Online Appendix to "Estimation of Characteristics-based Quantile Factor Models"

Liang Chen
HSBC Business School, Peking University
and
Juan J. Dolado
Department of Economics, Universidad Carlos III de Madrid
and
Jesús Gonzalo
Department of Economics, Universidad Carlos III de Madrid
and
Haozi Pan
HSBC Business School, Peking University

March 10, 2023

A1 Simulations

In this section, we run a few Monte Carlo simulations to study the behavior in finite samples of the QPPCA estimators regarding the estimation of the number of factors, the factors themselves and their loading functions. In most cases, unless otherwise explicitly said, we suppose that the number of characteristics is $D = 5$ and that all of them, $\{x_{id}, d = 1, \ldots, 5\}$, are drawn independently from the uniform distribution: $U[-1, 1]$.

A1.1 Estimating the number of factors

Consider the following DGP:

$$y_{it} = \sum_{r=1}^{3} \lambda_{ir} f_{tr} + \left(x_{i1}^2 + x_{i2}^2 + x_{i3}^2\right) u_{it},$$

where $f_{t1} = 1$, $f_{t2}, f_{t3} \sim$ *i.i.d* $N(0, 1)$. Note that the chosen DGP is a location-scale shift model where the scale is driven by a subset of the five characteristics. This type of heteroskedasticity implies that the quantile loading functions exhibit variations across quantiles, unlike a pure location-shift model where the loading functions would be the same (up to a constant) for different quantiles.

Let $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ and $g_3(x) = cos(\pi x)$, such that

$$\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id}), \quad \lambda_{i2} = \sum_{d=1,2} g_2(x_{id}), \quad \lambda_{i3} = \sum_{d=3,4} g_3(x_{id}).$$

As for the idiosyncratic component, u_{it} are i.i.d. draws from three alternative distributions: (i) the standard normal distribution, $N(0, 1)$, (ii) the Student's t distribution with 3 degrees of freedom, $t(3)$, and (iii) the standard Cauchy distribution, Cauchy(0,1). In the first-step, we set $k_n = n^{1/3}$ in the quantile sieve estimation, and make use of the *Chebyshev polynomials of the second kind* as the basis functions. Moreover, in order to implement the rank minimization estimator for the number of factors in Equation (8), the threshold p_n is chosen as in Equation (9), with $d = 1/4$.

First, Table A.1 displays the results of the number of factors estimated with the rank minimization criterion for $\tau \in \{0.25, 0.5, 0.75\}$, $T \in \{5, 10\}$ and $n \in \{50, 100, 200, 1000\}$ from 1000 simulation replications. For each combination of τ, n and T, the reported results represent: [frequency of $\hat{R} < R$; frequency of $\hat{R} = R$; frequency of $\hat{R} > R$]. Next, for comparison, Table A.2 reports the corresponding results when the number of factors is estimated using the Ahn and Horenstein (2013)'s eigen-ratio estimator discussed in Remark 7.

There are three main takeaways from these simulation results. First, both selection criteria accurately estimate the number of factors when T is small and n is large, supporting our previous claim about their consistency even when T is fixed. Second, when n is large (=1000), both estimators perform well, even when the errors follow the standard Cauchy distribution. Hence, this result also provides support for the claim that our estimator is consistent in the absence of moment restrictions on the error terms. Third, although both estimators yield similar results when $n = 1000$, the rank minimization estimator outperforms the eigen-ratio estimator when n is not sufficiently large.

A1.2 Estimating the factors

A1.2.1 Comparison of QPPCA with PCA, PPCA and QFA

Following Chen et al. (2021), we consider the following DGP:

$$y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it},$$

where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. As before, let $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ but now $g_3(x) = |cos(\pi x)|$. The factor loading functions and the error terms are also generated as in Subsection A1.1. Note that, in this DGP, there are two location shift factors, f_{t1} and f_{t2}, that affect the mean of y_{it} and only one scale shift factor f_{t3} that affects the variance of y_{it}.

First, we focus on the estimation of the two location factors: f_{t1} and f_{t2}. Four competing estimation methods are considered: (i) our the proposed method with $\tau = 0.5$ (QPPCA); (ii) the quantile factor analysis estimator (QFA) of Chen et al. (2021) with $\tau = 0.5$; (iii) the projection estimator proposed by Fan et al. (2016) (PPCA); and (iv) the standard estimator of Bai and Ng (2002) for AFM (PCA). For the first two methods, the choices of k_n and the basis functions are again the same as in Subsection A1.1.

Regarding the choices of n and T, two different scenarios are considered:

(i) Fix $T = 10, 50$ and let n increase from 50 to 500.

(ii) Fix $n = 100, 200$ and let T increase from 5 to 200.

For each estimation method, the number of factors ($R = 2$ at $\tau = 0.5$) is assumed to be known, and we report the average Frobenius error as a measure of fit: $\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T}$ from 1000 replications, where $\hat{\boldsymbol{H}}$ represents the associated rotation matrix for each estimator.

The results for the first scenario (fixed T and increasing n) are plotted in Figure A.1. As can be inspected, for small T ($T = 10$), the PCA and QFA estimators perform worse than the PPCA and QPPCA estimators when u_{it} is either drawn from the $N(0, 1)$ or $t(3)$ distributions. Moreover, when the distribution is a standard Cauchy, the QPPCA estimator performs much better than its competitors. These findings agree again with our previous theoretical results showing that this estimator is consistent even when T is fixed or the moments of u_{it} do not exist.

When T is relatively large ($T = 50$) and the distribution of u_{it} has a thin tail, like a $N(0, 1)$ random variable, all the estimators behave similarly, as long as $n \geq 100$. However, if u_{it} follows the $t(3)$ distribution, the PCA estimator is subject to a much larger estimation error than the alternative procedures. In the extreme case of the standard Cauchy distribution, the two methods based on quantile regressions are the obvious winners, with the performances of the QFA and QPPCA estimators being very similar insofar $n \geq 200$.

The results for the second scenario (fixed n and increasing T) are displayed in Figure A.2. The main takeaway from this simulation exercise is that the QPPCA estimator provides the most robust approach against heavy-tailed distributions when T is small, while only the QFA estimator performs slightly better as T increases.

Next, we proceed to estimate all the three factors jointly, paying particular attention to the results for the scale factor f_{t3}. Since this last factor is absent when $\tau = 0.5$, for brevity we only provide simulations for $\tau = 0.25, 0.75$, and sample sizes where $T \in \{10, 50\}$ and $n \in \{50, 100, 200\}$. In each of these setups, the three estimated factors by the four different approaches are denoted as $\hat{F}^{\tau}_{QPPCA}, \hat{F}^{\tau}_{QFA}, \hat{F}_{PPCA}, \hat{F}_{PCA}$. Subsequently, each of the true factors is regressed on these estimated factors and the adjusted R^2s are computed as a measure of goodness of fit. The whole procedure is repeated 1000 times and the averages of the adjusted R^2s are reported in Tables A.3 to A.5.

Table A.3 displays the results for the QPPCA estimator. As can be observed, it performs well in estimating all the three factors. It should be noted, however, that the estimates of the scale factor f_{t3} are not as good as the estimates of the two location factors, f_{t1}, f_{t2}, when n is small, though the fit improves substantially as n increases. Table A.4, in turn, presents the results for the QFA estimator, whereas Table A.5 presents the corresponding results for the PCA and PPCA estimators. The main finding from Table A.4 is that the QFA estimator performs poorly in estimating the scale factor f_{t3} when T is small ($T = 10$), while it performs similarly to the QPPCA estimator when T is relatively large ($T = 50$).

Finally and not surprisingly, the main conclusion from Table A.5 is that both the PCA and PPCA estimators fail to capture the scale factor f_{t3} in all instances since they are designed for AFM but not for QFM.

A1.2.2 Comparison of QPPCA with SQFA

In the previous subsection the SQFA estimator proposed by Ma et al. (2021) was not included in the set of comparisons since its performance is close to that of the QFA estimator whenever the number of characteristics is larger or equal to the number of factors ($D \geq R$). Yet, in this subsection, we study how they differ when the number of characteristics is smaller than the number of factors ($D < R$).

To do so, we consider the following location-scale model as the DGP:

$$y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it},$$

where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. Now, the number of characteristics is 2 and, as in the previous simulations, all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1,1]$. Let $g_1(x) = sin(2\pi x),\ g_2(x) = sin(\pi x)\ and\ g_3(x) = |cos(\pi x)|$. Moreover, let $\lambda_{i1} = \sum_{d=1,2} g_1(x_{id}),\ \lambda_{i2} = \sum_{d=1,2} g_2(x_{id})\ and\ \lambda_{i3} = \sum_{d=1,2} g_3(x_{id})$. Again, u_{it} are generated from three different distributions discussed in Subsection A1.1.

