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] *CEMFI, haozi.pan@cemfi.edu.es*

June 20, 2024

A Proofs of the Main Results

Proof of Proposition 1:

Proof. For any $\theta \in \Theta$, define $K(\theta, \theta_{0t}) = \mathbb{E}(L_n(\theta)) = \mathbb{E}[l(\theta, y_{it}, \boldsymbol{x}_i)]$. Under Assumption 1(iv), it can be shown that $K(\theta, \theta_{0t}) \asymp d(\theta, \theta_{0t})^2$. For the finite-dimensional linear sieve spaces Θ_n, it can be shown that Condition A.3 of Chen and Shen (1998) is satisfied with $\delta_n = \sqrt{k_n/n}$ (see Section 3.3 of Chen (2007)). By the definition of d and the properties of the check function, it is easy to see that,[1]

$$\sup_{\theta \in \Theta_n, d(\theta, \theta_{0t}) \leq \varepsilon} \operatorname{Var}\left[l(\theta, y_{it}, \boldsymbol{x}_i)\right] \leq \sup_{\theta \in \Theta_n, d(\theta, \theta_{0t}) \leq \varepsilon} \mathbb{E}\left[l(\theta, y_{it}, \boldsymbol{x}_i)\right]^2$$
$$\lesssim \sup_{\theta \in \Theta_n, d(\theta, \theta_{0t}) \leq \varepsilon} \mathbb{E}\left(\theta(\boldsymbol{x}_i) - \theta_{0t}(\boldsymbol{x}_i)\right)^2 \leq \varepsilon^2.$$

Thus, Condition A.2 of Chen and Shen (1998) is also satisfied. By Assumption 1(iii) we have $\sup_{\theta \in \Theta} |l(\theta, y_{it}, \boldsymbol{x}_i)| \lesssim \sup_{\theta \in \Theta} \sup_{\mathcal{X}} |\theta(\boldsymbol{x}) - \theta_{0t}(\boldsymbol{x})| < \infty$. Assumption 1(ii) implies that $d(\pi_n \theta_{0t}, \theta_{0t}) = \sqrt{\mathbb{E}\left(\pi_n \theta_{0t}(\boldsymbol{x}_i) - \theta_{0t}(\boldsymbol{x}_i)\right)^2} = O(k_n^{-\alpha})$. Therefore, it follows from Corollary 1 of Chen and Shen (1998) that

$$P\left[\max_t d(\hat{\theta}_{nt}, \theta_{0t}) \geq C\varepsilon_{nT}\right] \leq \sum_{t=1}^{T} P\left[d(\hat{\theta}_{nt}, \theta_{0t}) \geq C\varepsilon_{nT}\right] \leq c_1 \exp\left\{C^2 \ln T (1 - c_2 n \varepsilon_n^2)\right\}$$

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

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

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

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

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

where $c_1 > 0$, and the $O_P(\varepsilon_{nT}k_n^{-\alpha})$ in the above equation is uniform in t. It then 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

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

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. $\square$

Proof of Theorem 1:

Proof. Write $\hat{\boldsymbol{Y}} = \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}' + \hat{\boldsymbol{V}}$ where $\hat{\boldsymbol{V}}$ is as defined in Lemma 1. Let $\boldsymbol{\Omega}_R$ be the diagonal matrix whose elements are the eigenvalues of $\boldsymbol{\Sigma}_g \cdot \boldsymbol{F}'\boldsymbol{F}/T$. Note that

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

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

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

By the Wielandt-Hoffman inequality, we have $\|\hat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\| = o_P(1)$. 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

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

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

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

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

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

and thus

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

Then the second part of Theorem 1 follows.

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

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

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

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

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 2. *Let* $\xi_{it} = \theta_{0t}(\boldsymbol{x}_i) - \pi_n\theta_{0t}(\boldsymbol{x}_i) = \boldsymbol{g}(\boldsymbol{x}_i)'\boldsymbol{f}_t - \boldsymbol{a}_{0t}'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ *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 by 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,

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

Moreover, one can write

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

which gives (A.13).

