On-line supplement to Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity

Tomohiro Ando [1] and Jushan Bai [2]

Abstract: Appendix A provides the proof of Proposition 1. Proof of Theorem 1 is given in Appendix B. To prove Theorem 2, we need the uniform convergence rates of estimated parameters. Appendix C provides this result. Theorem 2 and Theorem 3 are proved in Appendix D and E, respectively. In Appendix F, we provide additional information on the empirical analysis. Appendix G reports the Monte Carlo simulation results.

Notation For notational simplicity, we suppress the dependency of τ such that $\varepsilon_{it,\tau} = \varepsilon_{it}$, $\boldsymbol{b}_{i,\tau} = \boldsymbol{b}_i$, $\boldsymbol{\lambda}_{i,\tau} = \boldsymbol{\lambda}_i$, $\boldsymbol{\gamma}_{i,\tau} = (\boldsymbol{b}'_{i,\tau}, \boldsymbol{\lambda}'_{i,\tau})'$, $\boldsymbol{f}_{t,\tau} = \boldsymbol{f}_t$, $B_\tau = B$, $\Lambda_\tau = \Lambda$, $F_\tau = F$, etc throughout the proof. Also, we denote the true parameters as $\boldsymbol{b}_{i,0}$, $\boldsymbol{\lambda}_{i,0}$, $\boldsymbol{\gamma}_{i,0}$, $\boldsymbol{f}_{t,0}$, F_0, Λ_0 and B_0, etc.

A Proof of Proposition 1

The following lemma is used in the proof of Proposition 1.

Lemma 1 *(Lemma 2.2.10 of Van der Vaart and Wellner (1996)) Let X_1,...,X_n be arbitrary random variables that satisfy the tail bound:*

$$P(|X_i| > z) \leq 2\exp\left(-\frac{1}{2} \times \frac{z^2}{a + bz}\right)$$

for all z (and all i) and fixed $a, b > 0$. Then,

$$E\left|\max_{1\leq i\leq n} X_i\right| \leq C\left(b \times \log(n+1) + \sqrt{a \times \log(n+1)}\right)$$

for some positive constant C.

[1]Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.

[2]Department of Economics, Columbia University, jb3064@columbia.edu. 1019 International Affairs Building 420 West 118 Street New York, NY 10027 USA

Lemma 2 *(Lemma 2.2.11 of Van der Vaart and Wellner (1996)) Suppose that X_1,...,X_n be independent random variables with zero mean such that $E|X_i|^m \leq m!M^{m-2}v_i/2$, for every $m \geq 2$ (and all i) and some constants M and v_i. Then:*

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq z\right) \leq 2\exp\left(-\frac{z^2}{v+Mz}\right),$$

for $v \geq v_1 + \cdots v_n$.

A.1 Proof of Proposition 1

We first show

$$\sup_{\boldsymbol{f}_t \in \mathcal{F}, \boldsymbol{\lambda}_i \in \mathcal{L}, \boldsymbol{b}_i \in \mathcal{B}} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i) - E\left[\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)\right] \right\} \right| = o_p(1), \quad (1)$$

where $E[\cdot]$ is the expectation of y_{it} conditioned on X, F_0, Λ_0 and B_0. For any $e > 0$, $\mathcal{F}$, $\mathcal{L}$, and $\mathcal{B}$ are covered by $\cup_{\boldsymbol{f} \in \mathcal{F}} B_e(\boldsymbol{f})$, $\cup_{\boldsymbol{\lambda} \in \mathcal{L}} B_e(\boldsymbol{\lambda})$ and $\cup_{\boldsymbol{b} \in \mathcal{B}} B_e(\boldsymbol{b})$, respectively. Here $B_r(z)$ is a closed ball with center z and radius $r > 0$. Because of their compactness by Assumptions A and B, there exist finite positive integers C_F, C_Λ, C_B, and $\{\bar{\boldsymbol{f}}_1, ..., \bar{\boldsymbol{f}}_{C_F}\} \in \mathcal{F}$, $\{\bar{\boldsymbol{\lambda}}_1, ..., \bar{\boldsymbol{\lambda}}_{C_\Lambda}\} \in \mathcal{L}$, $\{\bar{\boldsymbol{b}}_1, ..., \bar{\boldsymbol{b}}_{C_B}\} \in \mathcal{B}$ such that

$$\mathcal{F} \subset \cup_{k=1}^{C_F} B_e(\bar{\boldsymbol{f}}_k), \quad \mathcal{L} \subset \cup_{k=1}^{C_\Lambda} B_e(\bar{\boldsymbol{\lambda}}_k), \quad \mathcal{B} \subset \cup_{k=1}^{C_B} B_e(\bar{\boldsymbol{b}}_k).$$

Define

$$\mathcal{J} \equiv \{\boldsymbol{x}'\boldsymbol{b} + \boldsymbol{f}'\boldsymbol{\lambda} : \ \boldsymbol{b} \in \mathcal{B}, \ \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{\lambda} \in \mathcal{L}\}.$$

The set $\mathcal{J}$ is covered by $\cup_{k=1}^{C_B} \cup_{u=1}^{C_F} \cup_{v=1}^{C_\Lambda} H(\bar{\boldsymbol{b}}_k, \bar{\boldsymbol{f}}_u, \bar{\boldsymbol{\lambda}}_v)$, where $H(\bar{\boldsymbol{b}}_k, \bar{\boldsymbol{f}}_u, \bar{\boldsymbol{\lambda}}_v) = \{\boldsymbol{x}'\boldsymbol{b} + \boldsymbol{f}'\boldsymbol{\lambda}; \ \boldsymbol{b} \in B_e(\bar{\boldsymbol{b}}_k), \ \boldsymbol{f} \in B_e(\bar{\boldsymbol{f}}_u), \ \boldsymbol{\lambda} \in B_e(\bar{\boldsymbol{\lambda}}_v)\}$. Suppose $\boldsymbol{x}'\boldsymbol{b} + \boldsymbol{f}'\boldsymbol{\lambda} \in H(\bar{\boldsymbol{b}}_k, \bar{\boldsymbol{f}}_u, \bar{\boldsymbol{\lambda}}_v)$. We then have

$$\begin{aligned}
& |\boldsymbol{x}'\boldsymbol{b} + \boldsymbol{f}'\boldsymbol{\lambda} - (\boldsymbol{x}'\bar{\boldsymbol{b}}_k + \bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v)| \\
\leq \ & \|\boldsymbol{x}'\boldsymbol{b} - \boldsymbol{x}'\bar{\boldsymbol{b}}_k\| + \|\boldsymbol{f}'\boldsymbol{\lambda} - \bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v\| \\
\leq \ & \|\boldsymbol{x}\| \times \|\boldsymbol{b} - \bar{\boldsymbol{b}}_k\| + \|\boldsymbol{f} - \bar{\boldsymbol{f}}_u\| \times \|\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}}_v\| + \|\bar{\boldsymbol{f}}_u\| \times \|\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}}_v\| + \|\boldsymbol{f} - \bar{\boldsymbol{f}}_u\| \times \|\bar{\boldsymbol{\lambda}}_v\| \\
\leq \ & eM + e^2 + eM + eM \\
\equiv \ & \delta,
\end{aligned}$$

where $M < \infty$ is an upper bound (in terms of norm) for the elements in $\mathcal{L}$, $\mathcal{F}$ and $\mathcal{X}$ (by compactness assumption). Note that $H(\bar{\boldsymbol{b}}_k, \bar{\boldsymbol{f}}_u, \bar{\boldsymbol{\lambda}}_v) \subset B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_s + \bar{\boldsymbol{f}}'_u\bar{\boldsymbol{\lambda}}_v)$. Thus,

$$\mathcal{J} \subset \cup_{k=1}^{C_B} \cup_{u=1}^{C_F} \cup_{v=1}^{C_\Lambda} B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_k + \bar{\boldsymbol{f}}'_u\bar{\boldsymbol{\lambda}}_v),$$

which implies that, for any $\delta > 0$, there exist finite integers C_B, C_F and C_Λ such that $\mathcal{J}$ is covered by the union of $C_B \times C_F \times C_\Lambda$ closed balls $B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_1 + \bar{\boldsymbol{f}}'_1\bar{\boldsymbol{\lambda}}_1), ..., B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_{C_B} + \bar{\boldsymbol{f}}'_{C_F}\bar{\boldsymbol{\lambda}}_{C_\Lambda})$.

For each $\boldsymbol{x}_{it}$, $\boldsymbol{f}_t$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{b}_i$, we can identify $\bar{\boldsymbol{f}}_u$, $\bar{\boldsymbol{\lambda}}_v$ and $\bar{\boldsymbol{b}}_k$, such that $\boldsymbol{f}_t \in B_e(\bar{\boldsymbol{f}}_u)$, $\boldsymbol{\lambda}_i \in B_e(\bar{\boldsymbol{\lambda}}_v)$ and $\boldsymbol{b}_i \in B_e(\bar{\boldsymbol{b}}_k)$. Define $h_{it} = \boldsymbol{x}'_{it}\boldsymbol{b}_i + \boldsymbol{f}'_t\boldsymbol{\lambda}_i$ and

$$d_{it} := d_{it,k,u,v} = \boldsymbol{x}'_{it}\bar{\boldsymbol{b}}_k + \bar{\boldsymbol{f}}'_u\bar{\boldsymbol{\lambda}}_v,$$

then $|h_{it} - d_{it}| \leq \delta$. Thus,

$$\begin{aligned}
& \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_\tau(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{b}_i - \boldsymbol{f}'_t\boldsymbol{\lambda}_i) - E\left[\rho_\tau(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{b}_i - \boldsymbol{f}'_t\boldsymbol{\lambda}_i)\right]\} \right| \\
= & \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_\tau(y_{it} - h_{it}) - E\left[\rho_\tau(y_{it} - h_{it})\right]\} \right| \\
\leq & \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_\tau(y_{it} - h_{it}) - \rho_\tau(y_{it} - d_{it})\} \right| \\
& + \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{E[\rho_\tau(y_{it} - h_{it})] - E[\rho_\tau(y_{it} - d_{it})]\} \right| \\
& + \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_\tau(y_{it} - d_{it}) - E[\rho_\tau(y_{it} - d_{it})]\} \right| \\
= & I_1 + I_2 + I_3.
\end{aligned}$$

Now,

$$\begin{aligned}
I_1 & \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} |\rho_\tau(y_{it} - h_{it}) - \rho_\tau(y_{it} - d_{it})| \\
& \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{s_\tau(y_{it}, \delta) - E[s_\tau(y_{it}, \delta)]\} + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E[s_\tau(y_{it}, \delta)],
\end{aligned}$$

where $s_\tau(y, \delta) = \sup_{a,b \in \mathcal{J}, |a-b| \leq \delta} |\rho_\tau(y - a) - \rho_\tau(y - b)|$. The first term in the last line is $o_p(1)$ by a law of large numbers. By choosing a small enough δ, the second term can be made arbitrarily small. Similarly, I_2 can be made arbitrarily small by choosing a small enough δ.

Finally, we consider I_3. Note that $d_{it} = d_{it,k,u,v}$, I_3 is uniformly bounded by

$$\begin{aligned}
& \sup_{k,u,v} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_\tau(y_{it} - d_{it,k,u,v}) - E\left[\rho_\tau(y_{it} - d_{it,k,u,v})\right]\} \right| \\
= & \max_{d_j \in \mathcal{S}} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_\tau(y_{it} - d_j) - E\left[\rho_\tau(y_{it} - d_j)\right]\} \right|,
\end{aligned}$$

where

$$\mathcal{S} \equiv \left\{ \{d_{it,k,u,v}\}_{i=1,...,N,t=1,...,T} : \bar{\boldsymbol{f}}_u \in \{\bar{\boldsymbol{f}}_1, ..., \bar{\boldsymbol{f}}_{C_F}\}, \bar{\boldsymbol{\lambda}}_v \in \{\bar{\boldsymbol{\lambda}}_1, ..., \bar{\boldsymbol{\lambda}}_{C_\Lambda}\}, \bar{\boldsymbol{b}}_k \in \{\bar{\boldsymbol{b}}_1, ..., \bar{\boldsymbol{b}}_{C_B}\} \right\}$$

Note that there is an abuse of notation, but the idea should be clear. The maximum of number of elements of $\mathcal{S}$ is $C_B^N C_\Lambda^N C_F^T$ when k, u, v vary.

Note that

$$\begin{aligned}
& \sup_{h \in \mathcal{J}} |\rho_\tau(y_{it} - h) - E[\rho_\tau(y_{it} - h)]| \\
\leq \; & \sup_{h \in R} \left| \rho_\tau(\boldsymbol{x}_{it}' \boldsymbol{b}_{i,0} + \boldsymbol{f}_{t,0}' \boldsymbol{\lambda}_{i,0} + \varepsilon_{it} - h) - E\left[\rho_\tau(\boldsymbol{x}_{it}' \boldsymbol{b}_{i,0} + \boldsymbol{f}_{t,0}' \boldsymbol{\lambda}_{i,0} + \varepsilon_{it} - h)\right] \right| \\
= \; & \sup_{h \in R} |\rho_\tau(\varepsilon_{it} - h) - E[\rho_\tau(\varepsilon_{it} - h)]| \\
\leq \; & |\varepsilon_{it} - E[\varepsilon_{it}]| + E[|\varepsilon_{it} - E[\varepsilon_{it}]|], \qquad (2)
\end{aligned}$$

where the first inequality is obvious because $\mathcal{J} \subset R$. The second inequality can be obtained as follows. For any $X, z \in R$ with X being a random variable, we have $|\rho_\tau(z - h) - \rho_\tau(X - h)| \leq \max\{\tau, 1-\tau\}|z - X| \leq |z - X|$. This leads $|\rho_\tau(z-h) - E[\rho_\tau(X-h)]| \leq E[|\rho_\tau(z-h) - \rho_\tau(X-h)|] \leq E[|X - z|]$, which is further bounded by $|z - E[X]| + E[|X - E[X]|]$ (See Eq (13) in Bai (1998)). Thus, we have, for positive number $K(\geq 1)$,

$$\begin{aligned}
& E|\rho_\tau(y_{it} - d_j) - E[\rho_\tau(y_{it} - d_j)]|^K \\
\leq \; & E\left[\left| |\varepsilon_{it} - E[\varepsilon_{it}]| + E[|\varepsilon_{it} - E[\varepsilon_{it}]|] \right|^K \right] \\
\leq \; & 2^{K-1} \left[E\left| \varepsilon_{it} - E[\varepsilon_{it}] \right|^K + \left(E|\varepsilon_{it} - E[\varepsilon_{it}]| \right)^K \right],
\end{aligned}$$

where the first inequality follows from (2). From Assumption C, $E[|\varepsilon_{it} - E[\varepsilon_{it}]|^K]$ is bounded by $E[|\varepsilon_{it} - E[\varepsilon_{it}]|^K] \leq K! C_\varepsilon^K$ for some finite constant C_ε. Similarly, from Assumption C, $(E|\varepsilon_{it} - E[\varepsilon_{it}]|)^K$ is bounded by $(E|\varepsilon_{it} - E[\varepsilon_{it}]|)^K \leq C_\varepsilon^K$. Thus, we obtain

$$E\left[|\rho_\tau(y_{it} - d_j) - E[\rho_\tau(y_{it} - d_j)]|^K \right] \leq 2^{K-1} K! C_\varepsilon^K + 2^{K-1} C_\varepsilon^K \leq K! C_1^{K-2} C_2/2 \qquad (3)$$

for all i and t. Here C_1 and C_2 are positive constant. To obtain the last inequality, for example, we can take $C_1 = 2C_\varepsilon$ and $C_2 = (2C_\varepsilon)^2$. Define

$$Z_{it,j} \equiv \rho_\tau(y_{it} - d_j) - E[\rho_\tau(y_{it} - d_j)].$$

It then follows from Lemma 2 that, for all $z \geq 0$, all j,

$$P\left(\left| \sum_{i=1}^N \sum_{t=1}^T Z_{it,j} \right| \geq z \right) \leq 2 \exp\left(-\frac{z^2/2}{(NT) \times C_2 + C_1 z} \right) \qquad (4)$$

where we apply Lemma 2 with $M = C_1$ and $v = NTC_2$.

Recall the cardinality of $\mathcal{S}$ is

$$|\mathcal{S}| \leq {C_B}^N \times {C_F}^T \times {C_\Lambda}^N = O(C_{\max}^{2N+T}) \tag{5}$$

with $C_{\max} = \max\{C_B, C_F, C_\Lambda\} = O(1)$. From (4) and Lemma 1, we finally have

$$\begin{aligned}
& E\left[\max_{d_j \in \mathcal{S}} \frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it}-d_j) - E\left[\rho_\tau(y_{it}-d_j)\right]\}\right|\right] \\
\leq\ & C \times \frac{1}{NT}\left(C_1 \log(1+|\mathcal{S}|) + \sqrt{(NT)C_2}\{\log(1+|\mathcal{S}|)\}^{1/2}\right) \\
\leq\ & C' \times \frac{1}{NT}\left[(2N+T)\log C_{\max} + (NT)^{1/2}(2N+T)^{1/2}\log C_{\max}\right] \\
=\ & O\left(\frac{1}{N^{1/2}} + \frac{1}{T^{1/2}}\right),
\end{aligned}$$

where C and C' are some positive constant. Here the first inequality is obtained by applying Lemma 1 with $n = |\mathcal{S}|$, $a = NTC_2$, $b = C_1$ in view of (4). The second inequality is due to (5). Summarizing these results, we obtain (1).

