Supplementary Materials to "Generalized factor model for ultra-high dimensional correlated variables with mixed types"

Wei Liu[1], Huazhen Lin[1*], Shurong Zheng[2*] and Jin Liu[3]
[1]Center of Statistical Research and School of Statistics,
Southwestern University of Finance and Economics, Chengdu, China
[2] School of Mathematics and Statistics,
Northeast Normal University, Changchun, China
[3] Centre for Quantitative Medicine,
Program in Health Services & Systems Research,
Duke-NUS Medical School

A: Notations

Let $l(\boldsymbol{\theta}, \mathbf{X}) \hat{=} l(\mathbf{H}, \Upsilon) = \sum_{i=1}^n \sum_{j=1}^p L_j(\mathbf{h}_i, \boldsymbol{\gamma}_j; x_{ij}) + C$, as the conditional loglikelihood and recall $\boldsymbol{\theta} = (\mathbf{h}_1^T, \cdots, \mathbf{h}_n^T, \boldsymbol{\gamma}_1^T, \cdots, \boldsymbol{\gamma}_p^T)^T$ defined in Section 3. Recall $m_{ji} = m_j(\boldsymbol{\gamma}_j^T \boldsymbol{\kappa}_i)$, $m'_{ji} = m'_j(\boldsymbol{\gamma}_j^T \boldsymbol{\kappa}_i)$, $a_{ji} = \frac{x_{ij} - m_j(\boldsymbol{\gamma}_j^T \boldsymbol{\kappa}_i)}{c_j}$, $\mathbf{v}_{ji}(\mathbf{h}_i, \boldsymbol{\gamma}_j) = a_{ji}\mathbf{b}_j$ and $\mathbf{w}_{ji}(\mathbf{h}_i, \boldsymbol{\gamma}_j) = a_{ji}\boldsymbol{\kappa}_i$ for $i = 1, \cdots, n, j = 1, \cdots, p$ and
$\mathbf{f}(\boldsymbol{\theta}) = \left[\sum_{j=1}^{p_1} \mathbf{v}_{j1}(\mathbf{h}_1, \boldsymbol{\gamma}_j)^T, \cdots, \sum_{j=1}^{p_1} \mathbf{v}_{jn}(\mathbf{h}_n, \boldsymbol{\gamma}_j)^T, \sum_{i=1}^n \mathbf{w}_{1i}(\mathbf{h}_i, \boldsymbol{\gamma}_1)^T, \cdots, \sum_{i=1}^n \mathbf{w}_{pi}(\mathbf{h}_i, \boldsymbol{\gamma}_p)^T\right]^T$.

Denote $\tilde{\boldsymbol{\theta}} = \left(\tilde{\mathbf{h}}_1^T, \cdots, \tilde{\mathbf{h}}_n^T, \tilde{\boldsymbol{\gamma}}_1^T, \cdots, \tilde{\boldsymbol{\gamma}}_p^T\right)^T$, $\boldsymbol{\theta}^* = \left(\mathbf{h}_{1*}^T, \cdots, \mathbf{h}_{n*}^T, \boldsymbol{\gamma}_{1*}^T, \cdots, \boldsymbol{\gamma}_{p*}^T\right)^T$, where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\theta}} = \left(\hat{\mathbf{h}}_1^T, \cdots, \hat{\mathbf{h}}_n^T, \hat{\boldsymbol{\gamma}}_1^T, \cdots, \hat{\boldsymbol{\gamma}}_p^T\right)^T$. Let $a_{ji*} = \frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j*}^T \mathbf{h}_{i*})}{c_j}$, $m'_{ji*} = m'_j(\boldsymbol{\gamma}_{j*}^T \mathbf{h}_{i*}), i = 1, \cdots, n, j = 1, \cdots, p$, $\mathbf{d}_{i*} = (-\frac{m'_{1i*}}{c_1}\mathbf{b}_{1*}\boldsymbol{\kappa}_{i*}^T + a_{1i*}\mathbf{E}, \cdots, -\frac{m'_{p_1 i*}}{c_{p_1}}\mathbf{b}_{p_1*}\boldsymbol{\kappa}_{i*}^T + a_{p_1 i*}\mathbf{E}, \mathbf{0}_{q\times((q+1)(p-p_1))})$, and $\mathbf{u}_{j*} = (-\frac{m'_{j1*}}{c_j}\boldsymbol{\kappa}_{1*}\mathbf{b}_{j*}^T + a_{j1*}\mathbf{E}^T, \cdots, -\frac{m'_{jn*}}{c_j}\boldsymbol{\kappa}_{n*}\mathbf{b}_{j*}^T + a_{jn*}\mathbf{E}^T)$, where $\mathbf{E} = (\mathbf{0}_{\mathbf{q}\times\mathbf{1}}, \mathbf{I}_q) \in R^{q\times(q+1)}$. Define

$$A_{11}(\boldsymbol{\theta}) = \operatorname{diag}\left(-\sum_{j=1}^{p_1} \frac{m'_{j1}}{c_j}\mathbf{b}_j\mathbf{b}_j^T,\ -\sum_{j=1}^{p_1} \frac{m'_{j2}}{c_j}\mathbf{b}_j\mathbf{b}_j^T, \cdots, -\sum_{j=1}^{p_1} \frac{m'_{jn}}{c_j}\mathbf{b}_j\mathbf{b}_j^T\right),$$
$$A_{22}(\boldsymbol{\theta}) = \operatorname{diag}\left(-\sum_{i=1}^{n} \frac{m'_{1i}}{c_1}\boldsymbol{\kappa}_i\boldsymbol{\kappa}_i^T,\ -\sum_{i=1}^{n} \frac{m'_{2i}}{c_2}\boldsymbol{\kappa}_i\boldsymbol{\kappa}_i^T, \cdots, -\sum_{i=1}^{n} \frac{m'_{pi}}{c_p}\boldsymbol{\kappa}_i\boldsymbol{\kappa}_i^T\right),$$
$$A_{12}(\boldsymbol{\theta}) = \left(\{-\frac{m'_{ji}}{c_j}\mathbf{b}_j\boldsymbol{\kappa}_i^T + a_{ji}\mathbf{E}\}I_{\{1\leqslant j\leqslant p_1\}}\right)_{i=1,\cdots,n;j=1,\cdots,p},$$
$$A_{21}(\boldsymbol{\theta}) = \left(-\frac{m'_{ji}}{c_j}\boldsymbol{\kappa}_i\mathbf{b}_j^T + a_{ji}\mathbf{E}^T\right)_{i=1,\cdots,n;j=1,\cdots,p}.$$

Then we have $\nabla\mathbf{f}(\boldsymbol{\theta}) = (A_{ij}(\boldsymbol{\theta}))_{i,j=1,2}$. Denote $\mathbf{D}(\boldsymbol{\theta}) = \operatorname{diag}(A_{11}(\boldsymbol{\theta}), A_{22}(\boldsymbol{\theta}))$ and $\mathbf{D}_0(\boldsymbol{\theta}) = \nabla\mathbf{f}(\boldsymbol{\theta}) - \mathbf{D}(\boldsymbol{\theta})$. Here, we use the matrix blocking technique which separates the Hessian matrix $\nabla \mathbf{f}(\boldsymbol{\theta})$ of the log-likelihood function as a two-by-two symmetric matrix with $A_{11}(\boldsymbol{\theta})$ being for factors, $A_{22}(\boldsymbol{\theta})$ being for factor loadings and $A_{12}(\boldsymbol{\theta})$ being for relations between factors and factor loadings. The advantages of matrix blocking technique are to separate the rules of factors and factor loadings. Thus, it will make the researches of the rate of convergence for estimators of factors and factor loadings clearly in Lemma 3.

B: Proof of Propositions 1–2

Proof of Propositions 1. Denote $d_{ij} = \mathbf{b}_j^T \mathbf{h}_i + \mu_j$ and $\mathbf{D} = (d_{ij})$. We first show that $\mathbf{D}$ is identifiable. Assume that there are two $\mathbf{D}_1 = (d_{ij1})$ and $\mathbf{D}_2 = (d_{ij2})$ satisfying model (1), where $d_{ij1} = \mathbf{b}_{j1}^T \mathbf{h}_{i1} + \mu_{j1}$ and $d_{ij2} = \mathbf{b}_{j2}^T \mathbf{h}_{i2} + \mu_{j2}$. Then we have

$$\exp[\{x d_{ij1} - \nu_j(d_{ij1})\}/c_j] = \exp[\{x d_{ij2} - \nu_j(d_{ij2})\}/c_j].$$

Rearranging the equation, we obtain $x(d_{ij1} - d_{ij2}) = \nu_j(d_{ij1}) - \nu_j(d_{ij2})$ for any x. Thus, we have $d_{ij1} = d_{ij2}$, that is, $\mathbf{D}_1 = \mathbf{D}_2$.

Now we show that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. By $d_{ij1} = d_{ij2}$, we obtain $\mathbf{b}_{j1}^T \mathbf{h}_{i1} + \mu_{j1} = \mathbf{b}_{j2}^T \mathbf{h}_{i2} + \mu_{j2}$. Taking average over i, we have

$$1/n \sum_{i=1}^{n} \{\mathbf{b}_{j1}^T \mathbf{h}_{i1} + \mu_{j1}\} = 1/n \sum_{i=1}^{n} \{\mathbf{b}_{j2}^T \mathbf{h}_{i2} + \mu_{j2}\}.$$

By Condition (A1) that $1/n \sum_{i=1}^{n} \mathbf{h}_{i1} = 0$ and $1/n \sum_{i=1}^{n} \mathbf{h}_{i2} = 0$, we have $\mu_{j1} = \mu_{j2}$.

Finally, we show that $\mathbf{H}_1 = \mathbf{H}_2$ and $\mathbf{B}_1 = \mathbf{B}_2$. Let $\mathbf{1}_n$ be n-dimensional column vector of ones. Noting that

$$\mathbf{D}_1^T = \mathbf{B}_1 \mathbf{H}_1^T + \boldsymbol{\mu}_1 \mathbf{1}_n^T, \quad \mathbf{D}_2^T = \mathbf{B}_2 \mathbf{H}_2^T + \boldsymbol{\mu}_2 \mathbf{1}_n^T,$$

we obtain $\mathbf{B}_1 \mathbf{H}_1^T = \mathbf{B}_2 \mathbf{H}_2^T$, then $\mathbf{B}_1 \mathbf{H}_1^T \mathbf{H}_1/n = \mathbf{B}_2 \mathbf{H}_2^T \mathbf{H}_1/n$. Coupling with the identifiability condition (A1) that $\mathbf{H}_1^T \mathbf{H}_1/n = \mathbf{I}_q$, we obtain $\mathbf{B}_1 = \mathbf{B}_2 \mathbf{H}_2^T \mathbf{H}_1/n$. Denoting $\mathbf{H}_3 = \mathbf{H}_2^T \mathbf{H}_1/n$, we have $\mathbf{H}_3^T \mathbf{H}_3 = \mathbf{I}_p$ by (A1). That is, $\mathbf{H}_3$ is a orthogonal matrix. Now we show that $\mathbf{H}_3$ is an identity matrix. Then we will prove $\mathbf{B}_1 = \mathbf{B}_2$ by the equation $\mathbf{B}_1 = \mathbf{B}_2 \mathbf{H}_2^T \mathbf{H}_1/n$. Since $\mathbf{B}_1^T \mathbf{B}_1 = \mathbf{H}_3^T \mathbf{B}_2^T \mathbf{B}_2 \mathbf{H}_3$ and $\mathbf{B}_2^T \mathbf{B}_2$ is a diagonal matrix with decreasing diagonal elements by Condition (A2), $\mathbf{H}_3^T (\mathbf{B}_2^T \mathbf{B}_2) \mathbf{H}_3$ is a singular value decomposition of $\mathbf{B}_1^T \mathbf{B}_1$. Since $\mathbf{B}_1^T \mathbf{B}_1$

is also a diagonal matrix with decreasing diagonal elements from Condition (A2), thus $\mathbf{B}_1^T\mathbf{B}_1 = \mathbf{B}_2^T\mathbf{B}_2$ and $\mathbf{H}_3$ is a sign matrix (diagonal matrix with 1 or -1 on the diagonals). Let

$$\mathbf{B}_1 = (\tilde{\mathbf{b}}_{11}, \cdots, \tilde{\mathbf{b}}_{1q}), \quad \mathbf{B}_2 = (\tilde{\mathbf{b}}_{21}, \cdots, \tilde{\mathbf{b}}_{2q}), \quad \mathbf{H}_3 = \mathrm{diag}(r_1, \cdots, r_q)$$

where r_k is 1 or -1 for each k. Then $\tilde{\mathbf{b}}_{1k} = r_k\tilde{\mathbf{b}}_{2k}$ by $\mathbf{B}_1 = \mathbf{B}_2\mathbf{H}_3$. Since the first nonzero elements of $\mathbf{b}_{1k}$ and $\mathbf{b}_{2k}$ are both positive from (A3), thus $r_k = 1$ for each k. That is, $\mathbf{H}_3 = \mathbf{I}_q$. Therefore $\mathbf{B}_1 = \mathbf{B}_2$. Further, since $\mathbf{B}_1\mathbf{H}_1^T = \mathbf{B}_2\mathbf{H}_2^T$, we obtain $\mathbf{B}_1\mathbf{H}_1^T = \mathbf{B}_1\mathbf{H}_2^T$ and $\mathbf{B}_1^T\mathbf{B}_1\mathbf{H}_1^T = \mathbf{B}_1^T\mathbf{B}_1\mathbf{H}_2^T$. By the invertibility of $\mathbf{B}_1^T\mathbf{B}_1$, we conclude that $\mathbf{H}_1 = \mathbf{H}_2$.