For each estimator, we consider $\tau \in \{0.25, 0.5, 075\}$, $T \in \{10, 50\}$, $n \in \{50, 100, 200, 500\}$. Note that, when $\tau = 0.5$, there are only two location factors because f_{t3} does not affect the median of y_{it}. By contrast, when $\tau = 0.25, 0.75$ there will be two location factors and one scale factor. Moreover, to simplify the analysis, R is assumed to be known. For each τ, R factors are estimated using QPPCA. Note that the SQFA method chooses the number of factors as the number of characteristics by default, implying that only two factors will be estimated. Moreover, the choices of the basis functions and k_n are the same as in Subsection A1.1.

As before, we proceed to regress each of the true factors on the estimated factors and compute the adjusted R^2s. The whole procedure is repeated 1000 times and the averages of the adjusted R^2s are reported in Tables A.6 and A.7 for the QPPCA and SQFA estimators, respectively. When it comes to the estimation of the volatility factor, f_{t3}, it is not surprising to check that the QPPCA estimator outperforms the SQFA estimator since the latter is restricted to estimating only $D = 2$ factors. Thus, the main finding here is that the QPPCA estimator performs better than the SQFA estimator in estimating the location and scale shift factors whenever the number of factors exceeds the number of characteristics.

A2 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]

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

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

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

$d(\pi_n \theta_{0t}, \theta_{0t}) = \sqrt{\mathbb{E}\left(\pi_n \theta_{0t}(\boldsymbol{x}_i) - \theta_{0t}(\boldsymbol{x}_i)\right)^2} = O(k_n^{-\alpha})$. Therefore, it follows from Corollary 1 of Chen and Shen (1998) that

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

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

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. □

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

$$\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}/(nT) = \boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'/(nT) + \hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'/(nT) + \boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}/(nT) + \hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}/(nT). \quad \text{(A2.1)}$$

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

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

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 (A2.1) that

$$\hat{\boldsymbol{F}} = \boldsymbol{F}\hat{\boldsymbol{H}} + \hat{\boldsymbol{V}}'\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\hat{\boldsymbol{F}}/(nT)\hat{\boldsymbol{\Omega}}^{-1} + \boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}\hat{\boldsymbol{F}}/(nT)\hat{\boldsymbol{\Omega}}^{-1} + \hat{\boldsymbol{V}}'\hat{\boldsymbol{V}}/(nT)\hat{\boldsymbol{F}}\hat{\boldsymbol{\Omega}}^{-1}. \quad \text{(A2.2)}$$

Thus, it follows from (A2.2) and Lemma 1 that

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

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

$$\|\hat{\boldsymbol{B}} - \boldsymbol{B}_0(\boldsymbol{F}'\hat{\boldsymbol{F}}/T)\| \leq \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}_0\|}{\sqrt{T}} \cdot \frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}} = O_P(\varepsilon_{nT}). \tag{A2.3}$$

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 (A2.3) 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. □

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

$$\boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{f}\left(-\xi_{it}^*\right)\cdot\xi_{it}\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i). \tag{A2.4}$$

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

$$\boldsymbol{m}_t^*(\hat{\boldsymbol{a}}_t) = \boldsymbol{m}_t^*(\boldsymbol{a}_{0t}) - \boldsymbol{M}_t^*(\tilde{\boldsymbol{a}}_t)\cdot(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}) \tag{A2.5}$$

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

$$\boldsymbol{M}_t^*(\tilde{\boldsymbol{a}}_t) = -\frac{\partial \boldsymbol{m}_t^*(\boldsymbol{a})}{\partial \boldsymbol{a}'}|_{\boldsymbol{a}=\tilde{\boldsymbol{a}}_t} = \frac{1}{n}\sum_{i=1}^{n}\mathsf{f}\left((\tilde{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\right)\cdot\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'. \tag{A2.6}$$

By Assumption 3(ii) one can write

$$\boldsymbol{M}_t^*(\tilde{\boldsymbol{a}}_t) = \mathsf{f}(0)\cdot\hat{\boldsymbol{\Sigma}}_\phi + n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X}), \tag{A2.7}$$

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,

$$\begin{aligned}\max_{1\leq t\leq T}\|\boldsymbol{D}_t^*\|_S &\lesssim \max_{i,t}\left|(\tilde{\boldsymbol{a}}_t-\boldsymbol{a}_{0t})'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) - \xi_{it}\right| \\ &\leq \max_{1\leq t\leq T}\|\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t}\|\cdot O_P(\sqrt{k_n}) + O_P(k_n^{-\alpha}) = O_P(\sqrt{k_n}\varepsilon_{nT}).\end{aligned} \tag{A2.8}$$

Moreover, one can write

$$\boldsymbol{m}_t^*(\hat{\boldsymbol{a}}_t) = \boldsymbol{m}_t(\hat{\boldsymbol{a}}_t) - \tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) + \left[\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)\right] \tag{A2.9}$$

where $\tilde{\boldsymbol{m}}_t(\boldsymbol{a}) = \boldsymbol{m}_t(\boldsymbol{a}) - \boldsymbol{m}_t^*(\boldsymbol{a})$. It then follows from (A2.5) (A2.7) and (A2.9) that

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

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

$$\max_{1\leq t\leq T}\|\boldsymbol{m}_t^*(\boldsymbol{a}_{0t})\| = O_P(k_n^{-\alpha}), \tag{A2.10}$$

$$\max_{1\leq t\leq T}\|\boldsymbol{m}_t(\hat{\boldsymbol{a}}_t)\| = O_P(k_n^{3/2}/n), \tag{A2.11}$$

$$\frac{1}{T}\sum_{t=1}^{T}\|\tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t}) - \tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t)\|^2 = O_P\left(\eta_{nT}^2\right), \tag{A2.12}$$

$$\max_{1\leq t\leq T}\left\|n^{-1}\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{D}_t^*\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{a}}_t - \boldsymbol{a}_{0t})\right\| = O_P(\sqrt{k_n}\varepsilon_{nT}^2). \tag{A2.13}$$

Step 2: By (A2.4) 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

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

which gives (A2.10).

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 (A2.11).

Step 4: By (A2.8) 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 (A2.13).

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

$$\frac{1}{T}\sum_{t=1}^T \|\tilde{\boldsymbol{m}}_t(\hat{\boldsymbol{a}}_t) - \tilde{\boldsymbol{m}}_t(\boldsymbol{a}_{0t})\|^2 \le \frac{1}{T}\sum_{t=1}^T \left\|\tilde{\delta}_{1t}(\hat{\boldsymbol{a}}_t)\right\|^2 + \frac{1}{T}\sum_{t=1}^T \left\|\tilde{\delta}_{2t}(\hat{\boldsymbol{a}}_t)\right\|^2. \tag{A2.14}$$

In what follows, we will show that

$$\frac{1}{T}\sum_{t=1}^T \left\|\tilde{\delta}_{1t}(\hat{\boldsymbol{a}}_t)\right\|^2 = O_P\left(\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2}) \cdot k_n^{5/2}\varepsilon_{nT}n^{-1}\right), \tag{A2.15}$$

$$\frac{1}{T}\sum_{t=1}^T \left\|\tilde{\delta}_{2t}(\hat{\boldsymbol{a}}_t)\right\|^2 = O_P\left(\ln(k_n^{-1/2}\varepsilon_{nT}^{-1}) \cdot k_n^{3}\varepsilon_{nT}^2 n^{-1}\right), \tag{A2.16}$$

which imply (A2.12) and therefore complete the proof. We will focus on the proof of (A2.15) since the proof of (A2.16) 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

$$\mathbb{E}\left[\sup_{\|\boldsymbol{a}-\boldsymbol{a}_{0t}\|\leq C\varepsilon_{nT}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\alpha}, \boldsymbol{x}_i)\phi_{jd}(\boldsymbol{x}_i)]\}\right|^2\right]$$
$$= O\left(\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2}) \cdot k_n^{3/2}\varepsilon_{nT}\right) \quad \text{(A2.17)}$$

uniformly in t and j, from which (A2.15) 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 (A2.17) 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

$$\mathbb{E}\|\mathbb{G}_n h\|_{\mathcal{H}_{\varepsilon_{nT}}} \leq c_3\left[\cdot\ln(c_4 k_n^{-1/4}\varepsilon_{nT}^{-1/2}) \cdot k_n/\sqrt{n} + \sqrt{\ln(c_4 k_n^{-1/4}\varepsilon_{nT}^{-1/2})} \cdot k_n^{3/4}\varepsilon_{nT}^{1/2}\right]$$
$$\leq c_5\sqrt{\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2})} \cdot k_n^{3/4}\varepsilon_{nT}^{1/2}, \quad \text{(A2.18)}$$

where c_3, c_4, c_5 are bounded constants that do not depend on t and j. Finally, (A2.17) follows by noting that (see Chapter 6 of Ledoux and Talagrand 1991)

