Supplementary to "Identifiabliity of the Bifactor Models"

Guanhua Fang, Jinxin Guo, Xin Xu, Zhiliang Ying and Susu Zhang

In this supplementary, we provide simulation results, additional examples and technical proofs for all theoretical results stated in the main paper. In particular, Appendix A provides simulation results for validating our theories. Appendix B gives multiple examples to help readers to better understand the structure of bifactor models. Appendix C collects the proofs for the identifiability of standard bifactor models. Appendix D is for the extended bifactor models. Appendix E gives the proofs of two-tier model's identifiabiltiy. Finally, the proofs of results in Section 4 can be found in Appendix F.

A Simulation

This section presents the results from several simulation studies designed to verify the theoretical identifiability results. Specifically, section A.1 presents the numerical results on the probit bifactor model, and section A.2 presents the results on the probit extended bifactor model. For both classes of models, several tests were considered: Some of the tests have true parameters that satisfy the identifiability conditions, while others fail to meet the conditions. For each test, 500 sets of responses were randomly generated, with random samples of size $N = 1000, 2000$, or 4000.The stochastic expectation-maximization (StE) algorithm (Celeux, 1985; Ip, 2002) was employed to estimate both classes of models, which has been applied to IRT models (Diebolt & Ip, 1996; Fox, 2003; Zhang, Chen, & Liu, 2020). Specifically, a Gibbs sampler following Albert (1992) was adopted for the stochastic expectation (StE) step and a gradient descent algorithm for the maximization (M) step. To guarantee the convergence, the StE algorithm was iterated 10,000 times for each set and the first 5,000 iterations were discarded as burn-in. The estimated parameters $(\hat{A}, \hat{\mathbf{d}})$ were evaluated in terms of root mean squared error (RMSE) with respect to the true parameters. That is, for a particular entry of the A matrix, a_{jk}, its RMSE was given by

$$RMSE(\hat{a}_{jk}) = (\frac{1}{500}\sum_{r=1}^{500}(\hat{a}_{jk} - a_{jk})^2)^{1/2}. \tag{S1}$$

For tests that meet the identifiability requirements, parameter estimates are expected to converge to the true values as N increases, with the $RMSE$s approaching 0. This will not be the case for tests that fail to meet the identifiability conditions, in which case the parameters cannot be consistently estimated.

A.1 Study 1: Probit bifactor model

Under the probit bifactor model, parameter recovery was evaluated under the following 4 cases:

Table S1: True item parameters.

item	Main factor	Testlet-specific factors 1	2	3	d
1	1.00	2.00			1.51
2	1.00	2.00			.39
3	1.00	2.00			-.62
4	1.00	2.00			-2.21
5	1.00	2.00			1.12
6	1.00	2.00			-.04
7	1.00	2.00			-.02
8	1.00	2.00			.94
9	1.00	2.00			.82
10	1.00	2.00			.59
11	2.00		1.00		.92
12	2.00		1.00		.78
13	2.00		1.00		.07
14	2.00		1.00		-1.99
15	2.00		1.00		.62
16	2.00		1.00		-.06
17	2.00		1.00		-.16
18	2.00		1.00		-1.47
19	2.00		1.00		-.48
20	2.00		1.00		.42
21	1.00			-.63	1.36
22	1.00			.18	-.10
23	1.00			-.84	.39
24	1.00			1.60	-.05
25	1.00			.33	-1.38
26	1.00			-.82	-.41
27	1.00			.49	-.39
28	1.00			.74	-.06
29	1.00			.58	1.10
30	1.00			-.31	.76

Case 1 Consider a test with three testlets. The true parameters are listed in Table S1. By checking that $|\mathcal{H}_1| = 3$ and $|\mathcal{Q}_g| > 3$ for all g, the model is identifiable according to Theorem 1.

Case 2 Remove testlet 3 from Case 1. The model is no longer identifiable according to Theorem 1, because $|\mathcal{H}_1| = 2$ and $|\mathcal{H}_2| = 0$.

Case 3 Remove testlet 2 from Case 1. According to Theorem 1, the model is identifiable, because $|\mathcal{H}_1| = 2$, $|\mathcal{H}_2| = 1$ and $|\mathcal{Q}_g| > 3$ for all g.

Case 4 Based on the true parameters in Table S1, construct a new test containing item 1 from testlet 1, items 11 and 12 from testlet 2, and all the 10 items in testlet 4. The model is nonidentifiable according to Theorem 1 by checking that $|\mathcal{Q}_1|, |\mathcal{Q}_2| < 3$.

The average RMSEs for the 4 cases, across all non-zero $\hat{a}$s and $\hat{d}$s, are reported in Table S2. Compared to the two identifiable cases (Case 1 and Case 3), the RMSEs from the two unidentifiable cases (Case 2 and Case 4) were remarkably larger. This is consistent with the theoretical results on the sufficient and necessary condition for bifactor models. Moreover,

Table S2: RMSE of model parameters for bifactor models.

