Estimation of Characteristics-based Quantile Factor Models: Online Appendix

Liang Chen[1], Juan J. Dolado[2], Jesús Gonzalo[3], and Haozi Pan[4]

[1] *HSBC Business School, Peking University, chenliang@phbs.pku.edu.cn*
[2] *Department of Economics, Universidad Carlos III de Madrid, dolado@eco.uc3m.es*
[3] *Department of Economics, Universidad Carlos III de Madrid, jgonzalo@est-eco.uc3m.es*
[4] *Centro de Estudios Monetarios y Financieros (CEMFI), haozi.pan@cemfi.edu.es*

May 23, 2026

A Proofs and Additional Results

A.1 Proofs of the Main Results

Recall that the check function is given by $\rho_\tau(u) = (\tau - \mathbf{1}\{u \leq 0\})u$ and that, by definition, $l(\theta, y_{it}, \boldsymbol{x}_i) = \rho_\tau(y_{it} - \theta(\boldsymbol{x}_i)) - \rho_\tau(y_{it} - \theta_t(\boldsymbol{x}_i))$, $L_n(\theta) = n^{-1}\sum_{i=1}^n l(\theta, y_{it}, \boldsymbol{x}_i)$. Moreover, for any $\theta \in \Theta$, define $K(\theta, \theta_t) = \mathbb{E}(L_n(\theta)) = \mathbb{E}[l(\theta, y_{it}, \boldsymbol{x}_i)]$, and for any $\theta_1, \theta_2 \in \Theta$, define the pseudo-metric $d(\theta_1, \theta_2) \equiv \sqrt{\mathbb{E}\left(\theta_1(\boldsymbol{x}_i) - \theta_2(\boldsymbol{x}_i)\right)^2}$. Finally, Θ_n is the sieve space defined in the main text.

We first present the following lemma based on Corollary 1 of Chen and Shen (1998), and then prove Proposition 1 below using this lemma.

Lemma 1 (Corollary 1 of Chen and Shen (1998))**.** *Suppose that Assumption 1(i) holds and the following conditions hold for all $t \leq T$: (i) $K(\theta, \theta_t) \geq 0$ for $\theta \in \Theta$. (ii) $K(\theta, \theta_t) \asymp d(\theta, \theta_t)^2$ for $\theta \in \Theta$. (iii) $\sup_{\theta\in\Theta} |l(\theta, y_{it}, \boldsymbol{x}_i)| < \infty$ for all i, t. (iv) For all small $\epsilon > 0$,*

$$\sup_{\theta\in\Theta_n : d(\theta,\theta_t)\leq\epsilon} Var[l(\theta, y_{it}, \boldsymbol{x}_i)] \lesssim \epsilon^2.$$

(v) Let $\mathcal{F}_{n,T,\delta} = \{l(\theta, y_{it}, \boldsymbol{x}_i) - l(\theta_t, y_{it}, \boldsymbol{x}_i) : d(\theta, \theta_t) \leq \delta, \theta \in \Theta_n\}$ be a class of functions indexed by Θ_n. There exists $\delta_n \in (0, 1)$ such that

$$\delta_n = \sup\left\{\delta > 0 : \delta^{-2}\int_{a\delta^2}^{b\delta} \sqrt{H(w, \mathcal{F}_{n,T,\delta})}dw \lesssim n^{1/2}\right\},$$

where $H(w, \mathcal{F})$ is the bracketing L_2 metric entropy of the space $\mathcal{F}$. Then there exist constants $c, C > 0$ such that for any $\gamma \geq 1$ and any integer n,

$$P\left[d(\hat{\theta}_t, \theta_t) \geq \gamma \varepsilon_n\right] \leq 4c \exp(-Cn\varepsilon_n^2 \gamma^2)$$

for all t, where $\varepsilon_n = \max(\delta_n, d(\theta_t, \boldsymbol{a}_{0t}' \boldsymbol{\phi}_{k_n}))$.

Proposition 1. *Under Assumption 1, it holds that $\max_{1 \leq t \leq T} d(\hat{\theta}_t, \theta_t) = O_P(\varepsilon_{nT})$ when either T is fixed or $T \to \infty$ as $n \to \infty$.*

Proof. First, under Assumption 1(iv), it can be shown that $K(\theta, \theta_t) \asymp d(\theta, \theta_t)^2$, which implies that conditions (i) and (ii) of Lemma 1 hold. Second, by Assumption 1(iii) we have $\sup_{\theta \in \Theta} |l(\theta, y_{it}, \boldsymbol{x}_i)| \lesssim \sup_{\theta \in \Theta} \sup_{\mathcal{X}} |\theta(\boldsymbol{x}) - \theta_t(\boldsymbol{x})| < \infty$. Thus, condition (iii) of Lemma 1 holds. Third, by the definition of d and the properties of the check function, it is easy to see that,[1]

$$\begin{aligned} \sup_{\theta \in \Theta_n, d(\theta, \theta_t) \leq \varepsilon} \mathrm{Var}\left[l(\theta, y_{it}, \boldsymbol{x}_i)\right] &\leq \sup_{\theta \in \Theta_n, d(\theta, \theta_t) \leq \varepsilon} \mathbb{E}\left[l(\theta, y_{it}, \boldsymbol{x}_i)\right]^2 \\ &\lesssim \sup_{\theta \in \Theta_n, d(\theta, \theta_t) \leq \varepsilon} \mathbb{E}\left(\theta(\boldsymbol{x}_i) - \theta_t(\boldsymbol{x}_i)\right)^2 \leq \varepsilon^2. \end{aligned}$$

Thus, condition (iv) of Lemma 1 is also satisfied. Fourth, for the finite-dimensional linear sieve spaces Θ_n, it can be shown that condition (v) of Lemma 1 is satisfied with $\delta_n = \sqrt{k_n/n}$ (see Section 3.3 of Chen (2007)). Finally, Assumption 1(ii) implies that $d(\theta_t, \boldsymbol{a}_{0t}' \boldsymbol{\phi}_{k_n}) = O(k_n^{-\alpha})$. Therefore, it follows from Lemma 1 that (set $\gamma = \sqrt{\ln T}$)

$$P\left[\max_t d(\hat{\theta}_t, \theta_t) \geq C\varepsilon_{nT}\right] \leq \sum_{t=1}^{T} P\left[d(\hat{\theta}_t, \theta_t) \geq C\varepsilon_n \sqrt{\ln T}\right] \lesssim \exp\left\{C^2 \ln T(1 - c_2 n \varepsilon_n^2)\right\}$$

for any $C \geq 1$. Therefore, the desired result follows from the above inequality since $n\varepsilon_n^2 \geq k_n$. □

Lemma 2. *If Assumption 1 and Assumption 2(i) hold, then:*
(i) $\max_{1 \leq t \leq T} \|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\| = O_P(\varepsilon_{nT})$;
(ii) Let $\hat{\boldsymbol{V}} \equiv \hat{\boldsymbol{Y}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'$, then $(nT)^{-1/2}\|\hat{\boldsymbol{V}}\| = O_P(\varepsilon_{nT})$.

[1] Note that $|\rho_\tau(u_1) - \rho_\tau(u_2)| \leq 2|u_1 - u_2|$.

Proof. By Assumption 1 and Assumption 2(i),

$$\begin{aligned}
d(\hat\theta_t,\theta_t)^2 &= \int_{\mathcal{X}} \left(\hat\theta_t(\boldsymbol{x}) - \theta_t(\boldsymbol{x})\right)^2 d\mathsf{F}_x(\boldsymbol{x}) \\
&= \int_{\mathcal{X}} \left(\hat\theta_t(\boldsymbol{x}) - \boldsymbol{a}_{0t}'\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\right)^2 d\mathsf{F}_x(\boldsymbol{x}) + O_P(\varepsilon_{nT}k_n^{-\alpha}) \\
&= (\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})'\boldsymbol{\Sigma}_\phi(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}) + O_P(\varepsilon_{nT}k_n^{-\alpha}) \\
&\geq c_1\|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\|^2 + O_P(\varepsilon_{nT}k_n^{-\alpha})
\end{aligned}$$

where $c_1 > 0$, and the $O_P(\varepsilon_{nT}k_n^{-\alpha})$ term is uniform in t. Then it follows from Proposition 1 that $\max_{1\leq t\leq T}\|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\|^2 = O_P(\varepsilon_{nT}^2)$.

Next, note that

$$\begin{aligned}
(nT)^{-1}\|\hat{\boldsymbol{V}}\|^2 &\lesssim \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(\hat\theta_t(\boldsymbol{x}_i) - \boldsymbol{a}_{0t}'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\right)^2 + O_P(k_n^{-2\alpha}) \\
&= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left((\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})'\,\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\right)^2 + O_P(k_n^{-2\alpha}) \\
&\leq T^{-1}\sum_{t=1}^T \|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\|^2 \cdot \lambda_{\max}\left(\hat{\boldsymbol{\Sigma}}_\phi\right) + O_P(k_n^{-2\alpha}) \\
&\leq \max_{1\leq t\leq T}\|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\|^2 \cdot \lambda_{\max}\left(\hat{\boldsymbol{\Sigma}}_\phi\right) + O_P(k_n^{-2\alpha})
\end{aligned}$$

where $\hat{\boldsymbol{\Sigma}}_\phi \equiv n^{-1}\sum_{i=1}^n \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'$. Since Assumption 1(iii) implies that $\sup_{\mathcal{X}}\|\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\| = \sqrt{k_n}$, similar to the proof of Theorem 1 in Newey (1997), one can show that $\|\hat{\boldsymbol{\Sigma}}_\phi - \boldsymbol{\Sigma}_\phi\| = o_P(1)$ under Assumption 2, and therefore we have $\lambda_{\max}(\hat{\boldsymbol{\Sigma}}_\phi) = O_P(1)$. This completes the proof. □

Proof of Theorem 1:

Proof. Write $\hat{\boldsymbol{Y}} = \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0' + \hat{\boldsymbol{V}}$ where $\hat{\boldsymbol{V}}$ is as defined in Lemma 1. Note that

$$\begin{aligned}
\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}/(nT) = \boldsymbol{F}_0\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'/(nT) + \hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'/(nT) \\
+ \boldsymbol{F}_0\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}/(nT) + \hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}/(nT). \quad \text{(A.1)}
\end{aligned}$$

It then follows from Assumption 2(iv), Assumption 1(i) and Lemma 1 that:

$$\begin{aligned}
&\|\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}/(nT) - \boldsymbol{F}_0\boldsymbol{\Sigma}_g\boldsymbol{F}'/T\| \\
\leq\; & o_P(1) + 2\|\hat{\boldsymbol{V}}\|/\sqrt{nT}\cdot\|\boldsymbol{G}(\boldsymbol{X})\|/\sqrt{n}\cdot\|\boldsymbol{F}_0\|/\sqrt{T} + \|\hat{\boldsymbol{V}}\|^2/(nT) \\
=\; & o_P(1) + O_P(\varepsilon_{nT}).
\end{aligned}$$

Recall that $\hat{\boldsymbol{\Omega}}$ is the diagonal matrix whose elements are the eigenvalues of $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}/(nT)$, and let $\boldsymbol{\Omega}$ be the diagonal matrix whose elements are the eigenvalues of $\boldsymbol{\Sigma}_g \cdot \boldsymbol{F}_0'\boldsymbol{F}_0/T$. By the Wielandt-Hoffman inequality, the above result implies that $\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\| = o_P(1)$. It then follows from Assumption 2(iii) and 2(iv) that $\lambda_{\min}(\hat{\boldsymbol{\Omega}}) > 0$ with probability approaching 1.

By the definition of $\hat{\boldsymbol{F}}$, $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}/(nT)\hat{\boldsymbol{F}} = \hat{\boldsymbol{F}}\hat{\boldsymbol{\Omega}}$, it then follows from (A.1) that

$$\hat{\boldsymbol{F}} = \boldsymbol{F}_0\hat{\boldsymbol{H}} + \hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'\hat{\boldsymbol{F}}/(nT)\hat{\boldsymbol{\Omega}}^{-1} + \boldsymbol{F}_0\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/(nT)\hat{\boldsymbol{\Omega}}^{-1} + \hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}/(nT)\hat{\boldsymbol{F}}\hat{\boldsymbol{\Omega}}^{-1}. \quad \text{(A.2)}$$

Thus, it follows from (A.2) and Lemma 1 that

$$\|\hat{\boldsymbol{F}} - \boldsymbol{F}_0\hat{\boldsymbol{H}}\|/\sqrt{T} \leq O_P(1) \cdot \frac{\|\hat{\boldsymbol{V}}\|}{\sqrt{nT}} \cdot \frac{\|\boldsymbol{F}_0\|}{\sqrt{T}} \cdot \frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} \cdot \frac{\|\boldsymbol{G}(\boldsymbol{X})\|}{\sqrt{n}} + O_P(1) \cdot \frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} \cdot \frac{\|\hat{\boldsymbol{V}}\|^2}{nT} = O_P(\varepsilon_{nT}).$$

Then the first part of Theorem 1 follows.

Next, similar to the proof of Proposition 1 in Bai (2003) it can be shown that $\hat{\boldsymbol{H}} \to \boldsymbol{H} > 0$. Thus, $\hat{\boldsymbol{H}}$ is invertible with probability approaching 1. Note that $\hat{\boldsymbol{G}}(\boldsymbol{X}) = \hat{\boldsymbol{Y}}\hat{\boldsymbol{F}}/T = \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T + \hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/T$. Write $\boldsymbol{F}_0 = \hat{\boldsymbol{F}}\hat{\boldsymbol{H}}^{-1} + \boldsymbol{F}_0 - \hat{\boldsymbol{F}}\hat{\boldsymbol{H}}^{-1}$, then

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) = \boldsymbol{G}(\boldsymbol{X})(\hat{\boldsymbol{H}}')^{-1} + \boldsymbol{G}(\boldsymbol{X})(\boldsymbol{F}_0 - \hat{\boldsymbol{F}}\hat{\boldsymbol{H}}^{-1})'\hat{\boldsymbol{F}}/T + \hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/T,$$

and thus

$$\|\hat{\boldsymbol{G}}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X})(\hat{\boldsymbol{H}}')^{-1}\|/\sqrt{n} \leq \frac{\|\boldsymbol{G}(\boldsymbol{X})\|}{\sqrt{n}} \cdot \frac{\|\boldsymbol{F}_0 - \hat{\boldsymbol{F}}\hat{\boldsymbol{H}}^{-1}\|}{\sqrt{T}} \cdot \frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} + \frac{\|\hat{\boldsymbol{V}}\|}{\sqrt{nT}} \cdot \frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} = O_P(\varepsilon_{nT}).$$

Then the second part of Theorem 1 follows.

Finally, note that $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}\hat{\boldsymbol{F}}/T = \boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T) + (\hat{\boldsymbol{A}} - \boldsymbol{A}_0)\hat{\boldsymbol{F}}/T$. It follows from Proposition 1 that

$$\|\hat{\boldsymbol{B}} - \boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)\| \leq \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}_0\|}{\sqrt{T}} \cdot \frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} = O_P(\varepsilon_{nT}). \quad \text{(A.3)}$$

Thus, for any $\boldsymbol{x} \in \mathcal{X}$,

$$\begin{aligned}\hat{\boldsymbol{g}}(\boldsymbol{x})' &= \boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\hat{\boldsymbol{B}} = \boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T) + \boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\left(\hat{\boldsymbol{B}} - \boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)\right) \\ &= \boldsymbol{g}(\boldsymbol{x})'(\hat{\boldsymbol{H}}^{-1})' + (\boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\boldsymbol{B}_0 - \boldsymbol{g}(\boldsymbol{x})')(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T) + \boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\left(\hat{\boldsymbol{B}} - \boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)\right) + O_P(\varepsilon_{nT}).\end{aligned}$$

Thus, it follows from (A.3) and Assumption 1 that

$$\sup_{\mathcal{X}} \left\|\hat{\boldsymbol{g}}(\boldsymbol{x}) - \hat{\boldsymbol{H}}^{-1}\boldsymbol{g}(\boldsymbol{x})\right\| \leq O_P(k_n^{-\alpha}) + \sup_{\mathcal{X}} \|\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\| \cdot O_P(\varepsilon_{nT}) = O_P(\sqrt{k_n}\varepsilon_{nT}).$$

This completes the proof. $\square$

Lemma 3. *Let* $\xi_{it} = \boldsymbol{g}(\boldsymbol{x}_i)'\boldsymbol{f}_{0t} - \boldsymbol{a}_{0t}'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ *and* $\psi_{it} = \mathsf{F}(-\xi_{it}) - \mathbf{1}\{u_{it} \leq -\xi_{it}\}$*. If Assumptions 1 to 3 hold, then*

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t} - \mathsf{f}^{-1}(0)\cdot\hat{\boldsymbol{\Sigma}}_{\phi}^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n}\psi_{it}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\right\|^2} = O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right).$$

Proof. **Step 1:** For any $\boldsymbol{a} \in \mathbb{R}^{Dk_n}$ define:

$$\boldsymbol{m}_t(\boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^{n}\left[\tau - \mathbf{1}\{u_{it} \leq (\boldsymbol{a}-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\}\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i),$$

$$\boldsymbol{m}_t^*(\boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^{n}\left[\tau - \mathsf{F}\left((\boldsymbol{a}-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\right)\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i).$$

Since $\mathsf{F}(-\xi_{it}) = \tau - \mathsf{f}(-\xi_{it}^*)\cdot\xi_{it}$ where ξ_{it}^* is between 0 and ξ_{it}, it follows that

$$\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{f}(-\xi_{it}^*)\cdot\xi_{it}\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i). \tag{A.4}$$

Taylor Expansion of $\boldsymbol{m}_t^*(\hat{\boldsymbol{a}}_t)$ around $\boldsymbol{a}_{0t}$ gives

$$\boldsymbol{m}_t^*(\hat{\boldsymbol{a}}_t) = \boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) - \boldsymbol{M}_t^*(\tilde{\boldsymbol{a}}_t)\cdot(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}) \tag{A.5}$$

where $\tilde{\boldsymbol{a}}_t$ is between $\boldsymbol{a}_{0t}$ and $\hat{\boldsymbol{a}}_t$ and

$$\boldsymbol{M}_t^*(\tilde{\boldsymbol{a}}_t) = -\frac{\partial\boldsymbol{m}_t^*(\boldsymbol{a})}{\partial\boldsymbol{a}'}|_{\boldsymbol{a}=\tilde{\boldsymbol{a}}_t} = \frac{1}{n}\sum_{i=1}^{n}\mathsf{f}\left((\tilde{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\right)\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'. \tag{A.6}$$

By Assumption 3(ii) one can write

$$\boldsymbol{M}_t^*(\tilde{\boldsymbol{a}}_t) = \mathsf{f}(0)\cdot\hat{\boldsymbol{\Sigma}}_{\phi} + n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X}), \tag{A.7}$$

where $\hat{\boldsymbol{\Sigma}}_{\phi} = n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})$ and $\boldsymbol{D}_t^*$ is a $n \times n$ diagonal matrix whose diagonal elements are bounded in absolute values by $L\left|(\tilde{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\right|$. Note that by Lemma 1,

$$\begin{aligned}\max_{1\leq t\leq T}\|\boldsymbol{D}_t^*\|_S &\lesssim \max_{i,t}\left|(\tilde{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\right| \\ &\leq \max_{1\leq t\leq T}\|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\|\cdot O_P(\sqrt{k_n}) + O_P(k_n^{-\alpha}) = O_P(\sqrt{k_n}\varepsilon_{nT}).\end{aligned} \tag{A.8}$$

Moreover, one can write

$$\boldsymbol{m}_t^*(\hat{\boldsymbol{a}}_t) = \boldsymbol{m}_t(\hat{\boldsymbol{a}}_t) - \tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) + \left[\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)\right] \tag{A.9}$$

where $\tilde{\boldsymbol{m}}_t(\boldsymbol{a}) = \boldsymbol{m}_t(\boldsymbol{a}) - \boldsymbol{m}_t^*(\boldsymbol{a})$. It then follows from (A.5) (A.7) and (A.9) that

$$\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t} - \mathsf{f}^{-1}(0) \cdot \hat{\boldsymbol{\Sigma}}_\phi^{-1} \cdot \tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) = \mathsf{f}^{-1}(0) \cdot \hat{\boldsymbol{\Sigma}}_\phi^{-1}$$
$$\left\{\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) - \boldsymbol{m}_t(\hat{\boldsymbol{a}}_t) - [\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)] - n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})\right\},$$

where

$$\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) = \frac{1}{n}\sum_{i=1}^n \left[\mathsf{F}\left(-\xi_{it}\right) - \mathbf{1}\{u_{it} \leq -\xi_{it}\}\right] \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) = \frac{1}{n}\sum_{i=1}^n \psi_{it}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i).$$

Since $\mathsf{f}(0)$ is bounded below, and $\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_\phi)$ is bounded below with probability approaching 1, it suffices to show that

$$\max_{1\leq t\leq T} \|\boldsymbol{m}_t^*(\boldsymbol{a}_{0t})\| = O_P(k_n^{-\alpha}), \tag{A.10}$$

$$\max_{1\leq t\leq T} \|\boldsymbol{m}_t(\hat{\boldsymbol{a}}_t)\| = O_P(k_n^{3/2}/n), \tag{A.11}$$

$$\frac{1}{T}\sum_{t=1}^T \|\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)\|^2 = O_P\left(\eta_{nT}^2\right), \tag{A.12}$$

$$\max_{1\leq t\leq T} \left\|n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})\right\| = O_P(\sqrt{k_n}\varepsilon_{nT}^2). \tag{A.13}$$

Step 2: By (A.4) and Assumption 1,

$$\begin{aligned} & \max_{1\leq t\leq T} \|\boldsymbol{m}_t^*(\boldsymbol{a}_{0t})\| \\ = & \max_{1\leq t\leq T} \left\|\frac{1}{n}\sum_{i=1}^n \mathsf{f}\left(-\xi_{it}^*\right) \cdot \xi_{it} \cdot \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i).\right\| \\ \leq & \max_{1\leq t\leq T} \left\|\frac{1}{n}\sum_{i=1}^n \mathsf{f}\left(0\right) \cdot \xi_{it} \cdot \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\right\| + O_P\left(k_n^{1/2-2\alpha}\right). \end{aligned}$$

Define $z_{it} = \mathsf{f}\left(0\right) \cdot \xi_{it}$ and $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{nt})'$, then

$$\frac{1}{n}\sum_{i=1}^n \mathsf{f}\left(0\right) \cdot \xi_{it} \cdot \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) = n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{z}_t$$

and

$$\begin{aligned} & \max_{1\leq t\leq T} \left\|\frac{1}{n}\sum_{i=1}^n \mathsf{f}\left(0\right) \cdot \xi_{it} \cdot \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\right\| \\ = & \max_{1\leq t\leq T} \left\|n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{z}_t\right\| \leq \left\|n^{-1/2}\boldsymbol{\Phi}(\boldsymbol{X})\right\|_S \cdot \max_{1\leq t\leq T} \left\|n^{-1/2}\boldsymbol{z}_t\right\| = O_P(k_n^{-\alpha}). \end{aligned}$$

In sum, we have

$$\max_{1\le t\le T}\|\boldsymbol{m}_t^*(\boldsymbol{a}_{0t})\| = O_P(k_n^{1/2-2\alpha}) + O_P(k_n^{-\alpha}) = O_P(k_n^{-\alpha}),$$

which gives (A.10).