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

This completes the proof. □

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

$$\left\|\hat{\boldsymbol{A}} - \boldsymbol{A}_0 - \mathsf{f}(0)^{-1} \cdot \hat{\boldsymbol{\Sigma}}_\phi^{-1} \boldsymbol{\Phi}'(\boldsymbol{X})\boldsymbol{\Psi}/n\right\| / \sqrt{T} = O_P\left(k_n^{-\alpha}\right) + O_P\left(\eta_{nT}\right). \tag{A2.19}$$

From (A2.2) and Lemma 1 we have

$$\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T} \leq O_P(1) \cdot \|\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{V}}/(nT)\|_S + O_P(\varepsilon_{nT}^2). \tag{A2.20}$$

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 (A2.20) imply that

$$\|\hat{\boldsymbol{F}}-\boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T} \leq \|\boldsymbol{F}\boldsymbol{B}_0'\boldsymbol{\Phi}(\boldsymbol{X})'\boldsymbol{\Phi}(\boldsymbol{X})(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)/(nT)\|_S + O_P(k_n^{-\alpha}) + O_P(\varepsilon_{nT}^2). \quad \text{(A2.21)}$$

By (A2.19) 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 (A2.21) 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. $\square$

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

$$\begin{aligned}\boldsymbol{\Sigma}_{T,\tau}^{-1/2}(\hat{\boldsymbol{H}}')^{-1}\cdot\frac{\sqrt{nT}}{\sigma_{k_n}}\Big(\hat{\boldsymbol{g}}(\boldsymbol{x}) - (\boldsymbol{F}'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x})\Big) &= \boldsymbol{\Sigma}_{T,\tau}^{-1/2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^n\sqrt{T}\boldsymbol{d}_T(\boldsymbol{x}_i)q(\boldsymbol{x}_i)/\sigma_{k_n} \\ &+ O_P(k_n^{1/2-\alpha} + \sqrt{k_n}\eta_{nT})\sqrt{nT}\sigma_{k_n}^{-1}. \quad \text{(A2.22)}\end{aligned}$$

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

$$\|\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}'\boldsymbol{\Phi}(\boldsymbol{X})' - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'\boldsymbol{F}\boldsymbol{G}(\boldsymbol{X})'\|/(nT) = O_P(\varepsilon_{nT}). \tag{A2.23}$$

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 (A2.23) 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 $j = R+1, \ldots, \bar{R}$. □

Table A.1: Estimating the number of factors: rank minimization estimator

	T	n	$N(0,1)$	$t(3)$	Cauchy(0,1)
$\tau=0.25$	5	50	[0.13 0.65 0.23]	[0.03 0.41 0.56]	[0.01 0.10 0.89]
	5	100	[0.10 0.72 0.19]	[0.02 0.44 0.54]	[0.00 0.03 0.97]
	5	200	[0.23 0.77 0.00]	[0.12 0.82 0.06]	[0.00 0.17 0.83]
	5	1000	[0.17 0.83 0.00]	[0.16 0.84 0.00]	[0.06 0.81 0.13]
	10	50	[0.17 0.76 0.07]	[0.03 0.50 0.47]	[0.02 0.06 0.92]
	10	100	[0.08 0.89 0.03]	[0.03 0.65 0.46]	[0.00 0.03 0.97]
	10	200	[0.07 0.93 0.00]	[0.05 0.95 0.00]	[0.00 0.24 0.76]
	10	1000	[0.03 0.97 0.00]	[0.02 0.98 0.00]	[0.01 0.98 0.01]
$\tau=0.5$	5	50	[0.19 0.71 0.10]	[0.09 0.56 0.35]	[0.00 0.15 0.85]
	5	100	[0.17 0.76 0.08]	[0.07 0.59 0.34]	[0.00 0.20 0.80]
	5	200	[0.23 0.77 0.00]	[0.19 0.80 0.01]	[0.06 0.75 0.19]
	5	1000	[0.18 0.82 0.00]	[0.15 0.85 0.00]	[0.13 0.87 0.00]
	10	50	[0.20 0.78 0.03]	[0.08 0.76 0.15]	[0.00 0.13 0.87]
	10	100	[0.12 0.87 0.01]	[0.05 0.87 0.08]	[0.00 0.24 0.76]
	10	200	[0.05 0.95 0.00]	[0.05 0.95 0.00]	[0.03 0.94 0.03]
	10	1000	[0.01 0.99 0.00]	[0.02 0.98 0.00]	[0.02 0.99 0.00]
$\tau=0.75$	5	50	[0.11 0.68 0.21]	[0.04 0.41 0.56]	[0.01 0.09 0.90]
	5	100	[0.10 0.71 0.19]	[0.02 0.42 0.56]	[0.00 0.04 0.96]
	5	200	[0.22 0.78 0.00]	[0.14 0.81 0.05]	[0.00 0.15 0.85]
	5	1000	[0.18 0.82 0.00]	[0.17 0.83 0.00]	[0.04 0.82 0.15]
	10	50	[0.15 0.78 0.08]	[0.04 0.50 0.46]	[0.01 0.05 0.94]
	10	100	[0.11 0.86 0.04]	[0.03 0.65 0.32]	[0.00 0.03 0.97]
	10	200	[0.06 0.94 0.00]	[0.05 0.94 0.01]	[0.01 0.27 0.73]
	10	1000	[0.02 0.98 0.00]	[0.02 0.98 0.00]	[0.02 0.97 0.01]

Note: the DGP is $y_{it} = \sum_{r=1}^{3} \lambda_{ir} f_{tr} + \left(x_{i1}^2 + x_{i2}^2 + x_{i3}^2\right) u_{it}$, where $f_{t1} = 1$, $f_{t2}, f_{t3} \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are drawn independently from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ and $g_3(x) = cos(\pi x)$, and $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$, $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. u_{it} are i.i.d variables drawn from three different distributions. In the first step quantile sieve estimation, $k_n = n^{1/3}$ and we use the *Chebyshev polynomials of the second kind* as the basis functions. For the estimator of the number of factors, the threshold p_n is chosen as in Equation (9) with $d = 1/4$. The reported results are [frequency of $\hat{R} < R$; frequency of $\hat{R} = R$; frequency of $\hat{R} > R$] from 1000 replications.

Table A.2: Estimating the number of factors: eigen-ratio estimator

	T	n	$N(0,1)$			$t(3)$			Cauchy(0,1)		
$\tau = 0.25$	5	50	[0.57	0.25	0.19]	[0.54	0.22	0.25]	[0.54	0.17	0.29]
	5	100	[0.58	0.33	0.09]	[0.58	0.27	0.15]	[0.59	0.15	0.26]
	5	200	[0.44	0.54	0.01]	[0.54	0.43	0.04]	[0.62	0.24	0.14]
	5	1000	[0.23	0.77	0.00]	[0.31	0.69	0.00]	[0.56	0.42	0.02]
	10	50	[0.46	0.37	0.17]	[0.45	0.18	0.37]	[0.47	0.07	0.46]
	10	100	[0.37	0.59	0.04]	[0.46	0.42	0.11]	[0.60	0.09	0.31]
	10	200	[0.09	0.91	0.00]	[0.19	0.80	0.01]	[0.59	0.31	0.11]
	10	1000	[0.01	0.99	0.00]	[0.03	0.97	0.00]	[0.17	0.83	0.00]
$\tau = 0.5$	5	50	[0.58	0.28	0.14]	[0.57	0.22	0.20]	[0.50	0.20	0.30]
	5	100	[0.58	0.33	0.09]	[0.57	0.28	0.15]	[0.56	0.21	0.22]
	5	200	[0.42	0.57	0.01]	[0.46	0.51	0.03]	[0.54	0.41	0.06]
	5	1000	[0.21	0.79	0.00]	[0.23	0.77	0.00]	[0.28	0.72	0.00]
	10	50	[0.41	0.46	0.13]	[0.46	0.33	0.21]	[0.42	0.10	0.48]
	10	100	[0.30	0.66	0.04]	[0.36	0.57	0.07]	[0.51	0.24	0.26]
	10	200	[0.06	0.94	0.00]	[0.11	0.89	0.00]	[0.22	0.76	0.02]
	10	1000	[0.01	0.99	0.00]	[0.02	0.98	0.00]	[0.03	0.97	0.00]
$\tau = 0.75$	5	50	[0.58	0.25	0.17]	[0.54	0.22	0.24]	[0.55	0.17	0.28]
	5	100	[0.57	0.32	0.10]	[0.59	0.24	0.17]	[0.56	0.20	0.24]
	5	200	[0.43	0.55	0.02]	[0.52	0.43	0.04]	[0.65	0.21	0.14]
	5	1000	[0.24	0.76	0.00]	[0.33	0.67	0.00]	[0.55	0.44	0.01]
	10	50	[0.46	0.36	0.18]	[0.44	0.20	0.37]	[0.47	0.05	0.48]
	10	100	[0.36	0.59	0.06]	[0.46	0.40	0.14]	[0.63	0.09	0.28]
	10	200	[0.11	0.89	0.00]	[0.19	0.80	0.01]	[0.58	0.31	0.11]
	10	1000	[0.01	0.99	0.00]	[0.03	0.97	0.00]	[0.16	0.83	0.01]