Before proving Proposition 2, we introduce some notations, a definition and four assumptions. Let

$$l_j(\boldsymbol{\gamma}_j; \mathbf{H}) \hat{=} \sum_{i=1}^{n} L_j(\mathbf{h}_i, \boldsymbol{\gamma}_j; x_{ij}), \quad \tilde{l}_i(\mathbf{h}_i; \Upsilon) \hat{=} \sum_{j \in A_s} L_{P_s}(\boldsymbol{\gamma}_j, \mathbf{h}_i; x_{ij}).$$

Recalling $\boldsymbol{\theta} = (\mathbf{H}, \Upsilon)$, let $F(\cdot)$ be the mapping function of the iterative algorithm, that is, the algorithm generates the sequence $\{\boldsymbol{\theta}^{[r]}\}$ by $\boldsymbol{\theta}^{[r+1]} = F(\boldsymbol{\theta}^{[r]})$. Let

$$\Omega_0 = \{(\mathbf{H}, \Upsilon) \in \Omega : l(\mathbf{H}, \Upsilon) \geqslant l(\mathbf{H}^{[0]}, \Upsilon^{[0]})\}.$$

Definition 1. *(David G. Luenberger (2016), Page 199) A point-to-set mapping F from $\mathcal{X}$ to $\mathcal{Y}$ is said to be closed at $x \in \mathcal{X}$ if the assumptions 1) $x_k \to x$ and 2) $y_k \to y, y_k \in F(x_k)$ implies $y \in F(x)$. Furthermore, F is said to be closed over $\mathcal{X}$ if F is closed at every point of $\mathcal{X}$.*

Then we give three assumptions as follows:

- **Assumption S1:** Suppose that Ω_0 is compact for the initial value $(\mathbf{H}^{[0]}, \Upsilon^{[0]})$.
- **Assumption S2:** There is a constant $c_{10} > 0$, for any $(\mathbf{H}, \Upsilon) \in \Omega_0$, such that the minimum eigenvalue of $\frac{-\partial^2 l_j(\boldsymbol{\gamma}_j;\mathbf{H})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^T}$ is lower bounded by c_{10}.
- **Assumption S3:** There is a constant $c_{20} > 0$, for any $(\mathbf{H}, \Upsilon) \in \Omega_0$, such that the minimum eigenvalue of $\frac{-\partial^2 \tilde{l}_i(\mathbf{h}_i;\Upsilon)}{\partial\mathbf{h}_i\partial\mathbf{h}_i^T}$ is lower bounded by c_{20}.
- **Assumption S4:** F is closed over $\Omega_0 \backslash \mathcal{G}$, where $\Omega_0 \backslash \mathcal{G}$ is the difference of these two sets.

<u>Remark</u> 1. In a special case that all variables are normal with homoscedasticity, i.e.

$$c_j = \sigma^2, \quad m_j'(x) = 1, \quad j = 1, ..., n,$$

we can easily show that Assumptions S2 and S3 hold. Because

$$\frac{-\partial^2 l_j(\boldsymbol{\gamma}_j; \mathbf{H})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^T} = \sum_{i=1}^{n} \frac{m_{1i}'}{c_j} \boldsymbol{\kappa}_i \boldsymbol{\kappa}_i^T,$$

we have $\frac{-\partial^2 l_j(\boldsymbol{\gamma}_j; \mathbf{H})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^T} = \sigma^{-2} \mathbf{K}^T \mathbf{K}$ whose minimum eigenvalue has lower bounded by the identifiability condition on $\mathbf{H}$. Similarly,

$$\frac{-\partial^2 \tilde{l}_i(\mathbf{h}_i; \Upsilon)}{\partial \mathbf{h}_i \partial \mathbf{h}_i^T} = \sum_{j=1}^{p} \frac{m_{j1}'}{c_j} \mathbf{b}_j \mathbf{b}_j^T = \sigma^{-2} \mathbf{B}^T \mathbf{B}$$

whose minimum eigenvalue has lower bounded by the identifiability condition on $\mathbf{B}$.

<u>Remark</u> 2. We give a simple case where all variables are normal with homoscedasticity that Assumption S4 holds. In this case, $(\mathbf{H}^{[r+1]}, \Upsilon^{[r+1]})$ has the closed form given by

$$\mathbf{H}^{[r+1]} = \mathbf{X}\mathbf{B}^{[r]}(\mathbf{B}^{[r],T}\mathbf{B}^{[r]})^{-1}, \Upsilon^{[r+1]} = \mathbf{X}^T\mathbf{K}^{[r]}(\mathbf{K}^{[r],T}\mathbf{K}^{[r]})^{-1},$$

where $\mathbf{K}^{[r]} = (1_n, \mathbf{H}^{[r]})$. Thus, the mapping function is

$$F(\mathbf{H}^{[r]}, \Upsilon^{[r]}) = (\mathbf{X}\mathbf{B}^{[r]}(\mathbf{B}^{[r],T}\mathbf{B}^{[r]})^{-1}, \mathbf{X}^T\mathbf{K}^{[r]}(\mathbf{K}^{[r],T}\mathbf{K}^{[r]})^{-1}),$$

which is a point-to-point mapping, a special case of point-to-set mapping. Obviously, F is a continuous mapping, which indicates that F is closed by Definition 1.

Proof of Propositions 2. The proof includes five steps as follows:

- Step 1: Prove that $F(\cdot)$ is a point-to-point mapping;
- Step 2: Show that $l(\cdot)$ is nondecreasing with respect to the sequence $\{\boldsymbol{\theta}^{[r]}\}$, i.e. $l(\boldsymbol{\theta}^{[r]}) \leqslant l(\boldsymbol{\theta}^{[r+1]})$;

- Step 3: Show that if $\boldsymbol{\theta} \notin \mathcal{G}$, then $l(\boldsymbol{\theta}_1) < l(\boldsymbol{\theta}_2)$ for $\boldsymbol{\theta}_2 = F(\boldsymbol{\theta}_1)$, and if $\boldsymbol{\theta}_1 \in \mathcal{G}$, then $l(\boldsymbol{\theta}_1) \leqslant l(\boldsymbol{\theta}_2)$ for $\boldsymbol{\theta}_2 = F(\boldsymbol{\theta}_1)$;

  Step 4: Show that $l(\boldsymbol{\theta}^{[r]})$ converges monotonically to $L^* = l(\boldsymbol{\theta}^*)$ for some $\boldsymbol{\theta}^*$, where $\boldsymbol{\theta}^*$ is the limit of a subsequence of $\{\boldsymbol{\theta}^{[r]}\}$;

  Step 5: Show $\boldsymbol{\theta}^* \in \mathcal{G}$.

Step 1. First, we prove that $F(\cdot)$ is a point-to-point mapping.

By Assumption S2, given $\mathbf{H}^{[r]}$, $l_j(\boldsymbol{\gamma}_j; \mathbf{H}^{[r]})$ is a concave function with respect to $\boldsymbol{\gamma}_j$ where l_j can be regarded as a loglikelihood of a generalized linear model with covariate matrix $\mathbf{H}^{[r]}$. Therefore, we will obtain a unique iterative value $\boldsymbol{\gamma}^{[r+1]}$ by maximizing $l_j(\boldsymbol{\gamma}_j; \mathbf{H}^{[r]})$. Similarly, given $\Upsilon^{[r+1]}$ such that identifiability Conditions (A2)-(A3), $\tilde{l}_i(\mathbf{h}_i; \Upsilon^{[r+1]})$ is concave with respect to $\mathbf{h}_i$ by Assumption S3. Thus, a unique iterative value $\mathbf{H}^{[r+1]}$ can be obtained. That is, there exists a unique $\boldsymbol{\theta}^{[r+1]}$ such that

$$\boldsymbol{\theta}^{[r+1]} = F(\boldsymbol{\theta}^{[r]}).$$

Step 2: Show that $l(\boldsymbol{\theta})$ is nondecreasing with respect to the sequence $\{\boldsymbol{\theta}^{[r]}\}$, i.e.

$$l(\boldsymbol{\theta}^{[r-1]}) \leqslant l(\boldsymbol{\theta}^{[r]}).$$

From Equation (3) in the main paper, for any $j \in \{1, \cdots, p\}$ and Υ, given $\mathbf{H}^{[r-1]}$, it holds that

$$l_j(\boldsymbol{\gamma}_j; \mathbf{H}^{[r-1]}) \leqslant l_j(\boldsymbol{\gamma}_j^{[r]}; \mathbf{H}^{[r-1]}).$$

Taking summation over j, we have

$$l(\mathbf{H}^{[r-1]}, \Upsilon^{[r-1]}) \leqslant l(\mathbf{H}^{[r-1]}, \Upsilon^{[r]}). \tag{1}$$

Then, we consider two cases, where Case I represents that there is only one variable type in $\mathbf{x}_i$, and Case II represents that there are two or more variable types in $\mathbf{x}_i$.

Case I: In this case, by Equation (4) in the main paper, for any $\mathbf{h}_i$, we obtain

$$\sum_{j=1}^{p} L_j(\mathbf{h}_i, \boldsymbol{\gamma}_j^{[r]}; x_{ij}) \leqslant \sum_{j=1}^{p} L_j(\mathbf{h}_i^{[r]}, \boldsymbol{\gamma}_j^{[r]}; x_{ij}).$$

Taking summation over i, we obtain

$$l(\mathbf{H}^{[r-1]}, \Upsilon^{[r]}) \leqslant l(\mathbf{H}^{[r]}, \Upsilon^{[r]}). \tag{2}$$

Combing (1) and (2), we get $l(\mathbf{H}^{[r-1]}, \Upsilon^{[r-1]}) \leqslant l(\mathbf{H}^{[r]}, \Upsilon^{[r]})$.

Case II: In this case, we can obtain

$$\sum_{j \in A_s} L_j(\mathbf{h}_i^{[r-1]}, \boldsymbol{\gamma}_j^{[r]}; x_{ij}) \leqslant \sum_{j \in A_s} L_j(\mathbf{h}_i^{[r]}, \boldsymbol{\gamma}_j^{[r]}; x_{ij}).$$

Taking summation over i, we get $l_s(\mathbf{H}^{[r-1]}, \Upsilon^{[r]}) \leqslant l_s(\mathbf{H}^{[r]}, \Upsilon^{[r]})$. We can always ensure (2) by some coupling on updating of $\mathbf{H}$. For $\mathbf{H}^{[r]}$, if it also holds that

$$l(\mathbf{H}^{[r-1]}, \Upsilon^{[r]}) \leqslant l(\mathbf{H}^{[r]}, \Upsilon^{[r]}),$$

we update $\mathbf{H}$ in r-th iteration as $\mathbf{H}^{[r]}$; otherwise, $\mathbf{H}^{[r]}$ is retained as $\mathbf{H}^{[r-1]}$. Thus, Equation (2) also holds, and we obtain

$$l(\mathbf{H}^{[r-1]}, \Upsilon^{[r-1]}) \leqslant l(\mathbf{H}^{[r]}, \Upsilon^{[r]}).$$

Combing Case I with Case II, we showed that $l(\boldsymbol{\theta})$ is nondecreasing with respect to the sequence $\{\boldsymbol{\theta}^{[r]}\}$.