Step 3: Similar to the proof of Lemma A4 of Horowitz and Lee (2005) it can be shown that

$$\max_{1\le t\le T}\|\boldsymbol{m}_t(\hat{\boldsymbol{a}}_t)\| = O_P(k_n^{3/2}/n),$$

which gives (A.11).

Step 4: By (A.8) and Lemma 1

$$\max_{1\le t\le T}\left\|n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})\right\| \\ \le \|\boldsymbol{\Phi}(\boldsymbol{X})/\sqrt{n}\|_S^2\cdot\max_{1\le t\le T}\|\boldsymbol{D}_t^*\|_S\cdot\max_{1\le t\le T}\|\hat{\boldsymbol{a}}_t-\boldsymbol{a}_{0t}\| = O_P(\sqrt{k_n}\varepsilon_{nT}^2),$$

which gives (A.13).

Step 5: Define:

$$\delta_{1t}(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^n\left[\mathbf{1}\{u_{it}\le(\boldsymbol{a}-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)-\xi_{it}\}-\mathbf{1}\{u_{it}\le-\xi_{it}\}\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i),$$

$$\delta_{2t}(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^n\left[\mathsf{F}\left((\boldsymbol{a}-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)-\xi_{it}\right)-\mathsf{F}\left(-\xi_{it}\right)\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i),$$

$$\tilde{\delta}_{1t}(\boldsymbol{\alpha}) = \delta_{1t}(\boldsymbol{\alpha})-\mathbb{E}[\delta_{1t}(\boldsymbol{\alpha})],\qquad \tilde{\delta}_{2t}(\boldsymbol{\alpha}) = \delta_{2t}(\boldsymbol{\alpha})-\mathbb{E}[\delta_{2t}(\boldsymbol{\alpha})].$$

Note that $\mathbb{E}[\delta_{1t}(\boldsymbol{\alpha})] = \mathbb{E}[\delta_{2t}(\boldsymbol{\alpha})]$ because $\delta_{2t}(\boldsymbol{\alpha}) = \mathbb{E}[\delta_{1t}(\boldsymbol{\alpha})|\boldsymbol{x}_i]$. Then $\tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)-\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) = \tilde{\delta}_{2t}(\hat{\boldsymbol{a}}_t)-\tilde{\delta}_{1t}(\hat{\boldsymbol{a}}_t)$, and

$$\frac{1}{T}\sum_{t=1}^T\|\tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)-\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t})\|^2 \le \frac{1}{T}\sum_{t=1}^T\left\|\tilde{\delta}_{1t}(\hat{\boldsymbol{a}}_t)\right\|^2+\frac{1}{T}\sum_{t=1}^T\left\|\tilde{\delta}_{2t}(\hat{\boldsymbol{a}}_t)\right\|^2. \tag{A.14}$$

In what follows, we will show that

$$\frac{1}{T}\sum_{t=1}^T\left\|\tilde{\delta}_{1t}(\hat{\boldsymbol{a}}_t)\right\|^2 = O_P\left(\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2})\cdot k_n^{5/2}\varepsilon_{nT}n^{-1}\right), \tag{A.15}$$

$$\frac{1}{T}\sum_{t=1}^T\left\|\tilde{\delta}_{2t}(\hat{\boldsymbol{a}}_t)\right\|^2 = O_P\left(\ln(k_n^{-1/2}\varepsilon_{nT}^{-1})\cdot k_n^3\varepsilon_{nT}^2n^{-1}\right), \tag{A.16}$$

which imply (A.12) and therefore complete the proof. We will focus on the proof of (A.15) since the proof of (A.16) is similar.

Let $\phi_{jd}(\boldsymbol{x}_i)$ be the jdth element of $\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ for $j = 1, \ldots, k_n; d = 1, \ldots, D$, and define

$$\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i) = \mathbf{1}\{u_{it} \le (\boldsymbol{a} - \boldsymbol{a}_{0t})' \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\} - \mathbf{1}\{u_{it} \le -\xi_{it}\}.$$

Then for some $C > 0$, with probability approach 1,

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{\delta}_{1t}(\hat{\boldsymbol{a}}_t)\right\|^2 \le \frac{1}{n}\cdot\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{k_n}\sum_{d=1}^{D}\sup_{\|\boldsymbol{a}-\boldsymbol{a}_{0t}\|\le C\varepsilon_{nT}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i)]\}\right|^2$$

We will show that

$$\mathbb{E}\left[\sup_{\|\boldsymbol{a}-\boldsymbol{a}_{0t}\|\le C\varepsilon_{nT}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i)]\}\right|^2\right]$$
$$= O\left(\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2}) \cdot k_n^{3/2}\varepsilon_{nT}\right) \quad \text{(A.17)}$$

uniformly in t and j, from which (A.15) follows.

Define $\mathcal{H}_{\varepsilon_{nT}} = \{h(\boldsymbol{a}, \boldsymbol{x}_i) \equiv \Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i)] : \|\boldsymbol{a} - \boldsymbol{a}_{0t}\| \le C\varepsilon_{nT}\}$, and for any $h \in \mathcal{H}_{\varepsilon_{nT}}$ define $\mathbb{G}_n h = n^{-1/2}\sum_{i=1}^{n} h(\boldsymbol{a}, \boldsymbol{x}_i)$. Write

$$\sup_{\|\boldsymbol{a}-\boldsymbol{a}_{0t}\|\le C\varepsilon_{n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i)]\}\right| = \|\mathbb{G}_n h\|_{\mathcal{H}_{\varepsilon_{nT}}},$$

then the left-hand side of (A.17) can be written as $\mathbb{E}\,\|\mathbb{G}_n h\|^2_{\mathcal{H}_{\varepsilon_{nT}}}$. Let $N(\mathcal{H}_{\varepsilon_{nT}}, L_2(Q), \epsilon)$ be the covering number of $\mathcal{H}_{\varepsilon_{nT}}$, where $L_2(Q)$ is the L_2 norm for functions and Q is any probability measure on $\mathcal{X}$. Similar to the proof of (A.12) in Kato et al. (2012), it can be shown that $N(\mathcal{H}_{\varepsilon_{nT}}, L_2(Q), 2\epsilon) \le (A/\epsilon)^{c_1 k_n}$ for some bounded constant c_1 and $A \ge 3\sqrt{e}$ that do not depend on t and j. Moreover, it is easy to show that $\sup_{h\in\mathcal{H}_{\varepsilon_{nT}}}\mathbb{E}[h^2(\boldsymbol{a}, \boldsymbol{x}_i)] \le c_2^2\sqrt{k_n}\varepsilon_n$ for some bounded constant c_2. Then, applying Proposition B.1 of Kato et al. (2012), we have

$$\mathbb{E}\,\|\mathbb{G}_n h\|_{\mathcal{H}_{\varepsilon_{nT}}} \le c_3\left[\cdot\ln(c_4 k_n^{-1/4}\varepsilon_{nT}^{-1/2}) \cdot k_n/\sqrt{n} + \sqrt{\ln(c_4 k_n^{-1/4}\varepsilon_{nT}^{-1/2})} \cdot k_n^{3/4}\varepsilon_{nT}^{1/2}\right]$$
$$\le c_5\sqrt{\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2})} \cdot k_n^{3/4}\varepsilon_{nT}^{1/2}, \quad \text{(A.18)}$$

where c_3, c_4, c_5 are bounded constants that do not depend on t and j. Finally, (A.17) follows by noting that (see Chapter 6 of Ledoux and Talagrand 1991)

$$\mathbb{E}\,\|\mathbb{G}_n h\|^2_{\mathcal{H}_{\varepsilon_{nT}}} \le \left(\mathbb{E}\,\|\mathbb{G}_n h\|_{\mathcal{H}_{\varepsilon_{nT}}}\right)^2 + O(n^{-1}).$$

This completes the proof. $\square$

Proof of Theorem 2:

Proof. Let $\boldsymbol{\Psi}$ be the $n \times T$ matrix of ψ_{it}, then the result of Lemma 2 can be written as

$$\left\| \hat{\boldsymbol{A}} - \boldsymbol{A}_0 - \mathsf{f}(0)^{-1} \cdot \hat{\boldsymbol{\Sigma}}_\phi^{-1} \boldsymbol{\Phi}'(\boldsymbol{X}) \boldsymbol{\Psi}/n \right\| / \sqrt{T} = O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right). \tag{A.19}$$

From (A.2) and Lemma 1 we have

$$\|\hat{\boldsymbol{F}} - \boldsymbol{F}_0 \hat{\boldsymbol{H}}\|/\sqrt{T} \leq O_P(1) \cdot \|\boldsymbol{F}_0 \boldsymbol{G}(\boldsymbol{X})' \hat{\boldsymbol{V}}/(nT)\|_S + O_P(\varepsilon_{nT}^2). \tag{A.20}$$

Define $\boldsymbol{R}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}_0 - \boldsymbol{G}(\boldsymbol{X})$, then by Assumption 1(ii) $\|\boldsymbol{R}(\boldsymbol{X})\|/\sqrt{n} = O_P(k_n^{-\alpha})$. Moreover, we can write

$$\begin{aligned}
\hat{\boldsymbol{V}} &= \hat{\boldsymbol{Y}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0' \\
&= \boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0' \\
&= \boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 + \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0' \\
&= \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0) + \boldsymbol{R}(\boldsymbol{X})\boldsymbol{F}_0'.
\end{aligned}$$

Thus,

$$\begin{aligned}
& \boldsymbol{F}_0 \boldsymbol{G}(\boldsymbol{X})' \hat{\boldsymbol{V}}/(nT) \\
= \;& \boldsymbol{F}_0(\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}_0 - \boldsymbol{R}(\boldsymbol{X}))'[\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0) + \boldsymbol{R}(\boldsymbol{X})\boldsymbol{F}_0']/(nT) \\
= \;& \boldsymbol{F}_0 \boldsymbol{B}_0' \boldsymbol{\Phi}(\boldsymbol{X})' \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0)/(nT) - \boldsymbol{F}_0 \boldsymbol{R}(\boldsymbol{X})' \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0)/(nT) \\
& + \boldsymbol{F}_0 \boldsymbol{G}(\boldsymbol{X})' \boldsymbol{R}(\boldsymbol{X}) \boldsymbol{F}_0'/(nT).
\end{aligned}$$

It then follows from Theorem 1 and Lemma 1 that

$$\|\boldsymbol{F}_0 \boldsymbol{G}(\boldsymbol{X})' \hat{\boldsymbol{V}}/(nT)\|_S \leq \|\boldsymbol{F}_0 \boldsymbol{B}_0' \boldsymbol{\Phi}(\boldsymbol{X})' \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0)/(nT)\|_S + O_P(k_n^{-\alpha}).$$

The above inequality and (A.20) imply that

$$\|\hat{\boldsymbol{F}} - \boldsymbol{F}_0 \hat{\boldsymbol{H}}\|/\sqrt{T} \leq \|\boldsymbol{F}_0 \boldsymbol{B}_0' \boldsymbol{\Phi}(\boldsymbol{X})' \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0)/(nT)\|_S + O_P(k_n^{-\alpha}) + O_P(\varepsilon_{nT}^2). \tag{A.21}$$

By (A.19) and Assumption 1(ii), we have

$$
\begin{aligned}
& \|\boldsymbol{F}_0\boldsymbol{B}_0'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)/(nT)\|_S \\
\leq\ & \mathsf{f}(0)^{-1}\|\boldsymbol{B}_0'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{\Sigma}}_\phi^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}/(n^2T^{1/2})\|_S + O_P\left(k_n^{-\alpha}+\eta_{nT}\right) \\
=\ & \mathsf{f}(0)^{-1}\|\boldsymbol{B}_0'\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}/(nT^{1/2})\|_S + O_P\left(k_n^{-\alpha}+\eta_{nT}\right) \\
\leq\ & \mathsf{f}(0)^{-1}\|\boldsymbol{G}'(\boldsymbol{X})\boldsymbol{\Psi}/(nT^{1/2})\| + \|\boldsymbol{G}(\boldsymbol{X})-\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}_0\|/\sqrt{n}\cdot\|\boldsymbol{\Psi}\|/\sqrt{nT} + O_P\left(k_n^{-\alpha}+\eta_{nT}\right) \\
=\ & \mathsf{f}(0)^{-1}\|\boldsymbol{G}'(\boldsymbol{X})\boldsymbol{\Psi}/(nT^{1/2})\| + O_P\left(k_n^{-\alpha}+\eta_{nT}\right).
\end{aligned}
$$

Note that

$$
\|\boldsymbol{G}'(\boldsymbol{X})\boldsymbol{\Psi}/(nT^{1/2})\| = \frac{1}{\sqrt{n}}\cdot\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{x}_i)\psi_{it}\right\|^2} = O_P(n^{-1/2})
$$

because it is easy to see that $\mathbb{E}\left\|n^{-1/2}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{x}_i)\psi_{it}\right\|^2<\infty$ for all t. It then follows from (A.21) that

$$
\|\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}}\|/\sqrt{T} = O_P(n^{-1/2})+O_P(k_n^{-\alpha})+O_P(\eta_{nT})+O_P(\varepsilon_{nT}^2).
$$

This completes the proof. □

Lemma 4. *Under Assumptions 1, 2 and 4, we have*

$$
\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}_0-\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n\right\|/\sqrt{T} = O_P\left(k_n^{-\alpha}\right)+O_P\left(\eta_{nT}\right).
$$

where $\psi_{it}(\boldsymbol{x}_i)=\mathsf{F}\left(-\xi_{it}|\boldsymbol{x}_i\right)-\mathbf{1}\{u_{it}\leq-\xi_{it}\}$ *and* $\boldsymbol{\Psi}(\boldsymbol{X})$ *is the* $n\times T$ *matrix of* $\psi_{it}(\boldsymbol{x}_i)$*.*

Proof. The proof is similar to the proof of Lemma 3. Therefore, it is omitted to save space. □

Proof of Theorem 3:

Proof. By the proof of Theorem 1, for any $\boldsymbol{x}\in\mathcal{X}$,

$$
\hat{\boldsymbol{g}}(\boldsymbol{x}) = (\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x}) + (\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)'(\boldsymbol{B}_0'\boldsymbol{\phi}_{k_n}(\boldsymbol{x})-\boldsymbol{g}(\boldsymbol{x})) + (\hat{\boldsymbol{B}}-\boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T))'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}).
$$

Moreover,

$$
\hat{\boldsymbol{B}}-\boldsymbol{B}_0(\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T) = (\hat{\boldsymbol{A}}-\boldsymbol{A}_0)\boldsymbol{F}_0\hat{\boldsymbol{H}}/T + (\hat{\boldsymbol{A}}-\boldsymbol{A}_0)(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})/T.
$$

Thus, by Lemma 1 and Theorem 1,

$$
\hat{\boldsymbol{g}}(\boldsymbol{x}) - (\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x}) = \hat{\boldsymbol{H}}'\boldsymbol{F}_0'(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)'\boldsymbol{\phi}_{k_n}(\boldsymbol{x})/T + O_P(k_n^{-\alpha}) + O_P(\varepsilon_{nT}^2\sqrt{k_n}).
$$

It then follows from Lemma 3 that

$$\hat{\boldsymbol{g}}(\boldsymbol{x}) - (\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x}) = \hat{\boldsymbol{H}}'\boldsymbol{F}_0'\boldsymbol{\Psi}'(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x})/(nT) + O_P(k_n^{1/2-\alpha}) + O_P(\sqrt{k_n}\eta_{nT}).$$

Define $\boldsymbol{d}_T(\boldsymbol{x}_i) = T^{-1}\sum_{t=1}^T \boldsymbol{f}_{0t}\psi_{it}(\boldsymbol{x}_i)$, $q(\boldsymbol{x}_i) = \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x})$, then we can write

$$\boldsymbol{F}_0'\boldsymbol{\Psi}'(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x})/(nT) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{d}_T(\boldsymbol{x}_i)q(\boldsymbol{x}_i).$$

Note that $\mathbb{E}[\boldsymbol{d}_T(\boldsymbol{x}_i)q(\boldsymbol{x}_i)] = 0$ because $\mathbb{E}[\boldsymbol{d}_T(\boldsymbol{x}_i)|\boldsymbol{x}_i] = 0$, and it is easy to show that

$$\begin{aligned}\mathbb{E}[\boldsymbol{d}_T(\boldsymbol{x}_i)\boldsymbol{d}_T(\boldsymbol{x}_i)'q^2(\boldsymbol{x}_i)] &= \tau(1-\tau)(\boldsymbol{F}_0'\boldsymbol{F}_0/T^2)\boldsymbol{\phi}_{k_n}'(\boldsymbol{x})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Sigma}_\phi\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}) + o(1)\\ &= \tau(1-\tau)(\boldsymbol{F}_0'\boldsymbol{F}_0/T^2)\sigma_{k_n}^2 + o(1).\end{aligned}$$

Thus, we have

$$\begin{aligned}\boldsymbol{\Sigma}_{T,\tau}^{-1/2}(\hat{\boldsymbol{H}}')^{-1}\cdot\frac{\sqrt{nT}}{\sigma_{k_n}}\Big(\hat{\boldsymbol{g}}(\boldsymbol{x}) - (\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x})\Big) &= \boldsymbol{\Sigma}_{T,\tau}^{-1/2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^n\sqrt{T}\boldsymbol{d}_T(\boldsymbol{x}_i)q(\boldsymbol{x}_i)/\sigma_{k_n}\\ &\quad + O_P(k_n^{1/2-\alpha} + \sqrt{k_n}\eta_{nT})\sqrt{nT}\sigma_{k_n}^{-1}. \quad \text{(A.22)}\end{aligned}$$

Finally, it follows from Lyapunov's CLT and Assumption 4(iv) that

$$\boldsymbol{\Sigma}_{T,\tau}^{-1/2}(\hat{\boldsymbol{H}}')^{-1}\cdot\frac{\sqrt{nT}}{\sigma_{k_n}}\Big(\hat{\boldsymbol{g}}(\boldsymbol{x}) - (\boldsymbol{F}_0'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x})\Big) \xrightarrow{d} N(0, \boldsymbol{I}_R).$$

This completes the proof. □

Proof of Theorem 4:

Proof. Define $\boldsymbol{R}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}_0 - \boldsymbol{G}(\boldsymbol{X})$, we can write

$$\hat{\boldsymbol{Y}} = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 + \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0) = \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0' + \boldsymbol{R}(\boldsymbol{X})\boldsymbol{F}_0' + \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0).$$

Thus,

$$\begin{aligned}\tilde{\boldsymbol{F}}_0 &= \hat{\boldsymbol{Y}}'\hat{\boldsymbol{G}}(\boldsymbol{X})\cdot(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X}))^{-1} = \boldsymbol{F}_0(\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1}\\ &+ \boldsymbol{F}_0(\boldsymbol{R}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1} + (\hat{\boldsymbol{A}} - \boldsymbol{A}_0)'(\boldsymbol{\Phi}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1},\end{aligned}$$

and

$$
\begin{aligned}
\tilde{\boldsymbol{f}}_t - \tilde{\boldsymbol{H}}'\boldsymbol{f}_{0t} = (\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1}(\hat{\boldsymbol{G}}(\boldsymbol{X})'\boldsymbol{R}(\boldsymbol{X})/n)\boldsymbol{f}_{0t}& \\
+ (\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1}(\hat{\boldsymbol{G}}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})/n)(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}).&
\end{aligned}
$$

It is easy to see from Theorem 1 and Assumption 1(ii) that the first term on the right-hand side of the above equation is $O_P(k_n^{-\alpha})$. Moreover, by Lemma 3, the second term can be written as

$$
(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1} \cdot (\hat{\boldsymbol{G}}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})/n) \cdot \boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\psi_{it}(\boldsymbol{x}_i) + O_P(k_n^{-\alpha}) + O_P(\eta_{nT}).
$$

By Theorem 1 we can show that

$$
\|(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1} - \hat{\boldsymbol{H}}'\boldsymbol{\Sigma}_g^{-1}\hat{\boldsymbol{H}}\| = O_P(\varepsilon_{nT}),
$$

$$
\|(\hat{\boldsymbol{G}}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})/n) - \hat{\boldsymbol{H}}^{-1}\mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']\|_S = O_P(\varepsilon_{nT}),
$$

$$
\left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\psi_{it}(\boldsymbol{x}_i) \right\| = O_P(\sqrt{k_n/n}),
$$

it then follows from Assumption 4(iii) that

$$
\begin{aligned}
(\hat{\boldsymbol{H}}')^{-1}\sqrt{n}(\tilde{\boldsymbol{f}}_t - \tilde{\boldsymbol{H}}'\boldsymbol{f}_{0t}) = \boldsymbol{\Sigma}_g^{-1}\mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\psi_{it}(\boldsymbol{x}_i)\right)& \\
+ O_P(\varepsilon_{nT}k_n^{1/2}) + O_P(n^{1/2}k_n^{-\alpha}) + O_P(n^{1/2}\eta_{nT}).&
\end{aligned}
$$

By the Lyapunov's CLT we can show that

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\psi_{it}(\boldsymbol{x}_i) \xrightarrow{d} N(0, \tau(1-\tau)\boldsymbol{\Sigma}_\phi),
$$

then the desired result follows from Assumption 5. □

Lemma 5. *Denote the eigenvalues of $\hat{\boldsymbol{\Sigma}}_g \cdot \boldsymbol{F}_0'\boldsymbol{F}_0/T$ as $\rho_1, \ldots, \rho_R$. Then, for $j = 1, \ldots, R$, it holds that $\hat{\rho}_j = \rho_j + O_P(\varepsilon_{nT})$.*

Proof. Denote the estimation error of using $\hat{\boldsymbol{Y}}$ to estimate $\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'$ as $\hat{\boldsymbol{V}} = \hat{\boldsymbol{Y}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'$. From Lemma 2, it holds that $(nT)^{1/2}||\hat{\boldsymbol{V}}|| = O_P(\varepsilon_{nT})$. Define $\boldsymbol{F}^* = \boldsymbol{F} + \hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})(\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X}))^{-1}$. For any matrix $\boldsymbol{A}$, let $\varrho_j(A)$ be the j-th largest eigenvalue of $\boldsymbol{A}$. To complete the proof, it suffices to show the following two statements:

1. $\hat{\rho}_j = \varrho_j(\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}/nT) + O_P(\varepsilon_{nT}^2)$.

2. $\varrho_j(\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}/nT) = \rho_j + O_P(\varepsilon_{nT})$.