Note: the DGP is $y_{it} = \sum_{r=1}^{3} \lambda_{ir} f_{tr} + \left(x_{i1}^2 + x_{i2}^2 + x_{i3}^2\right) u_{it}$, where $f_{t1} = 1$, $f_{t2}, f_{t3} \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are drawn independently from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ and $g_3(x) = cos(\pi x)$, and $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$, $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. u_{it} are i.i.d variables drawn from three different distributions. In the first step quantile sieve estimation, $k_n = n^{1/3}$ and we use the *Chebyshev polynomials of the second kind* as the basis functions. The estimator for the number of factors is the integer that maximizes the eigen-ratios. The reported results are [frequency of $\hat{R} < R$; frequency of $\hat{R} = R$; frequency of $\hat{R} > R$] from 1000 replications.

Table A.3: Factor estimation using QPPCA

	T	n	$N(0,1)$ f_{1t}	$N(0,1)$ f_{2t}	$N(0,1)$ f_{3t}	$t(3)$ f_{1t}	$t(3)$ f_{2t}	$t(3)$ f_{3t}	Cauchy(0,1) f_{1t}	Cauchy(0,1) f_{2t}	Cauchy(0,1) f_{3t}
$\tau = 0.25$	10	50	0.859	0.879	0.574	0.738	0.745	0.630	0.386	0.370	0.609
	10	100	0.971	0.956	0.857	0.938	0.890	0.835	0.670	0.566	0.767
	10	200	0.989	0.983	0.924	0.978	0.959	0.911	0.862	0.767	0.867
	10	500	0.997	0.995	0.979	0.994	0.990	0.972	0.968	0.940	0.950
	50	50	0.893	0.909	0.417	0.751	0.796	0.499	0.086	0.069	0.375
	50	100	0.976	0.968	0.824	0.957	0.940	0.797	0.623	0.407	0.654
	50	200	0.990	0.986	0.901	0.982	0.977	0.892	0.919	0.838	0.821
	50	500	0.997	0.995	0.973	0.995	0.992	0.967	0.984	0.975	0.941
$\tau = 0.75$	10	50	0.861	0.876	0.581	0.749	0.749	0.623	0.383	0.362	0.605
	10	100	0.971	0.955	0.858	0.933	0.894	0.834	0.682	0.573	0.768
	10	200	0.989	0.983	0.921	0.979	0.960	0.905	0.867	0.777	0.867
	10	500	0.997	0.995	0.979	0.994	0.990	0.974	0.973	0.937	0.950
	50	50	0.893	0.911	0.420	0.749	0.794	0.493	0.081	0.066	0.380
	50	100	0.977	0.967	0.824	0.958	0.938	0.794	0.617	0.400	0.656
	50	200	0.990	0.986	0.901	0.982	0.976	0.894	0.915	0.832	0.818
	50	500	0.997	0.995	0.972	0.995	0.992	0.967	0.984	0.974	0.938

Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are independently drawn from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. 3 factors are estimated at each τ using the proposed method in this paper, and the reported results are the averages of the adjusted R^2 of regressing the true factors on the estimated factors from 1000 replications.

Table A.4: Factor estimation using QFA

	T	n	$N(0,1)$ f_{1t}	$N(0,1)$ f_{2t}	$N(0,1)$ f_{3t}	$t(3)$ f_{1t}	$t(3)$ f_{2t}	$t(3)$ f_{3t}	Cauchy(0,1) f_{1t}	Cauchy(0,1) f_{2t}	Cauchy(0,1) f_{3t}
$\tau = 0.25$	10	50	0.887	0.821	0.561	0.808	0.706	0.528	0.516	0.418	0.449
	10	100	0.898	0.833	0.586	0.822	0.727	0.574	0.525	0.427	0.501
	10	200	0.904	0.841	0.624	0.834	0.735	0.584	0.525	0.443	0.504
	10	500	0.908	0.840	0.643	0.841	0.740	0.608	0.513	0.420	0.512
	50	50	0.964	0.948	0.786	0.935	0.902	0.725	0.724	0.537	0.473
	50	100	0.983	0.976	0.884	0.972	0.956	0.848	0.871	0.767	0.669
	50	200	0.992	0.988	0.936	0.986	0.977	0.911	0.935	0.853	0.802
	50	500	0.996	0.994	0.965	0.994	0.989	0.951	0.963	0.906	0.880
$\tau = 0.75$	10	50	0.875	0.835	0.551	0.808	0.719	0.523	0.510	0.414	0.447
	10	100	0.898	0.938	0.595	0.820	0.730	0.583	0.523	0.420	0.506
	10	200	0.904	0.846	0.616	0.828	0.736	0.600	0.520	0.429	0.497
	10	500	0.899	0.838	0.625	0.843	0.742	0.616	0.528	0.433	0.489
	50	50	0.964	0.947	0.785	0.935	0.901	0.722	0.722	0.551	0.486
	50	100	0.983	0.975	0.884	0.972	0.956	0.846	0.874	0.760	0.672
	50	200	0.992	0.988	0.935	0.986	0.978	0.911	0.931	0.852	0.799
	50	500	0.996	0.994	0.964	0.994	0.989	0.949	0.964	0.903	0.878

Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are independently drawn from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. 3 factors are estimated at each τ using the QFA proposed by Chen et al. (2021), and the reported results are the averages of the adjusted R^2 of regressing the true factors on the estimated factors from 1000 replications.

Table A.5: Factor estimation using PCA and PPCA

	T	n	$N(0,1)$ f_{1t}	$N(0,1)$ f_{2t}	$N(0,1)$ f_{3t}	$t(3)$ f_{1t}	$t(3)$ f_{2t}	$t(3)$ f_{3t}	Cauchy(0,1) f_{1t}	Cauchy(0,1) f_{2t}	Cauchy(0,1) f_{3t}
PCA	10	50	0.955	0.921	0.420	0.847	0.723	0.455	0.271	0.250	0.392
	10	100	0.964	0.929	0.450	0.858	0.757	0.514	0.286	0.289	0.410
	10	200	0.970	0.944	0.478	0.871	0.751	0.530	0.289	0.285	0.422
	10	500	0.975	0.944	0.493	0.879	0.767	0.568	0.292	0.303	0.433
	50	50	0.973	0.957	0.084	0.894	0.781	0.079	0.003	-0.001	0.032
	50	100	0.986	0.977	0.131	0.937	0.862	0.116	0.032	0.031	0.066
	50	200	0.993	0.988	0.149	0.961	0.901	0.141	0.044	0.048	0.075
	50	500	0.997	0.994	0.166	0.977	0.933	0.161	0.055	0.054	0.091
PPCA	10	50	0.949	0.962	0.382	0.843	0.866	0.379	0.277	0.282	0.387
	10	100	0.989	0.984	0.374	0.960	0.930	0.379	0.321	0.314	0.406
	10	200	0.995	0.993	0.382	0.983	0.969	0.383	0.318	0.309	0.409
	10	500	0.998	0.997	0.400	0.994	0.989	0.402	0.321	0.317	0.417
	50	50	0.953	0.963	0.060	0.858	0.882	0.054	0.003	0.001	0.029
	50	100	0.987	0.982	0.095	0.962	0.947	0.085	0.036	0.031	0.062
	50	200	0.994	0.992	0.110	0.982	0.974	0.100	0.048	0.049	0.072
	50	500	0.998	0.997	0.130	0.994	0.990	0.114	0.058	0.056	0.090

Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are independently drawn from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. 3 factors are estimated using PCA and PPCA respectively, and the reported results are the averages of the adjusted R^2 of regressing the true factors on the estimated factors from 1000 replications.