Step 3. We show that if $\boldsymbol{\theta}_1 \notin \mathcal{G}$, then $l(\boldsymbol{\theta}_1) < l(\boldsymbol{\theta}_2)$ for $\boldsymbol{\theta}_2 = F(\boldsymbol{\theta}_1)$; and if $\boldsymbol{\theta}_1 \in \mathcal{G}$, then $l(\boldsymbol{\theta}_1) \leqslant l(\boldsymbol{\theta}_2)$ for $\boldsymbol{\theta}_2 = F(\boldsymbol{\theta}_1)$. It is followed directly by the proof of Step 2 and the definition of $\mathcal{G}$.

Step 4. Show that $l(\mathbf{H}^{[r]}, \Upsilon^{[r]})$ converges monotonically to $L^* = l(\mathbf{H}^*, \Upsilon^*)$ for some $(\mathbf{H}^*, \Upsilon^*) \in \mathcal{G}$.

By Assumption S1, for $\{\boldsymbol{\theta}^{[r]}\} \subset \Omega_0$, we can find a convergent subsequence $\{\boldsymbol{\theta}^{[r_i]}\}$ converging to the limit $\boldsymbol{\theta}^*$, where $\{r_i, r_i < r_{i+1}, i = 1, 2, ...\}$ is a subsequence of $\{1, 2, 3, ...\}$. Since $l(\boldsymbol{\theta})$ is continuous, it follows that

$$\lim_{i \to \infty} l(\boldsymbol{\theta}^{[r_i]}) = l(\boldsymbol{\theta}^*), \quad l(\boldsymbol{\theta}^{[r_i]}) \leqslant l(\boldsymbol{\theta}^*), \tag{3}$$

with $l(\boldsymbol{\theta}^{[r_i]})$ being monotonic about i. Moreover, we have

$$l(\boldsymbol{\theta}^{[r]}) \leqslant l(\boldsymbol{\theta}^{[r+1]}), \quad r = 1, 2, ...$$

Then for $r = 1, 2, ...,$ there exists a r_i satisfying $r < r_i$ and $l(\boldsymbol{\theta}^{[r]}) \leqslant l(\boldsymbol{\theta}^{[r_i]}) \leqslant l(\boldsymbol{\theta}^{*})$. That is, $l(\boldsymbol{\theta}^{[r]})$ is a monotonic and bounded sequence. Because the subsequence $\{l(\boldsymbol{\theta}^{[r_i]})\}$ converges to $l(\boldsymbol{\theta}^{*})$, then

$$\lim_{i\to\infty} l(\boldsymbol{\theta}^{[r]}) = l(\boldsymbol{\theta}^{*}). \tag{4}$$

Step 5. Finally, we use contradiction approach to show $\boldsymbol{\theta}^{*} \in \mathcal{G}$.

Suppose that $\boldsymbol{\theta}^{*}$ is not in $\mathcal{G}$. We consider the subsequence $\{\boldsymbol{\theta}^{[r_i+1]}\}$, where r_i is same as the above one in Step 4. Since all members of this sequence are contained in a compact set, there is a convergent subsequence $\{\boldsymbol{\theta}^{[r_{i_k}+1]}, i_1 < i_2 < \cdots\}$ such that

$$\lim_{k\to\infty} \boldsymbol{\theta}^{[r_{i_k}+1]} = \boldsymbol{\theta}^{**}. \tag{5}$$

Noting that $\{\boldsymbol{\theta}^{[r_{i_k}]}\}$ is a subsequence of $\{\boldsymbol{\theta}^{[r_i]}\}$, we have

$$\lim_{k\to\infty} \boldsymbol{\theta}^{[r_{i_k}]} = \boldsymbol{\theta}^{*}. \tag{6}$$

Using the monotonicity of $l(\cdot)$ and $r_{i_k} < r_{i_k} + 1 \leqslant r_{i_{k+1}}$, we obtain

$$l(\boldsymbol{\theta}^{[r_{i_k}]}) \leqslant l(\boldsymbol{\theta}^{[r_{i_k}+1]}) \leqslant l(\boldsymbol{\theta}^{[r_{i_{k+1}}]}),$$

which implies $l(\boldsymbol{\theta}^{*}) = l(\boldsymbol{\theta}^{**})$ by the continuity of $l(\cdot)$.

By Assumption S4 and the fact that $\boldsymbol{\theta}^{[r_{i_k}+1]} = F(\boldsymbol{\theta}^{[r_{i_k}]})$, it follows that $\boldsymbol{\theta}^{**} = F(\boldsymbol{\theta}^{*})$. However, if $\boldsymbol{\theta}^{*}$ is not in $\mathcal{G}$, we have $l(\boldsymbol{\theta}^{**}) > l(\boldsymbol{\theta}^{*})$ by the results of Step 3. So the contradiction happens. Therefore, $\boldsymbol{\theta}^{*} \in \mathcal{G}$.

Combing the results of Step 4 and Step 5, we complete the proofs of proposition 2.

C: Proofs of Theorems 1 and 2

Preliminary Lemma

Lemma 1. *Under Conditions (C1)–(C6), we have*

$$\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(1), \quad \sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1). \tag{7}$$

Proof. The proof includes 2 steps. The first step proves that one of $\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(1)$ and $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1)$ holds, and the second step is to prove (7).

Step 1. Assume $\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| \neq O_p(1)$ and $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| \neq O_p(1)$. That is, there exists an $\epsilon > 0$ satisfying $P(\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| > M_1) > \epsilon$ and $P(\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| > M_1) > \epsilon$ for any constant $M_1 > 0$, or

$$P(\sup_i \|\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}\| > M_1) > \epsilon, P(\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| > M_1) > \epsilon. \tag{8}$$

Hence, we have $P(\forall M_2 > 0, \exists 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p$ satisfying $|\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i| > M_2) \geqslant \epsilon$.

On the other hand, we have

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{X})}{\partial \boldsymbol{\gamma}_j} = \sum_{i=1}^{n} \left[\frac{x_{ij} - m_j(\boldsymbol{\gamma}_j^T \boldsymbol{\kappa}_i)}{c_j} \boldsymbol{\kappa}_i \right], \tag{9}$$

and

$$\frac{\partial l(\tilde{\boldsymbol{\theta}}, \mathbf{X})}{\partial \tilde{\boldsymbol{\gamma}}_j} = 0. \tag{10}$$

By (8) and (9), we obtain

$$P(\frac{\partial l(\tilde{\boldsymbol{\theta}}, \mathbf{X})}{\partial \tilde{\boldsymbol{\gamma}}_j} \neq 0) \geqslant \epsilon. \tag{11}$$

Then (10) and (11) contradicts. The proof of Step 1 is completed.

Step 2. Assume $\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| \neq O_p(1)$ and $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1)$. That is, there exists an $\epsilon > 0$ satisfying $P(\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| > M_1) > \epsilon$ for any $M_1 > 0$, or

$$P(\sup_i \|\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}\| > M_1) > \epsilon. \tag{12}$$

Because $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1)$, then for any $\epsilon_2 > 0$, there exists a M_2 satisfying $P(\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| > M_2) < \epsilon_2$, that is, $P(\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| \leqslant M_2) \geqslant 1 - \epsilon_2$. Thus, we have

$$P(\frac{\partial l(\tilde{\boldsymbol{\theta}}, \mathbf{X})}{\partial \tilde{\boldsymbol{\gamma}}_j} \neq 0) > 0. \tag{13}$$

Then (13) and (10) contradicts. Thus,

$$\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(1). \tag{14}$$

By using the same proof of (14), we have $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1)$. Then (7) holds.

Lemma 2. *Under Conditions (C1)–(C6), we have*

$$p^{-1}\sum_{j=1}^{p} a_{ji0}^4 = O_p(1), \quad p^{-1}\sum_{j=1}^{p} a_{ji0}^2/c_j^2 = O_p(1) \ \textit{uniformly over } i \ , \tag{15}$$

$$\sup_i \left[p^{-1}\sum_{j=1}^{p} \left\| \frac{m'_{ji*}}{c_j} \mathbf{b}_{j*} \boldsymbol{\kappa}_{i*}^T \right\|^2 \cdot \left\| n^{-1} \sum_{\ell=1}^{n} \mathbf{w}_{j\ell}(\mathbf{h}_{\ell 0}, \boldsymbol{\gamma}_{j0}) \right\|^2 \right] = O_p(n^{-1}), \tag{16}$$

$$\sup_i \left[p^{-1}\sum_{j=1}^{p} a_{ji0}^2 \left\| n^{-1} \sum_{\ell=1}^{n} \mathbf{w}_{j\ell}(\mathbf{h}_{\ell 0}, \boldsymbol{\gamma}_{j0}) \right\|^2 \right] = O_p(n^{-1}), \tag{17}$$

$$\sup_i \left[p^{-1}\sum_{j=1}^{p} [(m_{ji0} - m_{ji*})^2/c_j^2] \cdot \left\| n^{-1} \sum_{\ell=1}^{n} \mathbf{w}_{j\ell}(\mathbf{h}_{\ell 0}, \boldsymbol{\gamma}_{j0}) \right\|^2 \right] = O_p(n^{-1}). \tag{18}$$

Proof. The proof includes 2 steps. Step 1 is to prove (15), Step 2 is to prove (16). (17) and (18) can be similarly proved and omitted here.

Step 1: Denote $\varsigma_{is} = p^{-1}\sum_{j=1}^{p} a_{ji0}^s$. By (C3.2) and $\mathrm{E}a_{ji0}^s = s\int_0^{+\infty} t^{(s-1)} P(|a_{ji0}| > t)dt$, we obtain $\mathrm{E}a_{ji0}^4 \leqslant M$ uniformly over i. Then for any $\epsilon > 0$ and all i's, there exists $M_0 = M\epsilon^{-1} > 0$ satisfying $P(\varsigma_{i4} > M_0) \leqslant \frac{\mathrm{E}\varsigma_{i4}}{M_0} \leqslant \frac{M}{M_0} = \epsilon$. That is,

$$\varsigma_{i4} = p^{-1}\sum_{j=1}^{p} a_{ji0}^4 = O_p(1), \tag{19}$$

uniformly over i. By Condition (C6): $\inf_j c_j \geqslant c > 0$, $\mathrm{E}a_{ji0}^2 \leqslant 1 + \mathrm{E}a_{ji0}^4$ and Markov inequality, for any $\epsilon > 0$ and all i's, there exists $M_1 = (M+1)c^{-2}\epsilon^{-1} > 0$ satisfying

$$P(p^{-1}\sum_{j=1}^{p}(a_{ji0}^2 c_j^{-2}) > M_1) \leqslant \frac{\mathrm{E}\varsigma_{i2}}{c^2 M_1} \leqslant \frac{\mathrm{E}\varsigma_{i4}+1}{c^2 M_1} \leqslant \frac{M+1}{c^2 M_1} = \epsilon.$$

That is,

$$p^{-1}\sum_{j=1}^{p} a_{ji0}^2/c_j^2 = O_p(1), \tag{20}$$

uniformly over i. Combining (19) and (20), equation (15) holds.

