} = \sqrt{nm/(n+m)}(\widetilde{\mathcal{C}}_X - \widetilde{\mathcal{C}}_Y)$, we can show that under H_0 $(\mathcal{C}_X = \mathcal{C}_Y)$,

$$
\mathbb{E}\left(\|\mathcal{L}_{nm}\|_{HS}^4\right) \le \frac{4n^2m^2}{(n+m)^2}\left[\mathbb{E}\left(\|\widetilde{\mathcal{C}}_X - \mathcal{C}_X\|_{HS}^4\right) + \mathbb{E}\left(\|\widetilde{\mathcal{C}}_Y - \mathcal{C}_Y\|_{HS}^4\right)\right] = O(1),
$$

where we bound the fourth moment using similar arguments to (S.4). For quantifying the error of estimated eigenfunctions $\mathbb{E}(\|\tilde{e}_j - e_j\|^4)$ uniformly for $1 \le j \le K$, we apply the stochastic expansions (2.8) in Hall and Hosseini-Nasab (2006). By Assumptions 1-3 and on the event $\mathcal{E}_K$, we follow the same arguments as in (5.21) and (5.22) in Hall and Horowitz (2007) to attain that $\mathbb{E}(\|\tilde{e}_j - e_j\|^4) = O(j^4/n^2)$. As a consequence, under the condition $K = o(n^{1/(2\alpha+2)})$, we obtain that

$$
\mathbb{E}\left(\|\widetilde{S}(\mathcal{I}_K) - S(\mathcal{I}_K)\|^2\right) = \sum_{1\le j\le k\le K}\mathbb{E}\left(|\widetilde{S}_{jk} - S_{jk}|^2\right) = O\left(\frac{K^4}{n}\right) = o(1).
$$

Hence, on the event $\mathcal{E}_K$, we have $\mathbb{E}|\|\widetilde{S}(\mathcal{I}_K)\|^2 - \|S(\mathcal{I}_K)\|^2| = o(1)$ yielding $|\|\widetilde{S}(\mathcal{I}_K)\|^2 - \|S(\mathcal{I}_K)\|^2| = o_p(1)$ by applying Chebyshev's inequality. Since $\mathbb{P}(\mathcal{E}_K) \to 1$, we obtain that, uniformly for K,

$$
\|\widetilde{S}(\mathcal{I}_K)\|^2 \xrightarrow{d} \|S(\mathcal{I}_K)\|^2. \tag{S.5}
$$

Part (2): $\|S(\mathcal{I}_K)\|^2 \xrightarrow{d} \sum_{1\le j\le k\le K}\langle \mathcal{G}e_j, e_k\rangle^2$ **uniformly for** K**.**

Define that

$$\widetilde{\mathcal{L}}_{nm} = \sqrt{\frac{nm}{n+m}} \left[\frac{1}{n} \sum_{i=1}^{n} (X_i \otimes X_i) - \frac{1}{m} \sum_{i=1}^{m} (Y_i \otimes Y_i) \right],$$

where the average terms in $\mathcal{L}_{nm}$ are removed. Under Assumption 1, the CLT in Hilbert space (see Theorem 8.1.2 in Hsing and Eubank, 2015) implies that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i \otimes X_i - \mathcal{C}_X) \xrightarrow{d} \mathcal{L}_X,$$

where $\mathcal{L}_X$ is a Gaussian random element in $\mathcal{B}_{HS}(L^2[0,1])$ with mean zero and covariance

$$\mathscr{C}_X = \mathbb{E}\left[(X_1 \otimes X_1) \otimes_{HS} (X_1 \otimes X_1)\right] - \mathcal{C}_X \otimes_{HS} \mathcal{C}_X.$$

Analogously, it holds that

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} (Y_i \otimes Y_i - \mathcal{C}_Y) \xrightarrow{d} \mathcal{L}_Y,$$

where $\mathcal{L}_Y$ is a Gaussian random element in $\mathcal{B}_{HS}(L^2[0,1])$ with mean zero and covariance $\mathscr{C}_Y$. Then, under H_0, we obtain that

$$\widetilde{\mathcal{L}}_{nm} \xrightarrow{d} \mathcal{G} \triangleq \sqrt{1-\rho}\mathcal{L}_X - \sqrt{\rho}\mathcal{L}_Y.$$

Therefore, by noting that

$$\mathbb{E}\left(\|\mathcal{L}_{nm} - \widetilde{\mathcal{L}}_{nm}\|_{HS}^2\right) \le \frac{2nm}{n+m}\left[\mathbb{E}\left(\|\bar{X} \otimes \bar{X}\|_{HS}^2\right) + \mathbb{E}\left(\|\bar{Y} \otimes \bar{Y}\|_{HS}^2\right)\right] = O\left(\frac{1}{n}\right),$$

we have $\mathcal{L}_{nm} \xrightarrow{d} \mathcal{G}$ as $\min\{n, m\} \to \infty$. Then by the continuous mapping theorem, we have, uniformly for K,

$$\|S(\mathcal{I}_K)\|^2 = \sum_{1 \le j \le k \le K} \langle \mathcal{L}_{nm} e_j, e_k \rangle^2 \xrightarrow{d} \sum_{1 \le j \le k \le K} \langle \mathcal{G} e_j, e_k \rangle^2. \tag{S.6}$$

Part (3): $\sum_{1 \le j \le k \le K} \langle \mathcal{G} e_j, e_k \rangle^2 \xrightarrow{d} \langle \mathscr{M}\mathcal{G}, \mathcal{G} \rangle_{HS}$ as $K \to \infty$.

Lastly, we shall establish that $\sum_{1 \le j \le k \le K} \langle \mathcal{G} e_j, e_k \rangle^2$ converges to a quadratic form of the Gaussian measure $\mathcal{G}$ when K goes to infinity. To this end, we will show that $\mathcal{G}$, being a Gaussian process itself, admits a Karhunen-Loève decomposition, with respect to the corresponding eigenfunctions of its covariance. These eigenfunctions can be retrieved directly from the definition of $\mathscr{C}_X$ and $\mathscr{C}_Y$, and the Karhunen-Loève expansion of the typical processes X and Y.

According to the definition of $\mathcal{G}$, it has zero mean and its covariance is defined as

$$\mathscr{Q} = (1-\rho)\mathscr{C}_X + \rho\mathscr{C}_Y.$$

Since $\{e_j\}_{j=1}^{\infty}$ are eigenfunctions of both $\mathcal{C}_X$ and $\mathcal{C}_Y$ under H_0, i.e., $e_j = \phi_j = \psi_j$, we have

$$X = \sum_{j=1}^{\infty} a_j e_j, \qquad Y = \sum_{j=1}^{\infty} b_j e_j,$$

where $a_j = \xi_j$ and $b_j = \eta_j$. Denote $\Phi_{jk} = e_j \otimes e_k$ and $\tau_j := (1-\rho)\mathbb{E}(\xi_j^4) + \rho\mathbb{E}(\eta_j^4) - \lambda_j^2$. Under Assumptions 2, we can write $\mathscr{Q}$ as

$$\begin{aligned}
\mathscr{Q} &= (1-\rho)\mathscr{C}_X + \rho\mathscr{C}_Y \\
&= (1-\rho)\left\{\mathbb{E}\left[(X_1 \otimes X_1) \otimes_{HS} (X_1 \otimes X_1)\right] - \mathcal{C}_X \otimes_{HS} \mathcal{C}_X\right\} \\
&\quad + \rho\left\{\mathbb{E}\left[(Y_1 \otimes Y_1) \otimes_{HS} (Y_1 \otimes Y_1)\right] - \mathcal{C}_Y \otimes_{HS} \mathcal{C}_Y\right\} \\
&= (1-\rho)\left[\sum_{j_1,j_2,j_3,j_4} \mathbb{E}\left(\xi_{j_1}\xi_{j_2}\xi_{j_3}\xi_{j_4}\right)\left(\Phi_{j_1j_2} \otimes_{HS} \Phi_{j_3j_4}\right) - \sum_{j,k} \lambda_j\lambda_k \left(\Phi_{jj} \otimes_{HS} \Phi_{kk}\right)\right] \\
&\quad + \rho\left[\sum_{j_1,j_2,j_3,j_4} \mathbb{E}\left(\eta_{j_1}\eta_{j_2}\eta_{j_3}\eta_{j_4}\right)\left(\Phi_{j_1j_2} \otimes_{HS} \Phi_{j_3j_4}\right) - \sum_{j,k} \lambda_j\lambda_k \left(\Phi_{jj} \otimes_{HS} \Phi_{kk}\right)\right] \\
&= \sum_{j=1}^{\infty} \tau_j \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right) + \sum_{j \neq k} \lambda_j\lambda_k \left(\Phi_{jk} \otimes_{HS} \Phi_{jk} + \Phi_{jk} \otimes_{HS} \Phi_{kj}\right),
\end{aligned}$$

where the last equality holds by noting that $\mathbb{E}\left(\xi_{j_1}\xi_{j_2}\xi_{j_3}\xi_{j_4}\right)$ is $\lambda_j\lambda_k$ whenever pairs of indices are equal but not all indices are totally coincident, $\mathbb{E}(\xi_j^4)$ when all indices are equal, and zero otherwise.

If X and Y are Gaussian processes, the above representation reduces to that in Panaretos et al. (2010) (see p680) by noting that $\tau_j = 2\lambda_j^2$. Following the same arguments in Panaretos et al. (2010), we can obtain the eigen-decomposition of $\mathscr{Q}$. Specifically, regrouping the summation by adding the terms that are symmetric with respect to their indices, we obtain

$$
\begin{aligned}
\mathscr{Q} &= \sum_{j=1}^{\infty} \tau_j \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right) + \sum_{j<k} \lambda_j \lambda_k \left(\Phi_{jk} \otimes_{HS} \Phi_{jk} + \Phi_{jk} \otimes_{HS} \Phi_{kj} + \Phi_{kj} \otimes_{HS} \Phi_{jk} + \Phi_{kj} \otimes_{HS} \Phi_{kj}\right) \\
&= \sum_{j=1}^{\infty} \tau_j \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right) + \sum_{j<k} \lambda_j \lambda_k \left[\Phi_{jk} \otimes_{HS} \left(\Phi_{jk} + \Phi_{kj}\right) + \Phi_{kj} \otimes_{HS} \left(\Phi_{jk} + \Phi_{kj}\right)\right] \\
&= \sum_{j=1}^{\infty} \tau_j \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right) + \sum_{j<k} \lambda_j \lambda_k \left(\Phi_{jk} + \Phi_{kj}\right) \otimes_{HS} \left(\Phi_{jk} + \Phi_{kj}\right) \\
&= \sum_{j=1}^{\infty} \tau_j \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right) + \sum_{j<k} 2\lambda_j \lambda_k \left(\frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}} \otimes_{HS} \frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}}\right).
\end{aligned}
$$

Since $\langle \Phi_{jj}, \Phi_{jk} \rangle_{HS} = \langle e_j, e_j \rangle \langle e_j, e_k \rangle = 0$, and by noting that $\|\Phi_{jj}\|_{HS} = 1$, $\|(\Phi_{jk}+\Phi_{kj})/\sqrt{2}\|_{HS} = 1$ and $\langle \Phi_{jj}, (\Phi_{jk} + \Phi_{kj})/\sqrt{2} \rangle_{HS} = 0$, the sequence $\{\Phi_{jj}\}_{j=1}^{\infty} \cup \{(\Phi_{jk} + \Phi_{kj})/\sqrt{2}\}_{j<k}$ constitutes a complete orthonormal system of $\mathcal{B}_{HS}(L^2[0,1])$. Hence, we may represent $\mathcal{G}$ in a Karhunen-Loève expansion as

$$
\mathcal{G} = \sum_{j=1}^{\infty} \sqrt{\tau_j} \xi_j \Phi_{jj} + \sum_{j<k} \sqrt{2\lambda_j \lambda_k} \zeta_{jk} \frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}},
$$

where ξ_j and ζ_{jk} are i.i.d. random variables following standard normal distribution.

Recall the operator $\mathscr{M}$ on $\mathcal{B}_{HS}(L^2[0,1])$ defined in (12), as below

$$
\mathscr{M} = \sum_{j=1}^{\infty} \Phi_{jj} \otimes_{HS} \Phi_{jj} + \sum_{j \neq k} \frac{\left(\Phi_{jk} + \Phi_{kj}\right) \otimes_{HS} \left(\Phi_{jk} + \Phi_{kj}\right)}{8}.
$$

It is seen that $\mathscr{M}$ is self-adjoint (symmetric) by examining its indices. By direct and tedious calculations, we have

$$
\begin{aligned}
\mathscr{Q}^{1/2} \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right) \mathscr{Q}^{1/2} &= \tau_j \left(\Phi_{jj} \otimes_{HS} \Phi_{jj}\right), \\
\mathscr{Q}^{1/2} \left(\Phi_{kj} \otimes_{HS} \Phi_{jk}\right) \mathscr{Q}^{1/2} &= \lambda_j \lambda_k \left(\frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}} \otimes_{HS} \frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}}\right), \\
\mathscr{Q}^{1/2} \left(\Phi_{jk} \otimes_{HS} \Phi_{kj}\right) \mathscr{Q}^{1/2} &= \lambda_j \lambda_k \left(\frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}} \otimes_{HS} \frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}}\right), \\
\mathscr{Q}^{1/2} \left(\Phi_{jk} \otimes_{HS} \Phi_{jk}\right) \mathscr{Q}^{1/2} &= \lambda_j \lambda_k \left(\frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}} \otimes_{HS} \frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}}\right), \\
\mathscr{Q}^{1/2} \left(\Phi_{jk} \otimes_{HS} \Phi_{kj}\right) \mathscr{Q}^{1/2} &= \lambda_j \lambda_k \left(\frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}} \otimes_{HS} \frac{\Phi_{jk} + \Phi_{kj}}{\sqrt{2}}\right),
\end{aligned}
$$

where the last four equations hold by noting that $\langle \Phi_{jk}, (\Phi_{jk}+\Phi_{kj})/\sqrt{2}\rangle_{HS} = 1/\sqrt{2}$. Thus, we obtain that

$$\begin{aligned}
\mathscr{Q}^{1/2}\mathscr{M}\mathscr{Q}^{1/2} &= \mathscr{Q}^{1/2}\left(\sum_{j=1}^{\infty}\Phi_{jj}\otimes_{HS}\Phi_{jj} + \sum_{j\neq k}\frac{(\Phi_{jk}+\Phi_{kj})\otimes_{HS}(\Phi_{jk}+\Phi_{kj})}{8}\right)\mathscr{Q}^{1/2} \\
&= \sum_{j=1}^{\infty}\tau_j\left(\Phi_{jj}\otimes_{HS}\Phi_{jj}\right) + \sum_{j\neq k}\frac{\lambda_j\lambda_k}{2}\left(\frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\otimes_{HS}\frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\right) \\
&= \sum_{j=1}^{\infty}\tau_j\left(\Phi_{jj}\otimes_{HS}\Phi_{jj}\right) + \sum_{j<k}\lambda_j\lambda_k\left(\frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\otimes_{HS}\frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\right),
\end{aligned}$$

which implies that $\mathscr{Q}^{1/2}\mathscr{M}\mathscr{Q}^{1/2}$ shares the same eigenvectors as $\mathscr{Q}$ and $\{\tau_j\}_{j=1}^{\infty}\cup\{\lambda_j\lambda_k\}_{j<k}$ are the corresponding eigenvalues.

With the results above, next we follow the line of proof in Proposition 1.2.8 in Prato and Zabczyk (2002) (exactly Claim 2 there) to show that

$$\langle \mathscr{M}\mathcal{G}, \mathcal{G}\rangle_{HS} = \sum_{j=1}^{\infty}\tau_j\left|W_{\Phi_{jj}}(\mathcal{G})\right|^2 + \sum_{j<k}\lambda_j\lambda_k\left|W_{\left(\Phi_{jk}+\Phi_{kj}\right)/\sqrt{2}}(\mathcal{G})\right|^2, \tag{S.7}$$

where the isomorphism W is defined by $W_{\mathcal{F}}(\mathcal{G}) := \langle \mathscr{Q}^{-1/2}\mathcal{F}, \mathcal{G}\rangle_{HS}$ for any $\mathcal{F}\in\mathscr{Q}^{1/2}(\mathcal{B}_{HS}(L^2[0,1]))$, and the series being convergent in L^1. Using the eigen-decomposition of $\mathscr{Q}$ and Karhunen-Loève expansion of $\mathcal{G}$, we can see that $|W_{\Phi_{jj}}(\mathcal{G})|^2 = \xi_j^2$ and $|W_{\left(\Phi_{jk}+\Phi_{kj}\right)/\sqrt{2}}(\mathcal{G})|^2 = \zeta_{jk}^2$ are i.i.d. random variables following χ_1^2.