First, we can rewrite $\hat{\boldsymbol{Y}} = \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'} + \boldsymbol{Q}(\boldsymbol{G})\hat{\boldsymbol{V}}$ where $\boldsymbol{Q}(\boldsymbol{G}) = (\boldsymbol{I} - \boldsymbol{G}(\boldsymbol{X})(\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X}))^{-1}\boldsymbol{G}(\boldsymbol{X})')$. This implies that

$$\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT} = \frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT} + \frac{\hat{\boldsymbol{V}}'\boldsymbol{Q}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT}.$$

By Weyl inequality, for every $k = 1, \ldots, R$, we have that

$$\sum_{j=1}^{k} \varrho_j\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \leq \sum_{j=1}^{k} \varrho_j\left(\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT}\right) + k \cdot \varrho_1\left(\frac{\hat{\boldsymbol{V}}'\boldsymbol{Q}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT}\right),$$

where

$$k \cdot \varrho_1\left(\frac{\hat{\boldsymbol{V}}'\boldsymbol{Q}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT}\right) \leq k \cdot \varrho_1\left(\frac{\hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}}{nT}\right) \leq \left\|\frac{\hat{\boldsymbol{V}}}{\sqrt{nT}}\right\|^2 = O_P(\varepsilon_{nT}^2),$$

which means that

$$\sum_{j=1}^{k} \varrho_j\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \leq \sum_{j=1}^{k} \varrho_j\left(\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT}\right) + O_P(\varepsilon_{nT}^2). \tag{A.23}$$

Additionally, for $k = 1, \ldots, R$, suppose Ξ_k^* is the matrix of eigenvectors corresponding to the k largest eigenvalues of $\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}/nT$, such that $\Xi_k^{*'}\Xi_k^*/T = I_k$. Then, for any $k = 1, \ldots, R$, it holds that

$$\sum_{j=1}^{k} \varrho_j\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \geq \text{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT^2}\Xi_k^*\right) = \text{trace}\left(\Xi_k^{*'}\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT^2}\Xi_k^*\right) + \text{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{V}}'\boldsymbol{Q}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right),$$

where

$$\text{trace}\left(\Xi_k^{*'}\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT^2}\Xi_k^*\right) = \sum_{j=1}^{k} \varrho_j\left(\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT}\right),$$

and

$$\begin{aligned}
\text{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{V}}'\boldsymbol{Q}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right) =& \text{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right) - \text{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{V}}'(\boldsymbol{G}(\boldsymbol{X})(\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X}))^{-1}\boldsymbol{G}(\boldsymbol{X})')\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right) \\
\leq& \left\|\frac{\Xi_k^*}{\sqrt{T}}\right\|^2\left\|\frac{\hat{\boldsymbol{V}}}{\sqrt{nT}}\right\|^2 + \left\|\frac{\Xi_k^*}{\sqrt{T}}\right\|^2\left\|\frac{\hat{\boldsymbol{V}}}{\sqrt{nT}}\right\|^2\left\|\boldsymbol{G}(\boldsymbol{X})(\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X}))^{-1}\boldsymbol{G}(\boldsymbol{X})'\right\|^2 \\
=& O_P(\varepsilon_{nT}^2)
\end{aligned}$$

Hence,

$$\sum_{j=1}^{k} \varrho_j\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \geq \sum_{j=1}^{k} \varrho_j\left(\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT}\right) + O_P(\varepsilon_{nT}^2). \tag{A.24}$$

Therefore, (A.23) and (A.24) complete the proof of statement 1.

Next, to prove statement 2, let us denote $\boldsymbol{P}(\boldsymbol{G}) = \boldsymbol{G}(\boldsymbol{X})(\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X}))^{-1}\boldsymbol{G}(\boldsymbol{X})'$, and rewrite

$$\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'} = \boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{'} + \hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{'} + \boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}} + \hat{\boldsymbol{V}}'\boldsymbol{P}(\boldsymbol{G})\hat{\boldsymbol{V}}.$$

The proof proceeds as follows. First, define Ξ_k as the matrix of eigenvectors corresponding the k largest eigenvalues of $\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{'}/nT$ such that $\Xi_k'\Xi_k/T = I_k$. Then, for any $k = 1, \ldots, R$ we have that

$$\begin{aligned}\sum_{j=1}^{k} \varrho_j\left(\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT}\right) =& \operatorname{trace}\left(\Xi_k^{*'}\frac{\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{'}}{nT^2}\Xi_k^*\right) + \\ & 2\operatorname{trace}\left(\Xi_k^{*'}\frac{\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right) + \\ & \operatorname{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{V}}'\boldsymbol{P}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right),\end{aligned}$$

where

$$\operatorname{trace}\left(\Xi_k^{*'}\frac{\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{'}}{nT^2}\Xi_k^*\right) \leq \operatorname{trace}\left(\Xi_k^{'}\frac{\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{'}}{nT^2}\Xi_k\right) = \sum_{j=1}^{k}\rho_j,$$

$$\operatorname{trace}\left(\Xi_k^{*'}\frac{\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right) \leq \left\|\frac{\Xi_k^*}{\sqrt{T}}\right\|^2\left\|\frac{F}{\sqrt{T}}\right\|\,\left\|\frac{\boldsymbol{G}(\boldsymbol{X})}{\sqrt{n}}\right\|\,\left\|\frac{\hat{\boldsymbol{V}}}{\sqrt{nT}}\right\| = O_P(\varepsilon_{nT}),$$

and

$$\operatorname{trace}\left(\Xi_k^{*'}\frac{\hat{\boldsymbol{V}}'\boldsymbol{P}(\boldsymbol{G})\hat{\boldsymbol{V}}}{nT^2}\Xi_k^*\right) \leq \left\|\frac{\Xi_k^*}{\sqrt{T}}\right\|^2\left\|\frac{\hat{\boldsymbol{V}}}{\sqrt{nT}}\right\|^2\left\|\boldsymbol{G}(\boldsymbol{X})(\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X}))^{-1}\boldsymbol{G}(\boldsymbol{X})'\right\| = O_P(\varepsilon_{nT}^2).$$

Combining these inequalities, we get

$$\sum_{j=1}^{k} \varrho_j\left(\frac{\boldsymbol{F}^*\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^{*'}}{nT}\right) \leq \sum_{j=1}^{k}\rho_j + O_P(\varepsilon_{nT}). \tag{A.25}$$

Second, using similar reasoning, we have that

$$
\begin{aligned}
\sum_{j=1}^{k} \varrho_j \left(\frac{\boldsymbol{F}^* \boldsymbol{G}(\boldsymbol{X})' \boldsymbol{G}(\boldsymbol{X}) \boldsymbol{F}^{*'}}{nT} \right) \geq & \text{trace} \left(\Xi_k' \frac{\boldsymbol{F}^* \boldsymbol{G}(\boldsymbol{X})' \boldsymbol{G}(\boldsymbol{X}) \boldsymbol{F}^{*'}}{nT^2} \Xi_k \right) \\
= & \text{trace} \left(\Xi_k' \frac{\boldsymbol{F} \boldsymbol{G}(\boldsymbol{X})' \boldsymbol{G}(\boldsymbol{X}) \boldsymbol{F}'}{nT^2} \Xi_k \right) + 2\text{trace} \left(\Xi_k' \frac{\boldsymbol{F} \boldsymbol{G}(\boldsymbol{X})' \hat{\boldsymbol{V}}}{nT^2} \Xi_k \right) \\
& + \text{trace} \left(\Xi_k' \frac{\hat{\boldsymbol{V}}' \boldsymbol{P}(\boldsymbol{G}) \hat{\boldsymbol{V}}}{nT^2} \Xi_k \right) \\
= & \sum_{j=1}^{k} \rho_j + O_P(\varepsilon_{nT}),
\end{aligned} \tag{A.26}
$$

In this fashion, statement 2 is also proven by (A.25) and (A.26), so that the proof of Lemma 5 is completed. □

Lemma 6. *Define* $\tilde{\psi}_{it} = \tau - \mathbf{1}\{u_{it} \leq 0\}$, $\tilde{\boldsymbol{\Psi}}$ *as* $n \times T$ *matrix of* $\tilde{\psi}_{it}$, *and let* $\tilde{s}_t = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x_i}) \tilde{\psi}_{it}$. *Then, for* $j = 1, 2, \ldots, [b \min(T, Dk_n)]$ *where* $b \in (0, 1)$, *there exist two strictly positive constants* $\underline{c}^s$ *and* $\bar{c}^s$, *such that*

1. *When* $Dk_n/T \to c \in [0, \infty)$, $\underline{c}^s + o_P(1) < \varrho_j \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right) \leq \bar{c}^s + o_P(1)$
2. *When* $Dk_n/T \to \infty$, $\underline{c}^s + o_P(1) \leq \varrho_j \left(\frac{1}{Dk_n} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right) \leq \bar{c}^s + o_P(1)$

Proof. Note that $\mathbb{E}[(\tau - \mathbf{1}\{u_{it} \leq 0\})^2 | \boldsymbol{X}] = \tau(\tau - 1)^2 + (1 - \tau)\tau^2 = \tau - \tau^2$ for each $i = 1, \ldots, n$ and, by Assumption 6(i), $t = 1, \ldots, T$, and $\mathbb{E}[(\tau - \mathbf{1}\{u_{it} \leq 0\})(\tau - \mathbf{1}\{u_{jt} \leq 0\}) | \boldsymbol{X}] = 0$ for each $i \neq j$ and $t = 1, \ldots, T$. Then, we have

$$
\mathbb{E}[\tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' | \boldsymbol{X}] = \frac{1}{n} \sum_{i=1}^{n} [\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)' \mathbb{E}[(\tau - \mathbf{1}\{u_{it} \leq 0)^2 | \boldsymbol{X}]] = (\tau - \tau^2) \hat{\boldsymbol{\Sigma}}_{\phi}.
$$

Define the isotropic vector $\bar{\boldsymbol{s}}_t = ((\tau - \tau^2) \hat{\boldsymbol{\Sigma}}_{\phi})^{-1/2} \tilde{\boldsymbol{s}}_t$. Then, we have that

$$
\mathbb{E}[\bar{\boldsymbol{s}}_t | \boldsymbol{X}] = 0, \quad \mathbb{E}[\bar{\boldsymbol{s}}_t \bar{\boldsymbol{s}}_t' | \boldsymbol{X}] = \boldsymbol{I}_{Dk_n}.
$$

By Assumption 6(i), $\bar{\boldsymbol{s}}_t$ is i.i.d. conditional on $\boldsymbol{X}$. Rewrite

$$
\frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' = ((\tau - \tau^2) \hat{\boldsymbol{\Sigma}}_{\phi})^{1/2} \cdot \frac{1}{T} \sum_{t=1}^{T} \bar{\boldsymbol{s}}_t \bar{\boldsymbol{s}}_t' \cdot ((\tau - \tau^2) \hat{\boldsymbol{\Sigma}}_{\phi})^{1/2} \cdot .
$$

Then for any $j = 1, \ldots, \min(Dk_n, T)$, it holds that

$$(\tau - \tau^2)\lambda_{min}(\hat{\mathbf{\Sigma}}_\phi)\varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right) \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t'\right) \leq (\tau - \tau^2)\lambda_{max}(\hat{\mathbf{\Sigma}}_\phi)\varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right).$$

Further, note that $\hat{\mathbf{\Sigma}}_\phi = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'$ and each entry of $\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ is bounded above. It follows from Theorem 5.6.1 of Vershynin (2018) that

$$\|\hat{\mathbf{\Sigma}}_\phi - \mathbf{\Sigma}_\phi\|_S = O_P\left(\sqrt{\frac{k_n \log k_n}{n}} + \frac{k_n \log k_n}{n}\right) = o_P(1).$$

Consequently, $|\lambda_{min}(\hat{\mathbf{\Sigma}}_\phi) - \lambda_{min}(\mathbf{\Sigma}_\phi)| = o_P(1)$ and $|\lambda_{max}(\hat{\mathbf{\Sigma}}_\phi) - \lambda_{max}(\mathbf{\Sigma}_\phi)| = o_P(1)$ and it holds with probability approaching 1 that

$$(\tau - \tau^2)\lambda_{min}(\mathbf{\Sigma}_\phi)\varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right) \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t'\right) \leq (\tau - \tau^2)\lambda_{max}(\mathbf{\Sigma}_\phi)\varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right). \tag{A.27}$$

Note that $\bar{\boldsymbol{s}}_t = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}((\tau - \tau^2)\hat{\mathbf{\Sigma}}_\phi)^{-1/2}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\tilde{\psi}_{it}$ is a sum of independent random vector conditional on $\boldsymbol{X}$. Define them as $\bar{\boldsymbol{s}}_{it} = ((\tau - \tau^2)\hat{\mathbf{\Sigma}}_\phi)^{-1/2}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\tilde{\psi}_{it}$. Then, we have that

$$\mathbb{E}[\bar{\boldsymbol{s}}_{it}|\boldsymbol{X}] = 0, \quad \sum_{i=1}^{n}\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}'|\boldsymbol{X}] = ((\tau - \tau^2)\hat{\mathbf{\Sigma}}_\phi)^{-1}\sum_{i=1}^{n}(\tau - \tau^2)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)' = n\boldsymbol{I}_{Dk_n}.$$

Therefore, $||\bar{\boldsymbol{s}}_{it}|| \leq ||((\tau - \tau^2)\hat{\mathbf{\Sigma}}_\phi)^{-1/2}||_S \cdot ||\boldsymbol{\phi}(\boldsymbol{x}_i)|| \cdot |\tilde{\psi}_{it}| = O_P(\sqrt{k_n})$. Let ν_t and γ be the probability measures of $\bar{\boldsymbol{s}}_t$ and of a multivariate Gaussian $\mathcal{N}(0, \boldsymbol{I}_{Dk_n})$, respectively. Moreover, suppose $\Pi(\nu_t, \gamma)$ is the collection of all joint distributions π on $\mathbb{R}^{Dk_n} \times \mathbb{R}^{Dk_n}$ whose first marginal is ν_t and second marginal is γ, namely, $\pi(B, \mathbb{R}^{Dk_n}) = \nu_t(B)$ and $\pi(\mathbb{R}^{Dk_n}, B) = \gamma(B)$ for all Borel sets B in $\mathbb{R}^{Dk_n}$.

Next, define the Wasserstein distance between two probability measures

$$W_2(\nu, \gamma) = \left(\inf_{\pi \in \Pi(\nu,\gamma)}\int_{\mathbb{R}^{Dk_n}\times\mathbb{R}^{Dk_n}} ||y - x||^2\pi(dx, dy)\right)^{1/2}$$

Then, it follows from Theorem 12 in Bonis (2020) that there exists a constant $C > 0$ such that for each $t = 1, \ldots, T$,

$$\begin{aligned} C \cdot W_2(\nu_t, \gamma) \leq& \frac{(\sum_{i=1}^{n}\mathbb{E}[||\bar{\boldsymbol{s}}_{it}||^4])^{1/3}}{n^{2/3}} + \frac{(\sum_{i=1}^{n}\mathbb{E}[||\bar{\boldsymbol{s}}_{it}||^4])^{1/2}}{n} \\ &+ \frac{(\sum_{i=1}^{n}||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}']||^2)^{1/2}}{n} + \left(\frac{M(0)M(2)}{n}\right)^{1/2}, \end{aligned}$$

where, for $l \in [0,2]$,

$$M(l) = \left(\frac{1}{n} \sum_{i=1}^{n} ||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}']||^2 \mathbb{E}[||\bar{\boldsymbol{s}}_{it}||^l]^2 + ||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}'||\bar{\boldsymbol{s}}_{it}||^l]||^2 \right)^{1/2}.$$

Therefore,

$$\begin{aligned}
M(0) &= \left(\frac{1}{n} \sum_{i=1}^{n} 2||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}']||^2 \right)^{1/2} \leq (O(2k_n^2))^{1/2} = O(k_n), \\
M(2) &= \left(\frac{1}{n} \sum_{i=1}^{n} ||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}']||^2 \mathbb{E}[||\bar{\boldsymbol{s}}_{it}||^2]^2 + ||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}'||\bar{\boldsymbol{s}}_{it}||^2]||^2 \right)^{1/2} \leq (O(2k_n^4))^{1/2} = O(k_n^2).
\end{aligned}$$

We next analyze the following four terms:

$$\begin{aligned}
\frac{(\sum_{i=1}^n \mathbb{E}[||\bar{\boldsymbol{s}}_{it}||^4])^{1/3}}{n^{2/3}} &\leq \frac{(n \cdot O(k_n^2))^{1/3}}{n^{2/3}} = O\left(\frac{k_n^{2/3}}{n^{1/3}}\right) = o(1), \\
\frac{(\sum_{i=1}^n \mathbb{E}[||\bar{\boldsymbol{s}}_{it}||^4])^{1/2}}{n} &\leq \frac{(n \cdot O(k_n^2))^{1/2}}{n} = O\left(\frac{k_n}{n^{1/2}}\right) = o(1), \\
\frac{(\sum_{i=1}^n ||\mathbb{E}[\bar{\boldsymbol{s}}_{it}\bar{\boldsymbol{s}}_{it}']||^2)^{1/2}}{n} &\leq \frac{(n \cdot O(k_n^2))^{1/2}}{n} = O\left(\frac{k_n}{n^{1/2}}\right) = o(1), \\
\left(\frac{M(0)M(2)}{n}\right)^{1/2} &\leq \left(\frac{O(k_n)O(k_n^2)}{n}\right)^{1/2} = O\left(\frac{k_n^{3/2}}{n^{1/2}}\right) = o(1).
\end{aligned}$$

Suppose that $\delta_n = \frac{k_n^{2/3}}{n^{1/3}} + \frac{k_n}{n^{1/2}} + \frac{k_n^{3/2}}{n^{1/2}}$. Then, there exists $\bar{\bar{\boldsymbol{s}}}_t \overset{d}{=} \bar{\boldsymbol{s}}_t$ and multivariate Gaussian random vectors $\boldsymbol{d}_t \sim \mathcal{N}(0, \boldsymbol{I}_{Dk_n})$ such that $(\bar{\bar{\boldsymbol{s}}}_t, \boldsymbol{d}_t)$ are mutually independent across t, and

$$\mathbb{E}[||\bar{\bar{\boldsymbol{s}}}_t - \boldsymbol{d}_t||] \leq W_2(\nu_t, \gamma) = O(\delta_n) = o(1).$$

Define $\bar{\bar{\boldsymbol{S}}} = [\bar{\bar{\boldsymbol{s}}}_1, \bar{\bar{\boldsymbol{s}}}_2, \ldots, \bar{\bar{\boldsymbol{s}}}_T]$, and $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_T]$. Since $\boldsymbol{d}_t \sim \mathcal{N}(0, \boldsymbol{I}_{k_n})$ and $\{\boldsymbol{d}_t\}_{t=1}^T$ are mutually independent, $\boldsymbol{D}$ consists of real independent Gaussian random variables. In addition, noting that $||\bar{\bar{\boldsymbol{S}}} - \boldsymbol{D}||^2 \leq \sum_{t=1}^T ||\bar{\bar{\boldsymbol{s}}}_t - \boldsymbol{d}_t||^2 = O_P(T\delta_n^2)$, and $\frac{1}{T}\sum_{t=1}^T \bar{\bar{\boldsymbol{s}}}_t \bar{\bar{\boldsymbol{s}}}_t' = \frac{1}{T}\bar{\bar{\boldsymbol{S}}}\bar{\bar{\boldsymbol{S}}}'$, Weyl's inequality implies that, for $j = 1, \ldots, \min(T, Dk_n)$, the following result holds

$$\left| \varrho_j\left(\frac{1}{T}\bar{\bar{\boldsymbol{S}}}\bar{\bar{\boldsymbol{S}}}'\right) - \varrho_j\left(\frac{1}{T}\boldsymbol{D}\boldsymbol{D}'\right) \right| \leq 2\frac{||\boldsymbol{D}||_S}{\sqrt{T}} \cdot \frac{||\bar{\bar{\boldsymbol{S}}} - \boldsymbol{D}||}{\sqrt{T}} + \frac{||\bar{\bar{\boldsymbol{S}}} - \boldsymbol{D}||^2}{T} = \frac{||\boldsymbol{D}||_S}{\sqrt{T}} \cdot O_P(\delta_n) + O_P(\delta_n^2). \tag{A.28}$$

(i) When $Dk_n/T \to 0$, note that $\boldsymbol{d}_t$ is i.i.d standard normal and thus sub-Gaussian. Therefore, by Theorem 4.7.1 in Vershynin (2018), there exists a bounded constant C such that, for each

$u > 0$,

$$\Pr\left[\left|\left|\frac{1}{T}\boldsymbol{D}\boldsymbol{D}' - \boldsymbol{I}_{Dk_n}\right|\right|_S \leq C\left(\sqrt{\frac{Dk_n + u}{T}} + \frac{Dk_n + u}{T}\right)||\boldsymbol{I}_{Dk_n}||_S\right] \geq 1 - 2\exp(-u).$$

Moreover, since u is arbitrarily chosen and $||\boldsymbol{I}_{Dk_n}||_S = 1$, when $Dk_n/T \to 0$, we have that

$$\left|\left|\frac{1}{T}\boldsymbol{D}\boldsymbol{D}' - \boldsymbol{I}_{Dk_n}\right|\right|_S = o_P(1) \Rightarrow \frac{1}{T}\lambda_{max}(\boldsymbol{D}\boldsymbol{D}') \leq 1 + \left|\left|\frac{1}{T}\boldsymbol{D}\boldsymbol{D}' - \boldsymbol{I}_{Dk_n}\right|\right|_S = 1 + o_P(1)$$
$$\Rightarrow \frac{||\boldsymbol{D}||_S}{\sqrt{T}} = \sqrt{\frac{1}{T}\lambda_{max}(\boldsymbol{D}\boldsymbol{D}')} = O_P(1).$$

which implies that the left-hand side of (A.28) is $o_P(1)$, so that $\varrho_j\left(\frac{1}{T}\sum_{t=1}^T \bar{\bar{\boldsymbol{s}}}_t\bar{\bar{\boldsymbol{s}}}_t'\right) \xrightarrow{P} 1$ for $j = 1, 2, \ldots, Dk_n$. Since $\bar{\bar{\boldsymbol{s}}}_t \overset{d}{=} \bar{\boldsymbol{s}}_t$, we have $\varrho_j\left(\frac{1}{T}\sum_{t=1}^T \bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right) \xrightarrow{P} 1$. Then it follows from (A.27) that with probability approaching 1, for $j = 1, 2, \ldots, Dk_n$,

$$(\tau - \tau^2)\lambda_{min}(\boldsymbol{\Sigma}_\phi) \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^T \tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t'\right) \leq (\tau - \tau^2)\lambda_{max}(\boldsymbol{\Sigma}_\phi).$$

(ii) When $Dk_n/T \to c \in (0, 1]$, on one hand, it follows from Theorem II.13 in Davidson and Szarek (2001) that for any $u > 0$,

$$\Pr\left(\sqrt{\varrho_1\left(\frac{1}{T}\boldsymbol{D}\boldsymbol{D}'\right)} \geq 1 + \sqrt{\frac{Dk_n}{T}} + u\right) < \exp\left(-\frac{Tu^2}{2}\right). \tag{A.29}$$

Since u is arbitrary and in this case, $T \to \infty$, $\frac{||\boldsymbol{D}||_S}{\sqrt{T}} = O_P(1)$ so that (A.28) is still $o_P(1)$.

