Supplementary Material on “Functional instrumental variable regression with an application to estimating the impact of immigration on native wages”

Dakyung Seong[a*] Won-Ki Seo[a]

[a] School of Economics, University of Sydney, Australia

Abstract

Sections S1-S3 of this supplementary material provide mathematical preliminaries and proofs of the theoretical results given in the main article. In connection with the numerical examples therein, Section S4 provides some additional simulation results and discussions. Section S5 discusses computation of the FIVE and F2SLSE. Section S6 provides an extension of the FIVE with a general weighting operator, and Section S7 discusses a significance test in functional endogenous linear models.

Contents

*Corresponding author. Address: School of Economics, University of Sydney, Camperdown, 2006, NSW, Australia. E-mail addresses: `dakyung.seong@sydney.edu.au (D. Seong), won-ki.seo@sydney.edu.au (W.-K. Seo)`

S1 Preliminaries

Let $(\mathbb{S}, \mathbb{F}, \mathbb{P})$ denote the underlying probability space and let $\mathcal{H}$ be a separable Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle$ and the usual Borel σ-field.

S1.1 Random elements of Hilbert spaces

A $\mathcal{H}$-valued random variable X is defined by a measurable map from $\mathbb{S}$ to $\mathcal{H}$. We say that such a random variable X is integrable (resp. square-integrable) if $\mathbb{E}[\|X\|] < \infty$ (resp. $\mathbb{E}[\|X\|^2] < \infty$), where $\|\cdot\|$ is the norm induced by the inner product. If X is integrable, there exists a unique element $\mathbb{E}[X] \in \mathcal{H}$ satisfying $\mathbb{E}[\langle X, \zeta \rangle] = \langle \mathbb{E}[X], \zeta \rangle$ for every $\zeta \in \mathcal{H}$. The element $\mathbb{E}[X]$ is called the expectation of X.

Let Y be another $\mathcal{H}$-valued random variable. We let $\otimes$ denote the tensor product defined as follows: for all $\zeta_1, \zeta_2 \in \mathcal{H}$,

$$\zeta_1 \otimes \zeta_2(\cdot) = \langle \zeta_1, \cdot \rangle \zeta_2. \tag{S1.1}$$

Note that $\zeta_1 \otimes \zeta_2$ is a linear map from $\mathcal{H}$ to $\mathcal{H}$. If $\mathbb{E}[\|X\|\|Y\|] < \infty$, we may well define a linear map $\mathcal{C}_{XY}$ from $\mathcal{H}$ to $\mathcal{H}$ as follows: $\mathcal{C}_{XY} = \mathbb{E}[(X - \mathbb{E}[X]) \otimes (Y - \mathbb{E}[Y])]$. $\mathcal{C}_{XY}$ is called the cross-covariance operator of X and Y. If $X = Y$ and X is square-integrable, we then may define $\mathcal{C}_{XX}$ similarly, and this is called the covariance operator of X. If the cross-covariance operator of two random variables X and Y is a nonzero operator, X is said to be correlated with Y.

S1.2 Bounded linear operators on $\mathcal{H}$

Let $\mathcal{L}_{\mathcal{H}}$ denote the space of bounded linear operators acting on $\mathcal{H}$, equipped with the operator norm $\|\mathcal{T}\|_{\text{op}} = \sup_{\|\zeta\| \leq 1} \|\mathcal{T}\zeta\|$. For any $\mathcal{T} \in \mathcal{L}_{\mathcal{H}}$, the adjoint of $\mathcal{T}$ (denoted $\mathcal{T}^*$) is the unique element of $\mathcal{L}_{\mathcal{H}}$ satisfying that $\langle \mathcal{T}\zeta_1, \zeta_2 \rangle = \langle \zeta_1, \mathcal{T}^*\zeta_2 \rangle$ for all $\zeta_1, \zeta_2 \in \mathcal{H}$. The range (denoted $\operatorname{ran} \mathcal{T}$) and kernel (denoted $\ker \mathcal{T}$) of $\mathcal{T} \in \mathcal{L}_{\mathcal{H}}$ are respectively given by $\operatorname{ran} \mathcal{T} = \{\mathcal{T}\zeta : \zeta \in \mathcal{H}\}$ and $\ker \mathcal{T} = \{\zeta \in \mathcal{H} : \mathcal{T}\zeta = 0\}$. $\mathcal{T}$ is said to be nonnegative if $\langle \mathcal{T}\zeta, \zeta \rangle \geq 0$ for all $\zeta \in \mathcal{H}$, and positive if the inequality is strict. An element $\mathcal{T} \in \mathcal{L}_{\mathcal{H}}$ is called compact if there exist two orthonormal bases $\{\zeta_{1j}\}_{j \geq 1}$ and $\{\zeta_{2j}\}_{j \geq 1}$ of $\mathcal{H}$, and a sequence of real numbers $\{a_j\}_{j \geq 1}$ tending to zero, such that $\mathcal{T} = \sum_{j=1}^{\infty} a_j \zeta_{1j} \otimes \zeta_{2j}$. In this expression, it may be assumed that $\zeta_{1j} = \zeta_{2j}$ and $a_1 \geq a_2 \geq \ldots \geq 0$ if $\mathcal{T}$ is self-adjoint (i.e., $\mathcal{T} = \mathcal{T}^*$) and nonnegative (Bosq, 2000, p.35). In this case, a_j becomes an eigenvalue of $\mathcal{T}$ and hence ζ_{1j} is the corresponding eigenfunction, and moreover, we may define $\mathcal{T}^{1/2}$ by replacing a_j with $\sqrt{a_j}$. It is well known that the covariance of a $\mathcal{H}$-valued random variable is self-adjoint, nonnegative and compact if exists. A linear operator $\mathcal{T}$ is called a Hilbert-Schmidt operator if its Hilbert-Schmidt norm $\|\mathcal{T}\|_{\text{HS}} = (\sum_{j=1}^{\infty} \|\mathcal{T}\zeta_j\|^2)^{1/2}$ is finite, where $\{\zeta_j\}_{j \geq 1}$ is an arbitrary orthonormal basis of $\mathcal{H}$. It is well known that $\|\mathcal{T}\|_{\text{op}} \leq \|\mathcal{T}\|_{\text{HS}}$ holds and the collection of Hilbert-Schmidt operators consists of a strict subclass of $\mathcal{L}_{\mathcal{H}}$; see Section 1.5 of Bosq (2000).

S2 Appendix to Section 3 on "Functional IV estimator"

We will hereafter let $\mathsf{a} = \alpha^{-1}$ and use both interchangeably for convenience.

S2.1 Proofs of the results in Section 3.2

Proof of Theorem 1

Note that

$$\widehat{\mathcal{A}} = \widehat{\mathcal{C}}_{yz}^*\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1} = \mathcal{A}\widehat{\Pi}_{\mathrm{K}} + \widehat{\mathcal{C}}_{uz}^*\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}, \tag{S2.1}$$

where $\widehat{\Pi}_{\mathrm{K}} = \sum_{j=1}^{\mathrm{K}} \widehat{f}_j \otimes \widehat{f}_j$. Since $\|\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}\|_{\mathrm{op}} \leq \mathsf{a}^{-1/2}$ and $\|\widehat{\mathcal{C}}_{uz}\|_{\mathrm{HS}} = O_p(T^{-1/2})$, we find that $\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\|_{\mathrm{HS}} \leq O_p(\mathsf{a}^{-1/2}T^{-1/2})$. Thus the proof becomes complete if $\|\mathcal{A}\widehat{\Pi}_{\mathrm{K}} - \mathcal{A}\|_{\mathrm{HS}} \xrightarrow{p} 0$ is shown. From nearly identical arguments used to derive (8.63) of Bosq (2000), it can be shown that

$$\|\mathcal{A}\widehat{\Pi}_{\mathrm{K}} - \mathcal{A}\|_{\mathrm{HS}}^2 \leq \sum_{j=\mathrm{K}+1}^{\infty} \|\mathcal{A}\widehat{f}_j\|^2 \leq \sum_{j=\mathrm{K}+1}^{\infty} \|\mathcal{A}f_j^s\|^2 + 2\|\mathcal{A}\|_{\mathrm{op}}^2\|\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz}\|_{\mathrm{op}}\sum_{j=1}^{\mathrm{K}}\tau_j, \tag{S2.2}$$

where $f_j^s = \mathrm{sgn}\{\langle \widehat{f}_j, f_j\rangle\}f_j$. Since $\mathcal{A}$ is Hilbert-Schmidt, $\sum_{j=\mathrm{K}+1}^{\infty}\|\mathcal{A}f_j^s\|^2$ converges in probability to zero as T gets larger (note that K diverges almost surely as $T \to \infty$). In addition,

$$\|\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz}\|_{\mathrm{op}} \leq \|\widehat{\mathcal{C}}_{xz}^*\|_{\mathrm{op}}\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\mathrm{op}} + \|\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*\|_{\mathrm{op}}\|\mathcal{C}_{xz}\|_{\mathrm{op}} = O_p(T^{-1/2}), \tag{S2.3}$$

which in turn implies that the second term of the right hand side of (S2.2) is $o_p(1)$. Combining all these results, we find that $\|\mathcal{A}\widehat{\Pi}_{\mathrm{K}} - \mathcal{A}\|_{\mathrm{HS}}$ is $o_p(1)$, which implies the desired result. □

Proof of Theorem 2

To show (i), we note that $\mathcal{C}_{xz}\mathcal{C}_{xz}^* = \sum_{j=1}^{\infty}\lambda_j^2\xi_j \otimes \xi_j$ and there exists an orthonormal basis $\{\widehat{\xi}_j\}_{j\geq 1}$ such that $\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^* = \sum_{j=1}^{\infty}\widehat{\lambda}_j^2\widehat{\xi}_j \otimes \widehat{\xi}_j$. Moreover, the following can be shown: $\widehat{\mathcal{C}}_{xz}^*\widehat{\xi}_j = \widehat{\lambda}_j\widehat{f}_j$, $\widehat{\mathcal{C}}_{xz}\widehat{f}_j = \widehat{\lambda}_j\widehat{\xi}_j$, $\mathcal{C}_{xz}^*\xi_j = \lambda_j f_j$, and $\mathcal{C}_{xz}f_j = \lambda_j\xi_j$. We first note that $\mathcal{C}_{xz}(\mathcal{C}_{xz}^*\mathcal{C}_{xz})_{\mathrm{K}}^{-1} = \sum_{j=1}^{\mathrm{K}}(\lambda_j^s)^{-1}f_j^s \otimes \xi_j^s$, where $f_j^s = \mathrm{sgn}\{\langle\widehat{f}_j, f_j\rangle\}f_j$, $\xi_j^s = \mathrm{sgn}\{\langle\widehat{\xi}_j, \xi_j\rangle\}\xi_j$ and $\lambda_j^s = \mathrm{sgn}\{\langle\widehat{f}_j, f_j\rangle\}\,\mathrm{sgn}\{\langle\widehat{\xi}_j, \xi_j\rangle\}\lambda_j$. Let $a_j = \mathrm{sgn}\{\langle\widehat{f}_j, f_j\rangle\}\,\mathrm{sgn}\{\langle\widehat{\xi}_j, \xi_j\rangle\}$ and assume $\lambda_j \geq 0$ and $\widehat{\lambda}_j \geq 0$ without loss of generality (see Bosq 2000, p. 34). From the foregoing properties of f_j, $\widehat{f}_j$, ξ_j and $\widehat{\xi}_j$, we have $a_j = \mathrm{sgn}\{\langle\widehat{f}_j, f_j\rangle\}\,\mathrm{sgn}\{\langle\widehat{f}_j, f_j\rangle + \lambda_j^{-2}\langle\mathcal{C}_{xz}^*(\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz})\widehat{f}_j, f_j\rangle\}$. Observe that

$$\|\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1} - \mathcal{C}_{xz}(\mathcal{C}_{xz}^*\mathcal{C}_{xz})_{\mathrm{K}}^{-1}\|_{\mathrm{op}} \leq \|\sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})f_j^s \otimes \xi_j^s\|_{\mathrm{op}} + \|\sum_{j=1}^{\mathrm{K}}\widehat{\lambda}_j^{-1}(\widehat{f}_j \otimes \widehat{\xi}_j - f_j^s \otimes \xi_j^s)\|_{\mathrm{op}}. \tag{S2.4}$$

The first term of (S2.4) is bounded above by $\sup_{1\leq j\leq \mathrm{K}}|\widehat{\lambda}_j^{-1} - (\lambda_j^s)^{-1}|$, where $|\widehat{\lambda}_j^{-1} - (\lambda_j^s)^{-1}| = |(\lambda_j^s - \widehat{\lambda}_j)\widehat{\lambda}_j^{-1}\lambda_j^{-1}| = |(\lambda_j^2 - \widehat{\lambda}_j^2)\widehat{\lambda}_j^{-1}\lambda_j^{-1}(\lambda_j^s + \widehat{\lambda}_j)^{-1}|$. Since $\sup_{1\leq j\leq \mathrm{K}}|\widehat{\lambda}_j^2 - \lambda_j^2| \leq \|\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}\mathcal{C}_{xz}^*\|_{\mathrm{op}}$ (Bosq, 2000, Lemma 4.2) and $|\lambda_j^s\widehat{\lambda}_j| \leq |\widehat{\lambda}_j^2 + \lambda_j^s\widehat{\lambda}_j|$ uniformly in $j = 1, \ldots, \mathrm{K}$ for large enough T (because $\sup_{1\leq j\leq \mathrm{K}}|a_j - 1| = o_p(1)$, which can be deduced from that $\sup_{1\leq j\leq \mathrm{K}}|\lambda_j^{-2}\langle\mathcal{C}_{xz}^*(\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz})\widehat{f}_j, f_j\rangle| \leq$

$\lambda_{\mathrm{K}}^{-2} O_p(T^{-1/2}) \xrightarrow{p} 0)$, we find that

$$\| \sum_{j=1}^{\mathrm{K}} ((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1}) f_j^s \otimes \xi_j^s \|_{\mathrm{op}} \leq \sup_{1\leq j\leq \mathrm{K}} \frac{|\widehat{\lambda}_j^2 - \lambda_j^2|}{|\lambda_j^s||\widehat{\lambda}_j \lambda_j^s + \widehat{\lambda}_j^2|} \leq \sup_{1\leq j\leq \mathrm{K}} \frac{|\widehat{\lambda}_j^2 - \lambda_j^2|}{|\lambda_j^s||\widehat{\lambda}_j \lambda_j^s|} \leq \frac{\|\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}\mathcal{C}_{xz}^*\|_{\mathrm{op}}}{\sqrt{\mathsf{a}}\lambda_{\mathrm{K}}^2}, \tag{S2.5}$$

As $\|\widehat{f}_j - f_j^s\| \leq \tau_j \|\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz}\|_{\mathrm{op}}$ and $\|\widehat{\xi}_j - \xi_j^s\| \leq \tau_j \|\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}\mathcal{C}_{xz}^*\|_{\mathrm{op}}$ (Bosq, 2000, Lemma 4.3),

$$\begin{aligned}\| \sum_{j=1}^{\mathrm{K}} \widehat{\lambda}_j^{-1} (\widehat{f}_j \otimes \widehat{\xi}_j - f_j^s \otimes \xi_j^s)\|_{\mathrm{op}} &\leq \mathsf{a}^{-1/2} \sum_{j=1}^{\mathrm{K}} (\|\widehat{f}_j - f_j^s\| + \|\widehat{\xi}_j - \xi_j^s\|) \\ &\leq \mathsf{a}^{-1/2} \sum_{j=1}^{\mathrm{K}} \tau_j (\|\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz}\|_{\mathrm{op}} + \|\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}\mathcal{C}_{xz}^*\|_{\mathrm{op}}).\end{aligned} \tag{S2.6}$$

From (S2.3), we know that $\|\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz}\|_{\mathrm{op}} = O_p(T^{-1/2})$ and $\|\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}\mathcal{C}_{xz}^*\|_{\mathrm{op}} = O_p(T^{-1/2})$. Moreover, it can be shown that $\lambda_{\mathrm{K}}^{-2} \leq (\lambda_{\mathrm{K}}^2 - \lambda_{\mathrm{K}+1}^2)^{-1} \leq \tau_{\mathrm{K}} \leq \sum_{j=1}^{\mathrm{K}} \tau_j$, so the terms given in the right hand sides of (S2.5) and (S2.6) are $o_p(1)$ under our assumptions. We thus deduce from (S2.4) that

$$\sqrt{\frac{T}{\theta_{\mathrm{K}}(\zeta)}} (\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}})\zeta = \left(\frac{1}{\sqrt{T\theta_{\mathrm{K}}(\zeta)}} \sum_{t=1}^{T} z_t \otimes u_t \right) \mathcal{C}_{xz} (\mathcal{C}_{xz}^* \mathcal{C}_{xz})_{\mathrm{K}}^{-1} \zeta + o_p(1).$$

Let $\zeta_t = (\theta_{\mathrm{K}}(\zeta))^{-1/2} [z_t \otimes u_t] \mathcal{C}_{xz} (\mathcal{C}_{xz}^* \mathcal{C}_{xz})_{\mathrm{K}}^{-1} \zeta = (\theta_{\mathrm{K}}(\zeta))^{-1/2} \langle z_t, \mathcal{C}_{xz} (\mathcal{C}_{xz}^* \mathcal{C}_{xz})_{\mathrm{K}}^{-1} \zeta \rangle u_t$. Then, we have

$$\mathbb{E}[\zeta_t \otimes \zeta_t] = \mathcal{C}_{uu}, \tag{S2.7}$$

as in the proof of Theorem 3.10 of Park and Qian (2012) by Assumption M.(c). Thus, under Assumption M, $\{\langle \zeta_t, \psi \rangle\}_{t\geq 1}$ is a martingale difference sequence for any $\psi \in \mathcal{H}$. By the standard central limit theorem for such a sequence, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \langle \zeta_t, \psi \rangle \xrightarrow{d} \mathcal{N}(0, \langle \mathcal{C}_{uu}\psi, \psi \rangle). \tag{S2.8}$$

Let $\ddot{\zeta}_T = T^{-1/2} \sum_{t=1}^{T} \zeta_t$. If there exists an orthonormal basis $\{\ell_j\}_{j\geq 1}$ satisfying

$$\limsup_{n\to\infty} \limsup_{T} \mathbb{P} \left(\sum_{j=n+1}^{\infty} \langle \ddot{\zeta}_T, \ell_j \rangle^2 > m \right) = 0 \tag{S2.9}$$

for every $m > 0$, then (S2.8) implies that $\ddot{\zeta}_T \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{uu})$ (van der Vaart and Wellner, 1996, Theorem 1.8.4). Let $\{\ell_j\}_{j\geq 1}$ be the eigenfunctions of $\mathcal{C}_{uu}$ and let $\mathcal{L}_n = \sum_{j=n+1}^{\infty} \ell_j \otimes \ell_j$. Observe that

$$\mathbb{E}\left[\sum_{j=n+1}^{\infty} \langle \ddot{\zeta}_T, \ell_j \rangle^2 \right] \leq \frac{1}{T\theta_{\mathrm{K}}(\zeta)} \sum_{t=1}^{T} \mathbb{E}\left[\langle z_t, \mathcal{C}_{xz} (\mathcal{C}_{xz}^* \mathcal{C}_{xz})_{\mathrm{K}}^{-1} \zeta \rangle^2 \|\mathcal{L}_n u_t\|^2 \right] = \sum_{j=n+1}^{\infty} \langle \mathcal{C}_{uu} \ell_j, \ell_j \rangle^2, \tag{S2.10}$$

where the equality follows from that $\{u_t\}_{t\geq 1}$ is a martingale difference sequence (with respect to $\mathfrak{F}_{t-1}$). Since $\mathcal{C}_{uu}$ is Hilbert-Schmidt, the right hand side of (S2.10) converges to zero as $n \to \infty$. Combining this result with the Markov's inequality, we find that for any $m > 0$, $\mathbb{P}(\sum_{j=n+1}^{\infty}\langle \ddot{\zeta}_T, \ell_j\rangle^2 > m) \leq m^{-1}\sum_{j=n+1}^{\infty}\langle \mathcal{C}_{uu}\ell_j, \ell_j\rangle^2$, from which (S2.9) follows. Thus, the desired result is obtained.

(ii) follows from that $\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}}$ and $\|\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz})_{\text{K}}^{-1} - \mathcal{C}_{xz}(\mathcal{C}_{xz}^{*}\mathcal{C}_{xz})_{\text{K}}^{-1}\|_{\text{op}}$ are all $o_p(1)$. □

S2.2 Proofs of the results in Section 3.3

We first state a useful lemma and then provide our proofs of the main results given in Section 3.3.