Due to the eigenvectors $\{\Phi_{jj}\}_{j=1}^{\infty}\cup\{(\Phi_{jk}+\Phi_{kj})/\sqrt{2}\}_{j<k}$ of $\mathscr{Q}$, for $K\in\mathbb{N}$, define

$$\mathscr{P}_K = \sum_{j=1}^{K}\Phi_{jj}\otimes_{HS}\Phi_{jj} + \sum_{1\le j<k\le K}\frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\otimes_{HS}\frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}.$$

Also note that $\{\Phi_{jj}\}_{j=1}^{\infty}\cup\{(\Phi_{jk}+\Phi_{kj})/\sqrt{2}\}_{j<k}$ are the eigenvectors of $\mathscr{Q}^{1/2}\mathscr{M}\mathscr{Q}^{1/2}$. Then,

$$\begin{aligned}
\langle \mathscr{M}\mathscr{P}_K\mathcal{G}, \mathscr{P}_K\mathcal{G}\rangle_{HS} &= \langle\left(\mathscr{Q}^{1/2}\mathscr{M}\mathscr{Q}^{1/2}\right)\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}\rangle_{HS} \\
&= \sum_{j=1}^{\infty}\langle\left(\mathscr{Q}^{1/2}\mathscr{M}\mathscr{Q}^{1/2}\right)\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \Phi_{jj}\rangle_{HS}\langle\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \Phi_{jj}\rangle_{HS} \\
&\quad + \sum_{j<k}\left\langle\left(\mathscr{Q}^{1/2}\mathscr{M}\mathscr{Q}^{1/2}\right)\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\right\rangle_{HS}\left\langle\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\right\rangle_{HS} \\
&= \sum_{j=1}^{\infty}\tau_j\langle\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \Phi_{jj}\rangle_{HS}^2 + \sum_{j<k}\lambda_j\lambda_k\left\langle\mathscr{Q}^{-1/2}\mathscr{P}_K\mathcal{G}, \frac{\Phi_{jk}+\Phi_{kj}}{\sqrt{2}}\right\rangle_{HS}^2.
\end{aligned}$$

Consequently, for $K \in \mathbb{N}$,

$$\langle \mathscr{M} \mathscr{P}_K \mathcal{G}, \mathscr{P}_K \mathcal{G} \rangle_{HS} = \sum_{j=1}^{\infty} \tau_j \left| W_{\mathscr{P}_K \Phi_{jj}} (\mathcal{G}) \right|^2 + \sum_{j<k} \lambda_j \lambda_k \left| W_{\mathscr{P}_K \left(\Phi_{jk} + \Phi_{kj} \right)/\sqrt{2}} (\mathcal{G}) \right|^2 .$$

Under Assumption 1, we have $\sum_{j=1}^{\infty} \tau_j + \sum_{j<k} \lambda_j \lambda_k \leq \|\mathscr{Q}\|_{\mathcal{HS}}^2 < \infty$. Furthermore, as $K \to \infty$, $\mathscr{P}_K \mathcal{G} \to \mathcal{G}$, $W_{\mathscr{P}_K \Phi_{jj}} \to W_{\Phi_{jj}}$ and $W_{\mathscr{P}_K \left(\Phi_{jk} + \Phi_{kj} \right)/\sqrt{2}} \to W_{\left(\Phi_{jk} + \Phi_{kj} \right)/\sqrt{2}}$ in the sense of mean squared expectation. Then using the Fatou lemma, we obtain (S.7).

To get our result, note that

$$\mathscr{P}_K \mathscr{M} \mathscr{P}_K = \sum_{j=1}^{K} \Phi_{jj} \otimes_{HS} \Phi_{jj} + \sum_{1 \leq j < k \leq K} \frac{(\Phi_{jk} + \Phi_{kj}) \otimes_{HS} (\Phi_{jk} + \Phi_{kj})}{4},$$

and

$$\begin{aligned}
\langle \mathscr{M} \mathscr{P}_K \mathcal{G}, \mathscr{P}_K \mathcal{G} \rangle_{HS} &= \langle \mathscr{P}_K \mathscr{M} \mathscr{P}_K \mathcal{G}, \mathcal{G} \rangle_{HS} \\
&= \sum_{j=1}^{K} \langle (\Phi_{jj} \otimes_{HS} \Phi_{jj}) \mathcal{G}, \mathcal{G} \rangle_{HS} + \sum_{1 \leq j < k \leq K} \left\langle \frac{(\Phi_{jk} + \Phi_{kj}) \otimes_{HS} (\Phi_{jk} + \Phi_{kj})}{4} \mathcal{G}, \mathcal{G} \right\rangle_{HS} \\
&= \sum_{j=1}^{K} \langle \mathcal{G} e_j, e_j \rangle^2 + \sum_{1 \leq j < k \leq K} \langle \mathcal{G} e_j, e_k \rangle^2 \\
&= \sum_{j=1}^{\infty} \tau_j \left| W_{\mathscr{P}_K \Phi_{jj}} (\mathcal{G}) \right|^2 + \sum_{j<k} \lambda_j \lambda_k \left| W_{\mathscr{P}_K \left(\Phi_{jk} + \Phi_{kj} \right)/\sqrt{2}} (\mathcal{G}) \right|^2 .
\end{aligned}$$

Then as $K \to \infty$, we obtain that

$$\sum_{1 \leq j \leq k \leq K} \langle \mathcal{G} e_j, e_k \rangle^2 \longrightarrow \langle \mathscr{M} \mathcal{G}, \mathcal{G} \rangle_{HS} = \sum_{j=1}^{\infty} \tau_j \left| W_{\Phi_{jj}} (\mathcal{G}) \right|^2 + \sum_{j<k} \lambda_j \lambda_k \left| W_{\left(\Phi_{jk} + \Phi_{kj} \right)/\sqrt{2}} (\mathcal{G}) \right|^2 ,$$

where the convergence holds in the sense of mean squared expectation. Hence, the proof is complete. □

Proof of Theorem 2. Recall that $\widehat{\Delta}_X = \widetilde{\mathcal{C}}_X - \mathcal{C}_X$, where $\widetilde{\mathcal{C}}_X$ is the local smoothing estimator of covariance operator. Zhang and Wang (2016) have shown in their Theorem 4.2 that

$$\|\widehat{\Delta}_X\|_{HS} = O_p \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{nNh}} + \frac{1}{\sqrt{n}Nh} + h^2 \right).$$

Combining with $\varrho_{X,l} \asymp l^{-\alpha-1}$ under Assumption 3, it implies that uniformly for all $1 \leq l \leq K$,

$$\varrho_{X,l}^{-1} \|\widehat{\Delta}_X\|_{HS} = O_p\left(\frac{K^{\alpha+1}}{\sqrt{n}} + \frac{K^{\alpha+1}}{\sqrt{nNh}} + \frac{K^{\alpha+1}}{\sqrt{nNh}} + h^2 K^{\alpha+1}\right).$$

The same arguments hold for $\widehat{\Delta}_Y$. Then taking the bandwidth in (20) and choosing the truncation level under (21), it is easy to check that $\mathbb{P}(\mathcal{F}_K) \to 1$ as $\min\{n, m\} \to \infty$.

For the second assertion, similar to the proof of Theorem 1, we only need to show that $\|\widehat{S}(\mathcal{I}_K)\|^2 \xrightarrow{d} \|S(\mathcal{I}_K)\|^2$ uniformly for K and the remaining proof follows the same line. Recall (S.1) and observe that

$$|\widehat{S}_{jk} - S_{jk}|^2 = \frac{nm}{n+m}\left[\left(\hat{\theta}_{jk} - \hat{\zeta}_{jk}\right) - \left(\dot{\theta}_{jk} - \dot{\zeta}_{jk}\right)\right]^2 \leq 4A_1 + 2A_2,$$

where

$$A_1 = \frac{nm}{n+m}\left[\left(\hat{\theta}_{jk} - \check{\theta}_{jk}\right)^2 + \left(\hat{\zeta}_{jk} - \check{\zeta}_{jk}\right)^2\right],$$
$$A_2 = \frac{nm}{n+m}\left[\left(\check{\theta}_{jk} - \check{\zeta}_{jk}\right) - \left(\dot{\theta}_{jk} - \dot{\zeta}_{jk}\right)\right]^2.$$

By Lemma S.1 and $n \asymp m$, we have $\mathbb{E}(A_1) = O(N^{-1} + n^{-1})$ uniformly for $(j,k) \in \mathcal{I}_K$. Similar to (S.1), we can rewrite

$$\check{\theta}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, \hat{e}_j\rangle\langle X_i, \hat{e}_k\rangle - \langle \bar{X}, \hat{e}_j\rangle\langle \bar{X}, \hat{e}_k\rangle = \langle \widetilde{\mathcal{C}}_X \hat{e}_j, \hat{e}_k\rangle,$$
$$\check{\zeta}_{jk} = \frac{1}{m}\sum_{i=1}^{m}\langle Y_i, \hat{e}_j\rangle\langle Y_i, \hat{e}_k\rangle - \langle \bar{Y}, \hat{e}_j\rangle\langle \bar{Y}, \hat{e}_k\rangle = \langle \widetilde{\mathcal{C}}_Y \hat{e}_j, \hat{e}_k\rangle.$$

Hence, by (S.3),

$$A_2 = \left(\langle \mathcal{L}_{nm}\hat{e}_j, \hat{e}_k\rangle - \langle \mathcal{L}_{nm} e_j, e_k\rangle\right)^2.$$

By the independence between $\mathcal{L}_{nm}$ and $\hat{e}_j$, we have, on the event $\mathcal{F}_K$,

$$\begin{aligned}
\mathbb{E}(A_2) &= \mathbb{E}\left(|\langle \mathcal{L}_{nm}(\hat{e}_j - e_j), \hat{e}_k\rangle + \langle \mathcal{L}_{nm} e_j, \hat{e}_k - e_k\rangle|^2\right) \\
&\leq 2\mathbb{E}\left(\langle \mathcal{L}_{nm}(\hat{e}_j - e_j), \hat{e}_k\rangle^2\right) + 2\mathbb{E}\left(\langle \mathcal{L}_{nm} e_j, \hat{e}_k - e_k\rangle^2\right) \\
&\leq 2\mathbb{E}\left(\|\mathcal{L}_{nm}\|_{HS}^2\right)\mathbb{E}\left(\|\hat{e}_j - e_j\|^2\right) + 2\mathbb{E}\left(\|\mathcal{L}_{nm}\|_{HS}^2\right)\mathbb{E}\left(\|\hat{e}_k - e_k\|^2\right) \\
&= O\left(\omega_{n,N,K,h}\right),
\end{aligned}$$

where we have $\mathbb{E}\left(\|\mathcal{L}_{nm}\|_{HS}^2\right) = O(1)$ and $\omega_{n,N,j,h}$ denotes the convergence rates (19) of $\mathbb{E}\left(\|\hat{e}_j - e_j\|^2\right)$ in Proposition 1. Consequently, on the event $\mathcal{F}_K$ and under the conditions (20) and (21), we obtain that

$$\mathbb{E}\left(\|\widehat{S}(\mathcal{I}_K) - S(\mathcal{I}_K)\|^2\right) = \sum_{1\leq j\leq k\leq K} \mathbb{E}\left(|\widehat{S}_{jk} - S_{jk}|^2\right) = O\left(\frac{K^2}{N} + \frac{K^2}{n} + K^2\omega_{n,N,K,h}\right) = o(1).$$

Then on the event $\mathcal{F}_K$, we have $\mathbb{E}|\|\widehat{S}(\mathcal{I}_K)\|^2 - \|S(\mathcal{I}_K)\|^2| = o(1)$ and hence $|\|\widehat{S}(\mathcal{I}_K)\|^2 - \|S(\mathcal{I}_K)\|^2| = o_p(1)$, which combining with $\mathbb{P}(\mathcal{F}_K) \to 1$ completes the proof. □

Proof of Theorem 3. We have shown that $\mathbb{P}(\mathcal{F}_K) \to 1$ as $n \to \infty$, and the proof of first assertion is finished if one can show $\mathbb{P}(\mathcal{H}_K) \to 1$. Under H_0 and by Lemma S.2, we have, uniformly for $(j,k) \in \mathcal{I}_K$,

$$\lambda_j^{-1}\lambda_k^{-1}|\hat{\rho}_{jk} - \rho_{jk}| = O_p\left(\frac{1}{\sqrt{n}} + \frac{K^{2\alpha}}{\sqrt{nN}} + \frac{K^{2\alpha}}{N} + K^{\alpha}\sqrt{\omega_{n,N,K,h}}\right),$$

as $n \to \infty$, and this is $o_p(1)$ under the conditions (20) and (21). This shows that $\mathbb{P}(\mathcal{H}_K) \to 1$.

Now we proceed the second assertion. For each $(j,k) \in \mathcal{I}_K$, define

$$\widehat{Q}_{jk} = \sqrt{nm/[(n+m)\hat{\rho}_{jk}]}(\hat{\theta}_{jk} - \hat{\zeta}_{jk}), \tag{S.8}$$

and thus $\widehat{Q}_{jk} = \widehat{S}_{jk}/\sqrt{\hat{\rho}_{jk}}$. Further, denote that $\widehat{Q}^*_{jk} = S_{jk}/\sqrt{\hat{\rho}_{jk}}$ and $Q_{jk} = S_{jk}/\sqrt{\rho_{jk}}$. We first show $\pi(\widehat{Q}(\mathcal{I}_K), \widehat{Q}^*(\mathcal{I}_K)) \longrightarrow 0$ as $n \to \infty$. Under H_0, note that all eigenfunctions $e_j = \phi_j = \psi_j$ and the principle scores hold that $a_j = \xi_j$ and $b_j = \eta_j$. Recall that $\rho_{jk} = (1-\rho)[\mathbb{E}(\xi_j^2\xi_k^2) - \mathbb{E}^2(\xi_j\xi_k)] + \rho[\mathbb{E}(\eta_j^2\eta_k^2) - \mathbb{E}^2(\eta_j\eta_k)]$. For all $j,k \in \mathbb{N}$, by Assumption 2 we have $\mathrm{c}_0^{-1}\lambda_j\lambda_k \leq \rho_{jk} \leq (\mathrm{c}_0 - 1)\lambda_j\lambda_k$, where $\mathrm{c}_0 > 2$ is a constant. Then on the event $\mathcal{H}_K$, it holds that for all $(j,k) \in \mathcal{I}_K$,

$$\frac{1}{2\mathrm{c}_0}\lambda_j\lambda_k \leq \hat{\rho}_{jk} \leq \left(\mathrm{c}_0 - 1 + \frac{1}{2\mathrm{c}_0}\right)\lambda_j\lambda_k.$$

In the proof of Theorem 2, it has been shown that, on the event $\mathcal{F}_K$ and uniformly for $(j,k) \in \mathcal{I}_K$,

$$\mathbb{E}\left(|\widehat{S}_{jk} - S_{jk}|^2\right) = O\left(\frac{1}{N} + \frac{1}{n} + \omega_{n,N,K,h}\right).$$

Thus on the event $\mathcal{F}_K \cap \mathcal{H}_K$, by Assumption 3 one can show that

$$\mathbb{E}\left(|\widehat{Q}_{jk} - \widehat{Q}^*_{jk}|^2\right) = \mathbb{E}\left(\hat{\rho}_{jk}^{-1}|\widehat{S}_{jk} - S_{jk}|^2\right) \leq \mathrm{c}\lambda_j^{-1}\lambda_k^{-1}\mathbb{E}\left(|\widehat{S}_{jk} - S_{jk}|^2\right)$$
$$= O\left(\frac{K^{2\alpha}}{N} + \frac{K^{2\alpha}}{n} + K^{2\alpha}\omega_{n,N,K,h}\right),$$

from which it follows that

$$\mathbb{E}\left(\|\widehat{Q}(\mathcal{I}_K) - \widehat{Q}^*(\mathcal{I}_K)\|^2\right) = \sum_{(j,k)\in\mathcal{I}_K} \mathbb{E}\left(|\widehat{Q}_{jk} - \widehat{Q}^*_{jk}|^2\right) = O\left(\frac{K^{2\alpha+2}}{N} + \frac{K^{2\alpha+2}}{n} + K^{2\alpha+2}\omega_{n,N,K,h}\right).$$

Therefore, taking the bandwidth in (20) and by condition (21), $\mathbb{E}(\|\widehat{Q}(\mathcal{I}_K) - \widehat{Q}^*(\mathcal{I}_K)\|^2) = o(1)$ implies $\|\widehat{Q}(\mathcal{I}_K) - \widehat{Q}^*(\mathcal{I}_K)\|^2 = o_p(1)$ by Chebyshev's inequality on the event $\mathcal{F}_K \cap \mathcal{H}_K$. Since $\mathbb{P}(\mathcal{F}_K \cap \mathcal{H}_K) \to 1$, it holds that $\pi(\widehat{Q}(\mathcal{I}_K), \widehat{Q}^*(\mathcal{I}_K)) \to 0$ as $n \to \infty$.