On the other hand, note that $\boldsymbol{D}$ is a $Dk_n \times T$ matrix. Choose a real number $b \in (0, 1)$ and let $\boldsymbol{D}_{[bDk_n]}$ be the $[bDk_n] \times T$ row submatrix consisting of the first $[bDk_n]$ rows of $\boldsymbol{D}$. Then $\boldsymbol{D}_{[bDk_n]}\boldsymbol{D}_{[bDk_n]}'$ is the $[bDk_n] \times [bDk_n]$ principal submatrix of $\boldsymbol{D}\boldsymbol{D}'$. Due to the Sturmian Separation theorem (see page 64 of Rao et al. (1973)), it holds that

$$\varrho_{[bDk_n]}(\boldsymbol{D}_{[bDk_n]}\boldsymbol{D}_{[bDk_n]}') \leq \varrho_{[bDk_n]}(\boldsymbol{D}\boldsymbol{D}').$$

Further, it follows from Theorem II.13 in Davidson and Szarek (2001) that for any $u > 0$,

$$\Pr\left(\sqrt{\varrho_{[bDk_n]}\left(\frac{1}{T}\boldsymbol{D}_{[bDk_n]}\boldsymbol{D}_{[bDk_n]}'\right)} \leq 1 - \sqrt{\frac{[bDk_n]}{T}} - u\right) < \exp\left(-\frac{Tu^2}{2}\right),$$

which means that

$$\Pr\left(\sqrt{\varrho_{[bDk_n]}\left(\frac{1}{T}\boldsymbol{D}\boldsymbol{D}'\right)} \leq 1-\sqrt{\frac{[bDk_n]}{T}}-u\right) < \exp\left(-\frac{Tu^2}{2}\right). \tag{A.30}$$

Note that $\frac{[bDk_n]}{T} < b \cdot \frac{Dk_n}{T} \leq b < 1$ so that $1-\sqrt{\frac{[bDk_n]}{T}} > 0$. Consequently, when $Dk_n/T \to c \in (0,1]$, it is verified by (A.28) with probability approaching 1 that, for $j = 1, 2, \dots, Dk_n$,

$$0 < \left(1-\sqrt{\frac{[bDk_n]}{T}}\right)^2 \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\bar{\boldsymbol{s}}}_t\bar{\bar{\boldsymbol{s}}}_t'\right) \leq \left(1+\sqrt{\frac{Dk_n}{T}}\right)^2 < \infty.$$

Since $\bar{\bar{\boldsymbol{s}}}_t \stackrel{d}{=} \bar{\boldsymbol{s}}_t$, the above equation also holds for $\varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right)$. Then, it holds with probability approaching 1 that, for $j = 1, 2, \dots, [bDk_n]$,

$$0 < (\tau-\tau^2)\lambda_{min}(\boldsymbol{\Sigma}_\phi)\left(1-\sqrt{\frac{[bDk_n]}{T}}\right)^2 \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t'\right) \leq (\tau-\tau^2)\lambda_{max}(\boldsymbol{\Sigma}_\phi)\left(1+\sqrt{\frac{Dk_n}{T}}\right)^2 < \infty.$$

(iii) When $Dk_n/T \to c \in (1,\infty)$, $Dk_n > T$ and $\boldsymbol{D}\boldsymbol{D}'$ has at most T nonzero eigenvalues. Thus we focus on $\boldsymbol{D}'\boldsymbol{D}$. Similarly, it follows from Theorem II.13 in Davidson and Szarek (2001) that for any $u > 0$,

$$\Pr\left(\sqrt{\varrho_1\left(\frac{1}{Dk_n}\boldsymbol{D}'\boldsymbol{D}\right)} \geq 1+\sqrt{\frac{T}{Dk_n}}+\sqrt{\frac{T}{Dk_n}}u\right) < \exp\left(-\frac{Tu^2}{2}\right),$$
$$\Pr\left(\sqrt{\varrho_T\left(\frac{1}{Dk_n}\boldsymbol{D}'\boldsymbol{D}\right)} \leq 1-\sqrt{\frac{T}{Dk_n}}-\sqrt{\frac{T}{Dk_n}}u\right) < \exp\left(-\frac{Tu^2}{2}\right).$$

Note that $\boldsymbol{D}\boldsymbol{D}'$ and $\boldsymbol{D}'\boldsymbol{D}$ have the same non-zero eigenvalues. Multiplying $\frac{1}{Dk_n}\boldsymbol{D}'\boldsymbol{D}$ by $\frac{Dk_n}{T}$, we have that

$$\Pr\left(\sqrt{\varrho_1\left(\frac{1}{T}\boldsymbol{D}\boldsymbol{D}'\right)} \geq \sqrt{\frac{Dk_n}{T}}+1+u\right) < \exp\left(-\frac{Tu^2}{2}\right),$$
$$\Pr\left(\sqrt{\varrho_T\left(\frac{1}{T}\boldsymbol{D}\boldsymbol{D}'\right)} \leq \sqrt{\frac{Dk_n}{T}}-1-u\right) < \exp\left(-\frac{Tu^2}{2}\right).$$

Following the same analysis as in (ii), it holds with probability approaching 1 that, for $j = 1, 2, \dots, T$,

$$0 < (\tau-\tau^2)\lambda_{min}(\boldsymbol{\Sigma}_\phi)\left(\sqrt{\frac{Dk_n}{T}}-1\right)^2 \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t'\right) \leq (\tau-\tau^2)\lambda_{max}(\boldsymbol{\Sigma}_\phi)\left(1+\sqrt{\frac{Dk_n}{T}}\right)^2 < \infty.$$

Thus, the above proofs of cases (i), (ii) and (iii) prove statement 1.

(iv) When $Dk_n/T \to \infty$, We focus on $\boldsymbol{D}'\boldsymbol{D}$. Note that $\boldsymbol{D}$ has *i.i.d.* sub-gaussian entries. So we can apply again Theorem 4.7.1 in Vershynin (2018) to show that there exists a bounded constant C such that, for any $u > 0$

$$\Pr\left[\left|\left|\frac{1}{Dk_n}\boldsymbol{D}'\boldsymbol{D} - \boldsymbol{I}_T\right|\right|_S \leq C\left(\sqrt{\frac{T+u}{Dk_n}} + \frac{T+u}{Dk_n}\right)||\boldsymbol{I}_T||_S\right] \geq 1 - 2\exp(-u),$$

where $||A||_S$ is the spectral norm of a generic square matrix A. Then, $\varrho_j\left(\frac{1}{Dk_n}\boldsymbol{D}'\boldsymbol{D}\right) \xrightarrow{P} 1$ for $j = 1, \ldots, T$, because u is arbitrary. Following a similar analysis to that used in the case where $Dk_n/T \to 0$, it holds that $\varrho_j\left(\frac{1}{Dk_n}\sum_{t=1}^T \bar{\boldsymbol{s}}_t\bar{\boldsymbol{s}}_t'\right) \xrightarrow{P} 1$, and, with probability approaching 1, for $j = 1, 2, \ldots, T$,

$$(\tau - \tau^2)\lambda_{min}(\boldsymbol{\Sigma}_\phi) \leq \varrho_j\left(\frac{1}{Dk_n}\sum_{t=1}^T \tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t'\right) \leq (\tau - \tau^2)\lambda_{max}(\boldsymbol{\Sigma}_\phi),$$

which completes the proof of statement 2. $\square$

Lemma 7. *Let* $\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi} = \frac{1}{n}\sum_{i=1}^n \mathsf{f}(0|\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x_i})\boldsymbol{\phi}_{k_n}(\boldsymbol{x_i})'$, $\boldsymbol{A}_1 = \tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n$ *and* $\hat{\boldsymbol{\Sigma}}_{A_1\phi} = \boldsymbol{A}_1'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1$. *There exist strict positive number* $\underline{c}^A$ *and* $\bar{c}^A$, *such that for* $j = 1, 2, \ldots,$ $[b\min(Dk_n, T)]$ *where* $b \in (0, 1)$,

1. *When* $Dk_n/T \to c \in [0, \infty)$ $\underline{c}^A + o_P(1) \leq \frac{1}{T}\varrho_j\left(\hat{\boldsymbol{\Sigma}}_{A_1\phi}\right) \leq \bar{c}^A + o_P(1)$.
2. *When* $Dk_n/T \to \infty$, $\underline{c}^A + o_P(1) \leq \frac{1}{Dk_n}\varrho_j\left(\hat{\boldsymbol{\Sigma}}_{A_1\phi}\right) \leq \bar{c}^A + o_P(1)$.

Proof. For $j = 1, 2, \ldots, \min(Dk_n, T)$,

$$\begin{aligned}\frac{1}{T}\varrho_j(\hat{\boldsymbol{\Sigma}}_{A_1\phi}) &= \frac{1}{T}\varrho_j(\boldsymbol{\Psi}'(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})/n\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n) \\ &= \varrho_j\left(\frac{\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{\Psi}'(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})}{nT}\cdot\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\cdot\frac{\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})}{n}\cdot\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\right).\end{aligned}$$

Suppose $s_t = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\phi}_{k_n}(\boldsymbol{x_i})\psi_{it}$, then we have $\frac{\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{\Psi}'(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})}{nT} = \frac{1}{T}\sum_{t=1}^T s_ts_t'$. Consequently, it holds that

$$\varrho_j\left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{s}_t\boldsymbol{s}_t'\right)\lambda_{min}^2(\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1})\lambda_{min}(\hat{\boldsymbol{\Sigma}}_\phi) \leq \frac{1}{T}\varrho_j(\hat{\boldsymbol{\Sigma}}_{A_1\phi}) \leq \varrho_j\left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{s}_t\boldsymbol{s}_t'\right)\lambda_{max}^2(\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1})\lambda_{max}(\hat{\boldsymbol{\Sigma}}_\phi).$$

Define $\tilde{\psi}_{it} = \tau - \mathbf{1}\{u_{it} \leq 0\}$ and $\tilde{\boldsymbol{\Psi}}$ as $n \times T$ matrix of $\tilde{\psi}_{it}$. It is easy to show that $\Pr(\tilde{\psi}_{it} = \tau - 1) = \tau$ and $\Pr(\tilde{\psi}_{it} = \tau) = 1 - \tau$ so $\mathbb{E}[\tilde{\psi}_{it}] = 0$, $\mathbb{V}[\tilde{\psi}_{it}] = \tau - \tau^2$. Let $\tilde{s}_t = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\phi}_{k_n}(\boldsymbol{x_i})\tilde{\psi}_{it}$.

Then, we have

$$s_t - \tilde{s}_t = \quad \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x_i}) \left[(\mathsf{F}(0|x_i) - \mathsf{F}(-\xi_{it}|x_i)) - (\mathbf{1}\{u_{it} \le 0\} - \mathbf{1}\{u_{it} \le -\xi_{it}\}) \right],$$

where $\xi_{it} = \boldsymbol{a}_{0t}' \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \boldsymbol{g}(\boldsymbol{x}_i) \boldsymbol{f}_{0t}$ is deterministic given $\boldsymbol{x}_i$, while Assumption 6(iv) implies that $\max_{1 \le t \le T,\ 1 \le i \le n} |\xi_{it}| = O_P(k_n^{-\alpha})$ with $\alpha > 2$. Defining $p_{it} = \mathsf{F}(0|x_i) - \mathsf{F}(-\xi_{it}|x_i)$, we get

$$|p_{it}| = \left| \int_{-\xi_{it}}^{0} f(u|x_i) du \right| \le \left| \int_{-\xi_{it}}^{0} (f(0|x_i) + L\xi_{it}) du \right| \le f(0|x_i)|\xi_{it}| + L\xi_{it}^2.$$

which means that $\max_{1 \le t \le T,\ 1 \le i \le n} |p_{it}| = O_P(k_n^{-\alpha})$. Further, note that

$$\mathbb{E}[\mathbf{1}\{u_{it} \le 0\} - \mathbf{1}\{u_{it} \le -\xi_{it}\}|x_i] = p_{it}, \quad \mathbb{V}[\mathbf{1}\{u_{it} \le 0\} - \mathbf{1}\{u_{it} \le -\xi_{it}\}|x_i] = p_{it}(1 - p_{it}) \le p_{it},$$

and

$$\mathbb{E}[(\mathbf{1}\{u_{is} \le 0\} - \mathbf{1}\{u_{is} \le -\xi_{is}\})(\mathbf{1}\{u_{it} \le 0\} - \mathbf{1}\{u_{it} \le -\xi_{it}\})|x_i] = 0$$

since $s \neq t$, u_{is} and u_{it} are independent. Thus, it holds uniformly in t that

$$\mathbb{E}[\|s_t - \tilde{s}_t\|^2] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\|\boldsymbol{\phi}_{k_n}(\boldsymbol{x_i})\|^2 \cdot p_{it}(1 - p_{it})] = O(k_n^{1-\alpha}),$$

and

$$\mathbb{E}[\|\tilde{s}_t\|^2] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\|\boldsymbol{\phi}_{k_n}(\boldsymbol{x_i})\|^2 \cdot \tau(1 - \tau)] = O(k_n).$$

Then, it follows that $\boldsymbol{s}_t \boldsymbol{s}_t' - \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' = (\boldsymbol{s}_t - \tilde{\boldsymbol{s}}_t)\boldsymbol{s}_t' + \tilde{\boldsymbol{s}}_t(\boldsymbol{s}_t - \tilde{\boldsymbol{s}}_t)'$ satisfies

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{s}_t \boldsymbol{s}_t' - \frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right\| \le \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{s}_t - \tilde{\boldsymbol{s}}_t\| (\|\boldsymbol{s}_t\| + \|\tilde{\boldsymbol{s}}_t\|) = O_P(k_n^{1-\alpha/2}) = o_P(1).$$

Hence, by Weyl's inequality,

$$\left| \varrho_j \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{s}_t \boldsymbol{s}_t' \right) - \varrho_j \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right) \right| \le \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{s}_t \boldsymbol{s}_t' - \frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right\| = o_P(1).$$

Similarly, when $Dk_n/T \to \infty$,

$$\left| \varrho_j \left(\frac{1}{Dk_n} \sum_{t=1}^{T} \boldsymbol{s}_t \boldsymbol{s}_t' \right) - \varrho_j \left(\frac{1}{Dk_n} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right) \right| \le \frac{T}{Dk_n} \cdot \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{s}_t \boldsymbol{s}_t' - \frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{s}}_t \tilde{\boldsymbol{s}}_t' \right\| = o_P(1).$$

Further, note that $\hat{\boldsymbol{\Sigma}}_\phi = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'$. Under Assumption 6(iii), It follows from Theorem 5.6.1 of Vershynin (2018) that $\|\hat{\boldsymbol{\Sigma}}_\phi - \boldsymbol{\Sigma}_\phi\|_S = O_P\left(\sqrt{\frac{k_n \log k_n}{n}} + \frac{k_n \log k_n}{n} \right) = o_P(1)$.

Thus, $|\varrho_j(\hat{\mathbf{\Sigma}}_\phi) - \varrho_j(\mathbf{\Sigma}_\phi)| = o_P(1)$ for $j = 1, 2, \ldots, Dk_n$. Using the same analysis, it can be shown that $|\varrho_j(\tilde{\mathbf{\Sigma}}_{\mathsf{f}\phi}) - \varrho_j(\mathbf{\Sigma}_{\mathsf{f}\phi})| = o_P(1)$ for $j = 1, 2, \ldots, Dk_n$. Since all eigenvalues of $\mathbf{\Sigma}_{\mathsf{f}\phi}^{-1}$ and $\mathbf{\Sigma}_\phi$ are both bounded below from 0 and above from infinity, it follows Lemma 6 that

1. When $Dk_n/T \to c \in [0, \infty)\backslash\{1\}$, with probability approaching 1,

$$\underline{c}^s \lambda_{min}^2(\mathbf{\Sigma}_{\mathsf{f}\phi}^{-1})\lambda_{min}(\mathbf{\Sigma}_\phi) \le \frac{1}{T}\varrho_j\left(\hat{\mathbf{\Sigma}}_{A_1\phi}\right) \le \bar{c}^s \lambda_{max}^2(\mathbf{\Sigma}_{\mathsf{f}\phi}^{-1})\lambda_{max}(\mathbf{\Sigma}_\phi).$$

2. When $Dk_n/T \to \infty$, with probability approaching 1,

$$\underline{c}^s \lambda_{min}^2(\mathbf{\Sigma}_{\mathsf{f}\phi}^{-1})\lambda_{min}(\mathbf{\Sigma}_\phi) \le \frac{1}{Dk_n}\varrho_j\left(\hat{\mathbf{\Sigma}}_{A_1\phi}\right) \le \bar{c}^s \lambda_{max}^2(\mathbf{\Sigma}_{\mathsf{f}\phi}^{-1})\lambda_{max}(\mathbf{\Sigma}_\phi),$$

which completes the proof. □

Lemma 8. *For $j \in \{1, \ldots, [b\min(Dk_n, T)] - 2R\}$ where $b \in (0, 1)$, there exists $\underline{c}$ and $\bar{c}$ such that*

1. *When $Dk_n/T \to c \in [0, \infty)$, $\bar{c} + o_P(1) \le \ n \cdot \hat{\rho}_{R+j} \le \underline{c} + o_P(1)$.*
2. *When $Dk_n/T \to \infty$, $\bar{c} + o_P(1) \le \ \frac{nT}{Dk_n} \cdot \hat{\rho}_{R+j} \le \underline{c} + o_P(1)$.*

Proof. Note that $\hat{\rho}_{R+j} = \varrho_{R+j}\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right)$. We begin with

$$\hat{\boldsymbol{Y}} = \mathbf{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} = \mathbf{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 + \mathbf{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0),$$

where $\hat{\boldsymbol{A}} = (\hat{\boldsymbol{a}}_1, \hat{\boldsymbol{a}}_2, \ldots, \hat{\boldsymbol{a}}_T)$, and $\boldsymbol{A}_0 = (\boldsymbol{a}_{01}, \boldsymbol{a}_{02}, \ldots, \boldsymbol{a}_{0T})$. As shown in Lemma 3, we have that for $t = 1, 2, \ldots, T$,

$$\begin{aligned}&\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t} - \tilde{\mathbf{\Sigma}}_{\mathsf{f}\phi}^{-1} \cdot \tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) = \\ &\quad \tilde{\mathbf{\Sigma}}_{\mathsf{f}\phi}^{-1}\left\{\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) - \boldsymbol{m}_t(\hat{\boldsymbol{a}}_t) - [\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)] - n^{-1}\mathbf{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\mathbf{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})\right\}, \end{aligned} \quad \text{(A.31)}$$

where $\tilde{\mathbf{\Sigma}}_{\mathsf{f}\phi} = \frac{1}{n}\sum_{i=1}^n \mathsf{f}(0|\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'$, and for any $\boldsymbol{a} \in \mathbb{R}^{Dk_n}$,

$$\begin{aligned}\boldsymbol{m}_t(\boldsymbol{a}) =& \frac{1}{n}\sum_{i=1}^n \left[\tau - \mathbf{1}\{u_{it} \le (\boldsymbol{a} - \boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\}\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i),\\ \boldsymbol{m}_t^*(\boldsymbol{a}) =& \frac{1}{n}\sum_{i=1}^n \left[\tau - \mathsf{F}\left((\boldsymbol{a} - \boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}|\boldsymbol{x}_i\right)\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i),\\ \tilde{\boldsymbol{m}}_t(\boldsymbol{a}) =& \boldsymbol{m}_t(\boldsymbol{a}) - \boldsymbol{m}_t^*(\boldsymbol{a}).\end{aligned}$$

As shown in Lemma 3,

$$\begin{aligned}\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) =& \frac{1}{n}\sum_{i=1}^n \left[\tau - \mathsf{F}\left(-\xi_{it}|\boldsymbol{x}_i\right)\right]\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\\ =& \frac{1}{n}\sum_{i=1}^n \mathsf{f}\left(0|\boldsymbol{x}_i\right)\cdot \xi_{it}\cdot \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) + O_P\left(k_n^{1/2-2\alpha}\right).\end{aligned}$$

Note that $\xi_{it} = \boldsymbol{g}'(\boldsymbol{x}_i)\boldsymbol{f}_{0t} - \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'\boldsymbol{a}_{0t} = (\boldsymbol{g}'(\boldsymbol{x}_i) - \boldsymbol{\phi}_{k_n}'(\boldsymbol{x}_i)\boldsymbol{B}_0)\boldsymbol{f}_{0t}$. We rewrite it as

$$\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) = \tilde{\boldsymbol{B}}\cdot \boldsymbol{f}_{0t} + O_P\left(k_n^{1/2-2\alpha}\right), \tag{A.32}$$

where $\tilde{\boldsymbol{B}} = \frac{1}{n}\sum_{i=1}^n \mathsf{f}\left(0|\boldsymbol{x}_i\right)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)(\boldsymbol{g}'(\boldsymbol{x}_i) - \boldsymbol{\phi}_{k_n}'(\boldsymbol{x}_i)\boldsymbol{B}_0)$. Further, note that

$$\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) = \frac{1}{n}\sum_{i=1}^n \psi_{it}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_{\boldsymbol{i}}). \tag{A.33}$$

Now, define $\boldsymbol{A}_2$ as the $Dk_n\times T$ matrix with each column being $\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\{[\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{B}}\cdot\boldsymbol{f}_{0t}] - \boldsymbol{m}_t(\hat{\boldsymbol{a}}_t) - [\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)] - n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})\}$. Then, combining (A.31) to (A.33), we have

$$\hat{\boldsymbol{A}} - \boldsymbol{A}_0 = \tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n + \tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\tilde{\boldsymbol{B}}\boldsymbol{F}' + \boldsymbol{A}_2.$$

Then we can rewrite $\hat{\boldsymbol{Y}}$ as (note that $\boldsymbol{A}_0 = \boldsymbol{B}_0\boldsymbol{F}'$)

$$\begin{aligned}\hat{\boldsymbol{Y}} =& \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 + \boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0)\\ =& \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{B}_0 + \tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\tilde{\boldsymbol{B}})\boldsymbol{F}' + \boldsymbol{\Phi}(\boldsymbol{X})\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n + \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2\\ =& \boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}' + \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1 + \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2\\ =& \boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}' + \boldsymbol{E}(\boldsymbol{X}).\end{aligned}$$

where $\boldsymbol{\Lambda}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{B}_0 + \tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\tilde{\boldsymbol{B}})$, $\boldsymbol{A}_1 = \tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n$, and $\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1 + \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2$.