Step 5: Define:

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

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

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

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

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

In what follows, we will show that

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

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

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

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

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

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

We will show that

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

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

Define $\mathcal{H}_{\varepsilon_{nT}} = \{h(\boldsymbol{a}, \boldsymbol{x}_i) \equiv \Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i)] : \|\boldsymbol{a} - \boldsymbol{a}_{0t}\| \leq C\varepsilon_{nT}\}$, 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

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

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

$N(\mathcal{H}_{\varepsilon_{nT}}, L_2(Q), 2\epsilon) \leq (A/\epsilon)^{c_1 k_n}$ for some bounded constant c_1 and $A \geq 3\sqrt{e}$ that do not depend on t and j. Moreover, it is easy to show that $\sup_{h\in\mathcal{H}_{\varepsilon_{nT}}} \mathbb{E}[h^2(\boldsymbol{a}, \boldsymbol{x}_i)] \leq c_2^2\sqrt{k_n}\varepsilon_n$ for some bounded constant c_2. Then, applying Proposition B.1 of Kato et al. (2012), we have

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

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)

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

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

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

From (A.2) and Lemma 1 we have

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

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

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

Thus,

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

It then follows from Theorem 1 and Lemma 1 that

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

The above inequality and (A.20) imply that

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

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

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

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

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

This completes the proof. □

Lemma 3. *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 2. Therefore, it is omitted to save space. □

Proof of Theorem 3:

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

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

Moreover,

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

Thus, by Lemma 1 and Theorem 1,

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

It then follows from Lemma 3 that

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

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

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

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

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

Thus, we have

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

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

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

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

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

Thus,

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

and

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

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

$$(\hat{\boldsymbol{H}}')^{-1}\sqrt{n}(\tilde{\boldsymbol{f}}_t - \tilde{\boldsymbol{H}}'\boldsymbol{f}_t) = \boldsymbol{\Sigma}_g^{-1}\mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\psi_{it}(\boldsymbol{x}_i)\right)$$
$$+ O_P(\varepsilon_{nT}k_n^{1/2}) + O_P(n^{1/2}k_n^{-\alpha}) + O_P(n^{1/2}\eta_{nT}).$$

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. □

Proof of Theorem 5:

Proof. First, note that

$$
\begin{aligned}
& \|\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}'\boldsymbol{\Phi}(\boldsymbol{X})' - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\|/(nT) \\
\leq\ & 2\|\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\|/\sqrt{nT} \cdot \|\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\|/\sqrt{nT} + \|\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\|^2/(nT) \\
=\ & O_P(1) \cdot \|\hat{\boldsymbol{V}}\|/\sqrt{nT} + \|\hat{\boldsymbol{V}}\|^2/(nT).
\end{aligned}
$$

It then follows from Lemma 1(ii) that

$$
\|\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}'\boldsymbol{\Phi}(\boldsymbol{X})' - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\|/(nT) = O_P(\varepsilon_{nT}). \tag{A.23}
$$

Second, Assumption 2(iii) and (iv) imply that the largest R eigenvalues of $\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'/(nT)$, which are also the R eigenvalues of $(\boldsymbol{F}'\boldsymbol{F}/T)\cdot\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})/n$, converge in probability to the R eigenvalues of $(\boldsymbol{F}'\boldsymbol{F}/T)\cdot\boldsymbol{\Sigma}_g$. Also, note that the remaining eigenvalues of $\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'/(nT)$ are all 0, it then follows from (A.23) and the Wielandt-Hoffman inequality that $\hat{\rho}_j = O_P(\varepsilon_{nT})$ for $j = R+1, \ldots, \bar{R}$, and $\hat{\rho}_j$ converges in probability in some positive constant for $j = 1, \ldots, R$. The desired result then follows because $P[\hat{\rho}_j > p_n] \to 1$ for $j = 1, \ldots, R$ and $P[\hat{\rho}_j > p_n] \to 0$ for $j = R+1, \ldots, \bar{R}$. □