	Case 1			Case 2			Case 3			Case 4		
n	1000	2000	4000	1000	2000	4000	1000	2000	4000	1000	2000	4000
$\boldsymbol{a}$	.16	.10	.07	.77	.64	.55	.17	.11	.07	.49	.30	.20
d	.11	.07	.05	.18	.10	.06	.11	.07	.05	.30	.18	.12

under Case 4 where $|\mathcal{Q}_1|, |\mathcal{Q}_2| < 3$ and $|\mathcal{Q}_3| > 3$, the average RMSEs of $\hat{a}$s in testlet 1 were 1.65, .89 and .64, respectively, for $N = 1000, 2000$ and 4000. For testlet 2, the average RMSEs of $\hat{a}$s were $1.33, 0.91$ and 0.58, respectively. However, for testlet 3, the average RMSEs were $0.12, 0.07$ and 0.05. This suggests that, for a particular testlet g, the parameters were better recovered when the requirement of $|\mathcal{Q}_g| \geq 3$ was met.

A.2 Study 2: Probit extended bifactor model

Under the probit extended bifactor model, the following two cases were considered:

Case 5 The three-testlet extended bifactor model with loadings specified in Table S1. According to Theorem 3, we know the model is not identifiable by checking that $|\mathcal{H}_3| = 1$.

Case 6 Replace testlet 2 in Table S1 by testlet 4 in Table S3. According to Theorem 2, the model is identifiable by observing that $|\mathcal{H}_3| = 2$, $|\mathcal{H}_2| = 2$ and $|\mathcal{Q}_g| \geq 3$ for $g = 1, 3, 4$.

For both cases, the true covariance matrix is given in Table S4.

Table S5 reports the average RMSEs of the item parameters for the two cases with different sample sizes. Table S6 provides the RMSE for each entry of Σ_G, the covariance among the testlet factors. The numerical findings again corroborate the theoretical results, where both item and covariance parameters were recovered remarkably better under Case 6 compared to Case 5.

Table S3: True parameters of testlet 4 for Case 6.

		Testlet-specific factors	
item	Main factor	4	d
31	2.00	-.56	-.16
32	2.00	-.23	-.25
33	2.00	1.56	.70
34	2.00	.07	.56
35	2.00	.13	-.69
36	2.00	1.72	-.71
37	2.00	.46	.36
38	2.00	-1.27	.77
39	2.00	-.69	-.11
40	2.00	-.45	.88

Table S4: True covariance matrix for the extended bifactor model.

		Testlet-specific factors			
	Main factor	1	2	3	4
Main factor	1.00				
Testlet-specific factors		1.00			
		.44	1.00		
		.32	.52	1.00	
		.26	.21	.29	1.00

Table S5: RMSE of item parameters for extended bifactor model, under Case 5 and Case 6.

	Case 5			Case 6		
N	1000	2000	4000	1000	2000	4000
a	.46	.37	.32	.20	.11	.07
d	.11	.07	.05	.11	.07	.05

Table S6: RMSE of the covariance matrix under Cases 5 and 6.

$N = 1000$	Case 5			Case 6		
	.00			.00		
Σ_G	.11	.00		.04	.00	
	.16	.32	.00	.04	.09	.00

$N = 2000$	Case 5			Case 6		
	.00			.00		
Σ_G	.11	.00		.03	.00	
	.17	.30	.00	.03	.04	.00

$N = 4000$	Case 5			Case 6		
	.00			.00		
Σ_G	.12	.00		.02	.00	
	.18	.27	.00	.02	.02	.00

B Illustrative Examples

Example 1. *Consider a standard bifactor model for $J = 7$ items, where the true parameters are*

$$A = \begin{pmatrix} a & b & 0 & 0 \\ a & b & 0 & 0 \\ c & d & 0 & 0 \\ c & d & 0 & 0 \\ e & 0 & f & 0 \\ e & 0 & f & 0 \\ e & 0 & f & 0 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \end{pmatrix}, \tag{S2}$$

with $a, b, c, d, e, f \neq 0$ and $ad \neq bc$. One can check that (1) $|\mathcal{H}_1| = 2$, because both testlet 1 and testlet 2 have nonzero main factor loadings; (2) $|\mathcal{Q}_g| \geq 3$ for all g, because $|\mathcal{Q}_1| = 4, |\mathcal{Q}_2| = 3$; and (3) $|\mathcal{H}_2| \geq 1$, because testlet 1 can be partitioned to $\mathcal{B}_{g,1} = \{1, 3\}, \mathcal{B}_{g,2} = \{2, 4\}$, *each containing linearly independent columns. By Condition P2, we know that the model is identifiable.*

Example 2. *Even though the main factors by themselves satisfy the identification conditions of factor models, in a two-tier model context, the main factor loadings can still be indistinguishable. Consider a three-testlet model with two main factors, where*

$$A = \begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 1 & 0 & 4 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 3 \end{pmatrix}, \quad \Sigma_L = I_{3\times 3}. \tag{S3}$$