Recall that the estimator $\{\hat{B}, \hat{F}, \hat{\Lambda}\}$ is the minimizer of $\ell_\tau(B, F, \Lambda) = \sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)$. Then, for any given F, Λ and B, we have

$$\begin{aligned}
& U_\tau(B, F, \Lambda) \\
\equiv\ & \frac{1}{NT}\ell_\tau(B, F, \Lambda) - \frac{1}{NT}\ell_\tau(B_0, F_0, \Lambda_0) \\
=\ & \left[\frac{1}{NT}\ell_\tau(B, F, \Lambda) - \frac{1}{NT}E[\ell_\tau(B, F, \Lambda)]\right] + \left[\frac{1}{NT}\ell_\tau(B_0, F_0, \Lambda_0) - \frac{1}{NT}E[\ell_\tau(B_0, F_0, \Lambda_0)]\right] \\
& + \left[\frac{1}{NT}E[\ell_\tau(B, F, \Lambda)] - \frac{1}{NT}E[\ell_\tau(B_0, F_0, \Lambda_0)]\right] \\
=\ & J_1 + J_2 + J_3,
\end{aligned}$$

where $E[\ell_\tau(B, F, \Lambda)]$ is defined as $E[\ell_\tau(B, F, \Lambda)] \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E[\rho_\tau(y_{it} - \boldsymbol{x}_{it}\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)]$. From (1), we have

$$\sup_{\boldsymbol{f}_t \in \mathcal{F}, \boldsymbol{\lambda}_i \in \mathcal{L}, \boldsymbol{b}_i \in \mathcal{B}} |J_1| = o_p(1), \quad \text{and} \quad |J_2| = o_p(1).$$

About J_3, note that $\varepsilon_{it} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$, with conditional density $g_{it}(\cdot|x_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0})$. By Taylor expansion, we obtain

$$\begin{aligned}
& E\left[\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)\right] - E\left[\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})\right] \\
=\ & E\left[\rho_\tau(\varepsilon_{it} - \{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\})\right] - E\left[\rho_\tau(\varepsilon_{it})\right] \\
=\ & g_{it}\left(\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i \middle| \boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}\right) \times \left(\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2 \\
\geq\ & \bar{g} \times \left(\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2,
\end{aligned}$$

where $\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i$ is between 0 and $\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$. By assumption, $\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i$ belongs to a compact set. This leads to $0 < \bar{g} \leq g_{it}(\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i | x_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0})$ by Assumption C. Therefore, the last inequality was obtained. Then,

$$\begin{aligned} J_3 &= \frac{1}{NT} E\left[\ell_\tau\left(B, F, \Lambda\right)\right] - \frac{1}{NT} E\left[\ell_\tau\left(B_0, F_0, \Lambda_0\right)\right] \\ &\geq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \bar{g} \times \left(\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2. \end{aligned}$$

Without loss of generality, we assume that $\boldsymbol{b}_{i,0} = \boldsymbol{0}$, $i = 1, ..., N$ (for notational simplicity). Note that the centered objective function satisfies

$$U_\tau\left(B_0, F_0, \Lambda_0\right) = 0,$$

where we used that the function $h(\boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i) \equiv \boldsymbol{x}_{i,t}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$ evaluated at $\boldsymbol{b}_i = \boldsymbol{b}_{i,0}$, $\boldsymbol{f}_t = \boldsymbol{f}_{t,0}$ and $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{i,0}$ is zero. Note also that

$$U_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \leq U_\tau\left(B_0, F_0, \Lambda_0\right) = 0$$

by definition of $\{\hat{B}, \hat{F}, \hat{\Lambda}\}$. Therefore, we have

$$\begin{aligned} 0 &\geq U_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \\ &\geq \frac{\bar{g}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{i,t}'\hat{\boldsymbol{b}}_i + \hat{\boldsymbol{f}}_t'\hat{\boldsymbol{\lambda}}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2 + o_p(1). \end{aligned}$$

Combined with $\frac{\bar{g}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{i,t}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})^2 \geq 0$ for any $\boldsymbol{b}_i$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$, it must be true that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{i,t}'\hat{\boldsymbol{b}}_i + \hat{\boldsymbol{f}}_t'\hat{\boldsymbol{\lambda}}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2 = \frac{1}{NT}\sum_{i=1}^{N}\left\|X_i\hat{\boldsymbol{b}}_i + \hat{F}\hat{\boldsymbol{\lambda}}_i - F_0\boldsymbol{\lambda}_{i,0}\right\|^2 = o_p(1). \tag{6}$$

Define $M_F = I - F(F'F)^{-1}F'$ and

$$\tilde{U}_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \equiv \frac{1}{NT}\sum_{i=1}^{N}\left\|M_{\hat{F}}\left(X_i\hat{\boldsymbol{b}}_i - F_0\boldsymbol{\lambda}_{i,0}\right)\right\|^2 \tag{7}$$

then

$$\tilde{U}_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \leq \frac{1}{NT}\sum_{i=1}^{N}\left\|X_i\hat{\boldsymbol{b}}_i + \hat{F}\hat{\lambda}_i - F_0\boldsymbol{\lambda}_{i,0}\right\|^2 = o_p(1),$$

by equation (6). This is because M_F is a projection matrix, $\|M_F Z\| \leq \|Z\|$ for any Z, and also $M_{\hat{F}}\hat{F} = 0$.

Now, by (7)

$$\begin{aligned}
&\tilde{U}_\tau\left(\hat{B},\hat{F},\hat{\Lambda}\right)\\
=\;& \frac{1}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'X_i'M_{\hat{F}}X_i\hat{\boldsymbol{b}}_i-\frac{2}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'X_i'M_{\hat{F}}F_0\boldsymbol{\lambda}_{i,0}+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}'F_0{}'M_{\hat{F}}F_0\boldsymbol{\lambda}_{i,0}\\
=\;& \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'A_i\hat{\boldsymbol{b}}_i-\frac{2}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'C_i\boldsymbol{\eta}+\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}'B_i\boldsymbol{\eta},
\end{aligned}$$

where

$$A_i=\frac{1}{T}X_i'M_{\hat{F}}X_i,\quad B_i=(\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}_{i,0}')\otimes I_T,\quad C_i'=\frac{1}{\sqrt{T}}\boldsymbol{\lambda}_{i,0}'\otimes(X_i'M_{\hat{F}}),\quad \boldsymbol{\eta}=\frac{1}{\sqrt{T}}\mathrm{vec}(M_{\hat{F}}F_0).$$

Completing the square,

$$\tilde{U}_\tau\left(\hat{B},\hat{F},\hat{\Lambda}\right)=\boldsymbol{\eta}'\left(\frac{1}{N}\sum_{i=1}^{N}E_i\right)\boldsymbol{\eta}+\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i+A_i^{-1}C_i\boldsymbol{\eta}\right)'A_i\left(\hat{\boldsymbol{b}}_i+A_i^{-1}C_i\boldsymbol{\eta}\right),$$

where $E_i=B_i-C_i'A_i^{-1}C_i$. Because each of the two terms is non-negative, this implies that

$$\boldsymbol{\eta}'\left(\frac{1}{N}\sum_{i=1}^{N}E_i\right)\boldsymbol{\eta}=o_p(1), \tag{8}$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i+A_i^{-1}C_i\boldsymbol{\eta}\right)'A_i\left(\hat{\boldsymbol{b}}_i+A_i^{-1}C_i\boldsymbol{\eta}\right)=o_p(1). \tag{9}$$

From Assumption D, the matrix $N^{-1}\sum_{i=1}^{N}E_i$ is positive definite, and thus equation (8) implies that $\|\boldsymbol{\eta}\|^2=o_p(1)$. This result implies that

$$\|M_{\hat{F}}-M_{F_0}\|=\|P_{\hat{F}}-P_{F_0}\|=o_p(1), \tag{10}$$

where $P_F=F(F'F)^{-1}F'$. See Bai (2009, page 1265). That is, the space spanned by F_0 and the space spanned by the estimated factors $\hat{F}$ are asymptotically the same. We then have

$$\begin{aligned}
\|M_{F_0}\hat{F}\|/T^{1/2} &= \|(M_{F_0}-M_{\hat{F}})\hat{F}\|/T^{1/2}\\
&= \|(P_{F_0}-P_{\hat{F}})\hat{F}\|/T^{1/2}\\
&\le \|P_{F_0}-P_{\hat{F}}\|\times(\|\hat{F}\|/T^{1/2})\\
&= o_p(1)\times O_p(1),
\end{aligned}$$

where we used $\|\hat{F}\|/T^{1/2}=O_p(1)$. This implies that

$$\|\hat{F}-F_0(F_0'F_0)^{-1}F_0'\hat{F}\|/T^{1/2}=\|\hat{F}-F_0H\|/T^{1/2}=o_p(1),$$

where $H = (F_0'F_0)^{-1}F_0'\hat{F}$ is the rotation matrix. Because we omit this rotation matrix H in Proposition 1, as explained in the main text, we obtain $\|\hat{F} - F_0\|/\sqrt{T} = o_p(1)$.

From $\|\boldsymbol{\eta}\|^2 = o_p(1)$, equation (9) implies that

$$o_p(1) = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i' A_{i,0}\hat{\boldsymbol{b}}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'(A_i - A_{i,0})\hat{\boldsymbol{b}}_i \geq (\rho_A + o_p(1))\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'\hat{\boldsymbol{b}}_i,$$

where $0 < \rho_A$ is the lower bound of the eigenvalues of $A_{i,0} = \frac{1}{T}X_i'M_{F_0}X_i$ $i = 1, ..., N$. Because assumption (D2), $\rho_A > 0$ exists. We also used the fact that $\|A_i - A_{i,0}\| = o_p(1)$, which is proved in Ando and Bai (2015, Theorem 1). The average consistency of $\hat{\boldsymbol{b}}_i$ follows from $\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'\hat{\boldsymbol{b}}_i = o_p(1)$ (recall we normalize $b_{i,0} = 0$)). The average consistency of $\hat{\boldsymbol{f}}_t$ and the average consistency of $\hat{\boldsymbol{b}}_i$ further imply the average consistency of $\hat{\boldsymbol{\lambda}}_i$ (see Ando and Bai (2015)). That is, $N^{-1}\sum_{i=1}^{N}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|^2 = o_p(1)$. This completes the proof of Proposition 1.

B Proof of Theorem 1

Lemma 3 *(Theorem 2 of Hoeffding (1963)). Let X_1,...,X_n be independent random variables and bounded by the interval $[C_i, D_i]$. Then, for all $r > 0$, we have*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - E\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]\right| \geq r\right) \leq \exp\left(-\frac{2n^2r^2}{\sum_{i=1}^{n}(D_i - C_i)^2}\right).$$

B.1 Proof of Theorem 1

We first prove the uniform consistency of $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{\lambda}}_i$. Let $\boldsymbol{\gamma}_i = (\boldsymbol{b}_i', \boldsymbol{\lambda}_i')'$. We now define the following loss

$$L_{NT,i}(\boldsymbol{\gamma}_i, F) \equiv \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right),$$

and its centered version

$$\begin{aligned}\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) &\equiv L_{NT,i}(\boldsymbol{\gamma}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\\ &= \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right) - \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)\end{aligned}$$

For each $\delta > 0$, we also define

$$B_{i,T}(\delta) \equiv \left\{\boldsymbol{\gamma}_i : \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq \delta\right\}$$

and

$$B_T(e) \equiv \left\{F : \|F - F_0\|/T^{1/2} \leq e\right\},$$

where $e > 0$ is a small constant. We prove the uniform consistency of $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{\lambda}}_i$ in the sense of $\max_{1\leq i\leq N} \|\hat{\boldsymbol{b}}_i - \boldsymbol{b}_{i,0}\| = o_p(1)$ and $\max_{1\leq i\leq N} \|\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{i,0}\| = o_p(1)$. By Proposition 1, it is enough to prove the uniform consistency under the condition that the common factor F satisfies $\|F - F_0\|/T^{1/2} < e$, for small e.

For each $\boldsymbol{\gamma}_i \notin B_{i,T}(\delta)$, define

$$\tilde{\boldsymbol{\gamma}}_i = s_i\boldsymbol{\gamma}_i + (1 - s_i)\boldsymbol{\gamma}_{i,0},$$

where $\boldsymbol{\gamma}_{i,0} = (\boldsymbol{b}'_{i,0}, \boldsymbol{\lambda}'_{i,0})'$ and $s_i = \delta/\{\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\|\} \in (0, 1)$. The convexity of the objective function, given F, leads

$$\begin{aligned}
& s_i \left\{L_{NT,i}(\boldsymbol{\gamma}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F)\right\} \\
\geq\ & L_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F) \\
=\ & L_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) + L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F) \\
\geq\ & E\left[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)\right] + \left\{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E\left[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)\right]\right\} - C \times e, \qquad (11)
\end{aligned}$$

where $E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)] = E[L_{NT,i}(\boldsymbol{\gamma}_i, F)] - E[L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0)]$ and the expectation $E[\cdot]$ is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$ conditional on X, F_0, Λ_0 and B_0. Here, we have used, for $F \in B_T(e)$,

$$|L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F)| \leq C \times T^{-1}\sum_{t=1}^{T} \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq C \times \|F - F_0\|/T^{1/2} \leq C \times e$$

with C being a positive constant. By choosing a small e, this term is dominated by the term $E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] + \{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)]\}$. Thus, this term is negligible in our analysis below.

Let $\boldsymbol{\omega}_{it} = \{\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}\}$. Under $\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq \delta$, the identity of Knight (1998) leads to

$$\begin{aligned}
& E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)] \\
=\ & \frac{1}{T}\sum_{t=1}^{T} E\left[\rho_\tau\left(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{b}_i - \boldsymbol{f}'_t\boldsymbol{\lambda}_i\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{b}_{i,0} - \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0}\right)\right] \\
=\ & \frac{1}{T}\sum_{t=1}^{T} \int_0^{\boldsymbol{x}'_{it}(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}'_{t,0}(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}) + (\boldsymbol{f}_t - \boldsymbol{f}_{t,0})'\boldsymbol{\lambda}_i} \left\{G_{it}(s|\boldsymbol{\omega}_{it}) - \tau\right\} ds \\
=\ & \frac{1}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it}) \left\{\boldsymbol{x}'_{it}\left(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}\right) + \boldsymbol{f}'_{t,0}\left(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}\right) + \left(\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\right)'\boldsymbol{\lambda}_i\right\}^2
\end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{1}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it})\left\{\boldsymbol{x}_{it}'\left(\boldsymbol{b}_i-\boldsymbol{b}_{i,0}\right)+\boldsymbol{f}_{t,0}'\left(\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_{i,0}\right)\right\}^2+\frac{1}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it})\left\{\left(\boldsymbol{f}_t-\boldsymbol{f}_{t,0}\right)'\boldsymbol{\lambda}_i\right\}^2 \\
& +\frac{2}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it})\left\{\boldsymbol{x}_{it}'\left(\boldsymbol{b}_i-\boldsymbol{b}_{i,0}\right)+\boldsymbol{f}_{t,0}'\left(\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_{i,0}\right)\right\}\left\{\left(\boldsymbol{f}_t-\boldsymbol{f}_{t,0}\right)'\boldsymbol{\lambda}_i\right\} \qquad (12)
\end{aligned}
$$

where $\tilde{h}_{it}$ is between 0 and $\boldsymbol{x}_{it}'(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+\boldsymbol{f}_{t,0}'(\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_{i,0})+(\boldsymbol{f}_t-\boldsymbol{f}_{t,0})'\boldsymbol{\lambda}_i$, $G_{it}(\cdot|\boldsymbol{\omega}_{it})$ is the conditional cumulative distribution of $\varepsilon_{it}=y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0}-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$ conditional on $\boldsymbol{\omega}_{it}$. For $F\in B_T(e)$ with small e, the first term in (12) dominates the other two terms.

By assumptions, $\tilde{h}_{it}$ belongs to a compact set, which leads to $0<\bar{g}\le g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it})$. Together with Assumption (D2), there exists a constant $c_\delta>0$ such that

$$
\frac{1}{T}\sum_{t=1}^{T}\bar{g}\left\{\boldsymbol{x}_{it}'\left(\boldsymbol{b}_i-\boldsymbol{b}_{i,0}\right)+\boldsymbol{f}_{t,0}'\left(\boldsymbol{\lambda}_{i,0}-\boldsymbol{\lambda}_i\right)\right\}^2>c_\delta \qquad (13)
$$

for all $1\le i\le N$, all $\|\boldsymbol{\gamma}_i-\boldsymbol{\gamma}_{i,0}\|=\delta$. This means that, with large probability, $E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i,F)]\ge c_\delta/2$ for all $\|\boldsymbol{\gamma}_i-\boldsymbol{\gamma}_{i,0}\|=\delta$, and all $F\in B_T(e)$ with small e.

Note that $\tilde{\boldsymbol{\gamma}}_i$ is on the boundary of $B_{i,T}(\delta)$; i.e., $\tilde{\boldsymbol{\gamma}}_i\in\partial B_{i,T}(\delta)\equiv\{\boldsymbol{\gamma}_i:\|\boldsymbol{\gamma}_i-\boldsymbol{\gamma}_{i,0}\|=\delta\}$. This implies that with probability approaching to 1,

$$
E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i,F)]\ge\frac{c_\delta}{2} \qquad (14)
$$

for all $\tilde{\boldsymbol{\gamma}}_i\in\partial B_{i,T}(\delta)$, all $F\in B_T(e)$, and all small e.