B Additional Simulation Results

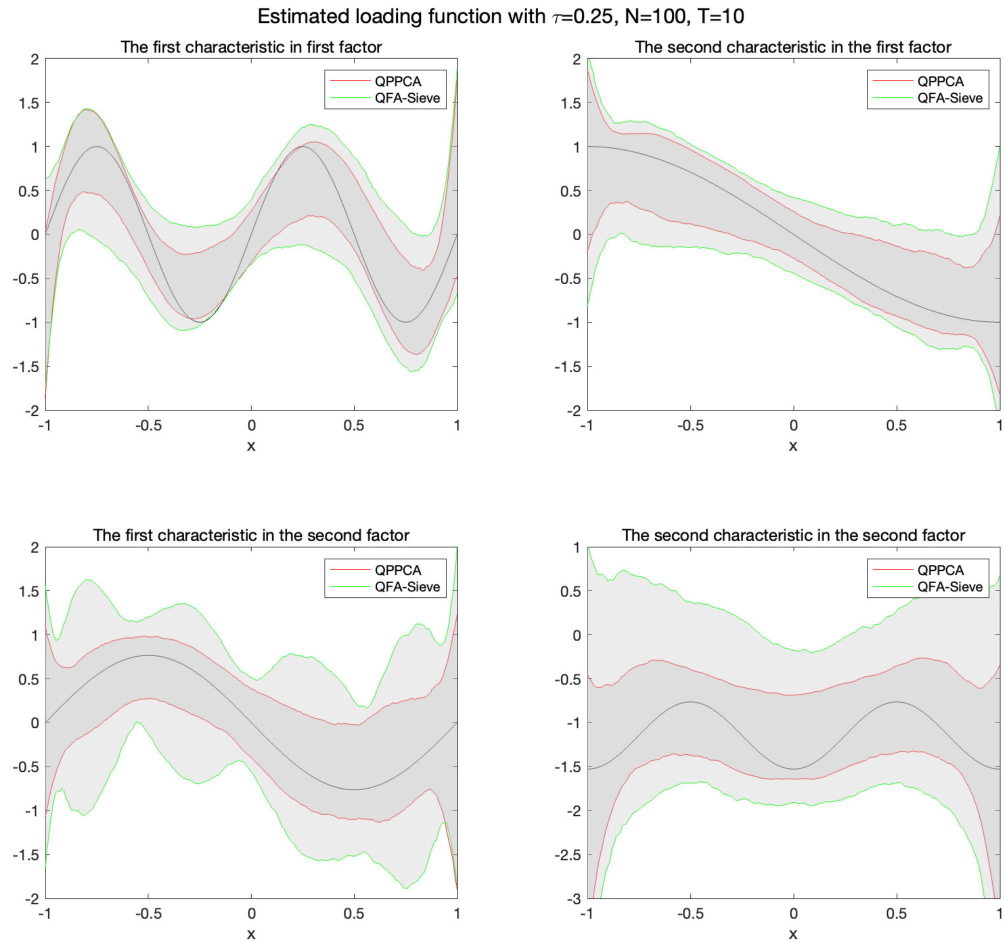
The goal of this section is to compare the estimated factor loadings using QPPCA and QFA-Sieve when T is small. To this end, consider the following DGP:

$$
y_{it} = (\sin(2\pi \cdot x_{i1}) - \sin(0.5\pi \cdot x_{i2}))f_{1t} + (\sin(\pi \cdot x_{i1}) + \cos^2(\pi \cdot x_{i2}) + 1)f_{2t} \cdot u_{it}
$$

where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n,$ and $d = 1, 2)$ are independently drawn from the uniform distribution $U[-1, 1]$ and $\{u_{it}\}(i = 1, 2, ..., n, t = 1, 2, ...T)$ are independently drawn from the $t(3)$ distribution. Moreover, we fix $T = 10$ and let $n = 100, 300, 500$.

