Supplement to “ Exploratory Hierarchical Factor Analysis with an Application to Psychological Measurement”

In the supplement, we provide technical proofs of all theoretical results, additional simulation studies, and further details of the real data analysis presented in the main paper. In particular, Section S1 provides the proof of Theorem 1, Section S2 proves Proposition 2, Section S3 discusses how Condition 3 of Theorem 1 can be relaxed under a simple hierarchical factor structure, Section S4 establishes Theorem 2, Section S5 presents numerical results demonstrating the convergence of Algorithm 3 to the global solution from multiple random starting points, Section S6 shows the numerical results of Algorithms 1 and 2 in learning the hierarchical factor structure when $c_{\max}$ is underspecified, Section S7 provides the construct of the data discussed in the real data analysis, and Section S8 presents the numerical results of the alternative models discussed in the real data analysis.

S1 Proof of Theorem 1

In this section, we give the proof of Theorem 1. For simplicity of notation, for any matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{S}_1 \subset \{1,\ldots,m\}$ and $\mathcal{S}_2 \subset \{1,\ldots,n\}$, we denote by $A_{[\mathcal{S}_1,:]} = A_{[\mathcal{S}_1,\{1,\ldots,n\}]}$ and $A_{[:,\mathcal{S}_2]} = A_{[\{1,\ldots,m\},\mathcal{S}_2]}$.

Proof. Suppose that there exists a hierarchical factor model satisfying the constraints C1-C4, and the corresponding loading matrix Λ and the unique variance matrix Ψ satisfy $\Sigma = \Lambda\Lambda^\top + \Psi$ and $\Sigma = \Sigma^*$. We prove Theorem 1 by induction on the depth of the hierarchy. It suffices to prove that $\mathrm{Ch}_1 = \mathrm{Ch}_1^*$, $v_k = v_k^*$ for all $k \in \mathrm{Ch}_1^*$ and $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1^*$ or $\boldsymbol{\lambda}_1 = -\boldsymbol{\lambda}_1^*$ hold, where $v_1, \ldots, v_K$ are the corresponding sets of variables for each factor according to Λ, $\mathrm{Ch}_1, \ldots, \mathrm{Ch}_K$ are the child factors of each factor according to the hierarchical factor model given Λ, and $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_1^*$ are the first columns of Λ and Λ^* respectively.

First, we establish that for each $k \in \mathrm{Ch}_1^*$, there exits $i \in \mathrm{Ch}_1$ such that $v_k^* \subset v_i$. By Condition 2, we have $\Lambda\Lambda^\top = \Lambda^*(\Lambda^*)^\top$ and $\Psi = \Psi^*$. If $\mathrm{Ch}_1^* = \emptyset$, the result holds trivially. Otherwise, suppose $\mathrm{Ch}_1^* \neq \emptyset$. For any $k \in \mathrm{Ch}_1^*$, define $\mathcal{B}_{k,i} = v_k^* \cap v_i, i \in \mathrm{Ch}_1$.

If $\mathrm{Ch}_k^* = \emptyset$, consider the following cases:

1. We have $|\{i \in \mathrm{Ch}_1 : |\mathcal{B}_{k,i}| \geq 1\}| \geq 4$, which implies the existence of four distinct factors i_1, i_2, i_3, i_4 such that $v_{i_j} \cap v_k^* \neq \emptyset$ for $j = 1, \ldots, 4$. In this case, choose $j_1 \in \mathcal{B}_{k,i_1}, \ldots, j_4 \in \mathcal{B}_{k,i_4}$. Consider $\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]}$, which is equivalent to

$$\Lambda_{[\{j_1,j_2\},\{1\}]}(\Lambda_{[\{j_3,j_4\},\{1\}]})^\top = \Lambda^*_{[\{j_1,j_2\},\{1,k\}]}(\Lambda^*_{[\{j_3,j_4\},\{1,k\}]})^\top. \quad \text{(S1.1)}$$

   Observe that the left-hand side of (S1.1) has rank at most 1, whereas, by Condition 3, the right-hand side has rank 2. This contradicts (S1.1). Hence, this case cannot occur.

2. There exist i_1 and $i_2 \neq i_1$ such that $|\mathcal{B}_{k,i_1}| \geq 2$ and $|\mathcal{B}_{k,i_2}| \geq 1$. In this case, choose distinct $j_1, j_2 \in \mathcal{B}_{k,i_1}$ and $j_3 \in \mathcal{B}_{k,i_2}$. Consider $\Sigma_{[\{j_1,j_2,j_3\},\{j_1,j_2,j_3\}]} = \Sigma^*_{[\{j_1,j_2,j_3\},\{j_1,j_2,j_3\}]}$, which is equivalent to

$$\Lambda_{[\{j_1,j_2,j_3\},:]}(\Lambda_{[\{j_1,j_2,j_3\},:]})^\top = \Lambda^*_{[\{j_1,j_2,j_3\},\{1,k\}]}(\Lambda^*_{[\{j_1,j_2,j_3\},\{1,k\}]})^\top. \quad \text{(S1.2)}$$

   By Condition 3, the right-hand side of (S1.2) has rank 2. Moreover, Condition 3 also implies that the submatrix $\Sigma^*_{[\{j_1,j_2\},\{j_1,j_2\}]}$ has rank 2, and hence the matrix $\Lambda_{[\{j_1,j_2\},:]}$ must have rank 2 as well. However, observe that for any $s \in \{i_2\} \cup D_{i_2}$, $\lambda_{j_1,s} = 0$ and $\lambda_{j_2,s} = 0$, whereas $\Lambda_{[\{j_3\},\{i_2\}\cup D_{i_2}]} \neq \mathbf{0}$. Consequently, $\Lambda_{[\{j_1,j_2,j_3\},:]}$ has rank 3. Thus, the left-hand side of (S1.2) has rank 3, which contradicts (S1.2). Therefore, this case cannot occur.

3. $|v_k^*| = 3$, and there exist distinct i_1, i_2, i_3 such that $|\mathcal{B}_{k,i_1}| = |\mathcal{B}_{k,i_2}| = |\mathcal{B}_{k,i_3}| = 1$. Let $\{j_1\} = \mathcal{B}_{k,i_1}$, $\{j_2\} = \mathcal{B}_{k,i_2}$, and $\{j_3\} = \mathcal{B}_{k,i_3}$. Consider

$$\Sigma_{[\{j_1,j_2,j_3\},\{j_1,j_2,j_3\}]} = \Sigma^*_{[\{j_1,j_2,j_3\},\{j_1,j_2,j_3\}]},$$

   which is equivalent to (S1.2). In this case, the left-hand side of (S1.2) has rank 3, whereas, by Condition 3, the right-hand side has rank 2. This contradicts (S1.2). Hence, this case cannot occur.

4. There exists a unique $i \in \mathrm{Ch}_1$ such that $\mathcal{B}_{k,i_1} = v_k^*$, which indicates that $v_k^* \subset v_i$.

When $\mathrm{Ch}_k^* \neq \emptyset$, consider the following cases:

1. There exist $s \in \mathrm{Ch}_k^*$ and $i \in \mathrm{Ch}_1$ such that $|\mathcal{B}_{k,i} \cap v_s^*| \geq 2$. In this case, we claim that

$$|(\cup_{i' \neq i, i' \in \mathrm{Ch}_1} \mathcal{B}_{k,i'}) \cap (\cup_{s' \neq s, s' \in \mathrm{Ch}_k^*} v_{s'}^*)| \leq 1, \tag{S1.3}$$

Otherwise, choose $j_1, j_2 \in \mathcal{B}_{k,i} \cap v_s^*$ and $j_3, j_4 \in (\cup_{i' \neq i, i' \in \mathrm{Ch}_1} \mathcal{B}_{k,i'}) \cap (\cup_{s' \neq s, s' \in \mathrm{Ch}_k^*} v_{s'}^*)$. Consider $\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]}$, which is equivalent to (S1.1). The left-hand side of (S1.1) has rank 1, whereas by Condition 3, the right-hand side has rank 2. This contradicts (S1.1), and thus the claim in (S1.3) holds.

Now observe that $|v_{s'}^*| \geq 3$ for all $s' \neq s, s' \in \mathrm{Ch}_k^*$. Combined with (S1.3), $|\mathcal{B}_{k,i} \cap v_{s'}^*| \geq 2$ for all $s' \in \mathrm{Ch}_k^*$. By an analogous argument, we also have,

$$|(\cup_{i' \neq i, i' \in \mathrm{Ch}_1} \mathcal{B}_{k,i'}) \cap v_s^*| \leq 1, \tag{S1.4}$$

holds. Combining (S1.3) with (S1.4) yields

$$|(\cup_{i' \neq i, i' \in \mathrm{Ch}_1} \mathcal{B}_{k,i'}) \cap (\cup_{s' \in \mathrm{Ch}_k^*} v_{s'}^*)| \leq 2. \tag{S1.5}$$

We now analyze the possible values of the left-hand side of (S1.5). If $|(\cup_{i' \neq i, i' \in \mathrm{Ch}_1} \mathcal{B}_{k,i'}) \cap (\cup_{s' \in \mathrm{Ch}_k^*} v_{s'}^*)| = 2$, there exists some $s' \neq s$ such that (S1.3) is tight. We choose $j_1, j_2 \in \mathcal{B}_{k,i} \cap v_s^*$, $j_3, j_4 \in \mathcal{B}_{k,i} \cap v_{s'}^*$,

$j_5 \in (\cup_{i'\neq i, i'\in \mathrm{Ch}_1}\mathcal{B}_{k,i'})\cap v_s^*$ and $j_6 \in (\cup_{i'\neq i, i'\in \mathrm{Ch}_1}\mathcal{B}_{k,i'})\cap v_{s'}^*$. Furthermore, when $\mathrm{Ch}_s^* \neq \emptyset$, we require that j_1, j_2 belong to different child factors of factor s with j_5. Similarly, when $\mathrm{Ch}_{s'}^* \neq \emptyset$, we require that j_3, j_4 belong to different child factors of factor s' with j_6. Such a choice is always possible due to the assumed structure of the hierarchical model. Now consider $\Sigma_{[\{j_1,j_2,j_3,j_4\},\{j_5,j_6\}]} = \Sigma^*_{[\{j_1,j_2,j_3,j_4\},\{j_5,j_6\}]}$, which is equivalent to

$$\begin{aligned}&\Lambda_{[\{j_1,j_2,j_3,j_4\},\{1\}]}\big(\Lambda_{[\{j_5,j_6\},\{1\}]}\big)^\top\\ =&\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,k,s,s'\}]}\big(\Lambda^*_{[\{j_5,j_6\},\{1,k,s,s'\}]}\big)^\top,\end{aligned} \tag{S1.6}$$

by the construction of $j_1,\ldots,j_6$. The left-hand side of (S1.6) has rank 1. On the other hand, by Condition 3, the matrix $\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,k,s,s'\}]}$ has rank 4, and $\Lambda^*_{[\{j_5,j_6\},\{1,k,s,s'\}]}$ has rank 2. By Sylvester's rank inequality (see, e.g., Horn and Johnson, 2012),

$$\begin{aligned}&\mathrm{rank}\big(\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,k,s,s'\}]}\big(\Lambda^*_{[\{j_5,j_6\},\{1,k,s,s'\}]}\big)^\top\big)\\ \geq&\mathrm{rank}\big(\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,k,s,s'\}]}\big)+\mathrm{rank}\big(\Lambda^*_{[\{j_5,j_6\},\{1,k,s,s'\}]}\big)-4\\ =&2,\end{aligned}$$

which contradicts (S1.6). Hence, this case cannot occur.

If $|(\cup_{i'\neq i, i'\in \mathrm{Ch}_1}\mathcal{B}_{k,i'})\cap(\cup_{s'\in \mathrm{Ch}_k^*}v_{s'}^*)| = 1$. Without loss of generality, assume $(\cup_{i'\neq i, i'\in \mathrm{Ch}_1}\mathcal{B}_{k,i'})\cap(\cup_{s'\in \mathrm{Ch}_k^*}v_{s'}^*) = \mathcal{B}_{k,i_1}\cap v_{s_1}^* = \{j\}$, where $i_1 \in \mathrm{Ch}_1, i_1 \neq i$ and $s_1 \in \mathrm{Ch}_k^*, s_1 \neq s$. Consider $\Sigma_{[v_k^*,v_k^*]} = \Sigma^*_{[v_k^*,v_k^*]}$, which is equivalent to

$$\Lambda_{[v_k^*,:]}(\Lambda_{[v_k^*,:]})^\top = \Lambda^*_{[v_k^*,\{1,k\}\cup D_k^*]}(\Lambda^*_{[v_k^*,\{1,k\}\cup D_k^*]})^\top. \tag{S1.7}$$

By Condition 3, $\Lambda^*_{[v_k^*\setminus\{j\},\{1,k\}\cup D_k^*]}$ has rank $2+|D_k^*|$. Thus, $\Lambda_{[v_k^*\setminus\{j\},:]}$ has rank $2+|D_k^*|$. Since $\Lambda_{[\{j\},\{i_1\}]} \neq 0$ and $\Lambda_{[v_k^*\setminus\{j\},\{i_1\}]} = \mathbf{0}$, $\Lambda_{[v_k^*,:]}$ has rank $3+|D_k^*|$, which contradicts (S1.7). Hence, this case cannot occur.

If $|(\cup_{i'\neq i, i'\in \mathrm{Ch}_1}\mathcal{B}_{k,i'}) \cap (\cup_{s'\in \mathrm{Ch}_k^*} v_{s'}^*)| = 0$, there exists a unique $i \in \mathrm{Ch}_1$ such that $\mathcal{B}_{k,i} = v_k^*$, which indicates $v_k^* \subset v_i$.

2. $|\mathcal{B}_{k,i} \cap v_s^*| \leq 1$ for all $i \in \mathrm{Ch}_1$ and $s \in \mathrm{Ch}_k^*$. If there exist some $i \in \mathrm{Ch}_1$ and $s \in \mathrm{Ch}_k^*$ such that $|\mathcal{B}_{k,i} \cap v_s^*| = 1$ and $|\mathcal{B}_{k,i} \cap v_{s'}^*| = 0$ for all $s' \in \mathrm{Ch}_k^*, s' \neq s$, assume $\{j\} = \mathcal{B}_{k,i} \cap v_s^*$. Similar to the proof in (S1.7), the matrices on both sides have unequal ranks. Thus, the assumption does not hold. We assume that there exist $i \in \mathrm{Ch}_1$, $s_1 \in \mathrm{Ch}_k^*$ and $s_2 \in \mathrm{Ch}_k^*, s_2 \neq s_1$ such that $|\mathcal{B}_{k,i} \cap v_{s_1}^*| = 1$ and $|\mathcal{B}_{k,i} \cap v_{s_2}^*| = 1$. If there further exists $s_3 \in \mathrm{Ch}_k^*, s_3 \neq s_1, s_2$ such that $|\mathcal{B}_{k,i} \cap v_{s_3}^*| = 0$, we denote by $\{j_1\} = \mathcal{B}_{k,i} \cap v_{s_1}^*$ and $\{j_2\} = \mathcal{B}_{k,i} \cap v_{s_2}^*$. Consider $\Sigma_{[v_{s_3}^*,\{j_1,j_2\}]} = \Sigma^*_{[v_{s_3}^*,\{j_1,j_2\}]}$, which is equivalent to

$$\Lambda_{[v_{s_3}^*,\{1\}]}(\Lambda_{[\{j_1,j_2\},\{1\}]})^\top = \Lambda^*_{[v_{s_3}^*,\{1,k\}]}(\Lambda^*_{[\{j_1,j_2\},\{1,k\}]})^\top.$$

Noticing that the rank of the matrix on the left side is 1, while according to Condition 3, the rank of the matrix on the right side is 2, the assumption does not hold.

Thus, for any $i \in \mathrm{Ch}_1$, if there exists some $s \in \mathrm{Ch}_k^*$ such that $|\mathcal{B}_{k,i} \cap v_s^*| = 1$, then $|\mathcal{B}_{k,i} \cap v_s^*| = 1$ for all $s \in \mathrm{Ch}_k^*$, which indicate that $|v_s^*|$ are the same for $s \in \mathrm{Ch}_k^*$. If $|\mathrm{Ch}_k^*| \geq 3$, let $s_1, s_2, s_3 \in \mathrm{Ch}_k^*$ and $i_1, i_2, i_3 \in \mathrm{Ch}_1$ such that $\{j_1\} = \mathcal{B}_{k,i_1} \cap v_{s_1}^*$, $\{j_2\} = \mathcal{B}_{k,i_2} \cap v_{s_1}^*$, $\{j_3\} = \mathcal{B}_{k,i_3} \cap v_{s_2}^*$, $\{j_4\} = \mathcal{B}_{k,i_3} \cap v_{s_3}^*$. Consider $\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]}$, which is equivalent to

$$\Lambda_{[\{j_1,j_2\},\{1\}]}(\Lambda_{[\{j_3,j_4\},\{1\}]})^\top = \Lambda^*_{[\{j_1,j_2\},\{1,k\}]}(\Lambda^*_{[\{j_3,j_4\},\{1,k\}]})^\top. \tag{S1.8}$$

Since the left-hand side has rank 1, while by Condition 3, the right-hand side has rank 2, the assumption does not hold.

Finally, if $|\mathrm{Ch}_k^*| = 2$, let $\{j_1\} = \mathcal{B}_{k,i_1} \cap v_{s_1}^*$, $\{j_2\} = \mathcal{B}_{k,i_1} \cap v_{s_2}^*$, $j_3, j_4 \in v_{s_1}^*, j_3, j_4 \neq j_1$ and $j_5, j_6 \in v_{s_2}^*, j_5, j_6 \neq j_2$. Furthermore, when $|\mathrm{Ch}_{s_1}^*| \neq 0$, we require that j_3, j_4 belong to different child factors of factor s_1 with j_1. Similarly, when $|\mathrm{Ch}_{s_2}^*| \neq 0$, j_5, j_6 belong to different child factors of factor s_2 with j_2. Such a choice is always possible due to the assumed structure of the hierarchical model. Consider $\Sigma_{[\{j_1,j_2\},\{j_3,j_4,j_5,j_6\}]} = \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4,j_5,j_6\}]}$, which is equivalent to

$$\begin{aligned}&\Lambda_{[\{j_1,j_2\},\{1\}]}(\Lambda_{[\{j_3,j_4,j_5,j_6\},\{1\}]})^\top \\ =&\Lambda^*_{[\{j_1,j_2\},\{1,k,s_1,s_2\}]}(\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,k,s_1,s_2\}]})^\top.\end{aligned} \tag{S1.9}$$

The left-hand side of (S1.9) has rank 1. On the other hand, by Condition 3, $\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,k,s_1,s_2\}]}$ has rank 4 and $\Lambda^*_{[\{j_1,j_2\},\{1,k,s_1,s_2\}]}$ has rank 2. By Sylvester's rank inequality,

$$\begin{aligned}&\operatorname{rank}\big(\Lambda^*_{[\{j_1,j_2\},\{1,k,s_1,s_2\}]}(\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,k,s_1,s_2\}]})^\top\big)\\ \geq&\operatorname{rank}\big(\Lambda^*_{[\{j_1,j_2\},\{1,k,s_1,s_2\}]}\big)+\operatorname{rank}\big(\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,k,s_1,s_2\}]}\big)-4\\ =&2,\end{aligned} \tag{S1.10}$$

which contradicts (S1.9). Thus, the assumption does not hold.