Furthermore, define

$$\begin{aligned}\boldsymbol{F}^{**} =& \boldsymbol{F} + \boldsymbol{E}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X})(\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X}))^{-1},\\ \boldsymbol{P}(\boldsymbol{\Lambda}(\boldsymbol{X})) =& \boldsymbol{\Lambda}(\boldsymbol{X})(\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X}))^{-1}\boldsymbol{\Lambda}(\boldsymbol{X})'\\ \boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X})) =& \boldsymbol{I} - \boldsymbol{P}(\boldsymbol{\Lambda}(\boldsymbol{X})).\end{aligned}$$

Then we have that

$$\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT} = \frac{\boldsymbol{F}^{**}\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}^{**'}}{nT} + \frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}.$$

On the one hand, since the rank of $\boldsymbol{F}^{**}\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}^{**'}$ is no more than R, we have that $\varrho_{R+1}(\boldsymbol{F}^{**}\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}^{**'}) = 0$. Then, by Weyl's inequality,

$$\begin{aligned}\varrho_{R+j}\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \leq & \varrho_{R+1}\left(\frac{\boldsymbol{F}^{**}\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}^{**'}}{nT}\right) + \varrho_j\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right) \\ = & \varrho_j\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right).\end{aligned}$$

On the other hand, since $\boldsymbol{F}^{**}\boldsymbol{\Lambda}(\boldsymbol{X})'\boldsymbol{\Lambda}(\boldsymbol{X})\boldsymbol{F}^{**'}$ and $\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})$ are both positive semidefinite, we have that

$$\varrho_{R+j}\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \geq \varrho_{R+j}\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right).$$

It then follows from the above two inequalities that

$$\varrho_{R+j}\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right) \leq \varrho_{R+j}\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \leq \varrho_j\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right). \tag{A.34}$$

Further, note that $\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X}) + \boldsymbol{E}(\boldsymbol{X})'\boldsymbol{P}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})$ and the rank of $\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{P}(\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}_0)\boldsymbol{E}(\boldsymbol{X})$ is also no more than R. A similar reasoning implies that

$$\begin{aligned}\varrho_j\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right) \leq \varrho_j\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})}{nT}\right), \\ \varrho_{R+j}\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{Q}(\boldsymbol{\Lambda}(\boldsymbol{X}))\boldsymbol{E}(\boldsymbol{X})}{nT}\right) \geq \varrho_{2R+j}\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})}{nT}\right).\end{aligned} \tag{A.35}$$

Therefore, from (A.34) and (A.35) we have that

$$\varrho_{2R+j}\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})}{nT}\right) \leq \varrho_{R+j}\left(\frac{\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}}{nT}\right) \leq \varrho_j\left(\frac{\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})}{nT}\right). \tag{A.36}$$

Now we focus on the eigenvalues of $\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})$. Recall that $\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1 + \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2$, where $\boldsymbol{A}_2$ is a $Dk_n \times T$ matrix with each column being $\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\{[\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{B}} \cdot \boldsymbol{f}_{0t}] - \boldsymbol{m}_t(\hat{\boldsymbol{a}}_t) - [\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)] - n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})\}$. First, as shown in Lemma 7, $|\varrho_j(\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}) - \varrho_j(\boldsymbol{\Sigma}_{\mathsf{f}\phi})| = o_P(1)$ for $j = 1, 2, \ldots, Dk_n$. Since $\lambda_{min}(\boldsymbol{\Sigma}_{\mathsf{f}\phi})$ is bounded below from 0, $\lambda_{min}(\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi})$ is also bounded below from 0 with probability approaching 1. Then, we have

$$\|\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi}^{-1}\|_S = \frac{1}{\lambda_{min}(\tilde{\boldsymbol{\Sigma}}_{\mathsf{f}\phi})} = O_P(1). \tag{A.37}$$

Further, as shown in Lemma 3

$$
\begin{aligned}
\max_{1\le t\le T}\|\boldsymbol{m}_t(\hat{\boldsymbol{a}}_t)\| =& O_P(k_n^{3/2}/n),\\
\frac{1}{T}\sum_{t=1}^{T}\|\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t})-\tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)\|^2 =& O_P\left(\eta_{nT}^2\right),\\
\max_{1\le t\le T}\left\|n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})\right\| =& O_P(\sqrt{k_n}\varepsilon_{nT}^2).
\end{aligned}
\tag{A.38}
$$

Then, it follows from (A.32), (A.37) and (A.38) that

$$
\frac{1}{\sqrt{T}}\|\boldsymbol{A}_2\| \le O_P\left(k_n^{1/2-2\alpha}+k_n^{3/2}/n+\eta_{nT}+\sqrt{k_n}\varepsilon_{nT}^2\right). \tag{A.39}
$$

Note that

$$
\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X}) =\hat{\boldsymbol{\Sigma}}_{A_1\phi}+\boldsymbol{A}_1'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2+\boldsymbol{A}_2'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1+\boldsymbol{A}_2'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2,
$$

where $\hat{\boldsymbol{\Sigma}}_{A_1\phi}=\boldsymbol{A}_1'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1$. Next, it follows from (A.39) that

$$
\frac{||\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2||}{\sqrt{nT}} \le \frac{||\boldsymbol{\Phi}(\boldsymbol{X})||_S}{\sqrt{n}}\frac{||\boldsymbol{A}_2||}{\sqrt{T}} = O_P\left(k_n^{1/2-2\alpha}+k_n^{3/2}/n+\eta_{nT}+\sqrt{k_n}\varepsilon_{nT}^2\right).
$$

Thus we have

$$
\begin{aligned}
\|\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})-\hat{\boldsymbol{\Sigma}}_{A_1\phi}\| \le& 2\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1\|_S\cdot\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2\|+\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_2\|^2\\
=& O_P\left(\sqrt{nT}\left(k_n^{1/2-2\alpha}+k_n^{3/2}/n+\eta_{nT}+\sqrt{k_n}\varepsilon_{nT}^2\right)\right)\cdot\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1\|_S+\\
& O_P\left(nT\left(k_n^{1-4\alpha}+k_n^3/n^2+\eta_{nT}^2+k_n\varepsilon_{nT}^4\right)\right)\\
=& o_P(\sqrt{T})\cdot\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1\|_S+o_P(T),
\end{aligned}
$$

where the last equation, follow from Assumption 6(ii) ($n^{1/2}k_n^{1/2-2\alpha}=o(1)$ and $n^{1/2}\eta_{nT}=o(1)$). Two cases are distinguished.

(i) When $Dk_n/T\to c\in[0,\infty)$, $\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1\|_S=O(\sqrt{T})$ by Lemma 7(1). Consequently,

$$
\frac{1}{T}\|\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})-\hat{\boldsymbol{\Sigma}}_{A_1\phi}\| = \frac{1}{T}\cdot o(\sqrt{T})\cdot O(\sqrt{T})=o(1).
$$

Then, it follows from Weyl's inequality that, for $j=1,\ldots,[b\min(Dk_n,T)]$,

$$
\left|\frac{1}{T}\varrho_j\left(\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})\right)-\frac{1}{T}\varrho_j\left(\hat{\boldsymbol{\Sigma}}_{A_1\phi}\right)\right| \le \frac{1}{T}\|\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})-\hat{\boldsymbol{\Sigma}}_{A_1\phi}\|=o(1).
$$

Then, statement 1 holds with $\underline{c}=\underline{c}^A$ and $\bar{c}=\bar{c}^A$ following Lemma 7(1) and (A.36).

(ii) When $Dk_n/T \to \infty$, $\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_1\|_S = O(\sqrt{Dk_n})$ by Lemma 7(2). Consequently,

$$\frac{1}{Dk_n}\|\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X}) - \hat{\boldsymbol{\Sigma}}_{A_1\phi}\| = \frac{1}{Dk_n}\cdot o(\sqrt{T})\cdot O(\sqrt{Dk_n}) = o\left(\sqrt{\frac{T}{Dk_n}}\right) = o(1).$$

Then, using Weyl's inequality again, yields that, for $j = 1, \ldots, [b\min(Dk_n, T)]$,

$$\left|\frac{1}{Dk_n}\varrho_j\left(\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X})\right) - \frac{1}{Dk_n}\varrho_j\left(\hat{\boldsymbol{\Sigma}}_{A_1\phi}\right)\right| \leq \frac{1}{Dk_n}\|\boldsymbol{E}(\boldsymbol{X})'\boldsymbol{E}(\boldsymbol{X}) - \hat{\boldsymbol{\Sigma}}_{A_1\phi}\| = o(1).$$

Then statement 2 holds with $\underline{c} = \underline{c}^A$ and $\bar{c} = \bar{c}^A$ following Lemma 7(2) and (A.36).

□

Proof of Theorem 5:

Proof. From Lemmas 5 and 8, it follows that

1. When $Dk_n/T \to c \in [0, \infty)$,
   - When $j < R$, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}} = \frac{\rho_j}{\rho_{j+1}} + o_P(1) = O_P(1)$;
   - When $j > R$, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}} \leq \frac{\bar{c}}{\underline{c}} + o_P(1) = O_P(1)$;
   - When $j = R$, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}} \geq n \cdot \frac{\rho_j}{\bar{c}} + o_P(1) \xrightarrow{P} +\infty$.
2. When $Dk_n/T \to \infty$,
   - When $j < R$, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}} = \frac{\rho_j}{\rho_{j+1}} + o_P(1) = O_P(1)$;
   - When $j > R$, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}} \leq \frac{\bar{c}}{\underline{c}} + o_P(1) = O_P(1)$;
   - When $j = R$, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}} \geq \frac{nT}{Dk_n} \cdot \frac{\rho_j}{\bar{c}} + o_P(1) \xrightarrow{P} +\infty$.

Thus, $\frac{\hat{\rho}_j}{\hat{\rho}_{j+1}}$ is maximized when $j = R$ in both cases with probability approaching 1, and R will be consistently estimated irrespective of k_n insofar as the choice of this tuning parameter satisfies Assumption 3 (iv) above, e.g. by cross-validation. □

Proof of Theorem 6:

Proof. By definition we have

$$\begin{aligned}
\check{\boldsymbol{\Lambda}} &= \boldsymbol{\Lambda}\boldsymbol{H} + (\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}) \\
&= \boldsymbol{G}(\boldsymbol{X})\boldsymbol{H} + \boldsymbol{\Gamma}\boldsymbol{H} + (\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}) \\
&= \boldsymbol{\Phi}\boldsymbol{B}_0\boldsymbol{H} + \boldsymbol{R}\boldsymbol{H} + \boldsymbol{\Gamma}\boldsymbol{H} + (\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}),
\end{aligned}$$

where $\boldsymbol{B}_0$ is defined in Section 2.2, and $\boldsymbol{R} = \boldsymbol{G}(\boldsymbol{X}) - \boldsymbol{\Phi}\boldsymbol{B}_0$. Denoting $\boldsymbol{P}_\Phi = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'$, it follows from the definition of $\check{\boldsymbol{G}}(\boldsymbol{X})$ that

$$\begin{aligned}
\check{\boldsymbol{G}}(\boldsymbol{X}) &= \boldsymbol{P}_\Phi \check{\boldsymbol{\Lambda}} = \boldsymbol{\Phi}\boldsymbol{B}_0\boldsymbol{H} + \boldsymbol{P}_\Phi \boldsymbol{R}\boldsymbol{H} + \boldsymbol{P}_\Phi \boldsymbol{\Gamma}\boldsymbol{H} + \boldsymbol{P}_\Phi(\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}) \\
&= \boldsymbol{G}(\boldsymbol{X})\boldsymbol{H} + (\boldsymbol{P}_\Phi - \boldsymbol{I})\boldsymbol{R}\boldsymbol{H} + \boldsymbol{P}_\Phi \boldsymbol{\Gamma}\boldsymbol{H} + \boldsymbol{P}_\Phi(\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}).
\end{aligned}$$

Then, by Assumption 6, we have

$$\begin{aligned}
\|\check{\boldsymbol{G}}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{H}\| &\leq \|(\boldsymbol{P}_\Phi - \boldsymbol{I})\boldsymbol{R}\| \cdot \|\boldsymbol{H}\| + \|\boldsymbol{P}_\Phi \boldsymbol{\Gamma}\| \cdot \|\boldsymbol{H}\| + \|\boldsymbol{P}_\Phi(\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H})\| \\
&\lesssim \|\boldsymbol{R}\| + \|\boldsymbol{P}_\Phi \boldsymbol{\Gamma}\| + \|\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}\|.
\end{aligned}$$

First, it follows from equation (4) that $\|\boldsymbol{R}\| = O_P(\sqrt{n}k_n^{-\alpha})$. Second, Assumption 6(i) implies that $\|\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}\| = O_P(\sqrt{n}l_{nT})$. Third, Assumptions 6(ii) and 6(iii) imply that $\|\boldsymbol{P}_\Phi \boldsymbol{\Gamma}\| = O_P(\sqrt{k_n})$. Thus, part (i) of Theorem 6 holds.

Next, it follows from the definition of $\check{\boldsymbol{g}}(\boldsymbol{x})$ and $\check{\boldsymbol{\Lambda}} = \boldsymbol{\Phi}\boldsymbol{B}_0\boldsymbol{H} + \boldsymbol{R}\boldsymbol{H} + \boldsymbol{\Gamma}\boldsymbol{H} + (\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H})$ that

$$\begin{aligned}
\check{\boldsymbol{g}}(\boldsymbol{x}) &= \boldsymbol{H}'\boldsymbol{B}_0'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}) + [\boldsymbol{R}\boldsymbol{H} + \boldsymbol{\Gamma}\boldsymbol{H} + (\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H})]'\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}) \\
&= \boldsymbol{H}'\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{H}'[\boldsymbol{B}_0'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})] + [\boldsymbol{R}\boldsymbol{H} + \boldsymbol{\Gamma}\boldsymbol{H} + (\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H})]'\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}).
\end{aligned}$$

Thus, by Assumption 6,

$$\begin{aligned}
\|\check{\boldsymbol{g}}(\boldsymbol{x}) - \boldsymbol{H}'\boldsymbol{g}(\boldsymbol{x})\| \lesssim\ & \|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{B}_0'\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\| + \|(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{R}\| \cdot \|\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\| \\
& + \|(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma}\| \cdot \|\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\| + \|(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'(\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H})\| \cdot \|\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\|,
\end{aligned}$$

whose RHS components behave as follows. First, $\|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{B}_0'\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\| = O_P(k_n^{-\alpha})$ by equation (4). Second, by Assumption 6(ii),

$$\begin{aligned}
\|(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{R}\|^2 &= \mathrm{Tr}\left[(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{R}\boldsymbol{R}'\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right] \\
&= \mathrm{Tr}\left[(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1/2}\boldsymbol{\Phi}'\boldsymbol{R}\boldsymbol{R}'\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1/2} \cdot (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right] \\
&\leq \lambda_{\max}[(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}] \cdot \mathrm{Tr}\left[(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1/2}\boldsymbol{\Phi}'\boldsymbol{R}\boldsymbol{R}'\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1/2}\right] \\
&= \lambda_{\max}[(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}] \cdot \mathrm{Tr}\left[\boldsymbol{R}\boldsymbol{R}'\boldsymbol{P}_\Phi\right] \\
&\leq \lambda_{\max}[(\boldsymbol{\Phi}'\boldsymbol{\Phi}/n)^{-1}] \cdot \lambda_{\max}(\boldsymbol{P}_\Phi) \cdot \|\boldsymbol{R}\|^2/n = O_P(k_n^{-2\alpha}).
\end{aligned}$$

Similarly, it can be shown that $\|(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'(\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H})\|^2 = O_P(l_{nT}^2)$. Third, by Assumption 6(ii) and 6(iii),

$$\begin{aligned}\|(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma}\|^2 &= \operatorname{Tr}\left[(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-2}\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\Phi}\right] \\ &\leq \lambda_{\max}[(\boldsymbol{\Phi}'\boldsymbol{\Phi}/n)^{-2}]\cdot \operatorname{Tr}\left[\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\Phi}/n^2\right] \\ &= n^{-1}\lambda_{\max}[(\boldsymbol{\Phi}'\boldsymbol{\Phi}/n)^{-2}]\cdot\|\boldsymbol{\Phi}'\boldsymbol{\Gamma}/\sqrt{n}\|^2 \\ &= O_P(k_n/n).\end{aligned}$$

Finally, part (ii) of Theorem 6 holds because $\sup_{x\in\mathcal{X}}\|\boldsymbol{\phi}_{k_n}(\boldsymbol{x})\| = \sqrt{k_n}$. □

A.2 Probability Limit of the Rotation Matrix

Lemma 9. $\frac{1}{T\sqrt{n}}\|\hat{\boldsymbol{V}}\boldsymbol{F}_0\| = O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right).$

Proof. First, recall that

$$\begin{aligned}\hat{\boldsymbol{V}} =&\hat{\boldsymbol{Y}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0' \\ =&\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 + (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0') \\ =&\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}} - \boldsymbol{A}_0) + O_P\left(\sqrt{nT}\cdot k_n^{-\alpha}\right).\end{aligned} \tag{A.40}$$

Further, it holds from Lemma 4 that

$$\hat{\boldsymbol{A}} - \boldsymbol{A}_0 = \boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})/n + O_P\left(\sqrt{T}k_n^{-\alpha}\right) + O_P\left(\sqrt{T}\eta_{nT}\right) \tag{A.41}$$

Next, we focus on $\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{F}_0$. Note that it is a $n\times R$ matrix with each row being $\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{F}_0$. By the proof of Theorem 3,

$$\|\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{F}_0\| = O_P\left(\sqrt{nTk_n}\right).$$

Consequently, we have that

$$\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{F}_0/n\| = \sqrt{n}\cdot O_P\left(\sqrt{nTk_n}/n\right) = O_P\left(\sqrt{Tk_n}\right) \tag{A.42}$$

Combining (A.40) to (A.42), it holds that

$$
\begin{aligned}
\frac{1}{T\sqrt{n}}\|\hat{\boldsymbol{V}}\boldsymbol{F}_0\| \leq & \frac{1}{T\sqrt{n}}\|\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)\boldsymbol{F}_0\| + O_P\left(k_n^{-\alpha}\right)\cdot\frac{\|\boldsymbol{F}_0\|}{\sqrt{T}} \\
\leq & \frac{1}{T\sqrt{n}}\|\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}(\boldsymbol{X})\boldsymbol{F}_0/n\| + \frac{\|\boldsymbol{\Phi}(\boldsymbol{X})\|_S}{\sqrt{n}}\cdot O_P\left(k_n^{-\alpha}+\eta_{nT}\right)\cdot\frac{\|\boldsymbol{F}_0\|}{\sqrt{T}} + O_P\left(k_n^{-\alpha}\right) \\
= & O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right)
\end{aligned}
$$

□

Lemma 10. *1.* $\frac{1}{T}\|\boldsymbol{F}_0'(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})\| = O_P\left(\varepsilon_{nT}^2\right) + O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right)$

2. $\frac{1}{T}\|(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})'\hat{\boldsymbol{F}}\| = O_P\left(\varepsilon_{nT}^2\right) + O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right)$

Proof. We first prove statement 1. From (A.2) we have that

$$
\boldsymbol{F}_0'(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})/T = \boldsymbol{F}_0'\hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'\hat{\boldsymbol{F}}/(nT^2)\hat{\boldsymbol{\Omega}}^{-1} + \boldsymbol{F}_0'\boldsymbol{F}_0\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/(nT^2)\hat{\boldsymbol{\Omega}}^{-1} + \boldsymbol{F}_0'\hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}/(nT^2)\hat{\boldsymbol{F}}\hat{\boldsymbol{\Omega}}^{-1}. \tag{A.43}
$$

In the proof of Theorem 1 we have shown that $\|\hat{\boldsymbol{\Omega}}^{-1}\| = O_P(1)$. Then we analyze (A.43) term by term. For the first term, it follows Lemma 9 that

$$
\begin{aligned}
\|\boldsymbol{F}_0'\hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}_0'\hat{\boldsymbol{F}}/(nT^2)\hat{\boldsymbol{\Omega}}^{-1}\| \leq & \frac{\|\boldsymbol{F}_0'\hat{\boldsymbol{V}}'\|}{T\sqrt{n}}\cdot\frac{\|\boldsymbol{G}(\boldsymbol{X})\|}{\sqrt{n}}\cdot\frac{\|\boldsymbol{F}_0\|}{\sqrt{T}}\cdot\frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}}\cdot\|\hat{\boldsymbol{\Omega}}^{-1}\| \\
= & O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right). \qquad \text{(A.44)}
\end{aligned}
$$

Then, the second term satisfies that

$$
\begin{aligned}
\|\boldsymbol{F}_0'\boldsymbol{F}_0\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/(nT^2)\hat{\boldsymbol{\Omega}}^{-1}\| = & \|\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/(nT)\hat{\boldsymbol{\Omega}}^{-1}\| \\
\leq & \|\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})/(nT)\hat{\boldsymbol{\Omega}}^{-1}\| + \|\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}\boldsymbol{F}_0\hat{\boldsymbol{H}}/(nT)\hat{\boldsymbol{\Omega}}^{-1}\| \\
\leq & \frac{\|\boldsymbol{G}(\boldsymbol{X})\|}{\sqrt{n}}\frac{\|\hat{\boldsymbol{V}}\|}{\sqrt{nT}}\frac{\|\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}}\|}{\sqrt{T}}\|\hat{\boldsymbol{\Omega}}^{-1}\| + \frac{\|\boldsymbol{G}(\boldsymbol{X})\|}{\sqrt{n}}\frac{\|\hat{\boldsymbol{V}}\boldsymbol{F}_0\|}{T\sqrt{n}}\|\hat{\boldsymbol{H}}\|\|\hat{\boldsymbol{\Omega}}^{-1}\| \\
= & O_P\left(\varepsilon_{nT}^2\right) + O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right). \qquad \text{(A.45)}
\end{aligned}
$$

As for the third term, we have that

$$
\|\boldsymbol{F}_0'\hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}/(nT^2)\hat{\boldsymbol{F}}\hat{\boldsymbol{\Omega}}^{-1}\| \leq \frac{\|\boldsymbol{F}_0\|}{\sqrt{T}}\frac{\|\hat{\boldsymbol{V}}\|^2}{nT}\frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} = O_P\left(\varepsilon_{nT}^2\right). \tag{A.46}
$$

Then the statement 1 follows from (A.43) to (A.46).