For any two events A and B, from $P(A)\le P(AB)+P(B^c)$, we have

$$
P(\max_i\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|>\delta)\le P\Big(\max_i\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|>\delta,\ \hat{F}\in B_T(e)\Big)+P\Big(\hat{F}\notin B_T(e)\Big)
$$

By Proposition 1, the second term converges to zero. We focus on the first term. Consider the event

$$
\begin{aligned}
& \left\{\max_i\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|>\delta,\ \hat{F}\in B_T(e)\right\} \\
= \ & \left\{\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|>\delta,\ \exists i,\ \hat{F}\in B_T(e)\right\} \\
= \ & \left\{L_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})\le L_{NT,i}(\boldsymbol{\gamma}_{i,0},\hat{F}),\ \exists i,\hat{\boldsymbol{\gamma}}_i\notin B_{i,T}(\delta),\ \hat{F}\in B_T(e)\right\} \\
\subset \ & \left\{L_{NT,i}(\boldsymbol{\gamma}_i,F)\le L_{NT,i}(\boldsymbol{\gamma}_{i,0},F),\ \exists i,\exists\boldsymbol{\gamma}_i\notin B_{i,T}(\delta),\exists F\in B_T(e)\right\} \\
\subset \ & \left\{\max_{1\le i\le N}\sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),\ F\in B_T(e)}\left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i,F)-E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i,F)]\right|\ge\frac{c_\delta}{3}\right\},
\end{aligned}
$$

The first two equalities and the first inclusion ($\subset$) follow easily. We explain why the second inclusion holds. From (11), we have

$$
s_i\left\{L_{NT,i}(\boldsymbol{\gamma}_i,F)-L_{NT,i}(\boldsymbol{\gamma}_{i0},F)\right\}\ge E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i,F)]+\{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i,F)-E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i,F)]\}-Ce.
$$

If $L_{NT,i}(\boldsymbol{\gamma}_i, F) \leq L_{NT,i}(\boldsymbol{\gamma}_{i,0}, F)$ for some i, some $\boldsymbol{\gamma}_i \notin B_{i,T}(\delta)$ and some $F \in B_T(e)$, this leads to $0 \geq E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] + \{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)]\} - Ce$. From (14), $E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] \geq c_\delta/2$, note that $\tilde{\boldsymbol{\gamma}}_i \in \partial B_{i,T}(\delta) \subset B_{i,T}(\delta)$. By choosing a small e, we have

$$\max_{1\leq i\leq N} \sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta),\ F\in B_T(e)} |\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]| \geq c_\delta/3.$$

The second inclusion is obtained. We next show that this event has a probability converging to zero. It suffices to show that for every $\varepsilon > 0$,

$$\lim_{N,T\to\infty} P\left\{\max_{1\leq i\leq N} \sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta),\ F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]\right| \geq \varepsilon\right\} = 0.$$

It further suffices to prove that

$$\max_{1\leq i\leq N} P\left\{\sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta),\ F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]\right| \geq \varepsilon\right\} = o(N^{-1}). \tag{15}$$

Let $h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) \equiv \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i) - \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})$. Observe that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\bar{\boldsymbol{\gamma}}_i, \boldsymbol{f}_{t,0})| \leq C(\|\boldsymbol{x}_{it}\| + \|\boldsymbol{f}_{t,0}\|)\|\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}_i\| + C(\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|)$ and $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_{t,0}) - h(\bar{\boldsymbol{\gamma}}_i, \boldsymbol{f}_{t,0})| \leq C(\|\boldsymbol{x}_{it}\| + \|\boldsymbol{f}_{t,0}\|)\|\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}_i\|$ for some universal constant $C > 0$. Put $W_{it} \equiv C(\|\boldsymbol{x}_{it}\| + \|\boldsymbol{f}_{t,0}\|)$ and $\kappa \equiv \sup_{i,t} W_{it}$. Since $B_{i,T}(\delta)$ is compact, there exist C_Γ balls with centers $\bar{\boldsymbol{\gamma}}_k = (\bar{\boldsymbol{b}}_k', \bar{\boldsymbol{\lambda}}_k)'$, $k = 1, ..., C_\Gamma$ and radius $\varepsilon/(8\kappa)$ such that the collection of these balls covers $B_{i,T}(\delta)$. Note that C_Γ can be chosen such that $C_\Gamma(\varepsilon) = O(1/\varepsilon^{p+r})$ and $\varepsilon \to 0$. For each $\boldsymbol{\gamma}_i \in B_{i,T}(\delta)$, there is $\bar{\boldsymbol{\gamma}}_k$ such that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\bar{\boldsymbol{\gamma}}_k, \boldsymbol{f}_{t,0})| \leq W_{it}\varepsilon/(8\kappa)$. These investigations lead to

$$\begin{aligned}
& |\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]| \\
\leq\ & |\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0)]| + \frac{C}{T}\sum_{t=1}^T \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \\
\leq\ & |\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0) - E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0)]| + \frac{\varepsilon}{8\kappa} \times \frac{1}{NT}\left|\sum_{t=1}^T\sum_{i=1}^N \{W_{it} + E[W_{it}]\}\right| + \frac{C}{T}\sum_{t=1}^T \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|,
\end{aligned}$$

For $F \in B_T(e)$, $T^{-1}\sum_{t=1}^T \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|^2 \leq e$. Thus, $T^{-1}\sum_{t=1}^T \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \sqrt{\sum_{t=1}^T \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|^2/T} \leq \sqrt{e}$. So the last term is small for small e. We have

$$\begin{aligned}
& P\left\{\sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta), F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0)]\right| \geq \varepsilon\right\} \\
\leq\ & \sum_{k=1}^{C_\Gamma(\varepsilon)} P\left\{\left|\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0) - E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0)]\right| \geq \frac{\varepsilon}{3}\right\} \\
& + P\left\{\left|\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \{W_{it} + E[W_{it}]\}\right| \geq \frac{8\kappa}{3}\right\} + P\left\{\left|\frac{1}{T}\sum_{t=1}^T \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\right| \geq \frac{\varepsilon}{3C}\right\},
\end{aligned}$$

where the second term is 0 because $N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\{W_{it}+E[W_{it}]\}$ is smaller than 2κ. The last term is also zero by choosing e smaller than $\varepsilon/(3C)$. Thus, we need to evaluate the first term.

Because of the uniform boundedness of $\boldsymbol{x}_{it}$ and $\boldsymbol{f}_t$, we have $|h(\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h(\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]|\leq M$, where M is a bounded constant. From the independence property of ε_{it}, Lemma 3 leads to

$$\begin{aligned}
& P\left\{\left|\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k,F_0)-E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k,F_0)]\right|\geq\frac{\varepsilon}{3}\right\}\\
= \; & P\left\{\left|\frac{1}{T}\sum_{t=1}^{T}\left\{\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\bar{\boldsymbol{b}}_k-\boldsymbol{f}_{t,0}'\bar{\boldsymbol{\lambda}}_k\right)-E\left[\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\bar{\boldsymbol{b}}_k-\boldsymbol{f}_{t,0}'\bar{\boldsymbol{\lambda}}_k\right)\right]\right\}\right|\geq\frac{\varepsilon}{3}\right\}\\
\leq \; & \exp\left(\frac{-T^2\varepsilon^2}{18\sum_{t=1}^{T}M^2}\right),
\end{aligned}$$

which leads

$$\begin{aligned}
& \sum_{k=1}^{C_\Gamma(\varepsilon)}P\left\{\left|\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k,F_0)-E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k,F_0)]\right|\geq\frac{\varepsilon}{3}\right\}\\
\leq \; & C_\Gamma(\varepsilon)\times\exp\left(\frac{-T^2\varepsilon^2}{18\sum_{t=1}^{T}M^2}\right)\\
= \; & O\left(\varepsilon^{-p-r}\times\exp(-T\varepsilon^2C)\right)\\
= \; & O\left(\exp\left[-T\left\{(\varepsilon^2C+(p+r)\frac{\log(\varepsilon)}{T}\right\}\right]\right)\\
= \; & \frac{1}{N}\times O\left(\exp\left[-T\varepsilon^2C/2+\log(N)\right]\right)\\
= \; & o(N^{-1}),
\end{aligned}\tag{16}$$

where the third equality is obtained because $\log(\varepsilon)/T\to 0$. The last line is obtained because $\log(N)/T^{1/2}\to 0$. Thus, we have

$$\max_{1\leq i\leq N}P\left\{\sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),F\in B_T(e)}\left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i,F)-E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i,F)]\right|\geq\varepsilon\right\}=o(N^{-1}),$$

which completes the proof of the uniform consistency of $\hat{\boldsymbol{\gamma}}_i$ for $i=1,...,N$.

Next, we prove that the estimated common factor is $\hat{\boldsymbol{f}}_t$ is uniformly consistent

$$\max_{1\leq t\leq T}\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|\to 0.\tag{17}$$

Let $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_1',...,\boldsymbol{\gamma}_N')'$. We define the following loss

$$S_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)\equiv\frac{1}{N}\sum_{i=1}^{N}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_i-\boldsymbol{f}_t'\boldsymbol{\lambda}_i\right),$$

and its centered version

$$\begin{aligned}\tilde{S}_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) &\equiv S_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) - S_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) \\ &= \frac{1}{N}\sum_{i=1}^{N} \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right) - \frac{1}{N}\sum_{i=1}^{N} \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right),\end{aligned}$$

where $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{1,0}', ..., \boldsymbol{\gamma}_{N,0}')'$. Fix any $\delta > 0$. For each $\boldsymbol{f}_t$ such that $\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| > \delta$, define $\tilde{\boldsymbol{f}}_t = s_t\boldsymbol{f}_t + (1 - s_t)\boldsymbol{f}_{t,0}$ with $s_t = \delta/\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \in (0, 1)$. Then, $\|\tilde{\boldsymbol{f}}_t - \boldsymbol{f}_t\| = \delta$. Similar to (13), for some $c_\delta > 0$, we obtain the inclusion relation

$$\begin{aligned}&\left\{\max_t \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| > \delta\right\} \\ \equiv\ &\left\{\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| > \delta,\ \exists t\right\} \\ \subset\ &\left\{S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) \le S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}),\ 1 \le \exists t \le T, \exists \boldsymbol{f}_t \text{ s.t. } \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| > \delta\right\} \\ \subset\ &\left\{\max_{1\le t\le T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\le\delta} \left|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \ge \frac{c_\delta}{2}\right\},\end{aligned}$$

where $E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t)] = E[S_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t)] - E[S_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})]$ and the expectation $E[\cdot]$ is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$ conditional on X, F_0 and Λ_0. The second inclusion is obtained as follows. Because of the convexity of the objective function (given $\hat{\boldsymbol{\gamma}}$), we have

$$\begin{aligned}&s_t\left\{S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0})\right\} \\ \ge\ &S_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \\ =\ &\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \\ =\ &\left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t)\right]\right] + E\left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0})\right] \\ &-\left\{\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) - E\left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0})\right]\right\} + E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t)\right].\end{aligned}$$

Similar to (14), for $\|\tilde{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = \delta$, the last term is greater than or equal to c_δ for some $c_\delta > 0$. By consistency of $\hat{\boldsymbol{\gamma}}$, the second and third terms in the last line are $o_p(1)$. Thus, we have

$$\begin{aligned}&s_t\left\{S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0})\right\} \\ \ge\ &\left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t)\right]\right] + c_\delta + o_p(1).\end{aligned}$$

If $S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) \le S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0})$ $1 \le \exists t \le T$ and $\exists \boldsymbol{f}_t$ s.t. $\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| > \delta$, then

$$0 \ge \left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t)\right]\right] + c_\delta + o_p(1).$$

Thus, under large N and T, with large probability, $|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t)]| \geq c_\delta/2$. In particular, noting that $\|\tilde{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = \delta$,

$$\max_{1\leq t\leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\leq\delta} |\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]| \geq c_\delta/2$$

This gives the second inclusion.

Therefore, it suffices to show that for every $\varepsilon > 0$,

$$\lim_{N,T\to\infty} P\left\{\max_{1\leq t\leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\leq\delta} \left|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \geq \varepsilon\right\} = 0.$$

Because

$$\begin{aligned} & \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)] \\ = \; & \left\{\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - \tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)\right\} + \left\{\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right\} \end{aligned}$$

it further suffices to prove that

$$\lim_{N,T\to\infty} P\left\{\max_{1\leq t\leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\leq\delta} \left|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - \tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \geq \varepsilon\right\} = 0, \tag{18}$$

and

$$\lim_{N,T\to\infty} P\left\{\max_{1\leq t\leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\leq\delta} \left|\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \geq \varepsilon\right\} = 0. \tag{19}$$

Since

$$\left|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - \tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)\right| \leq C \times \left\{\frac{1}{N}\sum_{i=1}^N \|\boldsymbol{x}_{it}\|\|\hat{\boldsymbol{b}}_i - \boldsymbol{b}_{i,0}\| + \frac{1}{N}\sum_{i=1}^N \|\boldsymbol{f}_t\|\|\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{i,0}\|\right\},$$

and $\sup_{it}\|\boldsymbol{x}_{it}\| < \infty$ and $\sup_t \|\boldsymbol{f}_t\| < \infty$, consistency of $\hat{\boldsymbol{\gamma}}$ implies (18).

Finally, we prove (19). It is enough to show

$$\max_{1\leq t\leq T} P\left\{\sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\leq\delta} \left|\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \geq \varepsilon\right\} = o(T^{-1}). \tag{20}$$

For each $\delta > 0$, we define

$$B_{t,N}(\delta) \equiv \left\{\boldsymbol{f}_t : \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \delta\right\}.$$

Observe that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\boldsymbol{\gamma}_i, \bar{\boldsymbol{f}}_t)| \leq C\|\boldsymbol{\lambda}_i\| \times \|\bar{\boldsymbol{f}}_t - \boldsymbol{f}_t\|$ for some universal constant $C > 0$. Put $K_i \equiv C\|\boldsymbol{\lambda}_i\|$ and $\beta \equiv \sup_i K_i$. Since $B_{t,N}(\delta)$ is compact subset, there exist

C_f balls with centers $\bar{\boldsymbol{f}}_k$, $k = 1, ..., C_f$ and radius ε/β such that the collection of these balls covers $B_{t,N}(\delta)$. Note that C_f can be chosen such that $C_f(\varepsilon) = O(1/\varepsilon^r)$ and $\varepsilon \to 0$. For each $\boldsymbol{f}_t \in B_{t,N}(\delta)$, there is $\bar{\boldsymbol{f}}_k$ such that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\boldsymbol{\gamma}_i, \bar{\boldsymbol{f}}_k)| \leq K_i\varepsilon/(8\beta)$. These investigations lead to

$$\begin{aligned} & |\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]| \\ \leq \; & |\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \bar{\boldsymbol{f}}_k) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \bar{\boldsymbol{f}}_k)]| + \frac{\varepsilon}{8\beta}\left|\frac{1}{N}\sum_{i=1}^{N}\{K(\boldsymbol{\lambda}_i) + E[K(\boldsymbol{\lambda}_i)]\}\right|, \end{aligned}$$

where C is universal constant. Therefore, we have

$$\begin{aligned} & P\left\{\sup_{F \in B_T(e)} \left|\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \geq \varepsilon\right\} \\ \leq \; & \sum_{k=1}^{C_\Gamma(\varepsilon)} P\left\{\left|\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, F_0) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, F_0)]\right| \geq \frac{\varepsilon}{2}\right\} + P\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\{K_i + E[K_i]\}\right| \geq 4\beta\right\}, \end{aligned}$$

where the second term can be made arbitrarily small under large N and T. The second term is zero because $|\frac{1}{N}\sum_{i=1}^{N}\{K_i + E[K_i]\}| < 2\beta$. Because of the uniform boundedness of $\boldsymbol{x}_{it}$ and $\boldsymbol{f}_t$, we have $|h(\boldsymbol{\gamma}_{i,0}, \boldsymbol{f}_t) - E[h(\boldsymbol{\gamma}_{i,0}, \boldsymbol{f}_t)]| \leq M$ where M is bounded constant. The first term is

$$\begin{aligned} & \sum_{k=1}^{C_f(\varepsilon)} P\left\{\left|\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \bar{\boldsymbol{f}}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \bar{\boldsymbol{f}}_t)]\right| \geq \frac{\varepsilon}{2}\right\} \\ \leq \; & C_f(\varepsilon) \times \exp\left(\frac{-N^2\varepsilon^2}{18\sum_{i=1}^{N} M^2}\right) \\ = \; & O\left(\varepsilon^{-r} \times \exp(-N\varepsilon^2 C)\right) \\ = \; & O\left(\exp\left[-N\left\{\varepsilon^2 C + r\frac{\log(\varepsilon)}{N}\right\}\right]\right) \\ = \; & \frac{1}{T} \times O\left(\exp\left[-N\varepsilon^2 C/2 + \log(T)\right]\right) \\ = \; & o(T^{-1}) \end{aligned} \tag{21}$$

where Lemma 3 is applied to obtain the second inequality. Because $(\log T)/\sqrt{N} \to 0$, the last line is obtained. This implies (20), or equivalently, (19). This completes the proof of Theorem 1.

C Lemma 4

Lemma 4 *Under Assumptions of Theorem 2, the following results hold.*

$$\max_{1 \leq i \leq N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p\left(\frac{\log(N)}{T^{1/2}}\right), \tag{22}$$

$$\max_{1\leq t\leq T}\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|=O_p\left(\frac{\log(T)}{N^{1/2}}\right). \tag{23}$$

These results give the rates for uniform consistency, they are used in the proof of Theorem 2. We first present some preliminary results in order to prove Lemma 4. The first result is Lemma 1 of Babu (1989), which is closely related to the Bernstein inequality.