Next, we show $\pi(\widehat{Q}^*(\mathcal{I}_K), Q(\mathcal{I}_K)) \longrightarrow 0$ as $n \to \infty$. On the event $\mathcal{H}_K$, it is observed that, for any $0 \leq \tau \leq 1$,

$$|\hat{\rho}_{jk}^{-1/2} - \rho_{jk}^{-1/2}| = \frac{|\hat{\rho}_{jk} - \rho_{jk}|}{\hat{\rho}_{jk}^{1/2}\rho_{jk}^{1/2}|\hat{\rho}_{jk}^{1/2} + \rho_{jk}^{1/2}|} = \frac{|\hat{\rho}_{jk} - \rho_{jk}|^{\tau}|\hat{\rho}_{jk} - \rho_{jk}|^{1-\tau}}{\hat{\rho}_{jk}^{1/2}\rho_{jk}^{1/2}|\hat{\rho}_{jk}^{1/2} + \rho_{jk}^{1/2}|} \leq \mathrm{c}\frac{|\hat{\rho}_{jk} - \rho_{jk}|^{\tau}}{\rho_{jk}^{\tau+1/2}}.$$

In particular, taking $\tau = 1/2$ such that $|\hat{\rho}_{jk}^{-1/2} - \rho_{jk}^{-1/2}| \leq c\rho_{jk}^{-1}|\hat{\rho}_{jk} - \rho_{jk}|^{1/2}$, we have

$$\mathbb{E}\left(|\widehat{Q}^*_{jk} - Q_{jk}|^2\right) = \mathbb{E}\left(|S_{jk}|^2|\hat{\rho}_{jk}^{-1/2} - \rho_{jk}^{-1/2}|^2\right) \leq \mathrm{c}\rho_{jk}^{-2}\mathbb{E}\left(|S_{jk}|^2|\hat{\rho}_{jk} - \rho_{jk}|\right)$$
$$\leq \mathrm{c}\rho_{jk}^{-2}\mathbb{E}^{1/2}\left(|S_{jk}|^4\right)\mathbb{E}^{1/2}\left(|\hat{\rho}_{jk} - \rho_{jk}|^2\right).$$

Under H_0 and by (S.2), it is seen that

$$S_{jk} = \langle \mathcal{L}_{nm} e_j, e_k\rangle = \sqrt{\frac{nm}{n+m}}\left(\langle(\widetilde{\mathcal{C}}_X - \mathcal{C}_X)e_j, e_k\rangle - \langle(\widetilde{\mathcal{C}}_Y - \mathcal{C}_Y)e_j, e_k\rangle\right).$$

By Assumption 2 and following the arguments of Section 5.3 in Hall and Horowitz (2007), we obtain that uniformly for $(j,k) \in \mathcal{I}_K$,

$$\mathbb{E}\left(|S_{jk}|^4\right) \leq \frac{4n^2m^2}{(n+m)^2}\left(\mathbb{E}\left[\langle(\widetilde{\mathcal{C}}_X - \mathcal{C}_X)e_j, e_k\rangle^4\right] + \mathbb{E}\left[\langle(\widetilde{\mathcal{C}}_Y - \mathcal{C}_Y)e_j, e_k\rangle^4\right]\right) \leq \mathrm{c}\lambda_j^2\lambda_k^2. \quad \text{(S.9)}$$

Together with Lemma S.2 and (S.9), it holds that

$$\mathbb{E}\left(|\widehat{Q}^*_{jk} - Q_{jk}|^2\right) = O\left(\frac{1}{\sqrt{n}} + \frac{K^{2\alpha}}{\sqrt{nN}} + \frac{K^{2\alpha}}{N} + K^{\alpha}\sqrt{\omega_{n,N,K,h}}\right),$$

from which, by the bandwidth (20) and condition (21), it follows that

$$\mathbb{E}\left(\|\widehat{Q}^*(\mathcal{I}_K) - Q(\mathcal{I}_K)\|^2\right) = \sum_{(j,k)\in\mathcal{I}_K} \mathbb{E}\left(|\widehat{Q}^*_{jk} - Q_{jk}|^2\right) = o(1).$$

Thus, it deduces that $\pi(\widehat{Q}^*(\mathcal{I}_K), Q(\mathcal{I}_K)) \to 0$ as $n \to \infty$.

Lastly, since $\widehat{D}_K = \|\widehat{Q}(\mathcal{I}_K)\|^2$, by the continuous mapping theorem, one has

$$\widehat{D}_K \xrightarrow{d} \|Q(\mathcal{I}_K)\|^2.$$

Since $Q_{jk} = S_{jk}/\sqrt{\rho_{jk}}$ and by (S.6) in the proof of Theorem 1, it holds uniformly for K,

$$\|Q(\mathcal{I}_K)\|^2 = \sum_{1\le j\le k\le K} \frac{\langle \mathcal{L}_{nm} e_j, e_k\rangle^2}{\rho_{jk}} \xrightarrow{d} \sum_{1\le j\le k\le K} \frac{\langle \mathcal{G} e_j, e_k\rangle^2}{\rho_{jk}}.$$

From the discussion in the proof of Theorem 1, we obtain the eigen-decomposition of the covariance operator of $\mathcal{G}$. By some calculations, we can show that $\langle \mathcal{G} e_j, e_k\rangle$ with $(j,k) \in \mathcal{I}_K$ are a sequence of i.i.d. Gaussian random variables with variances ρ_{jk}. Hence, $\sum_{1\le j\le k\le K}\langle \mathcal{G} e_j, e_k\rangle^2/\rho_{jk}$ follows a Chi-squared distribution with degrees of freedom $K(K+1)/2$. Therefore, we have $\pi(\widehat{D}_K, \chi^2_{K(K+1)/2}) \to 0$ due to the equivalence of weak convergence and Prokhorov metric (see p.28 in Huber, 1981). Then the proof is complete. □

S.3 Proof of Auxiliary Lemmas

Proof of Lemma S.1. We mainly focus on the proof of $\mathbb{E}[(\hat{\theta}_{jk} - \check{\theta}_{jk})^2]$ and the arguments hold for $\mathbb{E}[(\hat{\zeta}_{jk} - \check{\zeta}_{jk})^2]$ as well. For different indexes i_1 and i_2 from 1 to n, $\hat{\theta}_{jk,i_1}$ and $\hat{\theta}_{jk,i_2}$ are independent conditionally on $\hat{e}_j$'s, since $\hat{e}_j$'s are estimated from the sample with indexes from $n+1$ to $2n$ by the sample-splitting procedure. These conditional independence also holds for $\hat{\xi}_{ij}$ and $\check{\xi}_{ij}\check{\xi}_{ik}$. Hence, direct expansion leads to the following

$$\mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^n \left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right)\right]^2\right\} = \frac{1}{n}\mathbb{E}\left[\left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right)^2\right] + \frac{n-1}{n}\mathbb{E}^2\left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right).$$

Since s_{ip} $(1 \le p \le N)$ are i.i.d. following from uniform distribution on $[0,1]$ and independent from X_i, it is easy to deduce that

$$
\begin{aligned}
\mathbb{E}\left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right) &= \mathbb{E}\left[\mathbb{E}\left(\frac{1}{N(N-1)}\sum_{p\neq p'}^{N} X_{ip}X_{ip'}\hat{e}_j(s_{ip})\hat{e}_k(s_{ip'}) - \check{\xi}_{ij}\check{\xi}_{ik}\middle| X_i\right)\right] \\
&= \mathbb{E}\left\{\mathbb{E}\left[X_{ip}\hat{e}_j(s_{ip})\right]\mathbb{E}\left[X_{ip'}\hat{e}_k(s_{ip'})\right] - \check{\xi}_{ij}\check{\xi}_{ik}\right\} \\
&= \mathbb{E}\left(\check{\xi}_{ij}\check{\xi}_{ik} - \check{\xi}_{ij}\check{\xi}_{ik}\right) = 0.
\end{aligned}
$$

Moreover, we can similarly derive that $\mathbb{E}(\hat{\theta}_{jk,i}\check{\xi}_{ij}\check{\xi}_{ik}) = \mathbb{E}(\check{\xi}_{ij}^2\check{\xi}_{ik}^2)$.

For the term $n^{-1}\mathbb{E}(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik})^2$, note that

$$
\begin{aligned}
\mathbb{E}\left(\hat{\theta}_{jk,i}^2\right) &= \frac{1}{N^2(N-1)^2}\mathbb{E}\left[\left(\sum_{p\neq p'}^{N} X_{ip}X_{ip'}\hat{e}_j(s_{ip})\hat{e}_k(s_{ip'})\right)^2\right] \\
&:= B_1 + \frac{B_2 - 4B_1}{N} + \frac{2B_1 - B_2 + B_3}{N(N-1)},
\end{aligned}
\tag{S.10}
$$

where

$$
\begin{aligned}
B_1 &= \mathbb{E}\left(\check{\xi}_{ij}^2\check{\xi}_{ik}^2\right), \\
B_2 &= \mathbb{E}\left(\langle X_i^2, \check{\xi}_{ik}^2\hat{e}_j^2 + \check{\xi}_{ij}^2\hat{e}_k^2 + 2\check{\xi}_{ij}\check{\xi}_{ik}\hat{e}_j\hat{e}_k\rangle\right) + \sigma^2\mathbb{E}\left(\check{\xi}_{ik}^2 + \check{\xi}_{ij}^2 + 2\check{\xi}_{ij}\check{\xi}_{ik}\langle\hat{e}_j, \hat{e}_k\rangle\right), \\
B_3 &= \mathbb{E}\left[\left(\langle X_i^2, \hat{e}_j\hat{e}_k\rangle + \sigma^2\langle\hat{e}_j, \hat{e}_k\rangle\right)^2\right] + \mathbb{E}\left[\left(\langle X_i^2, \hat{e}_j^2\rangle + \sigma^2\right)\left(\langle X_i^2, \hat{e}_k^2\rangle + \sigma^2\right)\right].
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\mathbb{E}\left[\left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right)^2\right] &= \mathbb{E}\left(\hat{\theta}_{jk,i}^2\right) - 2\mathbb{E}\left(\hat{\theta}_{jk,i}\check{\xi}_{ij}\check{\xi}_{ik}\right) + \mathbb{E}\left(\check{\xi}_{ij}^2\check{\xi}_{ik}^2\right) \\
&= \mathbb{E}\left(\hat{\theta}_{jk,i}^2\right) - \mathbb{E}\left(\check{\xi}_{ij}^2\check{\xi}_{ik}^2\right) \\
&= \frac{B_2 - 4B_1}{N} + \frac{2B_1 - B_2 + B_3}{N^2}.
\end{aligned}
\tag{S.11}
$$

By Cauchy-Schwarz inequality we have $|\check{\xi}_{ij}| \le \|X_i\|$, which together with Assumption 1 implies that

$$
B_1 = \mathbb{E}\left(\check{\xi}_{ij}^2\check{\xi}_{ik}^2\right) \le \mathbb{E}\left(\|X_i\|^4\right) = O(1).
$$

Similarly, by noting that $\|\hat{e}_j\| = 1$, we obtain that

$$
\begin{aligned}
\mathbb{E}\left(\check{\xi}_{ij}^2\langle X_i^2, \hat{e}_k^2\rangle\right) &= \mathbb{E}\left(\langle\mathbb{E}\left(\check{\xi}_{ij}^2 X_i^2\right), \hat{e}_k^2\rangle\right) \le \mathbb{E}\left[\langle\mathbb{E}^{1/2}\left(\check{\xi}_{ij}^4\right)\mathbb{E}^{1/2}\left(X_i^4\right), \hat{e}_k^2\rangle\right] \\
&\le \mathbb{E}^{1/2}\left(\|X_i\|^4\right)\sup\mathbb{E}^{1/2}\left(X_i^4(s)\right) = O(1),
\end{aligned}
$$

and

$$\mathbb{E}\left(\check{\xi}_{ij}\check{\xi}_{ik}\langle X_i^2, \hat{e}_j\hat{e}_k\rangle\right) = \mathbb{E}\left(\langle \mathbb{E}\left(\check{\xi}_{ij}\check{\xi}_{ik}X_i^2\right), \hat{e}_j\hat{e}_k\rangle\right) \leq \mathbb{E}\left(\langle \mathbb{E}^{1/2}\left(\check{\xi}_{ij}^2\check{\xi}_{ik}^2\right)\mathbb{E}^{1/2}\left(X_i^4\right), |\hat{e}_j\hat{e}_k|\rangle\right)$$
$$\leq \mathbb{E}^{1/2}\left(\|X_i\|^4\right)\sup\mathbb{E}^{1/2}\left(X_i^4(s)\right) = O(1),$$

deducing that $B_2 = O(1)$. Lastly, by Fubini's theorem and Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\left(\langle X_i^2, \hat{e}_j\hat{e}_k\rangle^2\right) &= \mathbb{E}\left[\int_s\int_t \mathbb{E}\left(X_i^2(s)X_i^2(t)\right)\hat{e}_j(s)\hat{e}_k(s)\hat{e}_j(t)\hat{e}_k(t)dsdt\right] \\
&\leq \mathbb{E}\left[\int_s\int_t \mathbb{E}^{1/2}\left(X_i^4(s)\right)\mathbb{E}^{1/2}\left(X_i^4(t)\right)|\hat{e}_j(s)\hat{e}_k(s)\hat{e}_j(t)\hat{e}_k(t)|dsdt\right] \\
&\leq \sup\mathbb{E}^{1/2}\left(X_i^4(s)\right)\sup\mathbb{E}^{1/2}\left(X_i^4(t)\right) = O(1),
\end{aligned}$$

and similarly $\mathbb{E}(\langle X_i^2, \hat{e}_j^2\rangle\langle X_i^2, \hat{e}_k^2\rangle) \leq \sup\mathbb{E}(X_i^4(s)) = O(1)$, implying that $B_3 = O(1)$. Together with these, it holds that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right)\right]^2 = O\left(\frac{1}{nN}\right). \tag{S.12}$$

Now considering the average terms. Since $\mathbb{E}(\hat{\xi}_{ij}) = \mathbb{E}(\check{\xi}_{ij}) = 0$, it holds that

$$\mathbb{E}\left[\left(\sum_{i=1}^n\hat{\xi}_{ij}\sum_{i=1}^n\hat{\xi}_{ik}\right)^2\right] = n\mathbb{E}\left(\hat{\xi}_{ij}^2\hat{\xi}_{ik}^2\right) + n(n-1)\mathbb{E}\left(\hat{\xi}_{ij}^2\right)\mathbb{E}\left(\hat{\xi}_{ik}^2\right) + 2n(n-1)\mathbb{E}^2\left(\hat{\xi}_{ij}\hat{\xi}_{ik}\right),$$
$$\mathbb{E}\left[\left(\sum_{i=1}^n\check{\xi}_{ij}\sum_{i=1}^n\check{\xi}_{ik}\right)^2\right] = n\mathbb{E}\left(\check{\xi}_{ij}^2\check{\xi}_{ik}^2\right) + n(n-1)\mathbb{E}\left(\check{\xi}_{ij}^2\right)\mathbb{E}\left(\check{\xi}_{ik}^2\right) + 2n(n-1)\mathbb{E}^2\left(\check{\xi}_{ij}\check{\xi}_{ik}\right).$$