Table A.6: Factor estimation using QPPCA: $R = 3, D = 2$

			$N(0,1)$			$t(3)$			Cauchy(0,1)		
	T	n	f_{1t}	f_{2t}	f_{3t}	f_{1t}	f_{2t}	f_{3t}	f_{1t}	f_{2t}	f_{3t}
$\tau = 0.25$	10	50	0.731	0.880	0.606	0.588	0.806	0.617	0.348	0.481	0.602
	10	100	0.932	0.948	0.830	0.877	0.902	0.799	0.655	0.668	0.731
	10	200	0.969	0.978	0.916	0.947	0.960	0.901	0.778	0.837	0.847
	10	500	0.990	0.993	0.974	0.983	0.989	0.968	0.929	0.941	0.942
	50	50	0.665	0.875	0.485	0.488	0.782	0.473	0.100	0.202	0.390
	50	100	0.934	0.957	0.811	0.892	0.921	0.766	0.473	0.531	0.602
	50	200	0.969	0.982	0.906	0.949	0.970	0.889	0.785	0.839	0.796
	50	500	0.990	0.993	0.968	0.984	0.989	0.961	0.952	0.963	0.931
$\tau = 0.5$	10	50	0.643	0.889	0.152	0.524	0.837	0.165	0.364	0.664	0.201
	10	100	0.927	0.949	0.127	0.907	0.935	0.136	0.807	0.845	0.171
	10	200	0.968	0.981	0.128	0.955	0.974	0.136	0.917	0.940	0.158
	10	500	0.990	0.994	0.135	0.987	0.991	0.126	0.979	0.986	0.142
	50	50	0.697	0.913	-0.013	0.581	0.870	-0.011	0.279	0.682	0.005
	50	100	0.945	0.968	0.004	0.929	0.956	0.003	0.857	0.899	0.004
	50	200	0.973	0.984	0.011	0.967	0.980	0.012	0.945	0.968	0.014
	50	500	0.991	0.994	0.018	0.989	0.993	0.017	0.984	0.989	0.018
$\tau = 0.75$	10	50	0.718	0.878	0.603	0.609	0.804	0.629	0.356	0.473	0.596
	10	100	0.932	0.948	0.834	0.874	0.900	0.791	0.636	0.664	0.737
	10	200	0.970	0.980	0.922	0.943	0.962	0.907	0.796	0.833	0.848
	10	500	0.991	0.993	0.975	0.984	0.987	0.968	0.933	0.941	0.943
	50	50	0.663	0.872	0.498	0.485	0.779	0.476	0.102	0.203	0.392
	50	100	0.935	0.956	0.813	0.889	0.920	0.762	0.450	0.510	0.608
	50	200	0.969	0.981	0.906	0.951	0.970	0.890	0.792	0.845	0.800
	50	500	0.990	0.993	0.969	0.984	0.989	0.962	0.951	0.964	0.931

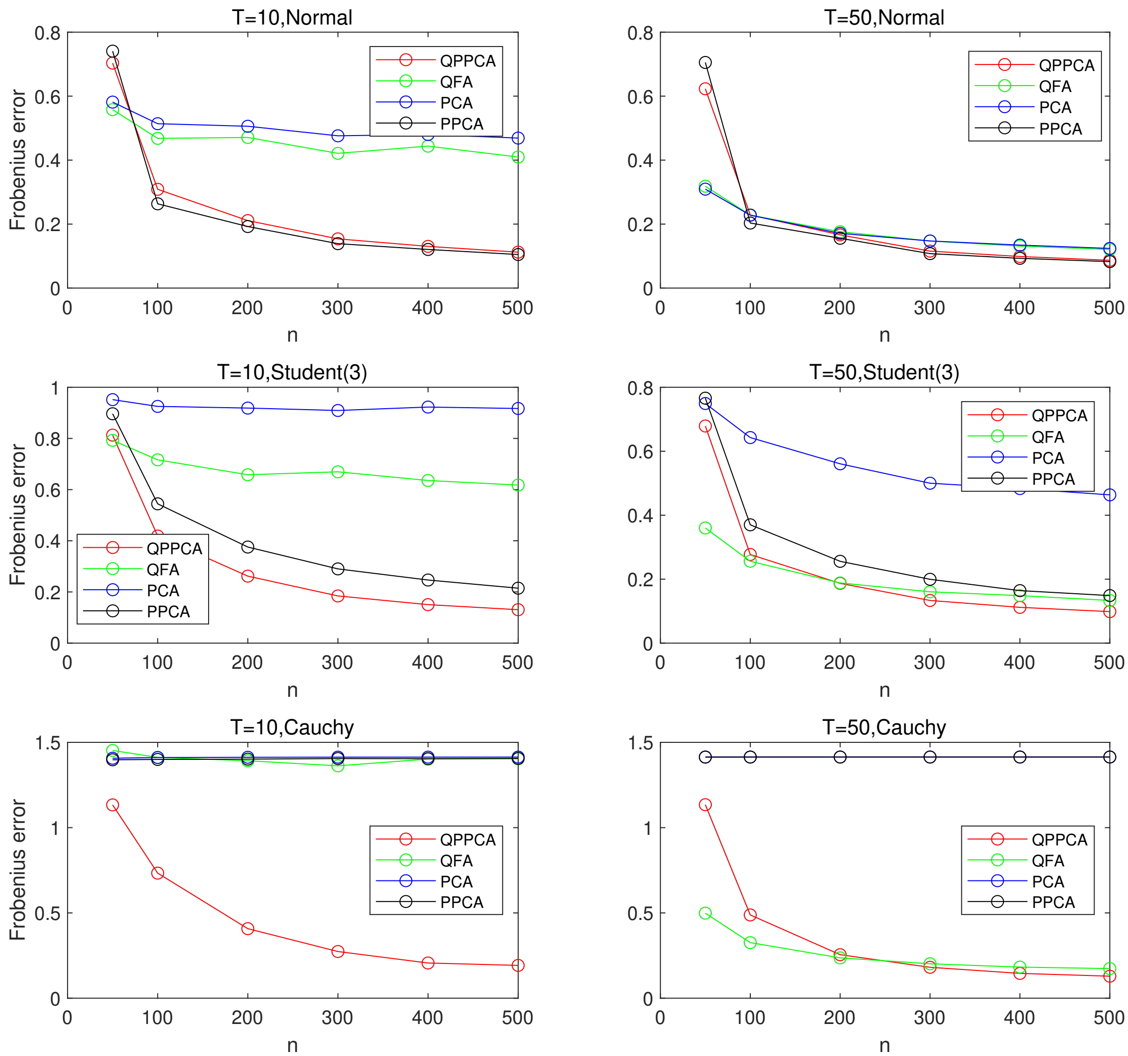
Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} are independently drawn from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. $\lambda_{i1} = \sum_{d=1,2} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ *and* $\lambda_{i3} = \sum_{d=1,2} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. 3 factors are estimated at each τ using the method proposed in this paper, and the reported results are the averages of the adjusted R^2 of regressing the true factors on the estimated factors from 1000 replications.