Step 2: Denote $\mathbf{G}_{ji*} = \frac{m'_{ji*}}{c_j}\mathbf{b}_{j*}\boldsymbol{\kappa}_{i*}^T, \mathbf{G}_{ji0} = \frac{m'_{ji0}}{c_j}\mathbf{b}_{j0}\boldsymbol{\kappa}_{i0}^T$ and $\mathbf{w}_{ji0} = \mathbf{w}_{ji}(\mathbf{h}_{i0}, \boldsymbol{\gamma}_{j0})$. Note

$$\begin{aligned}
&\sup_i \left[p^{-1}\sum_{j=1}^{p}\left\|\frac{m'_{ji*}}{c_j}\mathbf{b}_{j*}\boldsymbol{\kappa}_{i*}^T\right\|^2 \cdot \left\| n^{-1}\sum_{\ell=1}^{n}\mathbf{w}_{j\ell}(\mathbf{h}_{\ell 0}, \boldsymbol{\gamma}_{j0})\right\|^2\right] \\
&\leqslant \sup_i \left[\frac{2}{p}\sum_{j=1}^{p}\|\mathbf{G}_{ji*}\|^2 \cdot \left\|n^{-1}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2\right] + \sup_i\left[\frac{2}{p}\sum_{j=1}^{p}\|\mathbf{G}_{ji*}\|^2\cdot\left\|n^{-1}\mathbf{w}_{ji0}\right\|^2\right] \\
&\hat{=} 2II_1 + 2II_2.
\end{aligned} \tag{21}$$

Using the fact that $\|\mathbf{A}_1 + \mathbf{A}_2\|^2 \leqslant 2\|\mathbf{A}_1\|^2 + 2\|\mathbf{A}_2\|^2$, we have $\|\mathbf{G}_{ji*}\|^2 \leqslant 2\|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 + 2\|\mathbf{G}_{ji0}\|^2$. Thus, we obtain

$$\begin{aligned}
II_1 &\leqslant \sup_i\left[\frac{2}{p}\sum_{j=1}^{p}\|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \cdot \left\|\frac{1}{n}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2\right] + \sup_i\left[\frac{2}{p}\sum_{j=1}^{p}\|\mathbf{G}_{ji0}\|^2\cdot\left\|\frac{1}{n}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2\right] \\
&\hat{=} 2II_{11} + 2II_{12}.
\end{aligned} \tag{22}$$

Step 2.1: We aim to prove

$$II_{11} = \sup_i\left[\frac{1}{p}\sum_{j=1}^{p}\|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2\cdot\left\|\frac{1}{n}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2\right] = O_p(\frac{1}{n}). \tag{23}$$

By Lemma 1 that $\sup_i\|\hat{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(1), \sup_j\|\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1)$ and Condition (C5) that $m_j''(\cdot)$ is continuous, we have, for any $\epsilon_1 > 0$,

$$P(\sup_{i,j} \|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 > M_1) \leqslant \epsilon_1/2, \tag{24}$$

by which, there exists $\delta > 0$ such that

$$\begin{aligned}
& P\left[\sup_i \left\{p^{-1}\sum_{j=1}^{p} \|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \cdot \left\|n^{-1/2}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2\right\} > \delta\right] \\
\leqslant\ & P\left[\sup_i \left\{p^{-1}\sum_{j=1}^{p} \|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \cdot \left\|n^{-1/2}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2\right\} > \delta\ , \right. \\
& \left. \sup_{i,j}\|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \leqslant M_1\right] + P\left[\sup_{i,j}\|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \geqslant M_1\right] \\
\leqslant\ & \frac{\epsilon_1}{2} + P\left[p^{-1}M_1\sum_{j=1}^{p}\left\|n^{-1/2}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2 > \delta\right] \\
\leqslant\ & \frac{\epsilon_1}{2} + \frac{\mathrm{E}p^{-1}M_1\sum_{j=1}^{p}\left\|n^{-1/2}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}\right\|^2}{\delta} \leqslant \frac{\epsilon_1}{2} + \frac{M_2}{\delta} \leqslant \epsilon_1,
\end{aligned}$$

where we take $\delta = 2M_2/\epsilon_1$. Thus, we get (23).

Step 2.2: By (C2.2) and $\mathrm{E}h_{ik}^4 = 4\int_0^{+\infty} t^3 P(|h_{ik}| > t)dt$, we obtain $\mathrm{E}h_{ik}^4 \leqslant M$ for all k. And by (C3): $\mathrm{E}a_{ji0}^4 \leqslant M$ uniformly for j, we have $n^{-1}\sum_{\ell\neq i}\mathrm{E}[\mathbf{w}_{j\ell 0}] = 0$ and $\mathrm{E}\{[n^{-1}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0}]^2\} \leqslant n^{-1}C$, where the constant $C > 0$ for all j. Then we have

$$n^{-1}\sum_{\ell\neq i}\mathbf{w}_{j\ell 0} = O_p(n^{-1/2}), \tag{25}$$

uniformly for j. By the continuousness of m_j, $\mathbf{h}_i$'s are iid for all i, (C2.1): $\|\mathbf{b}_{j0}\|^2$ is bounded and (C6): $\inf_j c_j \geqslant c$, then

$$p^{-1}\sum_{j=1}^{p}\left\|\frac{m_{ji0}}{c_j}\mathbf{b}_{j0}\boldsymbol{\kappa}_{i0}^T\right\|^2 = O_p(1), \tag{26}$$

uniformly over i. Combining (25) and (26), we have,

$$II_{12} = \sup_i \left\{ p^{-1} \sum_{j=1}^{p} \left\| \frac{m_{ji0}}{c_j} \mathbf{b}_{j0} \boldsymbol{\kappa}_{i0}^T \right\|^2 \cdot \left\| n^{-1} \sum_{\ell \neq i} \mathbf{w}_{j\ell 0} \right\|^2 \right\} = O_p(n^{-1}). \tag{27}$$

Step 2.3: Using the same argument, we obtain

$$II_2 = \sup_i \left\{ p^{-1} \sum_{j=1}^{p} \left\| \frac{m'_{ji*}}{c_j} \mathbf{b}_{j*} \boldsymbol{\kappa}_{i*}^T \right\|^2 \cdot \left\| n^{-1} \mathbf{w}_{ji0} \right\|^2 \right\} = o_p(\frac{1}{n}). \tag{28}$$

Thus, combining (22), (23), (27), (28) and (21), we have (16). $\square$

Lemma 3 and its proof

The following lemma shows the convergent rate of the initial estimator $(\widetilde{\mathbf{H}}, \widetilde{\Upsilon})$.

Lemma 3. *Under Conditions (C1)–(C6), we have*

$$\|\widetilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(C_{np}^{-1}), \|\widetilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(C_{np}^{-1})$$

and

$$\sup_i \|\widetilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|^2 = O_p(n^{-1} + p^{-1} n^{1/\tau}), \sup_j \|\widetilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|^2 = O_p(p^{-1} + n^{-1} \ln p),$$

where $\tau \geqslant 4$ is defined in Condition (C3.2).

Proof. The proof includes 2 steps. Step 1 is to prove the order of $\|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|$, and Step 2 is to prove the order of $\|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|$.

Step 1: It is noted that $\nabla \mathbf{f}(\boldsymbol{\theta}) = \mathbf{D}(\boldsymbol{\theta}) + \mathbf{D}_0(\boldsymbol{\theta})$, we have

$$\nabla \mathbf{f}(\boldsymbol{\theta})^{-1} = \mathbf{D}^{-1}(\boldsymbol{\theta}) - \mathbf{D}^{-1}(\boldsymbol{\theta}) \mathbf{D}_0(\boldsymbol{\theta}) \nabla \mathbf{f}(\boldsymbol{\theta})^{-1}. \tag{29}$$

Here, the Hessian matrix $\nabla \mathbf{f}(\boldsymbol{\theta})$ is separated as a two-by-two matrix $(A_{ij}(\boldsymbol{\theta}))_{i,j=1,2}$ with $A_{11}(\boldsymbol{\theta})$ being for factors, $A_{22}(\boldsymbol{\theta})$ being for factor loadings and $A_{12}(\boldsymbol{\theta})$ being for relations between factors and factor loadings. The matrix blocking technique will make the researches of the rate of convergence of $\|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|$ and $\|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|$ clearly.

On the other hand, we have $0 = \mathbf{f}(\tilde{\boldsymbol{\theta}}) = \mathbf{f}(\boldsymbol{\theta}_0) + \nabla\mathbf{f}(\boldsymbol{\theta}^*)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}^*$, defined in Supplementary Materials A, lies between $\boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{\theta}}$. Combining it with (29), we have

$$\begin{aligned}\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 &= -\nabla\mathbf{f}(\boldsymbol{\theta}^*)^{-1}\mathbf{f}(\boldsymbol{\theta}_0) \\ &= -\mathbf{D}^{-1}(\boldsymbol{\theta}^*)\mathbf{f}(\boldsymbol{\theta}) + \mathbf{D}^{-1}(\boldsymbol{\theta}^*)\mathbf{D}_0(\boldsymbol{\theta}^*)\nabla\mathbf{f}(\boldsymbol{\theta}^*)^{-1}\mathbf{f}(\boldsymbol{\theta}_0). \qquad (30)\end{aligned}$$

Recalling $\mathbf{d}_{i*} = (-\frac{m'_{1i*}}{c_1}\mathbf{b}_{1*}\boldsymbol{\kappa}_{i*}^T + a_{1i*}\mathbf{E}, \cdots, -\frac{m'_{p_1i*}}{c_{p_1}}\mathbf{b}_{p_1*}\boldsymbol{\kappa}_{i*}^T + a_{p_1i*}\mathbf{E}, \mathbf{0}_{q\times((q+1)(p-p_1))})$, and $\mathbf{u}_{j*} = (\frac{m'_{j1*}}{c_j}\boldsymbol{\kappa}_{1*}\mathbf{b}_{j*}^T + a_{j1*}\mathbf{E}^T, \cdots, \frac{m'_{jn*}}{c_j}\boldsymbol{\kappa}_{n*}\mathbf{b}_{j*}^T + a_{jn*}\mathbf{E}^T)$. From (30), we have

$$\begin{aligned}\tilde{\mathbf{h}}_i - \mathbf{h}_{i0} &= \Phi_{i*}^{-1}\sum_{j=1}^{p_1}\mathbf{v}_{ji}(\mathbf{h}_{i0}, \boldsymbol{\gamma}_{j0}) - \Phi_{i*}^{-1}\left(\mathbf{0}_{q\times(nq)}, \mathbf{d}_{i*}\right)\nabla\mathbf{f}(\boldsymbol{\theta}^*)^{-1}\mathbf{f}(\boldsymbol{\theta}_0) \\ &\hat{=}\ \mathbf{r}_{i1} - \mathbf{r}_{i2}. \qquad (31)\end{aligned}$$

where $\Phi_{i*} = \sum_{j=1}^{p_1}\frac{m'_{ji*}}{c_j}\mathbf{b}_{j*}\mathbf{b}_{j*}^T$. Denote $\mathbf{A}_* = \mathrm{diag}(p^{-1}\mathbf{I}_{nq}, n^{-1}\mathbf{I}_{(pq+p)})\nabla\mathbf{f}(\boldsymbol{\theta}^*)$ and $\mathbf{a}_{\mathbf{h}i} = \left(\mathbf{0}_{q\times(nq)}, p^{-1}\mathbf{d}_{i*}\right)\mathrm{diag}(p^{-1}\mathbf{I}_{nq}, n^{-1}\mathbf{I}_{(pq+p)})\mathbf{f}(\boldsymbol{\theta}_0)$. By the compatible property of matrix norm, we have

$$\|\mathbf{r}_{i2}\|^2 \leqslant \left\|\mathbf{A}_*^{-1}\right\|^2 \cdot \left\|p\Phi_{i*}^{-1}\right\|^2 \|\mathbf{a}_{\mathbf{h}i}\|^2. \qquad (32)$$

Now, we prove

$$\left\|\mathbf{A}_*^{-1}\right\|^2 = O_p(1). \qquad (33)$$

For this, we only need to verify that for any $\epsilon > 0$, there exists $M > 0$ such that $P\left[\|\mathbf{A}_*^{-1}\| > M\right] \leqslant \epsilon$. Based on Condition (C4), we get

$$\begin{aligned}P\left[\left\|\mathbf{A}_*^{-1}\right\| > M\right] &\leqslant P\left[\left\|\mathbf{A}_*^{-1}\right\| > M, \sup_i\|\mathbf{h}_{i*} - \mathbf{h}_{i0}\| \leqslant M_2\right] + P\left[\sup_i\|\mathbf{h}_{i*} - \mathbf{h}_{i0}\| \geqslant M_2\right] \\ &\leqslant P\left[\left\|\mathbf{A}_*^{-1}\right\| > M, \sup_j\|\boldsymbol{\gamma}_{j*} - \boldsymbol{\gamma}_{j0}\| \leqslant M_1, \sup_i\|\mathbf{h}_{i*} - \mathbf{h}_{i0}\| \leqslant M_2\right] + \frac{2\epsilon}{3} \leqslant \epsilon,\end{aligned}$$

where the second inequality follows from Lemma 1, $\sup_i\|\mathbf{h}_{i*} - \mathbf{h}_{i0}\| \leqslant \sup_i\|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|$ and $\sup_j\|\boldsymbol{\gamma}_{j*} - \boldsymbol{\gamma}_{j0}\| \leqslant \sup_j\|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|$, the last inequality follows from Condition (C4) by taking