Lemma 5 *Let X_i be a sequence of independent random variables with mean zero and $|X_i|<d$ for some $d>0$. Let $V\geq\sum_{i=1}^N E[X_i^2]$. Then for all $0<s<1$ and $0\leq a\leq V/(sd)$,*

$$P\left(\left|\sum_{i=1}^N X_i\right|\geq a\right)\leq 2\exp(-a^2s(1-s)/V).$$

To state the next lemma, we introduce some notations. We let $\{\xi_t, t\geq 1\}$ be a stationary process taking values in a measurable space $(\mathcal{S},\mathcal{Q})$. Here $\mathcal{S}$ is a Polish space and $\mathcal{Q}$ is a Borel σ-field. We denote $\mathcal{H}$ being a class of measurable functions on the measurable space $(S,\mathcal{Q})$. For a process $Z(h)$ defined on $(\mathcal{S},\mathcal{Q})$, we define $\|Z(h)\|_{\mathcal{H}}\equiv\sup_{h\in\mathcal{H}}|Z(h)|$. The following lemma is a Bernstein type inequality for centered empirical processes (Talagrand (1996), Bousquet (2002)). The following Talagrand type inequality is due to Proposition B.2. of Kato et al. (2012).

Lemma 6 *Let $\mathcal{H}$ be a pointwise measurable class of functions on the measurable space $(\mathcal{S},\mathcal{Q})$ uniformly bounded by some constant U. Suppose that, for any $h(\cdot)\in\mathcal{H}$, (i) $E[h(\xi_t)]=0$, (ii) $\sup_{h\in\mathcal{H}}[h(\xi_t)^2]\leq\sigma^2$. For $Z=\|\sum_{t=1}^T h(\xi_t)\|_{\mathcal{H}}$, we have*

$$P\left(Z-E[Z]\geq\sqrt{2}s\left(T\sigma^2+2UE[Z]\right)^{1/2}+\frac{s^2U}{3}\right)\leq\exp(-s^2),$$

for all $s>0$.

C.1 Proof of Lemma 4

We define $\boldsymbol{z}_{it}=(\boldsymbol{x}_{it}',\boldsymbol{f}_t')'$, $\boldsymbol{z}_{it,0}=(\boldsymbol{x}_{it}',\boldsymbol{f}_{t,0}')'$, $\hat{\boldsymbol{z}}_{it}=(\boldsymbol{x}_{it}',\hat{\boldsymbol{f}}_t')'$, $\boldsymbol{\omega}_{it}=\{\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0}\}$ and

$$\begin{aligned}
Q_{NT,i}(\boldsymbol{\gamma}_i,F) &\equiv \frac{1}{T}\sum_{t=1}^T\left(\tau-I(\varepsilon_{it}\leq\boldsymbol{x}_{it}'(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+\boldsymbol{f}_t'\boldsymbol{\lambda}_i-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})\right)\boldsymbol{z}_{it},\\
\overline{Q}_{NT,i}(\boldsymbol{\gamma}_i,F) &\equiv E[Q_{NT,i}(\boldsymbol{\gamma}_i,F)]\\
&= \frac{1}{T}\sum_{t=1}^T E\left[\left(\tau-G_{it}(\boldsymbol{x}_{it}'(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+\boldsymbol{f}_t'\boldsymbol{\lambda}_i-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\middle|\boldsymbol{\omega}_{it}\right)\boldsymbol{z}_{it}\right],
\end{aligned}$$

where the expectation of $E[Q_{NT,i}(\boldsymbol{\gamma}_i,F)]$ is taken with respect to the true conditional distribution of $\{y_{it}: i=1,...,N,\ t=1,...,T\}$ conditional on X, F_0 and Λ_0, $G_{it}(\cdot|\boldsymbol{x}_{it})$

is the conditional cumulative distribution function of ε_{it}. Because of the computational property of the quantile regression estimator (Gutenbrunner and Jureckova (1992)), it is known that $\max_{1\leq i\leq N}|Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})|$ is bounded by $O_p(T^{-1}\max_{1\leq i\leq N,\ 1\leq t\leq T}\|\boldsymbol{x}_{it}\|)+O_p(T^{-1}\max_{1\leq t\leq T}\|\hat{\boldsymbol{f}}_t\|)=O_p(T^{-1})$. Here, we used $\max_{1\leq t\leq T}\|\hat{\boldsymbol{f}}_t\|=O_p(1)$. We thus have

$$\begin{aligned} O_p(T^{-1}) &= Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F}) \\ &= Q_{NT,i}(\boldsymbol{\gamma}_{i,0},F_0)+\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F}) \\ &\quad +\left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})-\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})-Q_{NT,i}(\boldsymbol{\gamma}_{i,0},F_0)\right\}. \end{aligned} \tag{24}$$

Expanding $\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})$ at $(\boldsymbol{\gamma}_{i,0},F_0)$, we obtain

$$\begin{aligned} &\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F}) \\ =\ & -\frac{1}{T}\sum_{t=1}^{T}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{z}'_{it,0}](\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0})-\frac{1}{T}\sum_{t=1}^{T}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}](\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}) \\ & +o_p(\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|)+o_p\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|\right), \end{aligned} \tag{25}$$

where we used the result of Theorem 1 such that $O_p(\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|^2)=o_p(\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|)$.

It then follows from (24) and (25) that

$$\begin{aligned} &\Gamma_i(\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0})+o_p(\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|) \\ =\ & Q_{NT,i}(\boldsymbol{\gamma}_{i,0},F_0)+\left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})-\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i,\hat{F})-Q_{NT,i}(\boldsymbol{\gamma}_{i,0},F_0)\right\} \\ & -\left[\frac{1}{T}\sum_{t=1}^{T}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}](\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0})\right]+O_p\left(\frac{1}{T}\right)+o_p\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|\right), \end{aligned} \tag{26}$$

where $\Gamma_i=T^{-1}\sum_{t=1}^{T}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{z}'_{it,0}]$ and we used the result of Theorem 1 such that $O_p(T^{-1}\sum_{t=1}^{T}\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|^2)=o_p(T^{-1}\sum_{t=1}^{T}\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|)$.

Next, we define

$$V_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)\ \equiv\ \frac{1}{N}\sum_{i=1}^{N}\left(\tau-I(\varepsilon_{it}\leq\boldsymbol{x}'_{it}(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+\boldsymbol{f}'_t\boldsymbol{\lambda}_i-\boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0})\right)\boldsymbol{\lambda}_i,$$

and

$$\overline{V}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)\equiv E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\tau-G_{it}(\boldsymbol{x}'_{it}(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+\boldsymbol{f}'_t\boldsymbol{\lambda}_i-\boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0}\big|\boldsymbol{\omega}_{it})\right)\boldsymbol{\lambda}_i\right],$$

where $G_{it}(\cdot|\boldsymbol{\omega}_{it})$ is the true cumulative distribution function of ε_{it} conditioned on $\boldsymbol{\omega}_{it}$. Noting that $\overline{V}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_{t,0})=0$, the expansion of $\overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{f}}_t)$ at $(\boldsymbol{\gamma}_0,\boldsymbol{f}_{t,0})$ leads

$$\begin{aligned} &\overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{f}}_t) \\ =\ & -\frac{1}{N}\sum_{i=1}^{N}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}'_{i,0}](\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0})-\frac{1}{N}\sum_{i=1}^{N}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}'_{it,0}](\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}) \\ & +o_p\left(\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_{t,0}\|\right)+o_p\left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\|\right). \end{aligned} \tag{27}$$

Viewing the factor loadings $\boldsymbol{\lambda}_i$ as regressors and $\boldsymbol{f}_t$ as regression coefficient, the similar argument that derived the equation (26) leads

$$\begin{aligned}
&\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0} + o_p(\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|) \\
= \;& \Psi_t^{-1} V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) - \Psi_t^{-1}\left\{V_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - \overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right\} \\
& + \Psi_t^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}'_{it,0}](\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})\right) + o_p\left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|\right) + O_p\left(\frac{1}{N}\right), \quad (28)
\end{aligned}$$

where $\Psi_t = N^{-1}\sum_{i=1}^{N} E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}'_{i,0}]$.

Putting (28) into (26), we have

$$\begin{aligned}
&(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) \\
= \;& \Gamma_i^{-1} Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) - \Gamma_i^{-1}\left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right\} \\
& - \frac{1}{T}\sum_{t=1}^{T}\Gamma_i^{-1} E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}]\Psi_t^{-1}\left(\frac{1}{N}\sum_{j=1}^{N} E[g_{jt}(0|\boldsymbol{\omega}_{jt})\boldsymbol{\lambda}_{j,0}\boldsymbol{z}'_{jt,0}](\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0})\right) \\
& + \frac{1}{T}\sum_{t=1}^{T}\Gamma_i^{-1} E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}]\Psi_t^{-1} V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) \\
& + \frac{1}{T}\sum_{t=1}^{T}\Gamma_i^{-1} E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}]\Psi_t^{-1}\left\{V_{NT,t}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{f}}_t) - \overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{f}}_t) - V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right\} \\
& + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + o_p(\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|), \quad (29)
\end{aligned}$$

From now, we evaluate each of the terms in (29). First, the first term is

$$\max_i |Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)| = O_p\left(\frac{\log(N)}{T^{1/2}}\right). \quad (30)$$

To have (30), it suffices to show that, for any $u > 0$,

$$\max_{1\le i\le N} P\left(\left|Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right| > \frac{\log(N)}{T^{1/2}}u\right) = o\left(\frac{1}{N}\right),$$

which can be obtained by applying Lemma 5 with $a = 2\log(N)T^{1/2}$.

The second term in (29) is bounded by $o_p(T^{-1/2})$. To evaluate the second term $\Gamma_i^{-1}\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\}$ in (29), we apply Lemma 6. Define $h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) = I(u \le \boldsymbol{x}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0} - \boldsymbol{f}'_t\boldsymbol{\lambda}_i) - I(u \le 0)$. From the result of Theorem 1, we define $\mathcal{H} = \{h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) | \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| < \kappa,\ \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| < \kappa\}$ with $\kappa \to 0$. It is obvious that $E[h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]] = 0$ and that each element in $\mathcal{H}$ is bounded by 2. Also, $E[\{h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}^2] \le C \times \kappa^2$ for $h \in \mathcal{H}$. Put

$Z_i = \| \sum_{t=1}^T \{h(\xi_{it}|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\xi_{it}|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\|_{\mathcal{H}}$ with $\xi_{it} = \varepsilon_{it}$ in Lemma 6. By Lemma 6, for all $s > 0$ with probability at least $1 - \exp(-s^2)$, we have

$$\frac{1}{T} Z_i \leq \frac{1}{T} E[Z_i] + \frac{s\sqrt{2}}{T} \sqrt{T\kappa^2 + 4E[Z_i]} + \frac{2s^2}{3T}. \tag{31}$$

Because of the independence property of the idiosyncratic errors over i and t, we see that $T^{-1}E[Z_i] = o(T^{-1/2})$. We now take $s = \sqrt{2 \log N}$ in (31). Then, it is seen that there exist a positive integer T_0 independent of i such that the right side on (31) is smaller than $o(T^{-1/2})$ for $T_0 < T$. This implies that $P(T^{-1}Z_i > T^{-1/2}) \leq N^{-2}$. Therefore, the second term in (29) is bounded by $o_p(T^{-1/2})$.

Next, we show that the fourth term in (29) is $o_p(T^{-1/2})$. The fourth term of (29) satisfies

$$\sqrt{T} \left[\frac{1}{T} \sum_{t=1}^T \Gamma_i^{-1} J_{it} \Psi_t^{-1} V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) \right] = \frac{1}{N\sqrt{T}} \sum_{i=1}^N \sum_{t=1}^T \Gamma_i^{-1} J_{it} \Psi_t^{-1} \left(\tau - I(\varepsilon_{it} \leq 0)\right) \boldsymbol{\lambda}_i = o_p(1), \tag{32}$$

where $J_{it} = E[g_{it}(0|\boldsymbol{\omega}_{it}) \boldsymbol{z}_{it,0} \boldsymbol{\lambda}_{i,0}']$. Because $\|\Gamma_i^{-1} J_{it} \Psi_t^{-1}\| < \infty$, it is enough to show that

$$\frac{1}{N\sqrt{T}} \sum_{i=1}^N \sum_{t=1}^T \left(\tau - I(\varepsilon_{it} \leq 0)\right) \boldsymbol{\lambda}_i = o_p(1). \tag{33}$$

The expected value of its second moment is

$$\begin{aligned} & \frac{1}{N^2 T} E \left[\sum_{i,j=1}^N \sum_{t,s=1}^T \left(\tau - I(\varepsilon_{it} \leq 0)\right) \boldsymbol{\lambda}_i \boldsymbol{\lambda}_j' \left(\tau - I(\varepsilon_{js} \leq 0)\right) \right] \\ = & \frac{1}{N^2 T} E \left[\sum_{i=1}^N \sum_{t,s=1}^T \left(\tau - I(\varepsilon_{it} \leq 0)\right) \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i' \left(\tau - I(\varepsilon_{is} \leq 0)\right) \right] \\ = & \frac{1}{N^2 T} E \left[\sum_{i=1}^N \sum_{t=1}^T \left(\tau - I(\varepsilon_{it} \leq 0)\right)^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i' \right], \end{aligned}$$

which converges to zero. Here the first and second equality used the fact that the idiosyncratic errors are independent over i and t. Therefore, we obtain the claim (32).

We next show the fifth term in (29) is $O_p(T^{-1/4})$. This rate will be improved later on. Because of the consistency of $\{\hat{\boldsymbol{\gamma}}, \hat{F}\}$ and the uniform boundedness of $\|\Gamma_i^{-1} E[g_{it}(0|\boldsymbol{\omega}_{it}) \boldsymbol{z}_{it,0}\|$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\lambda}_{i,0}$, it suffices to show that

$$P\left\{ \sup_{\boldsymbol{\gamma} \in B_\gamma(\kappa), F \in B_F(\kappa)} \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \{h_{it}(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h_{it}(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\} \right\| \geq \frac{1}{T^{1/4}} \right\} \tag{34}$$

converges to zero. Here $\zeta_{it} = \varepsilon_{it}$, $h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) = I(u \leq \boldsymbol{x}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_{t,0}' \boldsymbol{\lambda}_{i,0} - \boldsymbol{f}_t' \boldsymbol{\lambda}_i) - I(u \leq 0)$, $B_\gamma(\kappa) \equiv \{\gamma;\ \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq \kappa,\ i = 1, ..., N\}$, and $B_F(\kappa) \equiv \{F;\ \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \kappa,\ t = 1, ..., T\}$ with $\kappa \to 0$.

For some positive constant C_l, put $W_i \equiv C_l \times (\|\boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_{i,0}\|)$, $\alpha \equiv \sup_i E[W_i]$ and $\varepsilon = \frac{1}{T^{1/4}}$. There exist C_{γ_i} balls with centers $\bar{\boldsymbol{\gamma}}_{i_k}$, $k = 1, ..., C_{\gamma_i}$ and radius $\varepsilon/(8\alpha)$ such that the collection of these balls covers $B_{\gamma_i}(\kappa) \equiv \{\boldsymbol{\gamma}_i : \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq \kappa\}$. Similarly, because $B_{f_t}(e)$ is compact subset, there exist C_{f_t} balls with centers $\bar{\boldsymbol{f}}_{t,j}$ $j = 1, ..., C_{f_j}$ and radius $\varepsilon/(8\alpha)$ such that the collection of these balls covers $B_{f_t}(e)$. We note that C_{γ_i} and C_{f_t} can be chosen such that $C_{\gamma_i}(\varepsilon) = O(1/\varepsilon^{p+r})$ and $C_{f_t}(\varepsilon) = O(1/\varepsilon^r)$ with $\varepsilon \to 0$.

Because

$$\begin{aligned}
&\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}\right| \\
\leq\ &\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\}\right| + \left|\frac{\varepsilon}{8\alpha}\cdot\frac{1}{N}\sum_{i=1}^{N}\{W_i + E[W_i]\}\right|,
\end{aligned}$$

we have

$$\begin{aligned}
&P\left\{\sup_{\boldsymbol{\gamma}\in B_\Gamma(\delta),\ F\in B_F(e)}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}\right| \geq \frac{1}{T^{1/4}}\right\} \\
\leq\ &\sum_{k_1=1}^{C_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{C_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{C_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{C_{f,T}(\varepsilon)} P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\}\right| \geq \frac{1}{2T^{1/4}}\right\} \\
&+P\left\{\left|\frac{\varepsilon}{8\alpha}\frac{1}{N}\sum_{i=1}^{N}\{W_i + E[W_i]\}\right| \geq \frac{1}{2T^{1/4}}\right\},
\end{aligned}$$

where the second term is zero because $|\frac{1}{N}\sum_{i=1}^{N}\{W_i + E[W_i]\}| < 2\alpha$ and $\varepsilon = \frac{1}{T^{1/4}}$ from its definition.