Similar to B_1, we have $\mathbb{E}[(\sum_{i=1}^n\check{\xi}_{ij}\sum_{i=1}^n\check{\xi}_{ik})^2] \leq 3n^2\mathbb{E}(\|X_i\|^4)$. For the discrete parts, it is seen that

$$\mathbb{E}\left(\hat{\xi}_{ij}^2\hat{\xi}_{ik}^2\right) = \frac{(N-1)(N-2)(N-3)D_1}{N^3} + \frac{(N-1)(N-2)D_2}{N^3} + \frac{(N-1)D_3}{N^3} + \frac{D_4}{N^3},$$

where

$$D_1 = B_1,$$

$$D_2 = B_2 + 2\mathbb{E}\left(\langle X_i^2, \check{\xi}_{ij}\check{\xi}_{ik}\hat{e}_j\hat{e}_k\rangle\right) + 2\sigma^2\mathbb{E}\left(\check{\xi}_{ij}\check{\xi}_{ik}\langle\hat{e}_j, \hat{e}_k\rangle\right),$$

$$D_3 = B_3 + 2\mathbb{E}\left(\langle X_i^3, \check{\xi}_{ij}\hat{e}_j\hat{e}_k^2 + \check{\xi}_{ik}\hat{e}_k\hat{e}_j^2\rangle\right) + \mathbb{E}\left[\left(\langle X_i^2, \hat{e}_j\hat{e}_k\rangle + \sigma^2\langle\hat{e}_j, \hat{e}_k\rangle\right)^2\right],$$

$$D_4 = \mathbb{E}\left(\langle X_i^4, \hat{e}_j^2\hat{e}_k^2\rangle\right) + 6\sigma^2\mathbb{E}\left(\langle X_i^2, \hat{e}_j^2\hat{e}_k^2\rangle\right) + \mathbb{E}\left(\epsilon^4\right)\mathbb{E}\left(\langle\hat{e}_j^2, \hat{e}_k^2\rangle\right).$$

By Assumptions 1 and 5, we have

$$\mathbb{E}\left(\langle X_i^3, \check{\xi}_{ij}\hat{e}_j\hat{e}_k^2\rangle\right) \le \sup \mathbb{E}\left(X_i^4(s)\right) = O(1), \quad \mathbb{E}\left(\langle X_i^4, \hat{e}_j^2\hat{e}_k^2\rangle\right) \le \sup \mathbb{E}\left(X_i^4(s)\right) = O(1),$$

where we use the fact that $\int \hat{e}_j^2(s)\hat{e}_k^2(s)ds \le 1$. Since $B_i = O(1)$ for $i = 1, 2, 3$, it holds that $D_i = O(1)$ for $i = 1, 2, 3, 4$, implying that $\mathbb{E}(\hat{\xi}_{ij}^2\hat{\xi}_{ik}^2) = O(1)$. Similarly we can show that $\mathbb{E}(\hat{\xi}_{ij}^2) = O(1)$ and $\mathbb{E}(\hat{\xi}_{ij}\hat{\xi}_{ik}) = O(1)$. Consequently, we obtain that

$$\mathbb{E}\left[\left(\bar{\hat{\xi}}_j\bar{\hat{\xi}}_k - \bar{\check{\xi}}_j\bar{\check{\xi}}_k\right)^2\right] = O\left(\frac{1}{n^2}\right). \tag{S.13}$$

Combining (S.12) with (S.13), we complete the proof by noting that

$$\mathbb{E}\left[\left(\hat{\theta}_{jk} - \check{\theta}_{jk}\right)^2\right] \le 2\mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{jk,i} - \check{\xi}_{ij}\check{\xi}_{ik}\right)\right]^2\right\} + 2\mathbb{E}\left[\left(\bar{\hat{\xi}}_j\bar{\hat{\xi}}_k - \bar{\check{\xi}}_j\bar{\check{\xi}}_k\right)^2\right].$$

□

Proof of Lemma S.2. Recall that $\rho_{jk} = (1-\rho)[\mathbb{E}(\xi_j^2\xi_k^2) - \mathbb{E}^2(\xi_j\xi_k)] + \rho[\mathbb{E}(\eta_j^2\eta_k^2) - \mathbb{E}^2(\eta_j\eta_k)]$ and the definition of $\hat{\rho}_{jk}$ in (23), we have

$$\begin{aligned} |\hat{\rho}_{jk} - \rho_{jk}|^2 \le{} & \frac{2m}{n+m}\left|\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{jk,i}^2 - \bar{\hat{\theta}}_{jk}^2\right] - \left[\mathbb{E}\left(\xi_j^2\xi_k^2\right) - \mathbb{E}^2\left(\xi_j\xi_k\right)\right]\right|^2 \\ & + \frac{2n}{n+m}\left|\left[\frac{1}{m}\sum_{i=1}^{m}\hat{\zeta}_{jk,i}^2 - \bar{\hat{\zeta}}_{jk}^2\right] - \left[\mathbb{E}\left(\eta_j^2\eta_k^2\right) - \mathbb{E}^2\left(\eta_j\eta_k\right)\right]\right|^2. \end{aligned} \tag{S.14}$$

since $n/(n+m) \to \rho$ implies that $[n/(n+m) - \rho][\mathbb{E}(\eta_j^2\eta_k^2) - \mathbb{E}^2(\eta_j\eta_k)]$ can be arbitrarily close to 0 for sufficiently large n. Note that

$$\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{jk,i}^2 - \bar{\hat{\theta}}_{jk}^2\right] - \left[\mathbb{E}\left(\xi_j^2\xi_k^2\right) - \mathbb{E}^2\left(\xi_j\xi_k\right)\right] = E_1 + E_2 + E_3,$$

where

$$\begin{aligned} E_1 &= \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{jk,i}^2 - \bar{\hat{\theta}}_{jk}^2\right] - \left[\frac{1}{n}\sum_{i=1}^{n}\check{\xi}_{ij}^2\check{\xi}_{ik}^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\check{\xi}_{ij}\check{\xi}_{ik}\right)^2\right], \\ E_2 &= \left[\frac{1}{n}\sum_{i=1}^{n}\check{\xi}_{ij}^2\check{\xi}_{ik}^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\check{\xi}_{ij}\check{\xi}_{ik}\right)^2\right] - \left[\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}^2\xi_{ik}^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\xi_{ik}\right)^2\right], \\ E_3 &= \left[\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}^2\xi_{ik}^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\xi_{ik}\right)^2\right] - \left[\mathbb{E}\left(\xi_j^2\xi_k^2\right) - \mathbb{E}^2\left(\xi_j\xi_k\right)\right]. \end{aligned}$$

We first consider the discretization error E_1. Conditional on $\hat{e}_j$ and $\hat{e}_k$, and by the independence of X_i, we have

$$\mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)\right]^{2}\right\}=\frac{1}{n}\mathbb{E}\left[\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)^{2}\right]+\frac{n-1}{n}\mathbb{E}^{2}\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)$$
$$=\frac{1}{n}\mathbb{E}\left[\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)^{2}\right]+O\left(\frac{1}{N^{2}}\right),$$

where the last equality holds by (S.11). Similar to (S.10) but with more calculations and by Assumptions 1 and 5, we can show that

$$\mathbb{E}\left(\hat{\theta}_{jk,i}^{4}\right)=\mathbb{E}\left(\check{\xi}_{ij}^{4}\check{\xi}_{ik}^{4}\right)+O\left(\frac{1}{N}\right),\qquad \mathbb{E}\left(\hat{\theta}_{jk,i}^{2}\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)=\mathbb{E}\left(\check{\xi}_{ij}^{4}\check{\xi}_{ik}^{4}\right)+O\left(\frac{1}{N}\right).$$

Thus, it holds that

$$\mathbb{E}\left[\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)^{2}\right]=\mathbb{E}\left(\hat{\theta}_{jk,i}^{4}\right)-2\mathbb{E}\left(\hat{\theta}_{jk,i}^{2}\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)+\mathbb{E}\left(\check{\xi}_{ij}^{4}\check{\xi}_{ik}^{4}\right)=O\left(\frac{1}{N}\right).$$

Hence, we obtain that

$$\mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)\right]^{2}\right\}=O\left(\frac{1}{nN}+\frac{1}{N^{2}}\right).\tag{S.15}$$

Also note that

$$\begin{aligned}\mathbb{E}\left\{\left[\bar{\hat{\theta}}_{jk}^{2}-\left(\frac{1}{n}\sum_{i=1}^{n}\check{\xi}_{ij}\check{\xi}_{ik}\right)^{2}\right]^{2}\right\}&\leq\frac{2}{n^{4}}\mathbb{E}\left\{\left[\sum_{i=1}^{n}\left(\hat{\theta}_{jk,i}^{2}-\check{\xi}_{ij}^{2}\check{\xi}_{ik}^{2}\right)\right]^{2}\right\}\\&\quad+\frac{2}{n^{4}}\mathbb{E}\left\{\left[\sum_{i\neq i'}\left(\hat{\theta}_{jk,i}\hat{\theta}_{jk,i'}-\check{\xi}_{ij}\check{\xi}_{ik}\check{\xi}_{i'j}\check{\xi}_{i'k}\right)\right]^{2}\right\}\\&=O\left(\frac{1}{n^{3}N}+\frac{1}{n^{2}N^{2}}+\frac{1}{nN}\right),\end{aligned}\tag{S.16}$$

where we use (S.15) for the first term, and the second term is of order $(nN)^{-1}$ that we will show below. Due to $\mathbb{E}(\hat{\theta}_{jk,i})=\mathbb{E}(\check{\xi}_{ij}\check{\xi}_{ik})$, we have for $i\neq i'$,

$$\mathbb{E}\left(\hat{\theta}_{jk,i}\hat{\theta}_{jk,i'}-\check{\xi}_{ij}\check{\xi}_{ik}\check{\xi}_{i'j}\check{\xi}_{i'k}\right)=\mathbb{E}\left(\hat{\theta}_{jk,i}\right)\mathbb{E}\left(\hat{\theta}_{jk,i'}\right)-\mathbb{E}\left(\check{\xi}_{ij}\check{\xi}_{ik}\right)\mathbb{E}\left(\check{\xi}_{i'j}\check{\xi}_{i'k}\right)=0.$$

Hence, the n^4-order terms vanish in the expansion, i.e., for different i,i',i'',i''',

$$\mathbb{E}\left[\left(\hat{\theta}_{jk,i}\hat{\theta}_{jk,i'}-\check{\xi}_{ij}\check{\xi}_{ik}\check{\xi}_{i'j}\check{\xi}_{i'k}\right)\left(\hat{\theta}_{jk,i''}\hat{\theta}_{jk,i'''}-\check{\xi}_{i''j}\check{\xi}_{i''k}\check{\xi}_{i'''j}\check{\xi}_{i'''k}\right)\right]=0.$$

For the n^3-order terms, for instance, (i,i') and (i,i''), we have

$$\mathbb{E}\left[\left(\hat{\theta}_{jk,i}\hat{\theta}_{jk,i'} - \check{\xi}_{ij}\check{\xi}_{ik}\check{\xi}_{i'j}\check{\xi}_{i'k}\right)\left(\hat{\theta}_{jk,i}\hat{\theta}_{jk,i''} - \check{\xi}_{ij}\check{\xi}_{ik}\check{\xi}_{i''j}\check{\xi}_{i''k}\right)\right] = \left[\mathbb{E}\left(\hat{\theta}^2_{jk,i}\right) - \mathbb{E}\left(\check{\xi}^2_{ij}\check{\xi}^2_{ik}\right)\right]\mathbb{E}^2\left(\check{\xi}_{ij}\check{\xi}_{ik}\right)$$
$$= O\left(\frac{1}{N}\right).$$

Therefore, combining (S.15) and (S.16), we obtain that

$$\mathbb{E}\left(E_1^2\right) = O\left(\frac{1}{nN} + \frac{1}{N^2}\right). \tag{S.17}$$

Second, let $\widehat{\Phi}_{jk} := \hat{e}_j \otimes \hat{e}_k$ be an operator of $\mathcal{B}_{HS}(L^2[0,1])$ and $\|\widehat{\Phi}_{jk}\|_{HS} = 1$. Under Assumption 1, define $\mathcal{X}_i := X_i \otimes X_i - \mathcal{C}_X$ which is viewed as a random element with mean zero in $\mathcal{B}_{HS}(L^2[0,1])$. Moreover, let $\bar{\mathcal{X}} = n^{-1}\sum_{i=1}^n \mathcal{X}_i$ be the sample mean of $\mathcal{X}_i$ $(1 \leq i \leq n)$, and $\widehat{\mathscr{C}}$ be the sample covariance operator defined by

$$\widehat{\mathscr{C}} = \frac{1}{n}\sum_{i=1}^n \left[\left(\mathcal{X}_i - \bar{\mathcal{X}}\right) \otimes_{HS} \left(\mathcal{X}_i - \bar{\mathcal{X}}\right)\right] = \frac{1}{n}\sum_{i=1}^n \left(\mathcal{X}_i \otimes \mathcal{X}_i\right) - \bar{\mathcal{X}} \otimes_{HS} \bar{\mathcal{X}}.$$

With these notations and by the self-adjointness of $\widehat{\mathscr{C}}$, we can rewrite E_2 as

$$\begin{aligned}
E_2 &= \left[\frac{1}{n}\sum_{i=1}^n \check{\xi}^2_{ij}\check{\xi}^2_{ik} - \left(\frac{1}{n}\sum_{i=1}^n \check{\xi}_{ij}\check{\xi}_{ik}\right)^2\right] - \left[\frac{1}{n}\sum_{i=1}^n \xi^2_{ij}\xi^2_{ik} - \left(\frac{1}{n}\sum_{i=1}^n \xi_{ij}\xi_{ik}\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^n \left(\check{\xi}_{ij}\check{\xi}_{ik} - \frac{1}{n}\sum_{i=1}^n \check{\xi}_{ij}\check{\xi}_{ik}\right)^2 - \frac{1}{n}\sum_{i=1}^n \left(\xi_{ij}\xi_{ik} - \frac{1}{n}\sum_{i=1}^n \xi_{ij}\xi_{ik}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^n \langle \mathcal{X}_i - \bar{\mathcal{X}}, \widehat{\Phi}_{jk}\rangle^2_{HS} - \frac{1}{n}\sum_{i=1}^n \langle \mathcal{X}_i - \bar{\mathcal{X}}, \Phi_{jk}\rangle^2_{HS} \\
&= \langle \widehat{\mathscr{C}}\widehat{\Phi}_{jk}, \widehat{\Phi}_{jk}\rangle_{HS} - \langle \widehat{\mathscr{C}}\Phi_{jk}, \Phi_{jk}\rangle_{HS} \\
&= 2\langle \widehat{\mathscr{C}}\Phi_{jk}, \widehat{\Phi}_{jk} - \Phi_{jk}\rangle_{HS} + \langle \widehat{\mathscr{C}}(\widehat{\Phi}_{jk} - \Phi_{jk}), \widehat{\Phi}_{jk} - \Phi_{jk}\rangle_{HS}.
\end{aligned}$$