$M > 1/\lambda_2$. Hence, (33) is proved. Since for each i, $p^{-1}\Phi_{i*}$ is the submatrix of $\mathbf{A}_*$, we obtain

$$\sup_i \left\| p\Phi_{i*}^{-1} \right\|^2 = O_p(1). \tag{34}$$

Next, we consider the order of $\|\mathbf{a}_{\mathbf{h}i}\|^2$. It holds that

$$\|\mathbf{a}_{\mathbf{h}i}\|^2 \leqslant p^{-1}\sum_{j=1}^{p} \|\mathbf{G}_{ji*} + a_{ji*}\mathbf{E}\|^2 \cdot \left\| n^{-1}\sum_{\ell=1}^{n} \mathbf{w}_{j\ell 0} \right\|^2 .$$

Using the fact that $\|\mathbf{a}+\mathbf{b}\|^2 \leqslant 2\|\mathbf{a}\|^2 + 2\|\mathbf{b}\|^2$, we obtain

$$\begin{aligned}
\|\mathbf{a}_{\mathbf{h}i}\|^2 &\leqslant 2p^{-1}\sum_{j=1}^{p} \|\mathbf{G}_{ji*}\|^2 \cdot \left\| n^{-1}\sum_{\ell=1}^{n} \mathbf{w}_{j\ell 0} \right\|^2 \\
&+ 2p^{-1}\sum_{j=1}^{p} q[2a_{ji0}^2 + 2(m_{ji0} - m_{ji*})^2/c_j^2] \cdot \left\| n^{-1}\sum_{\ell=1}^{n} \mathbf{w}_{j\ell 0} \right\|^2 \\
&= 2p^{-1}\sum_{j=1}^{p} \|\mathbf{G}_{ji*}\|^2 \cdot \left\| n^{-1}\sum_{\ell=1}^{n} \mathbf{w}_{j\ell 0} \right\|^2 + 4p^{-1}\sum_{j=1}^{p} q a_{ji0}^2 \cdot \left\| n^{-1}\sum_{\ell=1}^{n} \mathbf{w}_{j\ell 0} \right\|^2 \\
&\quad +4p^{-1}\sum_{j=1}^{p} q[(m_{ji0} - m_{ji*})^2/c_j^2] \cdot \left\| n^{-1}\sum_{\ell=1}^{n} \mathbf{w}_{j\ell 0} \right\|^2 .
\end{aligned}$$

By (16), (17) and (18) in Lemma 2, we have

$$\|\mathbf{a}_{\mathbf{h}i}\|^2 = O_p(n^{-1}), \tag{35}$$

uniformly over i. By (32), (33), (34) and (35), we get

$$\|\mathbf{r}_{i2}\|^2 = O_p(n^{-1}) \text{ uniformly over } i. \tag{36}$$

Furthermore, by Condition (C3.1) and Markov inequality, we get $\|p^{-1}\sum_{j=1}^{p_1} \mathbf{v}_{ji}(\mathbf{h}_{i0}, \boldsymbol{\gamma}_{j0})\|^2 =$

$O_p(p^{-1})$, coupling with (34) and compatible property of matrix norm, we have

$$\|\mathbf{r}_{i1}\|^2 = \left\| p\Phi_{i*}^{-1} p^{-1} \sum_{j=1}^{p_1} \mathbf{v}_{ji}(\mathbf{h}_{i0}, \boldsymbol{\gamma}_{j0}) \right\|^2 = O_p(p^{-1}). \tag{37}$$

Combing with (31), (36) and (37), we conclude that

$$\|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(C_{np}^{-1}).$$

Now, we consider the order of $\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|$. Denote $\mathbf{v}_{ji0} = \mathbf{v}_{ji}(\mathbf{h}_{i0}, \boldsymbol{\gamma}_{j0})$. By the compatible property of matrix norm, we have

$$\sup_i \|\mathbf{r}_{i1}\|^2 \leqslant \sup_i \left\| \left(p^{-1}\Phi_{i*}\right)^{-1} \right\|^2 \sup_i \left\| p^{-1} \sum_{j=1}^{p_1} \mathbf{v}_{ji0} \right\|^2 .$$

By equation (33), we know $\sup_i \left\| (p^{-1}\Phi_{i*})^{-1} \right\|^2 = O_p(1)$. Denote $P_{\mathbf{h}} = P\left(\sup_i \left\| p^{-1} \sum_{j=1}^{p_1} \mathbf{v}_{ji0} \right\|^2 > \delta \right)$. By Bonferroni and Markov inequalities, we obtain

$$\begin{aligned} P_{\mathbf{h}} = P\left(\sup_i \|p^{-1} \sum_{j=1}^{p_1} \mathbf{v}_{ji0}\|^2 > \delta \right) &\leqslant n \sup_i P\left(\|p^{-1} \sum_{j=1}^{p_1} \mathbf{v}_{ji0}\|^{2\tau} > \delta^\tau \right) \\ \leqslant \quad \frac{nE\|p^{-1}\sum_{j=1}^{p_1} \mathbf{v}_{ji0}\|^{2\tau}}{\delta^\tau} &= O(\frac{n}{p^\tau \delta^\tau}), \end{aligned} \tag{38}$$

which implies $\sup_i \|p^{-1}\sum_{j=1}^{p_1} \mathbf{v}_{ji0}\|^2 = O_p(\frac{n^{1/\tau}}{p})$, where the second inequality follows from Condition (C3.1). Therefore, we obtain $\sup_i \|\mathbf{r}_{i1}\|^2 = O_p(\frac{n^{1/\tau}}{p})$. This coupling with (31) and (36), we have $\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|^2 = O_p(\frac{1}{n} + \frac{n^{1/\tau}}{p})$.

Step 2: Now we consider the rate of $\|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|$ when n and p are large enough. By simple algebra operation, we have

$$\begin{aligned}\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j &= \left(\sum_{i=1}^n \frac{m'_{ji*}}{c_j}\boldsymbol{\kappa}_{i*}\boldsymbol{\kappa}_{i*}^T\right)^{-1}\sum_{i=1}^n \mathbf{w}_{ji0} \\ &\quad - \left(\sum_{i=1}^n \frac{m'_{ji*}}{c_j}\boldsymbol{\kappa}_{i*}\boldsymbol{\kappa}_{i*}^T\right)^{-1}\left(\mathbf{u}_{j*}, \mathbf{0}_{(q+1)\times(pq+p)}\right)\nabla\mathbf{f}(\boldsymbol{\theta}^*)^{-1}\mathbf{f}(\boldsymbol{\theta}_0) \hat{=} \mathbf{s}_{j1} - \mathbf{s}_{j2}. \end{aligned} \tag{39}$$

It is obvious that

$$\|\mathbf{s}_{j2}\|^2 \leqslant \left\|\mathbf{A}_*^{-1}\right\|^2 \left\|\left(n^{-1}\mathbf{u}_{j*}, \mathbf{0}_{(q+1)\times(pq+p)}\right)\mathrm{diag}(p^{-1}\mathbf{I}_{nq}, n^{-1}\mathbf{I}_{(pq+p)})\mathbf{f}(\boldsymbol{\theta}_0)\right\|^2 \hat{=} \left\|\mathbf{A}_*^{-1}\right\|^2 \|\mathbf{a}_{\mathbf{b}j}\|^2.$$

Since

$$\begin{aligned}\|\mathbf{a}_{\mathbf{b}j}\|^2 &\leqslant n^{-1}\sum_{i=1}^n \left\|\mathbf{G}_{ji*}^T + a_{ji*}\mathbf{E}^T\right\|^2 \cdot \left\|p^{-1}\sum_{\ell=1}^p \mathbf{v}_{\ell i0}\right\|^2 \\ &\leqslant 2n^{-1}\sum_{i=1}^n \|\mathbf{G}_{ji*}\|^2 \cdot \left\|p^{-1}\sum_{\ell=1}^p \mathbf{v}_{\ell i0}\right\|^2 \\ &+ 2n^{-1}\sum_{i=1}^n q[2a_{ji0}^2 + 2(m_{ji0} - m_{ji*})^2/c_j^2] \cdot \left\|p^{-1}\sum_{\ell=1}^p \mathbf{v}_{\ell i0}\right\|^2 \\ &\leqslant 2n^{-1}\sum_{i=1}^n \|\mathbf{G}_{ji*}\|^2 \cdot \left\|p^{-1}\sum_{\ell=1}^p \mathbf{v}_{\ell i0}\right\|^2 + 4n^{-1}\sum_{i=1}^n \left\|a_{ji0}\cdot p^{-1}\sum_{\ell=1}^p \mathbf{v}_{\ell i0}\right\|^2 \\ &\quad +4n^{-1}\left\{(m_{ji0} - m_{ji*})^2/c_j^2\right\}\sum_{i=1}^n \left\|p^{-1}\sum_{\ell=1}^p \mathbf{v}_{\ell i0}\right\|^2. \end{aligned}$$

Following the same arguments in Lemma 2, we have

$$
\begin{aligned}
&\sup_j \left\{ n^{-1} \sum_{i=1}^n \|\mathbf{G}_{ji*}\|^2 \cdot \left\| p^{-1} \sum_{\ell=1}^p \mathbf{v}_{\ell i0} \right\|^2 \right\} = O_p(p^{-1}), \\
&\sup_j \left\{ n^{-1} \sum_{i=1}^n \left\| a_{ji0} \cdot p^{-1} \sum_{\ell=1}^p \mathbf{v}_{\ell i0} \right\|^2 \right\} = O_p(p^{-1}), \\
&\sup_j \left\{ n^{-1} \left\{ (m_{ji0} - m_{ji*})^2 / c_j^2 \right\} \sum_{i=1}^n \left\| p^{-1} \sum_{\ell=1}^p \mathbf{v}_{\ell i0} \right\|^2 \right\} = O_p(p^{-1}).
\end{aligned}
$$

Thus, $\|\mathbf{a}_{\mathbf{b}j}\|^2 = O_p(p^{-1})$ uniformly over j, coupling with (33), we have

$$
\sup_j \|\mathbf{s}_{j2}\|^2 = O_p(p^{-1}). \tag{40}
$$

Again by (33), we have $\left\| \left(\frac{1}{n} \sum_{i=1}^n \frac{m'_{ji*}}{c_j} \boldsymbol{\kappa}_{i*} \boldsymbol{\kappa}_{i*}^T \right)^{-1} \right\| = O_p(1)$. Coupling with $\|n^{-1} \sum_{i=1}^n \mathbf{w}_{ji0}\|^2 = O_p(\frac{1}{n})$ by the central limit Theorem, we get

$$
\|\mathbf{s}_{j1}\|^2 = O_p(n^{-1}). \tag{41}
$$

Then, by (39), (40) and (41), we conclude that for each j, $\|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(C_{np}^{-1})$.