From the independence property of the idiosyncratic errors over t and the consistency of the estimated parameters,

$$\begin{aligned}
&E\left[\left\{\frac{1}{T}\sum_{t=1}^{T}h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - \frac{1}{T}\sum_{t=1}^{T}E[h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\right\}^2\right] \\
=\ &\frac{1}{T^2}\sum_{t=1}^{T}E\left[\{h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}^2\right] \\
=\ &o(T^{-1})
\end{aligned}$$

Therefore,

$$\sum_{i=1}^{N}E\left\{T^{-1}\sum_{t=1}^{T}h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T}E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\right\}^2 \leq C \times N \times o(T^{-1})$$

where C is some positive constant. Take

$$Z_i = T^{-1}\sum_{t=1}^{T}h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T}E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]$$

in Lemma 5, we then have

$$
\begin{aligned}
& P\left\{ \frac{1}{NT} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\} \right| \geq \frac{1}{2T^{1/4}} \right\} \\
= \; & P\left\{ \left| \sum_{i=1}^{N} \left\{ \frac{1}{T} \sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - \frac{1}{T} \sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})] \right\} \right| \geq \frac{N}{2T^{1/4}} \right\} \\
\leq \; & \exp\left(-C \times \frac{N^2}{T^{1/2} \times N \times o(T^{-1})} \right)
\end{aligned}
$$

which leads

$$
\begin{aligned}
& \sum_{k_1=1}^{C_{\gamma_1}(\varepsilon)} \cdots \sum_{k_N=1}^{C_{\gamma_N}(\varepsilon)} \sum_{j_1=1}^{C_{f,1}(\varepsilon)} \cdots \sum_{j_T=1}^{C_{f,T}(\varepsilon)} P\left\{ \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\} \right| \geq \frac{1}{2T^{1/4}} \right\} \\
\leq \; & C_{\gamma_1}(\varepsilon) \times \cdots \times C_{\gamma_N}(\varepsilon) \times C_{f_1}(\varepsilon) \times \cdots \times C_{f_T}(\varepsilon) \times \exp\left(-C \times NT^{1/2} \right) \\
= \; & O\left(\varepsilon^{-N(p+r)} \times \varepsilon^{-Tr} \times \exp\left[-NT^{1/2} \right] \right) \\
= \; & O\left(\exp\left[-NT^{1/2} \left\{ 1 - \frac{N(p+r)\log(\varepsilon)}{NT^{1/2}} - \frac{Tr\log(\varepsilon)}{NT^{1/2}} \right\} \right] \right) \\
= \; & O\left(\exp\left[-NT^{1/2} \left\{ 1 - \frac{(p+r)\log(\varepsilon)}{T^{1/2}} - \frac{T^{1/2}}{N^{1-\gamma}} \times \frac{r\log(\varepsilon)}{N^{\gamma}} \right\} \right] \right) \\
= \; & o(1),
\end{aligned}
$$

where the last line is obtained by using $(p+r)\log(\varepsilon)/T^{1/2} \to 0$, $T^{1/2}/N^{1-\gamma} \to 0$ and $r\log(\varepsilon)/N^{\gamma} \to 0$. This is because $\varepsilon = T^{-1/4}$ and $T^{1/2}/N^{1-\gamma} \to 0$ $(1/16 < \gamma < 1/2)$. Therefore, the fifth term in (29) is $O_p(T^{-1/4})$.

Putting these results into (29), we have

$$
\begin{aligned}
& (\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) \\
= \; & -\frac{1}{T} \sum_{t=1}^{T} \Gamma_i^{-1} E[g_{it}(0|\boldsymbol{\omega}_{it}) \boldsymbol{z}_{it,0} \boldsymbol{\lambda}_{i,0}'] \Psi_t^{-1} \left(\frac{1}{N} \sum_{j=1}^{N} E[g_{jt}(0|\boldsymbol{\omega}_{jt}) \boldsymbol{\lambda}_{j,0} \boldsymbol{z}_{jt,0}'] (\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0}) \right) + O_p\left(\frac{1}{T^{1/4}} \right),
\end{aligned}
$$

which leads the following expression

$$
(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \frac{1}{N} K(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + O_p(1/T^{1/4}),
$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1, ..., \hat{\boldsymbol{\gamma}}_N')'$, $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{1,0}, ..., \boldsymbol{\gamma}_{N,0}')'$, and

$$
K = \begin{pmatrix} K_{11} & K_{12} & \cdots & K_{1N} \\ K_{21} & K_{22} & \cdots & K_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ K_{N1} & K_{N2} & \cdots & K_{NN} \end{pmatrix}, \tag{35}
$$

where $K_{ij} = T^{-1}\sum_{t=1}^{T}\Gamma_i^{-1}J_{it}\Psi_t^{-1}J_{jt}'$, $J_{it} = E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}']$. Because $(I - \frac{1}{N}K)$ is positive definite matrix, we obtain $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(1/T^{1/4})$. By the same argument, we also have $\max_{1\le t\le T}\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = O_p(1/N^{1/4})$.

From $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(1/T^{1/4})$, we can further improve the convergence rate of the fifth term in (29). We next show that

$$P\left\{\sup_{\boldsymbol{\gamma}\in B_\gamma(T^{-1/4}),F\in B_F(N^{-1/4})}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right\| \ge \frac{1}{T^{3/8}}\right\} \quad (36)$$

converges to zero. Here $\zeta_{it} = \varepsilon_{it}$, $\boldsymbol{\gamma}_i = (\boldsymbol{b}_i', \boldsymbol{\lambda}_i')'$, $h(u|\boldsymbol{x},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) = I(u \le \boldsymbol{x}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0} - \boldsymbol{f}_t'\boldsymbol{\lambda}_i) - I(u \le 0)$, $B_\gamma(T^{-1/4}) \equiv \{\gamma;\ \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \le T^{-1/4},\ i = 1,...,N\}$, and $B_F(N^{-1/4}) \equiv \{F;\ \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \le N^{-1/4},\ t = 1,...,T\}$.

For some positive constant C_l, put $M_i \equiv C_l \times (\|\boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_{i,0}\|)$, $\alpha \equiv \sup_i M_i$ and $\varepsilon = T^{-3/8}$. There exist a positive constant C and $D_{\gamma_i} = [C \times (T^{2/3})^{p+r}]$ balls with centers $\bar{\boldsymbol{\gamma}}_{i_k}$, $k = 1,...,D_{\gamma_i}$ and radius $T^{-3/8}/(3\alpha)$ such that the collection of these balls covers $B_{\gamma_i}(\kappa) \equiv \{\boldsymbol{\gamma}_i : \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \le T^{-1/4}\}$. Here $[a]$ denotes the maximum integer that does not exceed a. Similarly, because $B_{f_t}(N^{-1/4})$ is compact subset, there exist $D_{f_t} = [C \times (T^{3/8}/N^{1/4})^r]$ balls with centers $\bar{\boldsymbol{f}}_{t,j}$ $(j = 1,...,D_{f_t})$ and radius $T^{-3/8}/(8\alpha)$ such that the collection of these balls covers $B_{f_t}(N^{-1/4})$.

Because

$$\begin{aligned}
&\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right| \\
\le\ &\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| + \left|\frac{T^{-3/8}}{8\alpha}\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{M_i + E[M_i]\}\right|,
\end{aligned}$$

we have

$$\begin{aligned}
&P\left\{\sup_{\boldsymbol{\gamma}\in B_\Gamma(\delta),\ F\in B_F(e)}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right| \ge \frac{1}{T^{3/8}}\right\} \\
\le\ &\sum_{k_1=1}^{D_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{D_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{D_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{D_{f,T}(\varepsilon)}P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \ge \frac{1}{2T^{3/8}}\right\} \\
&+P\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\{M_i + E[M_i]\}\right| \ge 4\alpha\right\},
\end{aligned}$$

where the second term is zero because $\frac{1}{N}\sum_{i=1}^{N}\{M_i + E[M_i]\} < 2\alpha$.

From the independence property of the idiosyncratic errors over t and the consistency of the estimated parameters,

$$E\left[\left\{\frac{1}{T}\sum_{t=1}^{T}h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - \frac{1}{T}\sum_{t=1}^{T}E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\right\}^2\right]$$

$$
\begin{aligned}
&= \frac{1}{T^2}\sum_{t=1}^{T} E\left[\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}^2\right] \\
&= O\left(\frac{1}{T}\times\frac{T^{1/2}+N^{1/2}}{N^{1/2}T^{1/2}}\right),
\end{aligned}
$$

where we used the result of Lemma 4 such that $E[\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}^2] = (T^{1/2}+N^{1/2})/(N^{1/2}T^{1/2})$. Therefore,

$$
\sum_{i=1}^{N} E\left\{T^{-1}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}^2 \leq C\times\left(\frac{N^{1/2}(T^{1/2}+N^{1/2})}{T^{3/2}}\right),
$$

where C is some positive constant. Take $Z_i = T^{-1}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]$ in Lemma 5, we then have

$$
\begin{aligned}
& P\left\{\frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \geq \frac{1}{2T^{3/8}}\right\} \\
= \; & P\left\{\left|\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - \frac{1}{T}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}\right| \geq \frac{N}{2T^{3/8}}\right\} \\
\leq \; & \exp\left(-C\times\frac{N^2}{T^{3/4}}\times\frac{T^{3/2}}{N^{1/2}(T^{1/2}+N^{1/2})}\right),
\end{aligned}
$$

where C is some positive constant. This leads

$$
\begin{aligned}
& \sum_{k_1=1}^{D_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{D_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{D_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{D_{f,T}(\varepsilon)} P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \geq \frac{1}{2T^{3/8}}\right\} \\
\leq \; & D_{\gamma_1}(\varepsilon)\times\cdots\times D_{\gamma_N}(\varepsilon)\times D_{f_1}(\varepsilon)\times\cdots\times D_{f_T}(\varepsilon)\times\exp\left(-C\times\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right) \\
= \; & O\left(\left(T^{2/3}\right)^{N(p+r)}\times\left(\frac{T^{3/8}}{N^{1/4}}\right)^{Tr}\times\exp\left[-\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right]\right) \\
= \; & O\Bigg(\exp\Bigg[-\left(-\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right)\Bigg\{1-\frac{N(N^{1/2}+T^{1/2})\log(T^{2/3})}{N^{3/2}T^{3/4}} \\
& \qquad -\frac{T(N^{1/2}+T^{1/2})\log(T^{3/8}/N^{1/4})}{N^{3/2}T^{3/4}}\Bigg\}\Bigg]\Bigg) \\
= \; & O\Bigg(\exp\Bigg[-\left(-\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right)\Bigg\{1-\frac{\log(T)}{T^{3/4}}-\frac{\log(T)}{N^{1/2}T^{1/4}}-\frac{T^{1/4}\log(T^{3/8}/N^{1/4})}{N} \\
& \qquad -\frac{T^{3/4}\log(T^{3/8}/N^{1/4})}{N^{3/2}}\Bigg\}\Bigg]\Bigg) \\
= \; & o(1),
\end{aligned}
$$

where the last line is obtained by using $T^{1/2}/N \to 0$. Thus, the fifth term in (29) is bounded by $O_p(T^{-3/8})$. By repeating the argument that derived $\max_{1\leq i\leq N}\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\| =$

$O_p(1/T^{1/4})$, we obtain the claim $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\| = O_p(T^{-3/8})$. By using $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\| = O_p(T^{-3/8})$, we can further strengthen the result to $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_{i,0}\| = O_p(T^{-7/16})$. Similarly, we can prove that $\max_t\|\hat{\boldsymbol{f}}_t-\boldsymbol{f}_t\| = O_p(T^{-7/16})$.

We next further improve the rate for the fifth term in (29) to be $o_p(T^{-1/2})$ (this result will also be useful when deriving the limiting distribution, as it is negligible). It suffices to show that, for any $\eta>0$,

$$P\left\{\sup_{\boldsymbol{\gamma}\in B_\gamma(T^{-7/16}),F\in B_F(N^{-7/16})}\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\{h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right\|\ge\frac{\eta}{T^{1/2}}\right\} \quad (37)$$

converges to zero. Here ζ_{it} and $h(u|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)$ are defined before, $B_\gamma(T^{-7/16})$ and $B_F(N^{-7/16})$ are defined in the proof of Lemma 4. By using the same argument, we have

$$\begin{aligned}&P\left\{\sup_{\boldsymbol{\gamma}\in B_\Gamma(T^{-7/16}),\ F\in B_F(N^{-7/16})}\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right|\ge\frac{\eta}{T^{1/2}}\right\}\\ \le\ &\sum_{k_1=1}^{E_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{E_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{E_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{E_{f,T}(\varepsilon)}P\left\{\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right|\ge\frac{\eta}{2T^{1/2}}\right\}\end{aligned}$$

and

$$E\left[\left\{\frac{1}{T}\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-\frac{1}{T}\sum_{t=1}^T E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\right\}^2\right]=O\left(\frac{1}{T}\times\frac{T^{7/8}+N^{7/8}}{N^{7/8}T^{7/8}}\right).$$

Therefore,

$$\sum_{i=1}^N E\left\{T^{-1}\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-T^{-1}\sum_{t=1}^T E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}^2\le C\times\left(\frac{N^{1/8}(T^{7/8}+N^{7/8})}{T^{15/8}}\right),$$

where C is some positive constant. Again, we take Z_i as $Z_i=T^{-1}\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-T^{-1}\sum_{t=1}^T E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]$ in Lemma 5, we then have

$$\begin{aligned}&P\left\{\frac{1}{NT}\left|\sum_{i=1}^N\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right|\ge\frac{\eta}{2T^{1/2}}\right\}\\ \le\ &\exp\left(-C_\eta\cdot\frac{N^2}{T}\cdot\frac{T^{15/8}}{N^{1/8}(T^{7/8}+N^{7/8})}\right),\end{aligned}$$

where C_η is some positive constant. This leads

$$\begin{aligned}&\sum_{k_1=1}^{E_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{E_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{E_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{E_{f,T}(\varepsilon)}P\left\{\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right|\ge\frac{1}{2T^{1/2}}\right\}\\ \le\ &D_{\gamma_1}(\varepsilon)\times\cdots\times D_{\gamma_N}(\varepsilon)\times D_{f_1}(\varepsilon)\times\cdots\times D_{f_T}(\varepsilon)\times\exp\left(-C_\eta\times\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right)\end{aligned}$$

$$
\begin{aligned}
&= O\left(\left(T^{1/16}\right)^{N(p+r)} \times \left(\frac{T^{1/2}}{N^{7/16}}\right)^{Tr} \times \exp\left[-\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right]\right) \\
&= O\Bigg(\exp\Bigg[-\left(\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right)\Bigg\{1-\frac{\{N(p+r)\}(N^{7/8}+T^{7/8})\log(T^{1/16})}{N^{15/8}T^{7/8}} \\
&\qquad -\frac{(Tr)(N^{7/8}+T^{7/8})\log(T^{1/2}/N^{7/16})}{N^{15/8}T^{7/8}}\Bigg\}\Bigg]\Bigg) \\
&= O\Bigg(\exp\Bigg[-\left(\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right)\Bigg\{1-\frac{\log(T^{1/16})}{T^{7/8}}-\frac{\log(T^{1/16})}{N^{7/8}}-\frac{T^{1/8}\log(T^{1/2}/N^{7/16})}{N} \\
&\qquad -\frac{T\log(T^{1/2}/N^{7/16})}{N^{15/8}}\Bigg\}\Bigg]\Bigg) \\
&= o(1),
\end{aligned}
$$

where the last line is obtained by using $\sqrt{T}/N^{1-\gamma} \to 0$ for small value of γ $(1/16 < \gamma)$. Thus, the fifth term in (29) is bounded by $o_p(T^{-1/2})$. This completes the first claim of Lemma 4 in (22).

Next, we put (26) into (28)

$$
\begin{aligned}
&\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0} \\
&= \Psi_t^{-1}V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) - \Psi_t^{-1}\left\{V_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - \overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right\} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\Psi_t^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}_{it,0}']\Gamma_i^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}'](\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})\right] \\
&\quad - \frac{1}{N}\sum_{i=1}^{N}\Psi_t^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}_{it,0}']\Gamma_i^{-1}Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\Psi_t^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}_{it,0}']\Gamma_i^{-1}\left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right\} \\
&\quad + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + o_p(\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|). \qquad (38)
\end{aligned}
$$

By flipping the role of $\boldsymbol{f}_t$ and $\boldsymbol{\gamma}_i$, we can apply the same argument used to obtain (22). Thus, we have $\max_{1\le t\le T}\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = O_p(\log(T)/N^{1/2})$. This completes the proof of Lemma 4.

D Proof of Theorem 2

We first study the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_i$. Together with the analysis in the proof of Lemma 4, we can obtain the following expression.

$$
\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0} = \Gamma_i^{-1}Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) - \frac{1}{T}\sum_{t=1}^{T}\Gamma_i^{-1}J_{it}\Psi_t^{-1}\left(\frac{1}{N}\sum_{j=1}^{N}J_{jt}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0})\right) + o_p(1).
$$

This part of analysis is similar to Song (2013) and Ando and Bai (2015). We have

$$\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \frac{1}{N} K \sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \sqrt{T}\boldsymbol{\eta} + o_p(1), \tag{39}$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1, ..., \hat{\boldsymbol{\gamma}}_N')'$, $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{1,0}, ..., \boldsymbol{\gamma}_{N,0}')'$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', ..., \boldsymbol{\eta}_N')'$ with $\boldsymbol{\eta}_i = \Gamma_i^{-1} Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)$, and K is defined in (35). Solving (39) in terms of $\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$, we have

$$\begin{aligned} \sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) &= \left(I - \frac{1}{N}K\right)^{-1} \left(\sqrt{T}\boldsymbol{\eta} + o_p(1)\right), \\ &= \left(I + \frac{1}{N}K + \frac{1}{N^2}K^2 + \frac{1}{N^3}K^3 + \cdots\right) \left(\sqrt{T}\boldsymbol{\eta} + o_p(1)\right). \end{aligned}$$

Note that we can ignore the higher order terms related to $\sqrt{T}N^{-\alpha}K\boldsymbol{\eta}$ as $o_p(1)$ due to the increasing order of $N^{-\alpha}$. Similar technique is also employed in Song (2013) and Ando and Bai (2015). Then, we have

$$\sqrt{T}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) = \Gamma_i^{-1} \left(T^{1/2} Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right) + o_p(1).$$

We thus see that the asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})$ is normal with mean zero and variance-covariance matrix Σ_i.