Lemma S1. *Suppose that Assumptions E2.(a) and E2.(b) hold and $\delta > 1$. Then the following hold: (i) $\sum_{\ell\neq j}\frac{\lambda_\ell^2\ell^{-\delta}}{(\lambda_j^2-\lambda_\ell^2)^2} \leq O(j^{\rho+2-\delta})$ and (ii) $\sum_{\ell\neq j}\frac{\lambda_j^2\ell^{-\delta}}{(\lambda_j^2-\lambda_\ell^2)^2} \leq O(j^{\rho+2-\delta})$.*

Proof of Lemma S1. We only show (i), because the remaining result can be obtained in a similar manner. As in Imaizumi and Kato (2018), we can choose $j_0 \geq 1$ and $C > 1$ large enough so that $\lambda_j^2/\lambda_{\lfloor j/C\rfloor}^2 \leq 1/2$ and $\lambda_{\lfloor jC\rfloor+1}^2/\lambda_j^2 \leq 1/2$ for all $j \geq j_0$, where $\lfloor\cdot\rfloor$ denotes the floor function. In addition, because of Assumption E2.(b) we have $(\lambda_j^2-\lambda_\ell^2)^2 \geq O(1)j^{-2\rho-2}(j-\ell)^2$ for $\ell\neq j$ and $\lfloor j/C\rfloor < \ell < \lfloor Cj\rfloor$ (Imaizumi and Kato 2018, p. 29). Using these, for $\delta > 1$, we find that

$$\sum_{\ell=1}^{\lfloor j/C\rfloor}\frac{\lambda_\ell^2\ell^{-\delta}}{(\lambda_j^2-\lambda_\ell^2)^2} \leq \sum_{\ell=1}^{\lfloor j/C\rfloor}\frac{\lambda_\ell^2\ell^{-\delta}}{\lambda_\ell^4(1-\lambda_j^2/\lambda_{\lfloor j/C\rfloor}^2)^2} \leq 4\sum_{\ell=1}^{\lfloor j/C\rfloor}\frac{\ell^{-\delta}}{\lambda_j^2} \leq O(j^{\rho}),$$

and

$$\sum_{\ell=\lfloor jC\rfloor+1}^{\infty}\frac{\lambda_\ell^2\ell^{-\delta}}{(\lambda_j^2-\lambda_\ell^2)^2} \leq 4\lambda_j^{-2}\sum_{\ell=\lfloor jC\rfloor+1}^{\infty}\ell^{-\delta} \leq O(j^{\rho}).$$

Moreover, by using the inequality $\lambda_\ell^2 \leq |\lambda_\ell^2-\lambda_j^2| + \lambda_j^2$ and the property stated at the beginning of the proof, we have

$$\begin{aligned}
\sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\lambda_\ell^2\ell^{-\delta}}{(\lambda_j^2-\lambda_\ell^2)^2} &\leq \sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\ell^{-\delta}}{|\lambda_\ell^2-\lambda_j^2|} + \sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\lambda_j^2\ell^{-\delta}}{(\lambda_j^2-\lambda_\ell^2)^2}\\
&\leq j^{1+\rho}\sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\ell^{-\delta}}{|\ell-j|} + j^{2+\rho}\sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\ell^{-\delta}}{|\ell-j|^2}\\
&\leq j^{1+\rho}\sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\ell^{-\delta+1}}{|\ell-j|^2} + 2j^{2+\rho}\sum_{\ell=\lfloor j/C\rfloor+1,\neq j}^{\lfloor jC\rfloor}\frac{\ell^{-\delta}}{|\ell-j|^2} \leq O(j^{2+\rho-\delta}),
\end{aligned}$$

where the last two inequalities are obtained by using the fact that $|\ell-j| \leq \ell+j$ and $\ell^{-\delta+1} \leq (j/C)^{-\delta+1} \leq C^{\delta-1}j^{-\delta+1}$ for all $\ell > \lfloor j/C\rfloor + 1$. □

Proof of Theorem 3

We first note that $\mathsf{a} \leq \widehat{\lambda}_{\mathrm{K}}^2 = \widehat{\lambda}_{\mathrm{K}}^2 - \lambda_{\mathrm{K}}^2 + \lambda_{\mathrm{K}}^2 \leq \|\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz}\|_{\mathrm{op}} + c_\circ \mathrm{K}^{-\rho} \leq O_p(T^{-1/2}) + c_\circ \mathrm{K}^{-\rho}$ and $\mathsf{a}^{-1}T^{-1/2} = o(1)$. These imply that

$$\mathrm{K} \leq (1 + o_p(1))\mathsf{a}^{-1/\rho}. \tag{S2.11}$$

Using the fact that $\lambda_j^2 \geq \sum_{\ell=j}^{\infty}(\lambda_\ell^2 - \lambda_{\ell+1}^2) \geq \rho^{-1}c_\circ^{-1}j^{-\rho}$ and $\mathsf{a}^{-1}T^{-1/2} = o(1)$, we also find that

$$(c_\circ\rho)^{-1}(\mathrm{K}+1)^{-\rho} \leq \lambda_{\mathrm{K}+1}^2 = \lambda_{\mathrm{K}+1}^2 - \widehat{\lambda}_{\mathrm{K}+1}^2 + \widehat{\lambda}_{\mathrm{K}+1}^2 \leq O_p(T^{-1/2}) + \mathsf{a} \leq (o_p(1) + 1)\mathsf{a}. \tag{S2.12}$$

We will now obtain stochastic orders of $\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\|_{\mathrm{HS}}$, $\|\mathcal{A}\widehat{\Pi}_{\mathrm{K}} - \mathcal{A}\Pi_{\mathrm{K}}\|_{\mathrm{HS}}$, and $\|\mathcal{A}(\mathcal{I} - \Pi_{\mathrm{K}})\|_{\mathrm{HS}}$. From (S6.6), we know that $\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\|_{\mathrm{HS}} \leq \|\widehat{\mathcal{C}}_{uz}\|_{\mathrm{HS}}\|\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}\|_{\mathrm{op}} \leq O_p(\mathsf{a}^{-1/2}T^{-1/2})$. Moreover, we have

$$\|\mathcal{A}(\mathcal{I} - \Pi_{\mathrm{K}})\|_{\mathrm{HS}}^2 = \sum_{\ell=\mathrm{K}+1}^{\infty}\sum_{j=1}^{\infty}\langle\mathcal{A}f_\ell, \xi_j\rangle^2 \leq c_\circ \sum_{\ell=\mathrm{K}+1}^{\infty}\sum_{j=1}^{\infty}\ell^{-2\varsigma}j^{-2\gamma} \leq O((\mathrm{K}+1)^{-2\varsigma+1}) \leq O_p(\mathsf{a}^{(2\varsigma-1)/\rho}), \tag{S2.13}$$

where the first inequality immediately follows from Assumption E2.(c), and the second and third inequalities follow from (S2.12), Assumption E2, and the Euler-Maclaurin summation formula for the Riemann zeta-function (see e.g., (5.6) of Ibukiyama and Kaneko, 2014). We then focus on the remaining term $\|\mathcal{A}\widehat{\Pi}_{\mathrm{K}} - \mathcal{A}\Pi_{\mathrm{K}}\|_{\mathrm{HS}}$ and find that

$$\|\mathcal{A}\widehat{\Pi}_{\mathrm{K}} - \mathcal{A}\Pi_{\mathrm{K}}\|_{\mathrm{HS}}^2 \leq 2\|\sum_{j=1}^{\mathrm{K}}\widehat{f}_j \otimes \mathcal{A}(\widehat{f}_j - f_j^s)\|_{\mathrm{HS}}^2 + 2\|\sum_{j=1}^{\mathrm{K}}(\widehat{f}_j - f_j^s) \otimes \mathcal{A}f_j^s\|_{\mathrm{HS}}^2. \tag{S2.14}$$

We observe that

$$\|\widehat{f}_j - f_j^s\|^2 = O_p(j^2T^{-1}), \tag{S2.15}$$

$$\|\mathcal{A}(\widehat{f}_j - f_j^s)\|^2 = O_p(T^{-1})(j^{2-2\varsigma} + j^{\rho+2-2\varsigma}). \tag{S2.16}$$

These will be proved below after discussing the main result of interest. Specifically, we can show the second term in (S2.14) satisfies that

$$\begin{aligned}
\|\sum_{j=1}^{\mathrm{K}}(\widehat{f}_j - f_j^s) \otimes \mathcal{A}f_j^s\|_{\mathrm{HS}}^2 &= \sum_{\ell=1}^{\infty}\|\sum_{j=1}^{\mathrm{K}}\langle\mathcal{A}f_j, f_\ell\rangle(\widehat{f}_j - f_j^s)\|^2 \leq \sum_{\ell=1}^{\infty}\left(\sum_{j=1}^{\mathrm{K}}|\langle\mathcal{A}f_j, f_\ell\rangle|\|\widehat{f}_j - f_j^s\|\right)^2 \\
&\leq \sum_{\ell=1}^{\infty}\ell^{-2\gamma}\left(\sum_{j=1}^{\mathrm{K}}j^{-\varsigma}\|\widehat{f}_j - f_j^s\|\right)^2 = O_p(T^{-1})\left(\sum_{j=1}^{\mathrm{K}}j^{1-\varsigma}\right)^2 \\
&= \begin{cases} O_p(T^{-1}) & \text{if } \varsigma > 2, \\ O_p(T^{-1}\max\{\log^2\mathsf{a}^{-1}, \mathsf{a}^{(2\varsigma-4)/\rho}\}) & \text{if } \varsigma \leq 2, \end{cases}
\end{aligned} \tag{S2.17}$$

where the first equality follows from the properties of the Hilbert-Schmidt norm and the remaining relationships follow from (S2.15), Assumption E2, and the fact that $\sum_{j=1}^{\mathrm{K}} j^{1-\varsigma} = O_p(1)$ if $\varsigma > 2$ and $\sum_{j=1}^{\mathrm{K}} j^{1-\varsigma} = O_p(\max\{\log \mathsf{a}^{-1}, \mathsf{a}^{(\varsigma-2)/\rho}\})$ otherwise. Similarly, the first term in (S2.14) satisfies that

$$\|\sum_{j=1}^{\mathrm{K}} \widehat{f}_j \otimes \mathcal{A}(\widehat{f}_j - f_j)\|_{\mathrm{HS}}^2 = \sum_{j=1}^{\mathrm{K}} \|\mathcal{A}(\widehat{f}_j - f_j)\|^2 = O_p(T^{-1}) \sum_{j=1}^{\mathrm{K}} (j^{2-2\varsigma} + j^{\rho+2-2\varsigma})$$
$$= \begin{cases} O_p(T^{-1}) & \text{if } \varsigma > \rho/2 + 3/2, \\ O_p(T^{-1} \max\{\log \mathsf{a}^{-1}, \mathsf{a}^{(2\varsigma-\rho-3)/\rho}\}) & \text{if } \varsigma \leq \rho/2 + 3/2, \end{cases} \tag{S2.18}$$

where (S2.16) is used to establish the second equality. Since $2\varsigma - \rho - 3 < 2\varsigma - 4$, $\mathsf{a} \log \mathsf{a}^{-1} = o(1)$, and $\mathsf{a} \log^2 \mathsf{a}^{-1} = o(1)$, (S6.7) is deduced from (S2.13), (S2.17) and (S2.18).

Proofs of (S2.15) and (S2.16): We first show (S2.15). Note that for each j,

$$\widehat{f}_j - f_j^s = \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-1} \langle (\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz}) \widehat{f}_j, f_\ell^s \rangle f_\ell^s + \langle \widehat{f}_j - f_j^s, f_j^s \rangle f_j^s. \tag{S2.19}$$

Then, using the arguments used to derive (4.48) of Bosq (2000) and the expansion of $\langle \widehat{f}_j - f_j^s, f_\ell^s \rangle$ that was used to derive (S2.19), it can be shown that

$$\|\widehat{f}_j - f_j^s\|^2 \leq 4 \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz}) \widehat{f}_j, f_\ell \rangle^2. \tag{S2.20}$$

Since $\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz} = (\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*) \widehat{\mathcal{C}}_{xz} + \mathcal{C}_{xz}^* (\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz})$, the sum given in (S2.20) satisfies that

$$\sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\widehat{\mathcal{C}}_{xz}^* \widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^* \mathcal{C}_{xz}) \widehat{f}_j, f_\ell \rangle^2$$
$$\leq 2 \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*) \widehat{\lambda}_j \widehat{\xi}_j, f_\ell \rangle^2 + 2 \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}) \widehat{f}_j, \lambda_\ell \xi_\ell \rangle^2. \tag{S2.21}$$

The second term of (S2.21) satisfies that

$$\sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}) \widehat{f}_j, \lambda_\ell \xi_\ell \rangle^2$$
$$\leq 2 \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}) (\widehat{f}_j - f_j^s), \xi_\ell \rangle^2 + 2 \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}) f_j^s, \xi_\ell \rangle^2$$
$$\leq 2 \Delta_{1j} \|\widehat{f}_j - f_j^s\|^2 + 2 \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}) f_j^s, \xi_\ell \rangle^2, \tag{S2.22}$$

where $\Delta_{1j} = \sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\mathrm{op}}^2$. Similarly, for the first term of (S2.21), we have

$$\sum_{\ell \neq j} (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \widehat{\lambda}_j^2 \langle (\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*) \widehat{\xi}_j, f_\ell \rangle^2$$
$$\leq \sum_{\ell \neq j} \Big((\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*) \widehat{\xi}_j, f_\ell \rangle^2 + (\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-1} \langle (\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*) \widehat{\xi}_j, f_\ell \rangle^2 \Big)$$

$$
\begin{aligned}
&\le 2\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_\ell^2\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)(\widehat{\xi}_j-\xi_j^s),f_\ell\rangle^2+2\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_\ell^2\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)\xi_j^s,f_\ell\rangle^2\\
&\quad+2\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-1}\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)(\widehat{\xi}_j-\xi_j^s),f_\ell\rangle^2+2\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-1}\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)\xi_j^s,f_\ell\rangle^2\\
&\le 2(\Delta_{1j}+\Delta_{2j})\|\widehat{\xi}_j-\xi_j^s\|^2+2\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\widehat{\lambda}_j^2\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)\xi_j^s,f_\ell\rangle^2,
\end{aligned}
\tag{S2.23}
$$

where $\Delta_{2j}=\max_{\ell\neq j,1\le j\le \mathrm{K}}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-1}\|\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz}\|_{\mathrm{op}}^2$ and the second inequality simply follows from the decomposition $\widehat{\xi}_j=(\widehat{\xi}_j-\xi_j^s)+\xi_j^s$ and the fact that $(a+b)^2\le 2a^2+2b^2$ for $a,b\in\mathbb{R}$. Let

$$
\Delta_{3j}=\Delta_{3j,1}+\Delta_{3j,2}, \tag{S2.24}
$$

where

$$
\Delta_{3j,1}=\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_\ell^2\langle(\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz})f_j^s,\xi_\ell\rangle^2,\quad \Delta_{3j,2}=\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\widehat{\lambda}_j^2\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)\xi_j^s,f_\ell\rangle^2. \tag{S2.25}
$$

We then deduce from (S2.20)-(S2.25) that

$$
\|\widehat{f}_j-f_j^s\|^2\le 16\Delta_{1j}\|\widehat{f}_j-f_j^s\|^2+16(\Delta_{1j}+\Delta_{2j})\|\widehat{\xi}_j-\xi_j^s\|^2+16\Delta_{3j}. \tag{S2.26}
$$

A similar bound for $\|\widehat{\xi}_j-\xi_j^s\|^2$ can be obtained from nearly identical arguments to derive (S2.26), from which the following can be deduced with a little algebra:

$$
\|\widehat{f}_j-f_j^s\|^2\le\frac{16(1+16\Delta_{2j})}{(1-16\Delta_{1j})^2-16^2(\Delta_{1j}+\Delta_{2j})^2}\Delta_{3j}.
$$

From (S2.12), the condition $\alpha=\mathsf{a}^{-1}=o(T^{\rho/(2\rho+2)})$, and similar arguments used in the proof of Theorem 1 of Imaizumi and Kato (2018), we find that

$$
\mathbb{P}\{|\widehat{\lambda}_j^2-\lambda_\ell^2|\ge|\lambda_j^2-\lambda_\ell^2|/\sqrt{2},\text{for } j=1,\dots,\mathrm{K}\ \text{ and } \ell\neq j\}\to 1. \tag{S2.27}
$$

Because of (S2.11), (S2.27), and the condition $\alpha=\mathsf{a}^{-1}=o(T^{\rho/(2\rho+2)})$, we find that $\Delta_{1j}\le O_p(T^{-1})\max_{\ell\neq j,1\le j\le\mathrm{K}}|\widehat{\lambda}_j^2-\lambda_\ell^2|^{-2}\le O_p(T^{-1}\mathrm{K}^{2\rho+2})\le O_p(T^{-1}\mathsf{a}^{-(2\rho+2)/\rho})=o_p(1)$. Similarly, from (S2.11), (S2.27) and Assumption E2.(b), we also deduce that $\Delta_{2j}\le O_p(\mathrm{K}^{\rho+1}\|\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}\|_{\mathrm{op}}^2)\le O_p(\mathsf{a}^{-(\rho+1)/\rho}T^{-1})=o_p(1)$. Combining these results, the following is established:

$$
\|\widehat{f}_j-f_j^s\|^2\le 16(1+o_p(1))\Delta_{3j}. \tag{S2.28}
$$

We now focus on Δ_{3j}. First note that

$$
|\Delta_{3j,1}|\le O_p(1)\sum_{\ell\neq j}(\lambda_j^2-\lambda_\ell^2)^{-2}\lambda_\ell^2\langle(\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz})f_j^s,\xi_\ell\rangle^2\le O_p(j^2T^{-1});
$$

this may be deduced from (S2.27), Assumption E2, Lemma S1(i) and the fact that

$$T\mathbb{E}[\langle(\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz})f_j^s,\xi_\ell\rangle^2] \le \sum_{s=0}^{T}\mathbb{E}[\upsilon_t(j,\ell)\upsilon_{t-s}(j,\ell)] \le O(1)\mathbb{E}[\langle x_t,f_j\rangle^2\langle z_t,\xi_\ell\rangle^2]$$
$$\le \mathbb{E}[\|\zeta_\ell\|^2\|\langle x_t,f_j\rangle z_t\|^2] \le O(1)\lambda_j^2,$$

where the second inequality follows from Assumption E2.(d). To obtain the third inequality, we note that $\mathbb{E}[\langle x_t,f_j\rangle^2\langle z_t,\xi_\ell\rangle^2]=\mathbb{E}[\langle\langle x_t,f_j\rangle z_t,\xi_\ell\rangle\langle\langle z_t,\xi_\ell\rangle x_t,f_j\rangle]\le\mathbb{E}[\|\langle x_t,f_j\rangle z_t\|\|\langle z_t,\xi_\ell\rangle x_t\|]$ and then apply the Cauchy-Schwarz inequality and Assumption E2.(d). In a similar manner, we also find that $|\Delta_{3j,2}|\le O_p(j^2T^{-1})$, and thus $|\Delta_{3j}|\le O_p(j^2T^{-1})$. Combining this result with (S2.28), we find that the desired result (S2.15) holds.

We next show (S2.16). Note that

$$\mathcal{A}(\widehat{f}_j-f_j^s)=\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-1}\langle(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz}^*\mathcal{C}_{xz})\widehat{f}_j,f_\ell^s\rangle\mathcal{A}f_\ell^s+\langle\widehat{f}_j-f_j^s,f_j^s\rangle\mathcal{A}f_j^s, \tag{S2.29}$$

where $\|\langle\widehat{f}_j-f_j^s,f_j^s\rangle\mathcal{A}f_j^s\|^2\le O_p(T^{-1})j^{2-2\varsigma}$. For each $j=1,\ldots,\mathrm{K}$, the first term in (S2.29) is bounded above as follows:

$$\begin{aligned}
&(\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-1}\langle(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz}^*\mathcal{C}_{xz})\widehat{f}_j,f_\ell^s\rangle\mathcal{A}f_\ell^s)^2\le O(1)(\sum_{\ell\neq j}|\widehat{\lambda}_j^2-\lambda_\ell^2|^{-1}\ell^{-\varsigma}|\langle(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz}^*\mathcal{C}_{xz})\widehat{f}_j,f_\ell\rangle|)^2\\
&\le O(1)(\sum_{\ell\neq j}|\widehat{\lambda}_j^2-\lambda_\ell^2|^{-1}\ell^{-\varsigma}\lambda_\ell|\langle(\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz})\widehat{f}_j,\xi_\ell\rangle|)^2+O(1)(\sum_{\ell\neq j}|\widehat{\lambda}_j^2-\lambda_\ell^2|^{-1}\ell^{-\varsigma}\widehat{\lambda}_j|\langle(\widehat{\mathcal{C}}_{xz}^*-\mathcal{C}_{xz}^*)\widehat{\xi}_j,f_\ell\rangle|)^2\\
&\le O(1)\|\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz}\|_{\mathrm{op}}^2\left(\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_\ell^2\ell^{-2\varsigma}+\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\widehat{\lambda}_j^2\ell^{-2\varsigma}\right)\\
&\le O_p(T^{-1})\left(\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_\ell^2\ell^{-2\varsigma}+\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}(\widehat{\lambda}_j^2-\lambda_j^2)\ell^{-2\varsigma}+\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_j^2\ell^{-2\varsigma}\right)\\
&\le O_p(T^{-1})\left(j^{\rho-2\varsigma+2}+(O_p(T^{-1/2})\lambda_j^{-2}+1)\sum_{\ell\neq j}(\widehat{\lambda}_j^2-\lambda_\ell^2)^{-2}\lambda_j^2\ell^{-2\varsigma}\right)\\
&\le(1+O_p(T^{-1/2}\mathsf{a}^{-1}))O_p(T^{-1}j^{\rho-2\varsigma+2})\le(1+o_p(1))O_p(T^{-1}j^{\rho-2\varsigma+2}),
\end{aligned} \tag{S2.30}$$

where the second inequality follows from Assumption E2, and the remaining inequalities follow from the Hölder's inequality, Lemma S1(i)-(ii), and the fact that $\|\widehat{\mathcal{C}}_{xz}-\mathcal{C}_{xz}\|_{\mathrm{op}}=O_p(T^{-1/2})$. From (S2.29) and (S2.30), we find that (S2.16) holds. □

Proof of Theorem 4

In this proof, we consider the scenario where $\rho/2+2\ge\varsigma+\delta_\zeta$, thus incorporating the complementary result presented in Section S2.3.2. The whole proof is divided into two parts.

1. Proof of the convergence results: For the subsequent discussion, we first need to obtain an upper bound of $\langle \widehat{f}_j - f_j^s, \zeta\rangle$. From the expansion given in (S2.19), we have

$$\langle \widehat{f}_j - f_j^s, \zeta\rangle = \sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-1}\langle(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz})\widehat{f}_j, f_\ell\rangle\langle f_\ell, \zeta\rangle + \langle \widehat{f}_j - f_j^s, f_j^s\rangle\langle f_j^s, \zeta\rangle. \tag{S2.31}$$

Note that the second term in (S2.31) satisfies that $(\langle \widehat{f}_j - f_j, f_j^s\rangle\langle f_j^s, \zeta\rangle)^2 \leq O_p(T^{-1}j^{-2\delta_\zeta+2})$, for all $j = 1, \ldots, \mathrm{K}$. Moreover, the first term in (S2.31) satisfies the following:

$$\begin{aligned}
&(\sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-1}\langle(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz})\widehat{f}_j, f_\ell\rangle\langle f_\ell, \zeta\rangle)^2 \leq O(1)(\sum_{\ell\neq j}|\widehat{\lambda}_j^2 - \lambda_\ell^2|^{-1}\ell^{-\delta_\zeta}|\langle(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz})\widehat{f}_j, f_\ell\rangle|)^2\\
&\leq O(1)(\sum_{\ell\neq j}|\widehat{\lambda}_j^2 - \lambda_\ell^2|^{-1}\ell^{-\delta_\zeta}\lambda_\ell|\langle(\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz})\widehat{f}_j, \xi_\ell\rangle|)^2 + O(1)(\sum_{\ell\neq j}|\widehat{\lambda}_j^2 - \lambda_\ell^2|^{-1}\ell^{-\delta_\zeta}\widehat{\lambda}_j|\langle(\widehat{\mathcal{C}}_{xz}^* - \mathcal{C}_{xz}^*)\widehat{\xi}_j, f_\ell\rangle|)^2\\
&\leq O(1)\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\mathrm{op}}^2\left(\sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2}\lambda_\ell^2\ell^{-2\delta_\zeta} + \sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2}\widehat{\lambda}_j^2\ell^{-2\delta_\zeta}\right)\\
&\leq O_p(T^{-1})\left(\sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2}\lambda_\ell^2\ell^{-2\delta_\zeta} + \sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2}(\widehat{\lambda}_j^2 - \lambda_j^2)\ell^{-2\delta_\zeta} + \sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2}\lambda_j^2\ell^{-2\delta_\zeta}\right)\\
&\leq O_p(T^{-1})\left(j^{\rho-2\delta_\zeta+2} + (O_p(T^{-1/2})\lambda_j^{-2} + 1)\sum_{\ell\neq j}(\widehat{\lambda}_j^2 - \lambda_\ell^2)^{-2}\lambda_j^2\ell^{-2\delta_\zeta}\right)\\
&\leq O_p(T^{-1})j^{\rho-2\delta_\zeta+2}(1 + O_p(T^{-1/2})j^\rho),
\end{aligned}\tag{S2.32}$$

where the first inequality follows from the assumption on $\langle f_\ell, \zeta\rangle$ and the remaining inequalities are deduced from similar arguments that are used to obtain (S2.30). Because $T^{-1/2}j^\rho \leq O_p(T^{-1/2}\mathsf{a}^{-1}) = o_p(1)$ uniformly in $j = 1, \ldots, \mathrm{K}$, we conclude that, for each $j = 1, \ldots, \mathrm{K}$,

$$\langle \widehat{f}_j - f_j^s, \zeta\rangle^2 = O_p(T^{-1})j^{-2\delta_\zeta+2} + O_p(T^{-1})j^{-2\delta_\zeta+2+\rho}(1 + o_p(1)). \tag{S2.33}$$

We next show the following:

$$\|\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}\zeta - \mathcal{C}_{xz}(\mathcal{C}_{xz}^*\mathcal{C}_{xz})_{\mathrm{K}}^{-1}\zeta\|_{\mathrm{op}} = o_p(1). \tag{S2.34}$$

Given that $\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\mathrm{op}} = o_p(1)$, the asymptotic results given in Theorem 2.(i) and 2.(ii) are deduced without difficulty from (S2.34) and similar arguments used in our proofs given in Section S2. From the decomposition given in (S2.4), it can be deduced that the desired result (S2.34) is established if the following terms are all $o_p(1)$: $\|\sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})\langle f_j^s, \zeta\rangle\widehat{\xi}_j\|$, $\|\sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})\langle \widehat{f}_j - f_j^s, \zeta\rangle\widehat{\xi}_j\|$, $\|\sum_{j=1}^{\mathrm{K}}(\lambda_j^s)^{-1}\langle f_j^s, \zeta\rangle(\widehat{\xi}_j - \xi_j^s)\|$, and $\|\sum_{j=1}^{\mathrm{K}}(\lambda_j^s)^{-1}\langle \widehat{f}_j - f_j^s, \zeta\rangle\widehat{\xi}_j\|$.