Furthermore, following the same arguments of Lemma 4(c) in Fan et al. (2013), we have $\sup_j \|n^{-1} \sum_{i=1}^n \mathbf{w}_{ji0}\|^2 = O_p(\frac{\ln p}{n})$. Since for each j, $n^{-1} \sum_{i=1}^n \frac{m'_{ji*}}{c_j} \boldsymbol{\kappa}_{i*} \boldsymbol{\kappa}_{i*}^T$ is the submarix of $\mathbf{A}_*$, we obtain $\sup_j \left\| \left(n^{-1} \sum_{i=1}^n \frac{m'_{ji*}}{c_j} \boldsymbol{\kappa}_{i*} \boldsymbol{\kappa}_{i*}^T \right)^{-1} \right\| \leqslant \|\mathbf{A}_*^{-1}\|$. By (33), we obtain $\sup_j \|\mathbf{s}_{j1}\|^2 \leqslant \sup_j \left\| \left(n^{-1} \sum_{i=1}^n \frac{m'_{ji*}}{c_j} \boldsymbol{\kappa}_{i*} \boldsymbol{\kappa}_{i*}^T \right)^{-1} \right\|^2 \sup_j \|n^{-1} \sum_{i=1}^n \mathbf{w}_{ji0}\|^2 = O_p(\frac{\ln p}{n})$. Thus, coupling with (40), we have $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|^2 = O_p(\frac{1}{p} + \frac{\ln p}{n})$. □

Proofs of Theorems 1 and 2

Since the proofs of Theorem 1 and Theorem 2 are similar, we only give the proof of Theorem 1 here.

Proof of Theorem 1. Denote $\tilde{a}_{ji}=\frac{x_{ij}-m_j(\tilde{\boldsymbol{\gamma}}_j^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}$. By (5) in Subsection 2.2.2, we have

$$\hat{\mathbf{h}}_i=\tilde{\mathbf{h}}_i+\left\{\frac{1}{p}\sum_{j=1}^{p}\frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\mathbf{b}}_j\tilde{\mathbf{b}}_j^T\right\}^{-1}\frac{1}{p}\sum_{j=1}^{p}\left(\tilde{a}_{ji}\tilde{\mathbf{b}}_j\right)\hat{=}\tilde{\mathbf{h}}_i+\tilde{\Phi}_i^{-1}\frac{1}{p}\sum_{j=1}^{p}\left(\tilde{a}_{ji}\tilde{\mathbf{b}}_j\right),$$

where $\tilde{\boldsymbol{\kappa}}_i=(1,\tilde{\mathbf{h}}_i^T)^T$. The proof of Theorem 1 contains two steps. In the first step, we prove $\|\tilde{\Phi}_i^{-1}\|=O_p(1)$ and $\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j)\|=O_p(C_{np}^{-1})$. This combing with Lemma 3, we obtain $\|\widehat{\mathbf{h}}_i-\mathbf{h}_{i0}\|=O_p(C_{np}^{-1})$. In the second step, we prove $\sup_i\|\tilde{\Phi}_i^{-1}\|=O_p(1)$ and $\sup_i\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j)\|=O_p(\frac{1}{n}+\frac{n^{1/\tau}}{p})$. Then, again by Lemma 3, we get $\sup_i\|\widehat{\mathbf{h}}_i-\mathbf{h}_{i0}\|=O_p(\frac{1}{n}+\frac{n^{1/\tau}}{p})$.

Step 1: For any $\epsilon>0$, there exists $M>0$ such that

$$\begin{aligned}P\left(\|\tilde{\Phi}_i^{-1}\|>M\right)&\leqslant P\left(\|\Phi_{i0}^{-1}\|+\|\tilde{\Phi}_i^{-1}-\Phi_{i0}^{-1}\|>M\right)\\&\leqslant P\left(\|\Phi_{i0}^{-1}\|>M/2\right)+P\left(\|\tilde{\Phi}_i^{-1}-\Phi_{i0}^{-1}\|>M/2\right)\\&\leqslant P\left(\|\lambda_{\min}\big(-\nabla_s\mathbf{f}(\boldsymbol{\theta}_0)\big)^{-1}\|>M/2\right)+P\left(\|\tilde{\Phi}_i^{-1}-\Phi_{i0}^{-1}\|>M/2\right)\leqslant\epsilon,\end{aligned}$$

thus, $\|\tilde{\Phi}_i^{-1}\|=O_p(1)$, where $\Phi_{i0}=\frac{1}{p}\sum_{j=1}^p\frac{m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\mathbf{b}_{j0}\mathbf{b}_{j0}^T$ and the last inequality follows from Condition (C4) and Lemma 3 on the convergent rate of $\tilde{\mathbf{h}}_i$ and $\tilde{\boldsymbol{\gamma}}_j$.

Now we consider $\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j)\|$. Noting that $\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j)\|\leqslant\|\frac{1}{p}\sum_{j=1}^p(a_{ji0}\mathbf{b}_{j0})\|+\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j-a_{ji0}\mathbf{b}_{j0})\|$. By Condition (C3.1) and Markov inequality, we obtain that $\|p^{-1}\sum_{j=1}^p(a_{ji0}\mathbf{b}_{j0})\|^2=O_p(p^{-1})$. By the results that $\|\tilde{\mathbf{h}}_i-\mathbf{h}_{i0}\|=O_p(C_{np}^{-1})$ and $\|\tilde{\boldsymbol{\gamma}}_j-\boldsymbol{\gamma}_{j0}\|=O_p(C_{np}^{-1})$ from Lemma 3, we obtain $\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j-a_{ji0}\mathbf{b}_{j0})\|=O_p(C_{np}^{-1})$. Thus, we get $\|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j)\|=O_p(C_{np}^{-1})$.

Step 2: For any $\epsilon>0$, there exists $M>0$ such that

$$\begin{aligned}P\left(\sup_i\|\tilde{\Phi}_i^{-1}\|>M\right)&\leqslant P\left(\sup_i\|\Phi_{i0}^{-1}\|+\sup_i\|\tilde{\Phi}_i^{-1}-\Phi_{i0}^{-1}\|>M\right)\\&\leqslant P\left(\sup_i\|\Phi_{i0}^{-1}\|>M/2\right)+P\left(\sup_i\|\tilde{\Phi}_i^{-1}-\Phi_{i0}^{-1}\|>M/2\right)\\&\leqslant P\left(\|\lambda_{\min}\big(-\nabla_s\mathbf{f}(\boldsymbol{\theta}_0)\big)^{-1}\|>M/2\right)+P\left(\sup_i\|\tilde{\Phi}_i^{-1}-\Phi_{i0}^{-1}\|>M/2\right)\\&\leqslant\epsilon,\end{aligned}$$

where the last inequality follows from Condition (C4) and the uniform convergent rate of

$\tilde{\mathbf{h}}_i$. Thus, $\sup_i \|\tilde{\Phi}_i^{-1}\| = O_p(1)$.

Noting that $\sup_i \|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j)\| \leqslant \sup_i \|\frac{1}{p}\sum_{j=1}^p(a_{ji0}\mathbf{b}_{j0})\| + \sup_i \|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j - a_{ji0}\mathbf{b}_{j0})\|$. By (38), we have $\sup_i \|\frac{1}{p}\sum_{j=1}^p(a_{ji0}\mathbf{b}_{j0})\| = O_p(\frac{n^{1/\tau}}{p})$. Furthermore, by the continuousness of $m_j''(\cdot)$ of Condition (C5) and the uniform convergent rate of $\tilde{\mathbf{h}}_i$ from Lemma 3, we obtain $\sup_i \|\frac{1}{p}\sum_{j=1}^p(\tilde{a}_{ji}\tilde{\mathbf{b}}_j - a_{ji0}\mathbf{b}_{j0})\| = O_p(\frac{1}{n} + \frac{n^{1/\tau}}{p})$. Thus, we conclude that $\sup_i \|\widehat{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(\frac{1}{n} + \frac{n^{1/\tau}}{p})$. □

D: Proof of Theorem 3

By (6) in Subsection 2.2.2, we have

$$\hat{\boldsymbol{\gamma}}_j = \tilde{\boldsymbol{\gamma}}_j + \left\{\frac{1}{n}\sum_{i=1}^n \frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\boldsymbol{\kappa}}_i\tilde{\boldsymbol{\kappa}}_i^T\right\}^{-1}\frac{1}{n}\sum_{i=1}^n(\tilde{a}_{ji}\tilde{\boldsymbol{\kappa}}_i). \tag{42}$$

Reorganizing (42), we obtain

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) = \sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) + \left\{\frac{1}{n}\sum_{i=1}^n \frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\boldsymbol{\kappa}}_i\tilde{\boldsymbol{\kappa}}_i^T\right\}^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^n(\tilde{a}_{ji}\tilde{\boldsymbol{\kappa}}_i). \tag{43}$$

A Taylor-series expansion of $\sqrt{n}\frac{1}{n}\sum_{i=1}^n(\tilde{a}_{ji}\tilde{\boldsymbol{\kappa}}_i)$ around $\boldsymbol{\gamma}_{j0}$ gives

$$\frac{1}{n^{1/2}}\sum_{i=1}^n \frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\boldsymbol{\kappa}}_i - \left\{\frac{1}{n}\sum_{i=1}^n \frac{m_j'(\boldsymbol{\gamma}_{j*}^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\boldsymbol{\kappa}}_i\tilde{\boldsymbol{\kappa}}_i^T\right\}\sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) \tag{44}$$

By (43) and (44), we have

$$
\begin{aligned}
& \sqrt{n}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) & (45)\\
= \; & \sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) + \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T \right\}^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^{n} (\tilde{a}_{ji} \tilde{\boldsymbol{\kappa}}_i) & \\
= \; & \sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) + \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T \right\}^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i & \\
& - \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_j'(\boldsymbol{\gamma}_{j*}^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T \right\} \sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}). & (46)
\end{aligned}
$$

Moreover, completely similar to the proof of $\sup_{i,j} |G_{ij*} - G_{ij0}| = o_p(1)$ in Step 2.1 of Lemma 3, we have

$$
\begin{aligned}
& \sup_{i,j} |c_j^{-1} m_j'(\boldsymbol{\gamma}_{j*}^T \tilde{\boldsymbol{\kappa}}_i) \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T - c_j^{-1} m_{ji0}' \boldsymbol{\kappa}_{i0} \boldsymbol{\kappa}_{i0}^T| = o_p(1), \\
& \sup_{i,j} |c_j^{-1} m_j'(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i) \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T - c_j^{-1} m_{ji0}' \boldsymbol{\kappa}_{i0} \boldsymbol{\kappa}_{i0}^T| = o_p(1).
\end{aligned}
$$

That is,

$$
\begin{aligned}
& n^{-1} \sum_{i=1}^{n} c_j^{-1} m_j'(\boldsymbol{\gamma}_{j*}^T \tilde{\boldsymbol{\kappa}}_i) \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T - n^{-1} \sum_{i=1}^{n} c_j^{-1} m_{ji0}' \boldsymbol{\kappa}_{i0} \boldsymbol{\kappa}_{i0}^T = o_p(1), \\
& n^{-1} \sum_{i=1}^{n} c_j^{-1} m_j'(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i) \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T - n^{-1} \sum_{i=1}^{n} c_j^{-1} m_{ji0}' \boldsymbol{\kappa}_{i0} \boldsymbol{\kappa}_{i0}^T = o_p(1). \qquad (47)
\end{aligned}
$$

By (46)-(47), we have

$$
\begin{aligned}
& \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_j'(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_j'(\boldsymbol{\gamma}_{j*}^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i \tilde{\boldsymbol{\kappa}}_i^T \right\} \sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) \\
= \; & \sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0})(1 + o_p(1)).
\end{aligned}
$$

Thus, we have

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) = o_p(1)\sqrt{n}(\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}) + \left\{\frac{1}{n}\sum_{i=1}^{n}\frac{m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\boldsymbol{\kappa}_{i0}\boldsymbol{\kappa}_{i0}^T + o_p(1)\right\}^{-1} \times \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\boldsymbol{\kappa}}_i. \tag{48}$$

Since

$$\begin{aligned}
& m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i) - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0}) \\
= \; & m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*})\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_{i0} - \boldsymbol{\kappa}_{i0}) \\
= \; & m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_{i0} - \boldsymbol{\kappa}_{i0}) + [m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*}) - m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})]\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_{i0} - \boldsymbol{\kappa}_{i0}),
\end{aligned}$$

we have

$$\begin{aligned}
& \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i)}{c_j}\tilde{\boldsymbol{\kappa}}_i \\
= \; & \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\tilde{\boldsymbol{\kappa}}_i - \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i) - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\tilde{\boldsymbol{\kappa}}_i \\
= \; & \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\boldsymbol{\kappa}_{i0} + \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i) - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\boldsymbol{\kappa}_{i0} - \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{m_j(\boldsymbol{\gamma}_{j0}^T\tilde{\boldsymbol{\kappa}}_i) - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
= \; & \frac{1}{n^{1/2}}\sum_{i=1}^{n}a_{ji0}\boldsymbol{\kappa}_{i0} + \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\boldsymbol{\kappa}_{i0}\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{[m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*}) - m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})]}{c_j}\boldsymbol{\kappa}_{i0}\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0})\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{[m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*}) - m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})]}{c_j}(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0})\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0})
\end{aligned} \tag{49}$$