Then, it holds that

$$
\begin{aligned}
\frac{1}{T}\|(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})'\hat{\boldsymbol{F}}\| \leq & \frac{1}{T}\|(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})'(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})\| + \frac{1}{T}\|(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})'\boldsymbol{F}_0\hat{\boldsymbol{H}}\| \\
= & O_P\left(\varepsilon_{nT}^2\right) + O_P\left(\sqrt{\frac{k_n}{nT}}\right) + O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right),
\end{aligned}
$$

which proves the statement 2. □

Lemma 11. *If $\boldsymbol{F}_0'\boldsymbol{F}_0/T = \boldsymbol{I}_R$ and $\hat{\boldsymbol{\Sigma}}_g$ is diagonal with distinct elements, then $\hat{\boldsymbol{H}} = \boldsymbol{I}_R + O_P\left(\delta_{nT}\right)$, where $\delta_{nT} = \varepsilon_{nT}^2 + \sqrt{\frac{k_n}{nT}} + k_n^{-\alpha} + \eta_{nT}$.*

Proof. We begin with

$$
\frac{\boldsymbol{F}_0'\hat{\boldsymbol{F}}}{T} = \frac{\boldsymbol{F}_0'(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})}{T} + \frac{\boldsymbol{F}_0'\boldsymbol{F}_0}{T}\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}} + O_P(\delta_{nT}), \tag{A.47}
$$

where the second equality is due to Lemma 9. Left multiplying $\hat{\boldsymbol{H}}'$ on both sides, we have that

$$
\begin{aligned}
& \frac{\hat{\boldsymbol{H}}'\boldsymbol{F}_0'\hat{\boldsymbol{F}}}{T} = \hat{\boldsymbol{H}}'\hat{\boldsymbol{H}} + O_P(\delta_{nT}) \\
\Rightarrow & \frac{\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}}{T} - \frac{(\hat{\boldsymbol{F}}-\boldsymbol{F}_0\hat{\boldsymbol{H}})'\hat{\boldsymbol{F}}}{T} = \hat{\boldsymbol{H}}'\hat{\boldsymbol{H}} + O_P(\delta_{nT}) \\
\Rightarrow & \boldsymbol{I}_R + O_P(\delta_{nT}) = \hat{\boldsymbol{H}}'\hat{\boldsymbol{H}},
\end{aligned}
$$

where the last equality is due to Lemma 10. Ignoring $O_P(\delta_{nT})$, the above shows that $\hat{\boldsymbol{H}}$ is an orthogonal matrix so that its eigenvalues are either 1 or -1. We need to show that $\hat{\boldsymbol{H}}$ is a diagonal matrix. From the definition of $\hat{\boldsymbol{H}}$,

$$
\hat{\boldsymbol{H}} = \hat{\boldsymbol{\Sigma}}_g(\boldsymbol{F}_0'\hat{\boldsymbol{F}})/T\hat{\boldsymbol{\Omega}}^{-1} = \hat{\boldsymbol{\Sigma}}_g\hat{\boldsymbol{H}}\hat{\boldsymbol{\Omega}}^{-1} + O_P(\delta_{nT}) \Rightarrow \hat{\boldsymbol{\Sigma}}_g\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}}\hat{\boldsymbol{\Omega}} + O_P(\delta_{nT}).
$$

This equation implies that $\hat{\boldsymbol{H}}$ (up to a negligible term) is a matrix consisting of eigenvectors of $\hat{\boldsymbol{\Sigma}}_g$. The latter matrix is diagonal and has distinct eigenvalues by assumption. Thus, each eigenvalue is associated with a unique unitary eigenvector (up to a sign change) and each eigenvector has a single non-zero element. This implies that $\hat{\boldsymbol{H}}$ is a diagonal matrix up to an $O_P(\delta_{nT})$ order. It is already known that the eigenvalues of $\hat{\boldsymbol{H}}$ are 1 or -1, $\hat{\boldsymbol{H}}$ is a diagonal matrix with elements of 1 or -1 as its elements. Without loss of generality, we can assume all elements are 1. This implies $\hat{\boldsymbol{H}} = \boldsymbol{I}_R + O_P\left(\delta_{nT}\right)$. Moreover, from (A.47) we obtain

$$
\frac{\boldsymbol{F}_0'\hat{\boldsymbol{F}}}{T} = \boldsymbol{I}_R + O_P\left(\delta_{nT}\right).
$$

□

A.3 Estimating the Covariance Matrix of the QPPCA Estimator

Recall that $\boldsymbol{\Sigma}_{\mathsf{f}\phi} = \mathbb{E}[\mathsf{f}(0|\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']$ and $\sigma^2_{k_n}(\boldsymbol{x}) = \boldsymbol{\phi}'_{k_n}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}_{\mathsf{f}\phi}\boldsymbol{\Sigma}_{\phi}\boldsymbol{\Sigma}^{-1}_{\mathsf{f}\phi}\boldsymbol{\phi}_{k_n}(\boldsymbol{x})$. Here we make the dependency of σ_{k_n} on $\boldsymbol{x}$ explicit. Based on Theorem 3, for the purpose of inference and constructing confidence intervals, we propose to estimate the co-variance matrix of $\hat{\boldsymbol{g}}(\boldsymbol{x})$ by:

$$\hat{\boldsymbol{\Sigma}}_{T,\tau} \cdot \frac{\hat{\sigma}^2_{k_n}(\boldsymbol{x})}{nT},$$

where

$$\hat{\boldsymbol{\Sigma}}_{T,\tau} = \tau(1-\tau)(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}/T) = \tau(1-\tau)\boldsymbol{I}_R, \quad \hat{\sigma}^2_{k_n}(\boldsymbol{x}) = \boldsymbol{\phi}'_{k_n}(\boldsymbol{x})\hat{\boldsymbol{\Sigma}}^{-1}_{\mathsf{f}\phi}\hat{\boldsymbol{\Sigma}}_{\phi}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathsf{f}\phi}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}),$$

$$\hat{\boldsymbol{\Sigma}}_{\mathsf{f}\phi} = \frac{1}{n}\sum_{i=1}^n \hat{\mathsf{f}}(0|\boldsymbol{x}_i)\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)', \quad \hat{\boldsymbol{\Sigma}}_{\phi} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)',$$

$$\hat{\mathsf{f}}(0|\boldsymbol{x}_i) = \frac{1}{T}\sum_{t=1}^T \frac{1}{h}\mathcal{K}\left(\frac{y_{it}-\hat{y}_{it}}{h}\right),$$

$\mathcal{K}(\cdot)$ is a standard kernel function with bounded first-order derivative, and h is a bandwidth parameter.

We need to show that

$$\left\|\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}/T - \hat{\boldsymbol{H}}'(\boldsymbol{F}_0'\boldsymbol{F}_0/T)\hat{\boldsymbol{H}}\right\| = o_P(1), \tag{A.48}$$

$$(\hat{\sigma}_{k_n}(\boldsymbol{x}) - \sigma_{k_n}(\boldsymbol{x}))/\hat{\sigma}_{k_n}(\boldsymbol{x}) = o_P(1). \tag{A.49}$$

Note that (A.48) follows trivially from Theorem 1. To prove (A.49), it suffices to show that

$$\left\|\hat{\boldsymbol{\Sigma}}_{\mathsf{f}\phi} - \boldsymbol{\Sigma}_{\mathsf{f}\phi}\right\| = o_P(1). \tag{A.50}$$

and

$$\left\|\hat{\boldsymbol{\Sigma}}_{\phi} - \boldsymbol{\Sigma}_{\phi}\right\| = o_P(1). \tag{A.51}$$

First, write

$$\hat{\boldsymbol{\Sigma}}_{\mathsf{f}\phi} - \boldsymbol{\Sigma}_{\mathsf{f}\phi} = \frac{1}{n}\sum_{i=1}^n\left(\hat{\mathsf{f}}(0|\boldsymbol{x}_i) - \frac{1}{T}\sum_{t=1}^T\frac{1}{h}\mathcal{K}\left(\frac{u_{it}}{h}\right)\right)\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)' +$$
$$\frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=1}^T\frac{1}{h}\mathcal{K}\left(\frac{u_{it}}{h}\right) - \mathsf{f}(0|\boldsymbol{x}_i)\right)\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)' + \left[\frac{1}{n}\sum_{i=1}^n \mathsf{f}(0|\boldsymbol{x}_i)\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)' - \boldsymbol{\Sigma}_{\mathsf{f}\phi}\right].$$

It is standard to show that the first term on the right-hand side of the above equation is

$O_P(k_n\varepsilon_{nT}h^{-2})$, the second term is $O_P\left(k_n[(nh)^{-1/2}+h]\right)$, and the last term is $O_P(k_n n^{-1/2})$. Thus, (A.50) holds under the conditions that

$$k_n\varepsilon_{nT}h^{-2}=o(1), \quad k_n(nh)^{-1/2}=o(1), \quad k_n h=o(1).$$

In the case where k_n is chosen by cross-validation and therefore $k_n \asymp n^{1/(2\alpha+1)}$, the above conditions are satisfied if

$$n^{\frac{1}{2\alpha+1}} << h^{-1} << n^{\frac{\alpha-1}{4\alpha+2}},$$

which implicitly requires $\alpha > 3$. Finally, it is easy to see that $\|\hat{\mathbf{\Sigma}}_\phi - \mathbf{\Sigma}_\phi\| = O_P(k_n n^{-1/2})$ and thus (A.51) holds.

B Additional Simulation Results

This section contains some additional simulation results. We briefly summarize the results in what follows, and the detailed DGPs are described in the notes under each table or figure.

Figure B.1: estimation errors of the loading matrices using QPPCA, QFA-Sieve, PCA and PPCA. The DGP is the same as Figure 1 of the main text.

Figure B.2: estimation errors of the factors using QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$. The DGP is the same as Figure 2 of the main text.

Figures B.3 and B.4: estimation errors of the loading matrices using QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25, 0.75$. The DGP is the same as Figure 2 of the main text.

Figures B.5 and B.6: estimation errors of the factors using QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25, 0.75$, for the case $R = 2, D = 3$. These figures demonstrate the advantage of QPPCA and QFA-Sieve when $D > R$.

Figures B.7 and B.8: estimation errors of the loading matrices using QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25, 0.75$, for the case $R = 2, D = 3$. These figures demonstrate the advantage of QPPCA and QFA-Sieve when $D > R$.

Figures B.9 and B.10: estimation errors of the factors using QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25, 0.75$, for the case $R = 2, D = 1$. These figures demonstrate the advantage of QPPCA and QFA-Sieve when $D < R$.

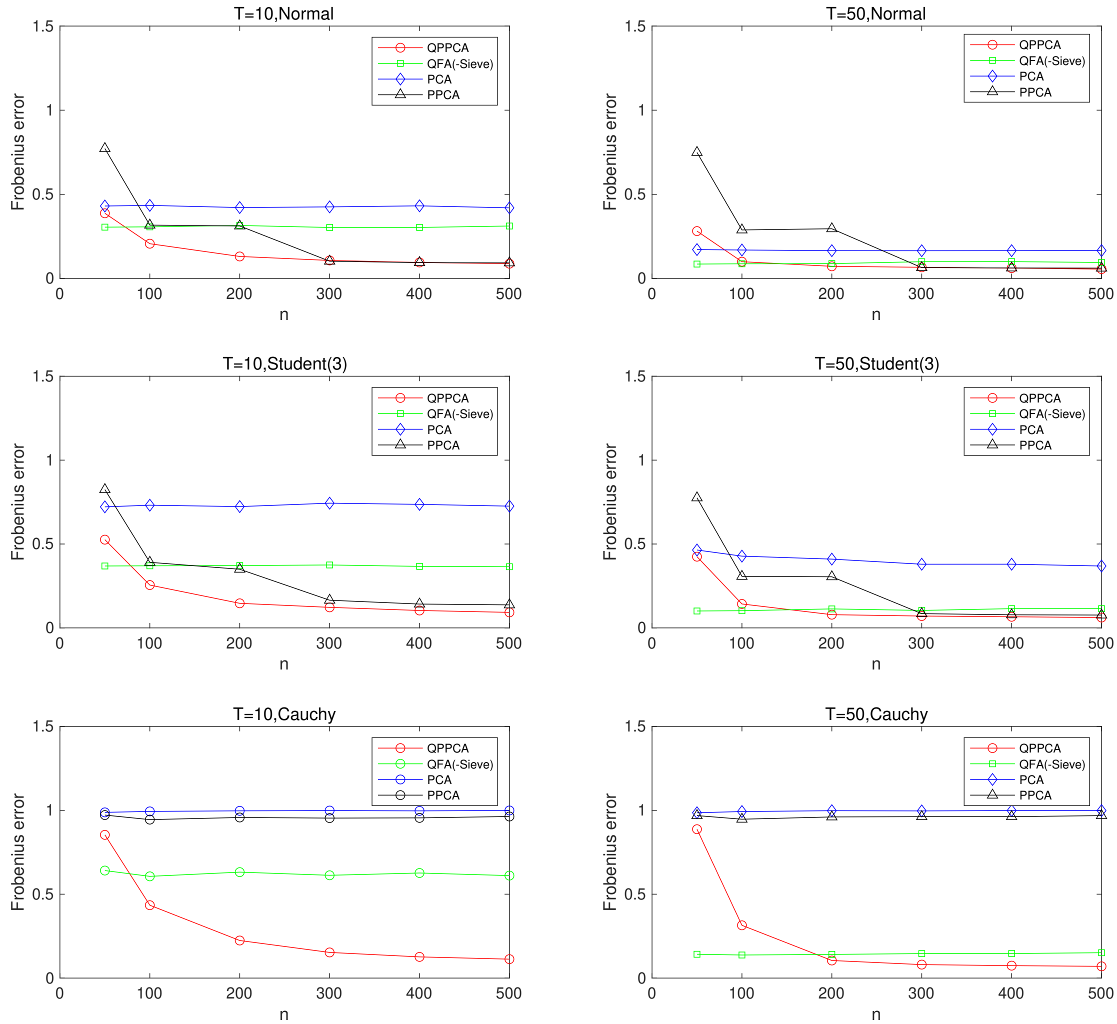
Figures B.11 and B.12: estimation errors of the loading matrices using QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25, 0.75$, for the case $R = 2, D = 1$. These figures demonstrate the advantage of QPPCA and QFA-Sieve when $D < R$.

Figure B.13: estimation errors of the loading functions using QPPCA and QFA-Sieve at $\tau = 0.75$, without missing variables, $n = 355, T = 62, R = 1, D = 4$. The DGP is the same as Figure 3 of the main text.

Figure B.14: estimation errors of the loading functions using QPPCA and QFA-Sieve at $\tau = 0.75$, with one missing variable, $n = 355, T = 10, R = 1, D = 5$. The DGP is the same as Figure 4 of the main text.

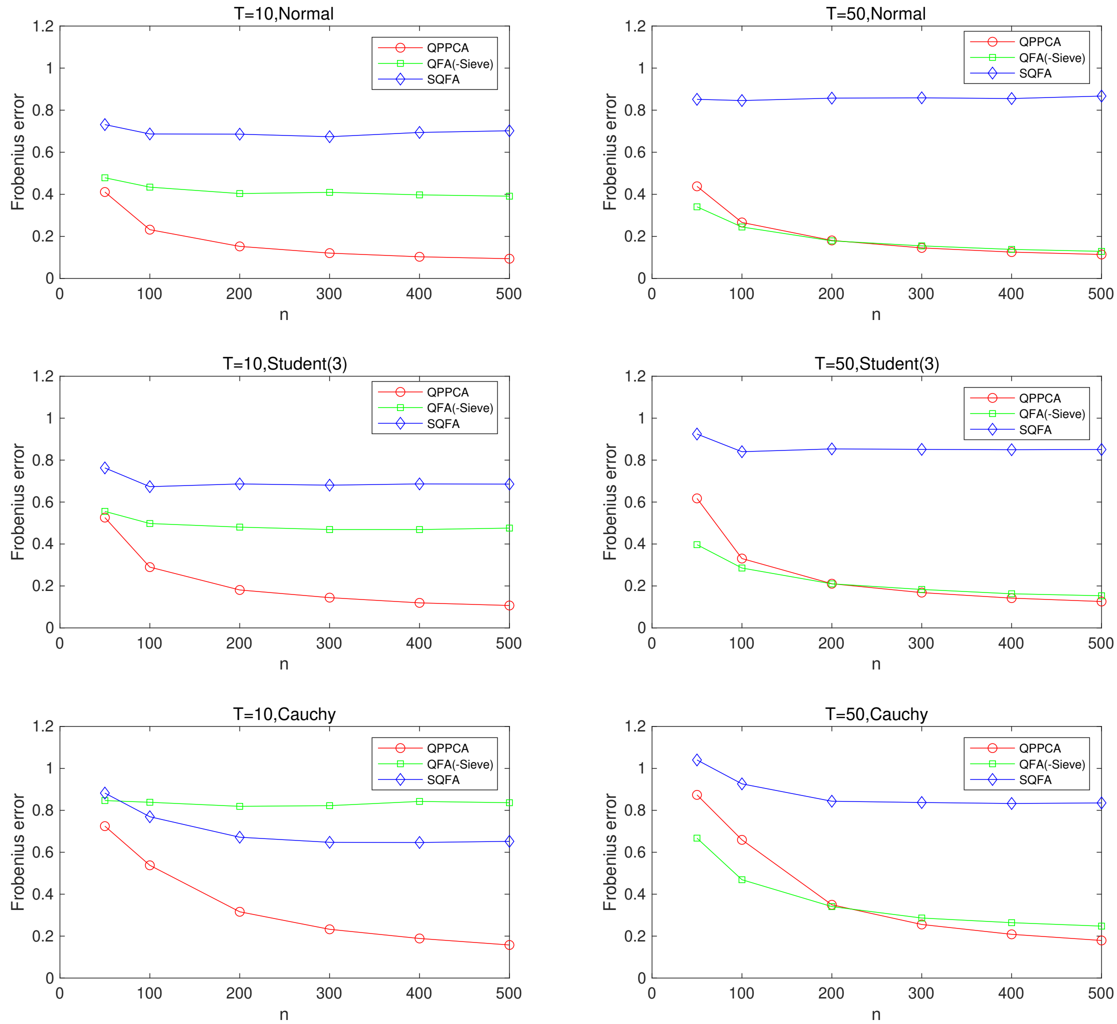
Figures B.15 and B.16: estimation errors of the loading functions using QPPCA and QFA-Sieve at $\tau = 0.25, 0.75$, without missing variables, $n = 355, T = 10, R = 2, D = 2$. These figures demonstrate the advantage of QPPCA when T is small.

Figure B.1: Estimation of loading matrix: QPPCA, PCA, PPCA and QFA-Sieve.



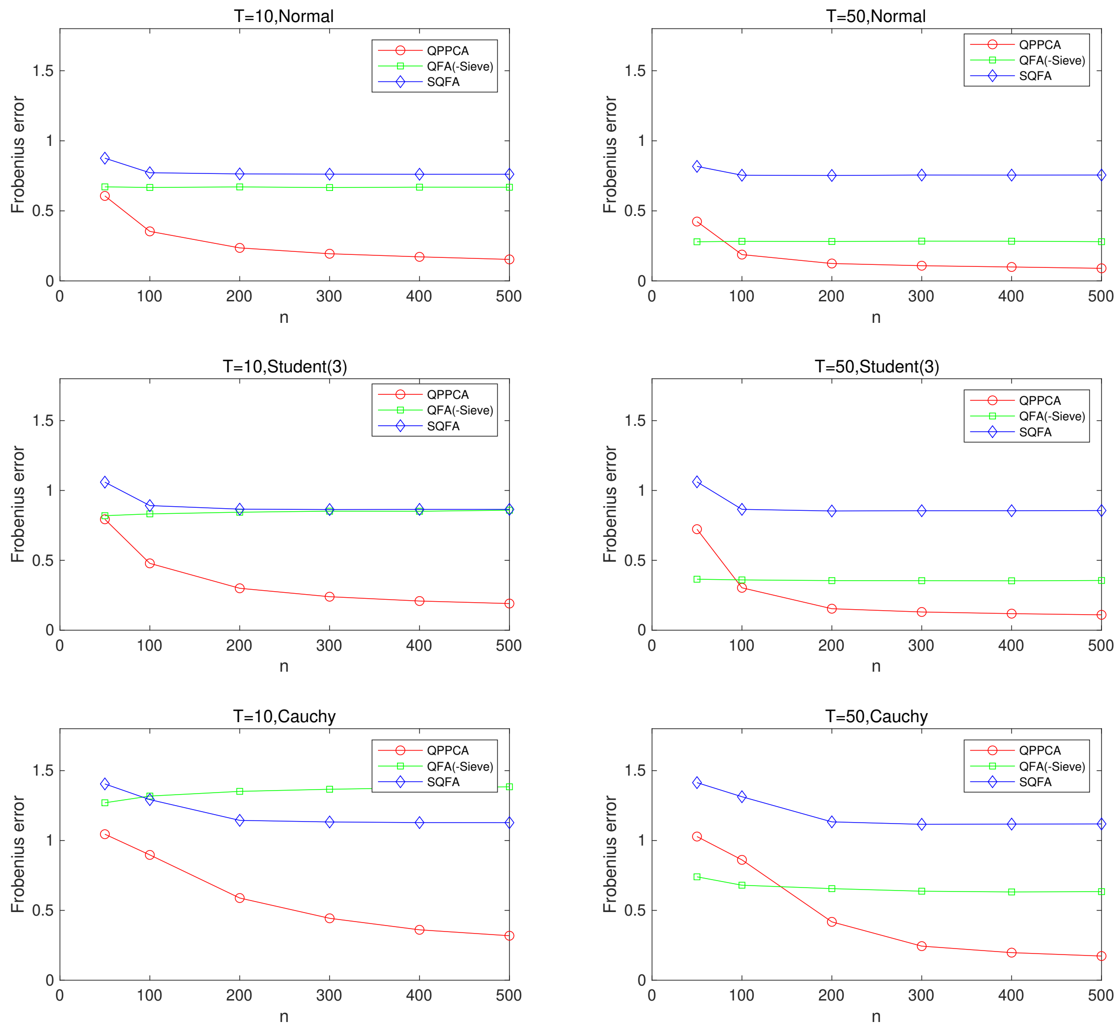
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrix is estimated by 3 methods: PCA, PPCA, QPPCA at $\tau = 0.5$ and QFA-Sieve at $\tau = 0.5$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{N}$ from 1000 repetitions.

Figure B.2: Estimation of factors: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$, with $D = R$.



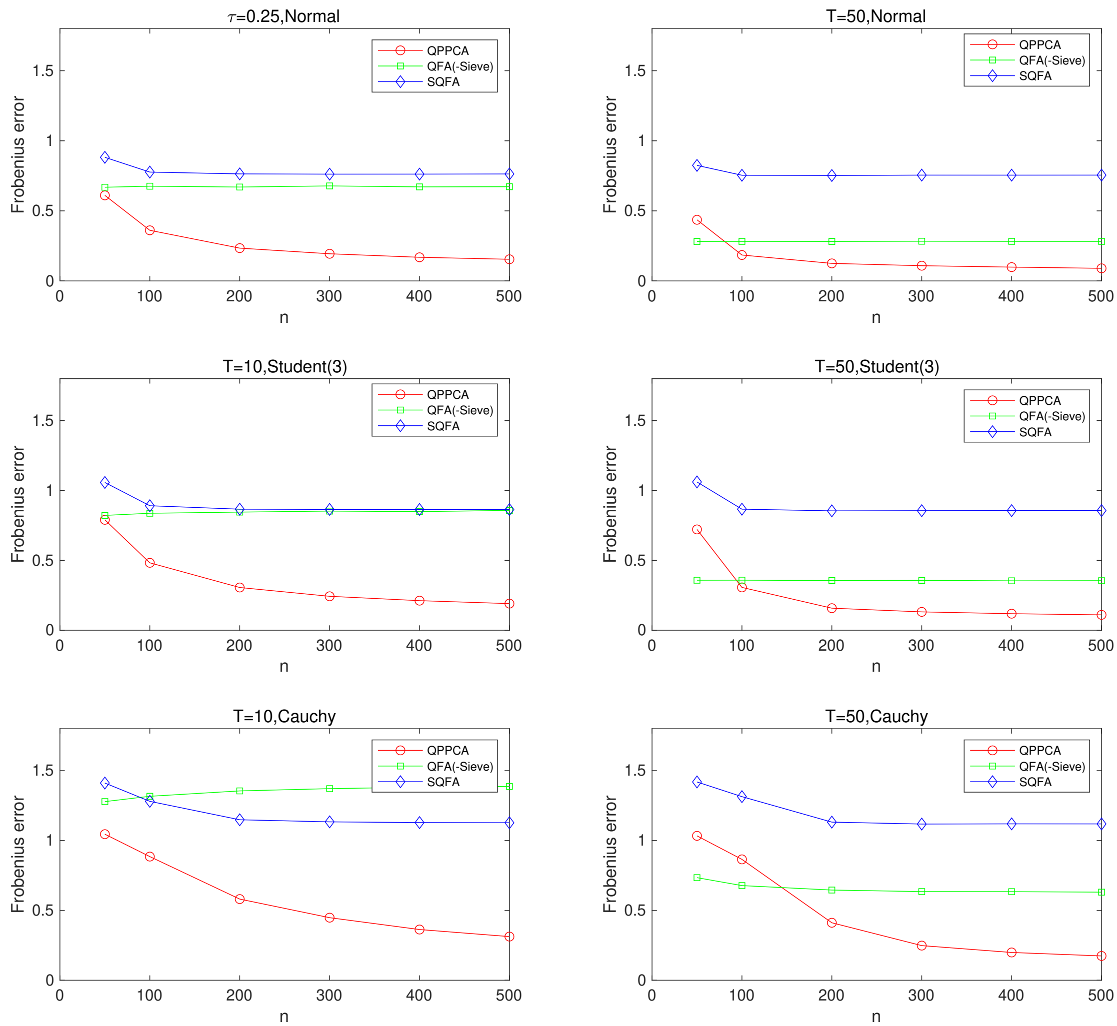
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{F}_0 - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}_0\|/\sqrt{T}$ from 1000 repetitions.