From the previous proof, for any $k \in \mathrm{Ch}_1^*$, there exists $i \in \mathrm{Ch}_1$ such that $v_k^* \subset v_i$. For any $i \in \mathrm{Ch}_1$, define $C_i = \{k \in \mathrm{Ch}_1^* : v_k^* \subset v_i\}$. Consider $\Sigma_{[v_i,v_i]} = \Sigma^*_{[v_i,v_i]}$, which is equivalent to

$$\Lambda_{[v_i,\{1,i\}\cup D_i]}(\Lambda_{[v_i,\{1,i\}\cup D_i]})^\top = \Lambda^*_{[v_i,\{1\}\cup C_i\cup(\cup_{k\in C_i} D_k^*)]}(\Lambda^*_{[v_i,\{1\}\cup C_i\cup(\cup_{k\in C_i} D_k^*)]})^\top.$$

According to Condition 3, the matrix $\Lambda^*_{[v_i,\{1\}\cup C_i\cup(\cup_{k\in C_i} D_k^*)]}$ has rank $1+|C_i|+\sum_{k\in C_i}|D_k^*|$. Thus, we must have $1+|D_i| \geq |C_i| + \sum_{k\in C_i}|D_k^*|$. Summing both sides over all $i \in \mathrm{Ch}_1$, we have

$$K-1 = \sum_{i\in\mathrm{Ch}_1}(1+|D_i|) \geq \sum_{i\in\mathrm{Ch}_1}\left(|C_i| + \sum_{k\in C_i}|D_k^*|\right) = \sum_{k\in\mathrm{Ch}_1^*}(1+|D_k^*|) = K-1.$$

Therefore,

$$1+|D_i| = |C_i| + \sum_{k\in C_i}|D_k^*|, \tag{S1.11}$$

for every $i \in \mathrm{Ch}_1$. According to Lemma 5.1 of Anderson and Rubin (1956), there exists an orthogonal matrix R_i such that

$$\Lambda_{[v_i,\{1,i\}\cup D_i]} = \Lambda^*_{[v_i,\{1\}\cup C_i\cup(\cup_{k\in C_i} D^*_k)]} R_i. \tag{S1.12}$$

On the other hand, for $i, i' \in \mathrm{Ch}_1$, consider $\Sigma_{[v_i,v_{i'}]} = \Sigma^*_{[v_i,v_{i'}]}$, which is equivalent to

$$\Lambda_{[v_i,\{1\}]}(\Lambda_{[v_{i'},\{1\}]})^\top = \Lambda^*_{[v_i,\{1\}]}(\Lambda^*_{[v_{i'},\{1\}]})^\top. \tag{S1.13}$$

Combining (S1.12) with (S1.13), either $\Lambda_{[v_i,\{1\}]} = \Lambda^*_{[v_i,\{1\}]}$ or $\Lambda_{[v_i,\{1\}]} = -\Lambda^*_{[v_i,\{1\}]}$ holds. Without loss of generality, we assume $\Lambda_{[v_i,\{1\}]} = \Lambda^*_{[v_i,\{1\}]}$, which further implies $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}^*_1$.

It remains to show that $|C_i| = 1$ for all $i \in \mathrm{Ch}_1$. Suppose, for contradiction, that there exists some $i \in \mathrm{Ch}_1$ such that $|C_i| \geq 2$. Since $|D_i| \geq 2$, for $s_1, s_2 \in \mathrm{Ch}_i$, there exist $k_1, k_2 \in C_i$ such that $v_{s_1} \cap v^*_{k_1} \neq \emptyset$ and $v_{s_2} \cap v^*_{k_2} \neq \emptyset$. Consider $\Sigma_{[v_{s_1}\cap v^*_{k_1}, v_{s_2}\cap v^*_{k_2}]} = \Sigma^*_{[v_{s_1}\cap v^*_{k_1}, v_{s_2}\cap v^*_{k_2}]}$. Combined with $\Lambda_{[v_{s_1}\cap v^*_{k_1},\{1\}]} = \Lambda^*_{[v_{s_1}\cap v^*_{k_1},\{1\}]}$ and $\Lambda_{[v_{s_2}\cap v^*_{k_2},\{1\}]} = \Lambda^*_{[v_{s_2}\cap v^*_{k_2},\{1\}]}$, we have

$$\Lambda_{[v_{s_1}\cap v^*_{k_1},\{i\}]}(\Lambda_{[v_{s_2}\cap v^*_{k_2},\{i\}]})^\top = \mathbf{0}.$$

Consequently, $\Lambda_{[v_{s_1}\cap v^*_{k_1},\{i\}]} = \mathbf{0}$ or $\Lambda_{[v_{s_2}\cap v^*_{k_2},\{i\}]} = \mathbf{0}$, which contradicts the definition of v_i. Thus, $|C_i| = 1$ for all $i \in \mathrm{Ch}_1$. Therefore, we have shown that $\mathrm{Ch}_1 = \mathrm{Ch}^*_1$, $v_k = v^*_k$ for all $k \in \mathrm{Ch}^*_1$.

Finally, combining $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}^*_1$ with $\Sigma = \Sigma^*$, the covariance equality decomposes into $|\mathrm{Ch}_1^*|$ independent equations

$$\Lambda_{[v_k^*,\{k\}\cup D_k]}(\Lambda_{[v_k^*,\{k\}\cup D_k]})^\top = \Lambda^*_{[v_k^*,\{k\}\cup D_k^*]}(\Lambda^*_{[v_k^*,\{k\}\cup D_k^*]})^\top,$$

$k \in \mathrm{Ch}_1^*$. By (S1.11), we have $|D_k| = |D_k^*|$ for all $k \in \mathrm{Ch}_1^*$. Thus, by applying the same argument recursively to factors on the tth layer, $t = 2, \ldots, T$, we conclude that $\Lambda = \Lambda^* Q$ and $\Psi = \Psi^*$ for some sign flip matrix Q. □

S2 Proof of Proposition 2

In this section, we give the proof of Proposition 2.

Proof. Since factor j and its descendant factors construct a hierarchical factor structure that satisfies constraint C1-C4, it suffices to prove that

$$|v_1^*| \geq 3 + |D_1^*|. \tag{S2.14}$$

Let L_t be the factors that belong to the tth layer, $t = 1, \ldots, T$. We divide L_t into $L_t^{(1)} = \{k \in L_t : \mathrm{Ch}_k^* \neq \emptyset\}$ and $L_t^{(2)} = \{k \in L_t : \mathrm{Ch}_k^* = \emptyset\}$ for $t = 2, \ldots, T$ so that $L_t^{(1)} \cup L_t^{(2)} = L_t$ and $L_t^{(1)} \cap L_t^{(2)} = \emptyset$. By definition, $L_T^{(1)} = \emptyset$. By constraint C3, we first have

$$|L_t^{(1)}| \leq \left\lfloor \frac{1}{2}|L_{t+1}| \right\rfloor = \left\lfloor \frac{1}{2}|L_{t+1}^{(1)}| + \frac{1}{2}|L_{t+1}^{(2)}| \right\rfloor, t = 2, \ldots, T-1. \tag{S2.15}$$

Iterating (S2.15) for $t+1 \leq j \leq T-1$ yields

$$|L_t^{(1)}| \leq \sum_{j=t+1}^{T} \frac{1}{2^{j-t}} |L_j^{(2)}|.$$

Consequently,

$$\begin{aligned}
|D_1^*| =& \sum_{t=2}^{T} |L_t| \\
=& \sum_{t=2}^{T} |L_t^{(1)}| + |L_t^{(2)}| \\
\leq& \sum_{t=2}^{T} (|L_t^{(2)}| + \sum_{j=t+1}^{T} \frac{1}{2^{j-t}} |L_j^{(2)}|) \qquad \text{(S2.16)} \\
=& \sum_{t=2}^{T} (2 - \frac{1}{2^{t-2}}) |L_t^{(2)}| \\
<& 2 \sum_{t=2}^{T} |L_t^{(2)}|.
\end{aligned}$$

On the other hand, constraint C4 implies

$$|v_1^*| \geq 3 \sum_{t=2}^{T} |L_t^{(2)}|. \qquad \text{(S2.17)}$$

Combining (S2.16) and (S2.17), we have $|v_1^*| > \frac{2}{3}|D_1^*|$. In particular, (S2.14) holds when $v_1^* \geq 9$. When $|v_1^*| = 7$ or 8, $|D_1^*| \leq 2$ by constraint C4 and (S2.14) holds. When $3 \leq |v_1^*| \leq 6$, $|D_1^*| = 0$ by constraint C4 and (S2.14) holds. □

S3 Further discussions of Condition 3

In this section, we discuss the identifiability of a bi-factor model with two group factors, which constructs a special case of a two-layer hierarchical factor model. Let Λ^* and Ψ^* be the true loading matrix and the unique variance matrix. Let v_1^*, v_2^* and v_3^* be the sets of variables loading on the general factor (Factor 1) and the two group factors (Factors 2 and 3) such that $v_1^* = v_2^* \cup v_3^*$ and $v_2^* \cap v_3^* = \emptyset$. Lemma 1 provides a counterexample showing that Conditions 1 and the following Condition S1—the latter being a sufficient condition for Condition 2 under this specific hierarchical structure—are not sufficient to guarantee identifiability of the model. Theorem S1 then establishes the identifiability under the additional Condition S2.

Condition S1. $|v_g^*| \geq 3$ and $\Lambda^*_{[v_g^*,\{1,g\}]}$ is of full rank for $g \in \{2,3\}$. Moreover, there exists $g \in \{2,3\}$ and disjoint partition E_1, E_2 such that $E_1 \cup E_2 = v_g^*$, $E_1 \cap E_2 = \emptyset$, and $\Lambda^*_{[E_1,\{1,g\}]}, \Lambda^*_{[E_2,\{1,g\}]}$ are of full column rank.

Lemma 1. *Suppose that Conditions 1 and S1 hold. There exists another hierarchical factor structure with the loading matrix Λ and the unique variance matrix Ψ such that $\Lambda\Lambda^\top + \Psi = \Sigma^*$.*

Proof. It is easy to check that Condition S1 is a sufficient condition for

Condition 2. Thus, we have $\Psi = \Psi^*$ and focus on constructing a loading matrix Λ that produces the same covariance matrix Σ^* but corresponds to a different hierarchical factor structure. The hierarchical factor structure decoded by Λ is still a bi-factor structure with two group factors. Let v_2 and v_3 be the sets of variables belonging to Factor 2 and 3 according to Λ. Moreover, let $\mathcal{B}_{2,2} = v_2^* \cap v_2$, $\mathcal{B}_{2,3} = v_2^* \cap v_3$, $\mathcal{B}_{3,2} = v_3^* \cap v_2$ and $\mathcal{B}_{3,3} = v_3^* \cap v_3$ and assume that $|\mathcal{B}_{i,j}| \neq 0$ for all $i, j \in \{2, 3\}$. Now we construct Λ^* and Λ by specifying their nonzero loadings as follows:

$$\begin{aligned}
&\Lambda^*_{[\mathcal{B}_{2,2},\{2\}]} = \frac{1}{2}\Lambda^*_{[\mathcal{B}_{2,2},\{1\}]}, \Lambda_{[\mathcal{B}_{2,2},\{1\}]} = \Lambda^*_{[\mathcal{B}_{2,2},\{1\}]}, \Lambda_{[\mathcal{B}_{2,2},\{2\}]} = \frac{1}{2}\Lambda^*_{[\mathcal{B}_{2,2},\{1\}]},\\
&\Lambda^*_{[\mathcal{B}_{2,3},\{2\}]} = 2\Lambda^*_{[\mathcal{B}_{2,3},\{1\}]}, \Lambda_{[\mathcal{B}_{2,3},\{1\}]} = 2\Lambda^*_{[\mathcal{B}_{2,3},\{1\}]}, \Lambda_{[\mathcal{B}_{2,3},\{3\}]} = -\Lambda^*_{[\mathcal{B}_{2,3},\{1\}]},\\
&\Lambda^*_{[\mathcal{B}_{3,2},\{3\}]} = \frac{1}{2}\Lambda^*_{[\mathcal{B}_{3,2},\{1\}]}, \Lambda_{[\mathcal{B}_{3,2},\{1\}]} = \frac{1}{2}\Lambda^*_{[\mathcal{B}_{3,2},\{1\}]}, \Lambda_{[\mathcal{B}_{3,2},\{2\}]} = \Lambda^*_{[\mathcal{B}_{3,2},\{1\}]},\\
&\Lambda^*_{[\mathcal{B}_{3,3},\{3\}]} = -\Lambda^*_{[\mathcal{B}_{3,3},\{1\}]}, \Lambda_{[\mathcal{B}_{3,3},\{1\}]} = \Lambda^*_{[\mathcal{B}_{3,3},\{1\}]}, \Lambda_{[\mathcal{B}_{3,3},\{3\}]} = \Lambda^*_{[\mathcal{B}_{3,3},\{1\}]}.
\end{aligned} \tag{S3.18}$$

As long as $|\mathcal{B}_{2,2}| \geq 2$ and $|\mathcal{B}_{2,3}| \geq 2$, Λ^* satisfies Conditions 1 and S1. However, $\Lambda\Lambda^\top = \Lambda^*\Lambda^{*\top}$ while Λ and Λ^* produce different hierarchical factor structures. □

To avoid the counterexample raised in Lemma 1, we need the following Condition S2.

Condition S2. There exists $g \in \{2, 3\}$ and $i, j, k \in v_g^*$ such that $\Lambda^*_{[\{i,j\},\{1,g\}]}$ $\Lambda^*_{[\{i,k\},\{1,g\}]}$ and $\Lambda^*_{[\{j,k\},\{1,g\}]}$ are of full rank.

Theorem S1. *Suppose that Conditions 1, S1, and S2 hold. If there exists some hierarchical factor structure with three factors such that its loading matrix* Λ *and unique variance matrix* Ψ *satisfy* $\Sigma^* = \Lambda\Lambda^\top + \Psi$*, there exists some sign flip matrix* $Q \in \mathcal{Q}$ *such that* $\Lambda = \Lambda^* Q$*, where* $\mathcal{Q}$ *consists of all the* 3×3 *diagonal matrix* Q *whose diagonal entries take values* 1 *or* -1*.*

Proof. We adopt the notation from the proof of Lemma 1. By Condition S1, we have $\Lambda\Lambda^\top = \Lambda^*\Lambda^{*\top}$ and $\Psi = \Psi^*$. We consider the following cases:

1. $|\mathcal{B}_{s_1,s_2}| \neq 0$ for all $s_1, s_2 \in \{2, 3\}$.

2. Without loss of generality, $v_2 = v_2^*$ but $v_3 \neq v_3^*$.

3. $v_2 = v_2^*$ and $v_3 = v_3^*$.

In the first case, since $\Lambda^*_{[v_2^*,\{1,2\}]}$ and $\Lambda^*_{[v_3^*,\{1,3\}]}$ are of rank-2, all the following submatrices are of rank-1:

$$\Lambda^*_{[\mathcal{B}_{2,2},\{1,2\}]}, \Lambda^*_{[\mathcal{B}_{2,3},\{1,2\}]}, \Lambda^*_{[\mathcal{B}_{3,2},\{1,3\}]}, \Lambda^*_{[\mathcal{B}_{3,3},\{1,3\}]},$$

$$\Lambda_{[\mathcal{B}_{2,2},\{1,2\}]}, \Lambda_{[\mathcal{B}_{2,3},\{1,3\}]}, \Lambda_{[\mathcal{B}_{3,2},\{1,2\}]}, \Lambda_{[\mathcal{B}_{3,3},\{1,3\}]}.$$

However, according to Condition S2, there exists at least one of the matrices $\Lambda^*_{[\mathcal{B}_{2,2},\{1,2\}]}, \Lambda^*_{[\mathcal{B}_{2,3},\{1,2\}]}, \Lambda^*_{[\mathcal{B}_{3,2},\{1,3\}]}$ and $\Lambda^*_{[\mathcal{B}_{3,3},\{1,3\}]}$ such that it is of rank-2. Thus, the first case is not allowed.

In the second case, we assume $v_2 = v_2^*$ while $v_3 \neq v_3^*$ without loss of generality. Since $\Lambda_{[v_2^*,\{1,2\}]}\Lambda^\top_{[v_2^*,\{1,2\}]} = \Lambda^*_{[v_2^*,\{1,2\}]}\Lambda^{*\top}_{[v_2^*,\{1,2\}]}$, there exits some orthogonal rotation matrix $R \in \mathbb{R}^{2\times 2}$ such that

$$\Lambda_{[v_2^*,\{1,2\}]} = \Lambda^*_{[v_2^*,\{1,2\}]} R. \tag{S3.19}$$

With $\Lambda_{[v_2^*,\{1\}]}\Lambda^\top_{[\mathcal{B}_{3,3},\{1\}]} = \Lambda^*_{[v_2^*,\{1\}]}\Lambda^{*\top}_{[\mathcal{B}_{3,3},\{1\}]}$, there exists some constant a such that $\Lambda_{[v_2^*,\{1\}]} = a\Lambda^*_{[v_2^*,\{1\}]}$. Combined with (S3.19), $a = 1$ or -1 since $\Lambda^*_{[v_2^*,\{1,2\}]}$ is of rank-2. Without loss of generality, we assume $\Lambda_{[v_2^*,\{1\}]} = \Lambda^*_{[v_2^*,\{1\}]}$ and $\Lambda_{[v_2^*,\{2\}]} = \Lambda^*_{[v_2^*,\{2\}]}$ further. Then, consider $\Lambda_{[v_2^*,\{1,2\}]}\Lambda^\top_{[\mathcal{B}_{3,2},\{1,2\}]} = \Lambda^*_{[v_2^*,\{1\}]}\Lambda^{*\top}_{[\mathcal{B}_{3,2},\{1\}]}$, which leads to $\Lambda_{[\mathcal{B}_{3,2},\{2\}]} = \mathbf{0}$. Thus, the second case is not allowed.

In the third case, similar to the proof in the second case, there exists two orthogonal rotation matrices $R_1, R_2 \in \mathbb{R}^{2\times 2}$ and such that

$$\Lambda_{[v_2^*,\{1,2\}]} = \Lambda^*_{[v_2^*,\{1,2\}]} R_1 \text{ and } \Lambda_{[v_3^*,\{1,3\}]} = \Lambda^*_{[v_3^*,\{1,3\}]} R_2. \tag{S3.20}$$

Combined with $\Lambda_{[v_2^*,\{1\}]}\Lambda^\top_{[v_3^*,\{1\}]} = \Lambda^*_{[v_2^*,\{1\}]}\Lambda^{*\top}_{[v_3^*,\{1\}]}$, there exists some sign flip matrix $Q \in \mathcal{Q}$ such that $\Lambda = \Lambda^* Q$. □

Remark S1. Theorem S1 establishes the identifiability of the bi-factor model with two group factors. Compared to the general hierarchical identifiability result in Theorem 1, it requires fewer structural assumptions, but still needs the additional rank condition (Condition S2). The proof of Theorem S1 is based on the specific hierarchical structure and we believe the requirement for Condition 3 can be simplified based on the true hierarchical factor structure.