One can observe that the main factor loadings, $A_{:,1:2}$*, satisfy the sufficient condition for factor model identifiability per the 3-indicator rule (see Bollen, 1989). In addition, there are three testlets with nonzero main factor loadings. However, we can easily construct another* *set of parameters with the same observed data distribution, say,*

$$A' = \begin{pmatrix} \sqrt{3/2} & 0 & \sqrt{1/2} & 0 & 0 \\ \sqrt{4/2} & 0 & \sqrt{2/2} & 0 & 0 \\ \sqrt{5/2} & 0 & \sqrt{3/2} & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 3 \end{pmatrix}, \quad \Sigma_L' = I_{2\times 2}. \tag{S4}$$

This is because main factor 1 only depends on one testlet 1 and is consequently mixed up with the testlet-specific factor.

We can easily construct another set of parameters.

$$A'' = \begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 1 & 0 & 4 & 0 & 0 \\ 0 & \sqrt{1/2} & 0 & \sqrt{3/2} & 0 \\ 0 & \sqrt{1/2} & 0 & \sqrt{3/2} & 0 \\ 0 & \sqrt{1/2} & 0 & \sqrt{3/2} & 0 \\ 0 & \sqrt{2} & 0 & 0 & 2\sqrt{2} \\ 0 & \sqrt{2} & 0 & 0 & 2\sqrt{2} \\ 0 & \sqrt{2} & 0 & 0 & 2\sqrt{2} \end{pmatrix}, \quad \Sigma''_L = I_{2\times 2}. \tag{S5}$$

This time, main factor 2 depends on two testlets and it is mixed up with second and third testlet-specific factors.

Example 3. *It should further be noted that having testlets that load on multiple main factors would not suffice for two-tier model identifiability. Consider a two-tier model with three* *testlets and three main factors, where*

$$A = \begin{pmatrix} 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 2 & 0 & 0 \\ 0 & 1 & -1 & 3 & 0 & 0 \\ 2 & 0 & 1 & 0 & 3 & 0 \\ 2 & 0 & 1 & 0 & 2 & 0 \\ 2 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 2 \\ 1 & 1 & 0 & 0 & 0 & 3 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_L = I_{3\times 3}. \tag{S6}$$

Though each of the main factors is associated with multiple testlets, we can construct *another set of parameters which implies the same distribution.*

$$A^{'} = \begin{pmatrix} 0 & -\frac{3}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & 1 & 0 & 0 \\ 0 & -\frac{3}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & 2 & 0 & 0 \\ 0 & -\frac{3}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & 3 & 0 & 0 \\ \sqrt{3} & 0 & \sqrt{2} & 0 & 3 & 0 \\ \sqrt{3} & 0 & \sqrt{2} & 0 & 2 & 0 \\ \sqrt{3} & 0 & \sqrt{2} & 0 & 1 & 0 \\ \frac{2}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 & 0 & 0 & 2 \\ \frac{2}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 & 0 & 0 & 3 \\ \frac{2}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_L^{'} = I_{3\times 3}. \tag{S7}$$

Example 4. *Consider a probit bifactor model for $J = 9$ items, where the true parameters* *are given by*

$$A = \begin{pmatrix} a & b & 0 & 0 \\ a & b & 0 & 0 \\ a & b & 0 & 0 \\ a & 0 & c & 0 \\ a & 0 & c & 0 \\ a & 0 & c & 0 \\ a & 0 & 0 & d \\ a & 0 & 0 & d \\ a & 0 & 0 & d \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \\ d_9 \end{pmatrix}, \tag{S8}$$

with $a, b, c, d \neq 0$. *The parameter is identifiable by checking that it satisfies Condition P1.*

Example 5. *Consider a probit bifactor model with $J = 8$ items and true parameters*

$$A = \begin{pmatrix} a & b & 0 & 0 \\ a & b & 0 & 0 \\ a & b & 0 & 0 \\ a & 0 & c & 0 \\ a & 0 & c & 0 \\ a & 0 & c & 0 \\ a & 0 & 0 & d \\ a & 0 & 0 & d \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix}, \tag{S9}$$

with $a, b, c, d \neq 0$. This setting is not identifiable by checking that it fails to satisfy either Condition P1 or Condition P2.