Because $\mathbf{h}_1, \cdots, \mathbf{h}_n$ are independent and identically distributed, and $\mathbf{h}_i$ has the finite second moment $\mathrm{E}\mathbf{h}_i\mathbf{h}_i^T = \mathbf{I}_q$ with $\mathrm{E}\mathbf{h}_i = \mathbf{0}_q$, then we have

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n} a_{ji0}\boldsymbol{\kappa}_{i0} \to N(\mathbf{0}_p, \boldsymbol{\Gamma}_j), \tag{50}$$

where $\boldsymbol{\Gamma}_j = var(a_{ji0}\boldsymbol{\kappa}_{i0})$. In addition, we have

$$\begin{aligned}
& \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij}-m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}(\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \\
= \ & \frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{x_{ij}-m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}(\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}) \\
= \ & \frac{1}{n^{1/2}}\sum_{i=1}^{n} a_{ji0}\Phi_{i*}^{-1}\sum_{\ell=1}^{p_1}\mathbf{v}_{\ell i0}(\mathbf{h}_{i0}, \boldsymbol{\gamma}_{\ell 0}) \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n} a_{ji0}\Phi_{i*}^{-1}\left(\mathbf{0}_{q\times(nq)}, \mathbf{d}_{i*}\right)\nabla\mathbf{f}(\boldsymbol{\theta}^*)^{-1}\mathbf{f}(\boldsymbol{\theta}_0) \\
= \ & \frac{1}{n^{1/2}}\sum_{i=1}^{n} a_{ji0}\Phi_{i*}^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell \\
& -\frac{1}{n^{1/2}}\sum_{i=1}^{n}\Phi_{i*}^{-1}\left(\mathbf{0}_{q\times(nq)}, \mathbf{d}_{i*}\right)\nabla\mathbf{f}(\boldsymbol{\theta}^*)^{-1}\mathbf{f}(\boldsymbol{\theta}_0)a_{ji0}
\end{aligned} \tag{51}$$

where $a_{ji0} = \frac{x_{ij}-m_j(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}$ and the second equality is from (31). Similar to (47), we have

$$\sup_i \|p_1^{-1}\sum_{j=1}^{p_1}\frac{m'_{ji*}}{c_j}\mathbf{b}_{j*}\mathbf{b}_{j*}^T - p_1^{-1}\sum_{j=1}^{p_1}\frac{m'_{ji}}{c_j}\mathbf{b}_j\mathbf{b}_j^T\| = o_p(1), \tag{52}$$

where $\Phi_{i*} = \sum_{j=1}^{p_1} \frac{m'_{ji*}}{c_j} \mathbf{b}_{j*} \mathbf{b}_{j*}^T$ and $\Phi_{i0} = \sum_{j=1}^{p_1} \frac{m'_{ji}}{c_j} \mathbf{b}_j \mathbf{b}_j^T$. By (51) and (52), we have

$$
\begin{aligned}
& \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T \boldsymbol{\kappa}_{i0})}{c_j} (\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) \right\| \\
= & \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} a_{ji0} \Phi_{i0}^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0} \mathbf{b}_\ell \right\| + o_p(1) \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\| a_{ji0} p_1^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0} \mathbf{b}_\ell \right\| \\
& + \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} \Phi_{i*}^{-1} \left(\mathbf{0}_{q\times(nq)}, \mathbf{d}_{i*} \right) \nabla \mathbf{f}(\boldsymbol{\theta}^*)^{-1} \mathbf{f}(\boldsymbol{\theta}_0) a_{ji0} \right\|. \qquad (53)
\end{aligned}
$$

If $n^{1/2} p_1^{-1} \to 0$, we have

$$
\mathrm{E}(n^{-1/2} \sum_{i=1}^{n} a_{ji0} p_1 \Phi_{i0}^{-1} p_1^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0} \mathbf{b}_\ell) = n^{1/2} p_1^{-1} \mathrm{E}(a_{ji0}^2 p_1 \Phi_{i0}^{-1} \mathbf{b}_j) \to 0, \qquad (54)
$$

and

$$
\begin{aligned}
& \mathrm{Var}(n^{-1/2} \sum_{i=1}^{n} a_{ji0} p_1 \Phi_{i0}^{-1} p_1^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0} \mathbf{b}_\ell) \\
= & \mathrm{Var}(a_{ji0} p_1 \Phi_{i0}^{-1} p_1^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0} \mathbf{b}_\ell) \\
\leqslant & \mathrm{E}[(a_{ji0} p_1 \Phi_{i0}^{-1} p_1^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0} \mathbf{b}_\ell)^2] \\
= & \mathrm{E}(a_{ji0}^2 p_1 \Phi_{i0}^{-1} p_1^{-2} \sum_{\ell_1, \ell_2} a_{\ell_1 i0} a_{\ell_2 i0} \mathbf{b}_{\ell_1} \mathbf{b}_{\ell_2} p_1 \Phi_{i0}^{-1}) \\
= & \mathrm{E}(a_{ji0}^2 p_1 \Phi_{i0}^{-1} p_1^{-2} \sum_{\ell=1}^{p_1} a_{\ell i0}^2 \mathbf{b}_\ell \mathbf{b}_\ell p_1 \Phi_{i0}^{-1}) \\
& + \mathrm{E}(a_{ji0}^2 p_1 \Phi_{i0}^{-1} p_1^{-2} \sum_{\ell_1 \neq \ell_2} a_{\ell_1 i0} a_{\ell_2 i0} \mathbf{b}_{\ell_1} \mathbf{b}_{\ell_2} p_1 \Phi_{i0}^{-1}) \\
= & p_1^{-1} \mathrm{E}(a_{ji0}^2 p_1 \Phi_{i0}^{-1} p_1^{-1} \sum_{\ell=1}^{p_1} a_{\ell i0}^2 \mathbf{b}_\ell \mathbf{b}_\ell p_1 \Phi_{i0}^{-1}) \to 0. \qquad (55)
\end{aligned}
$$

By (54) and (55), we have

$$n^{-1/2}\sum_{i=1}^{n} a_{ji0}p_1\Phi_{i0}^{-1}p_1^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell = o_p(1). \tag{56}$$

Then, we consider

$$\begin{aligned}
& \mathrm{E}(n^{-1/2}\sum_{i=1}^{n}\left\| a_{ji0}p_1^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell \right\|) \\
= \;& \mathrm{E}(n^{1/2}\left\| a_{ji0}p_1^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell \right\|) \\
\leqslant \;& n^{1/2}\sqrt{\mathrm{E}(a_{ji0}^2 p_1^{-2}\sum_{\ell_1,\ell_2} a_{\ell_1 i0}a_{\ell_2 i0}\mathbf{b}_{\ell_1}^T\mathbf{b}_{\ell_2})} \\
\leqslant \;& n^{1/2}\sqrt{\mathrm{E}(a_{ji0}^2 p_1^{-2}\sum_{\ell=1}^{p_1} a_{\ell i0}^2\mathbf{b}_\ell^T\mathbf{b}_\ell)} = O(1).
\end{aligned}$$

Similarly, we have

$$\mathrm{Var}(n^{-1/2}\sum_{i=1}^{n}\left\| a_{ji0}p_1^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell \right\|) = o(1).$$

Then we have

$$n^{-1/2}\sum_{i=1}^{n}\left\| a_{ji0}p_1^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell \right\| = O_p(1).$$

That is,

$$o_p(1)n^{-1/2}\sum_{i=1}^{n}\left\| a_{ji0}p_1^{-1}\sum_{\ell=1}^{p_1} a_{\ell i0}\mathbf{b}_\ell \right\| = o_p(1). \tag{57}$$

We have

$$
\begin{aligned}
& \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} \Phi_{i*}^{-1} \left(\mathbf{0}_{q\times(nq)}, \mathbf{d}_{i*} \right) \nabla \mathbf{f}(\boldsymbol{\theta}^*)^{-1} \mathbf{f}(\boldsymbol{\theta}_0) a_{ji0} \right\| \\
= & \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} p \Phi_{i*}^{-1} \mathbf{A}_*^{-1} \mathbf{f}(\boldsymbol{\theta}_0) a_{ji0} \right\| \\
= & \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} p \Phi_{i0}^{-1} \mathbf{A}_0^{-1} \mathbf{f}(\boldsymbol{\theta}_0) a_{ji0} \right\| + o_p(1) \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\| f(\boldsymbol{\theta}_0) a_{ji0} \right\| .
\end{aligned}
$$

Following the proof of (56), we have

$$
\frac{1}{n^{1/2}} \sum_{i=1}^{n} p \Phi_{i0}^{-1} \mathbf{A}_0^{-1} \mathbf{f}(\boldsymbol{\theta}_0) a_{ji0} = o_p(1). \tag{58}
$$

Following the proof of (57)

$$
o_p(1) \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\| f(\boldsymbol{\theta}_0) a_{ji0} \right\| = o_p(1). \tag{59}
$$

From (51)-(56)-(57)-(58)-(59), we have

$$
\frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{x_{ij} - m_j(\boldsymbol{\gamma}_{j0}^T \boldsymbol{\kappa}_{i0})}{c_j} (\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) = o_p(1). \tag{60}
$$

Following the proof of (60) we have

$$
\frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{m_j'(\boldsymbol{\gamma}_{j0}^T \boldsymbol{\kappa}_{i0})}{c_j} \boldsymbol{\kappa}_{i0} \boldsymbol{\gamma}_{j0}^T (\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}) = o_p(1). \tag{61}
$$

Similar to the proof of (24), we have $\sup_{i,j}\left\|m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*})-m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})\right\|=o_p(1)$. Then we have

$$\begin{aligned}&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{[m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*})-m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})]}{c_j}\boldsymbol{\kappa}_{i0}\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_{i0}-\boldsymbol{\kappa}_{i0})\\ \leqslant\;&\sup_{i,j}\left|m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*})-m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})\right|\cdot n^{-1}\sum_{i=1}^{n}\frac{|\boldsymbol{\kappa}_{i0}\boldsymbol{\gamma}_{j0}^T|}{|c_j|}\cdot\sup_i\|\tilde{\mathbf{h}}_{i0}-\mathbf{h}_{i0}\|n^{1/2}=o_p(1),\end{aligned}\tag{62}$$

where $n^{-1}\sum_{i=1}^{n}|\boldsymbol{\kappa}_{i0}\boldsymbol{\gamma}_{j0}^T|$ converges to $\mathrm{E}(|\boldsymbol{\kappa}_{i0}\boldsymbol{\gamma}_{j0}^T|)$ in probability. Similarly, we have

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{[m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0*})-m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})]}{c_j}(\tilde{\boldsymbol{\kappa}}_{i}-\boldsymbol{\kappa}_{i0})\boldsymbol{\gamma}_{j0}^T(\tilde{\boldsymbol{\kappa}}_{i0}-\boldsymbol{\kappa}_{i0})=o_p(1).\tag{63}$$

Combining (48)-(49)-(50)-(60)-(61)-(62)-(63) and noting $E\left(\frac{m_j'(\boldsymbol{\gamma}_{j0}^T\boldsymbol{\kappa}_{i0})}{c_j}\boldsymbol{\kappa}_{i0}\boldsymbol{\kappa}_{i0}^T\right)=$ $var(a_{ji0}\boldsymbol{\kappa}_i)=\Gamma_j$ in likelihood framework, we have

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_j-\boldsymbol{\gamma}_{j0})\rightarrow N(\mathbf{0}_q,\Gamma_j^{-1}).$$

E: Proof of Theorem 4

Proof. For the simplicity and without loss of generality, we suppose the model doesn't include intercepts. We prove Theorem 4 in two steps, including the cases of $k>q$ and $k<q$.