By using the similar argument that employed to derive the asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})$, we then have

$$N^{1/2}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}) = \Psi_t^{-1} \left(N^{1/2} V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right) + o_p(1),$$

which implies that the asymptotic distribution of $N^{1/2}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})$ is normal with mean zero and variance-covariance matrix Θ_t. This completes the proof of Theorem 2.

E Proof of Theorem 3

We prove Theorem 3 by investigating the following two cases. Case 1: $r_0 < r$ and Case 2: $r < r_0$.

Case 1: $r_0 < r$

First consider the case $r_0 < r$ with r_0 being the true number of common factors. Because the number of common factors used in the model, r, is different from the true number of common factors, r_0, we first define the true factor structure for the panel quantile model with the dimension of interactive effects when $r \neq r_0$. Recall that the true quantile function $Q(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0})$ with the true dimension of the interactive effects r_0 is given as

$$Q(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}) = \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} + \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0},$$

where the dependency of τ is dropped from $\boldsymbol{b}_{i,0,\tau}$, $\boldsymbol{f}_{t,0,\tau}$ and $\boldsymbol{\lambda}_{i,0,\tau}$. The true parameters $\{B_0, \Lambda_0, F_0\}$ are minimizer of the loss function;

$$\begin{aligned}
& \bar{\ell}_{NT}(B, \Lambda, F) \\
\equiv\ & E\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)\right] \\
=\ & E\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\boldsymbol{x}_{it}'(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+(\boldsymbol{f}_t'\boldsymbol{\lambda}_i-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})}\left\{G_{it}(s|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}) - \tau\right\}ds\right],
\end{aligned}$$

which is zero at $\{B, \Lambda, F\} = \{B_0, \Lambda_0, F_0\}$. The expectation is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$ conditional on X, F_0, B_0 and Λ_0.

Similar to Theorem 1 of Bai and Ng (2002), let H_r be an $r_0 \times r$ matrix with $\text{rank}(H_r) = \min\{r, r_0\}$ and H_r^+ be the generalized inverse of H_r such that $H_r H_r^+ = I_{r_0}$. Then, for $r_0 < r$, the interactive effects in the true quantile function can be re-expressed as

$$Q(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}) = \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} + (H_r'\boldsymbol{f}_{t,0})'(H_r^+\boldsymbol{\lambda}_{i,0}).$$

It is clear that these transformed true parameters $F_0(r) = (\boldsymbol{f}_{1,0}(r), ..., \boldsymbol{f}_{T,0}(r))'$ and $\Lambda_0(r) = (\boldsymbol{\lambda}_{1,0}(r), ..., \boldsymbol{\lambda}_{N,0}(r))'$ with $\boldsymbol{f}_{t,0}(r) \equiv H_r\boldsymbol{f}_{t,0}$ and $\boldsymbol{\lambda}_{i,0}(r) \equiv H_r^+\boldsymbol{\lambda}_{i,0}$ together with B_0, will let the loss $\bar{\ell}_{NT}(B, \Lambda, F)$ be zero when $r > r_0$. Therefore, we define $F_0(r)$ and $\Lambda_0(r)$ as the true factor structures when $r > r_0$.

Let $\hat{\boldsymbol{\gamma}}_i(r) = (\hat{\boldsymbol{b}}_i(r), \hat{\boldsymbol{\lambda}}_i(r))$ and $\hat{\boldsymbol{f}}_t(r)$ be the estimated model parameters under the number of common factors being r. Similar to the proofs of Theorem 1 and Theorem 2, we obtain $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i(r) - \boldsymbol{\gamma}_{i,0}(r)\| = O_p(\log(N)/\sqrt{T})$ and $\max_{1\le t\le T}\|\hat{\boldsymbol{f}}_t(r) - \boldsymbol{f}_{t,0}(r)\| = O_p(\log(T)/\sqrt{N})$. Here $\boldsymbol{\gamma}_{i,0}(r) = (\boldsymbol{b}_{i,0}', \boldsymbol{\lambda}_{i,0}(r)')'$. Note that $\boldsymbol{b}_{i,0}$ is the true parameter under the true model with the number of factors r_0.

Using Knight's identity $\rho_\tau(u - \nu) - \rho_\tau(u) = -\nu\psi_\tau(u) + \int_0^\nu (I(u \le s) - I(u \le 0))ds$ with $\psi_\tau(u) = \tau - I(u \le 0)$, we express

$$\begin{aligned}
& V(r) \\
\equiv\ & \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_i(r) - \hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)\right) \\
=\ & \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\Big(\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\} - \{\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r) - \boldsymbol{b}_{i,0}) + \hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r) - \boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\}\Big) \\
=\ & \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(\varepsilon_{it}\right) + \sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r) - \boldsymbol{b}_{i,0}) + \hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r) - \boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\right)\psi(\varepsilon_{\tau,i}) \\
& + \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+\hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\right)}(I(\varepsilon_{\tau,i} \le s) - I(\varepsilon_{\tau,i} \le 0))ds
\end{aligned}$$

$$= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(\varepsilon_{it}\right) + I_1(r) + I_2(r).$$

First, we investigate $I_2(r)$. For a notational simplicity, we denote $\min\{N,T\}$ as C_{NT}.

$$\begin{aligned}
&(NT)\times I_2(r)\\
=\;& \sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)}(I(\varepsilon_{\tau,i}\le s)-I(\varepsilon_{\tau,i}\le 0))ds\\
=\;& \frac{(\log N)^{1/2}}{C_{NT}^{1/2}}\sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\sqrt{C_{NT}/\log(N)}}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)\\
&\left(G_{it}\left(\varepsilon_{it}+\frac{s}{\sqrt{C_{NT}/\log N}}\middle|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0}\right)-G_{it}(\varepsilon_{it}\middle|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})\right)ds\\
=\;& \frac{\log N}{C_{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\int_0^{\sqrt{C_{NT}/\log(N)}}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)\right.\\
&\left. g_{it}(0|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})sds+o_p(1)\right]\\
=\;& \frac{\log N}{C_{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\left(\frac{C_{NT}}{2\log N}\right)g_{it}(0|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)\right.\right.\\
&\left.\left.+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)^2+o_p(1)\right]\\
=\;& C\times\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})\right)^2+C\times\sum_{t=1}^{T}\sum_{i=1}^{N}\left((\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)\right)^2\\
&+C\times\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)^2\\
=\;& O_p\left(N\right)+O_p\left(T\right),
\end{aligned}$$

where we used $N^{-1}\sum_{i=1}^{N}\|\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})\|^2=O_p(T^{-1})$, $N^{-1}\sum_{i=1}^{N}\|\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\|^2=O_p(T^{-1})$, $\sum_{i=1}^{T}\|(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))\|^2/T=O_p(N^{-1})$, and $0<g_{it}(0|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})<\infty$ for $i=1,...,N$, $t=1,...,T$.

Next, we evaluate the term $I_1(r)$. Noting that $\sum_{i=1}^{N}\|\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})\|^2/N=O_p(1/T)$, $\sum_{i=1}^{N}\|\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\|^2/N=O_p(1/T)$ and $\sum_{t=1}^{T}\|(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))\|^2/T=O_p(1/N)$, we have

$$(NT)\times I_1(r)\le C\times O_p\left(C_{NT}^{-1/2}\right)\times\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\psi(\varepsilon_{\tau,i})\right)=O_p\left(\sqrt{NT}/C_{NT}^{1/2}\right).$$

Thus, we obtain

$$V(r)-V(r_0)=O_p\left(C_{NT}^{-1}\right).$$

Using the same argument of the proof of Corollary 1 in Bai and Ng (2002), for $r > r_0$, this implies that $V(r)/V(r_0) = 1 + O_p(1/C_{NT})$. Thus, $\log(V(r)/V(r_0)) = O_p(C_{NT}^{-1})$. Because $(r - r_0)q(N,T) \geq q(N,T)$, which converges to zero slower rate than $O_p(1/C_{NT})$, it follows that

$$P\left(IC(r) - IC(r_0) < 0\right) \leq P\left(O_p\left(1/C_{NT}\right) + q(N,T) < 0\right) \to 0.$$

This indicates that the probability that $IC(r)$ selects the number of common factors $r > r_0$ is asymptotically 0.

Case 2: $r < r_0$

Because $r < r_0$, an $T \times r$ common factor $F(r)$ can not span the true space spanned by the true common factor F_0 with dimension $T \times r_0$. Therefore, regardless of the values of model parameters $\{B(r), \Lambda(r), F(r)\}$ with r dimensional interactive effects, the following loss function can never be zero:

$$\begin{aligned}
& \bar{\ell}_{NT}(B(r), \Lambda(r), F(r)) \\
\equiv\ & \frac{1}{NT} E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i(r) - \boldsymbol{f}_t(r)'\boldsymbol{\lambda}_i(r)\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)\right] \\
=\ & \frac{1}{NT} E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} \int_0^{\boldsymbol{x}_{it}'(\boldsymbol{b}_i(r)-\boldsymbol{b}_{i,0})+(\boldsymbol{f}_t(r)'\boldsymbol{\lambda}_i(r)-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})} \left\{G_{it}(s|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}) - \tau\right\} ds\right].
\end{aligned}$$

From the investigation of (13), for some positive constant $C > 0$, not depending on N and T,

$$\lim_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_i(r) - \hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_i(r_0) - \hat{\boldsymbol{f}}_t(r_0)'\hat{\boldsymbol{\lambda}}_i(r_0)\right)\right] > C,$$

where $\hat{\boldsymbol{b}}_i(r)$, $\hat{\boldsymbol{f}}_t(r)$ and $\hat{\boldsymbol{\lambda}}_i(r)$ are parameter estimates under the dimension of interactive effects r. Using the same argument of the proof of Corollary 1 in Bai and Ng (2002), we therefore have $V(r)/V(r_0) > 1 + c_0$ for some c_0 with large probability for all large N and T. This implies that $\log(V(r)/V(r_0)) \geq c_0'$ for some constant $0 < c_0'$, for large N and T. Because $q(N,T) \to 0$, we have $IC(r) - IC(r_0) > c_0 - (r_0 - r)q(N,T) \geq c_0''$ for some constant $0 < c_0''$, under large N and T with large probability. Thus

$$P\left(IC(r) - IC(r_0) < 0\right) \to 0.$$

This completes the proof of Theorem 3.

F Additional information on the empirical analysis

In Section 6, we analyzed the stock returns of publicly traded firms and firms traded in over-the-counter trading markets for over 6,000 international stocks from over 100 financial markets. The final samples for each period are summarized in Table 1.

To explore the effects of stock exchanges and industries on individual stock returns, Section 6.2.3 applied a clustering approach to the estimated regression coefficients and factor loadings $\{(\hat{\boldsymbol{b}}_{i,\tau}', \hat{\boldsymbol{\lambda}}_{i,\tau}');\ i = 1, ..., N\}$ to create a set of groups based on the similarities in the sensitivity to the common factors. If the source of the sensitivity to the factors (both observables and unobservables) is solely attributable to stock exchanges, it is expected that the two-way table of the assigned group membership from the clustering approach against the stock exchanges will be diagonal. Note that the industry classifications and listed stock exchanges are known. Therefore, it is easy to create a two-way table of the assigned group membership against these classifications.

First, we investigate the effect of stock exchanges. There are 36 stock exchanges where more than 40 stocks are listed, and we consider these 36 stock exchange markets. Setting the number of clusters as 36, the clustering approach is applied to $\{(\hat{\boldsymbol{b}}_{i,0.05}', \hat{\boldsymbol{\lambda}}_{i,0.05}');\ i = 1, ..., N\}$, the estimated regression coefficients and factor loadings for the lower tail. The left column in Figure 1 shows the distribution of the firms. An (i, j)-th element denotes the percentage of firms listed on stock exchange i such that they belong to the j-th group. Thus, each row represents the distribution of the firms listed on the same stock exchange. We can make the following observations. First, the degrees of similarity between Shanghai and Shenzhen are stable over the periods in both the upper and lower tails. Second, the New York Stock Exchange and NASDAQ tend to be very similar, while Non-NASDAQ OTC represents its dissimilarity to these two markets. The same observations can be seen from the right column of Figure 1, obtained from the upper tail's factor loadings $\{(\hat{\boldsymbol{b}}_{i,0.95}', \hat{\boldsymbol{\lambda}}_{i,0.95}');\ i = 1, ..., N\}$. Thus, investors should consider such market characteristics although all three markets are located in the U.S.

A similar approach is applied to determine the effect of industry. To see the effect of industry (5 industries), we also set the number of clusters at 5. Figure 2 shows the distribution of the firms in each industry obtained from $\{(\hat{\boldsymbol{b}}_{i,0.05}', \hat{\boldsymbol{\lambda}}_{i,0.05}');\ i = 1, ..., N\}$, the estimated regression coefficients and factor loadings for the lower tail. The i-th row represents the distribution of firms in industry i. Specifically, let $n_{i,j}$ denote the number of firms that belong to industry i and to group j. Then, the (i, j)-th element d_{ij} is calculated as $d_{ij} = n_{i,j}/\{\sum_{k=1}^{5} n_{i,k}\}$. Overall, there is one huge cluster that includes most of the firms from each of the 5 industries. Thus, investors regard these 5 industries as similar rather

than treating them as very different groups. The same observations can be seen from Figure 2, obtained from the upper tail's $\{(\hat{\boldsymbol{b}}'_{i,0.95}, \hat{\boldsymbol{\lambda}}'_{i,0.95});\ i = 1, ..., N\}$. There seems to be other sources of variability in stock returns in addition to the industry effects.

In summary, the firm industry and the stock exchange on which a firm is listed are important factors to be considered. However, we also note that these nominal factors are insufficient to fully capture the underlying market structures.

G Simulation study

G.1 Performance of the frequentist estimator

To demonstrate the usefulness of the proposed estimation procedure, we conduct a Monte Carlo simulation study. Because the data-generating process and the model parameters are known, we can evaluate the performance of our approach. Here, we report the results for the challenging case in which the variables $x_{it,\tau}$ are correlated with the unobservable factor structures $\boldsymbol{f}'_t\boldsymbol{\lambda}_{i,\tau}$.

G.1.1 Data generating process

For the first data-generating process, we first generate the uniform independent random variable $u_{it} \sim U[0,1]$ for $i = 1, ..., N$ and $t = 1, ..., T$. Then, we generate the data from the following structure:

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{b}_{i,u_{it}} + \boldsymbol{f}'_{t,u_{it}}\boldsymbol{\lambda}_{i,u_{it}} + \varepsilon_{it,u_{it}}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T,$$

where $\boldsymbol{x}_{it} = (x_{it,1}, ..., x_{it,p})'$ is a vector of regressors, the dimension of the common factor and the corresponding factor loading depends on the quantile u_{it}, and $\varepsilon_{it,u_{it}} = G^{-1}(u_{it})$. Here, $G(\cdot)$ is a cumulative distribution function of normal or Student-t distribution. The true quantile function of y_{it} at quantile point τ is

$$Q_{y_{it}}\left(\tau|\boldsymbol{x}_{it}\boldsymbol{f}_t(\tau), \boldsymbol{\lambda}_i(\tau)\right) = G^{-1}(\tau) + \boldsymbol{x}'_{it}\boldsymbol{b}_i(\tau) + \boldsymbol{f}_t(\tau)'\boldsymbol{\lambda}_i(\tau).$$

The quantile restriction $P(y_{it} - Q_{y_{it}}(\tau|\boldsymbol{\omega}_{it}) \leq 0) = \tau$ is satisfied.

We generate $T \times 5$ common factor matrix $F = (f_{tk})$ such that each element follows the uniform distribution over $[0, 2]$. Using the generated u_{it}, we define the common factor for the i-th unit at time t as

$$\boldsymbol{f}_{t,u_{it}} = \begin{cases} (f_{t1}, f_{t2}, f_{t3})' & \text{if } u_{it} \leq 0.2 \\ (f_{t1}, f_{t2}, f_{t3}, f_{t4})' & \text{if } 0.2 < u_{it} \leq 0.8 \\ (f_{t1}, f_{t2}, f_{t3}, f_{t4}, f_{t5})' & \text{if } 0.8 < u_{it} \end{cases}.$$

Note that the dimension of the common factor may vary over i and t because $u_{it} \sim U[0,1]$ for $i = 1, ..., N$ and $t = 1, ..., T$. Given u_{it}, the corresponding factor-loading vector is specified as

$$\boldsymbol{\lambda}_{i,u_{it}} = \begin{cases} (\zeta_{i1} + 0.1u_{it}, \zeta_{i2} + 0.1u_{it}, \zeta_{i3} + 0.1u_{it})' & \text{if } u_{it} \leq 0.2 \\ (\zeta_{i1} + 0.1u_{it}, \zeta_{i2} + 0.1u_{it}, \zeta_{i3} + 0.1u_{it}, \zeta_{i4} + 0.1u_{it})' & \text{if } 0.2 < u_{it} \leq 0.8 \\ (\zeta_{i1} + 0.1u_{it}, \zeta_{i2} + 0.1u_{it}, \zeta_{i3} + 0.1u_{it}, \zeta_{i4} + 0.1u_{it}, \zeta_{i5} + 0.1u_{it})' & \text{if } 0.8 < u_{it} \end{cases},$$

where ζ_{ik} is generated from the uniform distribution over $[0, 2]$ and then fixed over t. However, the quantile random variable u_{it} adds some variations over t because the factor loading depends on the quantile points.