Example 6. *The theoretical results from the current paper provide explanations to the findings in existing studies on bifactor identification with rigor and generality. For example, Green and Yang (2018) considered the empirical underidentification problem of bifactor model, which was encountered when fitting particular types of bifactor models to certain types of data sets. They demonstrated that the bifactor model can be underidentified in samples with homogenuous-within and homogenuous-between (HWHB) covariance structure, that is, $\sigma_{j_1j_2} = \sigma_{g_1g_2}$ where g_1, g_2 are the testlets that items j_1, j_2 belong to. In particular, they* *considered the following loading matrices.*

$$A = \begin{pmatrix} 0.8 & 0.3 & 0 \\ 0.8 & 0.3 & 0 \\ 0.8 & 0.3 & 0 \\ 0.8 & 0.3 & 0 \\ 0.8 & 0 & 0.3 \\ 0.8 & 0 & 0.3 \\ 0.8 & 0 & 0.3 \\ 0.8 & 0 & 0.3 \end{pmatrix}, \quad A = \begin{pmatrix} 0.7 & 0.4 & 0 \\ 0.7 & 0.4 & 0 \\ 0.7 & 0.4 & 0 \\ 0.7 & 0.4 & 0 \\ 0.8 & 0 & 0.3 \\ 0.8 & 0 & 0.3 \\ 0.8 & 0 & 0.3 \\ 0.8 & 0 & 0.3 \end{pmatrix}. \tag{S10}$$

They showed that the above two bifactor models were not identifable by constructing different solutions which lead to the same model-implied covariance matrix.

One can check that $|\mathcal{H}_1| = 2$, $|\mathcal{H}_2| = 0$ *for both settings. It follows from the sufficient and necessary conditions in Theorem 1 that the two models are not identifiable.*

C Proofs for Standard Bifactor Models

Proof of Theorem 1. We first introduce a few more notations.

- Define $\mathcal{Q}_{0,g} = \{j \mid g_j = g, \mathbf{a}_0[j] \neq 0\}$.
- Define $\mathcal{H}_6 = \{g \mid |\mathcal{Q}_{0,g}| \geq 2\}$.

It is easy to see that $\mathcal{H}_2 \subset \mathcal{H}_6 \subset \mathcal{H}_1$.

We prove the results by considering the follow cases. We aim to show that the model is identifiable if and only if Case 1.a or Case 2.d holds.

Case 1: $|\mathcal{H}_1| \geq 3$.

a $|\mathcal{Q}_g| \geq 3$ for all $g = 1, \ldots, G$.

b $|\mathcal{Q}_g| \leq 2$ for some $g \in \{1, \ldots, G\}$.

Case 2: $|\mathcal{H}_1| = 2$.

a $|\mathcal{Q}_g| \leq 2$ for some $g \in \{1, \ldots, G\}$.

b $|\mathcal{Q}_{0,g}| \leq 1$ for all $g \in \mathcal{H}_1$, i.e., $\mathcal{H}_6$ is empty.

c $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $\mathcal{H}_6$ is non-empty; $\mathcal{H}_2$ is empty.

d $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $\mathcal{H}_2$ is non-empty.

Case 3: $|\mathcal{H}_1| \leq 1$.

First, we can see that if two sets of parameters lead to the same marginal distribution, it must hold that

$$AA^T + \Lambda = A^{'}(A^{'})^T + \Lambda^{'} \tag{S11}$$

where $\Lambda = \mathrm{diag}((\lambda_1, \ldots \lambda_J))$. In other words, the off-diagonal elements are not collapsed with error variance.

Case 1 Suppose there is another set of parameters leading to the same model. Then we have $\mathbf{a}_0[\mathcal{B}_{g_1}]\mathbf{a}_0[\mathcal{B}_{g_2}]^T = \mathbf{a}_0^{'}[\mathcal{B}_{g_1}]\mathbf{a}_0^{'}[\mathcal{B}_{g_2}]^T$ for $g_1 \neq g_2 \in \mathcal{H}_1$. Thus it implies that $\mathbf{a}_0[\mathcal{B}_g] = \pm\mathbf{a}_0^{'}[\mathcal{B}_g]$ for $g \in \mathcal{H}_1$. By $\mathbf{a}_0[\mathcal{B}_{g_1}]\mathbf{a}_0[\mathcal{B}_g]^T = \mathbf{a}_0^{'}[\mathcal{B}_{g_1}]\mathbf{a}_0^{'}[\mathcal{B}_g]^T$ for $g_1 \in \mathcal{H}_1$ and $g \notin \mathcal{H}_1$, we further have $\mathbf{a}_0[\mathcal{B}_g] = \pm\mathbf{a}_0^{'}[\mathcal{B}_g]$ for $g \in \mathcal{H}_1$. This implies that main factor loading is identifiable.

Note the fact that $\mathbf{a}_g[j_1]\mathbf{a}_g[j_2] = \mathbf{a}_g^{'}[j_1]\mathbf{a}_g^{'}[j_2]$ for $j_1 \neq j_2 \in \mathcal{B}_g$ by comparing the correlations within testlet. By this, we consider the following.

If $\mathcal{Q}_g \geq 3$, then we must have that $\mathbf{a}_g[j] = \pm\mathbf{a}_g^{'}[j]$ for $j \in \mathcal{Q}_g$. Further, it implies $\mathbf{a}_g[j] = \pm\mathbf{a}_g^{'}[j]$ for $j \notin \mathcal{Q}_g$. Thus Case 1.a is identifiable.