S4 Proof of Theorem 2

We first introduce some notations and lemmas needed for the proof of Theorem 2. Suppose that $A, \varepsilon \in \mathbb{R}^{m\times n}$. We denote by $\sigma_1(A) \geq \ldots \geq \sigma_{\min(m,n)}(A) \geq 0$ are the singular values of A, and $U_1, \ldots, U_{\min(m,n)}$ are the corresponding right(left) singular vectors. Similarly, we denote by $\sigma_1(A+\varepsilon) \geq \ldots \geq \sigma_{\min(m,n)}(A+\varepsilon) \geq 0$ as the singular values of $A+\varepsilon$ and $U'_1, \ldots, U'_{\min(m,n)}$ the corresponding right(left) singular vectors. We use $\|A\|_2$ denote the spectral norm of a matrix A.

Lemma 2 (Weyl's bound, Weyl (1912))**.**

$$\max_{1\leq i\leq \min(m,n)} |\sigma_i(A) - \sigma_i(A+\varepsilon)| \leq \|\varepsilon\|_2.$$

We further assume that the rank of A is r. We denote by $U = (U_1, \ldots, U_j)$ and $U' = (U'_1, \ldots, U'_j)$, $1 \leq j \leq r$. The following Lemma 3 is a modification of Wedin's Theorem (Wedin, 1972).

Lemma 3. *There exists some orthogonal matrix R such that*

$$\|UR - U'\|_F \leq \frac{2^{3/2} r^{1/2} \|\varepsilon\|_F}{\delta}.$$

when $\delta = \sigma_j(A) - \sigma_{j+1}(A) > 0$.

Lemma 4. *Given a $J \times K$ dimensional matrix Λ following a hierarchical structure that satisfies constraints C1-C4 and a $J \times J$ dimensional diagonal matrix $\Psi = diag(\psi_1, \ldots, \psi_J)$ with $\psi_j > 0, j = 1, \ldots, J$. Assume that Λ satisfies Condition 5 and Condition 8. If there exist a series of $J \times K$ dimensional random matrices $\{\widehat{\Lambda}_N\}_{N=1}^{\infty}$ and a series of $J \times J$ dimensional diagonal random matrices $\{\widehat{\Psi}_N\}_{N=1}^{\infty}$, where $\widehat{\Psi}_N = diag(\widehat{\psi}_{N,1}, \ldots, \widehat{\psi}_{N,J})$ with $\widehat{\psi}_{N,j} \geq 0, j = 1, \ldots, J$, such that $\{\widehat{\Lambda}_N\}_{N=1}^{\infty}$ satisfies Condition 8 and*

$$\|\widehat{\Lambda}_N \widehat{\Lambda}_N^\top + \widehat{\Psi}_N - \Lambda\Lambda^\top - \Psi\|_F = O_{\mathbb{P}}(1/\sqrt{N}). \tag{S4.21}$$

Then we have $\|\widehat{\Lambda}_N \widehat{\Lambda}_N^\top - \Lambda\Lambda^\top\|_F = O_{\mathbb{P}}(1/\sqrt{N})$ and $\|\widehat{\Psi}_N - \Psi\|_F = O_{\mathbb{P}}(1/\sqrt{N})$.

Lemma 4 is a generalization of Theorem 5.1 in Anderson and Rubin (1956), and its proof proceeds along the same lines.

Proof. For $j = 1, \ldots, J$, by Condition 5, there exist $E_1, E_2 \in \{1, \ldots, J\} \backslash \{j\}$ with $|E_1| = |E_2| = K$ and $E_1 \cap E_2 = \emptyset$ such that $\Lambda_{[E_1,:]}$ and $\Lambda_{[E_2,:]}$ are full-rank matrices. Without loss of generality, we assume that Λ and $\widehat{\Lambda}_N$ can be expressed as

$$\Lambda = \begin{pmatrix} \Lambda_1 \\ \boldsymbol{\lambda}_j \\ \Lambda_2 \\ \Lambda_3 \end{pmatrix}, \quad \widehat{\Lambda}_N = \begin{pmatrix} \widehat{\Lambda}_{N,1} \\ \widehat{\boldsymbol{\lambda}}_{N,j} \\ \widehat{\Lambda}_{N,2} \\ \widehat{\Lambda}_{N,3} \end{pmatrix},$$

where we denote by $\Lambda_1 = \Lambda_{[E_1,:]}$, $\Lambda_2 = \Lambda_{[E_2,:]}$, $\boldsymbol{\lambda}_j = \Lambda_{[\{j\},:]}$ is the jth row of Λ, Λ_3 consists of the remaining rows in Λ with a slight abuse of notation. The blocks $\widehat{\Lambda}_{N,1}$, $\widehat{\Lambda}_{N,2}$, $\widehat{\boldsymbol{\lambda}}_{N,j}$, and $\widehat{\Lambda}_{N,3}$ are defined analogously for $\widehat{\Lambda}_N$, with the same row partitioning. Thus, we have

$$\Lambda_{[E_1\cup E_2\cup\{j\},:]}\Lambda_{[E_1\cup E_2\cup\{j\},:]}^\top = \begin{pmatrix} \Lambda_1\Lambda_1^\top & \Lambda_1\boldsymbol{\lambda}_j^\top & \Lambda_1\Lambda_2^\top \\ \boldsymbol{\lambda}_j\Lambda_1^\top & \boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top & \boldsymbol{\lambda}_j\Lambda_2^\top \\ \Lambda_2\Lambda_1^\top & \Lambda_2\boldsymbol{\lambda}_j^\top & \Lambda_2\Lambda_2^\top \end{pmatrix},$$

and

$$(\widehat{\Lambda}_N)_{[E_1\cup E_2\cup\{j\},:]}(\widehat{\Lambda}_N)_{[E_1\cup E_2\cup\{j\},:]}^\top = \begin{pmatrix} \widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,1}^\top & \widehat{\Lambda}_{N,1}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,2}^\top \\ \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\Lambda}_{N,1}^\top & \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\Lambda}_{N,2}^\top \\ \widehat{\Lambda}_{N,2}\widehat{\Lambda}_{N,1}^\top & \widehat{\Lambda}_{N,2}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\Lambda}_{N,2}\widehat{\Lambda}_{N,2}^\top \end{pmatrix}.$$

According to (S4.21), we have

$$\begin{aligned} \|\Lambda_1\boldsymbol{\lambda}_j^\top - \widehat{\Lambda}_{N,1}\widehat{\boldsymbol{\lambda}}_{N,j}^\top\| &= O_{\mathbb{P}}(1/\sqrt{N}), \\ \|\Lambda_2\boldsymbol{\lambda}_j^\top - \widehat{\Lambda}_{N,2}\widehat{\boldsymbol{\lambda}}_{N,j}^\top\| &= O_{\mathbb{P}}(1/\sqrt{N}), \\ \|\Lambda_1\Lambda_2^\top - \widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,2}^\top\|_F &= O_{\mathbb{P}}(1/\sqrt{N}). \end{aligned} \tag{S4.22}$$

Since each of the following $(K+1)\times(K+1)$ matrices has rank at most K

$$\begin{pmatrix} \Lambda_1\boldsymbol{\lambda}_j^\top & \Lambda_1\Lambda_2^\top \\ \boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top & \boldsymbol{\lambda}_j\Lambda_2^\top \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \widehat{\Lambda}_{N,1}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,2}^\top \\ \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\Lambda}_{N,2}^\top \end{pmatrix},$$

we have

$$\det\begin{pmatrix} \Lambda_1\boldsymbol{\lambda}_j^\top & \Lambda_1\Lambda_2^\top \\ \boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top & \boldsymbol{\lambda}_j\Lambda_2^\top \end{pmatrix} = \det\begin{pmatrix} \widehat{\Lambda}_{N,1}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,2}^\top \\ \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top & \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\Lambda}_{N,2}^\top \end{pmatrix} = 0.$$

Then, we have

$$\begin{aligned} &(-1)^K\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top \det(\Lambda_1\Lambda_2^\top) + f(\Lambda_1\boldsymbol{\lambda}_j^\top, \boldsymbol{\lambda}_j\Lambda_2^\top) \\ =&(-1)^K\widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top \det(\widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,2}^\top) + f(\widehat{\Lambda}_{N,1}\widehat{\boldsymbol{\lambda}}_{N,j}^\top, \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\Lambda}_{N,2}^\top) \\ =&0, \end{aligned} \tag{S4.23}$$

where $f(\cdot)$ is a scalar-valued function. Both $f(\cdot)$ and the determinant function $\det(\cdot)$ are Lipschitz continuous with respect to the entries of their matrix arguments, with Lipschitz constants depending only on K and τ. Combined with (S4.22) and (S4.23), we have

$$\begin{aligned} &|\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top - \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top||\det(\Lambda_1\Lambda_2^\top)| \\ \le&|f(\Lambda_1\boldsymbol{\lambda}_j^\top, \boldsymbol{\lambda}_j\Lambda_2^\top) - f(\widehat{\Lambda}_{N,1}\widehat{\boldsymbol{\lambda}}_{N,j}^\top, \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\Lambda}_{N,2}^\top)| \\ &+ |\widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top||\det(\Lambda_1\Lambda_2^\top) - \det(\widehat{\Lambda}_{N,1}\widehat{\Lambda}_{N,2}^\top)| \\ =&O_{\mathbb{P}}(1/\sqrt{N}). \end{aligned}$$

Noticing that $|\det(\Lambda_1\Lambda_2^\top)| > 0$, we have $|\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top - \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top| = O_{\mathbb{P}}(1/\sqrt{N})$. Combined with

$$|\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top + \psi_j - \widehat{\boldsymbol{\lambda}}_{N,j}\widehat{\boldsymbol{\lambda}}_{N,j}^\top - \widehat{\psi}_{N,j}| = O_{\mathbb{P}}(1/\sqrt{N}),$$

we have $|\psi_j - \widehat{\psi}_{N,j}| = O_{\mathbb{P}}(1/\sqrt{N})$ for $j = 1, \ldots, J$. Thus, we have $\|\widehat{\Psi}_N -$

$\Psi\|_F = O_{\mathbb{P}}(1/\sqrt{N})$ and furthermore we have $\|\widehat{\Lambda}_N\widehat{\Lambda}_N^\top - \Lambda\Lambda^\top\|_F = O_{\mathbb{P}}(1/\sqrt{N})$.

$\square$

For Factor $k \in L_{t-1}$, $t \geq 3$, let $\Sigma_{k,0}^* := \sum_{i=1}^{k_{t-2}}(\boldsymbol{\lambda}_i^*)_{[v_k^*]}(\boldsymbol{\lambda}_i^*)_{[v_k^*]}^\top$ when $k \in \widehat{L}_{t-1}$ and $\Sigma_k^* = (\boldsymbol{\lambda}_k^*)_{[v_k^*]}(\boldsymbol{\lambda}_k^*)_{[v_k^*]}^\top + \sum_{i \in D_k^*}(\boldsymbol{\lambda}_i^*)_{[v_k^*]}(\boldsymbol{\lambda}_i^*)_{[v_k^*]}^\top + \Psi^*_{[v_k^*, v_k^*]}$. We further define

$$
\begin{aligned}
\Theta_k(c,d) = \{&\Sigma = \Lambda_k\Lambda_k^\top + \Psi_k \in \mathbb{R}^{|v_k^*|\times|v_k^*|} : \Lambda_k \in \mathbb{R}^{|v_k^*|\times(1+cd)}, \\
&|\lambda_{k,ij}| \leq \tau, \lambda_{k,ij}\lambda_{k,ij'} = 0, \text{ for } i = 1,\ldots,|v_k^*|, j \in \mathcal{B}_s, j' \in \mathcal{B}_{s'}, s \neq s' \\
&\text{and } \Psi = \operatorname{diag}(\psi_{k1},\ldots,\psi_{k|v_k^*|}) \text{ with } \kappa_1 \leq \psi_{ki} \leq \kappa_2 \text{ for } i = 1,\ldots,|v_k^*|\},
\end{aligned} \tag{S4.24}
$$

where $\mathcal{B}_s = 2 + (s-1)d, \ldots, 1 + sd$ for $s = 1, \ldots, c$, and τ, κ_1 and κ_2 are those specified in Condition 8. Given a symmetric positive semi-definite matrix $\widetilde{\Sigma}_{k,0}$ serving as an estimator of $\Sigma_{k,0}^*$, we define

$$
\widehat{\Sigma}_k = \underset{\Sigma_k \in \Theta_k(c,d)}{\arg\min}\ l\left(\widetilde{\Sigma}_{k,0} + \Sigma_k, S_k\right). \tag{S4.25}
$$

The following Lemma 5 and 6 establish the consistency and convergence rate of $\widehat{\Sigma}_k$.

Lemma 5 (Consistency). *Suppose d is sufficiently large such that $\Sigma_k^* \in \Theta_k(c,d)$ and $\|\widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*\|_F = o_{\mathbb{P}}(1)$. If Conditions 7 and 8 hold, $\widehat{\Sigma}_k \xrightarrow{\mathbb{P}} \Sigma_k^*$.*

Proof. The proof of Lemma 5 follows Theorem 2.1 in Newey and McFadden (1994). First, we show that $\Theta_k(c,d)$ is a compact set in $\mathbb{R}^{|v_k^*|\times|v_k^*|}$. By definition, we directly have that $\Theta_k(c,d)$ is a bounded set. To prove that $\Theta_k(c,d)$ is also a closed set, we assume $\{\Sigma_k^{(n)} = \Lambda_k^{(n)}\Lambda_k^{(n)\top} + \Psi_k^{(n)}\}_{n=1}^{\infty}$ is an arbitrary convergent sequence in $\Theta_k(c,d)$. Since $\{\Lambda_k^{(n)}\}_{n=1}^{\infty}$ and $\{\Psi_k^{(n)}\}_{n=1}^{\infty}$ are bounded sequences, there exists subsequence $\{\Lambda_k^{(n_m)}\}_{m=1}^{\infty}$ and $\{\Psi_k^{(n_m)}\}_{m=1}^{\infty}$ such that

$$\lim_{m\to\infty} \Lambda_k^{(n_m)} = \Lambda_k^{\infty} \text{ and } \lim_{m\to\infty} \Psi_k^{(n_m)} = \Psi_k^{\infty}.$$

Since $\lambda_{k,ij}^{n_m}\lambda_{k,ij'}^{n_m} = 0$, $\lim_{m\to\infty} \lambda_{k,ij}^{n_m} = \lambda_{k,ij}^{\infty}$ and $\lim_{m\to\infty} \lambda_{k,ij'}^{n_m} = \lambda_{k,ij'}^{\infty}$, we have $\lambda_{k,ij}^{\infty}\lambda_{k,ij'}^{\infty} = 0$ for $i = 1,\ldots,|v_k^*|$, $j \in \mathcal{B}_s$, $j' \in \mathcal{B}_{s'}$, $1 \le s < s' \le c$. Thus,

$$\lim_{n\to\infty} \Sigma_k^{(n)} = \Lambda_k^{\infty}\Lambda_k^{\infty\top} + \Psi_k^{\infty} \in \Theta_k(c,d).$$

Then $\Theta_k(c,d)$ is a compact set.

Second, let

$$\begin{aligned}
a_k(x;\Sigma_k,\widetilde{\Sigma}_{k,0}) &= \log\det(\widetilde{\Sigma}_{k,0}+\Sigma_k) + \operatorname{tr}\left(x_{[v_k^*]}x_{[v_k^*]}^{\top}(\widetilde{\Sigma}_{k,0}+\Sigma_k)^{-1}\right)\\
M_k(\Sigma_k,\widetilde{\Sigma}_{k,0}) &= \log\det(\widetilde{\Sigma}_{k,0}+\Sigma_k) + \operatorname{tr}\left(S_k(\widetilde{\Sigma}_{k,0}+\Sigma_k)^{-1}\right)\\
M_{0,k}(\Sigma_k,\widetilde{\Sigma}_{k,0}) &= \log\det(\widetilde{\Sigma}_{k,0}+\Sigma_k) + \operatorname{tr}\left((\Sigma_{k,0}^*+\Sigma_k^*)(\widetilde{\Sigma}_{k,0}+\Sigma_k)^{-1}\right)
\end{aligned} \tag{S4.26}$$

We directly have

$$\widehat{\Sigma}_k = \mathop{\arg\min}_{\Sigma_k\in\Theta_k(c,d)} M_k(\Sigma_k,\widetilde{\Sigma}_{k,0}). \tag{S4.27}$$

Moreover,

$$
\begin{aligned}
&\frac{\partial}{\partial \Sigma_k} M_{0,k}(\Sigma_k, \Sigma_{k,0}^*) \\
=&(\Sigma_k + \Sigma_{k,0}^*)^{-1} - (\Sigma_k + \Sigma_{k,0}^*)^{-1}(\Sigma_k^* + \Sigma_{k,0}^*)(\Sigma_k + \Sigma_{k,0}^*)^{-1} \\
=&0,
\end{aligned} \tag{S4.28}
$$

when $\Sigma_k = \Sigma_k^* \in \Theta_k(c,d)$. Thus, $M_{0,k}(\Sigma_k, \Sigma_{k,0}^*)$ reaches its unique minimum at Σ_k^*.