Under Assumptions 1 and 2, by the independence between X_i and $\hat{e}_j$ and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}\left(\left\langle \frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_i\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\right\rangle_{HS}^{2}\right) &= \mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\langle\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_i\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}\right]^2\right\} \\
&= \frac{1}{n}\mathbb{E}\left[\langle\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_i\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right] \\
&\quad + \frac{n-1}{n}\mathbb{E}^2\left[\langle\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_i\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}\right] \\
&= \frac{1}{n}\mathbb{E}\left(\langle\mathcal{X}_i, \Phi_{jk}\rangle_{HS}^2\langle\mathcal{X}_i, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right) \\
&\quad + \frac{n-1}{n}\mathbb{E}^2\left(\langle\mathcal{X}_i, \Phi_{jk}\rangle_{HS}\langle\mathcal{X}_i, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}\right) \\
&\le \frac{1}{n}\mathbb{E}\left(\|\mathcal{X}_i\|_{HS}^2\langle\mathcal{X}_i, \Phi_{jk}\rangle_{HS}^2\right)\mathbb{E}\left(\|\widehat{\Phi}_{jk}-\Phi_{jk}\|_{HS}^2\right) \\
&\quad + \frac{n-1}{n}\mathbb{E}^2\left(\|\mathcal{X}_i\|_{HS}\langle\mathcal{X}_i, \Phi_{jk}\rangle_{HS}\right)\mathbb{E}^2\left(\|\widehat{\Phi}_{jk}-\Phi_{jk}\|_{HS}\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left(\langle\left(\bar{\mathcal{X}} \otimes_{HS} \bar{\mathcal{X}}\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right) &= \frac{1}{n^3}\mathbb{E}\left[\langle\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_i\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right] \\
&\quad + \frac{n-1}{n^3}\mathbb{E}^2\left[\langle\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_i\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}\right] \\
&\quad + \frac{2(n-1)}{n^3}\mathbb{E}^2\left[\langle\left(\mathcal{X}_i \otimes_{HS} \mathcal{X}_{i'}\right)\Phi_{jk}, \widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}\right] \\
&= O\left(n^{-2}\right).
\end{aligned}$$

Since $\|\mathcal{X}_i\|_{HS} = \|X_i\|^2$ and $\langle\mathcal{X}_i, \Phi_{jk}\rangle_{HS} = \langle X_i, e_j\rangle\langle X_i, e_k\rangle = \xi_{ij}\xi_{ik}$, and

$$\begin{aligned}
\|\widehat{\Phi}_{jk}-\Phi_{jk}\|_{HS} &= \|\hat{e}_j \otimes \hat{e}_k - e_j \otimes e_k\|_{HS} \\
&\le \|(\hat{e}_j - e_j) \otimes \hat{e}_k\|_{HS} + \|e_j \otimes (\hat{e}_k - e_k)\|_{HS} \\
&= \|\hat{e}_j - e_j\| + \|\hat{e}_k - e_k\|,
\end{aligned}$$

by the independence between $\widehat{\mathscr{C}}$ and $\widehat{\Phi}_{jk}$, we obtain that

$$\begin{aligned}
\mathbb{E}\left(\langle\widehat{\mathscr{C}}\Phi_{jk},\widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right) &\leq 2\mathbb{E}\left(\left\langle\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{X}_i\otimes_{HS}\mathcal{X}_i\right)\Phi_{jk},\widehat{\Phi}_{jk}-\Phi_{jk}\right\rangle_{HS}^2\right) \\
&\quad+2\mathbb{E}\left(\langle\left(\bar{\mathcal{X}}\otimes_{HS}\bar{\mathcal{X}}\right)\Phi_{jk},\widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right) \\
&\leq\frac{1}{n}\mathbb{E}\left(\|X_i\|^4\xi_{ij}^2\xi_{ik}^2\right)\mathbb{E}\left(2\|\hat{e}_j-e_j\|^2+2\|\hat{e}_k-e_k\|^2\right) \\
&\quad+\frac{n-1}{n}\mathbb{E}^2\left(\|X_i\|^2\xi_{ij}\xi_{ik}\right)\mathbb{E}^2\left(\|\hat{e}_j-e_j\|+\|\hat{e}_k-e_k\|\right)+O\left(n^{-2}\right) \\
&\leq \mathrm{c}\lambda_j\lambda_k\mathbb{E}\left(\|\hat{e}_j-e_j\|^2\right)+O\left(n^{-2}\right)=O\left(\lambda_j\lambda_k\omega_{n,N,K,h}\right).
\end{aligned}$$

Furthermore, similar to arguments above, we can show that

$$\begin{aligned}
\mathbb{E}\left[\langle\widehat{\mathscr{C}}(\widehat{\Phi}_{jk}-\Phi_{jk}),\widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\langle\mathcal{X}_i-\bar{\mathcal{X}},\widehat{\Phi}_{jk}-\Phi_{jk}\rangle_{HS}^2\right)\right]^2 \\
&\leq \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\|\mathcal{X}_i-\bar{\mathcal{X}}\|_{HS}^2\right)^2\right]\mathbb{E}\left(\|\widehat{\Phi}_{jk}-\Phi_{jk}\|_{HS}^4\right) \\
&\leq \mathbb{E}\left(\|X_i\|^8\right)\mathbb{E}\left(\|\tilde{e}_j-e_j\|^4\right)=O\left(\omega_{n,N,K,h}^2\right),
\end{aligned}$$

Combining these together we obtain that

$$\mathbb{E}(E_2^2)=O\left(\lambda_j\lambda_k\omega_{n,N,K,h}\right). \tag{S.18}$$

Lastly, since

$$\mathbb{E}(E_3^2)\leq 2\mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}^2\xi_{ik}^2-\mathbb{E}\left(\xi_j^2\xi_k^2\right)\right]^2\right\}+2\mathbb{E}\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\xi_{ik}\right)^2-\mathbb{E}^2\left(\xi_j\xi_k\right)\right]^2\right\},$$

by straightforward calculations and Assumption 2, we obtain that

$$\begin{aligned}
\mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}^2\xi_{ik}^2-\mathbb{E}\left(\xi_j^2\xi_k^2\right)\right]^2\right\} &= \frac{1}{n}\left[\mathbb{E}\left(\xi_j^4\xi_k^4\right)-\mathbb{E}^2\left(\xi_j^2\xi_k^2\right)\right] \\
&\leq \frac{1}{n}\mathbb{E}^{1/2}\left(\xi_j^8\right)\mathbb{E}^{1/2}\left(\xi_k^8\right)=O\left(\frac{\lambda_j^2\lambda_k^2}{n}\right),
\end{aligned}$$

and

$$\mathbb{E}\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\xi_{ik}\right)^2-\mathbb{E}^2\left(\xi_j\xi_k\right)\right]^2\right\}=\frac{4\mathbb{E}\left(\xi_j^2\xi_k^2\right)\mathbb{E}^2\left(\xi_j\xi_k\right)}{n}+\frac{F_1}{n^2}+\frac{F_2}{n^3}=O\left(\frac{\lambda_j^2\lambda_k^2}{n}\right),$$

where

$$F_1 = 8\mathbb{E}^4(\xi_j\xi_k) - 18\mathbb{E}\left(\xi_j^2\xi_k^2\right)\mathbb{E}^2(\xi_j\xi_k) + 3\mathbb{E}^2\left(\xi_j^2\xi_k^2\right) + 4\mathbb{E}\left(\xi_j^3\xi_k^3\right)\mathbb{E}(\xi_j\xi_k),$$

$$F_2 = -6\mathbb{E}^4(\xi_j\xi_k) + 12\mathbb{E}\left(\xi_j^2\xi_k^2\right)\mathbb{E}^2(\xi_j\xi_k) - 3\mathbb{E}^2\left(\xi_j^2\xi_k^2\right) - 4\mathbb{E}\left(\xi_j^3\xi_k^3\right)\mathbb{E}(\xi_j\xi_k) + \mathbb{E}\left(\xi_j^4\xi_k^4\right),$$

which together imply that

$$\mathbb{E}(E_3^2) = O\left(\frac{\lambda_j^2\lambda_k^2}{n}\right). \tag{S.19}$$

Together with (S.17), (S.18) and (S.19), we obtain the bound for the first term in (S.14), and the same results hold for the second term. Note that under H_0 it holds that $\lambda_j = \kappa_j$, then we obtain the bounds

$$\mathbb{E}\left(|\hat{\rho}_{jk} - \rho_{jk}|^2\right) = O\left(\frac{1}{nN} + \frac{1}{N^2} + \lambda_j\lambda_k\omega_{n,N,K,h} + \frac{\lambda_j^2\lambda_k^2}{n}\right),$$

which complete the proof. □

S.4 Discussion on Sampling Scheme

Regarding the sampling scheme, it is noticed that while our current focus is on the uniform random design for the sake of simplicity, our assumptions are not confined to this specific distribution. Indeed, the uniform distribution can be readily relaxed to more general distributions. In addition to the random design, another sampling scheme characterized by predetermined observation points aligned on a mesh grid, i.e., the regular design, is commonplace in numerous scientific endeavors and has garnered considerable attention in the literature (Cai and Yuan, 2011; Shao et al., 2022). It is pertinent to highlight that these two commonly-used sampling schemes include the majority of scenarios and real-world data applications encountered in FDA, and our test procedure and theoretical assertions remain valid for both cases. In the following, we provide comprehensive analysis and concise discussion on both sampling schemes.

In the current work, we employ MC-type integration as an effective estimator to approximate the Fourier coefficients under the assumption of uniform random design. The design can indeed be generalized to a broader class of random designs, as long as the density function of the sampling distribution remains bounded away from zero and infinity. Assume that the observed points $s_{ip}, 1 \le i \le n, 1 \le p \le N$, which are independent of the curve $X_i(s)$ and measurement error ϵ_{ip}, are i.i.d. sampled from a distribution on $[0, 1]$ with a density f satisfying that $0 < m \le \inf_s f(s) \le \sup_s f(s) \le M < \infty$. This condition posed on density function is commonly used for discretely observed function data in FDA (see Yao et al., 2005a,b; Li and Hsing, 2010; Cai and Yuan, 2011; Zhang and Wang, 2016, for example).

Let F be the distribution of f, and then $F(s_{ip})$ follows a uniform distribution on $[0, 1]$. Usually, the density f is unknown and we need to estimate it from samples. To avoid dependence between estimates, we can pick out a sub-sample of small size and combine the observed points together to estimate f. Let $\hat{f}$ be an estimator achieved by some standard nonparametric smoothing method (Tsybakov, 2009) and $\hat{F}$ be the corresponding empirical distribution function. On the remaining sample, we construct the test statistic following the same procedure of the main paper, by replacing s_{ip} in (8) with $\hat{F}(s_{ip})$ for calculating MC-type coefficients,

$$\hat{\xi}_{ij} = \frac{1}{N} \sum_{p=1}^{N} (X_i(\hat{F}(s_{ip})) + \epsilon_{ip}) \hat{e}_j(\hat{F}(s_{ip})). \tag{S.20}$$

It is seen that

$$\begin{aligned}
\mathbb{E}[\hat{\xi}_{ij} | X_i, \hat{f}, \hat{e}_j] &= \int X_i(\hat{F}(s)) \hat{e}_j(\hat{F}(s)) \hat{f}(s) ds + \int X_i(\hat{F}(s)) \hat{e}_j(\hat{F}(s)) (f(s) - \hat{f}(s)) ds \\
&= \int X_i(\hat{F}(s)) \hat{e}_j(\hat{F}(s)) d\hat{F}(s) + \int X_i(\hat{F}(s)) \hat{e}_j(\hat{F}(s)) \frac{f(s) - \hat{f}(s)}{\hat{f}(s)} d\hat{F}(s) \\
&= \langle X_i, \hat{e}_j \rangle + \int X_i(\hat{F}(s)) \hat{e}_j(\hat{F}(s)) \frac{f(s) - \hat{f}(s)}{\hat{f}(s)} d\hat{F}(s),
\end{aligned}$$

and

$$\left| \int X_i(\hat{F}(s)) \hat{e}_j(\hat{F}(s)) \frac{f(s) - \hat{f}(s)}{\hat{f}(s)} d\hat{F}(s) \right| \le \sup_s \left| \frac{f(s) - \hat{f}(s)}{\hat{f}(s)} \right| \langle X_i, \hat{e}_j \rangle.$$

By the boundness of f, we obtain that

$$\left|\mathbb{E}[\hat{\xi}_{ij}|X_i, \hat{e}_j] - \langle X_i, \hat{e}_j\rangle\right| \leq \mathbb{E}\left\{\sup_s \left|\frac{f(s) - \hat{f}(s)}{\hat{f}(s)}\right|\right\} \langle X_i, \hat{e}_j\rangle := \delta_n \langle X_i, \hat{e}_j\rangle,$$

where δ_n that only depends on n converges to 0 at the standard nonparametric convergence rate for density function, e.g., see Theorem 1.8 in Tsybakov (2009). Consequently, we have

$$(1 - \delta_n)\langle X_i, \hat{e}_j\rangle \leq \mathbb{E}[\hat{\xi}_{ij}|X_i, \hat{e}_j] \leq (1 + \delta_n)\langle X_i, \hat{e}_j\rangle, \tag{S.21}$$

and $\delta_n \to 0$. Following the similar routes of proof in the present paper, we can show that the theoretical assertions for the proposed statistic under the new estimated Fourier coefficients (S.20) still hold.

For the regular design, each sample path is observed on a deterministic grid $0 = s_0 < s_1, \ldots, s_{N-1} < s_N = 1$ (Cai and Yuan, 2011; Shao et al., 2022). In this design, it is usually assumed that the design points $\{s_p\}_{p=0}^N$ are nonrandom and equally spaced. It is remarkable that from a theoretical perspective, this design is different from the random design. Under suitable regularity conditions (Assumption S.1 below), Shao et al. (2022) and Zhou et al. (2022) have shown that the local linear estimators for mean/covariance functions are consistent and obtain the convergence rate as the same as those on the random design. Furthermore, Zhou et al. (2022) established the convergence rate of the estimated eigenfunctions with diverging indices under the fixed regular design, achieving the same results as in Proposition 1 of the main paper.

Assumption S.1. The design points $\{s_p\}_{p=1}^N$ are nonrandom, and there exist constants $c_2 \geq c_1 > 0$, such that for any intervals $A, B \in [0, 1]$,

(a) $c_1 N|A| - 1 \leq \sum_{p=1}^N \mathbb{I}_{\{s_p \in A\}} \leq \max\{c_2 N|A|, 1\}$;

(b) $c_1 N^2|A||B| - 1 \leq \sum_{p_1,p_2=1}^N \mathbb{I}_{\{s_{p_1} \in A\}}\mathbb{I}_{\{s_{p_2} \in B\}} \leq \max\{c_2 N^2|A||B|, 1\}$, where $|A|$ denotes the length of A. Moreover, the smoothness of covariance function is assumed that

(c) The second derivatives $\partial C(s,t)/\partial s^2$ and $\partial C(s,t)/\partial t^2$ are continuous on $[0,1]$.

For constructing the test statistic, we may use the Riemann sum instead of Monte-Carlo average to estimate the Fourier coefficients per subject, e.g., a formula based on the trapezoidal rule,

$$\hat{\xi}_{ij} = \sum_{p=1}^{N} \frac{X_{ip}\hat{e}_j(s_p) + X_{ip-1}\hat{e}_j(s_{p-1})}{2}(s_p - s_{p-1}). \tag{S.22}$$

In the classical book of numerical analysis (Gautschi, 2011), it is known that under suitable regularity conditions such that the integrated function has two continuously derivative, (S.22) has an error term $O(1/N^2)$. Thus, compared with the variance $O(1/(nN))$ in Lemma S.1, the bias of the estimated Fourier coefficient $\hat{\xi}_{ij}$ can be dominated if $N \gtrsim n^{1/3}$. This condition is reasonable since under the fixed regular design, $N \to \infty$ is required for the estimates to be consistent. We emphasize that the order $n^{1/3}$ is sufficiently lower than the obtained phase transition point $n^{(\alpha+1)/(\alpha+3)}$ with $\alpha > 1$ for standardized statistic (see condition (27) in Theorem 3). Other more efficient algorithms could be applied to the calculation of Fourier coefficients. For instance, the well-known Simpson's formula could be used if the integrated function has four continuously derivative, and the estimated score has a lower bias of order $O(1/N^4)$ (Gautschi, 2011). Based on the new estimated Fourier coefficients (S.22) and following the routes of proof in the present paper with slightly modified arguments, we can show that our theoretical results still hold under the fixed regular design.