If $\mathcal{Q}_g \leq 2$ for some g, we take out j_1 and j_2 from $\mathcal{Q}_g$ (if exist). Set $\mathbf{a}_g^{'}[j_1] = c \cdot \mathbf{a}_g[j_1]$ and $\mathbf{a}_g^{'}[j_2] = 1/c \cdot \mathbf{a}_g[j_2]$ for c satisfying that

$$|c^2 - 1| \cdot \mathbf{a}_g[j_1]^2 < \lambda_{j_1} \text{ and } |1 - 1/c^2| \cdot \mathbf{a}_g[j_2]^2 < \lambda_{j_2}. \tag{S12}$$

Such c exists since that $c = 1$ is one of the solution. By continuity, we know that any c sufficiently close to 1 satisfy (S12) and keep the same sign of $\mathbf{a}_g$. This tells that $\mathbf{a}_g[j]$ is not uniquely determined. Hence Case 1.b is not identifiable.

Case 2 Suppose $\mathcal{Q}_g \leq 2$ for some g. By the same construction in Case 1.b, parameter $\mathbf{a}_g[j]$ cannot be identified for $j \in \mathcal{Q}_g$. Thus Case 2.a is not identifiable.

Suppose $\mathcal{Q}_{0,g} \leq 1$ for $g \in \mathcal{H}_1$. We can set $\mathbf{a}_0^{'}[j_1] = c \cdot \mathbf{a}_0[j_1]$ for $j_1 \in \mathcal{Q}_{0g_1}$, $g_1 \in \mathcal{H}_1$; set $\mathbf{a}_0^{'}[j_2] = 1/c \cdot \mathbf{a}_0^{'}[j_2]$ for $j_2 \in \mathcal{Q}_{0g_2}$, $g_2 \in \mathcal{H}_1$ and keep other $a_g[j]$'s fixed. It is easy to check that $\mathbf{a}_g[j_1]\mathbf{a}_g[j_2] = \mathbf{a}_g^{'}[j_1]\mathbf{a}_g^{'}[j_2]$ holds for any $j_1 \neq j_2 \in \mathcal{B}_g$ and all g. We then choose c close to 1 enough such that $|c^2 - 1| \cdot \mathbf{a}_0[j_1]^2 < \lambda_{j_1}$ and $|1 - 1/c^2| \cdot \mathbf{a}_0[j_2]^2 < \lambda_{j_2}$. Then we can find $\lambda_{j_1}^{'}$ and $\lambda_{j_2}^{'}$ to keep (S11) hold, and the sign remains unchanged. Thus Case 2.b is not identifiable as the parameter $\mathbf{a}_0$ can not be determined.

Suppose $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $\mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6$ is non-empty. Let g be the testlet satisfying that $g \in \mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6$. By comparing the off-diagonals of the covariance matrix, it must hold that

$$(1 - c^2)\mathbf{a}_0[\mathcal{B}_g - \{j\}]\mathbf{a}_0[j] + \mathbf{a}_g[\mathcal{B}_g - \{j\}]\mathbf{a}_g[j] = \mathbf{a}_g^{'}[\mathcal{B}_g - \{j\}]\mathbf{a}_g^{'}[j] \quad (\forall j \in \mathcal{B}_g). \tag{S13}$$

By the property of $\mathcal{H}_2$, we can find a partition of $\mathcal{B}_g = \mathcal{B}_{g,1} \cup \mathcal{B}_{g,2}$. Hence (S13) can be written as

$$(1 - c^2)\mathbf{a}_0[\mathcal{B}_{g,1}]\mathbf{a}_0[\mathcal{B}_{g,2}]^T + \mathbf{a}_g[\mathcal{B}_{g,1}]\mathbf{a}_g[\mathcal{B}_{g,2}]^T = \mathbf{a}_g^{'}[\mathcal{B}_{g,1}]\mathbf{a}_g^{'}[\mathcal{B}_{g,2}]^T. \tag{S14}$$

When $c^2 \neq 1$, the left hand side of (S14) has rank 2, while the right hand side of (S14) has at most rank 1. Thus $c^2 \equiv 1$, which reduces to Case 1.a. Hence Case 2.d is identifiable.

Suppose $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $\mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6$ is empty. By Case 2.b, we only need to consider the situation that $|\mathcal{H}_1 \cap \mathcal{H}_6| \geq 1$ and $|\mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6| = 0$. Take $g \in \mathcal{H}_1 \cap \mathcal{H}_6$, we then know that $\bar{A}_g$ can only take one of the following forms (after suitable row ordering),

$$(1) \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}, \quad (2) \begin{pmatrix} a & b \\ c & d \\ c & d \\ \mathbf{c} & \mathbf{d} \end{pmatrix}, \tag{S15}$$

where the matrix of form (1) is 3 by 2 and satisfies that $b, d, f \neq 0$ and at most one of a, c, e is zero; the matrix of form (2) is J_g ($J_g \geq 4$) by 2 and contains at most two different rows (i.e. rows are not equal up to scaling). Under both cases, we only need to check (S13) for items corresponding to the first three rows. For notational simplicity, we denote three items as 1,2 and 3.