Third,

$$
\begin{aligned}
&|a_k(x; \Sigma_k, \Sigma_{k,0}^*)| \\
\leq&|\log\det(\Sigma_{k,0}^* + \Sigma_k)| + \left|\operatorname{tr}\left(x_{[v_k^*]} x_{[v_k^*]}^\top (\Sigma_{k,0}^* + \Sigma_k)^{-1}\right)\right| \\
\leq&|v_k^*| \max\left(|\log(\sigma_{\min}(\Sigma_{k,0}^* + \Sigma_k))|, |\log(\sigma_{\max}(\Sigma_{k,0}^* + \Sigma_k))|\right) + \frac{\left\|x_{[v_k^*]}\right\|^2}{\sigma_{\min}(\Sigma_{k,0}^* + \Sigma_k)} \\
\leq&|v_k^*| \max\left(|\log\kappa_1|, \left|\log\left(|v_k^*|\left((1+cd)^2 + K^2\right)\tau^2 + \kappa_2\right)\right|\right) + \frac{1}{\kappa_1}\left\|x_{[v_k^*]}\right\|^2.
\end{aligned} \tag{S4.29}
$$

Since $\mathbb{E}\left(\left\|x_{[v_k^*]}\right\|^2\right) < \infty$, by Lemma 2.4 of Newey and McFadden (1994)

$$
\sup_{\Sigma_k \in \Theta_k(c,d)} |M_k(\Sigma_k, \Sigma_{k,0}^*) - M_{0,k}(\Sigma_k, \Sigma_{k,0}^*)| \xrightarrow{\mathbb{P}} 0. \tag{S4.30}
$$

Now we have

$$
\begin{aligned}
& M_{0,k}(\widehat{\Sigma}_k, \Sigma^*_{k,0}) \\
=& M_{0,k}(\Sigma^*_k, \Sigma^*_{k,0}) + (M_{0,k}(\widehat{\Sigma}_k, \Sigma^*_{k,0}) - M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0})) + (M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) \\
& - M_k(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0})) + (M_k(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma^*_k, \widetilde{\Sigma}_{k,0})) + (M_k(\Sigma^*_k, \widetilde{\Sigma}_{k,0}) \\
& - M_{0,k}(\Sigma^*_k, \widetilde{\Sigma}_{k,0})) + (M_{0,k}(\Sigma^*_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\Sigma^*_k, \Sigma^*_{k,0})) \\
\leq & M_{0,k}(\Sigma^*_k, \Sigma^*_{k,0}) + 2 \sup_{\Sigma_k \in \Theta_k(c,d)} |M_{0,k}(\Sigma_k, \Sigma^*_{k,0}) - M_{0,k}(\Sigma_k, \widetilde{\Sigma}_{k,0})| \\
& + 2 \sup_{\Sigma_k \in \Theta_k(c,d)} |M_{0,k}(\Sigma_k, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\Sigma_k, \Sigma^*_{k,0}) + M_k(\Sigma_k, \Sigma^*_{k,0})| \\
& + 2 \sup_{\Sigma_k \in \Theta_k(c,d)} |M_k(\Sigma_k, \Sigma^*_{k,0}) - M_{0,k}(\Sigma_k, \Sigma^*_{k,0})|.
\end{aligned}
\tag{S4.31}
$$

For arbitrary $\Sigma_k \in \Theta_k(c,d)$, according to Taylor's expansion there exists some $\eta \in (0,1)$ such that

$$
\begin{aligned}
& |M_{0,k}(\Sigma_k, \Sigma^*_{k,0}) - M_{0,k}(\Sigma_k, \widetilde{\Sigma}_{k,0})| \\
\leq & \left| \mathrm{tr} \left((\widetilde{\Sigma}_{k,0} - \Sigma^*_{k,0}) \left(\Sigma_k + (1-\eta)\Sigma^*_{k,0} + \eta \widetilde{\Sigma}_{k,0} \right)^{-1} (\Sigma^*_{k,0} + \Sigma^*_k) \left(\Sigma_k + (1-\eta)\Sigma^*_{k,0} \right.\right.\right. \\
& \left.\left.\left. + \eta \widetilde{\Sigma}_{k,0} \right)^{-1} \right) \right| + \left| \mathrm{tr} \left((\widetilde{\Sigma}_{k,0} - \Sigma^*_{k,0}) \left((1-\eta)\Sigma^*_{k,0} + \eta \widetilde{\Sigma}_{k,0} \right)^{-1} \right) \right| \\
\leq & \frac{|v^*_k| K^2 \tau^2 + \kappa_2 + \kappa_1}{\kappa_1^2} \|\widetilde{\Sigma}_{k,0} - \Sigma^*_{k,0}\|_F,
\end{aligned}
\tag{S4.32}
$$

where the last inequality follows Ruhe's trace inequality (Ruhe, 1970). Similarly, with probability approaching 1 as N grows to infinity,

$$
\begin{aligned}
&|M_k(\Sigma_k, \Sigma_{k,0}^*) - M_k(\Sigma_k, \widetilde{\Sigma}_{k,0})| \\
\leq &\left| \mathrm{tr} \left((\widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*) \left(\Sigma_k + (1-\eta)\Sigma_{k,0}^* + \eta\widetilde{\Sigma}_{k,0} \right)^{-1} S_k \left(\Sigma_k + (1-\eta)\Sigma_{k,0}^* + \eta\widetilde{\Sigma}_{k,0} \right)^{-1} \right) \right| \\
&+ \left| \mathrm{tr} \left((\widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*) \left(\Sigma_k + (1-\eta)\Sigma_{k,0}^* + \eta\widetilde{\Sigma}_{k,0} \right)^{-1} \right) \right| \\
\leq &\frac{2\left(|v_k^*| K^2 \tau^2 + \kappa_2\right) + \kappa_1}{\kappa_1^2} \|\widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*\|_F,
\end{aligned}
\tag{S4.33}
$$

With (S4.32) and (S4.33)

$$
\sup_{\Sigma_k \in \Theta_k(c,d)} |M_{0,k}(\Sigma_k, \Sigma_{k,0}^*) - M_{0,k}(\Sigma_k, \widetilde{\Sigma}_{k,0})| = o_{\mathbb{P}}(1), \tag{S4.34}
$$

and

$$
\begin{aligned}
&\sup_{\Sigma_k \in \Theta_k(c,d)} |M_{0,k}(\Sigma_k, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\Sigma_k, \Sigma_{k,0}^*) + M_k(\Sigma_k, \Sigma_{k,0}^*)| \\
=&o_{\mathbb{P}}(1).
\end{aligned}
\tag{S4.35}
$$

For arbitrary $\epsilon > 0$, let

$$
\Delta(\epsilon) = \inf_{\Sigma_k \in \Theta_k(c,d), \|\Sigma_k - \Sigma_k^*\|_F \geq \epsilon} M_{0,k}(\Sigma_k, \Sigma_{k,0}^*) - M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*) > 0.
$$

Combined with (S4.30), (S4.31), (S4.34) and (S4.35), with probability approaching 1 as N grows to infinity,

$$
M_{0,k}(\widehat{\Sigma}_k, \Sigma_{k,0}^*) < M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*) + \Delta(\epsilon),
$$

which indicates $\|\widehat{\Sigma}_k - \Sigma_k^*\|_F < \epsilon$. Thus, $\widehat{\Sigma}_k \xrightarrow{\mathbb{P}} \Sigma_k^*$. $\square$

Lemma 6 (Convergence rate)**.** *Suppose d is sufficiently large such that $\Sigma_k^* \in \Theta_k(c,d)$ and $\|\widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. If Conditions 7 and 8 hold, $\|\widehat{\Sigma}_k - \Sigma_k^*\|_F = O_{\mathbb{P}}(1/\sqrt{N})$.*

Proof. Consider

$$\begin{aligned} & M_k(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k^*, \Sigma_{k,0}^*) \\ =& M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*) + M_k(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) \\ & - M_k(\Sigma_k^*, \Sigma_{k,0}^*) + M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*). \end{aligned} \tag{S4.36}$$

Let $\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}} = \widehat{\Sigma}_k + \widetilde{\Sigma}_{k,0} - \Sigma_k^* - \Sigma_{k,0}^*$. By Taylor's expansion, there exists some $\eta \in (0,1)$ such that

$$\begin{aligned} & M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*) \\ =& \frac{1}{2}\mathrm{tr}\left(\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}} \left(\Sigma_k^* + \Sigma_{k,0}^* + \eta\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\right)^{-1} \Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}} \left(\Sigma_k^* + \Sigma_{k,0}^* + \eta\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\right)^{-1} \right. \\ & \left. \left(2\left(\Sigma_k^* + \Sigma_{k,0}^*\right)\left(\Sigma_k^* + \Sigma_{k,0}^* + \eta\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\right)^{-1} - \mathbf{I}\right)\right). \end{aligned} \tag{S4.37}$$

For simplicity of the notations, let

$$\begin{aligned} \Delta_1 &= \Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}} \left(\Sigma_k^* + \Sigma_{k,0}^* + \eta\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\right)^{-1} \Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}, \\ \Delta_2 &= \left(\Sigma_k^* + \Sigma_{k,0}^* + \eta\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\right)^{-1} \left(2\left(\Sigma_k^* + \Sigma_{k,0}^*\right)\left(\Sigma_k^* + \Sigma_{k,0}^* + \eta\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\right)^{-1} - \mathbf{I}\right) \end{aligned}$$

According to Lemma 5, $\widehat{\Sigma}_k \xrightarrow{\mathbb{P}} \Sigma_k^*$. Combined with Lemma 2, with probability approaching 1 as N grows to infinity,

$$\begin{aligned}
&\frac{\|\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\|_F^2}{|v_k^*|K^2\tau^2+\kappa_2} \\
\leq&\frac{1}{2}\sigma_{\min}\left(\left(\Sigma_k^*+\Sigma_{k,0}^*\right)^{-1}\right)\|\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\|_F^2 \\
\leq&\mathrm{tr}(\Delta_1) \qquad\qquad (S4.38)\\
\leq&2\sigma_{\max}\left(\left(\Sigma_k^*+\Sigma_{k,0}^*\right)^{-1}\right)\|\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\|_F^2 \\
\leq&2\|\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\|_F^2/\kappa_1,
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{\min}(\Delta_2)&\geq\frac{1}{2}\sigma_{\min}\left(\left(\Sigma_k^*+\Sigma_{k,0}^*\right)^{-1}\right)\geq\frac{1}{|v_k^*|K^2\tau^2+\kappa_2}, \\
\sigma_{\max}(\Delta_2)&\leq 2\sigma_{\max}\left(\left(\Sigma_k^*+\Sigma_{k,0}^*\right)^{-1}\right)\leq\frac{2}{\kappa_1}.
\end{aligned} \qquad (S4.39)$$

By the Ruhe's trace inequality, we have

$$\|\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\|_F^2=O_{\mathbb{P}}\left(M_{0,k}(\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0})-M_{0,k}(\Sigma_k^*,\Sigma_{k,0}^*)\right). \qquad (S4.40)$$

Next, by Taylor's expansion, there exists some $\eta\in(0,1)$ such that

$$\begin{aligned}
&M_k(\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0})-M_{0,k}(\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0})-M_k(\Sigma_k^*,\Sigma_{k,0}^*)+M_{0,k}(\Sigma_k^*,\Sigma_{k,0}^*) \\
=&\mathrm{tr}\left(\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\left(\Sigma_k^*+\Sigma_{k,0}^*+\eta\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\right)^{-1}\left(S_k-\Sigma_{[v_k^*,v_k^*]}^*\right)\right. \qquad (S4.41)\\
&\left.\left(\Sigma_k^*+\Sigma_{k,0}^*+\eta\Delta_{\widehat{\Sigma}_k,\widetilde{\Sigma}_{k,0}}\right)^{-1}\right).
\end{aligned}$$

Combined with Condition 7, Lemma 2 and the Ruhe's trace inequality, with probability approaching 1 as N grows to infinity,

$$\begin{aligned}&\left|M_k(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k^*, \Sigma_{k,0}^*) + M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*)\right|\\ =&O_{\mathbb{P}}(\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F/\sqrt{N}).\end{aligned} \tag{S4.42}$$

Similarly, let $\Delta_{\Sigma_k^*, \widetilde{\Sigma}_{k,0}} = \Sigma_k^* + \widetilde{\Sigma}_{k,0} - \Sigma_k^* - \Sigma_{k,0}^* = \widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*$. We have

$$\begin{aligned}M_k(\Sigma_k^*, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k^*, \Sigma_{k,0}^*) &= O_{\mathbb{P}}(\|\Delta_{\Sigma_k^*, \widetilde{\Sigma}_{k,0}}\|_F^2) + O_{\mathbb{P}}(\|\Delta_{\Sigma_k^*, \widetilde{\Sigma}_{k,0}}\|_F/\sqrt{N})\\ &= O_{\mathbb{P}}(1/N)\end{aligned} \tag{S4.43}$$

Thus, we have

$$\begin{aligned}0 \le& M_{0,k}(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_{0,k}(\Sigma_k^*, \Sigma_{k,0}^*)\\ \le& M_k(\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k^*, \Sigma_{k,0}^*) + O_{\mathbb{P}}\left(\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F/\sqrt{N}\right)\\ \le& M_k(\Sigma_k^*, \widetilde{\Sigma}_{k,0}) - M_k(\Sigma_k^*, \Sigma_{k,0}^*) + O_{\mathbb{P}}\left(\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F/\sqrt{N}\right)\\ =& O_{\mathbb{P}}(1/N) + O_{\mathbb{P}}\left(\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F/\sqrt{N}\right).\end{aligned} \tag{S4.44}$$

Combined with (S4.40), we have $\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F^2 = O_{\mathbb{P}}(1/N) + O_{\mathbb{P}}(\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F/\sqrt{N})$, which leads to $\|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Furthermore, we have

$$\|\widehat{\Sigma}_k - \Sigma_k^*\|_F \le \|\Delta_{\widehat{\Sigma}_k, \widetilde{\Sigma}_{k,0}}\|_F + \|\widetilde{\Sigma}_{k,0} - \Sigma_{k,0}^*\|_F = O_{\mathbb{P}}(1/\sqrt{N}).$$

□

When the true hierarchical factor structure is known, the estimates of the loading matrix and the unique variance matrix are defined as follows:

$$\widehat{\Lambda}, \widehat{\Psi} = \arg\min_{\Lambda,\Psi} l(\Lambda\Lambda^\top + \Psi; S)$$
$$\text{s.t. } |\lambda_{ik}| \le \tau \text{ and } \lambda_{jk} = 0 \text{ for } k = 1, \ldots, K, i \in v_k^*, j \notin v_k^*,$$
$$\Psi = \text{diag}(\psi_1, \ldots, \psi_J), \kappa_1 \le |\psi_j| \le \kappa_2, j = 1, \ldots, J. \tag{S4.45}$$

Lemma 7. *Suppose that the hierarchical factor structure is known. If Conditions 1, 3, 5, 7 and 8 hold, we have*

$$\|\widehat{\Lambda} - \Lambda^*\widehat{Q}\|_F = O_{\mathbb{P}}(1/\sqrt{N}) \text{ and } \|\widehat{\Psi} - \Psi^*\|_F = O_{\mathbb{P}}(1/\sqrt{N}), \tag{S4.46}$$

where $\widehat{Q}$ is the diagonal matrix with diagonal entries consisting of the signs of the corresponding entries of $\widehat{\Lambda}^\top\Lambda^*$ defined in Theorem 2.

Proof. Similar to the proof of Lemma 6, we have $\|\widehat{\Lambda}\widehat{\Lambda}^\top + \widehat{\Psi} - \Sigma^*\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Furthermore, according to Lemma 4, we have $\|\widehat{\Lambda}\widehat{\Lambda}^\top - \Lambda^*\Lambda^{*\top}\|_F = O_{\mathbb{P}}(1/\sqrt{N})$ and $\|\widehat{\Psi} - \Psi^*\|_F = O_{\mathbb{P}}(1/\sqrt{N})$.

To prove that $\|\widehat{\Lambda} - \Lambda^*\widehat{Q}\|_F = O_{\mathbb{P}}(1/\sqrt{N})$, we first show that there exists some orthogonal rotation matrix R such that $\|\widehat{\Lambda} - \Lambda^* R\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Second, we show that $\|\widehat{\boldsymbol{\lambda}}_1 - \boldsymbol{\lambda}_1^*\text{sign}(\widehat{\boldsymbol{\lambda}}_1^\top\boldsymbol{\lambda}_1^*)\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Third, we conclude the proof by recursively applying the same argument to the factors in the tth layer, $t = 2, \ldots, T$.

Let $\Lambda^* = U^*\text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right) V^{*\top}$ be the singular value decomposition of Λ^* and $\widehat{\Lambda} = \widehat{U}\text{diag}\left(\sigma_1(\widehat{\Lambda}), \ldots, \sigma_K(\widehat{\Lambda})\right)\widehat{V}^\top$ be the singular value decomposition of $\widehat{\Lambda}$. Then $\Lambda^*\Lambda^{*\top} = U^*\text{diag}\left(\sigma_1^2(\Lambda^*), \ldots, \sigma_K^2(\Lambda^*)\right)U^{*\top}$ and $\widehat{\Lambda}\widehat{\Lambda}^\top = \widehat{U}\text{diag}\left(\sigma_1^2(\widehat{\Lambda}), \ldots, \sigma_K^2(\widehat{\Lambda})\right)\widehat{U}^\top$. By Lemma 2, $|\sigma_i^2(\Lambda^*) - \sigma_i^2(\widehat{\Lambda})| = O_{\mathbb{P}}(1/\sqrt{N})$ for $i = 1, \ldots, K$, which further leads to $|\sigma_i(\Lambda^*) - \sigma_i(\widehat{\Lambda})| = O_{\mathbb{P}}(1/\sqrt{N})$ for all i. By Lemma 3, there exits some orthogonal rotation matrix $\widetilde{R}$ such that $\|\widehat{U} - U^*\widetilde{R}\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Moreover, $\widetilde{R}$ satisfies $\widetilde{R}\text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right) = \text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right)\widetilde{R}$ with probability approaching 1 as N grows to infinity. Taking $R = V^*\widetilde{R}\widehat{V}^\top$, we have

$$\begin{aligned}
&\|\widehat{\Lambda} - \Lambda^* R\|_F \\
=&\left\|\widehat{U}\text{diag}\left(\sigma_1(\widehat{\Lambda}), \ldots, \sigma_K(\widehat{\Lambda})\right)\widehat{V}^\top - U^*\text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right)\widetilde{R}\widehat{V}^\top\right\|_F \\
=&\|\widehat{U}\text{diag}\left(\sigma_1(\widehat{\Lambda}), \ldots, \sigma_K(\widehat{\Lambda})\right) - U^*\widetilde{R}\text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right)\|_F \\
\leq&\left\|\widehat{U}\left(\text{diag}\left(\sigma_1(\widehat{\Lambda}), \ldots, \sigma_K(\widehat{\Lambda})\right) - \text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right)\right)\right\|_F + \left\|(\widehat{U} - U^*\widetilde{R}) \right. \\
&\left.\text{diag}\left(\sigma_1(\Lambda^*), \ldots, \sigma_K(\Lambda^*)\right)\right\|_F \\
=&O_{\mathbb{P}}(1/\sqrt{N}).
\end{aligned} \tag{S4.47}$$

For $i, j \in \text{Ch}_1^*$, $i \neq j$, by Lemma 3 and

$$\left\|\widehat{\Lambda}_{[v_i^*,\{1\}]}\widehat{\Lambda}_{[v_j^*,\{1\}]}^\top - \Lambda^*_{[v_i^*,\{1\}]}(\Lambda^*_{[v_j^*,\{1\}]})^\top\right\|_F = O_{\mathbb{P}}(1/\sqrt{N}),$$

we have

$$\left\|\frac{\widehat{\Lambda}_{[v_i^*,\{1\}]}}{\|\widehat{\Lambda}_{[v_i^*,\{1\}]}\|} - \frac{\Lambda^*_{[v_i^*,\{1\}]}}{\|\Lambda^*_{[v_i^*,\{1\}]}\|}\text{sign}\left(\widehat{\Lambda}_{[v_i^*,\{1\}]}^\top\Lambda^*_{[v_i^*,\{1\}]}\right)\right\| = O_{\mathbb{P}}(1/\sqrt{N}). \tag{S4.48}$$

Then, we further have

$$\left\|\frac{\widehat{\Lambda}_{[v_j^*,\{1\}]}}{\|\widehat{\Lambda}_{[v_j^*,\{1\}]}\|}-\frac{\Lambda^*_{[v_j^*,\{1\}]}}{\|\Lambda^*_{[v_j^*,\{1\}]}\|}\mathrm{sign}\left(\widehat{\Lambda}^\top_{[v_i^*,\{1\}]}\Lambda^*_{[v_i^*,\{1\}]}\right)\right\|=O_{\mathbb{P}}(1/\sqrt{N}) \qquad \text{(S4.49)}$$

for all $j\in\mathrm{Ch}_1^*$, which also leads to the fact that $\mathrm{sign}\left(\widehat{\Lambda}^\top_{[v_i^*,\{1\}]}\Lambda^*_{[v_i^*,\{1\}]}\right)=\mathrm{sign}(\widehat{\boldsymbol{\lambda}}_1^\top\boldsymbol{\lambda}_1^*)$ with probability approaching 1 as N grows to infinity. According to (S4.47), for each $i\in\mathrm{Ch}_1^*$, we have

$$\left\|\widehat{\Lambda}_{[v_i^*,\{1\}]}-\Lambda^*_{[v_i^*,\{1,i\}\cup D_i^*]}R_{[\{1,i\}\cup D_i^*,\{1\}]}\right\|_F=O_{\mathbb{P}}(1/\sqrt{N}).$$