Table A.7: Factor estimation using SQFA: $R = 3, D = 2$

	T	n	$N(0,1)$ f_{1t}	$N(0,1)$ f_{2t}	$N(0,1)$ f_{3t}	$t(3)$ f_{1t}	$t(3)$ f_{2t}	$t(3)$ f_{3t}	Cauchy(0,1) f_{1t}	Cauchy(0,1) f_{2t}	Cauchy(0,1) f_{3t}
$\tau = 0.25$	10	50	0.406	0.826	0.205	0.364	0.793	0.218	0.295	0.649	0.277
	10	100	0.595	0.698	0.195	0.578	0.690	0.208	0.556	0.656	0.231
	10	200	0.623	0.706	0.223	0.604	0.680	0.238	0.573	0.676	0.277
	10	500	0.630	0.682	0.233	0.614	0.687	0.239	0.596	0.686	0.301
	50	50	0.311	0.845	0.040	0.267	0.804	0.058	0.191	0.662	0.101
	50	100	0.516	0.680	0.036	0.503	0.660	0.043	0.471	0.627	0.063
	50	200	0.523	0.691	0.058	0.518	0.676	0.067	0.487	0.649	0.095
	50	500	0.578	0.656	0.061	0.553	0.651	0.068	0.543	0.626	0.102
$\tau = 0.5$	10	50	0.383	0.849	0.133	0.351	0.819	0.132	0.314	0.749	0.142
	10	100	0.584	0.695	0.150	0.573	0.688	0.143	0.531	0.674	0.154
	10	200	0.584	0.713	0.156	0.559	0.716	0.157	0.539	0.678	0.158
	10	500	0.615	0.689	0.157	0.600	0.663	0.157	0.598	0.649	0.152
	50	50	0.277	0.865	-0.014	0.236	0.843	-0.013	0.185	0.773	-0.014
	50	100	0.509	0.679	0.007	0.471	0.669	0.007	0.439	0.625	0.005
	50	200	0.514	0.688	0.015	0.493	0.680	0.016	0.452	0.669	0.016
	50	500	0.557	0.639	0.022	0.544	0.636	0.023	0.503	0.623	0.022
$\tau = 0.75$	10	50	0.402	0.816	0.200	0.374	0.804	0.213	0.316	0.630	0.268
	10	100	0.606	0.688	0.191	0.564	0.900	0.195	0.556	0.666	0.226
	10	200	0.590	0.691	0.221	0.584	0.962	0.215	0.582	0.672	0.281
	10	500	0.638	0.691	0.231	0.622	0.987	0.259	0.620	0.675	0.285
	50	50	0.318	0.837	0.039	0.268	0.779	0.049	0.191	0.658	0.099
	50	100	0.525	0.671	0.039	0.499	0.920	0.044	0.465	0.624	0.067
	50	200	0.528	0.688	0.057	0.510	0.970	0.064	0.481	0.654	0.101
	50	500	0.574	0.652	0.063	0.564	0.989	0.067	0.543	0.627	0.096

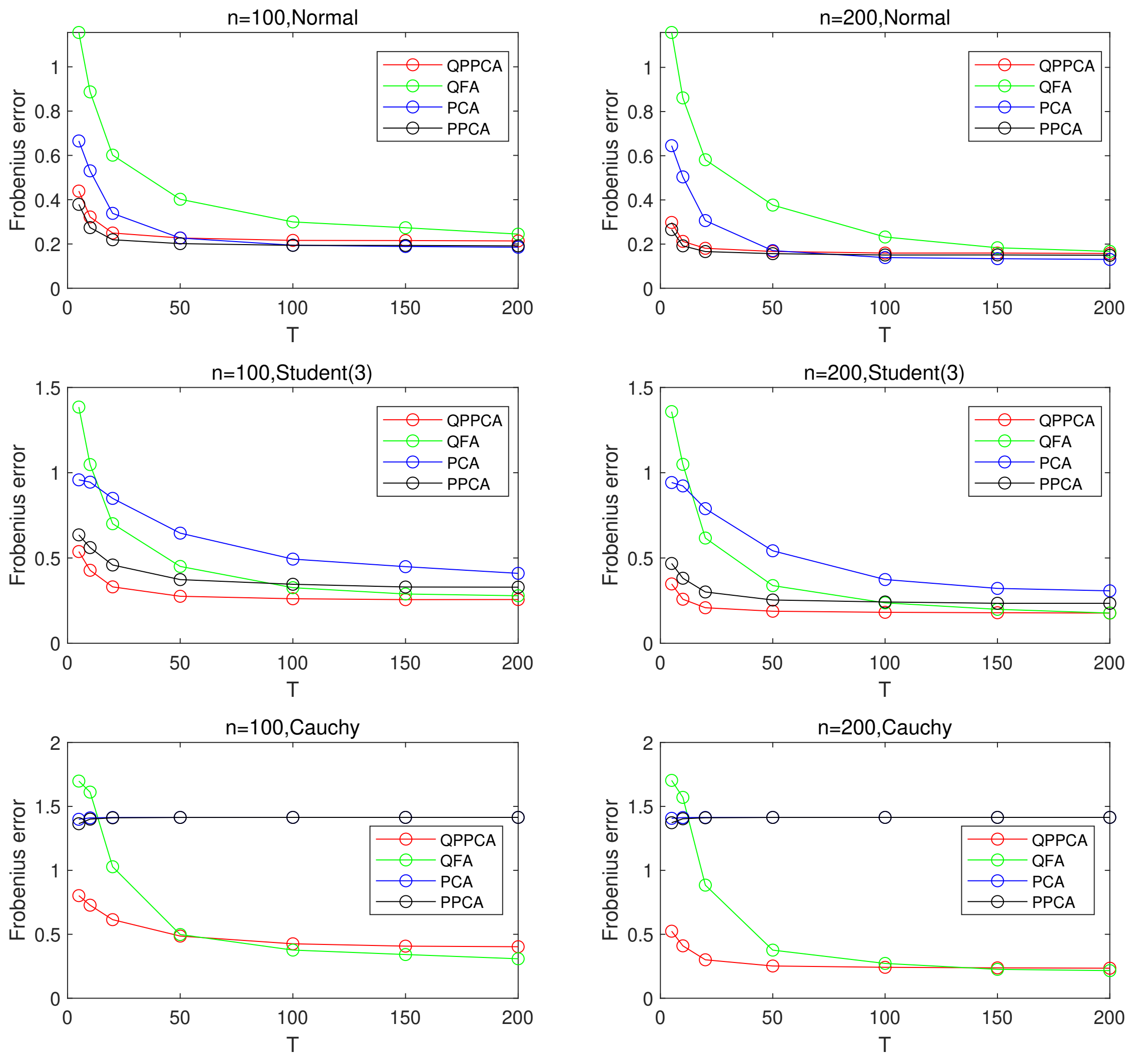
Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} are independently drawn from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. $\lambda_{i1} = \sum_{d=1,2} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ *and* $\lambda_{i3} = \sum_{d=1,2} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. 2 factors are estimated at each τ using the method proposed by Ma et al. (2021), and the reported results are the averages of the adjusted R^2 of regressing the true factors on the estimated factors from 1000 replications.

Figure A.1: Estimation of factors: fixed T and increasing n.



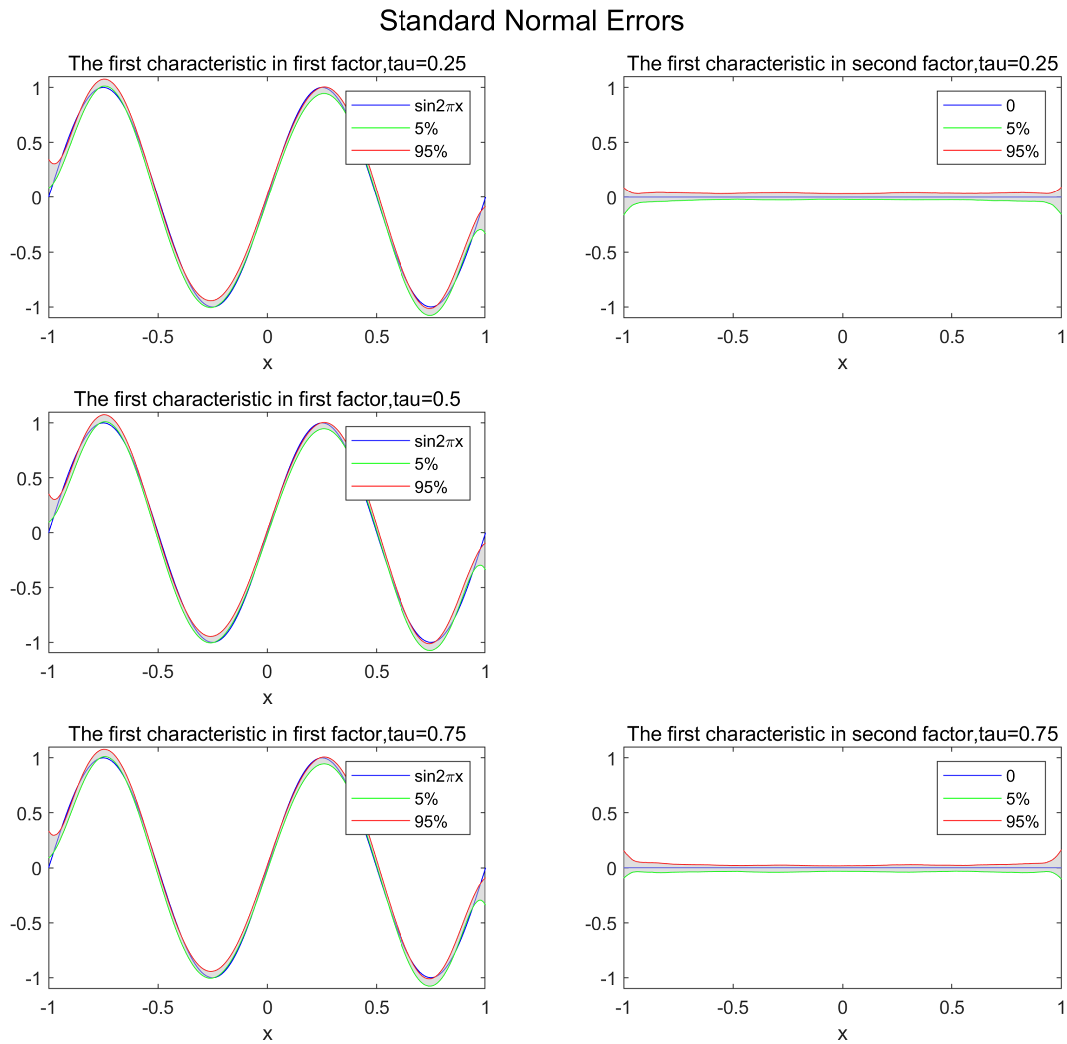
Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 5 and all characteristics x_{id} ($i = 1, ...N$ and $d = 1, 2, 3, 4, 5$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The mean factors (f_{t1} and f_{t2}) are estimated by four methods: PCA, PPCA, QFA and QPPCA at $\tau = 0.5$. The reported results are the average Frobenius errors: $\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T}$ from 1000 repetitions, where $\hat{\boldsymbol{H}}$ is the associated rotation matrix for each estimator.