We can construct another set of parameters, where

$$
\begin{aligned}
\mathbf{a}_g^{'}[1] &= \frac{(1-c^2)\mathbf{a}_0[1]\mathbf{a}_0[3]+\mathbf{a}_g[1]\mathbf{a}_g[3]}{\mathbf{a}_g[3]x};\\
\mathbf{a}_g^{'}[2] &= \frac{(1-c^2)\mathbf{a}_0[2]\mathbf{a}_0[3]+\mathbf{a}_g[2]\mathbf{a}_g[3]}{\mathbf{a}_g[3]x};\\
\mathbf{a}_g^{'}[3] &= \mathbf{a}_g[3]x;\\
x &= \frac{((1-c^2)\mathbf{a}_0[1]\mathbf{a}_0[3]+\mathbf{a}_g[1]\mathbf{a}_g[3])((1-c^2)\mathbf{a}_0[2]\mathbf{a}_0[3]+\mathbf{a}_g[2]\mathbf{a}_g[3])}{((1-c^2)\mathbf{a}_0[1]\mathbf{a}_0[2]+\mathbf{a}_g[1]\mathbf{a}_g[2])\mathbf{a}_g[3]^2};\\
\lambda_j^{'} &= \lambda_j+(1-c^2)a_0^2[j]+(a_g[j])^2-(a_g^{'}[j])^2,\ j=1,2,3.
\end{aligned}
\tag{S16}
$$

Notice that $\mathbf{a}_g[3]\neq 0$ according to the assumption that $|\mathcal{Q}_g|\geq 3$. Hence the above solution will be different from the true parameters when $c^2\neq 1$. Note that $x>0$ when c is sufficiently close to 1. Then $\mathbf{a}_g^{'}$ has the same sign as $\mathbf{a}_g$. Thus Case 2.c is not identifiable.

Case 3 Apparently, it is not identifiable. This is because we can construct another set of parameters, $\bar{A}_g^{'}=\bar{A}_g, g=2,\ldots,G$ and $\bar{A}_1^{'}=\bar{A}_1R$ with R being a 2 by 2 rotation matrix. It is easy to see that the two sets of parameters lead to the same distribution, since $\Sigma_{g_1g_2}=\Sigma_{g_1g_2}^{'}$ for all g_1 and g_2. In addition, we can easily choose the rotation matrix R such that it keeps sign of first non-zero element in each column of $\bar{A}_1$.

Identifiability of $\mathbf{d}$ is obvious by using the expectation of Y_j. Thus we conclude the proof. ■

Proof of Proposition 2. The sufficient part is straightforward by noticing that $P(Y_{j_1}=1,\ldots,Y_{j_k}=1)$ only depends on $d_j/(\mathbf{a}_j^T\Sigma\mathbf{a}_j+1)^{1/2}$'s and $(\mathbf{a}_{j_1}^T\Sigma\mathbf{a}_{j_2})/\big((\mathbf{a}_{j_1}^T\Sigma\mathbf{a}_{j_1}+1)(\mathbf{a}_{j_2}^T\Sigma\mathbf{a}_{j_2}+1)\big)^{1/2}$'s for all possible combinations of $j_1,\ldots j_k$.

The necessary part is also not hard. Notice that CDF function Φ is a strictly monotone increasing function. By (3.2), we must have $d_j/(\mathbf{a}_j^T\Sigma\mathbf{a}_j+1)^{1/2}=d_j^{'}/((\mathbf{a}_j^{'})^T\Sigma^{'}\mathbf{a}_j^{'}+1)^{1/2}$ for all j. In addition, $\Phi_2(a,b,\rho)$ is a strictly monotone increasing function of ρ for any fixed a,b. Thus, from (3.3), we get $(\mathbf{a}_{j_1}^T\Sigma\mathbf{a}_{j_2})/\big((\mathbf{a}_{j_1}^T\Sigma\mathbf{a}_{j_1}+1)(\mathbf{a}_{j_2}^T\Sigma\mathbf{a}_{j_2}+1)\big)^{1/2}=(\mathbf{a}_{j_1}^T\Sigma^{'}\mathbf{a}_{j_2})/\big(((\mathbf{a}_{j_1}^{'})^T\Sigma^{'}\mathbf{a}_{j_1}^{'}+1)((\mathbf{a}_{j_2}^{'})^T\Sigma^{'}\mathbf{a}_{j_2}^{'}+1)\big)^{1/2}$ for any $j_1\neq j_2$. Hence we prove the proposition. ■

Proof of Theorem 6. We keep using the notations of $\mathcal{Q}_{0,g}$ and $\mathcal{H}_6$. We still aim to show that the model is identifiable if and only if Case 1.a or Case 2.d holds.