Figure B.3: Estimation of loading matrices: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25$, with $D = R$.



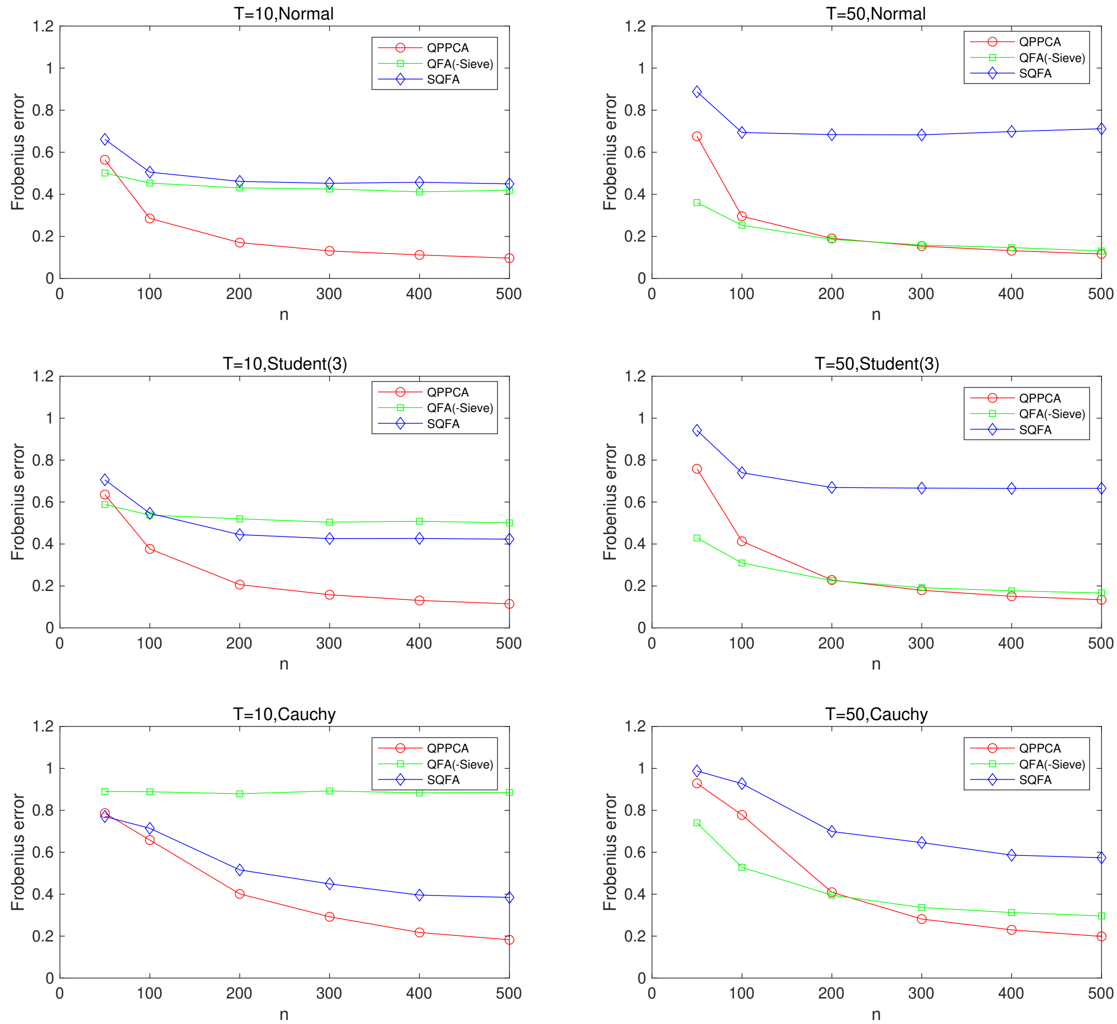
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrices are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.25$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$ from 1000 repetitions.

Figure B.4: Estimation of loading matrices: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$, with $D = R$.



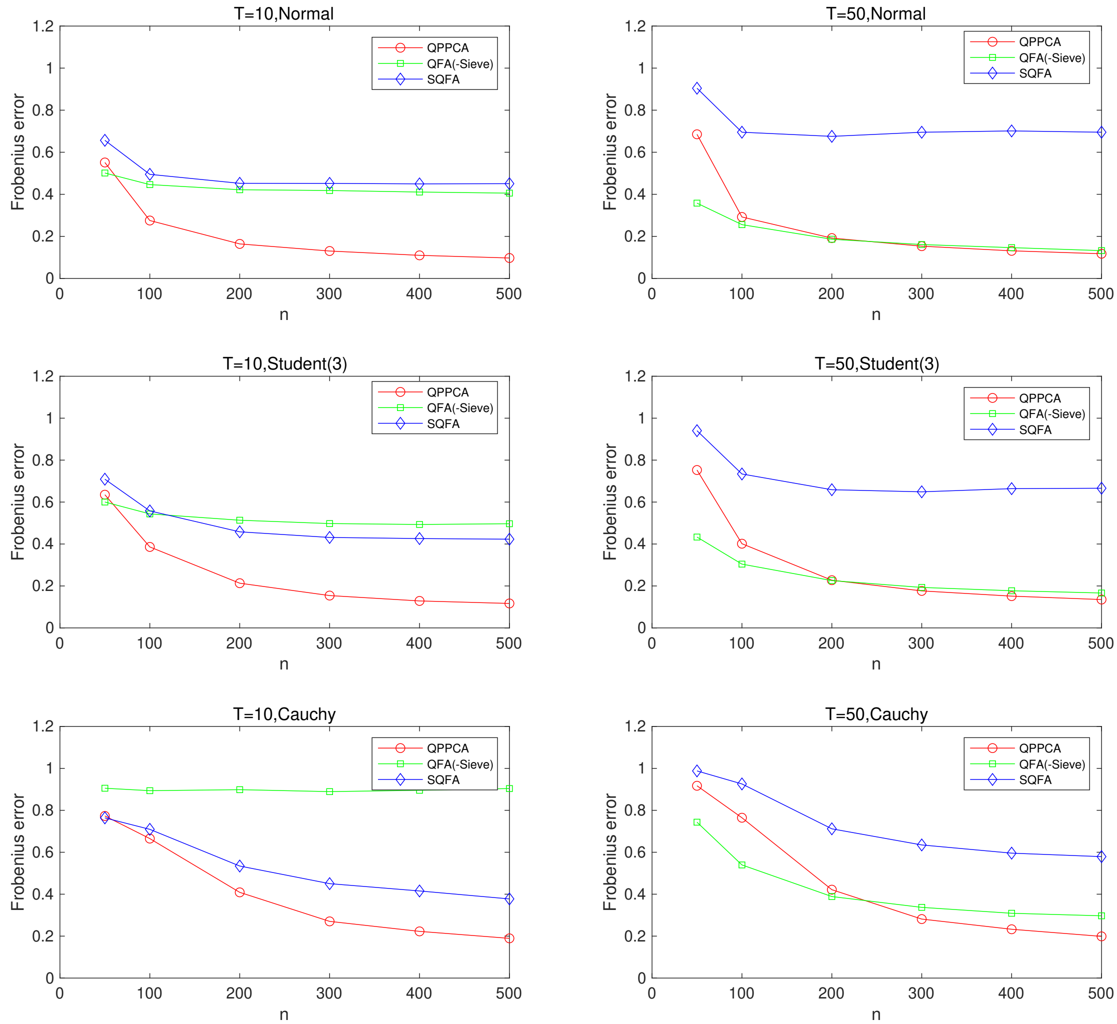
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrices are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$ from 1000 repetitions.

Figure B.5: Estimation of factors: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25$, with $D > R$.



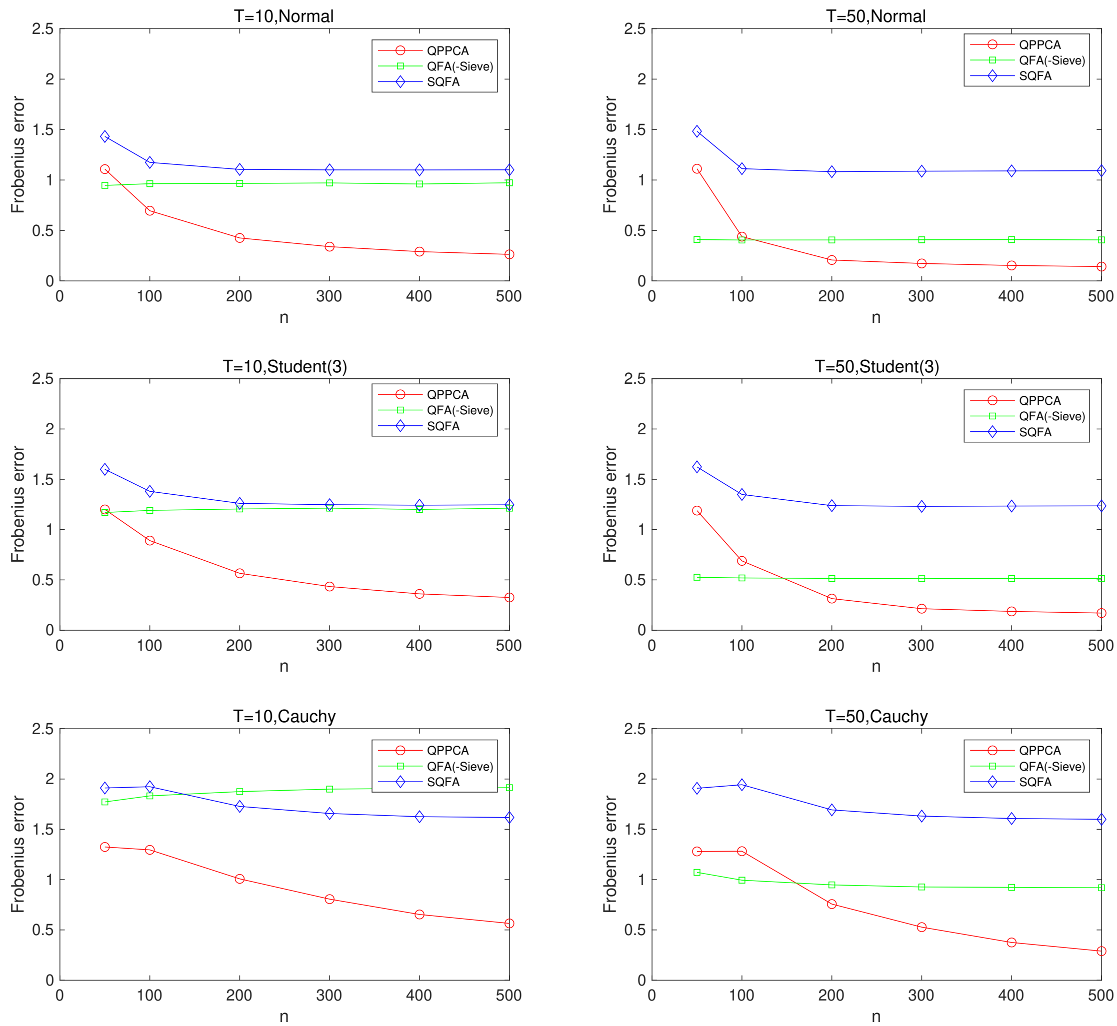
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 3 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3$) are independently drawn from the uniform distribution: $U[-1,1]$. $g_{11}(x) = g_{12}(x) = g_{13}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = g_{23}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}) + g_{13}(x_{i3})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2}) + g_{23}(x_{i3})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.25$. The reported results are the average Frobenius errors: $\|\boldsymbol{F}_0 - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}_0\|/\sqrt{T}$ from 1000 repetitions.

Figure B.6: Estimation of factors: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$, with $D > R$.



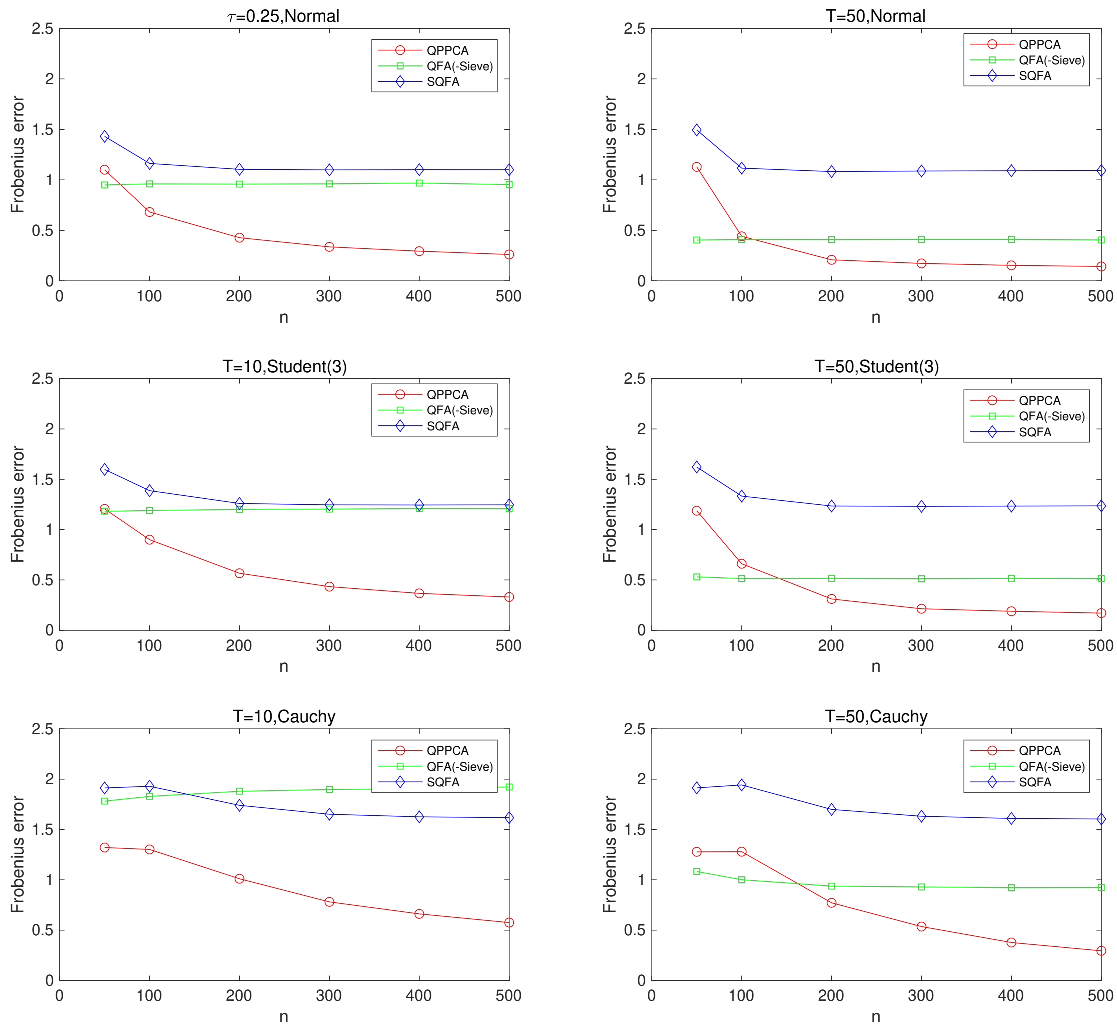
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 3 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = g_{13}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = g_{23}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}) + g_{13}(x_{i3})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2}) + g_{23}(x_{i3})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{F}_0 - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}_0\|/\sqrt{T}$ from 1000 repetitions.

Figure B.7: Estimation of loading matrices: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25$, with $D > R$.



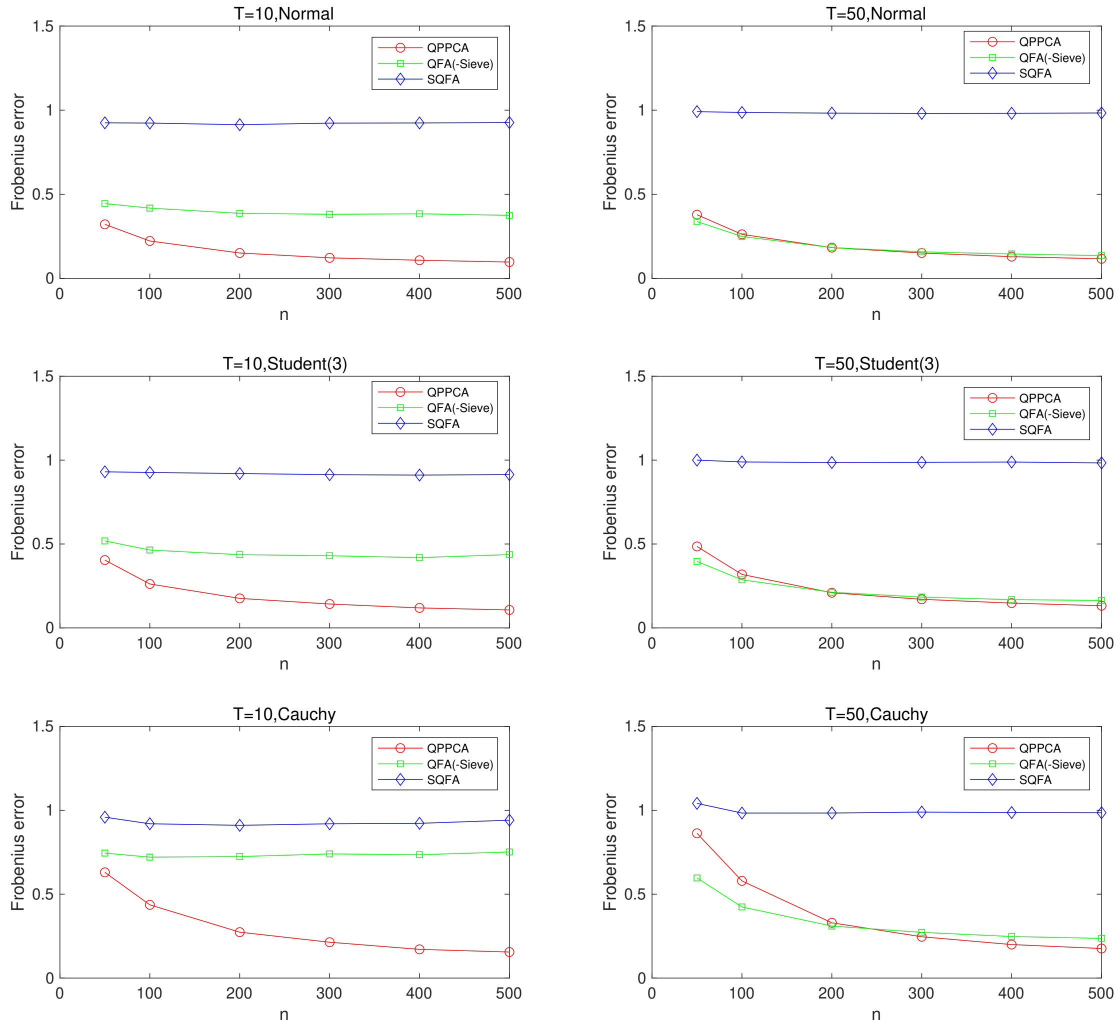
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 3 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = g_{13}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = g_{23}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}) + g_{13}(x_{i3})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2}) + g_{23}(x_{i3})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrices are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.25$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$ from 1000 repetitions.

Figure B.8: Estimation of loading matrices: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$, with $D > R$.



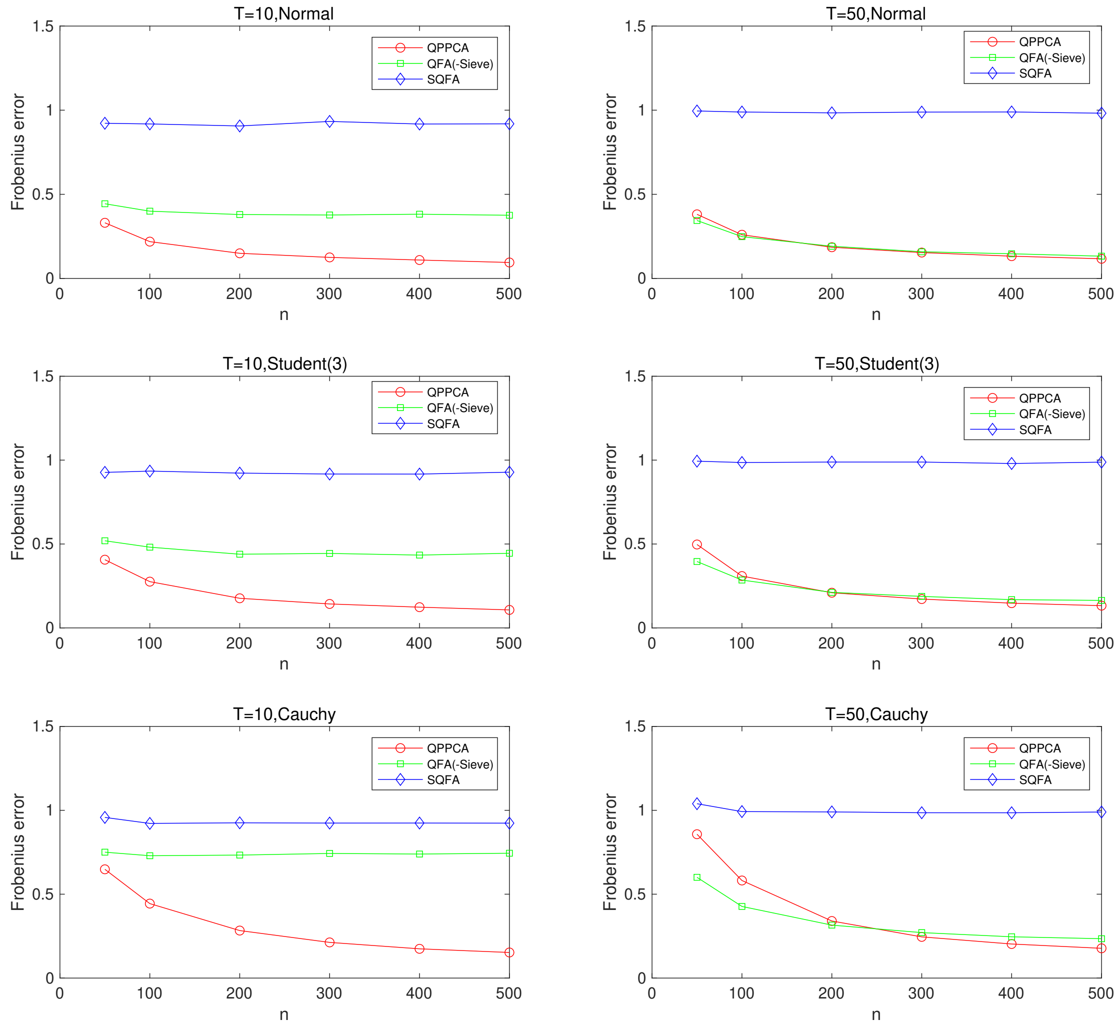
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 3 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = g_{13}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = g_{23}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}) + g_{13}(x_{i3})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2}) + g_{23}(x_{i3})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrices are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$ from 1000 repetitions.