S.5 Discussion on Existing Methods

In this paper, we employ pool-smoothing covariance estimator and MC-type integration for estimating second moment, to handle discretely observed data ranging from "sparse" to "dense" paradigms. Additionally, by employing diagonal-elimination and sample-splitting strategy, we can control bias and variance well at the same time to obtain the asymptotic null distribution. Furthermore, the phenomenon of phase transition is revealed, providing theoretical justification and practical guidance for testing implementation. These methodological challenges, computational complexities and theoretical barriers are notably absent from the existing literature about covariance test, even though the normalized statistic $\widehat{D}_{K,\mathbf{Z}'}$ in (24) appears similar to existing FPC-based statistics (Panaretos et al., 2010; Fremdt et al., 2013). Due to this similarity, many reviewers have expressed interest in the cross-effects of approaches applied on these statistics. To clarify this issue, we first introduce the statistics from Panaretos et al. (2010) and Fremdt et al. (2013), followed by a discussion and numerical simulations.

Let $\widetilde{\mathcal{C}}_X(s,s') = n^{-1}\sum_{i=1}^n \{X_i(s) - \bar{X}(s)\}\{X_i(s') - \bar{X}(s')\}$ be the sample covariance function of $X(s)$, where $\bar{X}(s) = n^{-1}\sum_{i=1}^n X_i(s)$ is the sample mean. The analogues for $Y(t)$ is defined as $\widetilde{C}_Y(t,t')$. Given a fixed truncation parameter K, define $\{(\tilde{\vartheta}_j, \tilde{e}_j)\}_{j=1}^K$ as the top K pairs of principal components associated with the empirical pooled covariance operator $\widetilde{\mathcal{C}}_P = (n+m)^{-1}(n\widetilde{\mathcal{C}}_X + m\widetilde{\mathcal{C}}_Y)$. Under the Gaussian assumption, Panaretos et al. (2010) proposed an FPC-based test statistic

$$T_{PK} = \frac{nm}{2(n+m)} \sum_{j=1}^K \sum_{k=1}^K \frac{\langle (\widetilde{\mathcal{C}}_X - \widetilde{\mathcal{C}}_Y)\tilde{e}_j, \tilde{e}_k\rangle^2}{\tilde{\rho}_{jk}}, \tag{S.23}$$

where $\tilde{\rho}_{jk} = \tilde{\vartheta}_j\tilde{\vartheta}_k$ $(1 \le j,k \le K)$ are the estimated variances for standardization. Note that for fully observed functions, $\tilde{\vartheta}_j$ can be rewritten as

$$\tilde{\vartheta}_j = \frac{n}{n+m}\left(\frac{1}{n}\sum_{i=1}^n \tilde{\xi}_{ij}^2\right) + \frac{m}{n+m}\left(\frac{1}{m}\sum_{i=1}^m \tilde{\eta}_{ij}^2\right),$$

where $\tilde{\xi}_{ij} = \langle X_i - \bar{X}, \tilde{e}_j\rangle$ and $\tilde{\eta}_{ij} = \langle Y_i - \bar{Y}, \tilde{e}_j\rangle$ are the Fourier coefficients of X_i and Y_i, respectively, with respect to $\tilde{e}_j$. Under the non-Gaussian case, Fremdt et al. (2013) extended the test statistic (S.23) to a quadratic form that

$$T_{FH} = \widetilde{S}^{\top}\widetilde{\Gamma}^{-1}\widetilde{S}, \tag{S.24}$$

where $\widetilde{S}$ is the vectorization of $\{\widetilde{S}_{jk} = \sqrt{nm/(n+m)}\langle(\widetilde{\mathcal{C}}_X - \widetilde{\mathcal{C}}_Y)\tilde{e}_j, \tilde{e}_k\rangle, 1 \leq j \leq k \leq K\}$, and $\widetilde{\Gamma} = (\tilde{\rho}_{jk})$ is the estimate of the asymptotic covariance matrix. Here the element $\tilde{\rho}_{jk}$ is defined as the empirical fourth moment

$$\tilde{\rho}_{jk} = \frac{m}{n+m}\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_{ij}^2\tilde{\xi}_{ik}^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_{ij}\tilde{\xi}_{ik}\right)^2\right] + \frac{n}{n+m}\left[\frac{1}{m}\sum_{i=1}^{m}\tilde{\eta}_{ij}^2\tilde{\eta}_{ik}^2 - \left(\frac{1}{m}\sum_{i=1}^{m}\tilde{\eta}_{ij}\tilde{\eta}_{ik}\right)^2\right].$$

When encountering discretely observed functional data, the statistics in (S.23) and (S.24) cannot be directly calculated from data. In the following, we will discuss two aspects. First, as mentioned in the main paper, the pool-smoothing estimator has more advantages then pre-smoothing estimator especially under the "sparse" design. So we investigate the effects when pool-smoothing covariance estimator is applied to the competing methods, i.e., T_{PK} and T_{FH}, by conducting an additional simulation. Second, the numerical studies in Section 5 show highly inflated size of T_{PK}. One possible reason is that T_{PK} uses the product of second moment estimation as the fourth moment estimation for standardization. This square integrability of fourth moment holds under the Gaussian assumption. However, for discretely observed functional data, especially under the "sparse" design, the product of second moment estimation performs poor than the fourth moment estimation. We conduct an additional simulation by replacing the fourth moment estimation in (23) with the product of second moment estimation, and investigate the effects.

(1) Improving existing statistics by pool-smoothing method: In the simulated study, we pre-smooth the discretely observed data and then apply the pre-smoothed curves to compute the statistics T_{FH} in Fremdt et al. (2013) and T_{PK} in Panaretos et al. (2010), which are constructed based on fully observed functional data. As pointed out by the referee, we conduct an additional simulation by using the local smoother of covariance based on pool-smoothing method to replace the cross-sectional estimator based on pre-smoothed curves. Specifically, we use pool-smoothing estimators $\widehat{\mathcal{C}}_X$ and $\widehat{\mathcal{C}}_Y$, and their pooled estimator $\widehat{\mathcal{C}}_P$ to obtain the empirical eigenfunctions $\hat{e}_j$. For the estimation of Fourier coefficients, we compute the integral $\langle \widehat{X}_i, \hat{e}_j \rangle$, where $\widehat{X}_i$ is the pre-smoothed curves, and use the estimated coefficients to compute the variances in the same way. We follow the same design as the simulated study in Section 5 of the main paper.

The performance of modified tests (denoted by $\widetilde{T}_{FH}$ and $\widetilde{T}_{PK}$) with original tests T_{FH} and T_{PK} are compared in Figure S.1 below, showing that the modified tests achieve improvements to some extent by the local smoother with pooling observations. For example, the size and power under $N = 6$ for T_{FH}, and the size and power under $N = 30$ for T_{PK}. Nevertheless, compared to Figure 1 in Section 5 of the main paper, our proposed test still outperforms the modified tests under sparse case ($N = 6$).

(2) Normalization by product of second moment: As mentioned above, for fully observed functions, Panaretos et al. (2010) use the product of second moment estimation as the estimates of asymptotic variances, since the asymptotic variances under Gaussian assumption and H_0 are

$$\rho_{jk} = (1-\rho)\mathrm{Var}\left(a_j a_k\right) + \rho \mathrm{Var}\left(b_j b_k\right) = \mathbb{E}\left(a_j^2 a_k^2\right) - \mathbb{E}^2\left(a_j a_k\right) = \begin{cases} \lambda_j \lambda_k, & j \neq k; \\ 2\lambda_j^2, & j = k. \end{cases}$$

In our paper, we use Assumption 2 to replace Gaussian assumption. extending the settings where testing procedure works. To see this, under Assumption 2, the asymptotic variances are

$$\rho_{jk} = (1-\rho)\mathrm{Var}\left(a_j a_k\right) + \rho \mathrm{Var}\left(b_j b_k\right) = \begin{cases} \lambda_j \lambda_k, & j \neq k; \\ (1-\rho)\mathbb{E}(a_j^4) + \rho \mathbb{E}(b_j^4) - \lambda_j^2, & j = k. \end{cases}$$

Comparing above two terms, we can see that under Assumption 2 (not necessarily Gaussian), the fourth moments are unknown. Therefore, in our setting the asymptotic variances cannot

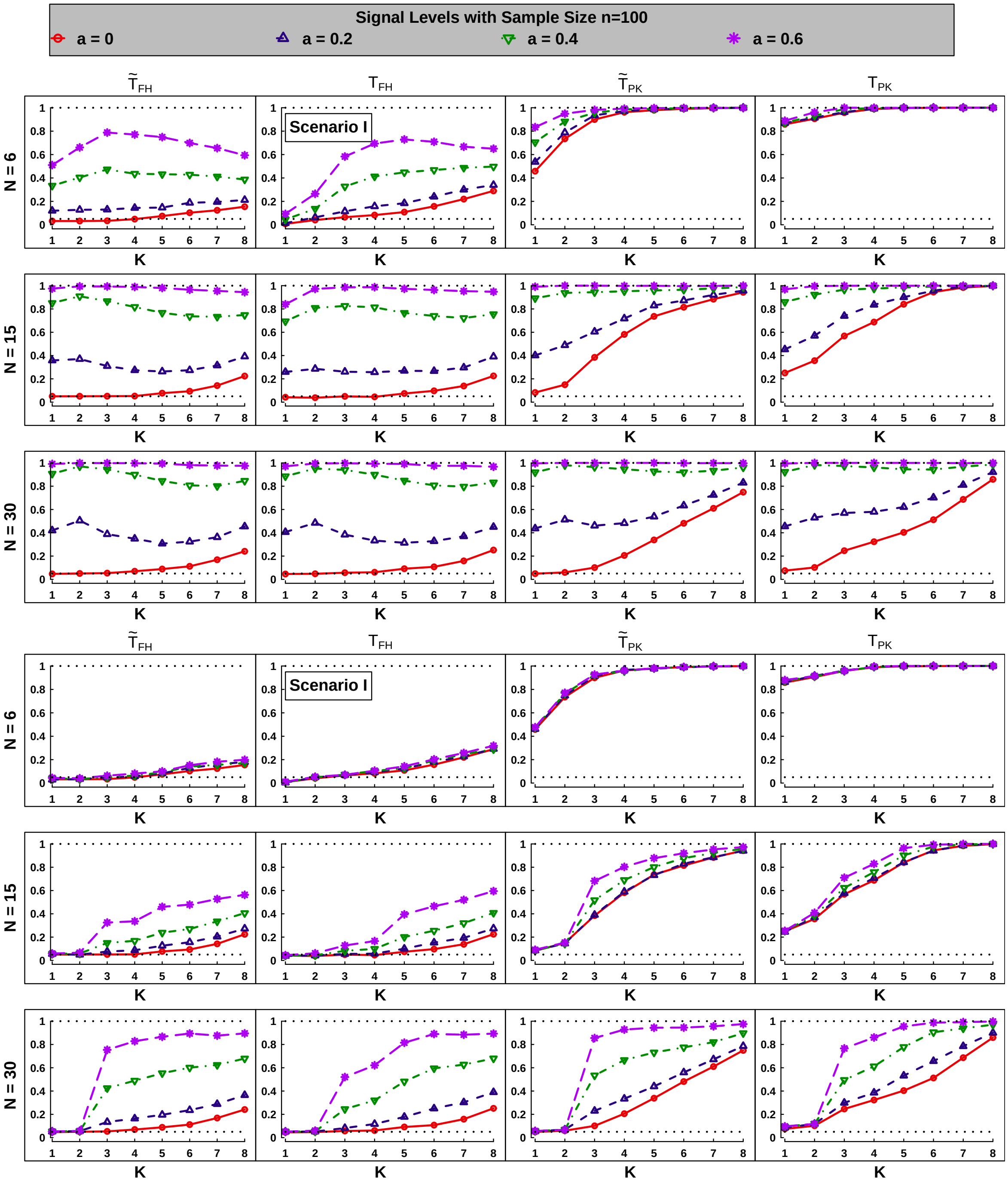

Figure S.1: Empirical sizes and powers under Scenario I and II (two sub-figures) for comparison of $\widetilde{T}_{FH}, \widetilde{T}_{PK}$ (pooling-based statistics) with T_{FH}, T_{PK} (presmoothing-based statistics). Rejection rates with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under sizes $n = m = 100$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

be estimated by second moment. This motivates us to adopt the estimation (23) for the proposed statistic.

Although all theoretical results of our paper hold without Gaussian assumption, in simulations we generate functional data following from Gaussian process which is referred as to Section 5. So we are interested in the performance of standardization under the product of second moment estimation as in Panaretos et al. (2010). Specifically, instead of (23), we use diagonal-eliminated second moment estimation defined in (10) to estimate the asymptotic variances as

$$\hat{\rho}_{jk} = \left(\frac{n}{n+m}\hat{\theta}_{jj} + \frac{m}{n+m}\hat{\zeta}_{jj}\right)\left(\frac{n}{n+m}\hat{\theta}_{kk} + \frac{m}{n+m}\hat{\zeta}_{kk}\right). \tag{S.25}$$

As a result, we modify the statistic in (24) by using (S.25) to replace $\hat{\rho}_{jk,\mathbf{Z}'}$ and obtain the modified test statistic, denoted by $\widetilde{T}_{pool}$. Following the same designs in Section 5, the comparison of performance are reported in Figure S.2 below. It is seen that the sizes of $\widetilde{T}_{pool}$ are apparently inflated, leading to unreliable higher power. Notably, the performance patterns of $\widetilde{T}_{pool}$ seems similar to that of T_{PK} which also utilize the product of second moment estimation as the estimates of variances. This reflects that this type estimator may be inappropriate for FPC-based test statistic with growing truncation levels based on discretely observed functional data, especially on sparse designs. On account of this, the bad behaviour of T_{PK} may be due to unstable estimation of variances.

S.6 Additional Simulations

In this section, we complete the simulations mentioned in the main paper and conduct some additional simulations on the performance of the proposed covariance test. Recalling the scenarios in Section 5 of the main paper, we report the plotted curves of rejection rates under the size $n = m = 200$ in Figure S.3. In addition, we investigate the performance of

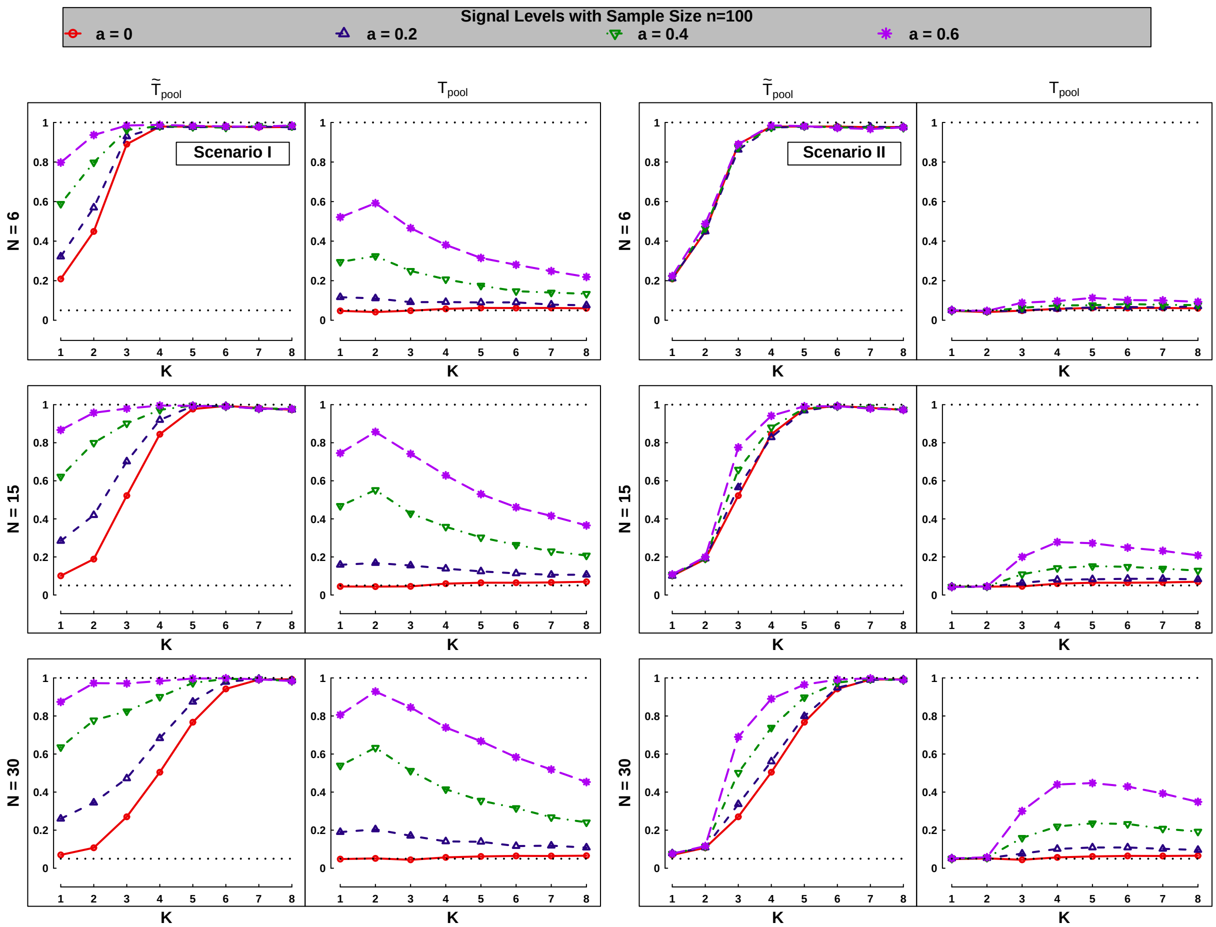

Figure S.2: Empirical sizes and powers under Scenario I and II (left and right two columns) for comparison of $\widetilde{T}_{pool}$ (second order) and T_{pool} (fourth order). Rejection rates with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under sizes $n = m = 100$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

proposed test under different settings, in comparison with T_{FH} and T_{PK}. In the following, we present the results in detail.