Case 1: $|\mathcal{H}_1|\geq 3$.

a $|\mathcal{Q}_g|\geq 3$ for all $g=1,\ldots,G$.

b $|\mathcal{Q}_g|\leq 2$ for some $g\in\{1,\ldots,G\}$.

Case 2: $|\mathcal{H}_1|=2$.

a $|\mathcal{Q}_g|\leq 2$ for some $g\in\{1,\ldots,G\}$.

b $|\mathcal{Q}_{0,g}| \le 1$ for all $g \in \mathcal{H}_1$, i.e., $\mathcal{H}_6$ is empty.

c $|\mathcal{Q}_g| \ge 3$ for $g = 1, \ldots, G$; $\mathcal{H}_2$ is empty.

d $|\mathcal{Q}_g| \ge 3$ for $g = 1, \ldots, G$; $\mathcal{H}_2$ is non-empty.

Case 3: $|\mathcal{H}_1| \le 1$.

The first step is to show that the mapping $(x, y) \to (\frac{x}{\sqrt{x^2+y^2+1}}, \frac{y}{\sqrt{x^2+y^2+1}})$ is one-to-one. It is easy to see that the mapping is onto. We only need to show it is injective. If there exists another $(x^{'}, y^{'})$ such that $\frac{x}{\sqrt{x^2+y^2+1}} = \frac{x^{'}}{\sqrt{(x^{'})^2+(y^{'})^2+1}}$ and $\frac{y}{\sqrt{x^2+y^2+1}} = \frac{y^{'}}{\sqrt{(x^{'})^2+(y^{'})^2+1}}$, we can find $(x^{'})^2 + (y^{'})^2 = x^2 + y^2$ which further implies $x = x^{'}$ and $y = y^{'}$. Therefore, in the following, we only need to work with $\tilde{\mathbf{a}}_j$, where $\tilde{\mathbf{a}}_j = \frac{\mathbf{a}_j}{\sqrt{\mathbf{a}_j^T \mathbf{a}_j + 1}}$.

Case 1 Suppose there is another set of parameters leading to the same distribution. Then we have $\tilde{\mathbf{a}}_0[\mathcal{B}_{g_1}]\tilde{\mathbf{a}}_0[\mathcal{B}_{g_2}]^T = \tilde{\mathbf{a}}_0^{'}[\mathcal{B}_{g_1}]\tilde{\mathbf{a}}_0^{'}[\mathcal{B}_{g_2}]^T$ for $g_1 \ne g_2 \in \mathcal{H}_1$. Thus it implies that $\tilde{\mathbf{a}}_0[\mathcal{B}_g] = \pm\tilde{\mathbf{a}}_0^{'}[\mathcal{B}_g]$ for $g \in \mathcal{H}_1$. By $\tilde{\mathbf{a}}_0[\mathcal{B}_{g_1}]\tilde{\mathbf{a}}_0[\mathcal{B}_g]^T = \tilde{\mathbf{a}}_0^{'}[\mathcal{B}_{g_1}]\tilde{\mathbf{a}}_0^{'}[\mathcal{B}_g]^T$ for $g_1 \in \mathcal{H}_1$ and $g \notin \mathcal{H}_1$, we further have $\tilde{\mathbf{a}}_0[\mathcal{B}_g] = \pm\tilde{\mathbf{a}}_0^{'}[\mathcal{B}_g]$ for $g \in \mathcal{H}_1$. This implies that main factor loading is identifiable.

Notice that $\tilde{\mathbf{a}}_g[j_1]\tilde{\mathbf{a}}_g[j_2] = \tilde{\mathbf{a}}_g^{'}[j_1]\tilde{\mathbf{a}}_g^{'}[j_2]$ for $j_1 \ne j_2 \in \mathcal{B}_g$ by comparing the correlations within testlet.

If $\mathcal{Q}_g \ge 3$, then we must have that $\tilde{\mathbf{a}}_g[j] = \pm\tilde{\mathbf{a}}_g^{'}[j]$ for $j \in \mathcal{Q}_g$. Further, it implies $\tilde{\mathbf{a}}_g[j] = \pm\tilde{\mathbf{a}}_g^{'}[j]$ for $j \notin \mathcal{Q}_g$. Thus Case 1.a is identifiable.

If $\mathcal{Q}_g \le 2$ for some g, we take out j_1 and j_2 from $\mathcal{Q}_g$ (if exist). Set $\tilde{\mathbf{a}}_g^{'}[j_1] = c \cdot \tilde{\mathbf{a}}_g[j_1]$ and $\tilde{\mathbf{a}}_g^{'}[j_2] = 1/c \cdot \tilde{\mathbf{a}}_g[j_2]$ for c satisfying that $\tilde{\mathbf{a}}_0[j_1]^2 + c^2 \cdot \tilde{\mathbf{a}}_g[j_1]^2 < 1$ and

$$\tilde{\mathbf{a}}_0[j_2]^2 + 1/c^2 \cdot \tilde{\mathbf{a}}_g[j_2]^2 < 1. \tag{S17}$$

Such c exists since that $c = 1$ is one of the solution. By continuity, we know that any c sufficiently close to 1 satisfy (S17) and keep the same sign of $\tilde{\mathbf{a}_g}$. This tells that $\mathbf{a}_g[j]$ is not uniquely determined. Hence Case 1.b is not identifiable.