Figure A.2: Estimation of factors: fixed n and increasing T.



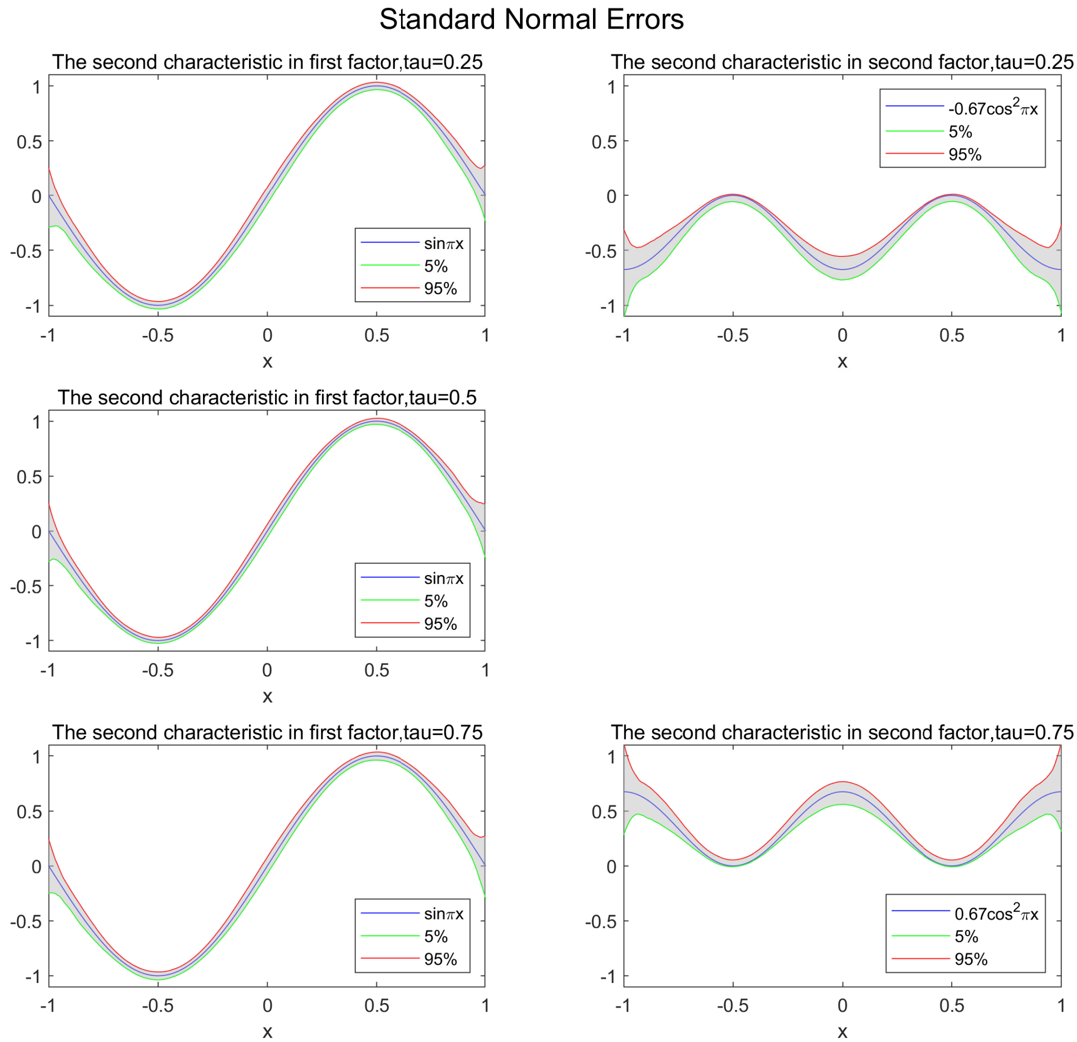
Note: the DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 5 and all characteristics x_{id} ($i = 1, ...N$ and $d = 1, 2, 3, 4, 5$) are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The mean factors (f_{t1} and f_{t2}) are estimated by four methods: PCA, PPCA, QFA and QPPCA at $\tau = 0.5$. The reported results are the average Frobenius errors: $\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T}$ from 1000 repetitions, where $\hat{\boldsymbol{H}}$ is the associated rotation matrix for each estimator.

Figure A.3: Loading function of the first characteristic when error term is $N(0, 1)$



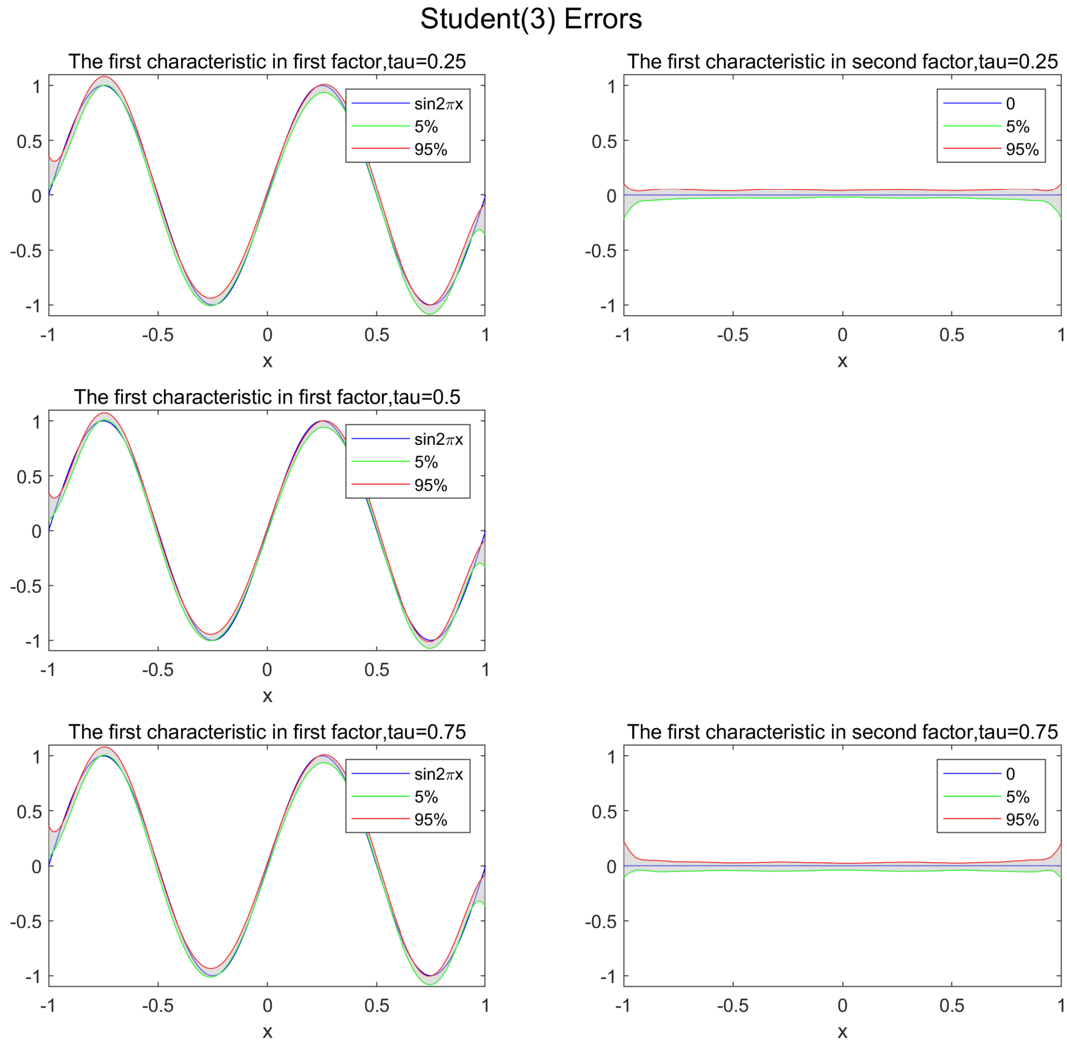
Note: the DGP is: $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |g_t|$ and $f_{t1}, g_t \sim i.i.d\ N(0, 1)$. $n = 500, T = 10$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ...N$ and $d = 1, 2$) are independently drawn from uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x), g_{21}(x) = 0, g_{12}(x) = sin(\pi x)$, $g_{22}(x) = cos^2(\pi x)$, and $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}), \lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. u_{it} are drawn independently from the standard normal distribution. The left panel are the estimation results for $g_{11,\tau}(x) = \sin(2\pi x)$ and the right panel are the estimation results for $g_{21,\tau}(x) = 0$ with $\tau \in \{0.25, 0.75\}$. For each graph, the blue line is the true function, the red line and the green line are the 95% and 5% empirical quantiles from 1000 replications.