First, from similar arguments used in our proof of Theorem 2, we may deduce the following:

$$\|\sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})\langle f_j^s, \zeta\rangle\widehat{\xi}_j\|^2 = \sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})^2\langle f_j^s, \zeta\rangle^2 \leq \sum_{j=1}^{\mathrm{K}}\frac{(\lambda_j^2 - \widehat{\lambda}_j^2)^2}{\lambda_j^2(\widehat{\lambda}_j^2 + \lambda_j^s\widehat{\lambda}_j)^2}c_\zeta j^{-2\delta_\zeta}$$

$$\le \sum_{j=1}^{\mathrm{K}} \frac{(\lambda_j^2 - \widehat{\lambda}_j^2)^2}{\lambda_j^4 \widehat{\lambda}_j^2} c_\zeta j^{-2\delta_\zeta} \le O_p(T^{-1}\mathsf{a}^{-1} \sum_{j=1}^{\mathrm{K}} j^{2\rho-2\delta_\zeta}) \le O_p(T^{-1} \max\{\mathsf{a}^{-(3\rho-2\delta_\zeta+1)/\rho}, \mathsf{a}^{-(1+\rho)/\rho}\}), \tag{S2.35}$$

where the last inequality follows from (S2.11). In addition, using (S2.33) and the arguments used to derive (S2.35), we find that

$$\begin{aligned}
&\|\sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})\langle \widehat{f}_j - f_j^s, \zeta\rangle \widehat{\xi}_j\|^2 = \sum_{j=1}^{\mathrm{K}}((\lambda_j^s)^{-1} - \widehat{\lambda}_j^{-1})^2 \langle \widehat{f}_j - f_j^s, \zeta\rangle^2 \\
&\le O_p(T^{-2}\mathsf{a}^{-1}) \sum_{j=1}^{\mathrm{K}} j^{2\rho-2\delta_\zeta+2} + O_p(T^{-2}\mathsf{a}^{-1}) \sum_{j=1}^{\mathrm{K}} j^{3\rho-2\delta_\zeta+2} \\
&\le O_p(T^{-2}) \max\{\mathsf{a}^{-(4\rho-2\delta_\zeta+3)/\rho}, \mathsf{a}^{-(1+\rho)/\rho}\}.
\end{aligned} \tag{S2.36}$$

Note that $T^{-1}\mathsf{a}^{-(1+\rho)/\rho} = o(1)$ and $T^{-2}\mathsf{a}^{-(4\rho-2\delta_\zeta+3)/\rho} = T^{-1}\mathsf{a}^{-(3\rho-2\delta_\zeta+1)/\rho}T^{-1}\mathsf{a}^{-(\rho+2)/\rho} = o(1)$ by (3.8). These imply that the terms given in (S2.35) and (S2.36) are all $o_p(1)$. We also find that

$$\begin{aligned}
&\|\sum_{j=1}^{\mathrm{K}}(\lambda_j^s)^{-1}\langle f_j^s, \zeta\rangle(\widehat{\xi}_j - \xi_j^s)\| \le \sum_{j=1}^{\mathrm{K}} \|\lambda_j^{-1}\langle f_j^s, \zeta\rangle(\widehat{\xi}_j - \xi_j^s)\| \\
&\le O_p(T^{-1/2}) \sum_{j=1}^{\mathrm{K}} j^{-\delta_\zeta+\rho/2+1} \le O_p(T^{-1/2} \max\{\mathsf{a}^{-1/\rho}, \mathsf{a}^{(\delta_\zeta-\rho/2-2)/\rho}\}) = o_p(1),
\end{aligned} \tag{S2.37}$$

where the first inequality follows from the triangular inequality and the second inequality is deduced from Assumption E2.(b) and the fact that $\|\widehat{\xi}_j - \xi_j^s\|^2 \le O_p(T^{-1}j^2)$ for $j = 1, \ldots, \mathrm{K}$ (this can obtained from nearly identical arguments used to derive (S2.15)). The remaining inequalities are deduced since $2\delta_\zeta > 1$ and (S2.11) holds. It only remains to show that $\|\sum_{j=1}^{\mathrm{K}}(\lambda_j^s)^{-1}\langle \widehat{f}_j - f_j^s, \zeta\rangle\widehat{\xi}_j\|^2 = o_p(1)$. This can be obtained from Assumption E2.(a) and (S2.33); specifically, we observe that

$$\|\sum_{j=1}^{\mathrm{K}}(\lambda_j^s)^{-1}\langle \widehat{f}_j - f_j^s, \zeta\rangle\widehat{\xi}_j\|^2 \le O_p(T^{-1}) \max\{\mathsf{a}^{-(2\rho-2\delta_\zeta+3)/\rho}, \mathsf{a}^{-1/\rho}\} = o_p(1). \tag{S2.38}$$

Hence, from the results given in (S2.35)-(S2.38), (S2.34) is established.

2. Analysis on the regularization bias: Next, we focus on the regularization bias term, $\|\mathcal{A}(\widehat{\Pi}_{\mathrm{K}} - \Pi_{\mathrm{K}})\zeta\|$. For convenience, we let

$$\mathcal{A}(\widehat{\Pi}_{\mathrm{K}} - \mathcal{I})\zeta = F_1 + F_2 + F_3 + F_4,$$

where $F_4 = \mathcal{A}(\Pi_{\mathrm{K}} - \mathcal{I})\zeta$,

$$F_1 = \sum_{j=1}^{\mathrm{K}}\langle \widehat{f}_j - f_j^s, \zeta\rangle\mathcal{A}(\widehat{f}_j - f_j^s), \quad F_2 = \sum_{j=1}^{\mathrm{K}}\langle f_j^s, \zeta\rangle\mathcal{A}(\widehat{f}_j - f_j^s), \quad F_3 = \sum_{j=1}^{\mathrm{K}}\langle \widehat{f}_j - f_j^s, \zeta\rangle\mathcal{A}f_j^s,$$

and thus $F_1 + F_2 + F_3 = \mathcal{A}(\widehat{\Pi}_{\mathrm{K}} - \Pi_{\mathrm{K}})\zeta$. Then, by using (S2.16) and (S2.33), we find that

$$\|F_1\| \leq \sum_{j=1}^{\mathrm{K}} |\langle \widehat{f}_j - f_j^s, \zeta\rangle| \|\mathcal{A}(\widehat{f}_j - f_j^s)\| \leq O_p(T^{-1}) \sum_{j=1}^{\mathrm{K}} j^{\rho-\varsigma-\delta_\zeta+2} \leq o_p(T^{-1/2}) \sum_{j=1}^{\mathrm{K}} j^{\rho/2-\varsigma-\delta_\zeta+1},$$

where the last bound is obtained since $\alpha = o(T^{\rho/(2\rho+2)})$. In a similar manner, it can be shown that

$$\|F_2\| \leq \sum_{j=1}^{\mathrm{K}} |\langle f_j^s, \zeta\rangle| \|\mathcal{A}(\widehat{f}_j - f_j^s)\| \leq O_p(T^{-1/2}) \sum_{j=1}^{\mathrm{K}} j^{\rho/2-\varsigma-\delta_\zeta+1},$$

and

$$\|F_3\| \leq \sum_{j=1}^{\mathrm{K}} |\langle \widehat{f}_j - f_j^s, \zeta\rangle| \|\mathcal{A} f_j^s\| \leq O_p(T^{-1/2}) \sum_{j=1}^{\mathrm{K}} j^{\rho/2-\varsigma-\delta_\zeta+1}.$$

Therefore, $\|F_1\|$, $\|F_2\|$ and $\|F_3\|$ are bounded by the following quantity:

$$O_p(T^{-1/2}) \sum_{j=1}^{\mathrm{K}} j^{\rho/2-\varsigma-\delta_\zeta+1} \leq \begin{cases} O_p(T^{-1/2}) & \text{if } \rho/2+2 < \varsigma + \delta_\zeta, \\ O_p(T^{-1/2} \max\{\log \mathsf{a}^{-1}, \mathsf{a}^{-(\rho/2-\varsigma-\delta_\zeta+2)/\rho}\}) & \text{if } \rho/2+2 \geq \varsigma + \delta_\zeta. \end{cases}$$

Lastly, the following can be shown:

$$\|F_4\|^2 \leq \sum_{j=\mathrm{K}+1}^{\infty} \|\langle f_j, \zeta\rangle \mathcal{A} f_j\|^2 \leq \sum_{j=\mathrm{K}+1}^{\infty} j^{-2\delta_\zeta} \|\mathcal{A} f_j\|^2 = O_p(\sum_{j=\mathrm{K}+1}^{\infty} j^{-2\delta_\zeta - 2\varsigma}) \leq O_p(\mathsf{a}^{(2\varsigma+2\delta_\zeta-1)/\rho}).$$

This concludes the proof. □

S2.3 Supplementary results

S2.3.1 Strong consistency of the FIVE

We first review essential mathematics to establish the strong consistency of our estimators. The space of Hilbert-Schmidt operators, denoted $\mathcal{S}_{\mathcal{H}}$, is a separable Hilbert space with respect to the inner product given by $\langle \mathcal{T}_1, \mathcal{T}_2\rangle_{\mathcal{S}_{\mathcal{H}}} = \sum_{j,k\geq 1} \langle \mathcal{T}_1 \zeta_{1j}, \zeta_{2k}\rangle \langle \mathcal{T}_2 \zeta_{1j}, \zeta_{2k}\rangle$ for two arbitrary orthonormal bases $\{\zeta_{1j}\}_{j\geq 1}$ and $\{\zeta_{2j}\}_{j\geq 1}$ of $\mathcal{H}$; this inner product does not depend on the choice of orthonormal bases (Bosq, 2000, Chapter 1). We then note that $\{x_t \otimes z_t - \mathcal{C}_{xz}\}_{t\geq 1}$ is a zero-mean stationary and geometrically strongly mixing sequence in $\mathcal{S}_{\mathcal{H}}$, and, in the sequel, employ the following assumption: below, $\{\Lambda_j\}_{j\geq 1}$ is the sequence of eigenvalues of the covariance operator of $d_t = x_t \otimes z_t - \mathcal{C}_{xz}$.

Assumption S1. (a) $\sup_{t\geq 1} \|x_t\| \leq m_x$, $\sup_{t\geq 1} \|z_t\| \leq m_z$, and $\sup_{t\geq 1} \|u_t\| \leq m_u$ a.s., (b) $\Lambda_j \leq ab^j$ for some $a > 0$ and $0 < b < 1$.

As shown by Corollaries 2.4 and 4.2 of Bosq (2000), Assumption S1 combined with Assumption M.(b) helps us obtain a stochastic bound of $\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\mathrm{op}}$, which is given as follows:

Lemma S2. *Under Assumptions M.(b) and S1, the following holds almost surely:*

$$\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\text{op}} = O(T^{-1/2}\log^{3/2} T).$$

We omit the proof of Lemma S2 since it is a direct consequence of Theorem 2.12 and Corollary 2.4 of Bosq (2000) and the fact that $\sup_{t\geq 1}\|d_t\| \leq m_d$ holds for some $m_d > 0$ under the employed assumptions. Based upon this result, we can establish the strong consistency of the FIVE as follows:

Theorem 1 (continued)**.** *If Assumption S1 is additionally satisfied, $\tau(\alpha) = o(T^{1/2}\log^{-3/2} T)$ a.s., and $\alpha T^{-1}\log T \to 0$, then $\|\widehat{\mathcal{A}} - \mathcal{A}\|_{\text{op}} \to 0$ a.s.*

Proof. It can be easily shown from our proof of Theorem 1 that $\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\text{K}}\|_{\text{op}} \leq \mathsf{a}^{-1/2}\|T^{-1}\sum_{t=1}^{T} z_t \otimes u_t\|_{\text{op}}$ holds a.s. Under Assumptions M and S1, the sequence of $z_t \otimes u_t$ is a martingale difference, and $\|z_t \otimes u_t\|_{\text{HS}}$ and $\mathbb{E}\|z_t \otimes u_t\|_{\text{HS}}^2$ are uniformly bounded, and we thus know from Theorem 2.14 of Bosq (2000) that $\|T^{-1}\sum_{t=1}^{T} z_t \otimes u_t\|_{\text{op}} = O(T^{-1/2}\log^{1/2} T)$, a.s. This implies that $\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\text{K}}\|_{\text{op}} = O(\mathsf{a}^{-1/2}T^{-1/2}\log^{1/2} T)$ a.s. Moreover, we note that $\|\mathcal{A}\widehat{\Pi}_{\text{K}} - \mathcal{A}\|_{\text{op}}^2$ is bounded above by the term given in the right hand side of (S2.2), and deduce from Lemma S2 that $\|\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}^*\mathcal{C}_{xz}\|_{\text{op}} = O(T^{-1/2}\log^{3/2} T)$ a.s. These results imply that $\|\mathcal{A}\widehat{\Pi}_{\text{K}} - \mathcal{A}\|_{\text{op}}^2 = o(1)$ a.s. □

S2.3.2 Refinements of the general asymptotic results for the FIVE

We now provide a complementary result to Theorem 4 for the case where $\rho/2 + 2 \geq \varsigma + \delta_\zeta$. Specifically, the following can be shown:

Theorem 4 (Continued)**.** *Let everything be as in Theorem 4 but with $\rho/2 + 2 \geq \varsigma + \delta_\zeta$. Then, Theorem 2 holds and*

$$\|\mathcal{A}(\widehat{\Pi}_{\text{K}} - \Pi_{\text{K}})\zeta\| = O_p(T^{-1/2}\max\{\log\alpha, \alpha^{(\rho/2-\varsigma-\delta_\zeta+2)/\rho}\}),$$
$$\|\mathcal{A}(\Pi_{\text{K}} - \mathcal{I})\zeta\| = O_p(\alpha^{(1/2-\varsigma-\delta_\zeta)/\rho}).$$

Our proof of the above result is already contained in the proof of Theorem 4 given in Section S2.2, and hence omitted.

S3 Appendix to Section 4 on "Functional two-stage least square estimator"

As in Section S2, we will hereafter let $\mathsf{a}_1 = \alpha_1^{-1}$ and $\mathsf{a}_2 = \alpha_2^{-1}$, and use them interchangeably. We first provide a useful lemma that is related to our discussion on the F2SLSE in Section 4.

Lemma S3. *There exist unique bounded linear operators $\mathcal{R}_{xz}$ and $\mathcal{R}_{yz}^*$ satisfying the following:*

$$\mathcal{C}_{zz}^{1/2}\mathcal{R}_{xz}\mathcal{C}_{xx}^{1/2} = \mathcal{C}_{xz}, \quad \mathcal{R}_{xz}[\text{ran}\,\mathcal{C}_{xx}^{1/2}]^\perp = \{0\}, \quad \mathcal{R}_{xz}^*[\text{ran}\,\mathcal{C}_{zz}^{1/2}]^\perp = \{0\},$$
$$\mathcal{C}_{yy}^{1/2}\mathcal{R}_{yz}^*\mathcal{C}_{zz}^{1/2} = \mathcal{C}_{yz}^*, \quad \mathcal{R}_{yz}^*[\text{ran}\,\mathcal{C}_{zz}^{1/2}]^\perp = \{0\}, \quad \mathcal{R}_{yz}[\text{ran}\,\mathcal{C}_{yy}^{1/2}]^\perp = \{0\},$$

where $V^\perp$ denotes the orthogonal complement of $V \subset \mathcal{H}$.

Lemma S3 directly follows from Theorem 1 of Baker (1973). From the properties of $\mathcal{R}_{xz}$ (resp. $\mathcal{R}_{yz}$) given above, it can be understood as the cross-correlation operator of x_t and z_t (resp. y_t and z_t). Let $\mathcal{C}_{zz}^{-1/2}$ be defined by $\sum_{j=1}^\infty \mu_j^{-1/2} g_j \otimes g_j$, which is not a bounded linear operator since $\mu_j \to 0$ as $j \to \infty$. However, even with this property, we know as a direct consequence of Lemma S3 that $\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz}$ and $\mathcal{C}_{yz}^*\mathcal{C}_{zz}^{-1/2}$ are well defined bounded linear operators and they are respectively given by

$$\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz} = \mathcal{R}_{xz}\mathcal{C}_{xx}^{1/2} \quad \text{and} \quad \mathcal{C}_{yz}^*\mathcal{C}_{zz}^{-1/2} = \mathcal{C}_{yy}^{1/2}\mathcal{R}_{yz}^*.$$

We thus find that

$$\mathcal{C}_{yz}^*\mathcal{C}_{zz}^{-1}\mathcal{C}_{xz} = \mathcal{C}_{yz}^*\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz} = \mathcal{C}_{yy}^{1/2}\mathcal{R}_{yz}^*\mathcal{R}_{yz}\mathcal{C}_{zz}^{1/2} =: \mathcal{P},$$
$$\mathcal{C}_{xz}^*\mathcal{C}_{zz}^{-1}\mathcal{C}_{xz} = \mathcal{C}_{xz}^*\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz} = \mathcal{C}_{xx}^{1/2}\mathcal{R}_{xz}^*\mathcal{R}_{xz}\mathcal{C}_{xx}^{1/2} =: \mathcal{Q}.$$

As desired, $\mathcal{P}$ and $\mathcal{Q}$ are uniquely defined elements of $\mathcal{L}_\mathcal{H}$, and moreover, they are compact since $\mathcal{C}_{xx}^{1/2}$ and $\mathcal{C}_{yy}^{1/2}$ are compact.

S3.1 Proofs of the results in Section 4.2

Proof of Theorem 5

Since $\|\widehat{\mathcal{C}}_{uz}\|_{\text{HS}} = O_p(T^{-1/2})$, we find that

$$\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\text{K}_2}\|_{\text{HS}} \leq \|\widehat{\mathcal{C}}_{uz}\|_{\text{HS}}\|(\widehat{\mathcal{C}}_{zz})_{\text{K}_1}^{-1/2}\|_{\text{op}}\|(\widehat{\mathcal{C}}_{zz})_{\text{K}_1}^{-1/2}\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{Q}}_{\text{K}_2}^{-1}\|_{\text{op}} \leq O_p(\mathtt{a}_1^{-1/4}\mathtt{a}_2^{-1/4}T^{-1/2}). \tag{S3.1}$$

Thus, $\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\text{K}_2}\|_{\text{HS}} = o_p(1)$, and hence it suffices to show that $\|\mathcal{A}\widetilde{\Pi}_{\text{K}_2} - \mathcal{A}\|_{\text{HS}}^2 = o_p(1)$. Note that

$$\|\mathcal{A}\widetilde{\Pi}_{\text{K}_2} - \mathcal{A}\|_{\text{HS}}^2 \leq \sum_{j=\text{K}_2+1}^{\infty} \|\mathcal{A}h_j^s\|^2 + |\mathcal{R}|, \tag{S3.2}$$

where $h_j^s = \text{sgn}\{\langle \widehat{h}_j, h_j\rangle\}h_j$ and $\mathcal{R} = \sum_{j=\text{K}_2+1}^\infty (\|\mathcal{A}\widehat{h}_j\|^2 - \|\mathcal{A}h_j^s\|^2)$. Since $\mathcal{A}$ is Hilbert-Schmidt, the first term of (S3.2) is $o_p(1)$. It thus only remains to verify that $|\mathcal{R}| = o_p(1)$. To show this, we first deduce the following inequality from similar arguments used to derive the equation between (8.62) and (8.63) of Bosq (2000):

$$|\mathcal{R}| \leq 2\|\mathcal{A}\|_{\text{op}}^2 \sum_{j=1}^{\text{K}_2} \|\widehat{h}_j - h_j^s\|. \tag{S3.3}$$

We find that $\mathcal{Q}\widehat{h}_j - \nu_j\widehat{h}_j = (\mathcal{Q} - \widehat{\mathcal{Q}})\widehat{h}_j + (\widehat{\nu}_j - \nu_j)\widehat{h}_j$. Hence, by Lemma 4.2 of Bosq (2000),

$$\|\mathcal{Q}\widehat{h}_j - \nu_j\widehat{h}_j\| \leq 2\|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\text{op}}. \tag{S3.4}$$

Moreover, it can be shown from similar arguments used in the proof of Lemma 4.3 of Bosq (2000) that $\|\widehat{h}_j - h_j^s\| \leq \tau_{2,j}\|\mathcal{Q}\widehat{h}_j - \nu_j\widehat{h}_j\|/2$, which, combined with (S3.4), implies that

$$\|\widehat{h}_j - h_j^s\| \leq \tau_{2,j}\|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\text{op}}. \tag{S3.5}$$

We then deduce from (S3.3) and (S3.5) that

$$|\mathcal{R}| \leq 2\|\mathcal{A}\|_{\text{op}}^2 \left(\sum_{j=1}^{\mathrm{K}_2} \tau_{2,j}\right) \|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\text{op}}. \tag{S3.6}$$