Step 1. We first consider the case of $k>q$. Without loss of generality, let $\mathbf{h}_{i0}$ and $\mathbf{b}_{j0}$ denote the k-dimensional vector $(\mathbf{h}_{i0}^T,\mathbf{0}_{k-q}^T)^T$ and $(\mathbf{b}_{j0}^T,\mathbf{0}_{k-q}^T)^T$, respectively. Let $\hat{\mathbf{h}}_i^{(q)}$ and $\hat{\mathbf{b}}_j^{(q)}$ denote the k-dimensional vector $(\hat{\mathbf{h}}_i^T,\mathbf{0}_{k-q}^T)$ and $(\hat{\mathbf{b}}_j^T,\mathbf{0}_{k-q}^T)$, respectively. For the clarity of expression, we denote $-(np)^{-1}l(\mathbf{H},\Upsilon)$ on $\boldsymbol{\theta}$ in (7) of Subsection 2.3 as $L(\boldsymbol{\theta},k)$ for emphasizing that it is related to k. We have

$$
\begin{aligned}
L(\hat{\boldsymbol{\theta}}^{(k)}, k) - L(\boldsymbol{\theta}_0, q) = & -(np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(k)} - \mathbf{b}_{j0})^T (\hat{\mathbf{h}}_i^{(k)} - \mathbf{h}_{i0})}{c_j} \\
& -(np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))\mathbf{b}_{j0}^T (\hat{\mathbf{h}}_i^{(k)} - \mathbf{h}_{i0})}{c_j} \\
& -(np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(k)} - \mathbf{b}_{j0})^T \mathbf{h}_{i0}}{c_j} \\
& + \frac{1}{2}(np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{m_j'((\hat{\mathbf{b}}_j^{(*k)})^T \hat{\mathbf{h}}_i^{(*k)})}{c_j} ((\hat{\mathbf{b}}_j^{(k)})^T \hat{\mathbf{h}}_i^{(k)} - \mathbf{b}_{j0}^T \mathbf{h}_{i0})^2 \\
& \hat{=} - I_1 - I_2 - I_3 + I_4,
\end{aligned} \tag{64}
$$

where $\hat{\mathbf{b}}_j^{(*k)}$ lies between $\hat{\mathbf{b}}_j^{(k)}$ and $\mathbf{b}_{j0}$ and $\hat{\mathbf{h}}_i^{(*k)}$ lies between $\hat{\mathbf{h}}_i^{(k)}$ and $\mathbf{h}_{i0}$.

Then, for I_1, we have

$$
\begin{aligned}
|I_1| &= \left| (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(k)} - \mathbf{b}_{j0})^T (\hat{\mathbf{h}}_i^{(k)} - \mathbf{h}_{i0})}{c_j} \right| \\
&\leqslant \{\frac{1}{n} \sum_{i=1}^{n} \|\hat{\mathbf{h}}_i^{(k)} - \mathbf{h}_{i0}\|^2\}^{1/2} \frac{1}{\sqrt{p}} \{\frac{1}{np} \sum_{i=1}^{n} \| \sum_{j=1}^{p} a_{ji} (\hat{\mathbf{b}}_j^{(k)} - \mathbf{b}_{j0})\|^2\}^{1/2} \\
&\leqslant O_p(C_{np}^{-1}) \cdot \frac{1}{\sqrt{p}} \cdot \{\sup_j \|\hat{\mathbf{b}}_j^{(k)} - \mathbf{b}_{j0}\|^2\}^{1/2} \cdot \{\frac{1}{n} \sum_{i=1}^{n} (\frac{1}{p} \sum_{j=1}^{p} a_{ji})^2\}^{1/2} = o_p(C_{np}^{-2}),
\end{aligned} \tag{65}
$$

where the second line is followed by Cauchy-Schwarz inequality and the third line is followed by Theorems 1-2 and the condition that the fourth moment of $p^{-1} \sum_{j=1}^{p} a_{ji}$ is bounded. For I_2, we have

$$
\begin{aligned}
|I_2| &= \left| (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))\mathbf{b}_{j0}^T (\hat{\mathbf{h}}_i^{(k)} - \mathbf{h}_{i0})}{c_j} \right| \\
&\leqslant \{\frac{1}{n} \sum_{i=1}^{n} \|\hat{\mathbf{h}}_i^{(k)} - \mathbf{h}_{i0}\|^2\}^{1/2} \frac{1}{\sqrt{p}} \{\frac{1}{np} \sum_{i=1}^{n} \| \sum_{j=1}^{p} \mathbf{v}_{ji}(\mathbf{h}_{i0}, \mathbf{b}_{j0})\|^2\}^{1/2} \\
&\leqslant O_p(C_{np}^{-1}) \cdot \frac{1}{\sqrt{p}} o_p(1) = o_p(C_{np}^{-2}),
\end{aligned} \tag{66}
$$

where the second line is followed by Cauchy-Schwarz inequality and the third line is followed by Theorem 1 and Condition (C3.4). For I_3, we have

$$\begin{aligned} |I_3| &= \left|(np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(x_{ij}-m_j(\mathbf{b}_{j0}^T\mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(k)}-\mathbf{b}_{j0})^T\mathbf{h}_{i0}}{c_j}\right| \\ &\leqslant \{\frac{1}{p}\sum_{j=1}^{p}\|\hat{\mathbf{b}}_j^{(k)}-\mathbf{b}_{j0}\|^2\}^{1/2}\frac{1}{\sqrt{n}}\{\frac{1}{np}\sum_{j=1}^{p}\|\sum_{i=1}^{n}\mathbf{w}_{ji}(\mathbf{h}_{i0},\mathbf{b}_{j0})\|^2\}^{1/2} \\ &\leqslant O_p(C_{np}^{-1})\cdot\frac{1}{\sqrt{n}}o_p(1)=o_p(C_{np}^{-2}), \end{aligned} \tag{67}$$

where the second line is followed by Cauchy-Schwarz inequality and the third line is followed by Theorem 2 and the condition that the second moment of $p^{-1}\sum_{j=1}^{p}a_{ji}$ is bounded. Similarly, We obtain $I_4 = O_p(C_{np}^{-2})$. Combining (64), (65), (66) and (67), we have

$$L(\hat{\boldsymbol{\theta}}^{(k)},k)-L(\boldsymbol{\theta}_0,q)=O_p(C_{np}^{-2}). \tag{68}$$

Similarly, we can get

$$\begin{aligned} L(\hat{\boldsymbol{\theta}}^{(q)},q)-L(\boldsymbol{\theta}_0,q) = &-(np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(x_{ij}-m_j(\mathbf{b}_{j0}^T\mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(q)}-\mathbf{b}_{j0})^T(\hat{\mathbf{h}}_i^{(q)}-\mathbf{h}_{i0})}{c_j} \\ &-(np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(x_{ij}-m_j(\mathbf{b}_{j0}^T\mathbf{h}_{i0}))\mathbf{b}_{j0}^T(\hat{\mathbf{h}}_i^{(q)}-\mathbf{h}_{i0})}{c_j} \\ &-(np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(x_{ij}-m_j(\mathbf{b}_{j0}^T\mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(q)}-\mathbf{b}_{j0})^T\mathbf{h}_{i0}}{c_j} \\ &+(np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{m_j'((\hat{\mathbf{b}}_j^{(*q)})^T\hat{\mathbf{h}}_i^{(*q)})}{c_j}((\hat{\mathbf{b}}_j^{(q)})^T\hat{\mathbf{h}}_i^{(q)}-\mathbf{b}_{j0}^T\mathbf{h}_{i0})^2 \\ \hat{=}\ &-J_1-J_2-J_3+J_4, \end{aligned} \tag{69}$$

where $\hat{\mathbf{b}}_j^{(*q)}$ lies between $\hat{\mathbf{b}}_j^{(q)}$ and $\mathbf{b}_{j0}$ and and $\hat{\mathbf{h}}_i^{(*q)}$ lies between $\hat{\mathbf{h}}_i^{(q)}$ and $\mathbf{h}_{i0}$.

Similar with the discussions on I_1, I_2, I_3 and I_4, we have

$$J_1 = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(q)} - \mathbf{b}_{j0})^T (\hat{\mathbf{h}}_i^{(q)} - \mathbf{h}_{i0})}{c_j} = o_p(C_{np}^{-2}), \tag{70}$$

$$J_2 = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))\mathbf{b}_{j0}^T (\hat{\mathbf{h}}_i^{(q)} - \mathbf{h}_{i0})}{c_j} = o_p(C_{np}^{-2}), \tag{71}$$

$$J_3 = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(x_{ij} - m_j(\mathbf{b}_{j0}^T \mathbf{h}_{i0}))(\hat{\mathbf{b}}_j^{(q)} - \mathbf{b}_{j0})^T \mathbf{h}_{i0}}{c_j} = o_p(C_{np}^{-2}), \tag{72}$$

$$J_4 = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{m_j'((\hat{\mathbf{b}}_j^{(*q)})^T \hat{\mathbf{h}}_i^{(*q)})}{c_j} ((\hat{\mathbf{b}}_j^{(q)})^T \hat{\mathbf{h}}_i^{(q)} - \mathbf{b}_{j0}^T \mathbf{h}_{i0})^2 = O_p(C_{np}^{-2}). \tag{73}$$

By (69), (70), (71) (72) and (73), we obtain

$$L(\hat{\boldsymbol{\theta}}^{(q)}, q) - L(\boldsymbol{\theta}_0, q) = O_p(C_{np}^{-2}). \tag{74}$$

By (68) and (74), we obtain $C_{np}^2(L(\hat{\boldsymbol{\theta}}^{(k)}, k) - L(\hat{\boldsymbol{\theta}}^{(q)}, q)) = O_p(1)$. If $k > q$, then we have $(k-q)g(n,p) > g(n,p)$. Since $C_{np}^2 g(n,p) \to +\infty$, then we have $P([L(\hat{\boldsymbol{\theta}}^{(k)}, k) + kg(n,p)] - [L(\hat{\boldsymbol{\theta}}^{(q)}, q) + qg(n,p)] > 0) \to 1$, as n and p tend to ∞.

Step 2. We consider the case of $k < q$ and then we have $(k-q)g(n,p) \to 0$. Without loss of generality and confusion, let $\hat{\mathbf{h}}_i^{(k)}$ and $\hat{\mathbf{b}}_j^{(k)}$ denote the q-dimensional vector $(\hat{\mathbf{h}}_i, \mathbf{0}_{q-k})$ and $(\hat{\mathbf{b}}_j, \mathbf{0}_{q-k})$, respectively. We have

$$\begin{aligned} L(\hat{\boldsymbol{\theta}}^{(k)}, k) - L(\hat{\boldsymbol{\theta}}^{(q)}, q) &= (\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)})^T \left. \frac{\partial L(\boldsymbol{\theta}, q)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(q)}} + \frac{1}{2}(\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)})^T \left. \frac{\partial^2 L(\boldsymbol{\theta}, q)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(*k)}} (\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)}) \\ &= \frac{1}{2}(\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)})^T \left. \frac{\partial^2 L(\boldsymbol{\theta}, q)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(*k)}} (\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)}), \end{aligned}$$

where $\hat{\boldsymbol{\theta}}^{(*k)}$ lies between $\hat{\boldsymbol{\theta}}^{(k)}$ and $\hat{\boldsymbol{\theta}}^{(q)}$. Since there exists a $j \in \{1, ..., p\}$ satisfying $\|\hat{\mathbf{b}}_j^{(k)} - \hat{\mathbf{b}}_j^{(q)}\| \geqslant c$ where c is a positive constant, then we have $L(\hat{\boldsymbol{\theta}}^{(k)}, k) - L(\hat{\boldsymbol{\theta}}^{(q)}, q) = \frac{1}{2}(\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)})^T \left. \frac{\partial^2 L(\boldsymbol{\theta}, q)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(*k)}} (\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(q)}) > c_0$, where c_0 is a constant. Then we have $P([L(\hat{\boldsymbol{\theta}}^{(k)}, k) + kg(n,p)] - [L(\hat{\boldsymbol{\theta}}^{(q)}, q) + qg(n,p)] > 0) \to 1$. Thus, we complete the proof of Theorem 4. $\square$

References

David G. Luenberger, Y. Y. (2016). *Linear and Nonlinear Programming*. Springer International Publishing.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.