Figure B.9: Estimation of factors: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25$, with $D < R$.



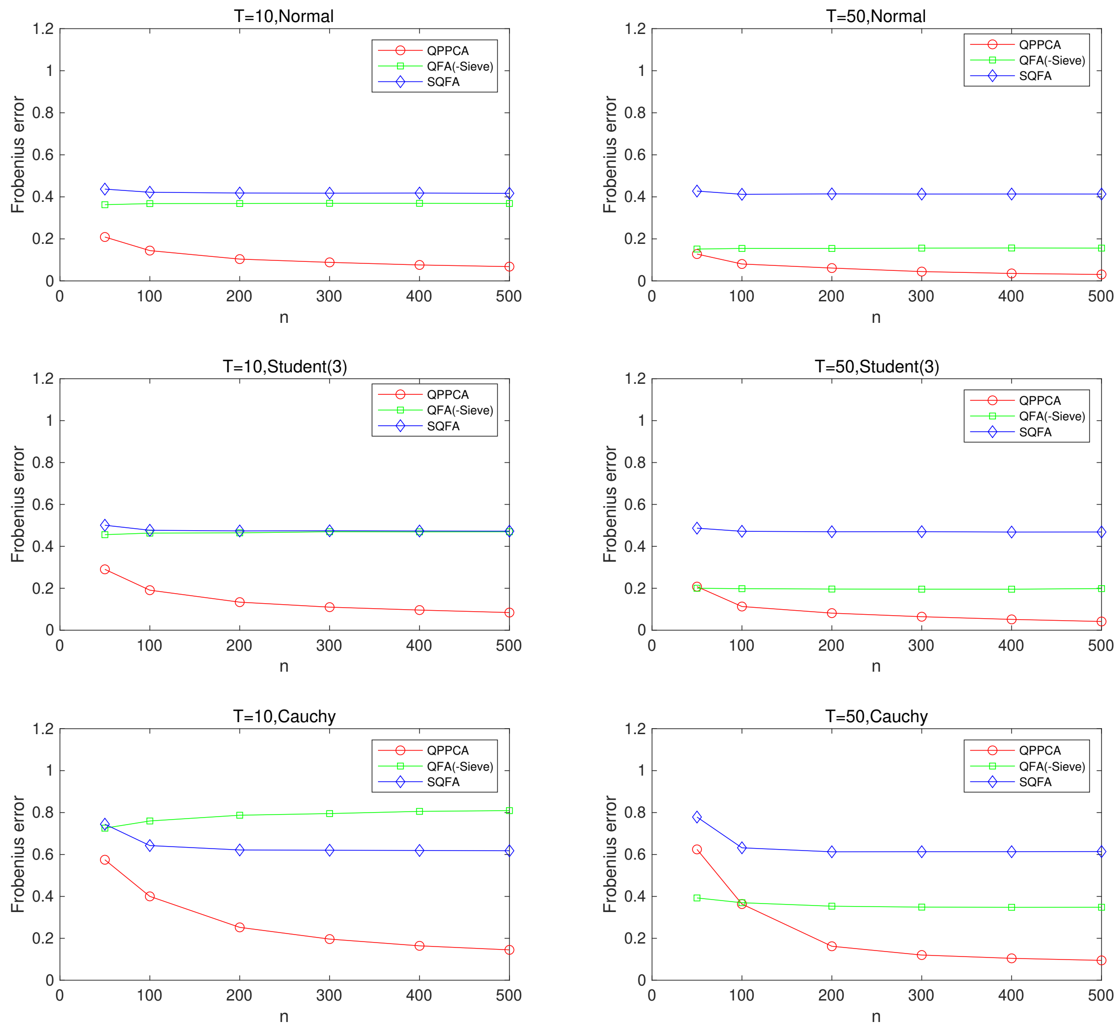
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 1 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x)$ and $g_{21}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1})$, $\lambda_{i2} = g_{21}(x_{i1})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.25$. The reported results are the average Frobenius errors: $\|\boldsymbol{F}_0 - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}_0\|/\sqrt{T}$ from 1000 repetitions.

Figure B.10: Estimation of factors: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$, with $D < R$.



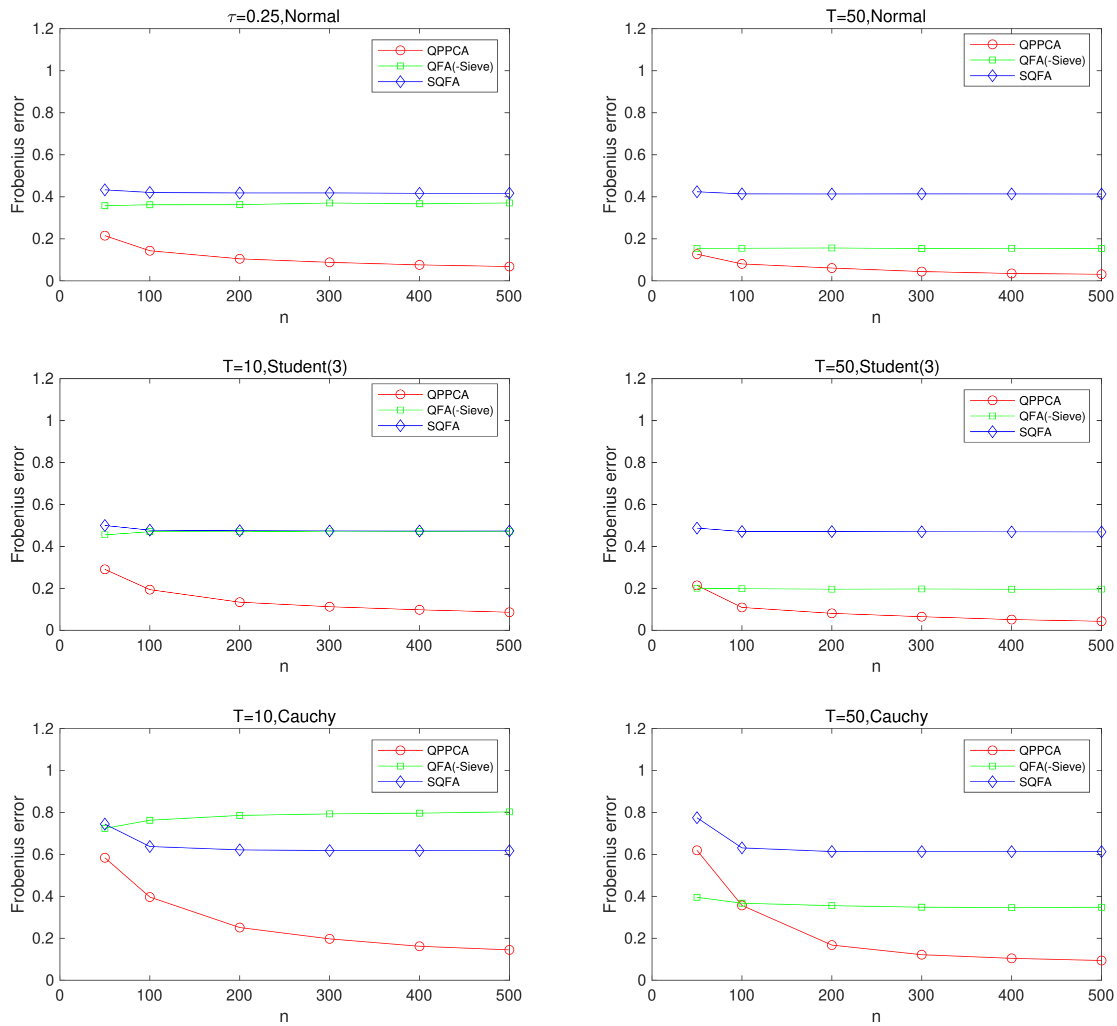
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 1 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x)$ and $g_{21}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1})$, $\lambda_{i2} = g_{21}(x_{i1})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{F}_0 - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}_0\|/\sqrt{T}$ from 1000 repetitions.

Figure B.11: Estimation of loading matrices: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.25$, with $D < R$.



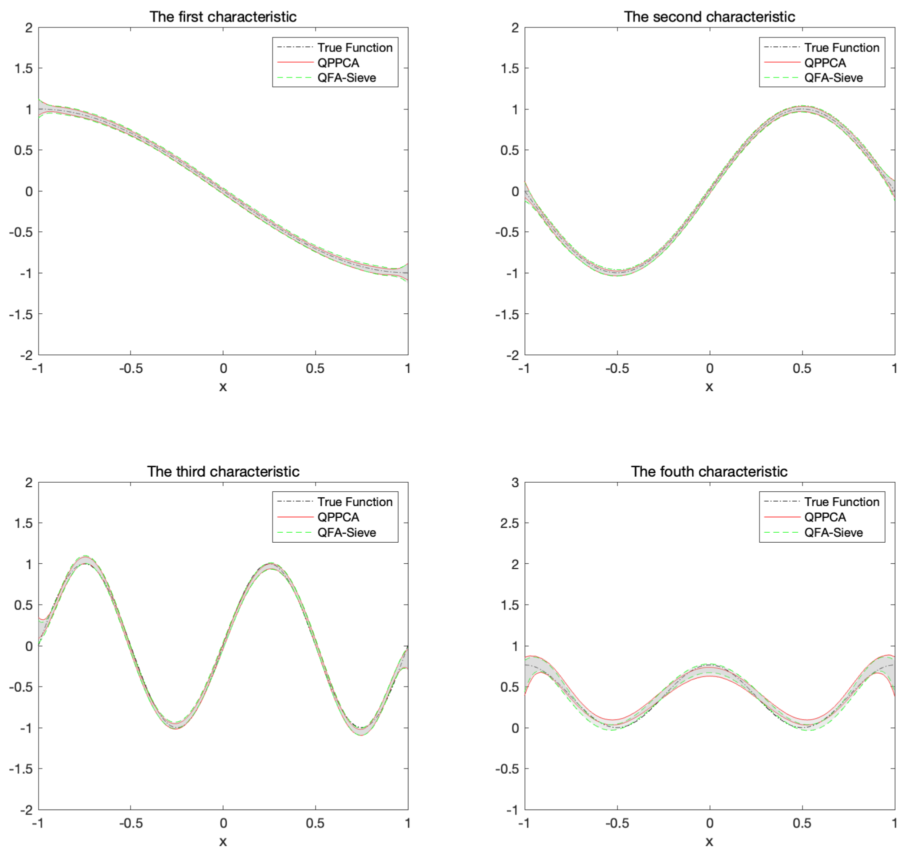
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 1 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x)$ and $g_{21}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1})$, $\lambda_{i2} = g_{21}(x_{i1})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrices are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.25$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$ from 1000 repetitions.

Figure B.12: Estimation of loading matrices: QPPCA, QFA-Sieve and SQFA, at $\tau = 0.75$, with $D < R$.



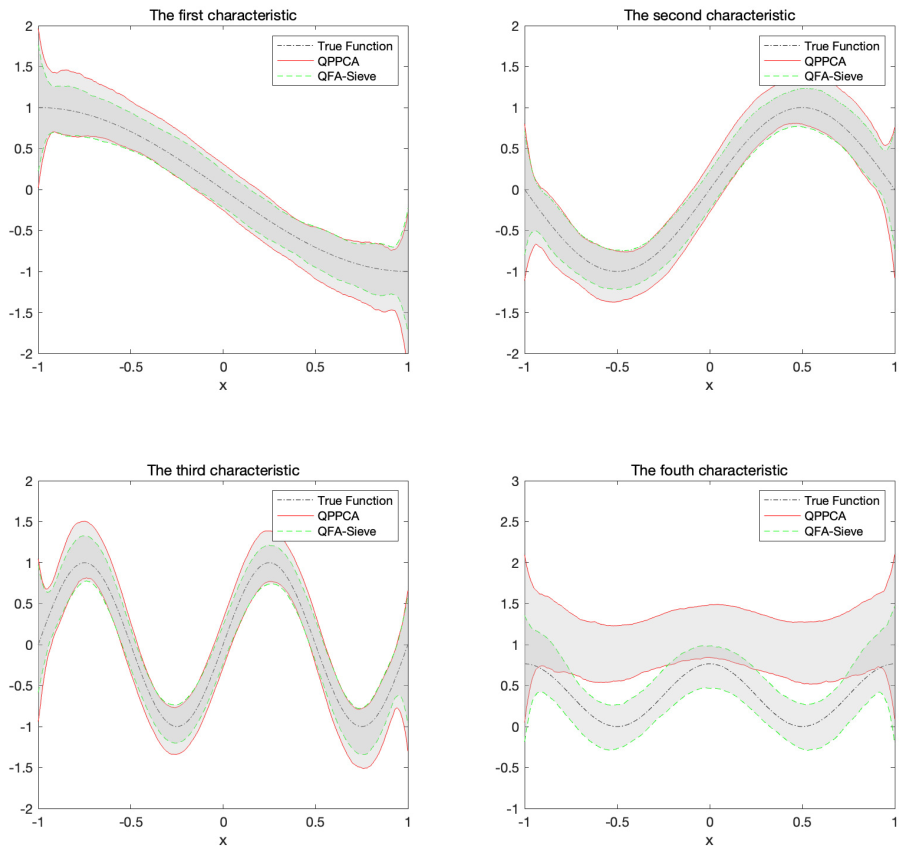
Note: The DGP is $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 1 and all characteristics x_{id} ($i = 1, ..., n$ and $d = 1$) are independently drawn from the uniform distribution: $U[-1,1]$. $g_{11}(x) = sin(2\pi x)$ and $g_{21}(x) = cos^2(\pi x)$. The factor loadings are generated as $\lambda_{i1} = g_{11}(x_{i1})$, $\lambda_{i2} = g_{21}(x_{i1})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The loading matrices are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$ from 1000 repetitions.

Figure B.13: Estimated Loading functions at $\tau = 0.75$, no missing variables: QPPCA and QFA-Sieve



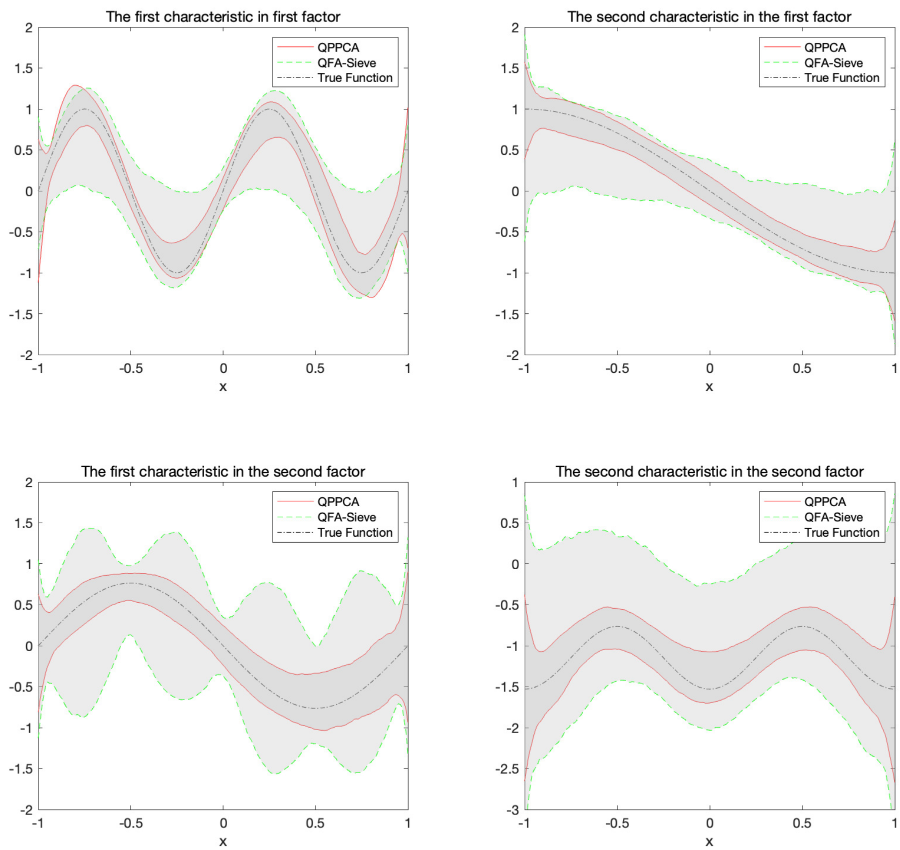
Note: $n = 355$, $T = 62$, $R = 1$ and $D = 4$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0,1)$, x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3, 4$) are independently drawn from the uniform distribution: $U[-1,1]$. $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black dash-dot lines) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red solid lines) and QFA-Sieve (the green dashed lines) from 1000 repetitions.

Figure B.14: Estimated Loading functions at $\tau = 0.75$, one missing variable: QPPCA and QFA-Sieve



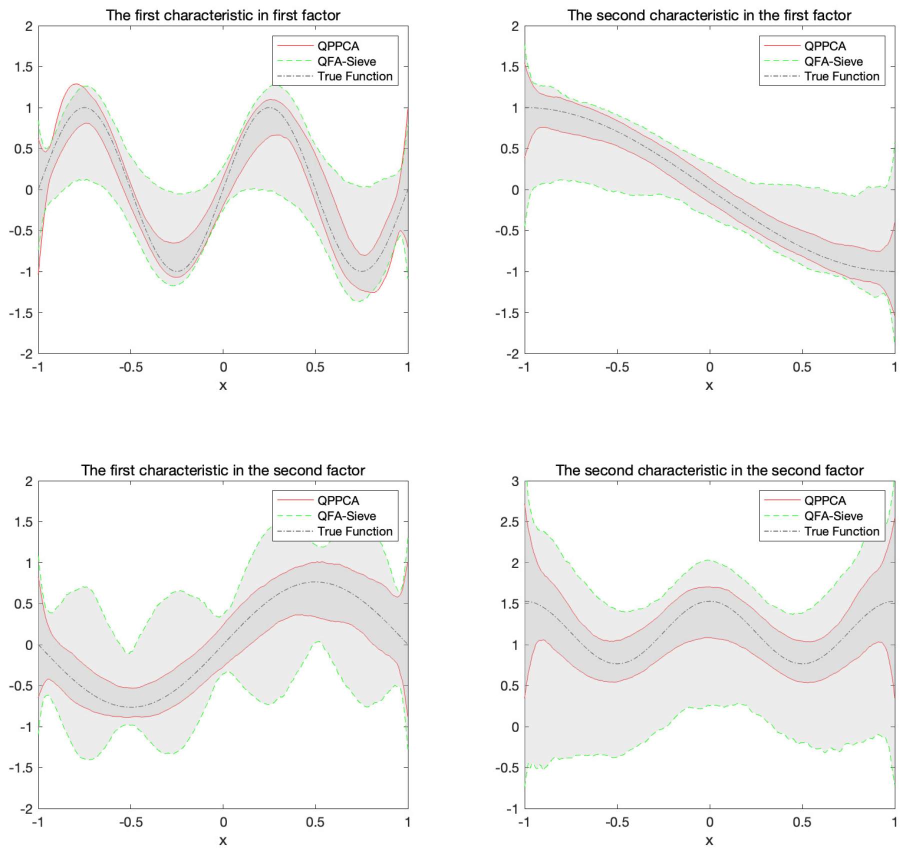
Note: $n = 355$, $T = 62$, $R = 1$ and $D = 5$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3}) + g_5(x_{i5})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0, 1)$, x_{id} ($i = 1, ..., n$ and $d = 1, \ldots, 5$) are independently drawn from the uniform distribution: $U[-1, 1]$. Assume that only the first 4 characteristics are observed while x_{i5} is unobserved. $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$ and $g_5(x) = 1.5cos(\pi x)$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black dash-dot lines) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red solid lines) and QFA-Sieve (the green dashed lines) from 1000 repetitions.

Figure B.15: Estimated Loading functions at $\tau = 0.25$, 2 factors, no missing variables: QPPCA and QFA-Sieve



Note: $n = 355$, $T = 10$, $R = 2$ and $D = 2$. The DGP is: $y_{it} = (g_{11}(x_{i1}) + g_{12}(x_{i2})) \cdot f_{1t} + (g_{21}(x_{i1}) + g_{22}(x_{i2}) \cdot f_{2t} \cdot u_{it}$, where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0,1)$, x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1,1]$. $g_{11}(x) = sin(2\pi x)$, $g_{12}(x) = -sin(0.5\pi x)$, $g_{21}(x) = sin(\pi x)$, $g_{22}(x) = cos^2(\pi x)+1$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black dash-dot line) at $\tau = 0.25$, and the empirical pointwise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red solid lines) and QFA-Sieve (the green dashed lines) from 1000 repetitions.

Figure B.16: Estimated Loading functions at $\tau = 0.75$, 2 factors, no missing variables: QPPCA and QFA-Sieve



Note: $n = 355$, $T = 10$, $R = 2$ and $D = 2$. The DGP is: $y_{it} = (g_{11}(x_{i1}) + g_{12}(x_{i2})) \cdot f_{1t} + (g_{21}(x_{i1}) + g_{22}(x_{i2}) \cdot f_{2t} \cdot u_{it}$, where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0, 1)$, x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x)$, $g_{12}(x) = -sin(0.5\pi x)$, $g_{21}(x) = sin(\pi x)$, $g_{22}(x) = cos^2(\pi x) + 1$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black dash-dot line) at $\tau = 0.75$, and the empirical pointwise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red solid lines) and QFA-Sieve (the green dashed lines) from 1000 repetitions.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bonis, T. (2020). Stein's method for normal approximation in wasserstein distances with application to the multivariate central limit theorem. *Probability Theory and Related Fields 178*(3), 827–860.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics 6*, 5549–5632.

Chen, X. and X. Shen (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 289–314.

Davidson, K. R. and S. J. Szarek (2001). Local operator theory, random matrices and banach spaces. In *Handbook of the geometry of Banach spaces*, Volume 1, pp. 317–366. Elsevier.

Horowitz, J. L. and S. Lee (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association 100*(472), 1238–1249.

Kato, K., A. F. Galvao, and G. V. Montes-Rojas (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics 170*(1), 76–91.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: isoperimetry and processes*, Volume 23. Springer Science & Business Media.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Rao, C. R., C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao (1973). *Linear statistical inference and its applications*, Volume 2. Wiley New York.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.