Then the following can be shown:

$$\|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\text{op}} \leq \|\mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{zz}\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^* - \mathcal{B}\mathcal{C}_{zz}\mathcal{B}^*\|_{\text{op}} + \|\mathcal{S}\|_{\text{op}}, \tag{S3.7}$$

where $\mathcal{S} = \widehat{\mathcal{C}}_{vz}^*\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^* + \mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{vz} + \widehat{\mathcal{C}}_{vz}^*(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{vz}$. Let $\Pi_{\mathrm{K}_1} = \sum_{j=1}^{\mathrm{K}_1} g_j \otimes g_j$ and let

$$\mathcal{T} = \mathcal{B}(\mathcal{I} - \Pi_{\mathrm{K}_1})\mathcal{C}_{zz}(\mathcal{I} - \Pi_{\mathrm{K}_1})\mathcal{B}^*.$$

We further find that

$$\begin{aligned}
&\|\mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{zz}\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^* - \mathcal{B}\mathcal{C}_{zz}\mathcal{B}^*\|_{\text{op}} \leq \|\mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{zz}\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^* - \mathcal{B}\Pi_{\mathrm{K}_1}\mathcal{C}_{zz}\Pi_{\mathrm{K}_1}\mathcal{B}^*\|_{\text{op}} + \|\mathcal{T}\|_{\text{op}} \\
&\leq \|\mathcal{B}\|_{\text{op}}^2\|\sum_{j=1}^{\mathrm{K}_1}(\widehat{\mu}_j - \mu_j)\widehat{g}_j \otimes \widehat{g}_j\|_{\text{op}} + 2\|\mathcal{B}\|_{\text{op}}^2\sum_{j=1}^{\mathrm{K}_1}\mu_j\|\widehat{g}_j - g_j^s\| + \|\mathcal{T}\|_{\text{op}} \\
&\leq \left(\|\mathcal{B}\|_{\text{op}}^2 + 2\|\mathcal{B}\|_{\text{op}}^2\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} + \|\mathcal{T}\|_{\text{op}} \leq O_p\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} + \|\mathcal{T}\|_{\text{op}},
\end{aligned} \tag{S3.8}$$

where $g_j^s = \text{sgn}\{\langle\widehat{g}_j, g_j\rangle\}g_j$. From (S3.6), (S3.7) and (S3.8), the following is established:

$$|\mathcal{R}| \leq \left\{O_p\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} + \|\mathcal{S}\|_{\text{op}} + \|\mathcal{T}\|_{\text{op}}\right\} O_p\left(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}\right). \tag{S3.9}$$

Since $\|\widehat{\mathcal{C}}_{vz}\|_{\text{HS}} = O_p(T^{-1/2})$, $\|\widehat{\Pi}_{\mathrm{K}_1}\|_{\text{op}} \leq 1$ and $\|(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\|_{\text{op}} \leq \mathtt{a}_1^{-1/2}$, we have

$$\|\mathcal{S}\|_{\text{op}} \leq O_p(T^{-1/2}) + O_p(\mathtt{a}_1^{-1/2}T^{-1}). \tag{S3.10}$$

Note that $\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} = O_p(T^{-1/2})$ and $\|\mathcal{T}\|_{\text{op}}\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} = o_p(1)$ (which follows from the fact that $\|\mathcal{T}\|_{\text{op}} \leq \|\mathcal{B}\|_{\text{op}}^2\sum_{j=\mathrm{K}_1+1}^{\infty}\mu_j$). Combining these results with (S3.9) and (S3.10), we find that $|\mathcal{R}| \leq O_p(T^{-1/2}(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j})(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j})) + O_p((T^{-1/2} + \mathtt{a}_1^{-1/2}T^{-1})\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j})$. Given that $\mathtt{a}_1^{-1}T^{-1} \to 0$ and $\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} \leq O_p((\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j})(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j})) = o_p(T^{1/2})$ (the inequality follows from that $\mu_1^{-1}\tau_{1,1}^{-1}(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}) \geq 1$), it may be easily deduced that $|\mathcal{R}| = o_p(1)$ as desired. □

Proof of Theorem 6

To show (i), we will first verify that

$$\|(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1} - (\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\mathcal{C}_{xz}\mathcal{Q}_{\mathrm{K}_2}^{-1}\|_{\text{op}} \leq E_1 + E_2 + E_3 = o_p(1), \tag{S3.11}$$

where E_1, E_2 and E_3 are defined as follows:

$$E_1 = \|((\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1} - (\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1})\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}}, \qquad E_2 = \|(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}(\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz})\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}},$$
$$E_3 = \|(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\mathcal{C}_{xz}(\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1} - \mathcal{Q}_{\mathrm{K}_2}^{-1})\|_{\mathrm{op}}.$$

Note first that $\|\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}} = O_p(\mathsf{a}_2^{-1/2})$, and thus

$$E_1 \le O_p(\mathsf{a}_2^{-1/2})\left(\|\sum_{j=1}^{\mathrm{K}_1}(\mu_j^{-1} - \widehat{\mu}_j^{-1})g_j^s \otimes g_j^s\|_{\mathrm{op}} + \|\sum_{j=1}^{\mathrm{K}_1}\widehat{\mu}_j^{-1}(\widehat{g}_j \otimes \widehat{g}_j - g_j^s \otimes g_j^s)\|_{\mathrm{op}}\right),$$

where $g_j^s = \mathrm{sgn}\{\langle \widehat{g}_j, g_j\rangle\}g_j$. We then find that

$$\|\sum_{j=1}^{\mathrm{K}_1}(\mu_j^{-1} - \widehat{\mu}_j^{-1})g_j^s \otimes g_j^s\|_{\mathrm{op}} \le \sup_{1\le j\le \mathrm{K}_1}|\widehat{\mu}_j^{-1} - \mu_j^{-1}| = \sup_{1\le j\le \mathrm{K}_1}\left|\frac{\widehat{\mu}_j - \mu_j}{\mu_j\widehat{\mu}_j}\right| \le \mathsf{a}_1^{-1/2}\mu_{\mathrm{K}_1}^{-1}\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\mathrm{op}} \tag{S3.12}$$

and

$$\|\sum_{j=1}^{\mathrm{K}_1}\widehat{\mu}_j^{-1}(\widehat{g}_j \otimes \widehat{g}_j - g_j^s \otimes g_j^s)\|_{\mathrm{op}} \le 2\mathsf{a}_1^{-1/2}\sum_{j=1}^{\mathrm{K}_1}\|\widehat{g}_j - g_j^s\| \le 2\mathsf{a}_1^{-1/2}\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\mathrm{op}}\sum_{j=1}^{\mathrm{K}_1}\tau_{1,j}. \tag{S3.13}$$

Since $\mu_{\mathrm{K}_1}^{-1} \le \sum_{j=1}^{\mathrm{K}_1}\tau_{1,j} = o_p(\mathsf{a}_1^{1/2}T^{1/2})$ and $\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\mathrm{op}} = O_p(T^{-1/2})$, the right hand sides of (S3.12) and (S3.13) are $o_p(1)$, and hence $E_1 = o_p(1)$. Since $\|(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\|_{\mathrm{op}} \le \mu_{\mathrm{K}_1}^{-1}$, $\|\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}} \le \mathsf{a}_2^{-1/2}$, and $\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\mathrm{op}} = O_p(T^{-1/2})$, we also find that

$$E_2 \le \mu_{\mathrm{K}_1}^{-1}O_p(\mathsf{a}_2^{-1/2}T^{-1/2}) \le O_p(\mathsf{a}_2^{-1/2}T^{-1/2})\sum_{j=1}^{\mathrm{K}_1}\tau_{1,j} = o_p(1).$$

Given that $\|(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\mathcal{C}_{xz}\|_{\mathrm{op}} \le \|\mathcal{B}^*\|_{\mathrm{op}} = O_p(1)$, it remains to show that $\|\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1} - \mathcal{Q}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}} = o_p(1)$ since this implies $E_3 = o_p(1)$ (and thus the desired result (S3.11) is obtained). To show this, we first note that

$$\|\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1} - \mathcal{Q}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}} \le \|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1} - \widehat{\nu}_j^{-1})h_j \otimes h_j\|_{\mathrm{op}} + \|\sum_{j=1}^{\mathrm{K}_2}\widehat{\nu}_j^{-1}(\widehat{h}_j \otimes \widehat{h}_j - h_j \otimes h_j)\|_{\mathrm{op}}.$$

Let $\mathcal{S} = \widehat{\mathcal{C}}_{vz}^*\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^* + \mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{vz} + \widehat{\mathcal{C}}_{vz}^*(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{vz}$. We then deduce from our proof of Theorem 5 that

$$\|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\mathrm{op}} \le O_p\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\mathrm{op}} + \|\mathcal{S}\|_{\mathrm{op}} + \|\mathcal{T}\|_{\mathrm{op}}. \tag{S3.14}$$

As in (S3.12), it can be shown that $\|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1}-\widehat{\nu}_j^{-1})h_j\otimes h_j\|_{\mathrm{op}} \le \mathsf{a}_2^{-1/2}\nu_{\mathrm{K}_2}^{-1}\|\widehat{\mathcal{Q}}-\mathcal{Q}\|_{\mathrm{op}}$, hence

$$\|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1}-\widehat{\nu}_j^{-1})h_j\otimes h_j\|_{\mathrm{op}} \le \mathsf{a}_2^{-1/2}\nu_{\mathrm{K}_2}^{-1}\left(O_p\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\|\widehat{\mathcal{C}}_{zz}-\mathcal{C}_{zz}\|_{\mathrm{op}}+\|\mathcal{S}\|_{\mathrm{op}}+\|\mathcal{T}\|_{\mathrm{op}}\right). \tag{S3.15}$$

We also deduce the following from (S3.5), (S3.13) and (S3.14):

$$\|\sum_{j=1}^{\mathrm{K}_2}\widehat{\nu}_j^{-1}(\widehat{h}_j\otimes\widehat{h}_j-h_j\otimes h_j)\|_{\mathrm{op}} \le 2\mathsf{a}_2^{-1/2}\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}\left(O_p\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\|\widehat{\mathcal{C}}_{zz}-\mathcal{C}_{zz}\|_{\mathrm{op}}+\|\mathcal{S}\|_{\mathrm{op}}+\|\mathcal{T}\|_{\mathrm{op}}\right). \tag{S3.16}$$

We find that $\nu_{\mathrm{K}_2}^{-1} \le \sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} \le O_p((\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j})(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j})) = o_p(\mathsf{a}_2^{1/2}T^{1/2})$, which follows from that $\mu_1^{-1}\tau_{1,1}^{-1}(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}) \ge 1$. Moreover, note that

$$\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j} \le \nu_1\nu_{\mathrm{K}_2}^{-1}\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j} \le \nu_1\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\left(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}\right) = o_p(\mathsf{a}_2^{1/2}T^{1/2}),$$

and

$$\nu_{\mathrm{K}_2}^{-1}\|\mathcal{T}\|_{\mathrm{op}} \le \sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}\|\mathcal{T}\|_{\mathrm{op}} \le \|\mathcal{B}\|_{\mathrm{op}}^2\left(\sum_{j=\mathrm{K}_1+1}^{\infty}\mu_j\right)\left(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}\right) = o_p(\mathsf{a}_2^{1/2}).$$

From these results, (S3.10) and the fact that $\|\widehat{\mathcal{C}}_{zz}-\mathcal{C}_{zz}\|_{\mathrm{op}} = O_p(T^{-1/2})$, we may deduce that the right hand sides of (S3.15) and (S3.16) are all $o_p(1)$, and thus $E_3 = o_p(1)$ and (S3.11) holds.

We thus know that

$$\sqrt{T}(\widetilde{\mathcal{A}}-\mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2})\zeta = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_t\otimes u_t\right)(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\mathcal{C}_{xz}\mathcal{Q}_{\mathrm{K}_2}^{-1}\zeta + o_p(1).$$

Define $\zeta_t = (\phi_{\mathrm{K}_2}(\zeta))^{-1/2}[z_t\otimes u_t](\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\mathcal{C}_{xz}\mathcal{Q}_{\mathrm{K}_2}^{-1}\zeta$ and let $\ddot{\zeta}_T = T^{-1/2}\sum_{t=1}^{T}\zeta_t$. Then from nearly identical arguments used to derive (S2.7), (S2.8) and (S2.9), we find that, for any $\psi\in\mathcal{H}$ and $m>0$, $T^{-1/2}\sum_{t=1}^{T}\langle\zeta_t,\psi\rangle \xrightarrow{d} \mathcal{N}(0,\langle\mathcal{C}_{uu}\psi,\psi\rangle)$ and $\limsup_{n\to\infty}\limsup_T \mathbb{P}(\sum_{j=n+1}^{\infty}\langle\ddot{\zeta}_T,\ell_j\rangle^2 > m) = 0$, where $\{\ell_j\}_{j\ge 1}$ denote the eigenfunctions of $\mathcal{C}_{uu}$. Hence (i) is established.

Given that $\|\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1}-\mathcal{Q}_{\mathrm{K}_2}^{-1}\|_{\mathrm{op}} = o_p(1)$, (ii) is immediately deduced. □

S3.2 Proofs of the results in Section 4.3

We hereafter define

$$\mathcal{Q}_{\mathrm{K}_1} = \mathcal{C}_{xz}^*(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1}\mathcal{C}_{xz},$$

which is repeatedly used in the subsequent proofs.

Proof of Theorem 7

We will show the following:

$$\mathrm{K}_2 \le (1+o_p(1))\mathsf{a}_2^{-1/\rho_\nu}, \tag{S3.17}$$
$$(c_\circ\rho)^{-1}(\mathrm{K}_2+1)^{-\rho_\nu} \le (1+o_p(1))\mathsf{a}_2, \tag{S3.18}$$
$$\|\widehat{h}_j - h_j^s\|^2 \le O_p(\mathsf{a}_1) j^{\rho_\nu-4\gamma_\mu+2}, \tag{S3.19}$$
$$\|\mathcal{A}(\widehat{h}_j - h_j^s)\|^2 \le O_p(\mathsf{a}_1) j^{\rho_\nu-2\varsigma_\nu-4\gamma_\mu+2} + O_p(d_T) j^{\rho_\nu-2\varsigma_\nu+2}, \tag{S3.20}$$

where h_j^s is defined as in our proof of Theorem 5 and d_T is defined by

$$d_T = \mathsf{a}_1^{(4\varsigma_\mu+\rho_\mu-2)/\rho_\mu} + T^{-1}\max\{\mathsf{a}_1^{-1/\rho_\mu}, \mathsf{a}_1^{-(\rho_\mu-2\varsigma_\mu+3)/\rho_\mu}\}. \tag{S3.21}$$

Note that

$$\|\widetilde{\mathcal{A}} - \mathcal{A}\|_{\mathrm{HS}} \le \|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\|_{\mathrm{HS}} + \|\mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2} - \mathcal{A}\Pi_{\mathrm{K}_2}\|_{\mathrm{HS}} + \|(\mathcal{I}-\Pi_{\mathrm{K}_2})\mathcal{A}\|_{\mathrm{HS}},$$

where $\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\|_{\mathrm{HS}} = O_p(\mathsf{a}_1^{-1/4}\mathsf{a}_2^{-1/4}T^{-1/2})$ as shown in (S3.1). Using (S3.18), we also find that

$$\|(\mathcal{I}-\Pi_{\mathrm{K}_2})\mathcal{A}\|_{\mathrm{HS}}^2 = \sum_{\ell=\mathrm{K}_2+1}^{\infty} \|\mathcal{A}h_\ell\|^2 \le O(1)\sum_{\ell=\mathrm{K}_2+1}^{\infty}\sum_{j=1}^{\infty} \ell^{-2\varsigma_\nu} j^{-2\gamma_\nu} \le O_p(\mathsf{a}_2^{(2\varsigma_\nu-1)/\rho_\nu}). \tag{S3.22}$$

We next focus on the remaining term $\|\mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2} - \mathcal{A}\Pi_{\mathrm{K}_2}\|_{\mathrm{HS}}$. Note that

$$2^{-1}\|\mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2} - \mathcal{A}\Pi_{\mathrm{K}_2}\|_{\mathrm{HS}}^2 \le \|\sum_{j=1}^{\mathrm{K}_2} \widehat{h}_j \otimes \mathcal{A}(\widehat{h}_j - h_j^s)\|_{\mathrm{HS}}^2 + \|\sum_{j=1}^{\mathrm{K}_2}(\widehat{h}_j - h_j^s)\otimes \mathcal{A}h_j\|_{\mathrm{HS}}^2. \tag{S3.23}$$

We know from (S3.19) and (S3.23) that

$$\begin{aligned}
&\|\sum_{j=1}^{\mathrm{K}_2}(\widehat{h}_j - h_j^s)\otimes \mathcal{A}h_j^s\|_{\mathrm{HS}}^2 = \sum_{\ell=1}^{\infty}\|\sum_{j=1}^{\mathrm{K}_2}\langle \mathcal{A}h_j, h_\ell\rangle(\widehat{h}_j - h_j^s)\|^2 \le \sum_{\ell=1}^{\infty}\left(\sum_{j=1}^{\mathrm{K}_2}|\langle \mathcal{A}h_j, h_\ell\rangle|\|\widehat{h}_j - h_j^s\|\right)^2 \\
&\le O_p(\mathsf{a}_1)\left(\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu/2-2\gamma_\mu-\varsigma_\nu+1}\right)^2 \\
&= \begin{cases} O_p(\mathsf{a}_1) & \text{if } \varsigma_\nu > 2+\rho_\nu/2-2\gamma_\mu, \\ O_p(\mathsf{a}_1\max\{\log^2 \mathsf{a}_2^{-1}, \mathsf{a}_2^{(2\varsigma_\nu-\rho_\nu+4\gamma_\mu-4)/\rho_\nu}\}) & \text{if } \varsigma_\nu \le 2+\rho_\nu/2-2\gamma_\mu. \end{cases}
\end{aligned} \tag{S3.24}$$

Moreover, from (S3.20) and the fact that $d_T\mathsf{a}_1^{-1} = o(1)$, the following may be deduced:

$$\|\sum_{j=1}^{\mathrm{K}_2}\widehat{h}_j \otimes \mathcal{A}(\widehat{h}_j - h_j^s)\|_{\mathrm{HS}}^2 \le \sum_{j=1}^{\mathrm{K}_2}\|\mathcal{A}(\widehat{h}_j - h_j^s)\|^2 \le O_p(\mathsf{a}_1)\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu-2\varsigma_\nu-4\gamma_\mu+2} + O_p(d_T)\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu-2\varsigma_\nu+2}$$

$$\leq \begin{cases} O_p(\mathsf{a}_1) & \text{if } \rho_\nu/2+3/2<\varsigma_\nu, \\ O_p(\mathsf{a}_1 \max\{\log \mathsf{a}_2^{-1}, \mathsf{a}_2^{(2\varsigma_\nu-\rho_\nu-3)/\rho_\nu}\}) & \text{if } \rho_\nu/2+3/2\geq\varsigma_\nu. \end{cases} \tag{S3.25}$$

Since $4\gamma_\mu - 4 > -3$, $\mathsf{a}_2 \log \mathsf{a}_2^{-1} = o(1)$, and $\mathsf{a}_2 \log^2 \mathsf{a}_2^{-1} = o(1)$, (4.2) may be deduced from (S3.22), (S3.24) and (S3.25).

Proofs of (S3.17)-(S3.20): To obtain the desired results, we first need to consider $\|(\widehat{\mathcal{Q}}-\mathcal{Q})h_\ell\|$ and $\|\widehat{\mathcal{Q}}-\mathcal{Q}\|_{\mathrm{HS}}$. Note that, for any ℓ,

$$2^{-1}\|(\widehat{\mathcal{Q}}-\mathcal{Q})h_\ell\|^2 \leq \|(\widehat{\mathcal{Q}}-\mathcal{Q}_{\mathrm{K}_1})h_\ell\|^2 + \|(\mathcal{Q}_{\mathrm{K}_1}-\mathcal{Q})h_\ell\|^2. \tag{S3.26}$$