(1) The signals on the components with large index: Besides Scenarios I and II, in the following we consider another two scenarios, where the signals take place at $K = 5, 6$ and $K = 7, 8$, respectively:

- Scenario III : $\gamma = a(\mathbf{e}_5 + \mathbf{e}_6)$, where a increases from 0 to 0.6;
- Scenario IV : $\gamma = a(\mathbf{e}_7 + \mathbf{e}_8)$, where a increases from 0 to 0.6.

In these scenarios, it is more difficult to detect the signals than those in Scenarios I and II. The results are reported in Figures S.4 and S.5 for $n = m = 100$ and 200, respectively. It is observed that all tests are insensitive to the signal strength. In Scenario III, the power of the proposed test T_{pool} exhibits a significant increase upon detecting the first anomaly signal ($K = 5$), and then stabilizes after detecting all anomaly signals ($K = 6$). Our test consistently outperforms the competitors, with T_{FH} and T_{PK} both have size inflations. In Scenario IV, all tests nearly have no powers when signal strength varies from 0 to 0.6. This implies that the signals occurring at smaller components are more difficult to detect, and more samples and measurements are required.

(2) The performance with different eigenvalue decay rates: From Theorems 2 and 3, we have seen that the phase transition and the maximum allowable range of truncation level are affected by the eigenvalue decay rates, i.e., the smoothing parameter α. In Section 5, we have investigated the finite sample performance of the proposed test in comparison with existing methods, under the setting of $\alpha = 1.5$. In this section, we perform additional simulations to investigate the performance of proposed test under different α, by considering the following three designs:

(a) Slow: $\lambda_1 = 1$, $\lambda_2 = 0.64$ and $\lambda_j = j^{-1.5}$ with $3 \le j \le 50$;

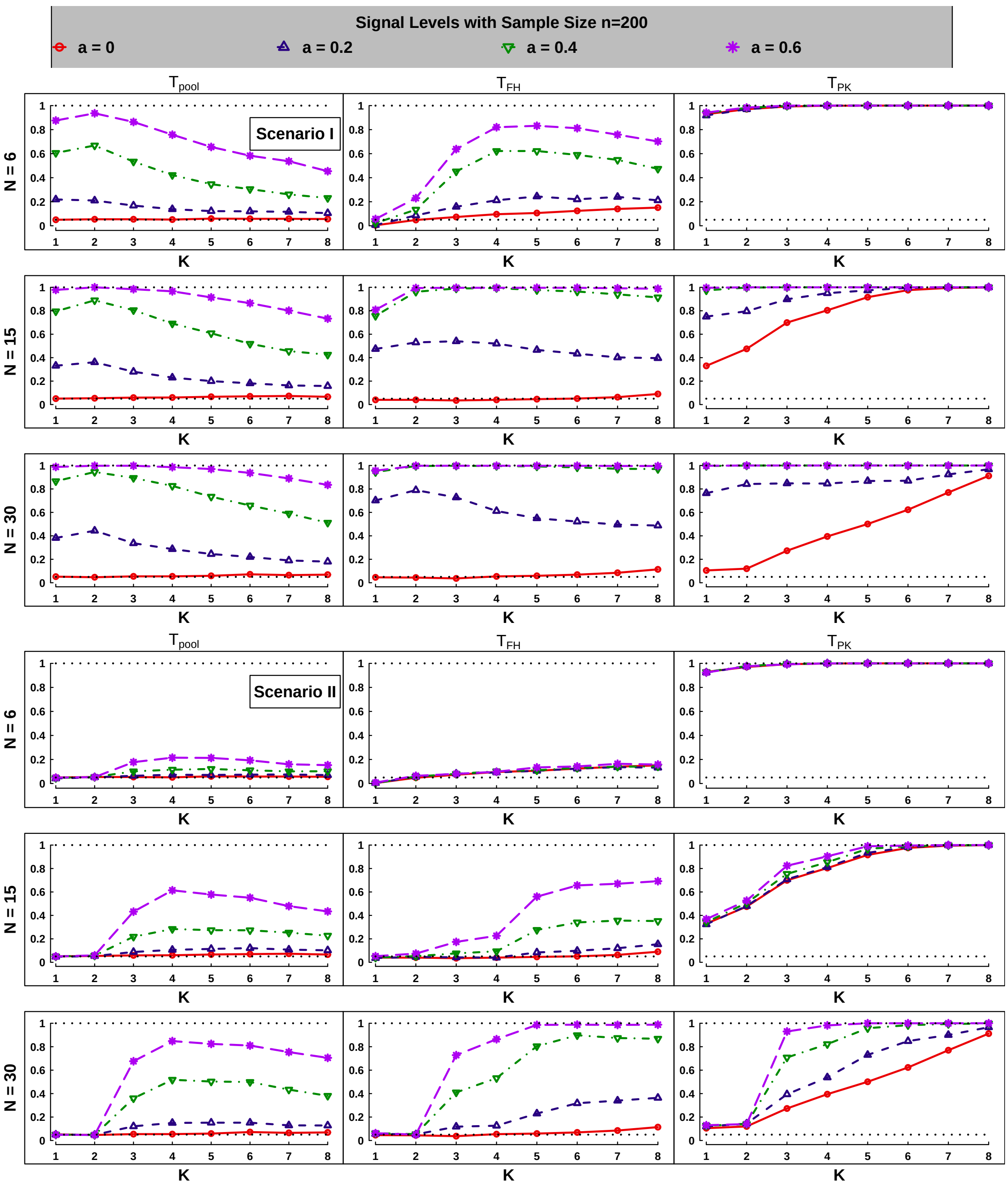

Figure S.3: Empirical sizes and powers under Scenarios I and II (two sub-figures). Rejection rates of three statistics (from left to right) with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under $n = m = 200$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

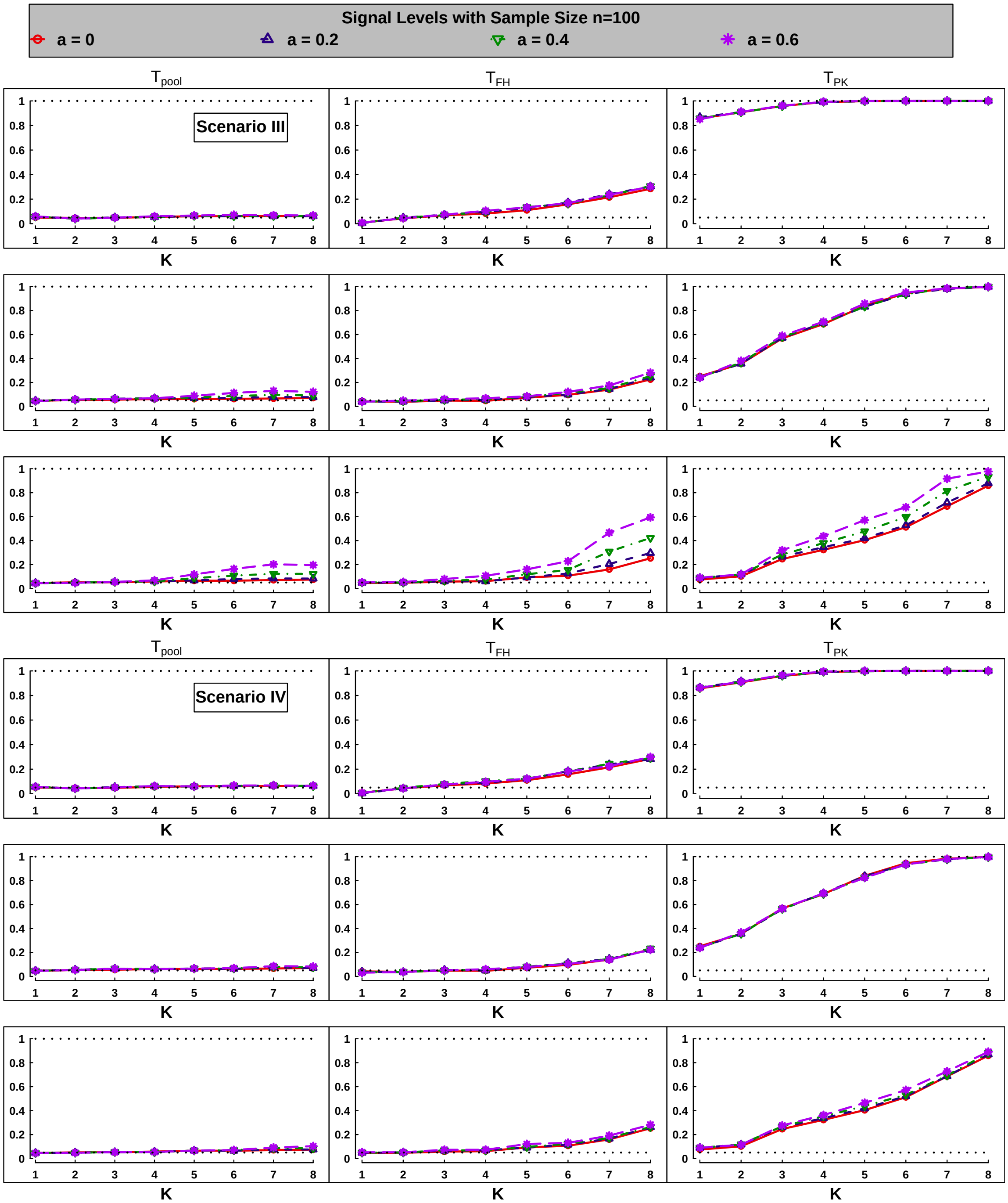

Figure S.4: Empirical sizes and powers under Scenarios III and IV (two sub-figures). Rejection rates of three statistics (from left to right) with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under $n = m = 100$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

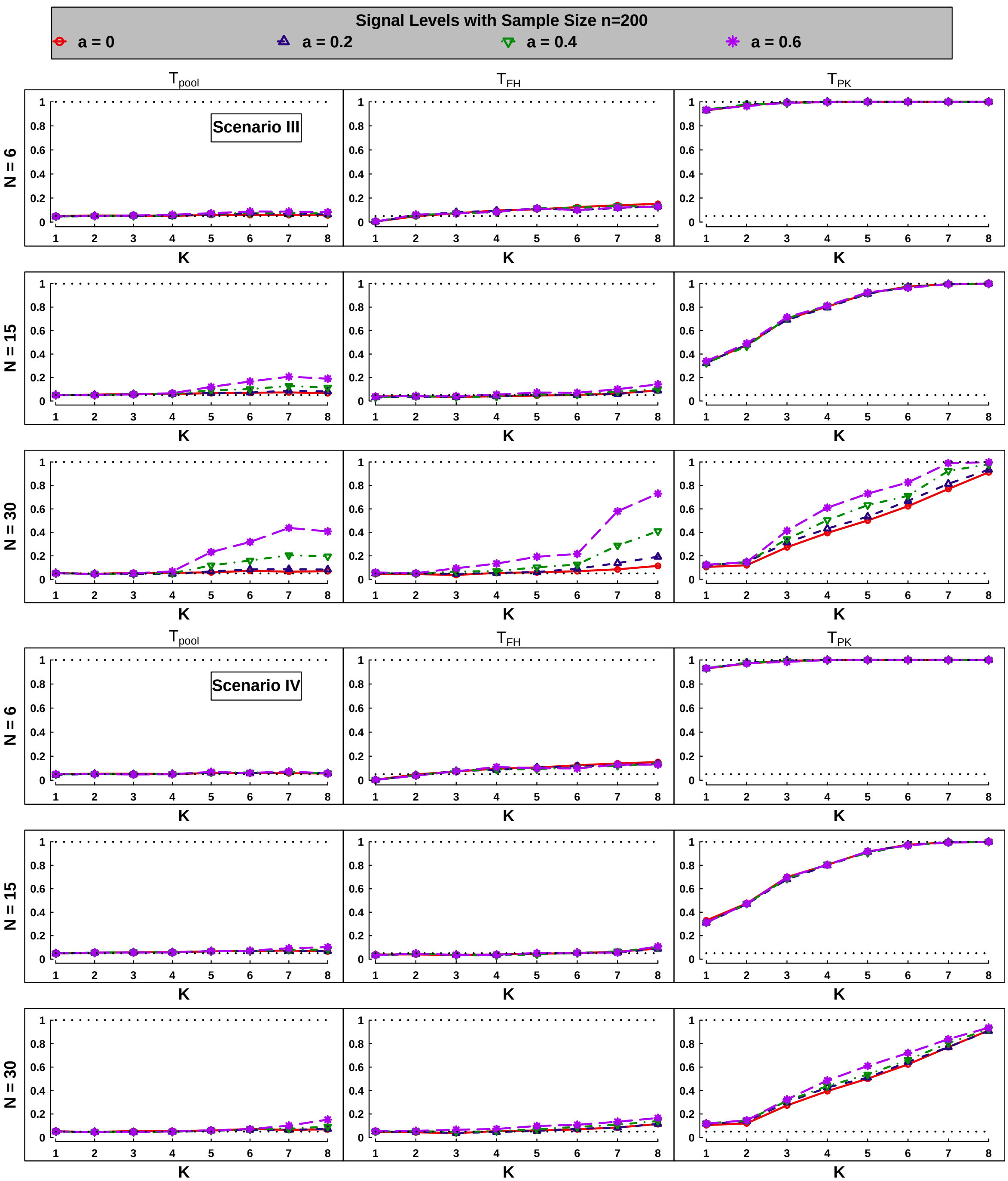

Figure S.5: Empirical sizes and powers under Scenarios III and IV (two sub-figures). Rejection rates of three statistics (from left to right) with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under $n = m = 200$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

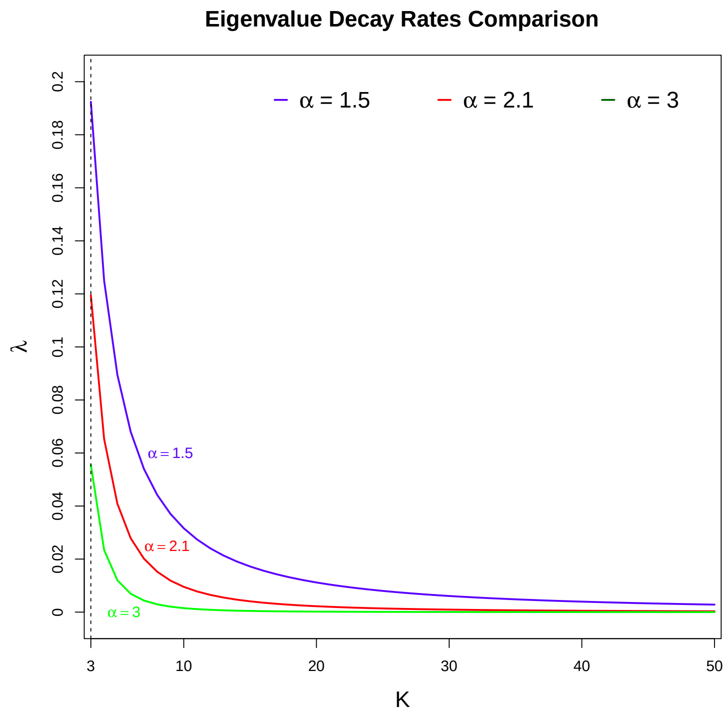

Figure S.6: Three kinds of eigenvalues under different decay rates, where the smoothing parameter α varies in $\{1.5, 2.1, 3\}$.