Let

$$P_i=\frac{\Lambda^*_{[v_i^*,\{1\}]}\left(\Lambda^*_{[v_i^*,\{1\}]}\right)^\top}{\left(\Lambda^*_{[v_i^*,\{1\}]}\right)^\top\Lambda^*_{[v_i^*,\{1\}]}}$$

and $\Lambda^*_{\mathrm{Proj},i}=(\mathbf{I}-P_i)\Lambda^*_{[v_i^*,\{i\}\cup D_i^*]}$. By Condition 3, $\sigma_{1+|D_i^*|}\left(\Lambda^*_{\mathrm{Proj},i}\right)>0$. We have

$$\begin{aligned}
&\left\|\widehat{\Lambda}_{[v_i^*,\{1\}]}-\Lambda^*_{[v_i^*,\{1,i\}\cup D_i^*]}R_{[\{1,i\}\cup D_i^*,\{1\}]}\right\|\\
\geq&\left\|\frac{\|\widehat{\Lambda}_{[v_i^*,\{1\}]}\|}{\|\Lambda^*_{[v_i^*,\{1\}]}\|}\Lambda^*_{[v_i^*,\{1\}]}\mathrm{sign}(\widehat{\boldsymbol{\lambda}}_1^\top\boldsymbol{\lambda}_1^*)-\Lambda^*_{[v_i^*,\{1,i\}\cup D_i^*]}R_{[\{1,i\}\cup D_i^*,\{1\}]}\right\|\\
&-\|\widehat{\Lambda}_{[v_i^*,\{1\}]}\|\left\|\frac{\widehat{\Lambda}_{[v_i^*,\{1\}]}}{\|\widehat{\Lambda}_{[v_i^*,\{1\}]}\|}-\frac{\Lambda^*_{[v_i^*,\{1\}]}}{\|\Lambda^*_{[v_i^*,\{1\}]}\|}\mathrm{sign}(\widehat{\boldsymbol{\lambda}}_1^\top\boldsymbol{\lambda}_1^*)\right\|\\
\geq&\left\|R_{[\{i\}\cup D_i^*,\{1\}]}\right\|\sigma_{1+|D_i^*|}\left(\Lambda^*_{\mathrm{Proj},i}\right)+O_{\mathbb{P}}(1/\sqrt{N}).
\end{aligned}$$

Thus, we have $\left\|R_{[\{i\}\cup D_i^*,\{1\}]}\right\|=O_{\mathbb{P}}(1/\sqrt{N})$ for all $i\in\mathrm{Ch}_i^*$, which leads to $\left\|R_{[\{2,\ldots,K\},\{1\}]}\right\|=O_{\mathbb{P}}(1/\sqrt{N})$ and $\left|R_{[\{1\},\{1\}]}-\mathrm{sign}(\widehat{\boldsymbol{\lambda}}_1^\top\boldsymbol{\lambda}_1^*)\right|=O_{\mathbb{P}}(1/\sqrt{N})$.

We then have

$$
\begin{aligned}
&\left\|\widehat{\boldsymbol{\lambda}}_1 - \boldsymbol{\lambda}_1^* \mathrm{sign}(\widehat{\boldsymbol{\lambda}}_1^\top \boldsymbol{\lambda}_1^*)\right\| \\
\leq &\left\|\widehat{\boldsymbol{\lambda}}_1 - \Lambda^* R_{[:,\{1\}]}\right\| + \left|R_{[\{1\},\{1\}]} - \mathrm{sign}(\widehat{\boldsymbol{\lambda}}_1^\top \boldsymbol{\lambda}_1^*)\right| \left\|\boldsymbol{\lambda}_1^*\right\| \\
&+ \left\|R_{[\{2,\ldots,K\},\{1\}]}\right\| \left\|\Lambda^*_{[:,\{2,\ldots,K\}]}\right\|_F \\
= &O_{\mathbb{P}}(1/\sqrt{N}).
\end{aligned}
\tag{S4.50}
$$

Finally, with (S4.50), we have

$$
\|\widehat{\Lambda}_{[v_i^*,\{i\}\cup D_i^*]}\widehat{\Lambda}^\top_{[v_i^*,\{i\}\cup D_i^*]} - \Lambda^*_{[v_i^*,\{i\}\cup D_i^*]}\Lambda^{*\top}_{[v_i^*,\{i\}\cup D_i^*]}\|_F = O_{\mathbb{P}}(1/\sqrt{N}),
$$

for all $i \in \mathrm{Ch}_1^*$. Then, Lemma 7 can be proved by recursively applying the same argument to the factors in the tth layer, $t = 2, \ldots, T$. □

We now give the proof of Theorem 2.

Proof. The proof follows in a recursive manner. We first prove that with probability approaching 1 as N grows to infinity,

$$
\widehat{\mathrm{Ch}}_1 = \mathrm{Ch}_1^*, \text{ and } \widehat{v}_i = v_i^* \text{ for all } i \in \mathrm{Ch}_1^*, \tag{S4.51}
$$

and as a by-product, $\|\widetilde{\boldsymbol{\lambda}}_1 - \boldsymbol{\lambda}_1^* \mathrm{sign}(\boldsymbol{\lambda}_1^{*\top} \widetilde{\boldsymbol{\lambda}}_1)\| = O_{\mathbb{P}}(1/\sqrt{N})$, which further implies $\|\widetilde{\boldsymbol{\lambda}}_1 \widetilde{\boldsymbol{\lambda}}_1^\top - \boldsymbol{\lambda}_1^* \boldsymbol{\lambda}_1^{*\top}\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Then Theorem 2 is proved by applying the same argument to the factors in the tth layer, $t = 2, \ldots, T$, in conjunction with Lemma 7.

For simplicity of the notation, we denote $c^* = |\mathrm{Ch}_1^*|$. When $c^* = 0$, the proof of (S4.51) is trivial. When $c^* \geq 2$, the proof of (S4.51) consists of two main steps:

1. For sufficiently large d such that $\Sigma^* \in \Theta_1(c,d)$, let $\bar{\Lambda}_{1,c}$ and $\bar{\Psi}_{1,c}$ be the estimates according to (18) and $v_1^{1,c},\dots,v_c^{1,c}$ be the sets of variables belonging to the child factors of factor 1 decoded by $\bar{\Lambda}_{1,c}$. The possible configurations for $v_1^{1,c},\dots,v_c^{1,c}$ are:

   A. For each $k \in \mathrm{Ch}_1^*$, there exists some $s \in \{1,\dots,c\}$ such that $v_k^* \subset v_s^{1,c}$.

   B. There exists some $j \in \{1,\dots,J\}$ such that

   $$(v_1^{1,c}, v_2^{1,c}) = (\{1,\dots,J\} \setminus \{j\}, \{j\}) \text{ and } v_s^{1,c} = \emptyset \text{ for } s > 2.$$

2. $c^* = \arg\min_{c=0,2,\dots,c_{\max}} \widetilde{\mathrm{IC}}_{1,c}$ and $\widetilde{d}_s^{c^*} = 1 + |D_{1+s}^*|$ for $s = 1,\dots,c^*$ with probability approaching 1 as N grows to infinity.

Given $v_1^{1,c},\dots,v_c^{1,c}$, we prove the first part by showing that for arbitrary $\Lambda \in \widetilde{\mathcal{A}}^1(c,d_1\dots,d_c)$ and Ψ, there exists some constant $C > 0$ depending only on Λ^* such that $\|\Lambda\Lambda^\top + \Psi - \Sigma^*\|_F \geq C$ if $v_1^{1,c},\dots,v_c^{1,c}$ are not in case A or B. Let $\mathcal{B}_{i,s} = v_i^* \cap v_s^{1,c^*}$ for $i = 2,\dots,1+c^*$ and $s = 1,\dots,c$. We first claim that such a constant C exists if there exists $i \in \{2,\dots,1+c^*\}$ such that the following cases do not hold: (1) $\mathcal{B}_{i,s} = v_i^*$ for some $s \in \{1,\dots,c\}$ and (2) $v_i^* = \mathcal{B}_{i,s_1} \cup \mathcal{B}_{i,s_2}$, $\mathcal{B}_{i,s_1}, \mathcal{B}_{i,s_2} \neq \emptyset$ for $s_1, s_2 \in \{1,\dots,c\}$ and $|v_{s_2}^{1,c}| = 1$.

We then claim that such a constant C exists if the second case holds for some i but case B does not hold.

Now we give the proof of the first claim. Let $\Sigma = \Lambda\Lambda^\top + \Psi$. For $i = 2, \dots, 1 + c^*$, consider the following cases where $\mathrm{Ch}_i^* = \emptyset$:

1. $|\{s : |\mathcal{B}_{i,s}| \geq 1\}| \geq 4$. Let $s_1, \dots, s_4 \in \{1, \dots, c\}$ such that $|\mathcal{B}_{i,s_1}| \geq 1, \dots, |\mathcal{B}_{i,s_4}| \geq 1$ and $j_1 \in \mathcal{B}_{i,s_1}, \dots, j_4 \in \mathcal{B}_{i,s_4}$. We have

$$\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Lambda_{[\{j_1,j_2\},\{1\}]}(\Lambda_{[\{j_3,j_4\},\{1\}]})^\top,$$

has rank 1, while by Condition 3,

$$\Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Lambda^*_{[\{j_1,j_2\},\{1,i\}]}(\Lambda^*_{[\{j_3,j_4\},\{1,i\}]})^\top,$$

has rank 2. By Lemma 2, we have

$$\begin{aligned} &\|\Sigma - \Sigma^*\|_F \\ \geq &\left\|\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} - \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]}\right\|_F \\ \geq &\sigma_2\left(\Lambda^*_{[\{j_1,j_2\},\{1,i\}]}(\Lambda^*_{[\{j_1,j_2\},\{1,i\}]})^\top\right) \\ > &0. \end{aligned} \tag{S4.52}$$

2. There exists some $1 \leq s \leq c$ such that $|\mathcal{B}_{i,s}| \geq 2$ and $|v_i^* \setminus \mathcal{B}_{i,s}| \geq 2$. In this case, choose $j_1, j_2 \in \mathcal{B}_{i,s}$ and $j_3, j_4 \in v_i^* \setminus \mathcal{B}_{i,s}$, (S4.52) also holds.

3. There exists some $1 \leq s \leq c$ such $|\mathcal{B}_{i,s}| = 1$ and $|v_s^{1,c}| > 1$. Let

$\{j\} = \mathcal{B}_{i,s}$ and we have

$$\begin{aligned} &\|\Sigma - \Sigma^*\|_F \\ \geq &\frac{1}{\sqrt{2}}\left(\left\|\Sigma_{[v_i^*\backslash\{j\},\{j\}]} - \Sigma^*_{[v_i^*\backslash\{j\},\{j\}]}\right\| + \left\|\Sigma_{[v_i^*\backslash\{j\},v_s^{1,c}\backslash\{j\}]} - \Sigma^*_{[v_i^*\backslash\{j\},v_s^{1,c}\backslash\{j\}]}\right\|_F\right). \end{aligned} \tag{S4.53}$$

Notice that $\Sigma_{[v_i^*\backslash\{j\},v_s^{1,c}\backslash\{j\}]} = \Lambda_{[v_i^*\backslash\{j\},\{1\}]}(\Lambda_{[v_s^{1,c}\backslash\{j\},\{1\}]})^\top$, and $\Sigma^*_{[v_i^*\backslash\{j\},v_s^{1,c}\backslash\{j\}]} = \Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}(\Lambda^*_{[v_s^{1,c}\backslash\{j\},\{1\}]})^\top$. We denote by

$$\delta = \left\|\Sigma_{[v_i^*\backslash\{j\},v_s^{1,c}\backslash\{j\}]} - \Sigma^*_{[v_i^*\backslash\{j\},v_s^{1,c}\backslash\{j\}]}\right\|_F.$$

By Lemma 3,

$$\left\|\frac{\Lambda_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda_{[v_i^*\backslash\{j\},\{1\}]}\right\|} - \frac{\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|}\right\| \leq \frac{2^{3/2}\delta}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|\left\|\Lambda^*_{[v_s^{1,c}\backslash\{j\},\{1\}]}\right\|}, \tag{S4.54}$$

or

$$\left\|\frac{\Lambda_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda_{[v_i^*\backslash\{j\},\{1\}]}\right\|} + \frac{\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|}\right\| \leq \frac{2^{3/2}\delta}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|\left\|\Lambda^*_{[v_s^{1,c}\backslash\{j\},\{1\}]}\right\|}$$

holds. Without loss of generality, we assume that (S4.54) holds. On the other hand, notice that

$$\Sigma_{[v_i^*\backslash\{j\},\{j\}]} = \lambda_{j,1}\Lambda_{[v_i^*\backslash\{j\},\{1\}]},$$

and

$$\Sigma^*_{[v_i^*\backslash\{j\},\{j\}]} = \lambda^*_{j,1}\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]} + \lambda^*_{j,i}\Lambda^*_{[v_i^*\backslash\{j\},\{i\}]}.$$

Let

$$P_i = \frac{\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\left(\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\right)^\top}{\left(\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\right)^\top \Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}}$$

and $\boldsymbol{\mu} = (\mathbf{I} - P_i)\Lambda^*_{[v_i^* \setminus \{j\}, \{j\}]}$. According to Condition 3, $\boldsymbol{\mu} \neq \mathbf{0}$. According to Condition 8, we have

$$\begin{aligned}
&\left\|\Sigma_{[v_i^* \setminus \{j\}, \{j\}]} - \Sigma^*_{[v_i^* \setminus \{j\}, \{j\}]}\right\| \\
=&\left\|\lambda_{j,1}\Lambda_{[v_i^* \setminus \{j\}, \{1\}]} - \lambda^*_{j,1}\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]} - \lambda^*_{j,i}\Lambda^*_{[v_i^* \setminus \{j\}, \{i\}]}\right\| \\
\geq&|\lambda^*_{j,i}|\|\boldsymbol{\mu}\| - |\lambda_{j,1}| \left\|\Lambda_{[v_i^* \setminus \{j\}, \{1\}]} - \frac{\left\|\Lambda_{[v_i^* \setminus \{j\}, \{1\}]}\right\|}{\left\|\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\right\|}\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\right\| \\
\geq&|\lambda^*_{j,i}|\|\boldsymbol{\mu}\| - \frac{2^{3/2}\tau^2\delta(|v_i^*| - 1)^{1/2}}{\left\|\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\right\|\left\|\Lambda^*_{[v_s^{1,c} \setminus \{j\}, \{1\}]}\right\|}.
\end{aligned} \tag{S4.55}$$

Combining (S4.53) and (S4.55), we have

$$\|\Sigma - \Sigma^*\|_F \geq \min\left(\frac{\sqrt{2}}{4}, \frac{\left\|\Lambda^*_{[v_i^* \setminus \{j\}, \{1\}]}\right\|\left\|\Lambda^*_{[v_s^{1,c} \setminus \{j\}, \{1\}]}\right\|}{8\tau^2(|v_i^*| - 1)^{1/2}}\right)|\lambda^*_{j,i}|\|\boldsymbol{\mu}\| > 0.$$

4. $|v_i^*| = \cup_{k=1,2,3}\mathcal{B}_{i,s_k}$ with $\{j_k\} = \mathcal{B}_{i,s_k}$ for $k = 1, 2, 3$. If there exists some k such that $|v_{s_k}^{1,c}| > 1$, with a similar argument from (S4.53) to (S4.55), we have

$$\|\Sigma - \Sigma^*\|_F \geq \min\left(\frac{\sqrt{2}}{4}, \frac{\left\|\Lambda^*_{[v_i^* \setminus \{j_k\}, \{1\}]}\right\|\left\|\Lambda^*_{[v_{s_k}^{1,c} \setminus \{j_k\}, \{1\}]}\right\|}{8\tau^2(|v_i^*| - 1)^{1/2}}\right)|\lambda^*_{j_k,i}|\|\boldsymbol{\mu}\| > 0,$$

where $\boldsymbol{\mu}$ is defined similarly in (S4.55). Otherwise, $\{j_k\} = v_{s_k}^{1,c}$ for

$k = 1, 2, 3$. Consider $i' \in \mathrm{Ch}_i^*$ and $i' \neq i$. We have

$$\begin{aligned} &\|\Sigma - \Sigma^*\|_F \\ \geq &\frac{1}{\sqrt{2}}\left(\left\|\Sigma_{[v_i^* \setminus \{j_k\}, \{j_k\}]} - \Sigma^*_{[v_i^* \setminus \{j_k\}, \{j_k\}]}\right\| + \left\|\Sigma_{[v_i^*, v_{i'}^*]]} - \Sigma^*_{[v_i^*, v_{i'}^*]]}\right\|_F\right). \end{aligned} \tag{S4.56}$$

With a similar argument from (S4.53) to (S4.55), we have

$$\|\Sigma - \Sigma^*\|_F \geq \min\left(\frac{\sqrt{2}}{4}, \frac{\left\|\Lambda^*_{[v_i^* \setminus \{j_k\}, \{1\}]}\right\| \left\|\Lambda^*_{[v_{i'}^*, \{1\}]}\right\|}{8\tau^2(|v_i^*| - 1)^{1/2}}\right) |\lambda^*_{j_k, i}| \|\boldsymbol{\mu}_k\| > 0,$$

where $\boldsymbol{\mu}_k$s are defined similarly in (S4.55) for $k = 1, 2, 3$.