The second term in (S3.26) is bounded above as follows:

$$\begin{aligned}\|(\mathcal{Q}_{\mathrm{K}_1}-\mathcal{Q})h_\ell\|^2 &= \|\sum_{j=\mathrm{K}_1+1}^{\infty} \mu_j \langle \mathcal{B}g_j, h_\ell\rangle \mathcal{B}g_j\|^2 \leq O(1)\sum_{j=\mathrm{K}_1+1}^{\infty} \mu_j^2 \|\mathcal{B}g_j\|^2 \sum_{j=\mathrm{K}_1+1}^{\infty} j^{-2\varsigma_\mu}\ell^{-2\gamma_\mu} \\ &\leq O(1)\mu_{\mathrm{K}_1+1}^2(\mathrm{K}_1+1)^{-4\varsigma_\mu+2}\ell^{-2\gamma_\mu} \leq O_p(1)\mathsf{a}_1^{(4\varsigma_\mu+\rho_\mu-2)/\rho_\mu}\ell^{-2\gamma_\mu},\end{aligned} \tag{S3.27}$$

where the first inequality follows from the Hölder's inequality and the second is obtained because $2\gamma_\mu > 1$, $\sum_{j=\mathrm{K}_1+1}^{\infty} j^{-2\varsigma_\mu} \leq (\mathrm{K}_1+1)^{-2\varsigma_\mu+1}$, and $\mu_j^2 \leq \mu_{\mathrm{K}_1+1}^2$ for $j > \mathrm{K}_1$. The last inequality is obtained using the arguments that are used to derive (S2.12). We now focus on the first term in (S3.26). Note that

$$\begin{aligned}4^{-1}\|(\widehat{\mathcal{Q}}-\mathcal{Q}_{\mathrm{K}_1})h_\ell\|^2 \leq \|\widehat{\mathcal{C}}_{vz}^*(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{vz}h_\ell\|^2 + \|\widehat{\mathcal{C}}_{vz}^*\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^*h_\ell\|^2 + \|\mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{vz}h_\ell\|^2 \\ + \|(\mathcal{B}(\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{zz}\widehat{\Pi}_{\mathrm{K}_1}-\Pi_{\mathrm{K}_1}\mathcal{C}_{zz}\Pi_{\mathrm{K}_1})\mathcal{B}^*)h_\ell\|^2,\end{aligned}$$

where $\|\widehat{\mathcal{C}}_{zv}\widehat{\Pi}_{\mathrm{K}_1}\mathcal{B}^*h_\ell\|^2 \leq \ell^{-2\gamma_\mu}O_p(T^{-1})$. Moreover, we have $\|\mathcal{B}\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{vz}h_\ell\|^2 \leq \ell^{-2\gamma_\mu}O_p(T^{-1})$ and $\|\widehat{\mathcal{C}}_{vz}^*(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{vz}h_\ell\|^2 \leq O_p(\mathsf{a}_1^{-1}T^{-1})\|\widehat{\mathcal{C}}_{vz}h_\ell\|^2 \leq O_p(T^{-2}\mathsf{a}_1^{-1})\ell^{-\rho_\nu/2}$ since

$$T\mathbb{E}[\|\widehat{\mathcal{C}}_{vz}h_j\|^2] = T^{-1}\mathbb{E}[\|\sum_{t=1}^{T}\langle v_t, h_j\rangle z_t\|^2] \leq O(1)\mathbb{E}[\|\langle v_t, h_j\rangle z_t\|^2] \leq O(1)\nu_j \leq O(j^{-\rho_\nu/2}), \tag{S3.28}$$

where the inequalities are obtained from Assumption E2*; specifically, under the assumption, we have that $\mathbb{E}[\|\langle v_t, h_j\rangle z_t\|^2] \leq \mathbb{E}[\|\langle x_t, h_j\rangle z_t\|^2] \leq c_\circ\|\mathcal{C}_{xz}h_j\|^2 \leq \|\Pi_{\mathrm{K}_1}\mathcal{C}_{zz}\|_{\mathrm{op}}\|(\mathcal{C}_{zz})_{\mathrm{K}_1}^{-1/2}\mathcal{C}_{xz}h_j\|^2 \leq O(1)\nu_j$. Lastly, using the arguments used to obtain (S2.16), we can show that $\|\widehat{g}_j - g_j^s\|^2 \leq O_p(T^{-1})j^2$ and $\|\mathcal{B}(\widehat{g}_j - g_j^s)\|^2 \leq O_p(T^{-1})(j^{2-2\varsigma_\mu} + j^{\rho_\mu+2-2\varsigma_\mu})$. Using this bound, we find that

$$\begin{aligned}&4^{-1}\|(\mathcal{B}(\widehat{\Pi}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{zz}\widehat{\Pi}_{\mathrm{K}_1}-\Pi_{\mathrm{K}_1}\mathcal{C}_{zz}\Pi_{\mathrm{K}_1})\mathcal{B}^*)h_\ell\|^2 \\ &\leq \|\sum_{j=1}^{\mathrm{K}_1}(\widehat{\mu}_j-\mu_j)\langle\widehat{g}_j, \mathcal{B}^*h_\ell\rangle\mathcal{B}\widehat{g}_j\|^2 + \|\sum_{j=1}^{\mathrm{K}_1}\mu_j\langle\widehat{g}_j-g_j^s, \mathcal{B}^*h_\ell\rangle\mathcal{B}g_j\|^2 \\ &\quad + \|\sum_{j=1}^{\mathrm{K}_1}\mu_j\langle g_j^s, \mathcal{B}^*h_\ell\rangle\mathcal{B}(\widehat{g}_j-g_j^s)\|^2 + \|\sum_{j=1}^{\mathrm{K}_1}\mu_j\langle\widehat{g}_j-g_j^s, \mathcal{B}^*h_\ell\rangle\mathcal{B}(\widehat{g}_j-g_j^s)\|^2\end{aligned}$$

$$
\begin{aligned}
&\leq \mathrm{K}_1 \|\mathcal{B}\|_{\mathrm{op}}^2 \|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\mathrm{op}}^2 \|\mathcal{B}^* h_\ell\|^2 + \|\mathcal{C}_{zz}\|_{\mathrm{HS}}^2 \sum_{j=1}^{\mathrm{K}_1} \|\widehat{g}_j - g_j^s\|^2 \|\mathcal{B}^* h_\ell\|^2 \|\mathcal{B} g_j\|^2 \\
&\quad + \|\mathcal{C}_{zz}\|_{\mathrm{HS}}^2 \sum_{j=1}^{\mathrm{K}} \|\mathcal{B}^* h_\ell\|^2 \|\mathcal{B}(\widehat{g}_j - g_j^s)\|^2 + \|\mathcal{C}_{zz}\|_{\mathrm{HS}}^2 \|\mathcal{B}^* h_\ell\|^2 \sum_{j=1}^{\mathrm{K}_1} \|\widehat{g}_j - g_j^s\|^4 \\
&\leq \ell^{-2\gamma_\mu} \mathsf{a}^{-1/\rho_\mu} O_p(T^{-1}) + \ell^{-2\gamma_\mu} O_p(T^{-1}) \sum_{j=1}^{\mathrm{K}_1} j^{-2\varsigma_\mu + 2} \\
&\quad + \ell^{-2\gamma_\mu} O_p(T^{-1}) \sum_{j=1}^{\mathrm{K}_1} (j^{-2\varsigma_\mu+2} + j^{\rho_\mu - 2\varsigma_\mu + 2}) + \ell^{-2\gamma_\mu} O_p(T^{-2}) \sum_{j=1}^{\mathrm{K}_1} j^4 \\
&\leq \ell^{-2\gamma_\mu} O_p(T^{-1} \max\{\mathsf{a}_1^{-1/\rho_\mu}, \mathsf{a}_1^{-(\rho_\mu - 2\varsigma_\mu + 3)/\rho_\mu}\}),
\end{aligned} \tag{S3.29}
$$

where the second inequality is obtained by using Lemma 4.2 in Bosq (2000) and noting that $\| \sum_{j=1}^{\mathrm{K}_1} \mu_j \langle \widehat{g}_j - g_j, \mathcal{B}^* h_\ell \rangle \mathcal{B} g_j\|^2 \leq (\sum_{j=1}^{\mathrm{K}_1} \mu_j^2)(\sum_{j=1}^{\mathrm{K}_1} \|\mathcal{B} g_j\|^2 \langle \widehat{g}_j - g_j, \mathcal{B}^* h_\ell\rangle^2)$ holds by the Hölder's inequality. The last two inequalities are deduced from Assumption M, Assumption E2*, and that $\|\mathcal{B}^* h_\ell\|^2 \leq O(1)\ell^{-2\gamma_\mu}$ and $\mathsf{a}_1^{-1} = \alpha_1 = o(T^{\rho_\mu/(2\rho_\mu+2)})$. Then, from the results given in (S3.26)-(S3.29) and the definition of d_T given in (S3.21), we conclude the following: for any $\ell \leq \mathrm{K}_2$,

$$
\|(\widehat{\mathcal{Q}} - \mathcal{Q}) h_\ell\|^2 \leq \ell^{-2\gamma_\mu} O_p(d_T), \tag{S3.30}
$$

from which we also find that

$$
\|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\mathrm{HS}}^2 = O_p(d_T). \tag{S3.31}
$$

We now verify (S3.17) and (S3.18). It can be shown without difficulty that $d_T = O(\mathsf{a}_1)$. Given that $\mathsf{a}_1^{-1} = o(T^{1/2})$ and $\mathsf{a}_2^{-1}\mathsf{a}_1^{1/2} = o(1)$, we find that $\mathsf{a}_2^{-1} d_T^{1/2} = o(1)$, from which the following is deduced:

$$
\mathsf{a}_2 = \widehat{\nu}_{\mathrm{K}_2}^2 - \nu_{\mathrm{K}_2}^2 + \nu_{\mathrm{K}_2}^2 \leq \|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\mathrm{op}} + c_\circ \mathrm{K}_2^{-\rho} \leq o(1)\mathsf{a}_2 + c_\circ \mathrm{K}_2^{-\rho_\nu}.
$$

Then (S3.17) follows from the above. Similarly as in (S2.12), it can be shown that

$$
(c_\circ \rho)^{-1} (\mathrm{K}_2 + 1)^{-\rho_\nu} \leq \nu_{\mathrm{K}_2+1}^2 = \nu_{\mathrm{K}_2+1}^2 - \widehat{\nu}_{\mathrm{K}_2+1}^2 + \widehat{\nu}_{\mathrm{K}_2+1} \leq (1 + o_p(1))\mathsf{a}_2,
$$

and thus we find that (S3.18) holds.

We then show (S3.19). To this end, it should first be noted that, under the employed assumptions, (S2.27) holds if $\widehat{\lambda}_j$ (resp. λ_ℓ) is replaced by $\widehat{\nu}_j$ (resp. ν_ℓ) . Moreover, note that the eigenfunctions of $\mathcal{Q}^2$ and $\widehat{\mathcal{Q}}^2$ are, respectively, equivalent to those of $\mathcal{Q}$ and $\widehat{\mathcal{Q}}$. Therefore, by applying the arguments that are used in our proof of Theorem 3, we can show that

$$
8^{-1} \|\widehat{h}_j - h_j^s\|^2 \leq \sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \langle (\widehat{\mathcal{Q}} - \mathcal{Q}) \widehat{h}_j, h_\ell \rangle^2 + \sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-2} \widehat{\nu}_j^2 \langle (\widehat{\mathcal{Q}} - \mathcal{Q}) \widehat{h}_j, h_\ell \rangle^2. \tag{S3.32}
$$

From similar arguments used to derive (S2.22), the first term in (S3.32) is bounded above as follows:

$$\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q})\widehat{h}_j,h_\ell\rangle^2 \leq 2\widetilde{\Delta}_{1j}\|\widehat{h}_j-h_j^s\|^2+2\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\langle h_j,(\widehat{\mathcal{Q}}-\mathcal{Q})h_\ell\rangle^2, \quad \text{(S3.33)}$$

where $\widetilde{\Delta}_{1j}=\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\|(\widehat{\mathcal{Q}}-\mathcal{Q})h_\ell\|^2$. Moreover, by using similar arguments that are used to obtain (S2.23), we can show that

$$\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\widehat{\nu}_j^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q})\widehat{h}_j,h_\ell\rangle^2 \leq 2\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\widehat{\nu}_j^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q})h_j,h_\ell\rangle^2+2(\widetilde{\Delta}_{1j}+\widetilde{\Delta}_{2j})\|\widehat{h}_j-h_j^s\|^2, \quad \text{(S3.34)}$$

where $\widetilde{\Delta}_{2j}=\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-1}\|(\widehat{\mathcal{Q}}-\mathcal{Q})h_\ell\|^2$. We then use the results given in (S3.30) and (S3.31) to obtain the following bounds of $\widetilde{\Delta}_{1j}$ and $\widetilde{\Delta}_{2j}$: for $j=1,\ldots,\mathrm{K}_2$,

$$\widetilde{\Delta}_{1j}\leq O_p(d_T)\sum_{\ell\neq j}(\nu_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\ell^{-2\gamma_\mu}\leq j^{\rho_\nu+2-2\gamma_\mu}O_p(d_T)\leq O_p(\mathsf{a}_2^{-(2\gamma_\mu-2-\rho_\nu)/\rho_\nu}d_T)=o_p(1), \quad \text{(S3.35)}$$

where the second inequality follows from Lemma S1(i) and the last equality follows from that $\gamma_\mu\leq 1+\frac{3}{2}\rho_\nu$ and $\mathsf{a}_2^{-1}d_T^{1/2}=o_p(1)$. Similarly, for $j=1,\ldots,\mathrm{K}_2$,

$$\widetilde{\Delta}_{2j}\leq\max_{1\leq\ell\leq\mathrm{K}_2}(\nu_j^2-\nu_\ell^2)^{-1}\sum_{\ell\neq j}\|(\widehat{\mathcal{Q}}-\mathcal{Q})h_\ell\|^2\leq j^{1+\rho_\nu}O_p(d_T)\leq O(\mathsf{a}_2^{-(1+\rho_\nu)/\rho_\nu}d_T)=o_p(1). \quad \text{(S3.36)}$$

From (S3.32)-(S3.36), we have

$$\|\widehat{h}_j-h_j^s\|^2\leq O(1)(1+o_p(1))\widetilde{\Delta}_{3j} \quad \text{(S3.37)}$$

for $j=1,\ldots,\mathrm{K}_2$, where

$$\widetilde{\Delta}_{3j}=\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\widehat{\nu}_j^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q})h_j,h_\ell\rangle^2+\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q})h_j,h_\ell\rangle^2. \quad \text{(S3.38)}$$

We will analyze the above term using the decomposition $\widehat{\mathcal{Q}}-\mathcal{Q}=\widehat{\mathcal{Q}}-\mathcal{Q}_{\mathrm{K}_1}+\mathcal{Q}_{\mathrm{K}_1}-\mathcal{Q}$. Note that

$$\langle(\mathcal{Q}_{\mathrm{K}_1}-\mathcal{Q})h_j,h_\ell\rangle^2\leq\left(\sum_{i=\mathrm{K}_1+1}^\infty\mu_i^2\langle g_i,\mathcal{B}^*h_j\rangle^2\right)\left(\sum_{i=\mathrm{K}_1+1}^\infty\langle g_i,\mathcal{B}^*h_\ell\rangle^2\right)\leq O_p(\mathsf{a}_1)\ell^{-2\gamma_\mu}j^{-2\gamma_\mu}, \quad \text{(S3.39)}$$

where the first inequality follows from the Hölder's inequality, and the second is deduced from Assumption E2* and similar arguments used to derive (S2.12). From (S3.39), Lemma S1(i) and a result similar to (S2.27), we find that

$$\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\langle(\mathcal{Q}_{\mathrm{K}_1}-\mathcal{Q})h_j,h_\ell\rangle^2\leq O_p(\mathsf{a}_1)j^{\rho_\nu-4\gamma_\mu+2}. \quad \text{(S3.40)}$$

Similarly, for $j = 1, \ldots, \mathrm{K}_2$, we have

$$\begin{aligned}
&\sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-2} \widehat{\nu}_j^2 \langle (\mathcal{Q}_{\mathrm{K}_1} - \mathcal{Q}) h_j, h_\ell \rangle^2 \\
&\leq |\widehat{\nu}_j^2 - \nu_j^2| \sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_j^2 \nu_j^{-2} \langle (\mathcal{Q}_{\mathrm{K}_1} - \mathcal{Q}) h_j, h_\ell \rangle^2 + \sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_j^2 \langle (\mathcal{Q}_{\mathrm{K}_1} - \mathcal{Q}) h_j, h_\ell \rangle^2 \\
&\leq (O_p(d_T^{1/2}) j^{\rho_\nu} + 1) O_p(\mathsf{a}_1) j^{\rho_\nu - 4\gamma_\mu + 2}, \qquad \text{(S3.41)}
\end{aligned}$$

where the second inequality follows from that $|\widehat{\nu}_j^2 - \nu_j^2| \leq \|\widehat{\mathcal{Q}}^2 - \mathcal{Q}^2\|_{\mathrm{op}} \leq O_p(1) \|\widehat{\mathcal{Q}} - \mathcal{Q}\|_{\mathrm{op}}$. Given that $\mathsf{a}_2^{-1} d_T^{1/2} = o(1)$, (S3.41) is bounded above by $O_p(\mathsf{a}_1) j^{\rho_\nu - 4\gamma_\mu + 2}$. Next, we will obtain an upper bound of $\langle (\widehat{\mathcal{Q}} - \mathcal{Q}_{\mathrm{K}_1}) h_j, h_\ell \rangle^2$. Observe that

$$\begin{aligned}
4^{-1} \langle (\widehat{\mathcal{Q}} - \mathcal{Q}_{\mathrm{K}_1}) h_\ell, h_j \rangle^2 &\leq \langle \widehat{\mathcal{C}}_{vz}^* (\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1} \widehat{\mathcal{C}}_{vz} h_\ell, h_j \rangle^2 + \langle \widehat{\mathcal{C}}_{vz}^* \widehat{\Pi}_{\mathrm{K}_1} \mathcal{B}^* h_\ell, h_j \rangle^2 + \langle \mathcal{B} \widehat{\Pi}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{vz} h_\ell, h_j \rangle^2 \\
&\quad + \langle (\mathcal{B}(\widehat{\Pi}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{zz} \widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1} \mathcal{C}_{zz} \Pi_{\mathrm{K}_1}) \mathcal{B}^*) h_\ell, h_j \rangle^2 \\
&\leq \mathsf{a}_1^{-1} \nu_j \nu_\ell O_p(T^{-2}) + \nu_j \ell^{-2\gamma_\mu} O_p(T^{-1}) + \nu_\ell j^{-2\gamma_\mu} O_p(T^{-1}) \\
&\quad + \langle (\mathcal{B}(\widehat{\Pi}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{zz} \widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1} \mathcal{C}_{zz} \Pi_{\mathrm{K}_1}) \mathcal{B}^*) h_\ell, h_j \rangle^2, \qquad \text{(S3.42)}
\end{aligned}$$

where the last inequality is obtained from (S3.28). The last term in (S3.42) satisfies the following:

$$\begin{aligned}
&4^{-1} \langle (\mathcal{B}(\widehat{\Pi}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{zz} \widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1} \mathcal{C}_{zz} \Pi_{\mathrm{K}_1}) \mathcal{B}^*) h_\ell, h_j \rangle^2 \\
&\leq \Big(\sum_{i=1}^{\mathrm{K}_1} (\widehat{\mu}_i - \mu_i) \langle \widehat{g}_i, \mathcal{B}^* h_\ell \rangle \langle \mathcal{B} \widehat{g}_i, h_j \rangle\Big)^2 + \Big(\sum_{i=1}^{\mathrm{K}_1} \mu_i \langle \widehat{g}_i - g_i, \mathcal{B}^* h_\ell \rangle \langle \mathcal{B} \widehat{g}_i, h_j \rangle\Big)^2 \\
&\quad + \Big(\sum_{i=1}^{\mathrm{K}_1} \mu_i \langle g_i, \mathcal{B}^* h_\ell \rangle \langle \mathcal{B}(\widehat{g}_i - g_i), h_j \rangle\Big)^2 \\
&\leq \max_{1 \leq i \leq \mathrm{K}_1} |\widehat{\mu}_i - \mu_i|^2 \|\widehat{\Pi}_{\mathrm{K}_1} \mathcal{B}^* h_\ell\|^2 \|\widehat{\Pi}_{\mathrm{K}_1} \mathcal{B}^* h_j\|^2 + \sum_{i=1}^{\mathrm{K}_1} \mu_i^2 \|\widehat{g}_i - g_i\|^2 \|\mathcal{B}^* h_\ell\|^2 \|\widehat{\Pi}_{\mathrm{K}_1} \mathcal{B}^* h_j\|^2 \\
&\quad + \sum_{i=1}^{\mathrm{K}_1} \mu_i^2 \|\widehat{g}_i - g_i\|^2 \|\mathcal{B}^* h_\ell\|^2 \|\Pi_{\mathrm{K}_1} \mathcal{B}^* h_j\|^2 \\
&\leq \ell^{-2\gamma_\mu} j^{-2\gamma_\mu} O_p(T^{-1} \mathsf{a}_1^{-1/\rho_\mu}), \qquad \text{(S3.43)}
\end{aligned}$$

where the last inequality follows from Assumptions M* and E2*. Combining the results given in (S3.42) and (S3.43), we find that

$$\begin{aligned}
&4^{-1} \langle (\widehat{\mathcal{Q}} - \mathcal{Q}_{\mathrm{K}_1}) h_\ell, h_j \rangle^2 \\
&\leq O_p(T^{-1}) (\mathsf{a}_1^{-1} T^{-1} j^{-\rho_\nu/2} \ell^{-\rho_\nu/2} + j^{-\rho_\nu/2} \ell^{-2\gamma_\mu} + j^{-2\gamma_\mu} \ell^{-\rho_\nu/2} + \mathsf{a}_1^{-1/\rho_\mu} \ell^{-2\gamma_\mu} j^{-2\gamma_\mu}) \\
&\leq O_p(T^{-1}) \mathsf{a}_1^{-1/\rho_\mu} \ell^{-2\gamma_\mu} j^{-2\gamma_\mu} + O_p(T^{-1}) (j^{-\rho_\nu/2} \ell^{-2\gamma_\mu} + j^{-2\gamma_\mu} \ell^{-\rho_\nu/2}).
\end{aligned}$$

Together with Lemma S1(i), this implies that

$$\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q}_{\mathrm{K}_1})h_\ell,h_j\rangle^2 \leq O_p(T^{-1}\mathtt{a}_1^{-1/\rho_\mu})j^{\rho_\nu-4\gamma_\mu+2}+O_p(T^{-1})j^{\rho_\nu/2-2\gamma_\mu+2}$$
$$\leq O_p(T^{-1}\mathtt{a}_1^{-1/\rho_\mu})j^{\rho_\nu-4\gamma_\mu+2}, \tag{S3.44}$$

where the inequalities follow from Lemma S1(i) and the fact that the first term is dominant under the condition $\gamma_\mu \leq \rho_\nu/4+1/2$. In addition, from the arguments that are used to derive (S3.41) and the fact that $\mathtt{a}_2^{-1}d_T^{1/2}=o(1)$, the following is deduced: for $j=1,\ldots,\mathrm{K}_2$,

$$\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\widehat{\nu}_j^2\langle(\widehat{\mathcal{Q}}-\mathcal{Q}_{\mathrm{K}_1})h_\ell,h_j\rangle^2 \leq (O_p(\mathtt{a}_2^{-1}d_T^{1/2})+1)O_p(T^{-1}\mathtt{a}_1^{-1/\rho_\mu})j^{\rho_\nu-4\gamma_\mu+2}$$
$$= O_p(T^{-1}\mathtt{a}_1^{-1/\rho_\mu})j^{\rho_\nu-4\gamma_\mu+2}. \tag{S3.45}$$

From (S3.37), (S3.38), (S3.40), (S3.41), (S3.44), (S3.45), and the decomposition $\widehat{\mathcal{Q}}-\mathcal{Q}=\widehat{\mathcal{Q}}-\mathcal{Q}_{\mathrm{K}_1}+\mathcal{Q}_{\mathrm{K}_1}-\mathcal{Q}$, we conclude that

$$\|\widehat{h}_j-h_j^s\|^2\leq(1+o_p(1))\widetilde{\Delta}_{3j}\leq O_p(\mathtt{a}_1)j^{\rho_\nu-4\gamma_\mu+2}+O_p(T^{-1}\mathtt{a}_1^{-1/\rho_\mu})j^{\rho_\nu-4\gamma_\mu+2}\leq O_p(\mathtt{a}_1)j^{\rho_\nu-4\gamma_\mu+2},$$

where the last inequality follows from $\mathtt{a}_1^{-1}=o(T^{\rho_\mu/(2\rho_\mu+2)})$. This completes our proof of (S3.19).

Lastly, to show (S3.20), we note that

$$\mathcal{A}(\widehat{h}_j-h_j^s)=\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-1}\langle(\widehat{\mathcal{Q}}^2-\mathcal{Q}^2)\widehat{h}_j,h_\ell\rangle\mathcal{A}h_\ell+\langle\widehat{h}_j-h_j^s,h_j\rangle\mathcal{A}h_j, \tag{S3.46}$$

where $\|\langle\widehat{h}_j-h_j^s,h_j\rangle\mathcal{A}h_j\|^2\leq O_p(\mathtt{a}_1)j^{\rho_\nu-2\varsigma_\nu-4\gamma_\mu+2}$. The first term in (S3.46) is bounded above as follows:

$$\begin{aligned}
&(\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-1}\langle(\widehat{\mathcal{Q}}^2-\mathcal{Q}^2)\widehat{h}_j,h_\ell\rangle\mathcal{A}h_\ell)^2\leq(\sum_{\ell\neq j}|\widehat{\nu}_j^2-\nu_\ell^2|^{-1}|\langle(\widehat{\mathcal{Q}}^2-\mathcal{Q}^2)\widehat{h}_j,h_\ell\rangle|\|\mathcal{A}h_\ell\|)^2\\
&\leq O(1)\|\widehat{\mathcal{Q}}-\mathcal{Q}\|_{\mathrm{op}}^2\left(\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\ell^{-2\varsigma_\nu}+\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\widehat{\nu}_j^2\ell^{-2\varsigma_\nu}\right)\\
&\leq O_p(d_T)\left(\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_\ell^2\ell^{-2\varsigma_\nu}+\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}(\widehat{\nu}_j^2-\nu_j^2)\ell^{-2\varsigma_\nu}+\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_j^2\ell^{-2\varsigma_\nu}\right)\\
&\leq O_p(d_T)\left(j^{\rho_\nu-2\varsigma_\nu+2}+(O_p(d_T^{1/2})\nu_j^{-2}+1)\sum_{\ell\neq j}(\widehat{\nu}_j^2-\nu_\ell^2)^{-2}\nu_j^2\ell^{-2\varsigma_\nu}\right)\\
&\leq(1+o_p(1))O_p(d_T)j^{\rho_\nu-2\varsigma_\nu+2}.
\end{aligned}\tag{S3.47}$$

Combining (S3.46) and (S3.47), we obtain (S3.20) as desired. □

Proof of Theorem 8

In this proof, we allow the case where $\rho_\nu/2+2 \geq \varsigma_\nu + \delta_\zeta$, thereby encompassing the complementary result provided in Section S3.3.2 as a special case. The whole proof is divided into two parts.