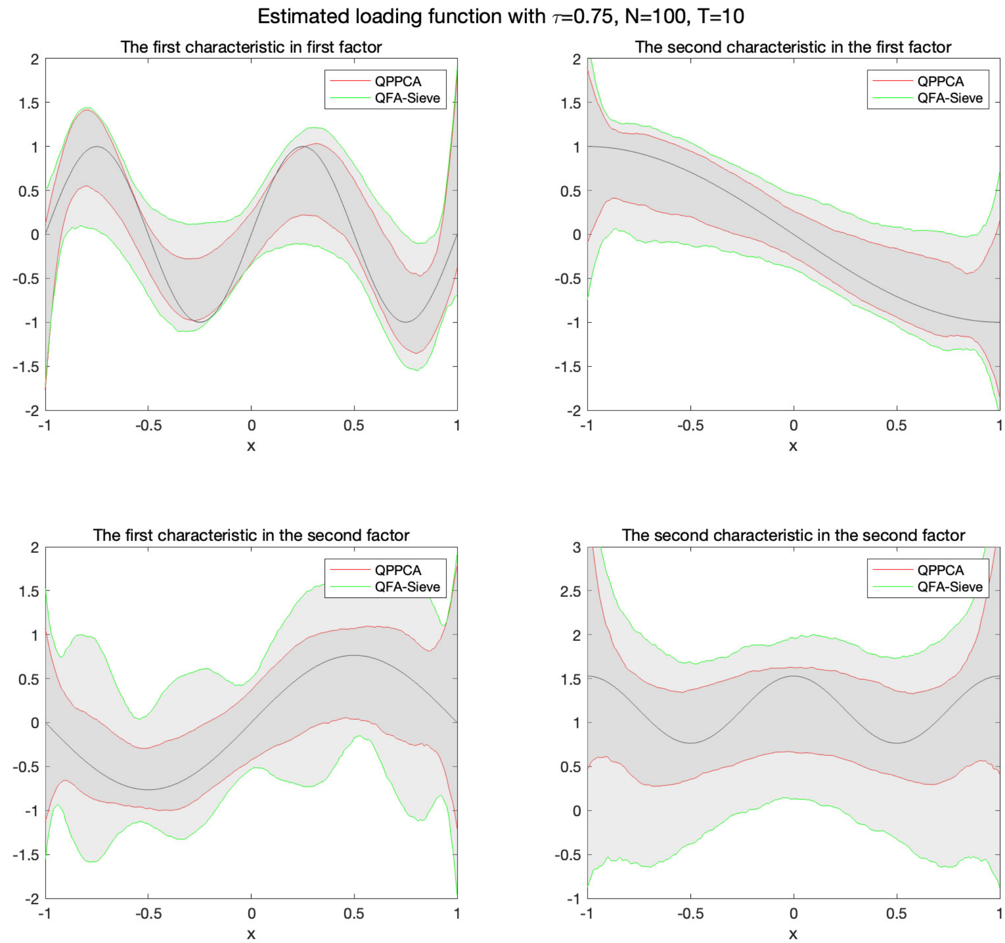
In Figures A.1 to A.6, the true loading functions (black lines) and the 5% and 95% pointwise quantiles of the QPPCA (red lines) and QFA-Sieve (green lines) estimators from 1000 replications at $\tau = 0.25, 0.75$ are plotted. It can be seen that, in almost all cases, the confidence intervals of the QPPCA estimators contain the true loading functions, and their widths shrink as n increases. In contrast, the confidence intervals of the QFA-Sieve estimators are much wider and they do not shrink as n increases. This is as expected since the convergence rate of the QFA estimator is determined by $\min\{\sqrt{n}, \sqrt{T}\}$. In general, we can conclude that the QPPCA estimators perform much better than the QFA-Sieve estimators when T is not big.