5. The rest of the cases are included in the two cases of our first claim.

When $\mathrm{Ch}_i^* \neq \emptyset$, consider the following cases:

1. There exist $k \in \mathrm{Ch}_i^*$ and $s = 1, \ldots, c$ such that $|\mathcal{B}_{i,s} \cap v_k^*| \geq 2$. If we further have

$$|(\cup_{1 \leq s' \leq c, s' \neq s} \mathcal{B}_{i,s'}) \cap (\cup_{k' \neq k, k' \in \mathrm{Ch}_i^*} v_{k'}^*)| \geq 2,$$

choose $j_1, j_2 \in \mathcal{B}_{i,s} \cap v_k^*$ and $j_3, j_4 \in (\cup_{1 \leq s' \leq c^*, s' \neq s} \mathcal{B}_{i,s'}) \cap (\cup_{k' \neq k, k' \in \mathrm{Ch}_1^*} v_{k'}^*)$. We have

$$\Sigma_{[\{j_1, j_2\}, \{j_3, j_4\}]} = \Lambda_{[\{j_1, j_2\}, \{1\}]} (\Lambda_{[\{j_3, j_4\}, \{1\}]})^\top,$$

which has rank 1, while by Condition 3

$$\Sigma^*_{[\{j_1, j_2\}, \{j_3, j_4\}]} = \Lambda^*_{[\{j_1, j_2\}, \{1, i\}]} (\Lambda^*_{[\{j_3, j_4\}, \{1, i\}]})^\top,$$

has rank 2. By Lemma 2, we also have

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq &\big\|\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} - \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]}\big\|_F \\
\geq &\sigma_2\big(\Lambda^*_{[\{j_1,j_2\},\{1,i\}]}(\Lambda^*_{[\{j_1,j_2\},\{1,i\}]})^\top\big) \\
> &0.
\end{aligned}$$

If

$$|(\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap (\cup_{k'\neq k, k'\in \mathrm{Ch}_i^*} v_{k'}^*)| \leq 1, \tag{S4.57}$$

since $|v_{k'}^*| \geq 3$ for $k' \neq k, k' \in \mathrm{Ch}_i^*$, by (S4.57) we also have $|\mathcal{B}_{i,s} \cap v_{k'}^*| \geq 2$ for all $k' \in \mathrm{Ch}_i^*$. Similar to (S4.57), we have

$$|(\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap v_k^*| \leq 1. \tag{S4.58}$$

Combining (S4.57) and (S4.58), we have

$$|(\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap (\cup_{k'\in \mathrm{Ch}_i^*} v_{k'}^*)| \leq 2.$$

First, if $|(\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap (\cup_{k'\in \mathrm{Ch}_i^*} v_{k'}^*)| = 2$, we denote by $k' \neq k$ such that (S4.57) is tight. Choose $j_1, j_2 \in \mathcal{B}_{i,s} \cap v_k^*$, $j_3, j_4 \in \mathcal{B}_{i,s} \cap v_{k'}^*$, $j_5 \in (\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap v_k^*$ and $j_6 \in (\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap v_{k'}^*$. Furthermore, we require that when $\mathrm{Ch}_k^* \neq \emptyset$, j_1, j_2 belong to different child factors of factor k with j_5 and when $\mathrm{Ch}_{k'}^* \neq \emptyset$, j_3, j_4 belong to different child factors of factor k' with j_6. Such a choice is always possible due to the assumed structure of the hierarchical model. It is easy to check that

$$\Sigma_{[\{j_1,j_2,j_3,j_4\},\{j_5,j_6\}]} = \Lambda_{[\{j_1,j_2,j_3,j_4\},\{1\}]}(\Lambda_{[\{j_5,j_6\},\{1\}]})^\top$$

has rank 1. On the other hand,

$$\Sigma^*_{[\{j_1,j_2,j_3,j_4\},\{j_5,j_6\}]} = \Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,i,k,k'\}]}(\Lambda^*_{[\{j_5,j_6\},\{1,i,k,k'\}]})^\top.$$

According to Condition 3, the rank of $\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,i,k,k'\}]}$ is 4 and the rank of $\Lambda^*_{[\{j_5,j_6\},\{1,i,k,k'\}]}$ is 2. By Sylvester's rank inequality,

$$\begin{aligned}
&\text{rank}\big(\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,i,k,k'\}]}(\Lambda^*_{[\{j_5,j_6\},\{1,i,k,k'\}]})^\top\big)\\
\geq&\text{rank}\big(\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,i,k,k'\}]}\big) + \text{rank}\big(\Lambda^*_{[\{j_5,j_6\},\{1,i,k,k'\}]}\big) - 4\\
=&2.
\end{aligned}$$

Thus, by Lemma 2,

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F\\
\geq&\big\|\Sigma_{[\{j_1,j_2,j_3,j_4\},\{j_5,j_6\}]} - \Sigma^*_{[\{j_1,j_2,j_3,j_4\},\{j_5,j_6\}]}\big\|_F\\
\geq&\sigma_2\big(\Lambda^*_{[\{j_1,j_2,j_3,j_4\},\{1,i,k,k'\}]}(\Lambda^*_{[\{j_5,j_6\},\{1,i,k,k'\}]})^\top\big)\\
>&0.
\end{aligned}$$

Second, if $|(\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap (\cup_{k'\in \text{Ch}^*_i} v^*_{k'})| = 1$, let $(\cup_{1\leq s'\leq c, s'\neq s}\mathcal{B}_{i,s'}) \cap (\cup_{k'\in \text{Ch}^*_i} v^*_{k'}) = \mathcal{B}_{i,s_1} \cap v^*_{k_1} = \{j\}$ without loss of generality. When $|v^{1,c}_{s_1}| = 1$, the second case of our first claim holds. Otherwise, it is easy to check that

$$\begin{aligned}&\|\Sigma-\Sigma^*\|_F\\ \geq&\frac{1}{\sqrt{2}}\left(\left\|\Sigma_{[v_i^*\backslash\{j\},\{j\}]}-\Sigma^*_{[v_i^*\backslash\{j\},\{j\}]}\right\|+\left\|\Sigma_{[v_i^*\backslash\{j\},v_{s_1}^{1;c}\backslash\{j\}]}-\Sigma^*_{[v_i^*\backslash\{j\},v_{s_1}^{1;c}\backslash\{j\}]}\right\|_F\right).\end{aligned} \tag{S4.59}$$

Notice that

$$\Sigma_{[v_i^*\backslash\{j\},v_{s_1}^{1;c}\backslash\{j\}]}=\Lambda_{[v_i^*\backslash\{j\},\{1\}]}(\Lambda_{[v_{s_1}^{1;c}\backslash\{j\},\{1\}]})^\top,$$

while

$$\Sigma^*_{[v_i^*\backslash\{j\},v_{s_1}^{1;c}\backslash\{j\}]}=\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}(\Lambda^*_{[v_{s_1}^{1;c}\backslash\{j\},\{1\}]})^\top.$$

We denote by $\delta=\left\|\Sigma_{[v_i^*\backslash\{j\},v_{s_1}^{1;c}\backslash\{j\}]}-\Sigma^*_{[v_i^*\backslash\{j\},v_{s_1}^{1;c}\backslash\{j\}]}\right\|_F$. By Lemma 3, either

$$\left\|\frac{\Lambda_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda_{[v_i^*\backslash\{j\},\{1\}]}\right\|}-\frac{\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|}\right\|\leq\frac{2^{3/2}\delta}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|\left\|\Lambda^*_{[v_{s_1}^{1;c}\backslash\{j\},\{1\}]}\right\|}, \tag{S4.60}$$

or

$$\left\|\frac{\Lambda_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda_{[v_i^*\backslash\{j\},\{1\}]}\right\|}+\frac{\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|}\right\|\leq\frac{2^{3/2}\delta}{\left\|\Lambda^*_{[v_i^*\backslash\{j\},\{1\}]}\right\|\left\|\Lambda^*_{[v_{s_1}^{1;c}\backslash\{j\},\{1\}]}\right\|}.$$

holds. Without loss of generality, we assume that (S4.60) holds. On the other hand, notice that

$$\Sigma_{[v_i^*\backslash\{j\},\{j\}]}=\lambda_{j,1}\Lambda_{[v_i^*\backslash\{j\},\{1\}]},$$

and

$$\Sigma^*_{[v_i^*\backslash\{j\},\{j\}]}=\Lambda^*_{[v_i^*\backslash\{j\},\{1,i\}\cup D_i^*]}(\Lambda^*_{[\{j\},\{1,i\}\cup D_i^*]})^\top.$$

Let

$$P_i = \frac{\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\left(\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\right)^\top}{\left(\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\right)^\top \Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}}$$

and $\Lambda^*_{\mathrm{Proj},i} = (\mathbf{I} - P_i)\Lambda^*_{[v_i^*\setminus\{j\},\{i\}\cup D_i^*]}$. By Condition 3, $\sigma_{1+|D_i^*|}\left(\Lambda^*_{\mathrm{Proj},i}\right) > 0$. By Condition 8,

$$\begin{aligned}
&\left\|\Sigma_{[v_i^*\setminus\{j\},\{j\}]} - \Sigma^*_{[v_i^*\setminus\{j\},\{j\}]}\right\| \\
\geq& -|\lambda_{j,1}| \left\|\Lambda_{[v_i^*\setminus\{j\},\{1\}]} - \frac{\left\|\Lambda_{[v_i^*\setminus\{j\},\{1\}]}\right\|}{\left\|\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\right\|}\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\right\| \\
&+ \left\|\Lambda^*_{[\{j\},\{i\}\cup D_i^*]}\right\|\sigma_{1+|D_i^*|}\left(\Lambda^*_{\mathrm{Proj},i}\right) \\
\geq& \left\|\Lambda^*_{[\{j\},\{i\}\cup D_i^*]}\right\|\sigma_{1+|D_i^*|}\left(\Lambda^*_{\mathrm{Proj},i}\right) - \frac{2^{3/2}\tau^2\delta(|v_i^*|-1)^{1/2}}{\left\|\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\right\|\left\|\Lambda^*_{[v_{s_1}^{1,c}\setminus\{j\},\{1\}]}\right\|}.
\end{aligned} \tag{S4.61}$$

Combining (S4.59) and (S4.61), we have

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq& \min\left(\frac{\sqrt{2}}{4}, \frac{\left\|\Lambda^*_{[v_i^*\setminus\{j\},\{1\}]}\right\|\left\|\Lambda^*_{[v_s^{1,c}\setminus\{j\},\{1\}]}\right\|}{8\tau^2(|v_i^*|-1)^{1/2}}\right)\left\|\Lambda^*_{[\{j\},\{i\}\cup D_i^*]}\right\|\sigma_{1+|D_i^*|}\left(\Lambda^*_{\mathrm{Proj},i}\right) \\
>& 0.
\end{aligned} \tag{S4.62}$$

Finally, when $|(\cup_{1\leq s'\leq c^*, s'\neq s}\mathcal{B}_{i,s'}) \cap (\cup_{k'\in \mathrm{Ch}_i^*} v^*_{k'})| = 0$, the first case of our first claim holds.

2. $|\mathcal{B}_{i,s} \cap v_k^*| \leq 1$ for all $1 \leq s \leq c$ and $k \in \mathrm{Ch}_i^*$. First, consider the case when there exist some $1 \leq s \leq c$ such that $|\mathcal{B}_{i,s} \cap v^*_{k_1}| = 1$ and $|\mathcal{B}_{i,s} \cap v^*_{k_2}| = 1$ for some $k_1, k_2 \in \mathrm{Ch}_i^*$. we denote $\{j_1\} = \mathcal{B}_{i,s} \cap v^*_{k_1}$

and $\{j_2\} = \mathcal{B}_{i,s} \cap v_{k_2}^*$. Moreover, choose $j_3, j_4 \in v_{k_1}^*$ and $j_3, j_4 \in v_{k_2}^*$. Furthermore, if $|\mathrm{Ch}_{k_1}^*| \neq 0$, we further require that j_3, j_4 belong to different child factors of factor k_1 with j_1. Similarly, if $|\mathrm{Ch}_{k_2}^*| \neq 0$, j_5, j_6 belong to different child factor of factor k_2 with j_2. Such a choice is always possible due to the assumed structure of the hierarchical model. It is easy to check that

$$\Sigma_{[\{j_1,j_2\},\{j_3,j_4,j_5,j_6\}]} = \Lambda_{[\{j_1,j_2\},\{1\}]}(\Lambda_{[\{j_3,j_4,j_5,j_6\},\{1\}]})^\top,$$

has rank 1. By Condition 3, $\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,i,k_1,k_2\}]}$ has rank 4 and $\Lambda^*_{[\{j_1,j_2\},\{1,i,k_1,k_2\}]}$ has rank 2. By Sylvester's rank inequality,

$$\begin{aligned}
&\mathrm{rank}\big(\Lambda^*_{[\{j_1,j_2\},\{1,i,k_1,k_2\}]}(\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,i,k_1,k_2\}]})^\top\big) \\
\geq&\mathrm{rank}\big(\Lambda^*_{[\{j_1,j_2\},\{1,i,k_1,k_2\}]}\big) + \mathrm{rank}\big(\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,i,k_1,k_2\}]}\big) - 4 \\
=&2.
\end{aligned}$$

By Lemma 2,

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq&\big\|\Sigma_{[\{j_1,j_2\},\{j_3,j_4,j_5,j_6\}]} - \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4,j_5,j_6\}]}\big\|_F \\
\geq&\sigma_2\big(\Lambda^*_{[\{j_1,j_2\},\{1,i,k_1,k_2\}]}(\Lambda^*_{[\{j_3,j_4,j_5,j_6\},\{1,i,k_1,k_2\}]})^\top\big) \\
>&0.
\end{aligned} \tag{S4.63}$$

Second, for each $1 \leq s \leq c$, if $|\mathcal{B}_{i,s} \cap v_k^*| = 1$ for some $k \in \mathrm{Ch}_i^*$, $|\mathcal{B}_{i,s} \cap v_{k'}^*| = 0$ for all $k' \in \mathrm{Ch}_i^*$, $k' \neq k$, which indicates $|\mathcal{B}_{i,s} \cap v_i^*| \leq 1$

for $1 \leq s \leq c$. Since $|v_i^*| \geq 7$ by constraint C4, choose s_1, s_2, s_3, s_4 such that $\{j_k\} = \mathcal{B}_{i,s_k} \cap v_i^*$ for $k = 1, \ldots, 4$. Moreover, we require that j_1, j_2 and j_3, j_4 belong to different child factors of Factor i. We have

$$\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Lambda_{[\{j_1,j_2\},\{1\}]}(\Lambda_{[\{j_3,j_4\},\{1\}]})^\top,$$

has rank 1, while by Condition 3,

$$\Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]} = \Lambda^*_{[\{j_1,j_2\},\{1,i\}]}(\Lambda^*_{[\{j_3,j_4\},\{1,i\}]})^\top,$$

has rank 2. By Lemma 2, we have

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq &\left\|\Sigma_{[\{j_1,j_2\},\{j_3,j_4\}]} - \Sigma^*_{[\{j_1,j_2\},\{j_3,j_4\}]}\right\|_F \\
\geq &\sigma_2\big(\Lambda^*_{[\{j_1,j_2\},\{1,i\}]}(\Lambda^*_{[\{j_3,j_4\},\{1,i\}]})^\top\big) \\
\geq &0.
\end{aligned}$$

Now the first claim is proved, and we focus on our second claim. We assume that there exist $i_1 \in \{2, \ldots, 1 + c^*\}$ and $s_1, s_2 \in \{1, \ldots, c\}$ and $s_1 \neq s_2$ that satisfy $v_{i_1}^* = \mathcal{B}_{i_1,s_1} \cup \mathcal{B}_{i_1,s_2}$ and $v_{s_2}^{1,c} = \{j_1\}$ for some $j_1 \in \{1, \ldots, J\}$. Furthermore, for each $i_2 \in \{2, \ldots, 1+c^*\}$ and $i_2 \neq i_1$, we denote by $v_{i_2}^* = \mathcal{B}_{i_2,s_3} \cup \mathcal{B}_{i_2,s_4}$ for some $s_3, s_4 \in \{1, \ldots, c\}$ that satisfy $\mathcal{B}_{i_2,s_4} = \emptyset$ or $\mathcal{B}_{i_2,s_4} = v_{s_4}^{1,c} = \{j_2\}$ for some $j_2 \in \{1, \ldots, J\}$. We will first show that $s_3 = s_1$ for all $i_2 \neq i_1$ and second, show that $\mathcal{B}_{i_2,s_4} = \emptyset$ for all $i_2 \neq i_1$, which finally leads to case B.

First, when $s_3 \neq s_1$ for some i_2, it is easy to check that

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq &\frac{1}{\sqrt{2}} \left(\left\| \Sigma_{[v_{i_1}^* \setminus \{j_1\}, \{j_1\}]} - \Sigma^*_{[v_{i_1}^* \setminus \{j_1\}, \{j_1\}]} \right\| + \left\| \Sigma_{[v_i^* \setminus \{j_1\}, \mathcal{B}_{i_2, s_3}]} - \Sigma^*_{[v_i^* \setminus \{j_1\}, \mathcal{B}_{i_2, s_3}]} \right\|_F \right).
\end{aligned} \tag{S4.64}$$

Similarly to the proof in (S4.59) to (S4.62), we have

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq &\min \left(\frac{\sqrt{2}}{4}, \frac{\left\| \Lambda^*_{[v_{i_1}^* \setminus \{j_1\}, \{1\}]} \right\| \left\| \Lambda^*_{[\mathcal{B}_{i_2, s_3}, \{1\}]} \right\|}{8\tau^2 (|v_i^*| - 1)^{1/2}} \right) \left\| \Lambda^*_{[\{j_1\}, \{i_1\} \cup D_{i_1}^*]} \right\| \sigma_{1+|D_{i_1}^*|} (\Lambda^*_{\mathrm{Proj}, i_1}) \\
> &0.
\end{aligned} \tag{S4.65}$$

Second, when $\mathcal{B}_{i_2, s_4} \neq \emptyset$ for some i_2

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq &\frac{1}{\sqrt{2}} \left(\left\| \Sigma_{[v_{i_1}^* \setminus \{j_1\}, \{j_1\}]} - \Sigma^*_{[v_{i_1}^* \setminus \{j_1\}, \{j_1\}]} \right\| + \left\| \Sigma_{[v_i^* \setminus \{j_1\}, \mathcal{B}_{i_2, s_4}]} - \Sigma^*_{[v_i^* \setminus \{j_1\}, \mathcal{B}_{i_2, s_4}]} \right\|_F \right).
\end{aligned} \tag{S4.66}$$

Again, similarly to the proof in (S4.59) to (S4.62), we have

$$\begin{aligned}
&\|\Sigma - \Sigma^*\|_F \\
\geq &\min \left(\frac{\sqrt{2}}{4}, \frac{\left\| \Lambda^*_{[v_{i_1}^* \setminus \{j_1\}, \{1\}]} \right\| \left\| \Lambda^*_{[\mathcal{B}_{i_2, s_4}, \{1\}]} \right\|}{8\tau^2 (|v_i^*| - 1)^{1/2}} \right) \left\| \Lambda^*_{[\{j_1\}, \{i_1\} \cup D_{i_1}^*]} \right\| \sigma_{1+|D_{i_1}^*|} (\Lambda^*_{\mathrm{Proj}, i_1}) \\
> &0.
\end{aligned} \tag{S4.67}$$

Now we have finished the first part of our proof.

For the second part, we mainly focus on case A and omit the proof when $v_1^{1,c}, \ldots, v_c^{1,c}$ are included in case B for two reasons: (1) case B does not satisfy constraint C4 and will never be selected in our algorithms and (2) by a similar argument below, the information criterion brought by such case will be strictly larger than the optimal solution with probability approaching 1 as N grows to infinity. We also assume that $d_{\max}$ is sufficiently large to avoid further discussions.