1. Proof of the convergence results: We need an upper bound of $\langle \widehat{h}_j - h_j^s, \zeta\rangle$, which is importantly used in the following discussion. Using the expansion in (S2.19), we find that

$$\langle \widehat{h}_j - h_j^s, \zeta\rangle = \sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-1} \langle (\widehat{\mathcal{Q}} - \mathcal{Q})\widehat{h}_j, h_\ell\rangle \langle h_\ell, \zeta\rangle + \langle \widehat{h}_j - h_j^s, h_j^s\rangle\langle h_j^s, \zeta\rangle.$$

Note that $(\langle \widehat{h}_j - h_j^s, h_j^s\rangle\langle h_j^s, \zeta\rangle)^2 \leq O_p(\mathsf{a}_1) j^{\rho_\nu - 4\gamma_\mu - 2\delta_\zeta + 2}$ for $j = 1, \ldots, \mathrm{K}_2$, because of (S3.19). Moreover, using similar arguments that are used to derive (S2.32) and (S3.47), we find that

$$(\sum_{\ell \neq j} (\widehat{\nu}_j^2 - \nu_\ell^2)^{-1} \langle (\widehat{\mathcal{Q}} - \mathcal{Q})\widehat{h}_j, h_\ell\rangle \langle h_\ell, \zeta\rangle)^2 \leq (1 + o_p(1)) O_p(d_T) j^{\rho_\nu - 2\delta_\zeta + 2}. \tag{S3.48}$$

Hence, we conclude that, for $j = 1, \ldots, \mathrm{K}_2$,

$$\langle \widehat{h}_j - h_j^s, \zeta\rangle^2 \leq O_p(\mathsf{a}_1) j^{\rho_\nu - 4\gamma_\mu - 2\delta_\zeta + 2} + O_p(d_T) j^{\rho_\nu + 2 - 2\delta_\zeta} \leq O_p(\mathsf{a}_1) j^{\rho_\nu - 2\delta_\zeta + 2}, \tag{S3.49}$$

where the last inequality follows from that $d_T \mathsf{a}_1^{-1} = o(1)$ and $j^{-\gamma_\mu} \leq 1$.

Using the result given in (S3.49), we will show that

$$\|(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} \widehat{\mathcal{Q}}^{-1}_{\mathrm{K}_2}\zeta - (\mathcal{C}_{zz})^{-1}_{\mathrm{K}_1} \mathcal{C}_{xz} \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta\| = o_p(1), \tag{S3.50}$$

$$\|\widehat{\mathcal{Q}}^{-1}_{\mathrm{K}_2}\zeta - \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta\| = o_p(1). \tag{S3.51}$$

To show (S3.50), we observe that

$$\|(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} \widehat{\mathcal{Q}}^{-1}_{\mathrm{K}_2}\zeta - (\mathcal{C}_{zz})^{-1}_{\mathrm{K}_1} \mathcal{C}_{xz} \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta\| \leq \|(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} (\widehat{\mathcal{Q}}^{-1}_{\mathrm{K}_2} - \mathcal{Q}^{-1}_{\mathrm{K}_2})\zeta\| + \|((\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} - (\mathcal{C}_{zz})^{-1}_{\mathrm{K}_1} \mathcal{C}_{xz}) \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta\|. \tag{S3.52}$$

Because $(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} = \widehat{\Pi}_{\mathrm{K}_1} \mathcal{B}^* + (\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{vz}$ and $\|(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{vz}\|_{\mathrm{op}} = O_p(\mathsf{a}_1^{-1/2} T^{-1/2})$, the first term in (S3.52) satisfies that

$$\|(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} (\widehat{\mathcal{Q}}^{-1}_{\mathrm{K}_2} - \mathcal{Q}^{-1}_{\mathrm{K}_2})\zeta\| \leq O_p(1) \|(\widehat{\mathcal{Q}}^{-1}_{\mathrm{K}_2} - \mathcal{Q}^{-1}_{\mathrm{K}_2})\zeta\|. \tag{S3.53}$$

Moreover, under the employed assumptions, the following holds:

$$\|\mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta\|^2 = \|\sum_{j=1}^{\mathrm{K}_2} \nu_j^{-1} \langle h_j, \zeta\rangle h_j\|^2 \leq O(1) \sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu - 2\delta_\zeta} \leq O(\max\{\mathsf{a}_2^{-1/\rho_\nu}, \mathsf{a}_2^{-(\rho_\nu - 2\delta_\zeta + 1)/\rho_\nu}\}). \tag{S3.54}$$

Given that $\mathsf{a}_1^{1/2}\mathsf{a}_2^{-1} = o(1)$, we have $\mathsf{a}_1 \mathsf{a}_2^{-1/\rho_\nu} \leq \mathsf{a}_1 \mathsf{a}_2^{-1} \mathsf{a}_2^{(\rho_\nu - 1)/\rho_\nu} = o(1)$ and $\mathsf{a}_1 \mathsf{a}_2^{-(\rho_\nu - 2\delta_\zeta + 1)/\rho_\nu} = \mathsf{a}_1 \mathsf{a}_2^{-1} \mathsf{a}_2^{(2\delta_\zeta - 1)/\rho_\nu} = o(1)$. This implies that $\|\mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta\|^2 \leq o_p(\mathsf{a}_1^{-1})$. Note also that

$$((\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{xz} - (\mathcal{C}_{zz})^{-1}_{\mathrm{K}_1} \mathcal{C}_{xz}) \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta = (\widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1}) \mathcal{B}^* \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta + (\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1} \widehat{\mathcal{C}}_{vz} \mathcal{Q}^{-1}_{\mathrm{K}_2}\zeta, \tag{S3.55}$$

where the second term on the right hand side is $o_p(1)$ since $\|(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1}\widehat{\mathcal{C}}_{vz}\|_{\mathrm{op}} = O_p(\mathsf{a}_1^{-1/2}T^{-1/2})$, $\|\mathcal{Q}_{\mathrm{K}_2}^{-1}\zeta\| \leq o_p(\mathsf{a}_1^{-1/2})$, and $\mathsf{a}_1^{-1}T^{-1/2} = o(1)$. Furthermore, since $\|\widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1}\|_{\mathrm{op}} \leq O(1)\sum_{j=1}^{\mathrm{K}_1}\|\widehat{g}_j - g_j^s\| \leq O_p(T^{-1/2}\,\mathrm{K}_1^2) = O_p(T^{-1/2}\mathsf{a}_1^{-2/\rho_\mu})$ and $\mathsf{a}_1^{-1} = o(T^{\rho_\mu/(2\rho_\mu+2)})$, we find that

$$\begin{aligned}\|(\widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1})\mathcal{B}^*\mathcal{Q}_{\mathrm{K}_2}^{-1}\zeta\| &\leq \|\widehat{\Pi}_{\mathrm{K}_1} - \Pi_{\mathrm{K}_1}\|_{\mathrm{op}}\|\mathcal{B}\|_{\mathrm{op}}\|\mathcal{Q}_{\mathrm{K}_2}^{-1}\zeta\| = O_p(T^{-1/2}\mathsf{a}_1^{-2/\rho_\mu-1/2}) \\ &= O_p(T^{-1/2}\mathsf{a}_1^{-(\rho_\mu+1)/\rho_\mu})\mathsf{a}_1^{(\rho_\mu/2-1)/\rho_\mu} = o_p(1). \end{aligned} \tag{S3.56}$$

From (S3.53)-(S3.56), it is deduced that (S3.51) implies (S3.50) and thus we only need to show (S3.51) for the desired results. From a similar decomposition to that given in (S2.4), it can be shown that (S3.51) holds if the following terms are all $o_p(1)$: $\|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1} - \widehat{\nu}_j^{-1})\langle h_j, \zeta\rangle\widehat{h}_j\|$, $\|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1} - \widehat{\nu}_j^{-1})\langle \widehat{h}_j - h_j^s, \zeta\rangle\widehat{h}_j\|$, $\|\sum_{j=1}^{\mathrm{K}}\nu_j^{-1}\langle h_j, \zeta\rangle(\widehat{h}_j - h_j^s)\|$, and $\|\sum_{j=1}^{\mathrm{K}}\nu_j^{-1}\langle \widehat{h}_j - h_j^s, \zeta\rangle\widehat{h}_j\|$.

As in (S2.35) and (S2.36), we obtain the following:

$$\begin{aligned}&\|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1} - \widehat{\nu}_j^{-1})\langle h_j, \zeta\rangle\widehat{h}_j\|^2 = \sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1} - \widehat{\nu}_j^{-1})^2\langle h_j, \zeta\rangle^2 \leq \sum_{j=1}^{\mathrm{K}_2}\frac{(\nu_j^2 - \widehat{\nu}_j^2)^2}{\nu_j^2(\widehat{\nu}_j^2 + \nu_j\widehat{\nu}_j)^2}c_\zeta j^{-2\delta_\zeta} \\ &\leq \sum_{j=1}^{\mathrm{K}_2}\frac{(\nu_j^2 - \widehat{\nu}_j^2)^2}{\nu_j^4\widehat{\nu}_j^2}c_\zeta j^{-2\delta_\zeta} \leq O_p(d_T\max\{\mathsf{a}_2^{-(3\rho_\nu-2\delta_\zeta+1)/\rho_\nu}, \mathsf{a}_2^{-(1+\rho_\nu)/\rho_\nu}\}), \end{aligned} \tag{S3.57}$$

$$\begin{aligned}&\|\sum_{j=1}^{\mathrm{K}_2}(\nu_j^{-1} - \widehat{\nu}_j^{-1})\langle \widehat{h}_j - h_j^s, \zeta\rangle\widehat{h}_j\|^2 = O_p(d_T\mathsf{a}_2^{-1})\sum_{j=1}^{\mathrm{K}_2}j^{2\rho_\nu}\langle \widehat{h}_j - h_j^s, \zeta\rangle^2 \leq O_p(d_T\mathsf{a}_1\mathsf{a}_2^{-1})\sum_{j=1}^{\mathrm{K}_2}j^{3\rho_\nu-2\delta_\zeta+2} \\ &\leq O_p(d_T\mathsf{a}_1)\max\{\mathsf{a}_2^{-(4\rho_\nu-2\delta_\zeta+3)/\rho_\nu}\mathsf{a}_2^{-(1+\rho_\nu)/\rho_\nu}\}. \end{aligned} \tag{S3.58}$$

Under the conditions on a_1 and a_2 given in Theorem 8, we have $d_T = O(\mathsf{a}_1)$, and $\mathsf{a}_1^{1/2}\mathsf{a}_2^{-1} = o(1)$. These imply that the right hand sides of (S3.57) and (S3.58) are $o_p(1)$. We then use the arguments that are used to show (S2.37) and find that

$$\|\sum_{j=1}^{\mathrm{K}_2}\nu_j^{-1}\langle h_j, \zeta\rangle(\widehat{h}_j - h_j^s)\| \leq O_p(\mathsf{a}_1^{1/2})\sum_{j=1}^{\mathrm{K}_2}j^{-\delta_\zeta+\rho_\nu-2\gamma_\mu+1} \leq O_p(\mathsf{a}_1^{1/2}\max\{\mathsf{a}_2^{-1/\rho_\nu}, \mathsf{a}_2^{(\delta_\zeta-\rho_\nu+2\gamma_\mu-2)/\rho_\nu}\}). \tag{S3.59}$$

From (4.3) and the fact that $\mathsf{a}_1^{1/2}\mathsf{a}_2^{-1} = o(1)$, we can show that the above term is $o_p(1)$. We then note that

$$\|\sum_{j=1}^{\mathrm{K}_2}\nu_j^{-1}\langle \widehat{h}_j - h_j^s, \zeta\rangle\widehat{h}_j\|^2 \leq O_p(\mathsf{a}_1)\sum_{j=1}^{\mathrm{K}_2}j^{2\rho_\nu-2\delta_\zeta+2} \leq O_p(\mathsf{a}_1\max\{\mathsf{a}_2^{-1/\rho_\nu}, \mathsf{a}_2^{(2\delta_\zeta-2\rho_\nu-3)/\rho_\nu}\}), \tag{S3.60}$$

where the inequalities can be deduced from (S3.49). Because of (4.3), the right hand side of (S3.60) is also $o_p(1)$. Thus, by combining the results given in (S3.57)-(S3.60), we conclude that (S3.51) holds. Then the results given in Theorem 6.(i) and 6.(ii) immediately follow from (S3.50) and (S3.51), and hence the details are omitted.

2. Analysis on the regularization bias: We next focus on the regularization bias term,

$\|\mathcal{A}(\widetilde{\Pi}_{\mathrm{K}_2} - \Pi_{\mathrm{K}_2})\zeta\|$. Note that $\|\mathcal{A}(\widetilde{\Pi}_{\mathrm{K}_2} - \Pi_{\mathrm{K}_2})\zeta\| \leq G_1 + G_2 + G_3 + \|\mathcal{A}(\Pi_{\mathrm{K}_2} - I)\zeta\|$, where

$$G_1 = \|\sum_{j=1}^{\mathrm{K}_2}\langle \widehat{h}_j - h_j^s, \zeta\rangle \mathcal{A}(\widehat{h}_j - h_j^s)\|, \quad G_2 = \|\sum_{j=1}^{\mathrm{K}_2}\langle h_j^s, \zeta\rangle \mathcal{A}(\widehat{h}_j - h_j^s)\|, \quad G_3 = \|\sum_{j=1}^{\mathrm{K}_2}\langle \widehat{h}_j - h_j^s, \zeta\rangle \mathcal{A}h_j^s\|.$$

Then, by using (S3.20), (S3.49) and the fact that $d_T\mathtt{a}_1^{-1} = o(1)$, we find that

$$G_1 \leq \sum_{j=1}^{\mathrm{K}_2}|\langle \widehat{h}_j - h_j^s, \zeta\rangle|\|\mathcal{A}(\widehat{h}_j - h_j^s)\| \leq O_p(\mathtt{a}_1)\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu - \delta_\zeta - \varsigma_\nu + 2} \leq o_p(\mathtt{a}_1^{1/2})\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu/2 - \varsigma_\nu - \delta_\zeta + 1},$$

$$G_2 \leq \sum_{j=1}^{\mathrm{K}_2}|\langle h_j^s, \zeta\rangle|\|\mathcal{A}(\widehat{h}_j - h_j^s)\| \leq O_p(\mathtt{a}_1^{1/2})\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu/2 - \varsigma_\nu - \delta_\zeta + 1},$$

and

$$G_3 \leq \sum_{j=1}^{\mathrm{K}_2}|\langle \widehat{h}_j - h_j^s, \zeta\rangle|\|\mathcal{A}h_j^s\| \leq O_p(\mathtt{a}_1^{1/2})\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu/2 - \varsigma_\nu - \delta_\zeta + 1}.$$

Hence, G_1, G_2 and G_3 are bounded by the following.

$$O_p(\mathtt{a}_1^{1/2})\sum_{j=1}^{\mathrm{K}_2} j^{\rho_\nu/2 - \varsigma_\nu - \delta_\zeta + 1} \leq \begin{cases} O_p(\mathtt{a}_1^{1/2}) & \text{if } \rho_\nu/2 + 2 < \varsigma_\nu + \delta_\zeta, \\ O_p(\mathtt{a}_1^{1/2})\max\{\log \mathtt{a}_2^{-1}, \mathtt{a}_2^{-(\rho_\nu/2 - \varsigma_\nu - \delta_\zeta + 2)/\rho_\nu}\} & \text{if } \rho_\nu/2 + 2 \geq \varsigma_\nu + \delta_\zeta. \end{cases}$$

Lastly, we have

$$\|\mathcal{A}(\Pi_{\mathrm{K}_2} - I)\zeta\|^2 \leq \sum_{j=\mathrm{K}_2+1}^{\infty}\|\langle h_j, \zeta\rangle \mathcal{A}h_j\|^2 \leq O(\sum_{j=\mathrm{K}_2+1}^{\infty} j^{-2\delta_\zeta - 2\varsigma_\nu}) \leq O_p(\mathtt{a}_2^{(2\varsigma_\nu + 2\delta_\zeta - 1)/\rho_\nu}),$$

from which the desired result follows. □

S3.3 Supplementary results

S3.3.1 Strong consistency of the F2SLSE

As in the case of the FIVE, we need some additional assumptions: below, as we did for the sequence of d_t in Section S2.3.1, we let $\{M_j\}_{j\geq 1}$ be the sequence of eigenvalues of the covariance of $z_t \otimes z_t - \mathcal{C}_{zz}$.

Assumption S2. (a) $\sup_{t\geq 1}\|x_t\| \leq m_x$, $\sup_{t\geq 1}\|z_t\| \leq m_z$, and $\sup_{t\geq 1}\|u_t\| \leq m_u$ a.s., (b) $M_j \leq ab^j$ for some $a > 0$ and $0 < b < 1$, (c) the sequence of v_t is a martingale difference with respect to $\mathcal{G}_t = \sigma(\{z_s\}_{s\leq t+1}, \{v_s\}_{s\leq t})$.

We may establish the following preliminary results:

Lemma S4. *Under Assumptions M.(b), M*(b) and S2, the following hold almost surely:*

$$\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\mathrm{op}} = O(T^{-1/2}\log^{3/2} T) \quad \text{and} \quad \|\widehat{\mathcal{C}}_{vz}\|_{\mathrm{op}} = O(T^{-1/2}\log^{1/2} T).$$

Proof. The first result follows from Theorem 2.12 and Corollary 2.4 of Bosq (2000). We then note that under Assumptions M.(b) and M*(b), $\sup_{t\geq 1}\|v_t\| \leq \sup_{t\geq 1}\|x_t\| + \|\mathcal{B}\|_{\text{op}} \sup_{t\geq 1}\|z_t\| < \infty$, and apply Theorem 2.14 of Bosq (2000) to find that $\|\widehat{\mathcal{C}}_{zv}\|_{\text{op}} = O(T^{-1/2}\log^{1/2} T)$ a.s. $\square$

In our proof of the strong consistency of the F2SLSE, what we want to have by employing Assumption S2.(c) is the asymptotic order of $\|\widehat{\mathcal{C}}_{vz}\|_{\text{op}}$ given in Lemma S4; in fact, our proof does not require any change once the following weaker condition holds:

$$\|\widehat{\mathcal{C}}_{vz}\|_{\text{op}} = O(T^{-1/2}\log T), \quad a.s. \tag{S3.61}$$

Hence, in the sequel, (S3.61) may replace Assumption S2.(c). We now establish the strong consistency.

Theorem 5 (continued)**.** *If Assumption S2 is additionally satisfied,* $(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j})(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}) = o(T^{1/2}\log^{-3/2}T)$ *a.s.,* $(\sum_{j=\mathrm{K}_1+1}^{\infty}\mu_j)(\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}) = o(1)$ *a.s.,* $\alpha_1^{-1}\alpha_2 \to 0$*, and* $\alpha_1 T^{-1}\log T \to 0$*, then* $\|\widetilde{\mathcal{A}} - \mathcal{A}\|_{\text{op}} \to 0$ *a.s.*

Proof. From (S3.1), we know that $\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\|_{\text{op}} \leq \mathsf{a}_1^{-1/4}\mathsf{a}_2^{-1/4}\|T^{-1}\sum_{t=1}^{T} z_t\otimes u_t\|_{\text{op}}$. Moreover, $\{z_t\otimes u_t\}_{t\geq 1}$ is a martingale difference sequence satisfying that $\sup_t\|z_t\otimes u_t\|_{\text{HS}} < \infty$ a.s and $\sup_t \mathbb{E}\|z_t\otimes u_t\|_{\text{HS}}^2 < \infty$. We therefore deduce from Theorem 2.14 of Bosq (2000) that $\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\|_{\text{op}} = O(\mathsf{a}_1^{-1/4}\mathsf{a}_2^{-1/4}T^{-1/2}\log^{1/2}T)$ a.s., and hence, $\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\|_{\text{op}} = o(1)$ a.s. Note also that $\|\mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2} - \mathcal{A}\|_{\text{op}}^2 \leq \sum_{j=\mathrm{K}_2+1}^{\infty}\|\mathcal{A}h_j^s\|^2 + |\mathcal{R}|$, where h_j^s and $\mathcal{R}$ are defined as in our proof of Theorem 5. Since $\mathcal{A}$ is Hilbert-Schmidt, we have $\sum_{j=\mathrm{K}_2+1}^{\infty}\|\mathcal{A}h_j^s\|^2 = o(1)$ a.s. It thus only remains to show that $|\mathcal{R}| = o(1)$ a.s. We know from Lemma S4 that $\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} = O(T^{-1/2}\log^{3/2}T)$ a.s., and hence

$$O\left(\sum_{j=1}^{\mathrm{K}_1}\mu_j\tau_{1,j}\right)\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j}\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} = o(1), \quad a.s. \tag{S3.62}$$

As shown in Lemma S4, we have $\|\widehat{\mathcal{C}}_{zv}\|_{\text{op}} = O(T^{-1/2}\log^{1/2}T)$, a.s., and also find that

$$\begin{aligned}\|\mathcal{S}\|_{\text{op}}\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} &\leq (O(T^{-1/2}\log^{1/2}T) + O(\mathsf{a}_1^{-1/2}T^{-1}\log T))\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} \\ &= o(\log^{-1}T) + o(\mathsf{a}_1^{-1/2}T^{-1/2}\log^{-1/2}T) = o(1), \quad a.s.\end{aligned} \tag{S3.63}$$

by the definition of $\mathcal{S}$. Moreover,

$$\|\mathcal{T}\|_{\text{op}}\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} \leq O\left(\sum_{j=\mathrm{K}_1+1}^{\infty}\mu_j\right)\sum_{j=1}^{\mathrm{K}_2}\tau_{2,j} = o(1), \quad a.s. \tag{S3.64}$$

From (S3.9), (S3.62), (S3.63), and (S3.64), it immediately follows that $|\mathcal{R}| = o(1)$ a.s. $\square$

S3.3.2 Refinements of the general asymptotic results for the F2SLSE

We in this section provide a complementary result to Theorem 8 for the case where $\rho_\nu/2+2 \geq \varsigma_\nu+\delta_\zeta$. Specifically, the following can be shown:

Theorem 8 (Continued)**.** *Let everything as in Theorem 8 but with $\rho_\nu/2+2 \geq \varsigma_\nu + \delta_\zeta$. Then, Theorem 6 holds and*

$$\|\mathcal{A}(\widetilde{\Pi}_{\mathrm{K}_2} - \Pi_{\mathrm{K}_2})\zeta\| = O_p(\alpha_1^{-1/2}\max\{\log\alpha_2, \alpha_2^{(\rho_\nu/2-\varsigma_\nu-\delta_\zeta+2)/\rho_\nu}\}),$$
$$\|\mathcal{A}(\Pi_{\mathrm{K}_2} - \mathcal{I})\zeta\| = O_p(\alpha_2^{(1/2-\varsigma_\nu-\delta_\zeta)/\rho_\nu}).$$

Our proof of the above result is contained in the proof of Theorem 8 given in Section S3.1, and hence omitted.