Figure A.1: Estimated Loading functions with $T = 10$: QPPCA and QFA-Sieve



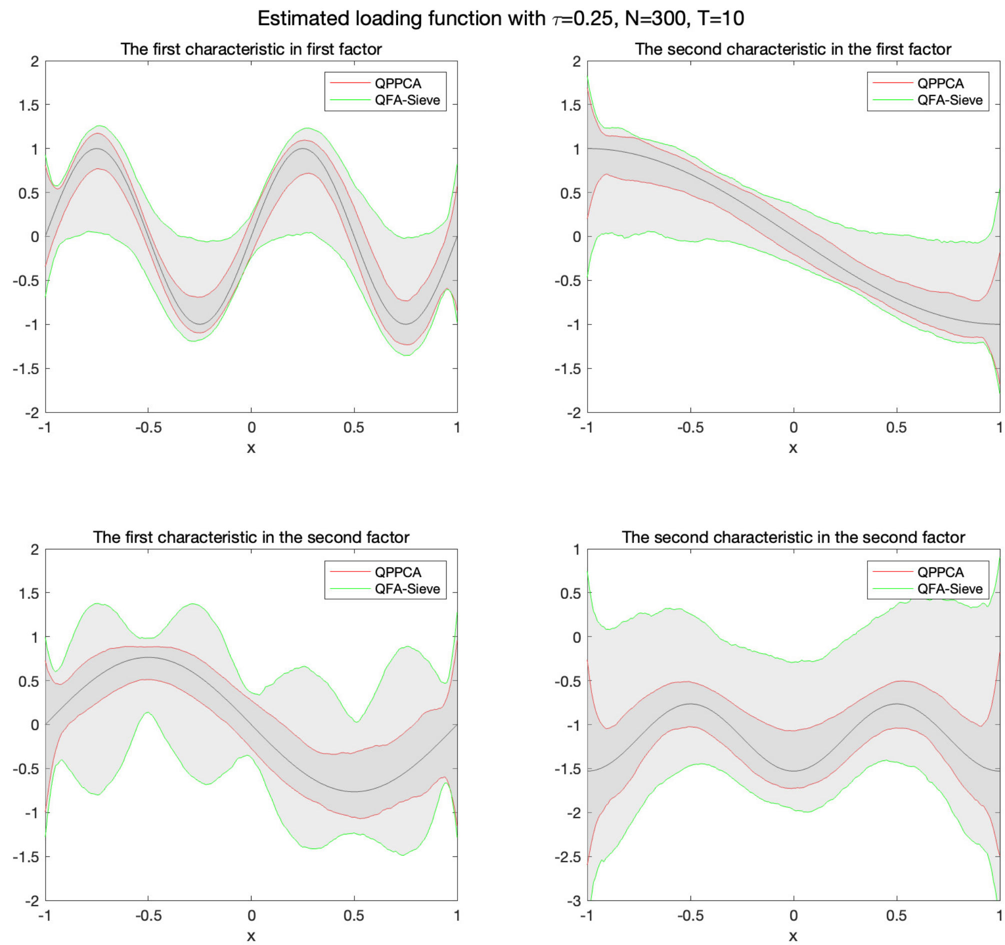
Note: $n = 100$, $T = 10$, $R = 2$ and $D = 2$. The DGP is: $y_{it} = (g_{11}(x_{i1}) + g_{12}(x_{i2})) \cdot f_{1t} + (g_{21}(x_{i1}) + g_{22}(x_{i2})) \cdot f_{2t} \cdot u_{it}$, where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n,$ and $d = 1, 2)$ 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 line) at $\tau = 0.25$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure A.2: Estimated Loading functions with $T = 10$: QPPCA and QFA-Sieve



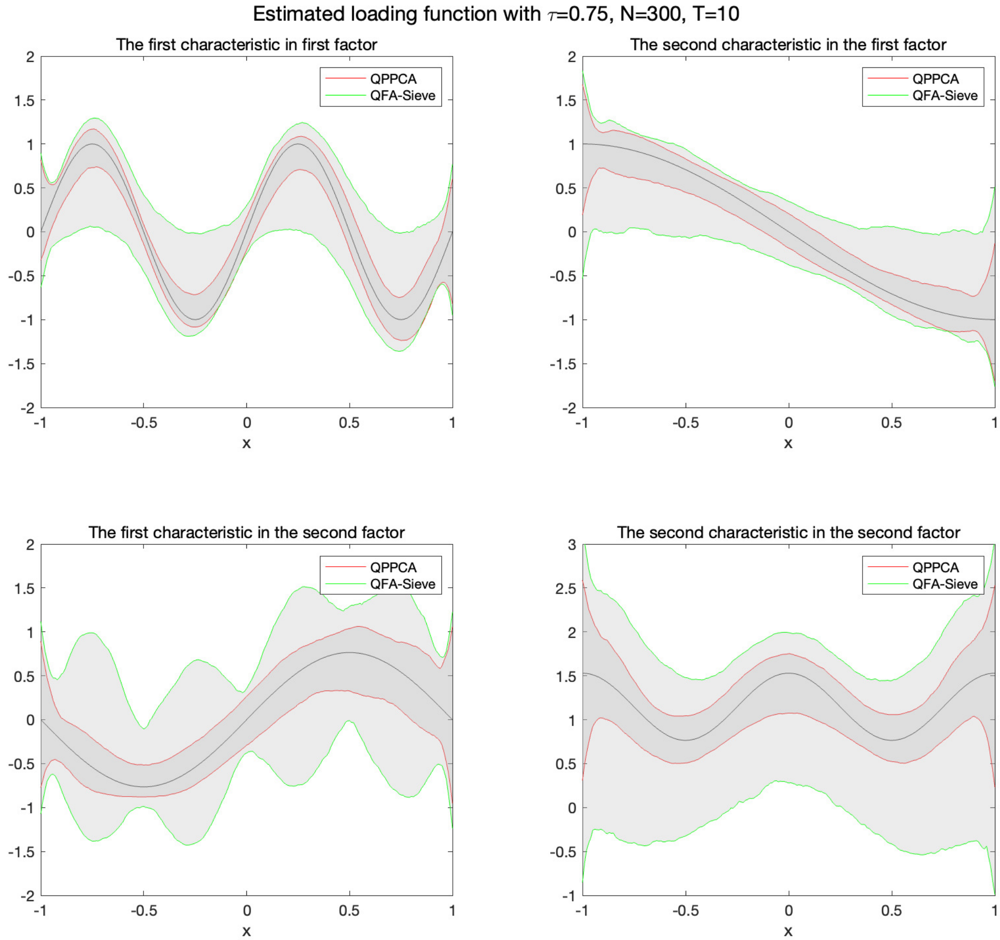
Note: $n = 100$, $T = 10$, $R = 2$ and $D = 2$. The DGP is: $y_{it} = (g_{11}(x_{i1}) + g_{12}(x_{i2})) \cdot f_{1t} + (g_{21}(x_{i1}) + g_{22}(x_{i2})) \cdot f_{2t} \cdot u_{it}$, where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n$, and $d = 1, 2)$ 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 line) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure A.3: Estimated Loading functions with $T = 10$: QPPCA and QFA-Sieve