Now, we focus on case A and we only discuss the case when $v_s^{1,c}$ are nonempty for $s = 1, \ldots, c$. First, we show that when $c = c^*$ in case A,

$$\widetilde{\mathrm{IC}}_{1,c^*} = \sum_{k \in \mathrm{Ch}_1^*} (|v_k^*|(|D_k^*| + 1) - |D_k^*|(|D_k^*| + 1)/2) \log N + O_{\mathbb{P}}(1). \quad \text{(S4.68)}$$

In such a case, we have $v_s^{1,c^*} = v_{1+s}^*$ for $s = 1, \ldots, c^*$. We claim that Step 6 of Algorithm 2 outputs $\widetilde{d}_s^{c^*} = 1 + |D_{1+s}^*|$ for $s = 1, \ldots, c^*$ with probability approaching 1 as N grows to infinity. When $s = 1$, for $d_1 \geq 1 + |D_2^*|$, let $\underline{\Lambda}_{d_1}$ and $\underline{\Psi}_{d_1}$ be the solution to

$$\widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_3^*|, d), \ldots, \min(|v_{1+c^*}^*|, d)). \quad \text{(S4.69)}$$

Similar to the proof of Lemma 6, $\left\| \underline{\Lambda}_{d_1} \underline{\Lambda}_{d_1}^\top + \underline{\Psi}_{d_1} - \Sigma^* \right\|_F = O_{\mathbb{P}}(1/\sqrt{N})$, and we further have

$$
\begin{aligned}
&\widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_3^*|, d), \ldots, \min(|v_{c^*}^*|, d)) \\
=&l\big(\underline{\Lambda}_{d_1}\underline{\Lambda}_{d_1}^\top + \underline{\Psi}_{d_1}; S\big) + p_1\big(\underline{\Lambda}_{d_1}\big)\log N \\
=&O_{\mathbb{P}}(1) + \big(|v_2^*|d_1 - d_1(d_1-1)/2\big)\log N + \sum_{2\le s\le c^*}\big(|v_{s+1}^*|d_s - d_s(d_s-1)/2\big)\log N,
\end{aligned}
\tag{S4.70}
$$

where we define $d_s = \min(|v_{1+s}^*|, d)$, $s = 2, \ldots, c^*$ for simplicity. Noticing that the third term of (S4.70) is independent of the choice of d_1 and the second term is strictly increasing with respect to d_1 for $1 + |D_2^*| \le d_1 \le \min(|v_2^*|, d)$, we then have

$$
1 + |D_2^*| = \operatorname*{arg\,min}_{1+|D_2^*|\le d_1\le \min(|v_2^*|,d)} \widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_3^*|, d), \ldots, \min(|v_{c^*}^*|, d)),
\tag{S4.71}
$$

with probability approaching 1 as N grows to infinity.

When $d_1 < 1+|D_2^*|$, for any $\Lambda \in \widetilde{\mathcal{A}}^1(c^*, d_1, \min(|v_3^*|, d), \ldots, \min(|v_{1+c^*}^*|, d))$ and Ψ, we denote by $\Sigma = \Lambda\Lambda^\top + \Psi$. According to Condition 6, there exist $E_1, E_2 \subset v_2^*$ with $|E_1| = 2 + |D_2^*|$, $|E_2| = 1 + |D_2^*|$ and $E_1 \cap E_2 = \emptyset$ such that $\Lambda^*_{[E_1,\{1,2\}\cup D_2^*]}$ and $\Lambda^*_{[E_2,\{2\}\cup D_2^*]}$ are of full rank. We further denote by $B_1 = \{2, \ldots, 1 + d_1\}$. First we have

$$
\|\Sigma - \Sigma^*\|_F \ge \frac{1}{\sqrt{2}}\left(\big\|\Sigma_{[v_2^*,v_i^*]} - \Sigma^*_{[v_2^*,v_i^*]}\big\|_F + \big\|\Sigma_{[E_1,E_2]} - \Sigma^*_{[E_1,E_2]}\big\|_F\right),
\tag{S4.72}
$$

for any $i = 3, \dots, 1 + c^*$.We denote by $\delta = \left\|\Sigma_{[v_2^*, v_i^*]} - \Sigma^*_{[v_2^*, v_i^*]}\right\|_F$. Notice that

$$\Sigma_{[v_2^*, v_i^*]} = \Lambda_{[v_2^*, \{1\}]}(\Lambda_{[v_i^*, \{1\}]})^\top,$$

and

$$\Sigma^*_{[v_2^*, v_i^*]} = \Lambda^*_{[v_2^*, \{1\}]}(\Lambda^*_{[v_i^*, \{1\}]})^\top.$$

According to Lemma 3, either

$$\left\| \frac{\Lambda_{[v_2^*, \{1\}]}}{\left\|\Lambda_{[v_2^*, \{1\}]}\right\|} - \frac{\Lambda^*_{[v_2^*, \{1\}]}}{\left\|\Lambda^*_{[v_2^*, \{1\}]}\right\|} \right\| \leq \frac{2^{3/2}\delta}{\left\|\Lambda^*_{[v_2^*, \{1\}]}\right\| \left\|\Lambda^*_{[v_i^*, \{1\}]}\right\|}, \tag{S4.73}$$

or

$$\left\| \frac{\Lambda_{[v_2^*, \{1\}]}}{\left\|\Lambda_{[v_2^*, \{1\}]}\right\|} + \frac{\Lambda^*_{[v_2^*, \{1\}]}}{\left\|\Lambda^*_{[v_2^*, \{1\}]}\right\|} \right\| \leq \frac{2^{3/2}\delta}{\left\|\Lambda^*_{[v_2^*, \{1\}]}\right\| \left\|\Lambda^*_{[v_i^*, \{1\}]}\right\|},$$

holds. Without loss of generality, we assume that (S4.73) holds. On the other hand, notice that

$$\begin{aligned}
&\Sigma_{[E_1, E_2]} - \Sigma^*_{[E_1, E_2]} \\
=&\Lambda_{[E_1, \{1\}]}(\Lambda_{[E_2, \{1\}]})^\top + \Lambda_{[E_1, B_1]}(\Lambda_{[E_2, B_1]})^\top - \Lambda^*_{[E_1, \{1\}]}(\Lambda^*_{[E_2, \{1\}]})^\top \\
&- \Lambda^*_{[E_1, \{2\} \cup D_2^*]}(\Lambda^*_{[E_2, \{2\} \cup D_2^*]})^\top \\
=&\Lambda_{[E_1, \{1\}]}(\Lambda_{[E_2, \{1\}]})^\top - \frac{\left\|\Lambda_{[v_2^*, \{1\}]}\right\|^2}{\left\|\Lambda^*_{[v_2^*, \{1\}]}\right\|^2}\Lambda^*_{[E_1, \{1\}]}(\Lambda^*_{[E_2, \{1\}]})^\top \\
&+ \Lambda_{[E_1, B_1]}(\Lambda_{[E_2, B_1]})^\top - \left(1 - \frac{\left\|\Lambda_{[v_2^*, \{1\}]}\right\|^2}{\left\|\Lambda^*_{[v_2^*, \{1\}]}\right\|^2}\right)\Lambda^*_{[E_1, \{1\}]}(\Lambda^*_{[E_2, \{1\}]})^\top \\
&- \Lambda^*_{[E_1, \{2\} \cup D_2^*]}(\Lambda^*_{[E_2, \{2\} \cup D_2^*]})^\top.
\end{aligned} \tag{S4.74}$$

Combined with (S4.73), we have

$$
\begin{aligned}
&\left\|\Lambda_{[E_1,\{1\}]}(\Lambda_{[E_2,\{1\}]})^\top - \frac{\left\|\Lambda_{[v_2^*,\{1\}]}\right\|^2}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|^2}\Lambda^*_{[E_1,\{1\}]}(\Lambda^*_{[E_2,\{1\}]})^\top\right\|_F \\
\leq &\left\|\left(\Lambda_{[E_1,\{1\}]} - \frac{\left\|\Lambda_{[v_2^*,\{1\}]}\right\|}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|}\Lambda^*_{[E_1,\{1\}]}\right)(\Lambda_{[E_2,\{1\}]})^\top\right\|_F \\
&+ \frac{\left\|\Lambda_{[v_2^*,\{1\}]}\right\|}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|}\left\|\Lambda^*_{[E_1,\{1\}]}\left(\Lambda_{[E_2,\{1\}]} - \frac{\left\|\Lambda_{[v_2^*,\{1\}]}\right\|}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|}\Lambda^*_{[E_2,\{1\}]}\right)^\top\right\|_F \\
\leq &\frac{2^{3/2}\delta\left\|\Lambda_{[v_2^*,\{1\}]}\right\|}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|\left\|\Lambda^*_{[v_i^*,\{1\}]}\right\|}\left(\left\|\Lambda_{[E_2,\{1\}]}\right\| + \frac{\left\|\Lambda_{[v_2^*,\{1\}]}\right\|}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|}\left\|\Lambda^*_{[E_1,\{1\}]}\right\|\right) \\
\leq &\frac{2^{5/2}\tau^2|v_2^*|\delta}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|\left\|\Lambda^*_{[v_i^*,\{1\}]}\right\|}.
\end{aligned}
\tag{S4.75}
$$

We denote by

$$
\Lambda_E^\top = \left(\Lambda^*_{[E_1,\{1,2\}\cup D_2^*]}\right)^{-1}\Lambda_{[E_1,B_1]}(\Lambda_{[E_2,B_1]})^\top,
$$

whose rank is at most $d_1 < 1 + |D_2^*|$, and

$$
\Lambda_E^* = \left(\left(1 - \frac{\left\|\Lambda_{[v_2^*,\{1\}]}\right\|^2}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|^2}\right)\Lambda^*_{[E_2,\{1\}]}, \quad \Lambda^*_{[E_2,\{2\}\cup D_2^*]} \right).
$$

By Condition 6, $\Lambda^*_{[E_2,\{2\}\cup D_2^*]}$ has rank $1 + |D_2^*|$. Thus, by Lemma 2

$$
\begin{aligned}
&\left\|\Lambda_E - \Lambda_E^*\right\|_F \\
\geq &\left\|\left(\Lambda_E\right)_{[:,B_1]} - \left(\Lambda_E^*\right)_{[:,B_1]}\right\|_F \\
\geq &\sigma_{1+|D_2^*|}\left(\Lambda^*_{[E_2,\{2\}\cup D_2^*]}\right).
\end{aligned}
\tag{S4.76}
$$

Combined with (S4.74), (S4.75) and (S4.76), we have

$$\begin{aligned}&\left\|\Sigma_{[E_1,E_2]}-\Sigma^*_{[E_1,E_2]}\right\|_F\\ \geq&\sigma_{2+|D_2^*|}\big(\Lambda^*_{[E_1,\{1,2\}\cup D_2^*]}\big)\sigma_{1+|D_2^*|}\big(\Lambda^*_{[E_2,\{2\}\cup D_2^*]}\big)-\frac{2^{5/2}\tau^2|v_2^*|\delta}{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|\left\|\Lambda^*_{[v_i^*,\{1\}]}\right\|}.\end{aligned} \tag{S4.77}$$

Combined with (S4.72) we further have

$$\begin{aligned}&\|\Sigma-\Sigma^*\|_F\\ \geq&\min\left(\frac{\sqrt{2}}{4},\frac{\left\|\Lambda^*_{[v_2^*,\{1\}]}\right\|\left\|\Lambda^*_{[v_i^*,\{1\}]}\right\|}{16\tau^2|v_2^*|}\right)\sigma_{2+|D_2^*|}\big(\Lambda^*_{[E_1,\{1,2\}\cup D_2^*]}\big)\sigma_{1+|D_2^*|}\big(\Lambda^*_{[E_2,\{2\}\cup D_2^*]}\big).\end{aligned}$$

Thus, the derived information criterion satisfies

$$\widetilde{\mathrm{IC}}_1(c,d_1,\min(|v_3^*|,d),\ldots,\min(|v_{c^*}^*|,d))=O_{\mathbb{P}}(N).$$

Thus, with probability approaching 1 as N grows to infinity, we have

$$1+|D_2^*|=\underset{1\leq d_1\leq 1+|D_2^*|}{\arg\min}\ \widetilde{\mathrm{IC}}_1(c,d_1,\min(|v_3^*|,d),\ldots,\min(|v_{c^*}^*|,d)). \tag{S4.78}$$

Combining (S4.71) with (S4.78), we have $\widetilde{d}_1^{c^*}=1+|D_2^*|$. Similarly, we have $\widetilde{d}_s^{c^*}=1+|D_{1+s}^*|$, for $s=1,\ldots,c^*$. Then we have

$$\begin{aligned}&\widetilde{\mathrm{IC}}_1(c^*,1+|D_2^*|,\ldots,1+|D_{1+c^*}^*|)\\ =&\sum_{k\in\mathrm{Ch}_1^*}(|v_k^*|(|D_k^*|+1)-|D_k^*|(|D_k^*|+1)/2)\log N+O_{\mathbb{P}}(1),\end{aligned}$$

and (S4.68) holds.

Second, when $c<c^*$ in case A. We will show that the $\widetilde{d}_s^c$ given by Step 6 of Algorithm 2 satisfies $\widetilde{d}_s^c=\sum_{v_i^*\subset v_s^{1,c}}1+|D_i^*|$ for $s=1,\ldots,c$ with

probability approaching 1 as N grows to infinity.

For $s = 1$, when $d_1 \geq \sum_{v_i^* \subset v_s^{1,c}} 1 + |D_i^*|$, let $\underline{\Lambda}_{d_1}$ and $\underline{\Psi}_{d_1}$ be the solution to

$$\widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_2^{1,c}|, d), \ldots, \min(|v_c^{1,c}|, d)).$$

Similarly to Lemma 6, we have $\left\|\underline{\Lambda}_{d_1}\underline{\Lambda}_{d_1}^\top + \underline{\Psi}_{d_1} - \Sigma^*\right\|_F = O_{\mathbb{P}}(1/\sqrt{N})$ and by Taylor's expansion, we have

$$\begin{aligned}
&\widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_2^{1,c}|, d), \ldots, \min(|v_c^{1,c}|, d)) \\
=& l\big(\underline{\Lambda}_{d_1}\underline{\Lambda}_{d_1}^\top + \underline{\Psi}_{d_1}; S\big) + p_1\big(\underline{\Lambda}_{d_1}\big)\log N \\
=& O_{\mathbb{P}}(1) + \big(|v_1^{1,c}|d_1 - d_1(d_1-1)/2\big)\log N + \sum_{2\leq s\leq c}\big(|v_s^{1,c}|d_s - d_s(d_s-1)\big)\log N,
\end{aligned} \tag{S4.79}$$

where we denoted by $d_s = \min(|v_s^{1,c}|, d)$, $s = 2, \ldots, c$ for simplicity. Notice that the third term in (S4.79) is independent of the choice of d_1 and the second term is strictly increasing with respect to d_1 when $\sum_{v_i^* \subset v_1^{1,c}} 1 + |D_i^*| \leq d_1 \leq \min(|v_1^{1,c}|, d)$. Thus, with probability approaching 1, as N grows to infinity, we have

$$\begin{aligned}
&\sum_{v_i^* \subset v_1^{1,c}} 1 + |D_i^*| \\
=& \underset{\sum_{v_i^* \subset v_1^{1,c}} 1 + |D_i^*| \leq d_1 \leq \min(|v_1^{1,c}|, d)}{\arg\min} \widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_2^{1,c}|, d), \ldots, \min(|v_c^{1,c}|, d)).
\end{aligned} \tag{S4.80}$$

When $d_1 < \sum_{v_i^* \subset v_1^{1,c}} 1 + |D_i^*|$, similar to the proof in (S4.72)-(S4.77), we have

$$\sum_{v_i^* \subset v_1^{1,c}} 1+|D_i^*| = \mathop{\arg\min}_{1 \le d_1 \le \sum_{v_i^* \subset v_1^{1,c}} 1+|D_i^*|} \widetilde{\mathrm{IC}}_1(c, d_1, \min(|v_2^{1,c}|, d), \ldots, \min(|v_c^{1,c}|, d)), \tag{S4.81}$$

with probability approaching 1 as N grows to infinity. Combining (S4.80) with (S4.81), we have $\widetilde{d}_1^c = \sum_{v_i^* \subset v_1^{1,c}} (1 + |D_i^*|)$. Similarly, we also have $\widetilde{d}_s^c = \sum_{v_i^* \subset v_s^{1,c}} (1 + |D_i^*|), s = 1, \ldots, c$. However, it is obvious that

$$\sum_{s=1}^{c} \big(|v_s^{1,c}|\widetilde{d}_s^c - \widetilde{d}_s^c(\widetilde{d}_s^c - 1)/2\big) > \sum_{i \in \mathrm{Ch}_i^*} \big(|v_s^*|(|D_s^*| + 1) - |D_s^*|(|D_s^*| + 1)/2\big),$$

when $\widetilde{d}_s^c = \sum_{v_i^* \subset v_s^{1,c}} (1 + |D_i^*|), s = 1, \ldots, c$. Thus, with probability approaching 1 as N grows to infinity, the derived $\widetilde{\mathrm{IC}}_1(c, \widetilde{d}_1^c, \ldots, \widetilde{d}_c^c)$ is larger than (S4.68).

Finally, when $v_1^{1,c}, \ldots, v_c^{1,c}$ are not included in case A or B, the derived information criterion is strictly larger than (S4.68) with probability approaching 1 as N grows to infinity by the first part of our proof. Thus, the second part is proved.

At the end of the proof, we conclude that with the same argument in Lemma 7, $\|\widetilde{\boldsymbol{\lambda}}_1 - \boldsymbol{\lambda}_1^* \mathrm{sign}(\boldsymbol{\lambda}_1^{*\top} \widetilde{\boldsymbol{\lambda}}_1)\| = O_{\mathbb{P}}(1/\sqrt{N})$, which indicating $\|\widetilde{\boldsymbol{\lambda}}_1 \widetilde{\boldsymbol{\lambda}}_1^\top - \boldsymbol{\lambda}_1^* \boldsymbol{\lambda}_1^{*\top}\|_F = O_{\mathbb{P}}(1/\sqrt{N})$. Then Theorem 2 is proved by applying the same argument to the factors in the tth layer, $t = 2, \ldots, T$, together with Lemma 7.

□

S5 Simulation studies for Algorithm 3 with correctly estimated number of child factors

As discussed in Section 3, Algorithm 3 may converge only to a local optimum, and the local solution may not satisfy constraint C4. In this section, we examine the performance of Algorithm 3 to find a global optimum and decode the structure of the child factors of Factor k given $c = |\mathrm{Ch}_k^*|$ with multiple random starts in detail. We consider the hierarchical structure shown in Figure S1 with $J \in \{24, 36\}$, $v_1^* = \{1, \ldots, J\}$, $v_2^* = \{1, \ldots, J/3\}$, $v_3^* = \{1 + J/3, \ldots, 2J/3\}$, $v_3^* = \{1 + 2J/3, \ldots, J\}$, $v_5^* = \{1, \ldots, J/6\}$ and $v_6^* = \{1 + J/6, \ldots, J/3\}$.

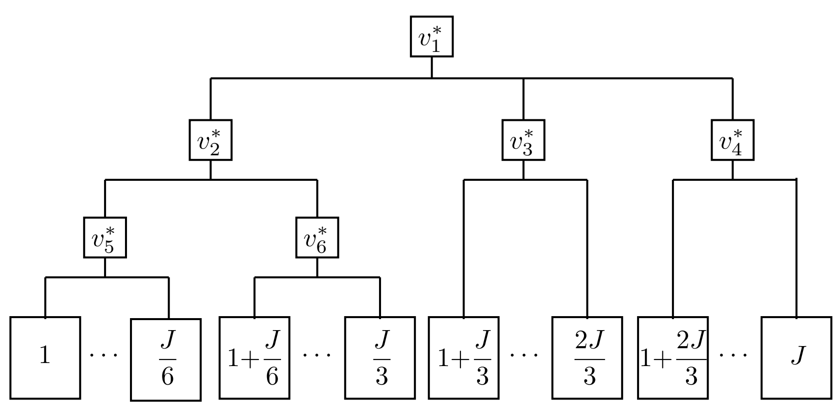

Figure S1: The hierarchical factor structure in the simulation studies of Section S5.