S4 Appendix to Section 5 on "Numerical studies"

S4.1 Appendix to Section 5.1

We show that the simulation DGP considered in Section 5.1 satisfies the employed assumptions. From our construction of the variables ($\{y_t, x_t, z_t, u_t, v_t\}_{t\geq 1}$) and operators ($\mathcal{A}$ and $\mathcal{B}$) in Section 5.1 and from Theorem 2.7 of Bosq (2000), it is not difficult to show that Assumptions M and M* are satisfied. Moreover, this setup simplifies the representation of the eigenvalues and eigenfunctions associated with the FIVE and F2SLSE. Specifically, we have $\lambda_j^2 = b_j^2\mu_j^2$, $\nu_j^2 = b_j^4\mu_j^2$, $f_j = \xi_j = h_j = g_j$, $\langle z_t, g_j\rangle \sim \mathcal{N}(0,\mu_j^2)$ and $\langle x_t, g_j\rangle = \langle \vartheta\mathcal{B}z_t, g_j\rangle + \langle v_t, g_j\rangle \sim \mathcal{N}(0,(\vartheta^2 b_j^2+1)\mu_j^2)$. Since $\mathbb{E}[\|x_t\|^4] < \infty$ and $\mathbb{E}[\|z_t\|^4] < \infty$, we find the following: for some $c_\circ > 0$,

$$\mathbb{E}[\|\langle z_t, \xi_j\rangle x_t\|^2] \leq \mathbb{E}[|\langle z_t, \xi_j\rangle|^4]^{1/2}\mathbb{E}[\|x_t\|^4]^{1/2} \leq c_\circ\lambda_j^2, \tag{S4.1}$$
$$\mathbb{E}[\|\langle x_t, \xi_j\rangle z_t\|^2] \leq \mathbb{E}[|\langle x_t, \xi_j\rangle|^4]^{1/2}\mathbb{E}[\|z_t\|^4]^{1/2} \leq c_\circ\lambda_j^2. \tag{S4.2}$$

Moreover, we have

$$|\langle \mathcal{A}f_j, \xi_\ell\rangle| = 2j^{-n_a}1\{j=\ell\} \leq 2j^{-n_a/2}\ell^{-n_a/2}.$$

From these results and the fact that $\{x_t, z_t\}_{t\geq 1}$ is an iid sequence, Assumptions E1, E1* and E2 are also satisfied. It is obvious that Assumptions E2*.(a)-(d) are satisfied from our construction of the variables and operators, and Assumption E2*.(e) can be shown to hold from the fact $h_j = g_j$. We next note that

$$|\langle h_j, \mathcal{B}g_\ell\rangle| \leq b_j 1\{j=\ell\} \leq j^{-n_b/2}\ell^{-n_b/2},$$

and, in the considered setup, $\rho_\nu = 4n_b + 4$ and thus $n_b/2 \leq \rho_\nu/4 + 1/2$. Therefore, Assumption E2*.(f) is satisfied. Lastly, it can also be shown that Assumption E2*.(g) holds from similar arguments used to show (S4.1) and (S4.2) and that $\{z_t\}_{t\geq 1}$ is an iid sequence and $\mathcal{C}_{xz}h_j = \mathcal{C}_{xz}g_j = \lambda_j$.

S4.2 Appendix to Section 5.3

In Section 5.3, we compute the local likelihood estimate of $\log p_t^\circ$ using random samples $\{s_{i,t}\}_{i=1}^n$ drawn from the distribution p_t°. Consider the following log-likelihood:

$$l(\{s_{i,t}\}_{i=1}^n) = \sum_{i=1}^n \log p_t(s_{i,t}) - n \left(\int p_t(v)dv - 1 \right). \tag{S4.3}$$

Under some local smoothness assumptions (Loader, 1996), we can obtain a localized version of (S4.3) by approximating $\log p_t(s)$ using polynomial functions, as follows.

$$l(\{s_{i,t}\}_{i=1}^n)(s) = \sum_{i=1}^n w\left(\frac{s_{i,t}-s}{h_s}\right) H(s_{i,t}-s;\beta_t) - n\int w\left(\frac{v-s}{h_s}\right)\exp(H(v-s;\beta_t))dv, \tag{S4.4}$$

where $w(\cdot)$ is a suitable weight function, h_s is a bandwidth, and $H(v;\beta_t)$ is polynomial in v with coefficients β_t, i.e., $H(v;\beta_t) = \sum_{j=0}^q \beta_{j,t} v^j$ for some nonnegative integer q. For a fixed $s \in [0,1]$, let $\widehat{\beta}_t$ be the maximizer of (S4.4), then the local likelihood log-density estimate is given by $\widehat{\log p_t}(s) = \widehat{\beta}_{0,t}$. By repeating this procedure for a fine grid of points and interpolating the results as described by Loader (2006, Chapter 12), we can obtain $\widehat{\log p_t}$. In our simulation experiment in Section 5.3, $w(\cdot)$ is set to the tricube kernel that is used in many examples given by Loader (2006), $q = 1$, and h_s is set to the nearest neighbor bandwidth covering 33.3% of observations (Loader, 2006, Section 2.2.1).

S4.3 Appendix to Section 5.4

We here define a measure of worker's skill similar to Peri and Sparber's (2009) measure of occupation-specific relative provision of communication versus manual skills. Specifically, we use the O*Net ability survey data,[1] in which the importance of each of 52 distinct abilities required by each occupation is quantified. Using the data, we construct the communication skill measure (c_j°) and the manual skill measure (m_j°) for each occupation j, where the definitions of communication and manual skills are equivalent to the extended definitions in Table A.1 of Peri and Sparber (2009). We merge the values of c_j° and m_j° to individuals in the 2000 census using the monthly US Current Population Survey (CPS) data. Then, as done similarly by Peri and Sparber (2009), the measure of occupation-specific skill intensity (s_j) is obtained by converting the value of c_j°/m_j° to its percentile score (s_{jt}°) for each month in 2000 and averaging the monthly scores for each occupation j, i.e., $s_j = 12^{-1}\sum_{t=1}^{12} s_{jt}^\circ$. The number of distinct skill levels, s_j, is 223, and, by construction, each occupation is uniquely identified by the skill score $s_j \in [0,1]$.[2] In Table S3, we report occupations with the lowest and highest scores of relative communication skill provision.

In addition to estimation results reported in Section 5.4, we examine the null hypothesis H_0 :

[1] Version 24, provided by the US Department of Labor.

[2] A similar rescaling procedure is taken by Peri and Sparber (2009). Specifically, they first converted the values of c_j° and m_j° into their percentile scores in 2000 for each j, and then their measure of relative provision of communication versus manual skills is given by the ratio of the percentile scores for each j. In contrast, we first take the ratio of c_j° and m_j° and then convert the ratio into the percentile score for each j. This is to ensure that our measure of relative communication skill provision takes values in [0,1].

$\mathcal{A}^*\psi = 0$ using the significance test given in Section S7. Note that ψ can be any arbitrary element of $\mathcal{H}$, but here we consider only a few cases, where $\psi \in \{\zeta_{j(3)}\}_{j=1}^{3}$, for the purpose of illustration. Then, by testing H_0, we can examine whether the average of changes in the wages of native workers in the occupations of which $s \in [(j-1)/3, j/3]$ is affected by an inflow of immigrants. Similarly to the simulation experiments in Section S7, we set D to $\lceil T^{1/3} \rceil$ and compute the critical values based on 10,000 Monte Carlo simulations. The testing results are reported in Table S1. In the table, we found that an inflow of immigrants significantly affects the wages of native workers who are in occupations intensive in either manual or communication skills.

Table S1: Significance testing results

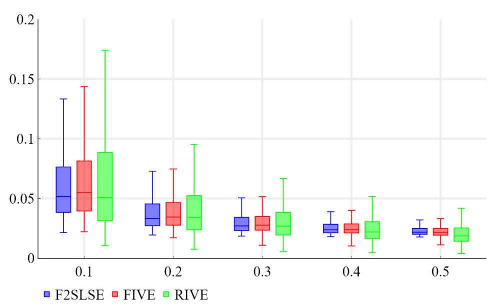
ψ	$\zeta_{1(3)}$	$\zeta_{2(3)}$	$\zeta_{3(3)}$
Test statistic	0.00050*	0.00038	0.00084**

Notes: We use * and ** to denote rejection at 10% and 5% significance levels, respectively.

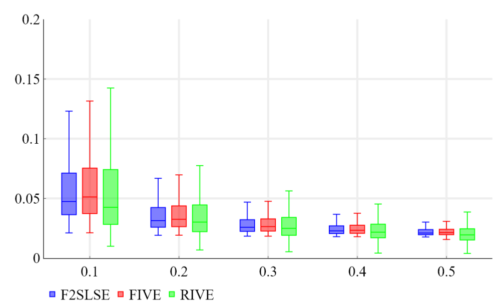
S4.4 Additional tables

Figure S1: Simulation results for Experiment 1: boxplots of the empirical MSEs ($T = 250$)

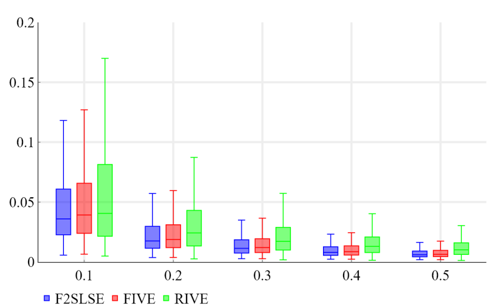
(a) $(n_a, n_b) = (3, 3/4)$



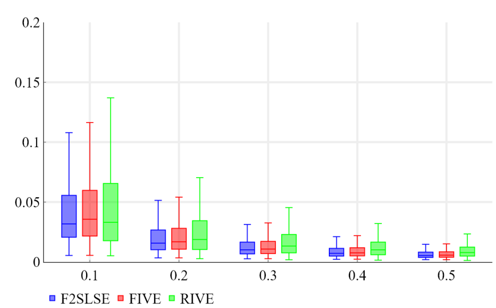
(b) $(n_a, n_b) = (3, 1.5)$



(c) $(n_a, n_b) = (5, 3/4)$



(d) $(n_a, n_b) = (5, 1.5)$



Notes: Boxplots of the empirical MSEs of the FIVE (red), the F2SLSE (blue) and the RIVE (green) are reported for each value of the first-stage functional coefficient of determinations $\mathtt{r}^2 \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

Table S2: Simulation results for Experiment 3: empirical MSEs ($a_1, a_2 \geq 0.6$)

		Sparse Design				Exponential Design			
	n	100		150		100		150	
	T	200	500	200	500	200	500	200	500
Loader's	FIVE	0.185	0.152	0.174	0.149	0.314	0.194	0.255	0.173
	F2SLSE	0.188	0.151	0.174	0.148	0.313	0.188	0.256	0.170
	RIVE	0.194	0.154	0.179	0.150	0.342	0.205	0.276	0.174
	FLSE	0.248	0.221	0.206	0.183	0.419	0.347	0.322	0.260
Silverman's	FIVE	0.273	0.231	0.222	0.194	0.339	0.242	0.271	0.207
	F2SLSE	0.275	0.228	0.225	0.193	0.352	0.234	0.277	0.201
	RIVE	0.271	0.226	0.224	0.192	0.342	0.237	0.278	0.199
	FLSE	0.351	0.318	0.267	0.240	0.422	0.350	0.319	0.257

Notes: Based on 1,000 replications. Each cell reports the empirical mean squared error (MSE) of the four considered estimators: FIVE, F2SLSE, Benatia et al.'s (2017) RIVE and Park and Qian's (2012) FLSE.

Table S3: Occupations with the lowest and highest communication skill intensity in 2000 (denoted s)

Four occupations with the lowest s	s
Pressing machine operators (clothing)	0.0010
Construction Trades	0.0035
Machine operators	0.0235
Garbage and recyclable material collectors	0.0242
Four occupations with the highest s	s
Chief executives and public administrators	0.9926
Operations and systems researchers and analysts	0.9954
Management Analysts	0.9985
Economists, market researchers, and survey researchers	1.0000

S5 Computation

We here only describe how to compute the FIVE $\widehat{\mathcal{A}}$ from observations $\{y_t, x_t, z_t\}_{t=1}^T$; in fact, computation of the F2SLSE $\widetilde{\mathcal{A}}$ can be done with only slight modifications and is thus omitted. Specifically, for each T, $\widehat{\mathcal{A}}$ is a finite rank operator acting on the Hilbert space of square-integrable functions defined on $[0, 1]$, so it allows the following representation (Gohberg, Goldberg, and Kaashoek, 2013, Chapter 8):

$$\widehat{\mathcal{A}}\upsilon(s_1) = \int_0^1 \widehat{\kappa}(s_1, s_2)\upsilon(s_2)ds_2, \quad s_1, s_2 \in [0, 1], \tag{S5.1}$$

where υ is any arbitrary random or nonrandom element taking values in $\mathcal{H}$ (for example, υ can be x_t or any fixed element in $\mathcal{H}$). Therefore, computation of $\widehat{\mathcal{A}}$ reduces to obtaining an explicit formula for the integral kernel $\widehat{\kappa}(s_1, s_2)$ for $s_1, s_2 \in [0,1]$. We here present a way to compute this integral kernel from the eigenelements of $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ and $\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^*$, which can be obtained by the standard functional principal component method; see e.g., Ramsay and Silverman (2005, Chapter 8) and Horváth and Kokoszka (2012, Chapter 3).

Let $\{\widehat{\xi}_j\}_{j\geq 1}$ be the collection of the eigenfunctions of $\widehat{\mathcal{C}}_{xz}\widehat{\mathcal{C}}_{xz}^*$, and then note that $\widehat{\mathcal{C}}_{xz}\widehat{f}_j = \widehat{\lambda}_j\widehat{\xi}_j$ and $\widehat{\mathcal{C}}_{xz}^*\widehat{\xi}_j = \widehat{\lambda}_j\widehat{f}_j$, where $\widehat{\lambda}_j = \langle \widehat{\mathcal{C}}_{xz}\widehat{f}_j, \widehat{\xi}_j\rangle$ (Bosq, 2000, Section 4.3). Then, $\widehat{\mathcal{A}}$ is given by

$$\widehat{\mathcal{A}} = \widehat{\mathcal{C}}_{yz}^*\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1} = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{\mathrm{K}} \widehat{\lambda}_j^{-1}\langle \widehat{\xi}_j, z_t\rangle \widehat{f}_j \otimes y_t. \tag{S5.2}$$

It is quite obvious from (S5.2) that the integral kernel $\widehat{\kappa}(s_1, s_2)$ for $s_1, s_2 \in [0,1]$ is given by $T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{\mathrm{K}} \widehat{\lambda}_j^{-1}\langle \widehat{\xi}_j, z_t\rangle \widehat{f}_j(s_1)y_t(s_2)$, and this can be equivalently expressed as follows:

$$\widehat{\kappa}(s_1, s_2) = T^{-1}\widehat{F}_{\mathrm{K}}(s_1)'\mathrm{diag}(\{\widehat{\lambda}_j^{-1}\}_{j=1}^{\mathrm{K}})\widehat{G}_{\mathrm{K}}Y_T(s_2), \tag{S5.3}$$

where $Y_T(s) = (y_1(s), \ldots, y_T(s))'$, $\widehat{F}_{\mathrm{K}}(s) = (\widehat{f}_1(s), \ldots, \widehat{f}_{\mathrm{K}}(s))'$ for $s \in [0,1]$, and $\widehat{G}_{\mathrm{K}}$ is the $\mathrm{K} \times T$ matrix whose (i,t)-th element is given by $\langle \widehat{\xi}_i, z_t\rangle$. Thus, for each choice of s_1 and s_2, $\widehat{\kappa}(s_1, s_2)$ can be obtained by simple matrix multiplications. In view of (S5.2) and (S5.3), $\widehat{\mathcal{A}}\upsilon$, for any arbitrary random or nonrandom element υ taking values in $\mathcal{H}$, is computed as follows: $\widehat{\mathcal{A}}\upsilon(s) = T^{-1}\widehat{F}_{\mathrm{K}}(\upsilon)'\mathrm{diag}(\{\widehat{\lambda}_j^{-1}\}_{k=1}^{\mathrm{K}})\widehat{G}_{\mathrm{K}}Y_T(s)$, where $s \in [0,1]$ and $\widehat{F}_{\mathrm{K}}(\upsilon)' = (\langle \widehat{f}_1, \upsilon\rangle, \ldots, \langle \widehat{f}_{\mathrm{K}}, \upsilon\rangle)'$.

Computing the FIVE requires choosing $\alpha = \mathsf{a}^{-1}$ defined in (3.2). The eigenvalue $\widehat{\lambda}_j^2$ of $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ depends on the scales of x_t and z_t, and this needs to be considered in choosing α. In practice, it thus may be of interest to have a scale-invariant choice of α. This can be done by computing the contribution of each of the eigenvalues to a magnitude of the operator $\widehat{\mathcal{C}}_{xz}$ and then viewing α^{-1} as the threshold parameter for such computed contributions. We illustrate an easy-to-implement way here. Define $\widehat{r}_k = \widehat{\lambda}_k^2 \big/ \sum_{j=1}^{\infty}\widehat{\lambda}_j^2$. Since $\|\widehat{\mathcal{C}}_{xz}\|_{\mathrm{HS}}^2 = \sum_{j=1}^{T}\widehat{\lambda}_j^2$, the ratio $\widehat{r}_k$ computes the contribution of the k-th eigenvalue to the squared Hilbert-Schmidt norm of $\widehat{\mathcal{C}}_{xz}$. Of course, the above quantity does not depend on the scales of x_t and z_t, and hence, a scale-invariant version of (3.2) may be written as

$$\mathrm{K} = \#\{j : \widehat{r}_k > 1/\alpha_T\}, \tag{S5.4}$$

where α_T (> 1) depends only on T and diverges to infinity as T increases. An alternative way is to directly choose K, instead of α, as the minimal number of the eigenvalues whose sum exceeds a pre-specified proportion of $\|\widehat{\mathcal{C}}_{xz}\|_{\mathrm{HS}}^2$. (Of course, even in this case, it is more natural to understand K as a random variable.) To be more specific, let

$$\widehat{\mathrm{R}}_k = \sum_{j=1}^{k}\widehat{\lambda}_j^2 \Big/ \sum_{j=1}^{\infty}\widehat{\lambda}_j^2 \quad \text{and} \quad \mathrm{K} = \min_k\{k : \widehat{\mathrm{R}}_k > (1 - 1/\alpha_T)\}, \tag{S5.5}$$

where α_T is similarly defined as in (S5.4). This choice is obviously scale-invariant as well.

It is also possible to pursue a data-driven selection of α_T in (S5.4) and (S5.5), such as a cross-validation approach, proposed by Benatia et al. (2017) developed in an iid setting. Such a procedure may be adapted for dependent non-iid data, but this will not be further studied in this paper.

S6 FIVE with a general weighting operator

The theoretical results given for the FIVE in Section 3 can be extended to the case involving a general weighting operator as in Euclidean space setting. Let $\mathcal{W}$ be a self-adjoint positive definite operator, and define

$$\bar{z}_t = \mathcal{W}^{1/2} z_t.$$

If $\ker \mathcal{C}_{xz} = \{0\}$, due to the positive definiteness of $\mathcal{W}$, $\mathcal{A}$ is uniquely identified from the equation: $\mathcal{C}^*_{y\bar{z}}\mathcal{C}_{x\bar{z}} = \mathcal{A}\mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}}$. A natural extension of the FIVE can be defined as follows:

$$\widehat{\mathcal{A}}(\mathcal{W}) = \widehat{\mathcal{C}}^*_{y\bar{z}}\widehat{\mathcal{C}}_{x\bar{z}} \left(\widehat{\mathcal{C}}^*_{x\bar{z}}\widehat{\mathcal{C}}_{x\bar{z}}\right)^{-1}_{\mathrm{K}} = \widehat{\mathcal{C}}^*_{yz}\mathcal{W}\widehat{\mathcal{C}}_{xz} \left(\widehat{\mathcal{C}}^*_{xz}\mathcal{W}\widehat{\mathcal{C}}_{xz}\right)^{-1}_{\mathrm{K}}. \tag{S6.1}$$

In this case, the theoretical results given in Section 3 can easily be extended by an obvious conversion of the assumptions given for z_t into those for $\bar{z}_t$. This can further be extended to the case with a possibly random weighting operator $\widehat{\mathcal{W}}$ satisfying some conditions, which will be discussed in this section.

It will be useful to define additional notation. We let

$$\breve{z}_t = \widehat{\mathcal{W}}^{1/2} z_t, \tag{S6.2}$$

and let $\widehat{\mathcal{A}}(\widehat{\mathcal{W}})$ (resp. K) be defined by replacing $\mathcal{W}$ (resp. λ_j) with $\widehat{\mathcal{W}}$ in (S6.1) (the j-th ordered eigenvalue $\breve{\lambda}_j$ of $\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}} = \widehat{\mathcal{C}}^*_{xz}\widehat{\mathcal{W}}\widehat{\mathcal{C}}_{xz}$ in (3.2)). Let $\mathfrak{L}_+$ be the collection of self-adjoint positive definite operators in $\mathcal{L}_{\mathcal{H}}$, and write the spectral representations of $\mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}}$ and $\mathcal{C}_{x\bar{z}}\mathcal{C}^*_{x\bar{z}}$ as

$$\mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}} = \sum_{j=1}^{\infty} \bar{\lambda}_j^2 \bar{f}_j \otimes \bar{f}_j \quad \text{and} \quad \mathcal{C}_{x\bar{z}}\mathcal{C}^*_{x\bar{z}} = \sum_{j=1}^{\infty} \bar{\lambda}_j^2 \bar{\xi}_j \otimes \bar{\xi}_j, \tag{S6.3}$$

respectively. We also let $\bar{\Pi}_{\mathrm{K}}$ denote the projection onto the span of the first K eigenfunctions of $\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}}$.