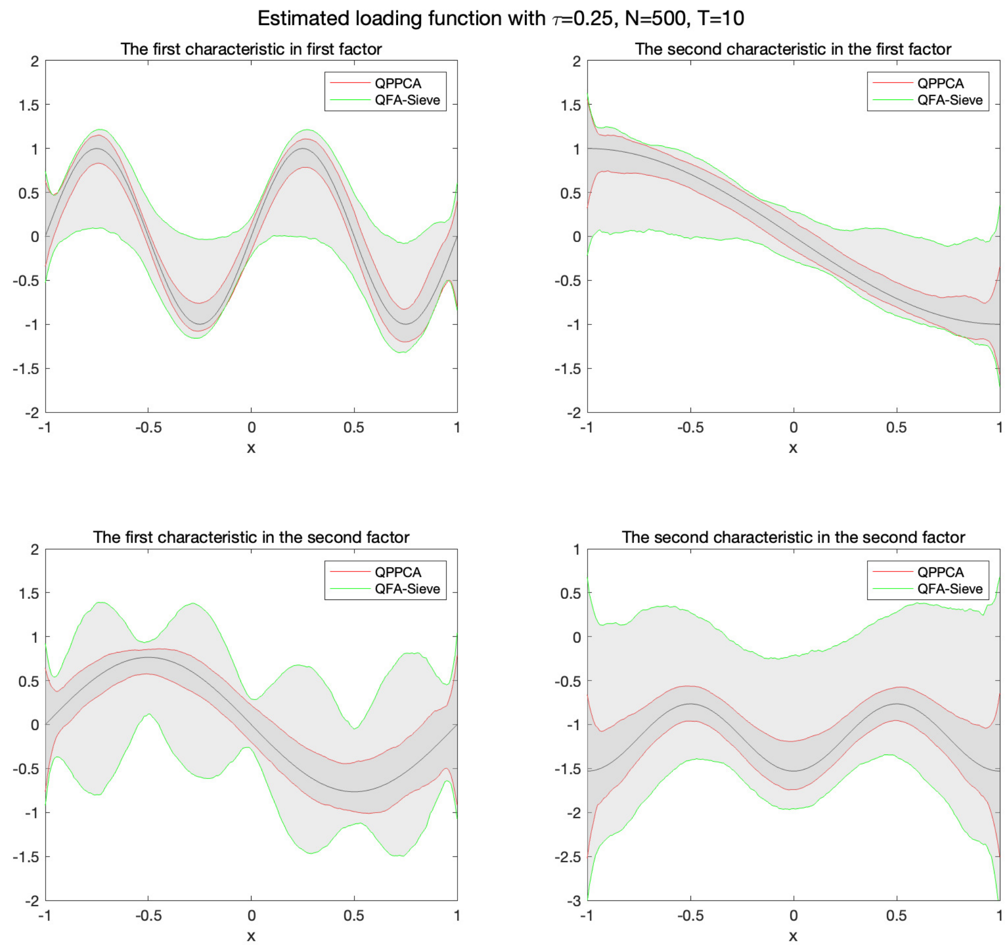
Note: $n = 300$, $T = 10$, $R = 2$ and $D = 2$. The DGP is: $y_{it} = (g_{11}(x_{i1}) + g_{12}(x_{i2})) \cdot f_{1t} + (g_{21}(x_{i1}) + g_{22}(x_{i2})) \cdot f_{2t} \cdot u_{it}$, where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n$, and $d = 1, 2)$ 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 line) at $\tau = 0.25$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure A.4: Estimated Loading functions with $T = 10$: QPPCA and QFA-Sieve