(b) Medium: $\lambda_1 = 1$, $\lambda_2 = 0.6$ and $\lambda_j = 1.2 \cdot j^{-2.1}$ with $3 \leq j \leq 50$;

(c) Fast: $\lambda_1 = 1$, $\lambda_2 = 0.5$ and $\lambda_j = 1.5 \cdot j^{-3}$ with $3 \leq j \leq 50$.

The case (a) is the setting of the main paper, while cases (b) and (c) are the new settings. It is easy to check that the 95% PVE threshold for three settings are 25, 8 and 4, respectively, representing different decay rates from slow to fast. The curves of values are plotted in Figure S.6 below. The results of cases (b) and (c) for $n = m = 100$ are reported in Figures S.7 and S.8, respectively, and the results of case $n = m = 200$ are omitted due to similarity. We can see that the results remain consistent with the conclusions drawn for case (a), implying the proposed test is insensitive to the smoothness parameter α. The comparisons among all tests also align with the discussions presented before.

For more straightforward comparative analysis across various decay rates, we focus on

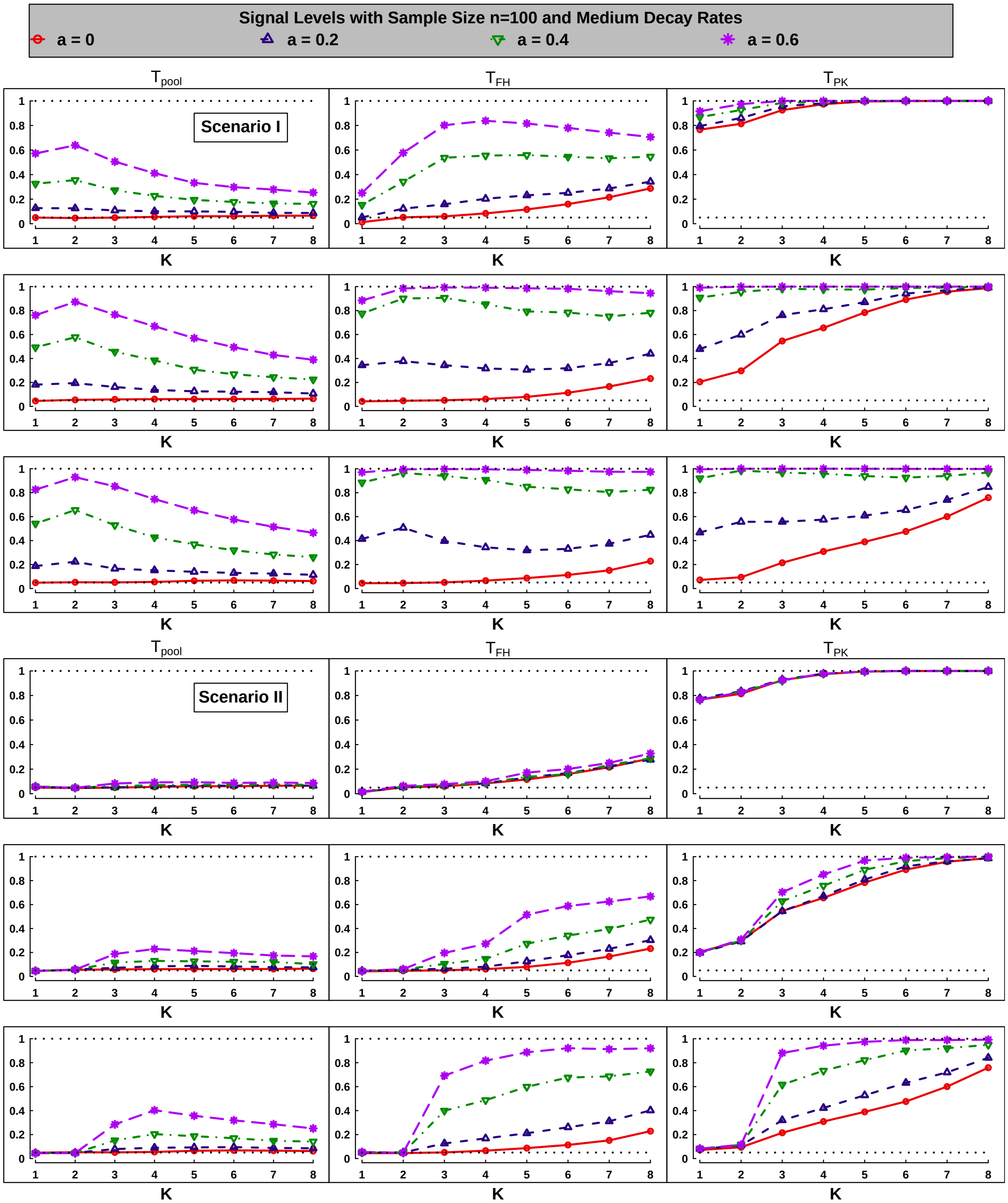

Figure S.7: Empirical sizes and powers under decay rate $\alpha = 2.1$ in Scenarios I and II (two sub-figures). Rejection rates of three statistics (from left to right) with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under $n = m = 100$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

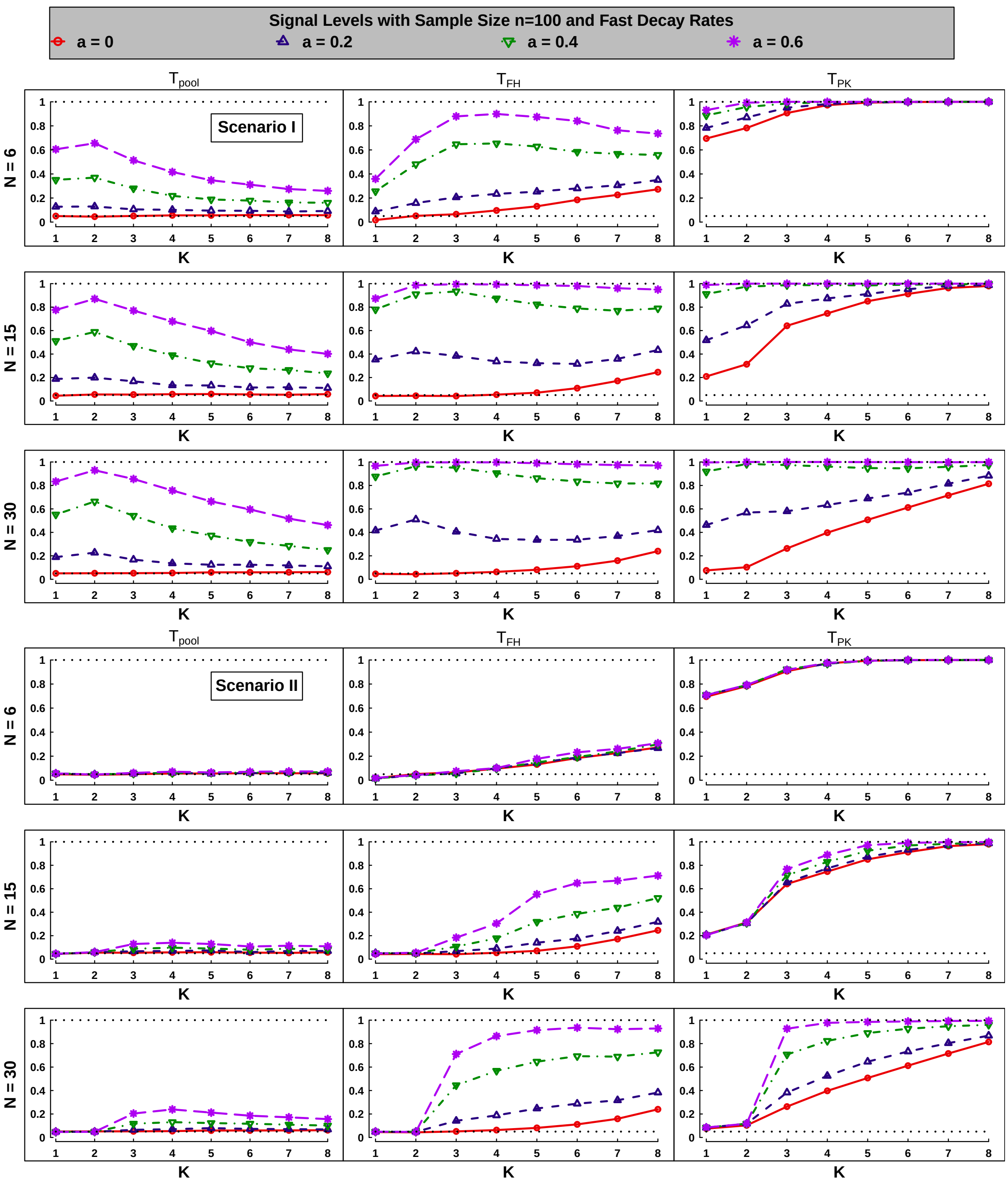

Figure S.8: Empirical sizes and powers under decay rate $\alpha = 3$ in Scenarios I and II (two sub-figures). Rejection rates of three statistics (from left to right) with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under $n = m = 100$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

the performance of the proposed test. The result is reported in Figure S.9. In Scenario I, i.e., the signals occur at $K = 1$ and $K = 2$, the sizes and powers of T_{pool} appear essentially unchanged across different α. This is reasonable that in our settings the first two eigenvalues are obviously larger than the remaining eigenvalues. When it comes to Scenario II ($K = 3$ and $K = 4$), it is seen that the sizes keep aligning with the nominal significance level 0.05. However, the powers gradually decrease as α increases from 1.5 to 3, since small eigenvalues leads to unstable estimation of variances.

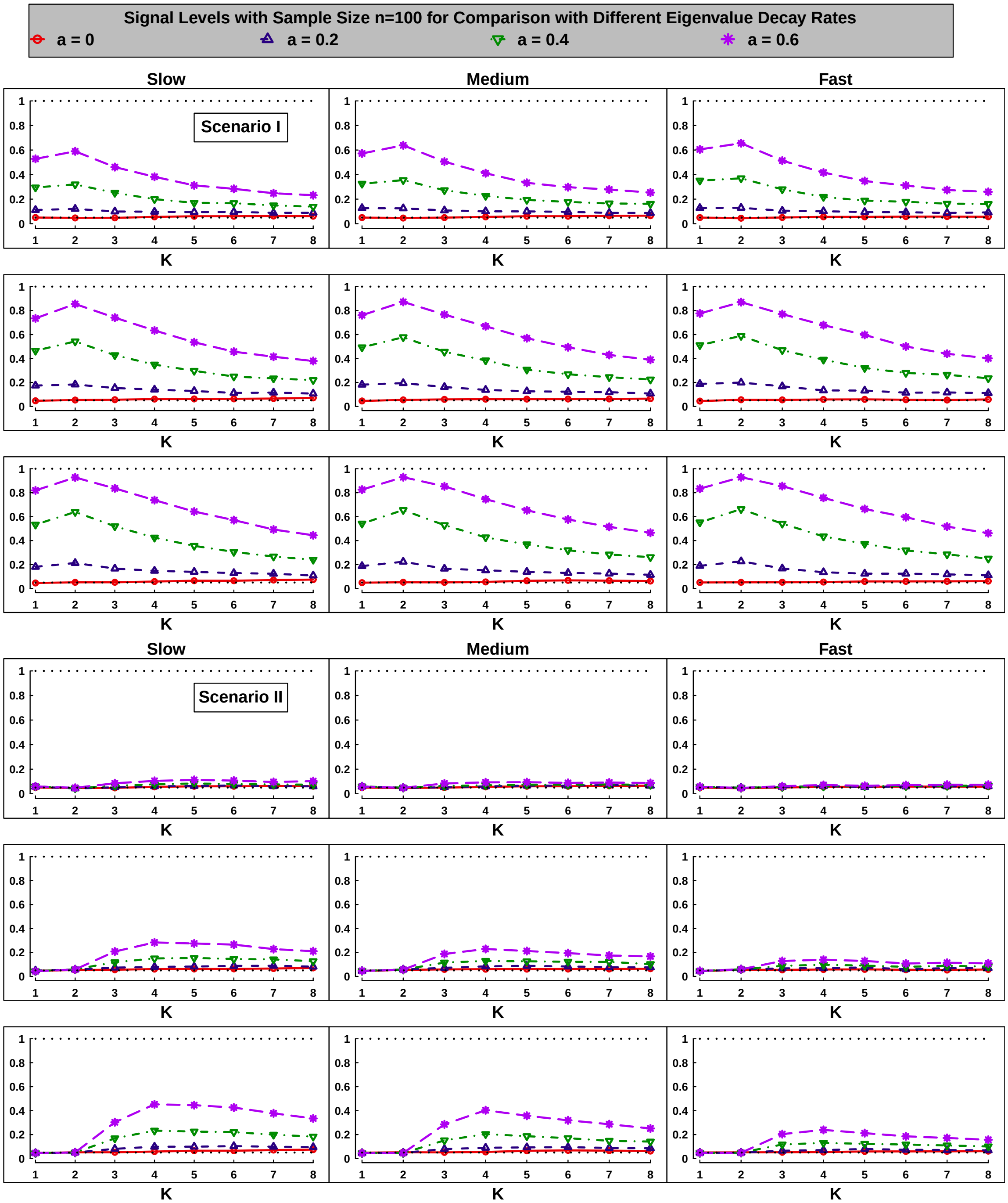

Figure S.9: Empirical sizes and powers under various decay rates $\alpha = 1.5, 2.1, 3$ in Scenarios I and II (two sub-figures). Rejection rates of three statistics (from left to right) with sampling frequencies $N = M = 6, 15, 30$ (from top to bottom) under $n = m = 100$ are depicted at the significant level 0.05 (dashed line). In each panel, power curves across different signal levels of $a = 0, 0.2, 0.4, 0.6$ vary with truncation level K from 1 to 8. In all cases, the tests are repeated 1000 times for size and 500 times for power.

References

Cai, T. T. and Yuan, M. (2011), "Optimal Estimation of the Mean Function Based on Discretely Sampled Functional Data: Phase Transition," *The Annals of Statistics*, 39, 2330–2355.

Fremdt, S., Steinebach, J., Horváth, L., and Kokoszka, P. (2013), "Testing the Equality of Covariance Operators in Functional Samples," *Scandinavian Journal of Statistics*, 40, 138–152.

Gautschi, W. (2011), *Numerical Analysis*, Birkhüser Boston, MA, 2nd ed.

Hall, P. and Horowitz, J. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91.

Hall, P. and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 68, 109–126.

Hsing, T. and Eubank, R. L. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley, West Sussex.

Huber, P. J. (1981), *Robust Statistics*, John Wiley & Sons, New York.

Li, Y. and Hsing, T. (2010), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *The Annals of Statistics*, 38, 3321–3351.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010), "Second-order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles," *Journal of the American Statistical Association*, 105, 670–682.

Prato, G. D. and Zabczyk, J. (2002), *Second Order Partial Differential Equations in Hilbert Spaces*, Cambridge, London.

Shao, L. X., Lin, Z. H., and Yao, F. (2022), "Intrinsic Riemannian Functional Data Analysis for Sparse Longitudinal Observations," *The Annals of Statistics*, 50, 1696–1721.

Tsybakov, A. B. (2009), *Introduction to Nonparametric Estimation*, Springer Publishing Company, 1st ed.

Yao, F., Müller, H. G., and Wang, J. L. (2005a), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590.

— (2005b), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903.

Zhang, X. and Wang, J. L. (2016), "From Sparse to Dense Functional Data and Beyond," *The Annals of Statistics*, 44, 2281–2321.

Zhou, H., Wei, D., and Yao, F. (2022), "Theory of Functional Principal Components Analysis for Discretely Observed Data," *arXiv:2209.08768v4*.