Assumption S3. $\widehat{\mathcal{W}} \in \mathfrak{L}_+$ a.s. and $\|\widehat{\mathcal{W}} - \mathcal{W}\|_{\mathrm{op}} = O_p(T^{-1/2})$ for some fixed $\mathcal{W} \in \mathfrak{L}_+$.

Note that the F2SLSE discussed in Section 4 does not satisfy Assumption S3 as $(\widehat{\mathcal{C}}_{zz})^{-1}_{\mathrm{K}_1}$ diverges in operator norm.

Theorem S1. *Suppose that Assumption S3 holds.*

(i) If the assumptions in Theorem 1 when λ_j is replaced with $\bar{\lambda}_j$ are also satisfied,

$$\|\widehat{\mathcal{A}}(\widehat{\mathcal{W}}) - \mathcal{A}\bar{\Pi}_{\mathrm{K}}\|^2_{\mathrm{op}} = O_p(T^{-1}\alpha) \quad \textit{and} \quad \|\mathcal{A}(\mathcal{I} - \bar{\Pi}_{\mathrm{K}})\|^2_{\mathrm{op}} = o_p(1).$$

(ii) If the assumptions in Theorem 2 when λ_j is replaced with $\bar{\lambda}_j$ are also satisfied,

$$\sqrt{T/\bar{\theta}_{\mathrm{K}}(\zeta)}(\widehat{\mathcal{A}}(\widehat{\mathcal{W}}) - \mathcal{A}\bar{\Pi}_{\mathrm{K}})\zeta \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{uu}) \quad \text{and} \quad |\breve{\theta}_{\mathrm{K}}(\zeta) - \bar{\theta}_{\mathrm{K}}(\zeta)| \xrightarrow{p} 0, \tag{S6.4}$$

where

$$\bar{\theta}_{\mathrm{K}}(\zeta) \coloneqq \langle \zeta, (\mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}})^{-1}_{\mathrm{K}}\mathcal{C}^*_{x\bar{z}}\mathcal{C}_{\bar{z}\bar{z}}\mathcal{C}_{x\bar{z}}(\mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}})^{-1}_{\mathrm{K}}\zeta \tag{S6.5}$$

and $\breve{\theta}_{\mathrm{K}}(\zeta)$ is defined by replacing $\bar{z}_t$ with $\breve{z}_t$ in the above.

Proof. We note that

$$\widehat{\mathcal{A}}(\widehat{\mathcal{W}}) = \widehat{\mathcal{C}}^*_{y\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}}(\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}})^{-1}_{\mathrm{K}} = \mathcal{A}\bar{\Pi}_{\mathrm{K}} + \widehat{\mathcal{C}}^*_{u\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}}(\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}})^{-1}_{\mathrm{K}}. \tag{S6.6}$$

From Assumption S3 and our construction of K, the following can be shown: (a) $\|\widehat{\mathcal{C}}_{u\breve{z}}\|_{\mathrm{HS}} = O_p(T^{-1/2})$, (b) $\|\widehat{\mathcal{C}}_{x\breve{z}}(\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}})^{-1}_{\mathrm{K}}\|_{\mathrm{op}} = \mathsf{a}^{-1/2} = \alpha^{1/2}$, and (c) $\|\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}} - \mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}}\|_{\mathrm{op}} = \|\widehat{\mathcal{C}}^*_{xz}\widehat{\mathcal{W}}\widehat{\mathcal{C}}_{xz} - \mathcal{C}^*_{xz}\mathcal{W}\mathcal{C}_{xz}\|_{\mathrm{op}} = O_p(T^{-1/2})$. Combining these with similar arguments used in our proofs of Theorems 1 and 2, the desired results are obtained. □

Theorem S2. *Suppose that Assumption S3 holds.*

(i) If the assumptions in Theorem 3 when λ_j, f_j and ξ_j are, respectively, replaced with $\bar{\lambda}_j$, $\bar{f}_j$ and $\bar{\xi}_j$ are also satisfied, then $\|\widehat{\mathcal{A}}(\widehat{\mathcal{W}}) - \mathcal{A}\bar{\Pi}_{\mathrm{K}}\|^2_{\mathrm{op}} = O_p(T^{-1}\alpha)$ as in Theorem S1 and

$$\|\mathcal{A}(\mathcal{I} - \bar{\Pi}_{\mathrm{K}})\|^2_{\mathrm{op}} = O_p(T^{-1}\alpha \max\{1, \alpha^{(3-2\varsigma)/\rho}\} + \alpha^{(1-2\varsigma)/\rho}). \tag{S6.7}$$

Thus, $\|\widehat{\mathcal{A}}(\widehat{\mathcal{W}}) - \mathcal{A}\|_{\mathrm{op}} = o_p(1)$ for any $\rho > 2$ and $\varsigma > 1/2$.

(ii) If the assumptions in Theorem 4 when λ_j, f_j and ξ_j are, respectively, replaced with $\bar{\lambda}_j$, $\bar{f}_j$ and $\bar{\xi}_j$ are also satisfied, then Theorem S1.(ii) holds and

$$\|\mathcal{A}(\bar{\Pi}_{\mathrm{K}} - \Pi_{\mathrm{K}})\zeta\| = \begin{cases} O_p(T^{-1/2}) & \text{if } \rho/2 + 2 < \varsigma + \delta_\zeta, \\ O_p(T^{-1/2}\max\{\log\alpha, \alpha^{(\rho/2-\varsigma-\delta_\zeta+2)/\rho}\}) & \text{if } \rho/2 + 2 \geq \varsigma + \delta_\zeta, \end{cases}$$
$$\|\mathcal{A}(\Pi_{\mathrm{K}} - \mathcal{I})\zeta\| = O_p(\alpha^{(1/2-\varsigma-\delta_\zeta)/\rho}).$$

Proof. Let $\breve{f}_j$ be the eigenfunction corresponding to the j-th largest eigenvalue of $\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}}$ and let $\bar{f}^s_j = \mathrm{sgn}\{\langle \breve{f}_j, \bar{f}_j\rangle\}\bar{f}_j$. Using the fact that $\|\widehat{\mathcal{C}}^*_{x\breve{z}}\widehat{\mathcal{C}}_{x\breve{z}} - \mathcal{C}^*_{x\bar{z}}\mathcal{C}_{x\bar{z}}\|_{\mathrm{op}} = \|\widehat{\mathcal{C}}^*_{xz}\widehat{\mathcal{W}}\widehat{\mathcal{C}}_{xz} - \mathcal{C}^*_{xz}\mathcal{W}\mathcal{C}_{xz}\|_{\mathrm{op}} = O_p(T^{-1/2})$ and nearly identical arguments used in our proof of (S2.15) and (S2.16), the following can be shown under the assumptions employed for (i):

$$\|\breve{f}_j - \bar{f}^s_j\|^2 = O_p(j^2 T^{-1}), \tag{S6.8}$$
$$\|\mathcal{A}(\breve{f}_j - \bar{f}^s_j)\|^2 = O_p(T^{-1})(j^{2-2\varsigma} + j^{\rho+2-2\varsigma}). \tag{S6.9}$$

Then, the desired result given in (i) can easily be obtained from (S6.8), (S6.9) and similar arguments that we used to establish (S2.17) and (S2.18).

Moreover, under the assumptions employed for (ii), the following can be shown by nearly identical arguments that are used to obtain (S2.33):

$$\langle \breve{f}_j - \bar{f}_j^s, \zeta\rangle^2 = O_p(T^{-1})j^{-2\delta_\zeta+2} + O_p(T^{-1})j^{-2\delta_\zeta+2+\rho}(1+o_p(1)). \tag{S6.10}$$

The remaining proof is almost identical to that of Theorem 4; by using (S6.8)-(S6.10), it can be shown that $\|\widehat{\mathcal{C}}_{x\check{z}}(\widehat{\mathcal{C}}_{x\check{z}}^*\widehat{\mathcal{C}}_{x\check{z}})_{\mathrm{K}}^{-1}\zeta - \mathcal{C}_{x\bar{z}}(\mathcal{C}_{x\bar{z}}^*\mathcal{C}_{x\bar{z}})_{\mathrm{K}}^{-1}\zeta\| = o_p(1)$. We can also analyze the term $\|\mathcal{A}(\bar{\Pi}_{\mathrm{K}} - \mathcal{I})\zeta\|$, as $\|\mathcal{A}(\widehat{\Pi}_{\mathrm{K}} - \mathcal{I})\zeta\|$ in our proof of Theorem 4. The details are omitted. □

S7 Significance testing in functional endogenous linear model

Practitioners may often be interested in examining if various characteristics of y_t depend on x_t. For example, in our empirical application where $y_t(s)$ represents the wage of workers of skill level $s \in [0,1]$, practitioners might be interested in examining if the average wage ($\int_0^1 y_t(s)ds$) is affected by the considered explanatory variable x_t. Likewise, because various characteristics of y_t can be written as $\langle y_t, \psi\rangle$ for $\psi \in \mathcal{H}$, we in this section develop a significance test for examining if $\langle y_t, \psi\rangle$ is affected by x_t. Specifically, for any $\psi \in \mathcal{H}$, observe that

$$\langle y_t, \psi\rangle = \langle x_t, \mathcal{A}^*\psi\rangle + \langle u_t, \psi\rangle.$$

We then want to test the following null and alternative hypotheses:

$$H_0 : \mathcal{A}^*\psi = 0 \quad \text{v.s.} \quad H_1 : \mathcal{A}^*\psi \neq 0. \tag{S7.1}$$

The null hypothesis means that the characteristic $\langle y_t, \psi\rangle$ of y_t does not linearly depend on x_t. Note that $\widehat{\mathcal{C}}_{yz}\psi = \widehat{\mathcal{C}}_{xz}\mathcal{A}^*\psi + \widehat{\mathcal{C}}_{uz}\psi$, and hence $\widehat{\mathcal{C}}_{yz}\psi$ reduces to $\widehat{\mathcal{C}}_{uz}\psi$ if the null is true; moreover, in this case, $\sqrt{T}\widehat{\mathcal{C}}_{uz}\psi$ turns out to weakly converge to a $\mathcal{H}$-valued Gaussian random element under relevant assumptions. Using this property, we develop a significance test, which is described by Theorem S3.

Theorem S3. *Suppose that (i) $\mathcal{C}_{uu}$ is positive definite, (ii) either Assumption M or Assumption M* holds, (iii) $\bar{\mathcal{A}}$ is the FIVE (if Assumption M holds) or the F2SLSE (if Assumption M* holds) and the other assumptions for $\|\bar{\mathcal{A}} - \mathcal{A}\|_{\mathrm{op}} \xrightarrow{p} 0$ are satisfied (see Theorems 1, 3, 5 and 7). Let $\bar{u}_t = y_t - \bar{\mathcal{A}}x_t$ and let $\mathcal{J}$ denote $T\|\widehat{c}_\psi^{-1}\widehat{\mathcal{C}}_{yz}\psi\|^2$, where $\widehat{c}_\psi^2 = \langle T^{-1}\sum_{t=1}^T \bar{u}_t \otimes \bar{u}_t(\psi), \psi\rangle$. Then, the following hold (below $\varkappa_j \sim_{iid} \mathcal{N}(0,1)$).*

(i) $\mathcal{J} \xrightarrow{d} \sum_{j=1}^\infty \mu_j \varkappa_j^2$ under H_0 of (S7.1) while $\mathcal{J} \xrightarrow{p} \infty$ under H_1 of (S7.1).

(ii) Let $\widehat{q}_{1-a_0}$ be the $100(1-a_0)\%$ quantile of $\sum_{j=1}^D \widehat{\mu}_j \varkappa_j^2$ for $a_0 \in (0,1)$, $D \to \infty$ and $D = o(T^{1/2})$. Then $\mathbb{P}\{\mathcal{J} > \widehat{q}_{1-a_0}\} \to a_0$ under H_0 of (S7.1) while $\mathbb{P}\{\mathcal{J} > \widehat{q}_{1-a_0}\} \to 1$ under H_1 of (S7.1).

Proof. For notational convenience, let $c_\psi = \langle \mathcal{C}_{uu}\psi, \psi\rangle^{1/2}$. To show (i), first note that

$$\sqrt{T}\widehat{\mathcal{C}}_{yz}\psi = \sqrt{T}\widehat{\mathcal{C}}_{xz}\mathcal{A}^*\psi + \sqrt{T}\widehat{\mathcal{C}}_{uz}\psi. \tag{S7.2}$$

Under H_0, the first term in (S7.2) is equal to zero, and thus $\sqrt{T}\widehat{\mathcal{C}}_{yz}\psi = T^{-1/2}\sum_{t=1}^{T} c_\psi \psi_t$, where $\psi_t = c_\psi^{-1}\langle u_t, \psi\rangle z_t$. Then, note that $\mathbb{E}[\psi_t] = 0$, $\mathbb{E}[\psi_t \otimes \psi_t] = \mathcal{C}_{zz}$ and $\{\langle \psi_t, \zeta\rangle\}_{t\geq 1}$ is a real-valued martingale difference sequence for any $\zeta \in \mathcal{H}$. Thus, by applying nearly identical arguments that are used to show (S2.8) and (S2.9), it can be shown that, for any $\zeta \in \mathcal{H}$ and $m > 0$, $T^{-1/2}\sum_{t=1}^{T}\langle \psi_t, \zeta\rangle \xrightarrow{d} \mathcal{N}(0, \langle \mathcal{C}_{zz}\zeta, \zeta\rangle)$ and $\limsup_{n\to\infty}\limsup_T \mathbb{P}(\sum_{j=n+1}^{\infty}\langle T^{-1/2}\sum_{t=1}^{T}\psi_t, g_j\rangle^2 > m) = 0$ since $\mathcal{C}_{zz}$ is Hilbert-Schmidt. Therefore, $T^{-1/2}\sum_{t=1}^{T}\psi_t \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{zz})$, and we note that $\mathcal{N}(0, \mathcal{C}_{zz}) \overset{d}{=} \sum_{j=1}^{\infty}\sqrt{\mu_j}\varkappa_j g_j$, where $\varkappa_j \sim_{\text{iid}} \mathcal{N}(0,1)$ across j. The rest of the proof follows from the consistency of $\widehat{c}_\psi$ (Corollaries 1 and 2), continuous mapping theorem, and orthonormality of $\{g_j\}_{j\geq 1}$.

Under H_1, the first term in (S7.2) is not equal to zero, and, by combining the results given above, we find that $\|\sqrt{T}\widehat{\mathcal{C}}_{yz}\psi\|^2 = T\|\mathcal{C}_{xz}\mathcal{A}^*\psi\|^2 + O_p(1)$ holds in this case. Since $\ker \mathcal{C}_{xz} = \{0\}$ under either of Assumptions M or M*, we find that $\|\mathcal{C}_{xz}\mathcal{A}^*\psi\|^2 > 0$ under H_1, and therefore $\mathcal{J} \xrightarrow{p} \infty$.

To show (ii), note that $|\sum_{j=1}^{\infty}\mu_j\varkappa_j^2 - \sum_{j=1}^{D}\widehat{\mu}_j\varkappa_j^2| \leq \sum_{j=1}^{D}|\mu_j - \widehat{\mu}_j||\varkappa_j^2| + \sum_{j=D+1}^{\infty}\mu_j\varkappa_j^2$. The first term in the right hand side is bounded above by $\sup_{j\geq 1}|\mu_j - \widehat{\mu}_j|\sum_{j=1}^{D}\varkappa_j^2$, where $D^{-1}\sum_{j=1}^{D}\varkappa_j^2 = O_p(1)$ by the Markov's inequality, and $\sup_{j\geq 1}|\mu_j - \widehat{\mu}_j| \leq \|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{op}} \leq O_p(T^{-1/2})$ under Assumption M.(f) and Lemma 4.2 in Bosq (2000). Since $\mathcal{C}_{zz}$ is nonnegative, we also have $\mathbb{E}[|\sum_{j=D+1}^{\infty}\mu_j\varkappa_j^2|] \leq \sum_{j=D+1}^{\infty}\mu_j \to 0$ as $D \to \infty$, which implies that $\sum_{j=D+1}^{\infty}\mu_j\varkappa_j^2$ is $o_p(1)$. Since $D \to \infty$ and $D/\sqrt{T} \to 0$ as $T \to \infty$, we find that $|\sum_{j=1}^{\infty}\mu_j\varkappa_j^2 - \sum_{j=1}^{D}\widehat{\mu}_j\varkappa_j^2| = o_p(1)$. From this result, we find that $\widehat{\eta}_{1-a_0}$ converges to η_{1-a_0} (see e.g., Lemma 21.2 of van der Vaart, 1998), where η_{1-a_0} is the $100(1-a_0)\%$ quantile of the distribution function G of $\sum_{j=1}^{\infty}\mu_j\varkappa_j^2$. It is then obvious that $1 - \mathbb{P}\{\mathcal{J} > \widehat{\eta}_{1-a_0}\}$ converges to $G(\eta_{1-a_0})$. By combining this with the previous result Theorem S3.(i), the desired limiting behavior of $\mathbb{P}\{\mathcal{J} > \widehat{q}_{1-a_0}\}$ is established. $\square$

Theorem S3.(i) shows that the asymptotic null distribution of the proposed statistic does not depend on $\psi \in \mathcal{H}$, but does depend on all the eigenvalues of $\mathcal{C}_{zz}$; this means that there are infinitely many nuisance parameters. However, we can approximate the limiting distribution using the estimated eigenvalues of $\mathcal{C}_{zz}$ and thus implement a valid asymptotic test without any significant difficulty, as detailed in Theorem S3.(ii); once the estimated eigenvalues $\widehat{\mu}_j$ are obtained, then the approximate quantile $\widehat{q}_{1-a_0}$ for any significance level $a_0 \in (0,1)$ can easily be computed by Monte Carlo simulations. Thus, implementation of the proposed test in practice is straightforward, which will be further illustrated in Section 5.

Remark S1. The test proposed in Theorem S3 can obviously be extended to examine the following hypotheses:

$$H_0 : \mathcal{A}^*\psi = \psi_0 \quad \text{v.s.} \quad H_1 : \mathcal{A}^*\psi \neq \psi_0, \tag{S7.3}$$

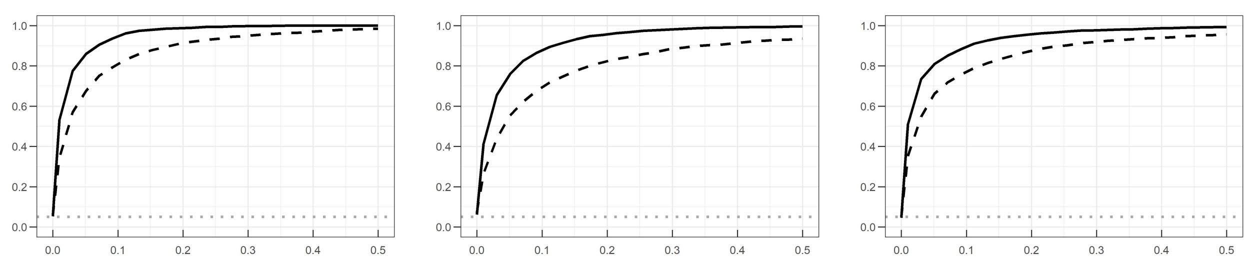
for any $\psi_0 \in \mathcal{H}$. The extension only requires redefining $\mathcal{J}$ as $T\|\widehat{c}_\psi^{-1}(\widehat{\mathcal{C}}_{yz}\psi - \widehat{\mathcal{C}}_{xz}\psi_0)\|^2$, and this does not make any change in the convergence results given in Theorem S3.

Finite sample performance

We hereby explore the finite sample performance of the test for examining the null and alternative hypotheses (S7.3), which is proposed in Remark S1 of Section S7. To this end, we consider the DGP employed in Section 5.2. The test statistic $\mathcal{J}$ is computed with $\widehat{c}_\psi$ obtained from the FIVE (see Theorem S3 and Corollary 1). For each realization of the DGP, the critical value at 5% significance level is obtained from 500 Monte Carlo simulations of the distribution given in Theorem S3.(ii) with $D = \lceil T^{1/3} \rceil$. The finite sample properties of the test are investigated by computing its rejection probabilities when $\mathcal{A}^*\zeta = \psi_0 + c_\zeta \tilde{\psi}$ holds, where ζ is defined in Section 5.2, $c_\zeta \in \{0, 0.01, \ldots, 0.5\}$, and $\tilde{\psi}$ is a perturbation element with unit norm and is randomly generated for each realization of the DGP. Specifically, $\tilde{\psi} = \ddot{\psi}/\|\ddot{\psi}\|$ and $\ddot{\psi} = \sum_{j=1}^{11} \ddot{q}_{3,j}\xi_j$, where $\ddot{q}_{3,j} \sim_{\text{iid}} \mathrm{N}(0, 0.5^{2(j-1)})$ across j and $\mathbb{E}[\ddot{q}_{1,i}\ddot{q}_{3,j}] = 0$ for all i and j. This section only considers the case where $\sigma_\eta = 0.5$; in unreported simulations, we also investigated the performance of the test (i) when $\widehat{c}_\psi$ is computed from the F2SLSE and (ii) when σ_η is given by 0.9, but found no significant difference.

Figure S2: Simulation results for Experiment 2: rejection probability of $\mathcal{J}$ when $\mathcal{A}^*\psi = \psi_0 + c_\zeta\tilde{\psi}$

(a) Sparse Design (b) Exponential Design (c) Noisy Design



Notes: Based on 1,000 Monte Carlo replications. The rejection probability of $\mathcal{J}$ when $\widehat{c}_\psi$ is computed from the FIVE is reported according to the value of $c_\zeta \in \{0, 0.01, \ldots, 0.5\}$.

Figure S2 shows the rejection probability of the test depending on c_ζ. The dashed and solid lines indicate the rejection probabilities when $T = 200$ and 500, respectively, and the dotted horizontal lines indicate the nominal size of the test. As expected, the proposed test exhibits a higher power as c_ζ gets deviated from zero, and it seems that the power of the test more rapidly increases in the sparse design compared to those in the other designs. Moreover, in all the considered cases, the test has excellent size control, as can be seen from the case where $c_\zeta = 0$.

References

Baker, C. R. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society 186*, 273–289.

Benatia, D., M. Carrasco, and J.-P. Florens (2017). Functional linear regression with functional response. *Journal of Econometrics 201*(2), 269–291.

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

Gohberg, I., S. Goldberg, and M. Kaashoek (2013). *Classes of Linear Operators Vol. I.* Birkhäuser.

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

Ibukiyama, T. and M. Kaneko (2014). The Euler–Maclaurin summation formula and the Riemann zeta function. In *Bernoulli Numbers and Zeta Functions*, pp. 65–74. Springer, Tokyo.

Imaizumi, M. and K. Kato (2018). PCA-based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis 163*, 15–36.

Loader, C. R. (1996). Local likelihood density estimation. *Annals of Statistics 24*(4), 1602–1618.

Loader, C. R. (2006). *Local Regression and Likelihood*. Springer, New York.

Park, J. Y. and J. Qian (2012). Functional regression of continuous state distributions. *Journal of Econometrics 167*(2), 397–412.

Peri, G. and C. Sparber (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics 1*(3), 135–169.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.