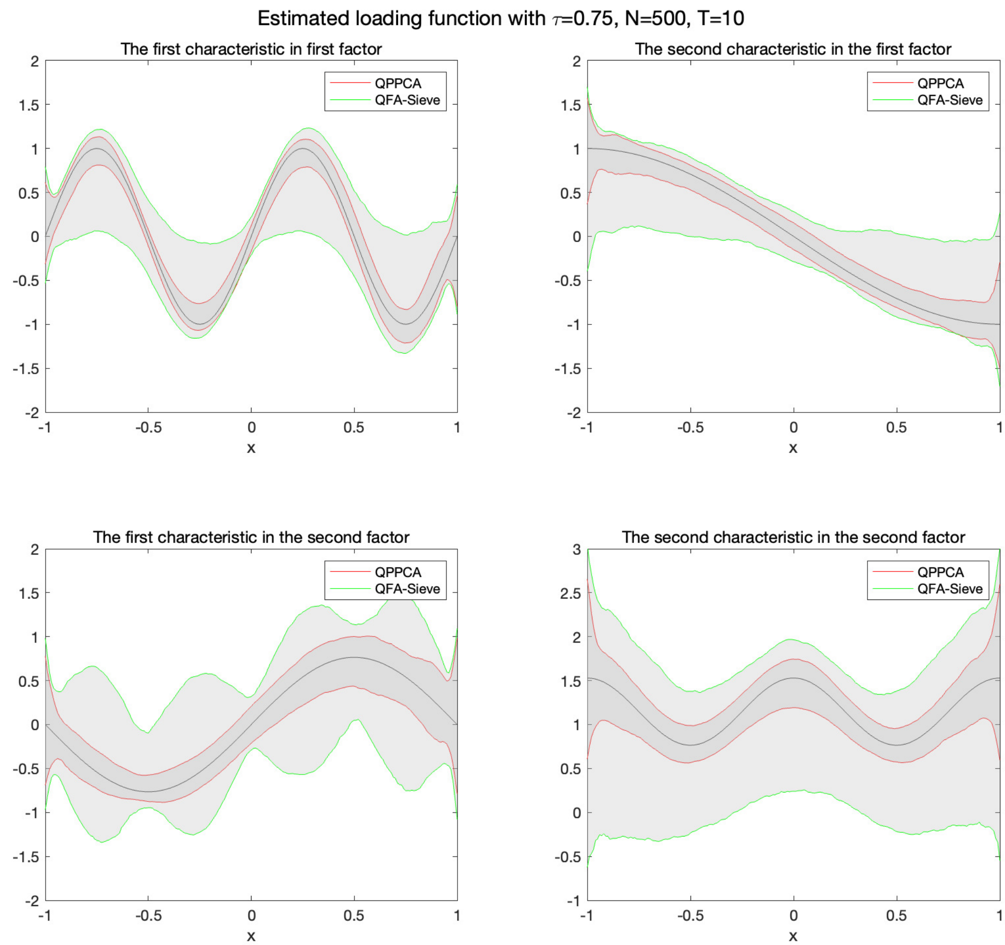
Note: $n = 300$, $T = 10$, $R = 2$ and $D = 2$. The DGP is: $y_{it} = (g_{11}(x_{i1}) + g_{12}(x_{i2})) \cdot f_{1t} + (g_{21}(x_{i1}) + g_{22}(x_{i2})) \cdot f_{2t} \cdot u_{it}$, where $f_{2t} = |h_t|$ and f_{1t}, h_t are independently drawn from $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n$, and $d = 1, 2)$ 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 line) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure A.5: Estimated Loading functions with $T = 10$: QPPCA and QFA-Sieve



Note: $n = 500$, $T = 10$, $R = 2$ 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 $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n$, 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 line) at $\tau = 0.25$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure A.6: Estimated Loading functions with $T = 10$: QPPCA and QFA-Sieve



Note: $n = 500$, $T = 10$, $R = 2$ 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 $N(0,1)$, $\{x_{id}\}(i = 1, 2, ...n$, 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 line) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

References

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

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.

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.