Figure A.4: Loading function of the second characteristic when error term is N(0,1)



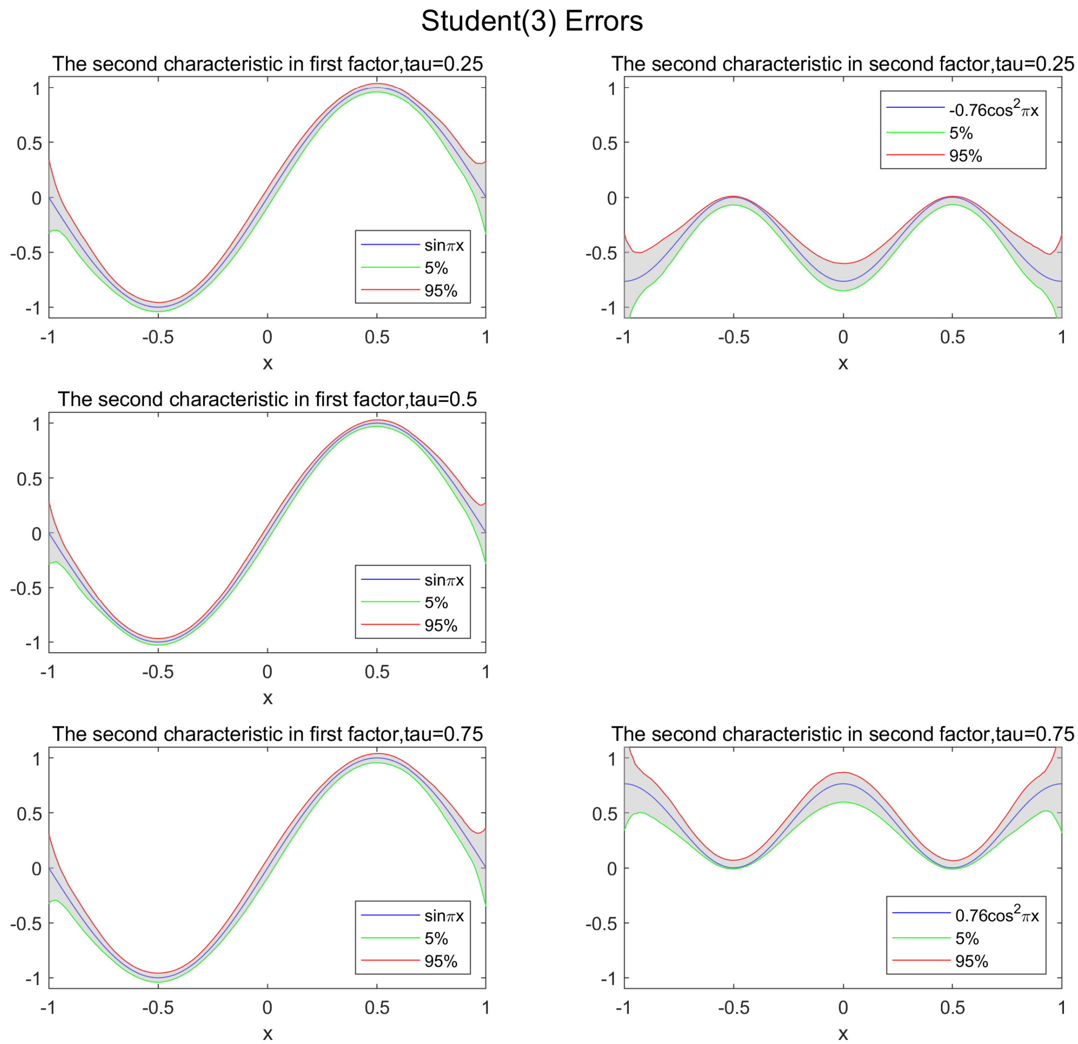
Note: the DGP is: $y_{it} = \lambda_{1i} f_{1t} + (\lambda_{2i} f_{2t}) u_{it}$, where $f_{2t} = |g_t|$ and $f_{1t}, g_t \sim i.i.d\ N(0,1)$. $n = 500, T = 10$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ...N$ and $d = 1, 2$) are independently drawn from uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x), g_{21}(x) = 0, g_{12}(x) = sin(\pi x)$, $g_{22}(x) = cos^2(\pi x)$, and $\lambda_{1i} = g_{11}(x_{1i}) + g_{12}(x_{2i}), \lambda_{2i} = g_{21}(x_{1i}) + g_{22}(x_{2i})$. u_{it} are drawn independently from the standard normal distribution. The left panel are the estimation results for $g_{12,\tau}(x) = \sin(\pi x)$ and the right panel are the estimation results for $g_{22,\tau}(x)$ with $\tau \in \{0.25, 0.75\}$. For each graph, the blue line is the true function, the red line and the green line are the 95% and 5% empirical quantiles from 1000 replications.

Figure A.5: Loading function of first characteristic when error term is $t(3)$



Note: the DGP is: $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |g_t|$ and $f_{t1}, g_t \sim i.i.d\ N(0,1)$. $n = 500, T = 10$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ...N$ and $d = 1, 2$) are independently drawn from uniform distribution: $U[-1, 1]$. $g_{11}(x) = sin(2\pi x), g_{21}(x) = 0, g_{12}(x) = sin(\pi x)$, $g_{22}(x) = cos^2(\pi x)$, and $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}), \lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. u_{it} are drawn independently from the student's t distribution with 3 degrees of freedom. The left panel are the estimation results for $g_{11,\tau}(x) = \sin(2\pi x)$ and the right panel are the estimation results for $g_{21,\tau}(x) = 0$ with $\tau \in \{0.25, 0.75\}$. For each graph, the blue line is the true function, the red line and the green line are the 95% and 5% empirical quantiles from 1000 replications.

Figure A.6: Loading function of second characteristic when error term is $t(3)$



Note: the DGP is: $y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |g_t|$ and $f_{t1}, g_t \sim i.i.d\ N(0,1)$. $n = 500, T = 10$. The number of characteristics is 2 and all characteristics x_{id} ($i = 1, ...N$ and $d = 1, 2$) are independently drawn from uniform distribution: $U[-1,1]$. $g_{11}(x) = sin(2\pi x), g_{21}(x) = 0, g_{12}(x) = sin(\pi x)$, $g_{22}(x) = cos^2(\pi x)$, and $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}), \lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. u_{it} are drawn independently from the student's t distribution with 3 degrees of freedom. The left panel are the estimation results for $g_{12,\tau}(x) = \sin(\pi x)$ and the right panel are the estimation results for $g_{22,\tau}(x)$ with $\tau \in \{0.25, 0.75\}$. For each graph, the blue line is the true function, the red line and the green line are the 95% and 5% empirical quantiles from 1000 replications.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

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

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

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.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of Statistics 44*(1), 219.

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.

Ma, S., O. Linton, and J. Gao (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics 222*(1), 295–323.

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