In the data generation model, Λ^* is generated by

$$\lambda^*_{jk} = \begin{cases} u_{jk} \ \text{ if } \ k = 1; \\ 0 \ \text{ if } \ k > 1, j \notin v^*_k; \\ (1 - 2x_{jk})u_{jk} \ \text{ if } \ k > 1, j \in v^*_k, \end{cases} \tag{S5.82}$$

for $j = 1, \ldots, J$, and $k = 1, \ldots, K$. Here, u_{jk}s are i.i.d., following a Uniform$(0.5, 2)$ distribution and x_{jk}s are i.i.d., following a Bernoulli(0.5) distribution. Ψ^* is either an identity matrix or $\Psi^* = \text{diag}(\psi_1^{*2}, \ldots, \psi_J^{*2})$ with $\psi^*_j, j = 1, \ldots, J$ i.i.d following a Uniform$(0.5, 1.5)$ distribution.

Let $\widehat{\Lambda}$ and $\widehat{\Psi}$ be the estimates of Λ^* and Ψ^* given by Algorithm 3 and $\{\widehat{v}_{1+i}\}_{i=1}^c$ be the estimated set of variables belonging to child factors of factor 1. To define a global optimal solution to the optimization problem in (18), we consider the ideal case when $S = \Sigma^*$. It is easy to notice that the objective function

$$\begin{aligned} &\widetilde{l}\left(\Lambda\Lambda^\top + \Psi, \Sigma^*\right) \\ =& \log(\det(\Lambda\Lambda^\top + \Psi)) + \text{tr}(\Sigma^*(\Lambda\Lambda^\top + \Psi)^{-1}) - \log(\det(\Sigma^*)) - J \end{aligned}$$

reach its global minimum at 0. Thus, for each optimization result from a random starting point, we define the following criterion

1. GS(Global Solution): a binary variable equal to 1 if $|\widetilde{l}(\widehat{\Lambda}\widehat{\Lambda}^\top + \widehat{\Psi}, \Sigma^*)| < \delta$ and 0 otherwise, where δ is a tolerance parameter.

2. CR(Correctness Rate): a binary variable equal to 1 if $\{\widehat{v}_{1+i}\}_{i=1}^c =$

$\{v_2^*, v_3^*, v_4^*\}$ and 0 otherwise.

We apply Algorithm 3 with $c = 3$ and $d = 5$ and further denote $\widehat{v}_2$, $\widehat{v}_3$, and $\widehat{v}_4$ as the estimated set of variables belonging to the child factors of factor 1. In this simulation study, we consider 4 simulation settings, given by the combinations of $J = 24, 36$ and two generation processes of Ψ. For each setting, 100 independent simulations are generated, and in each simulation, we use 100 random starting points with the tolerance parameter $\delta = 10^{-4}$. The numerical results are given in Table S1. As shown in Table S1, when $J = 24$, around 57% of the random starting points converge to a global optimum and 15% of the estimation results correctly decode the underlying hierarchical factor structure. When $J = 36$, there exists a decrease in both GS and CR, with around 38% and 11% of the random starting points converging to a global optimum, respectively.

Table S1: The mean value and standard deviation of GS and CR in the simulation study.

Ψ^*	J	GS	CR
Identity	24	$57.55_{(16.32)}$	$15.42_{(5.39)}$
	36	$37.52_{(12.11)}$	$11.19_{(4.23)}$
Heterogeneous	24	$56.48_{(15.61)}$	$14.88_{(5.48)}$
	36	$39.49_{(12.29)}$	$11.58_{(4.18)}$

Remark S2. We emphasize that when the optimization problem (18) reaches a global solution, the estimated sets of variable $\widehat{v}_2$, $\widehat{v}_3$, $\widehat{v}_4$ are not necessarily equal to v_2^*, v_3^*, v_4^*. In the current setting, the following configurations can yield equivalent covariance structures while satisfying the constraints of the optimization problem:

A. $\widehat{v}_2, \widehat{v}_3$, $\widehat{v}_4$ are equal to $\{1, \dots, 2J/3\}$, $\{1 + 2J/3, \dots, J\}$, $\emptyset$.

B. $\widehat{v}_2, \widehat{v}_3$, $\widehat{v}_4$ are equal to $\{1, \dots, J/3, 1+2J/3, \dots, J\}$, $\{1+J/3, \dots, 2J/3\}$, $\emptyset$.

C. $\widehat{v}_2, \widehat{v}_3$, $\widehat{v}_4$ are equal to $\{1, \dots, J/3\}$, $\{1 + J/3, \dots, J\}$, $\emptyset$.

D. $\widehat{v}_2, \widehat{v}_3$, $\widehat{v}_4$ are equal to $\{1, \dots, J\}$, $\emptyset$, $\emptyset$.

E. $\widehat{v}_2, \widehat{v}_3$, $\widehat{v}_4$ are equal to $\{1, \dots, J\} \setminus \{i\}$, $\{i\}$, $\emptyset$ for some $i \in \{1, \dots, J\}$.

These cases correspond precisely to the cases discussed in the proof of Theorem 2. To be more exact, case A, B, C are the cases when two of v_2^*, v_3^*, v_4^* are merged into one set, and Case D is the case when v_2^*, v_3^*, v_4^* are merged. Case E constructs the following parametric space for the loading matrix Λ:

$$\{\Lambda \in \mathbb{R}^{J\times 16} : \lambda_{jk} = 0 \text{ for } j \neq i, 7 \leq k \leq 16 \text{ and } \lambda_{ik} = 0$$
$$\text{for } k = 2, \dots, 6, 12, \dots, 16\}.$$

Given an arbitrary $\Lambda^* \in \mathbb{R}^{J\times 6}$ and unique variance matrix Ψ^*, we construct the loading matrix $\widetilde{\Lambda}$ and unique variance matrix $\widetilde{\Psi} = \text{diag}(\widetilde{\psi}_1, \dots, \widetilde{\psi}_J)$

belonging to the parametric space defined in Case E such that $\widetilde{\Lambda}\widetilde{\Lambda}^\top + \widetilde{\Psi} = \Lambda^*\Lambda^{*\top} + \Psi^*$ as follows:

1. Let $R \in \mathbb{R}^{6\times 6}$ be an orthogonal matrix such that $R^\top_{[\{1,\dots,6\},\{1\}\}]} = \frac{\Lambda^*_{[\{i\},\{1,\dots,6\}]}}{\left\|\Lambda^*_{[\{i\},\{1,\dots,6\}]}\right\|}$.
2. Let $\widetilde{\Lambda}_{[\{1,\dots,J\}\setminus\{i\},\{1,\dots,6\}]} = \Lambda^*_{[\{1,\dots,J\}\setminus\{i\},\{1,\dots,6\}]}R$ and $\widetilde{\Lambda}_{[\{i\},\{1\}]} = \left\|\Lambda^*_{[\{i\},\{1,\dots,6\}]}\right\|$.
3. Let $\widetilde{\Lambda}_{[\{i\},\{7,\dots,11\}]}$ be an arbitrary vector such that $\left\|\widetilde{\Lambda}_{[\{i\},\{7,\dots,11\}]}\right\|^2 < \psi^*_i$, $\widetilde{\psi}_j = \psi^*_j$ for $j \neq i$ and $\widetilde{\psi}_i = \psi^*_i - \left\|\widetilde{\Lambda}_{[\{i\},\{7,\dots,11\}]}\right\|^2$.

This construction shows that Case E also yields a global minimizer of the objective function. However, all cases A-E have at least one empty set among $\widehat{v}_2, \widehat{v}_3$ and $\widehat{v}_4$. Since our goal is to recover the structure of three non-empty child factors of factor 1, such solutions violate the intention of the modeling and are excluded in Steps 5–8 of Algorithm 2.

S6 Simulation studies for underestimated number of child factors

In this section, we examine the performance of Algorithm 1 and 2 when $c_{\max}$, the upper bound for the possible number of child factors of each factor, is underestimated. We adopt the same hierarchical structure and data generation model used in Section S5. As illustrated in Figure S1, $c_{\max}$

should be at least 3. However, in this simulation study, we deliberately set $c_{\max} = 2$ and $d_{\max} = 5$ when applying Algorithms 1 and 2. In this simulation study, we consider 8 simulation settings, given by the combinations of $J = 24, 36$, two sample sizes $N = 500, 2000$ and two generation processes of Ψ used in Section S5. For each setting, we generate the loading matrix and the unique variance matrix once, and then 100 independent simulations are generated.

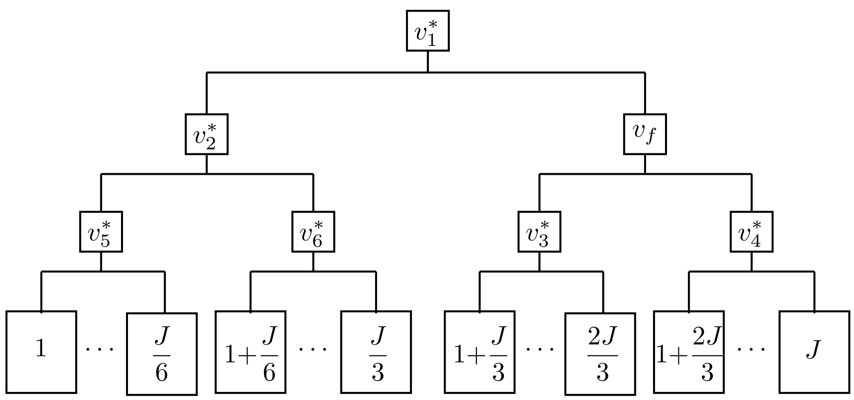

Figure S2: The hierarchical factor structure learned with underestimated $c_{\max}$.

Figure S2 displays the most frequently estimated hierarchical structure, which is selected in more than 60% of the 100 replications across all settings. As shown, Algorithms 1 and 2 recover a correctly specified but less parsimonious representation of the true hierarchy. To be more exact, a redundant factor, whose sets of variables $v_f = v_3^* \cup v_4^*$, is learned due to the choice of $c_{\max} = 2$ in the current simulation settings.

S7 Real Data Analysis: Agreeableness Scale Item Key

Table S2: Agreeableness Item Key

Item	Sign	Facet	Item detail
1	+	Trust(A1)	Trust others.
2	+	Trust(A1)	Believe that others have good intentions.
3	+	Trust(A1)	Trust what people say.
4	−	Trust(A1)	Distrust people.
5	−	Morality(A2)	Use others for my own ends.
6	−	Morality(A2)	Cheat to get ahead.
7	−	Morality(A2)	Take advantage of others.
8	−	Morality(A2)	Obstruct others' plans.
9	+	Altruism(A3)	Love to help others.
10	+	Altruism(A3)	Am concerned about others.
11	−	Altruism(A3)	Am indifferent to the feelings of others.
12	−	Altruism(A3)	Take no time for others.
13	−	Cooperation(A4)	Love a good fight.
14	−	Cooperation(A4)	Yell at people.
15	−	Cooperation(A4)	Insult people.
16	−	Cooperation(A4)	Get back at others.
17	−	Modesty(A5)	Believe that I am better than others.
18	−	Modesty(A5)	Think highly of myself.
19	−	Modesty(A5)	Have a high opinion of myself.
20	−	Modesty(A5)	Boast about my virtues.
21	+	Sympathy(A6)	Sympathize with the homeless.
22	+	Sympathy(A6)	Feel sympathy for those who are worse off than myself.
23	−	Sympathy(A6)	Am not interested in other people's problems.
24	−	Sympathy(A6)	Try not to think about the needy.

S8 Real Data Analysis: Additional Results

In this section, we present the estimated loading matrix and correlation matrix of the three competing models. The estimated correlation matrix of the three models, denoted by $\widehat{\Phi}_{\mathrm{CFA}}$, $\widehat{\Phi}_{\mathrm{CBF}}$, $\widehat{\Phi}_{\mathrm{EBF}}$, are shown in (S8.83), (S8.84), and (S8.85). The estimated loading matrix of the three models, denoted by $\widehat{\Lambda}_{\mathrm{CFA}}$, $\widehat{\Lambda}_{\mathrm{CBF}}$, $\widehat{\Lambda}_{\mathrm{EBF}}$, are shown in (S8.86), (S8.87), and (S8.88).

$$\widehat{\Phi}_{\mathrm{CFA}} = \begin{pmatrix} 1 & 0.33 & 0.44 & 0.43 & -0.06 & 0.37 \\ 0.33 & 1 & 0.42 & 0.62 & 0.25 & 0.37 \\ 0.44 & 0.42 & 1 & 0.39 & 0.15 & 0.80 \\ 0.43 & 0.62 & 0.39 & 1 & 0.11 & 0.30 \\ -0.06 & 0.25 & 0.15 & 0.11 & 1 & 0.16 \\ 0.37 & 0.37 & 0.80 & 0.30 & 0.16 & 1 \end{pmatrix}, \tag{S8.83}$$

$$\widehat{\Phi}_{\mathrm{CBF}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.01 & 0.24 & 0.03 & -0.07 & 0.25 \\ 0 & 0.01 & 1 & 0.12 & 0.27 & 0.34 & 0.22 \\ 0 & 0.24 & 0.12 & 1 & -0.08 & 0.18 & 0.74 \\ 0 & 0.03 & 0.27 & -0.08 & 1 & 0.25 & 0.05 \\ 0 & -0.07 & 0.34 & 0.18 & 0.25 & 1 & 0.17 \\ 0 & 0.25 & 0.22 & 0.74 & 0.05 & 0.17 & 1 \end{pmatrix}, \tag{S8.84}$$

$$\widehat{\Phi}_{\mathrm{EBF}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.12 & 0.18 & 0.24 & 0.12 & -0.02 \\ 0 & 0.12 & 1 & 0.50 & 0.11 & 0.95 & 0.33 \\ 0 & 0.18 & 0.50 & 1 & 0.13 & 0.74 & 0.24 \\ 0 & 0.24 & 0.11 & 0.13 & 1 & 0.09 & -0.14 \\ 0 & 0.12 & 0.95 & 0.74 & 0.09 & 1 & 0.31 \\ 0 & -0.02 & 0.33 & 0.24 & -0.14 & 0.31 & 1 \end{pmatrix}. \tag{S8.85}$$

$$\widehat{\Lambda}_{\text{CFA}} = \begin{pmatrix} 0.85 & 0 & 0 & 0 & 0 & 0 \\ 0.73 & 0 & 0 & 0 & 0 & 0 \\ 0.76 & 0 & 0 & 0 & 0 & 0 \\ 0.87 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.89 & 0 & 0 & 0 & 0 \\ 0 & 0.64 & 0 & 0 & 0 & 0 \\ 0 & 0.92 & 0 & 0 & 0 & 0 \\ 0 & 0.39 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.51 & 0 & 0 & 0 \\ 0 & 0 & 0.61 & 0 & 0 & 0 \\ 0 & 0 & 0.67 & 0 & 0 & 0 \\ 0 & 0 & 0.57 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0 & 0.81 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.71 & 0 \\ 0 & 0 & 0 & 0 & 0.90 & 0 \\ 0 & 0 & 0 & 0 & 1.12 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.70 \\ 0 & 0 & 0 & 0 & 0 & 0.71 \\ 0 & 0 & 0 & 0 & 0 & 0.65 \\ 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}, \tag{S8.86}$$

$$\widehat{\Lambda}_{\text{CBF}} = \begin{pmatrix} 0.42 & 0.73 & 0 & 0 & 0 & 0 & 0 \\ 0.35 & 0.64 & 0 & 0 & 0 & 0 & 0 \\ 0.30 & 0.72 & 0 & 0 & 0 & 0 & 0 \\ 0.53 & 0.69 & 0 & 0 & 0 & 0 & 0 \\ 0.46 & 0 & 0.83 & 0 & 0 & 0 & 0 \\ 0.49 & 0 & 0.41 & 0 & 0 & 0 & 0 \\ 0.57 & 0 & 0.71 & 0 & 0 & 0 & 0 \\ 0.47 & 0 & 0.11 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0.45 & 0 & 0 & 0 \\ 0.24 & 0 & 0 & 0.60 & 0 & 0 & 0 \\ 0.43 & 0 & 0 & 0.51 & 0 & 0 & 0 \\ 0.41 & 0 & 0 & 0.41 & 0 & 0 & 0 \\ 0.28 & 0 & 0 & 0 & 0.70 & 0 & 0 \\ 0.54 & 0 & 0 & 0 & 0.43 & 0 & 0 \\ 0.68 & 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0.66 & 0 & 0 & 0 & 0.27 & 0 & 0 \\ 0.30 & 0 & 0 & 0 & 0 & 0.73 & 0 \\ -0.14 & 0 & 0 & 0 & 0 & 0.93 & 0 \\ -0.03 & 0 & 0 & 0 & 0 & 1.09 & 0 \\ 0.38 & 0 & 0 & 0 & 0 & 0.35 & 0 \\ 0.15 & 0 & 0 & 0 & 0 & 0 & 0.73 \\ 0.16 & 0 & 0 & 0 & 0 & 0 & 0.75 \\ 0.38 & 0 & 0 & 0 & 0 & 0 & 0.52 \\ 0.27 & 0 & 0 & 0 & 0 & 0 & 0.58 \end{pmatrix}, \tag{S8.87}$$

$$
\widehat{\Lambda}_{\mathrm{EBF}} = \begin{pmatrix}
0.35 & 0 & 0 & 0 & 0 & 0 & 0.77 \\
0.29 & 0 & 0 & 0 & 0 & 0 & 0.67 \\
0.27 & 0 & 0 & 0 & 0 & 0 & 0.73 \\
0.45 & 0 & 0 & 0 & 0 & 0 & 0.74 \\
0.65 & 0.65 & 0 & 0 & 0 & 0 & 0 \\
0.55 & 0.32 & 0 & 0 & 0 & 0 & 0 \\
0.69 & 0.61 & 0 & 0 & 0 & 0 & 0 \\
0.45 & 0 & 0.10 & 0 & 0 & 0 & 0 \\
0.18 & 0 & 0.52 & 0 & 0 & 0 & 0 \\
0.19 & 0 & 0 & 0 & 0 & 0.57 & 0 \\
0.37 & 0 & 0 & 0 & 0 & 0.55 & 0 \\
0.32 & 0 & 0.52 & 0 & 0 & 0 & 0 \\
0.50 & 0 & 0 & 0 & 0.18 & 0 & 0 \\
0.68 & -0.27 & 0 & 0 & 0 & 0 & 0 \\
0.80 & -0.21 & 0 & 0 & 0 & 0 & 0 \\
0.69 & 0 & 0 & 0 & 0 & 0 & 0.10 \\
0.38 & 0 & 0 & 0 & 0.67 & 0 & 0 \\
0.01 & 0 & 0 & 0 & 0.93 & 0 & 0 \\
0.13 & 0 & 0 & 0 & 1.09 & 0 & 0 \\
0.44 & 0 & 0 & 0 & 0.29 & 0 & 0 \\
0.20 & 0 & 0 & 0.74 & 0 & 0 & 0 \\
0.18 & 0 & 0 & 0.75 & 0 & 0 & 0 \\
0.29 & 0 & 0 & 0 & 0 & 0.62 & 0 \\
0.27 & 0 & 0 & 0.59 & 0 & 0 & 0
\end{pmatrix}, \tag{S8.88}
$$

References

Anderson, T. and H. Rubin (1956). Statistical inference in factor analysis. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 111 – 150. University of California Press.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis.* Cambridge University Press.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Ruhe, A. (1970). Perturbation bounds for means of eigenvalues and invariant subspaces. *BIT Numerical Mathematics 10*(3), 343–354.

Wedin, P.-A. (1972). Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics 12*, 99–111.

Weyl, H. (1912). Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen 71*(4), 441–479.