Setting $p = 8$ for $i = 1, ..., N$, we generate the set of regressors as follows:

$$x_{it,1} = v_{it,1} + 0.02f_{t1}^2 + 0.02\zeta_{i1}^2, \quad x_{it,3} = v_{it,3} - 0.01f_{t2}^2 + 0.02\zeta_{i2}^2,$$
$$x_{it,5} = v_{it,5} - 0.01f_{t3}^2 + 0.03\zeta_{i3}^2, \quad x_{it,k} = v_{it,k} \quad (k \neq 1, 3, 5).$$

where $v_{it,k}$ is generated from the uniform distribution over $[0, 2]$. The k-th element of true parameter values of regression coefficients $\boldsymbol{b}_{i,0,u_{it}}$ are set to be

$$b_{ik,0,u_{it}} = \begin{cases} 1 + i/N + 0.1u_{it} & \text{if } k = 2, 4, 6 \\ -1 + i/N + 0.1u_{it} & \text{if } k \neq 2, 4, 6 \end{cases},$$

Similar to the factor loadings, the quantile random variable u_{it} adds some variations. Finally, cumulative distribution function of $\varepsilon_{it,u_{it}}$ is the normal distribution $N(0, 1)$.

The second data-generating process modifies the first data-generating process. We let the cumulative distribution function of $\varepsilon_{it,u_{it}}$ as the Student-t distribution with degrees of freedom 8. Thus, the error terms have a fat-tail property.

G.1.2 Results

We simulate a large panel with N individuals and T time periods. We consider various combinations of T and N. We base our estimate on the true number of factors and assess the robustness of the proposed strategy to endogeneity. The dimension of the interactive effects is set as its true dimension. For example, the dimension of the interactive effects is set as $r = 3$ when we estimate the $\tau = 5\%$ quantile structure,

The estimation results are averaged over 100 simulated data sets and reported in Table 2 $\sim$ Table 3. Tables show the mean squared error (MSE) between the true structure and the estimates

$$\begin{aligned} \text{MSE}_1 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2, \\ \text{MSE}_2 &= \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2, \end{aligned}$$

where $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $\boldsymbol{b}_{i,0,\tau}$ and the true τ-th quantile function $Q_{it}(\tau)$, respectively. These measures are computed for the estimators both with and without the factor structure. The panel quantile model without the factor structure is estimated by minimizing the standard loss function $\ell_\tau(Y|X, B_\tau) = \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau})$. Table 2 $\sim$ Table 3 indicate that our estimator with the factor structure performs better than the standard approach. Similar results are also obtained under the second data generating process.

G.2 Performance of the model selection criterion

We investigate the performance of the proposed model selection criterion to select the dimension of the interactive effects. Using the two data generating processes in the previous section, we generate the dataset under the various combinations of N and T. We set the possible dimension of the interactive effects (i.e., the number of common factors) to range from 0 to 8. Calculating the scores of $IC(r)$ over all possible r, we can detect the number of r.

Table 4 reports the histogram of the selected number of common factors $\hat{r}$ over 200 simulation runs. As shown in the tables, the proposed criterion is capable of selecting the true number of factors. When the size of panel N and T increases, the procedure achieves better performance in terms of identifying the true dimension of the interactive effects.

G.3 Estimation under a small panel

Although we developed the asymptotic theory for the frequentist estimator, it was developed under large N and T. In this section, we compare the small sample properties of the frequentist estimator and the proposed Bayesian approach. We use the first data generating process described in Section G.1. We set total number of Markov chain Monte Carlo iterations to 3,000. If one aims to obtain the samples from the posterior distribution, the first iterations are usually discarded to ensure adequate dissipation of initial conditions, or burn-in. However, Geyer (2011) pointed out that Markov chain started anywhere near the center of the posterior distribution needs no burn-in. Because our frequentist estimator corresponds to the Bayesian maximum a posteriori estimator, our initial parameter value is already a good starting point for MCMC. Thus, burn-in period is not considered. We follow Gerlach et al. (2011) by examining trace plots from the MCMC sampler.

Figure 4 (a) shows the MCMC sampling path for the regression coefficient $b_{20,3,\tau}$ with $\tau = 0.05$, $N = T = 100$. We see that the sampling behaviors of each of MCMC sample are already stable from the beginning. Also, the generated posterior sample distributes

around the true value of the regression coefficient. Figure 4 (b) compares of the asymptotic distribution of the frequentist estimator (based on Theorem 2) and the posterior density from Bayesian approach. From Theorem 2, we know that the asymptotic distribution of the frequentist estimator is normal distribution. Dashed blue line is the constructed asymptotic distribution from Theorem 2, Solid black line is the constructed posterior density from MCMC output. We see that the constructed posterior density has the wider tail than that from the asymptotic distribution. Thus, the Bayesian posterior credible interval is wider than that of the constructed 95% percent confidence interval of the frequentist estimator. This is commonly known because Bayesian approach takes account parameter uncertainty.

Next, we compare the performance of the following estimators: the frequentist estimator and the Bayesian estimators based on posterior mean, posterior mode and posterior median. We set the length of time series and the number of units as $T = 100, 300, 900$ and $N = 100$, respectively. We note that the similar results are obtained under the different data generating processes described in Section G.1 as well as the different quantile points τ.

Figure 5 shows the boxplots of the average mean squared error between the true parameter vectors $\boldsymbol{b}_{i,0,\tau}$ and its estimates $\hat{\boldsymbol{b}}_{i,\tau}$ over i; MSE $= \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$. These results are obtained based on 200 repetitions at $\tau = 0.05$. Note that similar results are obtained under the different quantile points $\tau = 0.5$ and $\tau = 0.95$, and thus these results are omitted.

We can make the following observations. First, MSE decreases as T increases. Second, the Bayesian estimator performs well in the sense that the median of MSEs are smaller than the frequantist estimator when the length of time series T is small. Although the computational time of our proposed Bayesian estimator is slower than the frequentist estimator, it provides better MSE than the others even when the panel size is small. Third, under $T = 900$, the performance of Bayesian estimators (the posterior mean, mode, median) and frequestist estimators became very similar. Because prior is dominated by the pseudo likelihood $L(Y|X, F_\tau, \Lambda_\tau, B_\tau)$, this property can be observed in the estimation results. Thus, Bayesian estimators (the posterior mean, mode, median) and frequestist estimators are asymptotically equivalent as long as the prior information is dominated by the pseudo likelihood $L(Y|X, F_\tau, \Lambda_\tau, B_\tau)$.

References

Ando, T., and Bai, J. (2015). "Asset pricing with a general multifactor structure," *Journal*

of Financial Econometrics, **13**, 556–604.

Ando, T. and Bai, J. (2017). "Clustering huge number of time series: A panel data approach with high-dimensional predictors and factor structures," *Journal of the American Statistical Association*, **112**, 1182–1198.

Babu, G.J. (1989). "Strong representations for LAD estimators in linear models," *Probability Theory and Related Fields*, **83**, 547–558.

Bai, J. (1998). "Estimation of multiple-regime regressions with least absolutes deviation," *Journal of Statistical Planning and Inference*, **74**, 103–134.

Bai, J. (2009). "Panel data models with interactive fixed effects," *Econometrica*, **77**, 1229–1279.

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

Bousquet, O. (2002). "A bennett concentration inequality and its application to suprema of empirical processes," *Comptes Rendus Mathematique*, **334**, 495-500.

Chen, L., Dolado, J. and Gonzalo, J. (2017). "Quantile Factor Models," Working Paper.

Geyer, C.J. (2011). "Introduction to Markov Chain Monte Carlo," In *Handbook of Markov Chain Monte Carlo* (ed. Gelman, A., Brooks, S., Jones, G. and Meng, X.-L.), CRC Press, 3–48.

Gerlach, R, Chen, C.W.S. and Chan, N.Y.C. (2011). "Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets," *Journal of Business and Economic Statistics*, **29**, 481–492.

Gutenbrunner, C. and Jureckova, J. (1992). "Regression rank scores and regression quantiles," *Annals of Statistics*, **20**, 305-330.

Hoeffding, W. (1963). "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, **58**, 13-30.

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

Knight, K. (1998). "Limiting distributions for L1 regression estimators under general conditions," *Annals of Statistics*, **26**, 755-770.

Song, M. (2013). "Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications," Unpublished manuscript, Department of Economics, Columbia University.

Talagrand, M. (1996). "New concentration inequalities in product spaces," *Inventiones Mathematicae*, **126**, 505-563.

van der Vaart A. W. (1998). *Weak convergence and empirical processes.* Springer, New York.

Table 1: Distributions of the number of listed financial firms. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015).

	Period 1	Period 2	Period 3
Amman	68	68	67
Australian	99	98	98
Bangkok	86	85	85
Berlin	140	144	146
Borsa Istanbul	46	46	47
BSE Ltd	285	286	277
Copenhagen Stock Exchange	34	34	34
Dhaka	69	68	69
Euronext.liffe Paris	83	83	81
Frankfurt	531	536	529
Hong Kong	203	202	199
Indonesia	76	73	75
Johannesburg	40	39	40
Karachi	56	56	56
Korea Stock Exchange	53	52	53
Kuala Lumpur	99	99	99
Kuwait City	74	72	74
London	144	146	149
Milan	44	44	44
NASDAQ	374	375	370
National India	103	101	103
New York Stock Exchange (NYSE)	186	187	183
Non NASDAQ OTC	1370	1364	1289
OTC Bulletin Board	49	51	52
Philippine Stock Exchange	82	82	81
Santiago	42	40	41
Shanghai	73	73	73
Shenzen	55	55	54
Singapore	49	49	47
SIX Swiss	68	68	68
Stockholm	45	44	44
Stuttgart	63	61	63
Taiwan	61	61	60
Tel Aviv	107	105	106
Thailand	58	59	60
Tokyo Stock Exchange	177	175	177
Toronto	81	80	77
TSX Ventures	64	60	56
XETRA	83	83	82
Others	781	785	780

Table 2: First data-generating process. Comparison of the proposed estimator with factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$, and the standard estimator without factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$. The mean squared errors are defined as $\text{MSE}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2$, $\text{MSE}_2 = \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$, where $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $\boldsymbol{b}_{i,0,\tau}$ and the true τ-th quantile function $Q_{it}(\tau)$, respectively. Averages over 200 simulation data sets are reported. The second lines are corresponding to the standard deviation of the mean squared errors.

			With factor structure		Without factor structure	
τ	T	N	MSE_1	MSE_2	MSE_1	MSE_2
0.05	300	300	0.7140	0.2397	0.9571	0.3046
			0.0524	0.0137	0.0308	0.0123
	300	900	0.7004	0.2389	0.9558	0.3028
			0.0451	0.0088	0.0278	0.0075
	900	300	0.5649	0.0822	0.8701	0.1085
			0.0374	0.0039	0.0338	0.0045
	900	900	0.5063	0.0820	0.8689	0.1089
			0.0130	0.0024	0.0200	0.0027
0.5	300	300	0.3382	0.1279	0.6602	0.1589
			0.0116	0.0053	0.0174	0.0063
	300	900	0.3178	0.1305	0.6567	0.1586
			0.0117	0.0056	0.0139	0.0049
	900	300	0.3054	0.0745	0.6258	0.0828
			0.0090	0.0049	0.0083	0.0053
	900	900	0.2697	0.0766	0.6203	0.0822
			0.0049	0.0022	0.0120	0.0023
0.95	300	300	0.7438	0.2779	1.1611	0.4097
			0.0240	0.0140	0.0483	0.0185
	300	900	0.6892	0.2651	1.1526	0.4063
			0.0220	0.0075	0.0375	0.0105
	900	300	0.6171	0.0904	1.0466	0.1447
			0.0227	0.0041	0.0395	0.0061
	900	900	0.5506	0.0890	1.0272	0.1437
			0.0169	0.0032	0.0266	0.0037

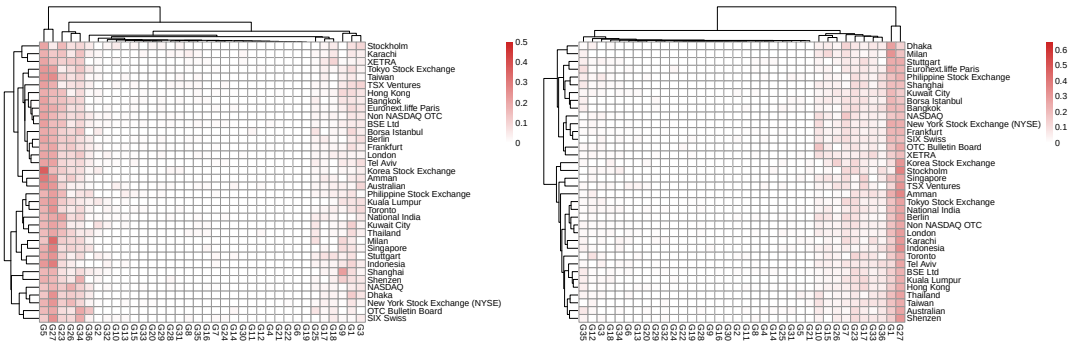
Table 3: Second data-generating process. Comparison of the proposed estimator with factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$, and the standard estimator without factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$. The mean squared errors are defined as $\text{MSE}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2$, $\text{MSE}_2 = \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$, where $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $\boldsymbol{b}_{i,0,\tau}$ and the true τ-th quantile function $Q_{it}(\tau)$, respectively. Averages over 200 simulation data sets are reported. The second lines are corresponding to the standard deviation of the mean squared errors.

			With factor structure		Without factor structure	
τ	T	N	MSE_1	MSE_2	MSE_1	MSE_2
0.05	300	300	0.9283	0.3532	1.1423	0.4061
			0.0400	0.0140	0.0423	0.0144
	300	900	0.8691	0.3517	1.1262	0.4029
			0.0408	0.0121	0.0225	0.0106
	900	300	0.6723	0.1209	0.9908	0.1445
			0.0330	0.0042	0.0349	0.0061
	900	900	0.6012	0.1217	0.9994	0.1461
			0.0139	0.0028	0.0190	0.0036
0.5	300	300	0.3429	0.1338	0.6480	0.1665
			0.0110	0.0087	0.0191	0.0077
	300	900	0.3283	0.1375	0.6708	0.1681
			0.0091	0.0060	0.0162	0.0052
	900	300	0.3062	0.0735	0.6238	0.0822
			0.0146	0.0041	0.0149	0.0043
	900	900	0.2703	0.0761	0.6253	0.0838
			0.0041	0.0038	0.0118	0.0032
0.95	300	300	0.9736	0.4000	1.3840	0.5130
			0.0403	0.0172	0.0555	0.0208
	300	900	0.8970	0.3878	1.3884	0.5138
			0.0237	0.0081	0.0425	0.0114
	900	300	0.7750	0.1309	1.2211	0.1831
			0.0283	0.0053	0.0439	0.0078
	900	900	0.6833	0.1300	1.2190	0.1810
			0.0217	0.0041	0.0349	0.0055

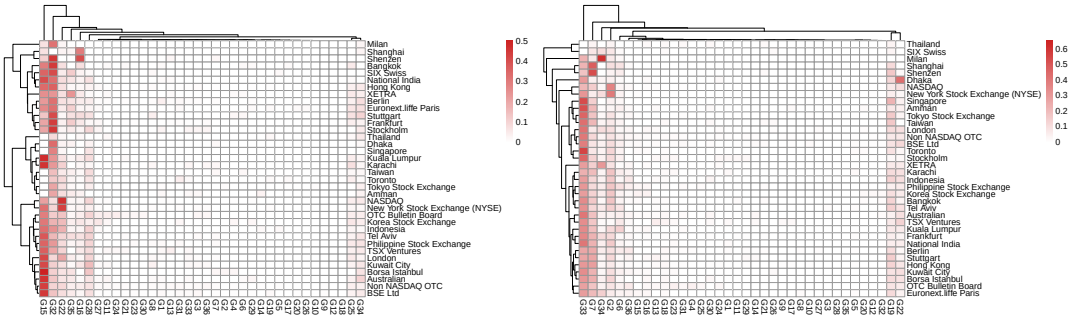
Table 4: The histogram of the selected number of common factors $\hat{r}$ over 100 simulation runs. The results are for $\tau = 5\%$ and $\tau = 95\%$ quantile points. The true number of common factors are $r_{0,\tau} = 3$ at $\tau = 5\%$ and $r_{0,\tau} = 5$ at $\tau = 95\%$.

First data generating process											
		$\hat{r}$	0	1	2	3	4	5	6	7	8
τ		T	N								
0.05 ($r_{0,\tau} = 3$)	300	300	0	0	28	41	31	0	0	0	0
	300	900	0	0	0	46	54	0	0	0	0
	900	300	0	0	0	90	10	0	0	0	0
	900	900	0	0	0	96	4	0	0	0	0
0.95 ($r_{0,\tau} = 5$)	300	300	0	0	0	0	2	98	0	0	0
	300	900	0	0	0	0	0	100	0	0	0
	900	300	0	0	0	0	0	100	0	0	0
	900	900	0	0	0	0	0	100	0	0	0

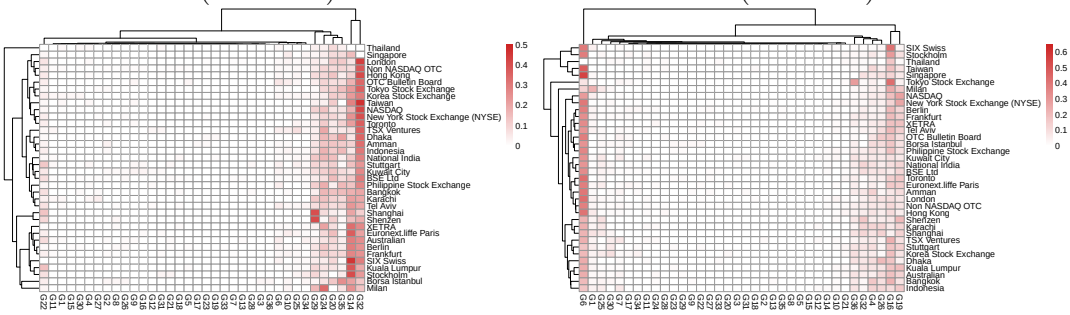
Second data generating process											
		$\hat{r}$	0	1	2	3	4	5	6	7	8
τ	T	N									
0.05 ($r_{0,\tau} = 3$)	300	300	0	10	22	38	30	0	0	0	0
	300	900	0	0	14	35	51	0	0	0	0
	900	300	0	0	8	82	10	0	0	0	0
	900	900	0	0	0	95	5	0	0	0	0
0.95 ($r_{0,\tau} = 5$)	300	300	0	0	0	16	29	55	0	0	0
	300	900	0	0	0	0	0	100	0	0	0
	900	300	0	0	0	0	0	100	0	0	0
	900	900	0	0	0	0	0	100	0	0	0



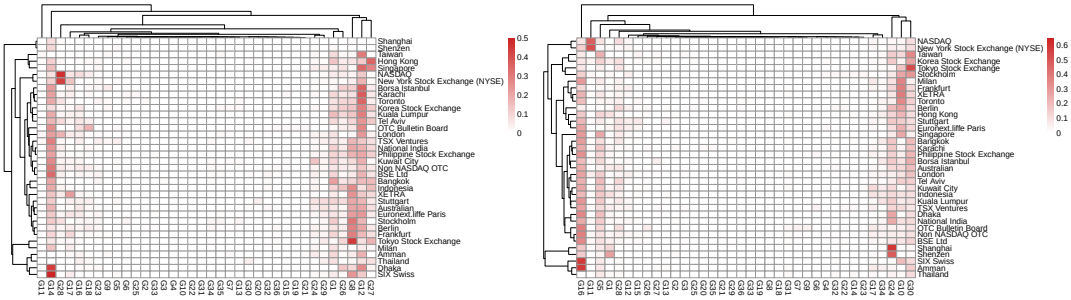
Period 1 ($\tau = 0.05$). Period 1 ($\tau = 0.95$).



Period 2 ($\tau = 0.05$). Period 2 ($\tau = 0.95$).



Period 3 ($\tau = 0.05$). Period 3 ($\tau = 0.95$).



Period 4 ($\tau = 0.05$). Period 4 ($\tau = 0.95$).

Figure 1: Distribution of firms in each of the stock exchanges (See Section 6.2.3). An (i, j)-th element denotes the percentage of firms in industry i such that they belong to the j-th group. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

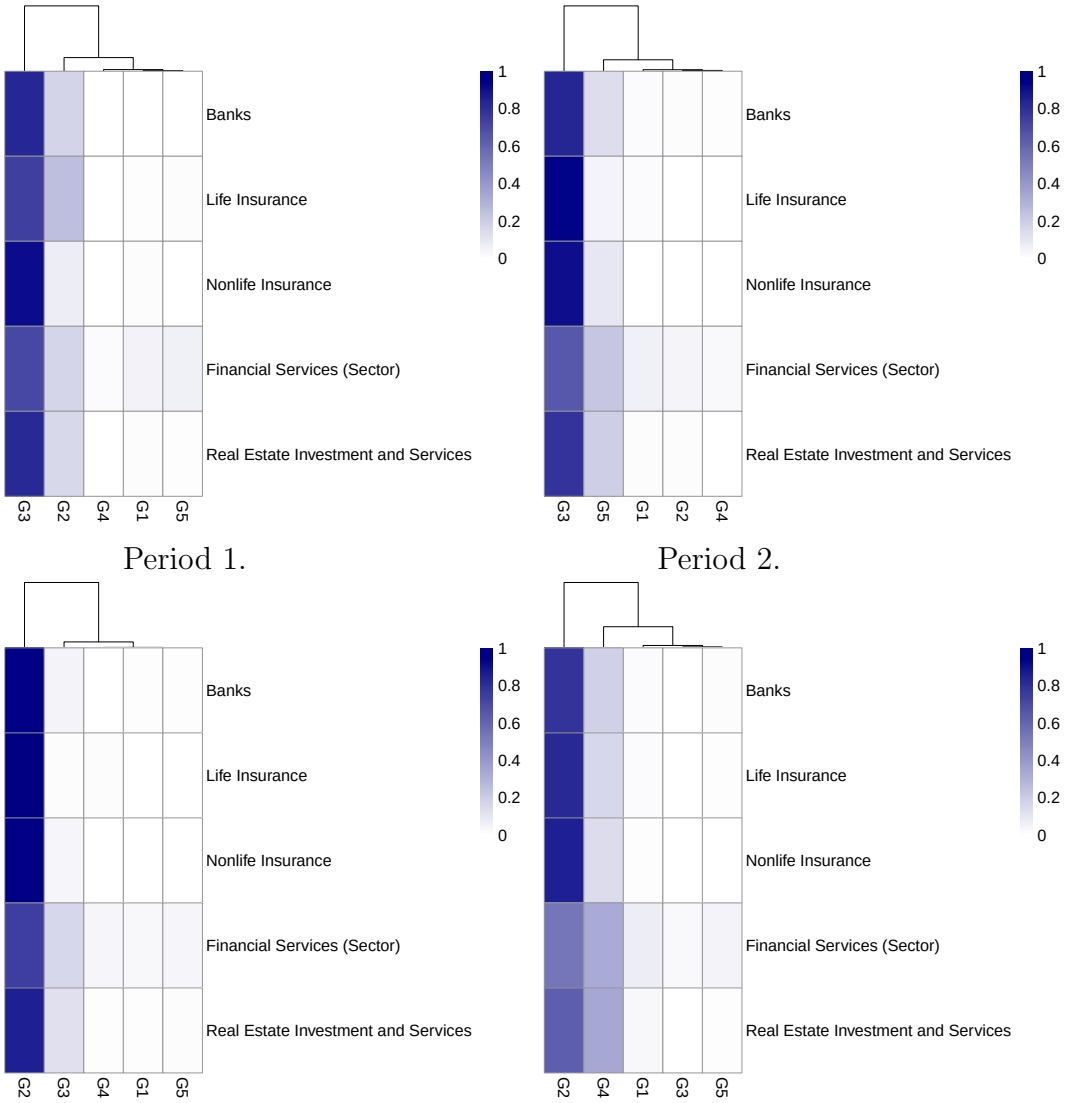
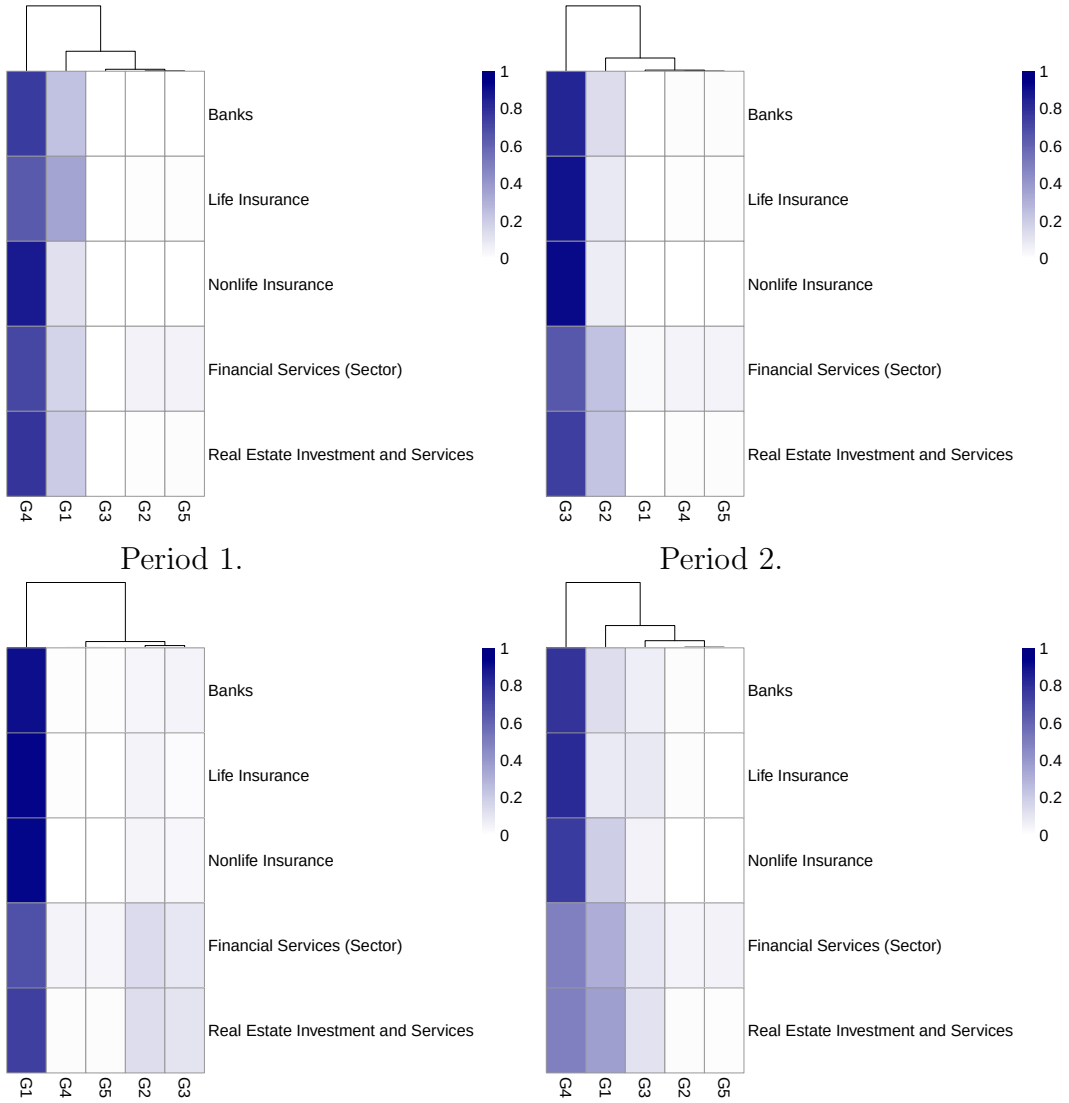


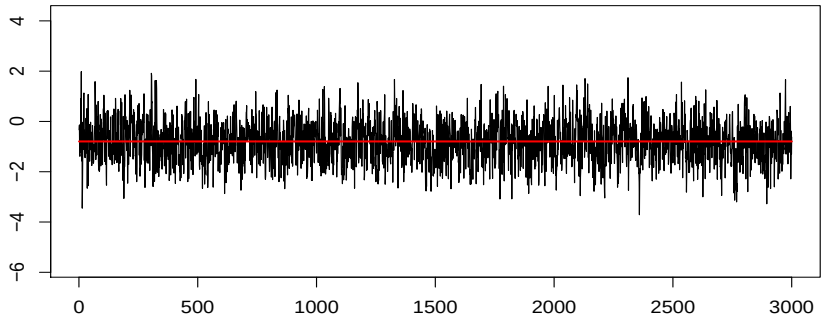
Figure 2: Lower quantile $\tau = 0.05$. Distribution of firms in each of the sectors. An (i, j)-th element denotes the percentage of firms in industry i such that they belong to the j-th group. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).



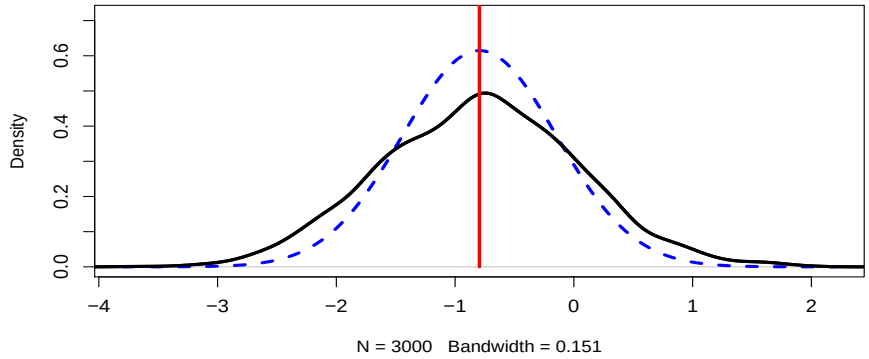
Period 1. Period 2.

Period 3. Period 4.

Figure 3: Upper quantile $\tau = 0.95$. Distribution of firms in each of the sectors. An (i, j)-th element denotes the percentage of firms in industry i such that they belong to the j-th group. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

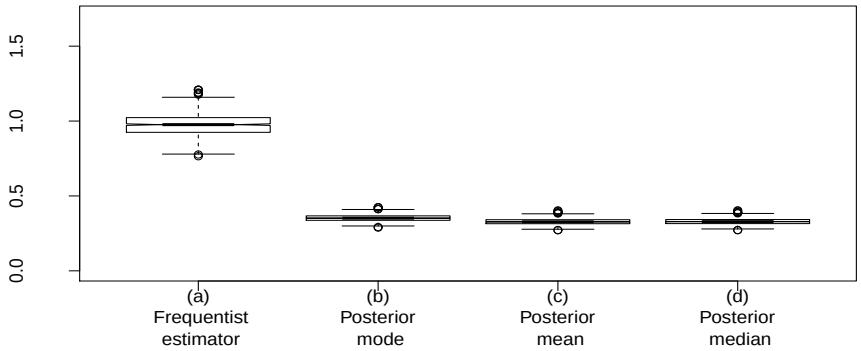


(a): Trance plot

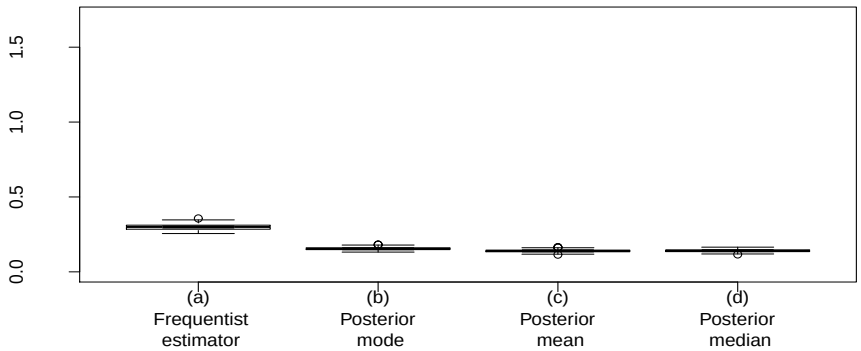


(b): Constructed distribution (from Theorem 2) for the frequestist estimator and posterior distribution

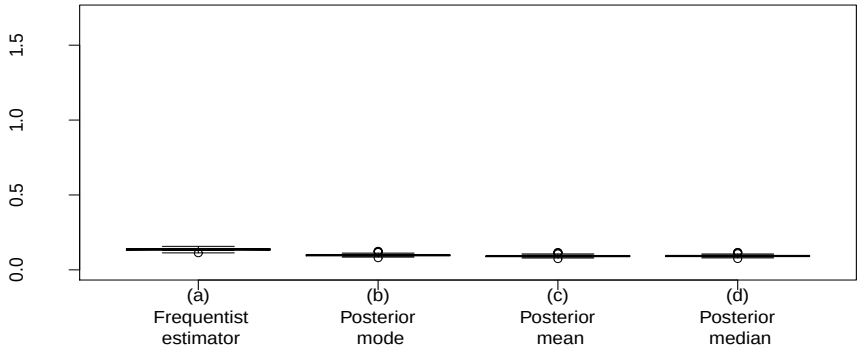
Figure 4: Summary of MCMC posterior sampling results with respect to the regression coefficient of $b_{20,3,\tau}$ at $\tau = 0.05$. A set of 3,000 samples were generated by the proposed data-augmentation algorithm. (a) Black line: trace plot of MCMC sample. Red line: true value of the regression coefficient. (b) Comparison of the constructed distribution (from Theorem 2) for the frequestist estimator and posterior distribution from MCMC for the regression coefficient. Solid black line: the constructed posterior density from MCMC output, Dashed blue line: the constructed asymptotic distribution from Theorem 2, Solid vertical line: true value of the regression coefficient.



(a): $T = 100$



(b): $T = 300$



(c): $T = 900$

Figure 5: Boxplot of the average mean squared errors: MSE $= \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$, between the true parameter vector $\boldsymbol{b}_{i,0,\tau}$ and its estimate $\hat{\boldsymbol{b}}_{i,\tau}$. (a) Frequentist estimator denotes our proposed estimator given in Section 3.1 (b) $\sim$ (d) Bayesian estimators based on the proposed data-augmentation strategy in Section 3.2. (e) Without factor structure is based on the standard quantile regression that ignores the factor structures.