Case 2 Suppose $\mathcal{Q}_g \le$ for some g. By the same construction in Case 1.b, parameter $\mathbf{a}_g[j]$ cannot be identified for $j \in \mathcal{Q}_g$. Thus Case 2.a is not identifiable.

Suppose $\mathcal{Q}_{0,g} \le 1$ for $g \in \mathcal{H}_1$. We can set $\tilde{\mathbf{a}}_0^{'}[j_1] = c \cdot \tilde{\mathbf{a}}_0[j_1]$ for $j_1 \in \mathcal{Q}_{0g_1}$, $g_1 \in \mathcal{H}_1$; set $\tilde{\mathbf{a}}_0^{'}[j_2] = 1/c \cdot \tilde{\mathbf{a}}_0^{'}[j_2]$ for $j_2 \in \mathcal{Q}_{0g_2}$, $g_2 \in \mathcal{H}_1$ and keep other $\tilde{a}_g[j]$'s fixed. It is easy to check that $\tilde{\mathbf{a}}_g[j_1]\tilde{\mathbf{a}}_g[j_2] = \tilde{\mathbf{a}}_g^{'}[j_1]\tilde{\mathbf{a}}_g^{'}[j_2]^T$ holds for any $j_1 \ne j_2 \in \mathcal{B}_g$ and all g. Here c is chosen to be positive to keep the sign. Thus Case 2.b is not identifiable as the parameter $\mathbf{a}_0[j_1]$ can not be determined.

Suppose $|\mathcal{Q}_g| \ge 3$ for $g = 1, \ldots, G$; $|\mathcal{Q}_g| \ge 3$ for $g = 1, \ldots, G$; $\mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6$ is non-empty. Let g be the testlet satisfying that $g \in \mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6$. By comparing the correlation within matrix, it must hold that

$$(1 - c^2)\tilde{\mathbf{a}}_0[\mathcal{B}_g - \{j\}]\tilde{\mathbf{a}}_0[j] + \tilde{\mathbf{a}}_g[\mathcal{B}_g - \{j\}]\tilde{\mathbf{a}}_g[j] = \tilde{\mathbf{a}}_g^{'}[\mathcal{B}_g - \{j\}]\tilde{\mathbf{a}}_g^{'}[j] \quad (\forall j \in \mathcal{B}_g). \tag{S18}$$

By the property of $\mathcal{H}_2$, we can find a partition of $\mathcal{B}_g = \mathcal{B}_{g,1} \cup \mathcal{B}_{g,2}$. Hence (S18) can be written as

$$(1-c^2)\tilde{\mathbf{a}}_0[\mathcal{B}_{g,1}]\tilde{\mathbf{a}}_0[\mathcal{B}_{g,2}]^T + \tilde{\mathbf{a}}_g[\mathcal{B}_{g,1}]\tilde{\mathbf{a}}_g[\mathcal{B}_{g,2}]^T = \tilde{\mathbf{a}}_g^{'}[\mathcal{B}_{g,1}]\tilde{\mathbf{a}}_g^{'}[\mathcal{B}_{g,2}]^T. \tag{S19}$$

When $c^2 \neq 1$, the left hand side of (S19) has rank 2, while the right hand side of (S19) has at most rank 1. Thus $c^2 \equiv 1$, which reduces to Case 1.a. Hence Case 2.d is identifiable.

Suppose $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $|\mathcal{Q}_g| \geq 3$ for $g = 1, \ldots, G$; $\mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6$ is empty. By Case 2.b, we only need to consider the situation that $|\mathcal{H}_1 \cap \mathcal{H}_6| \geq 1$ and $|\mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_6| = 0$. Take $g \in \mathcal{H}_1 \cap \mathcal{H}_6$, we then know that $\bar{A}_g$ can only take one of the following forms (after suitable row ordering),

$$(1) \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}, \quad (2) \begin{pmatrix} a & b \\ c & d \\ c & d \\ \mathbf{c} & \mathbf{d} \end{pmatrix}, \tag{S20}$$

where the matrix of form (1) is 3 by 2 and satisfies that $b, d, f \neq 0$ and at most one of a, c, e is zero; the matrix of form (2) is J_g ($J_g \geq 4$) by 2 and contains at most two different rows (i.e. rows are not equal up to scaling). Under both cases, we only need to check (S18) for items corresponding to the first three rows. For notational simplicity, we denote three items as 1,2 and 3.

We can construct another set of parameters,

$$\begin{aligned}
\tilde{\mathbf{a}}_g^{'}[1] &= \frac{(1-c^2)\tilde{\mathbf{a}}_0[1]\tilde{\mathbf{a}}_0[3] + \tilde{\mathbf{a}}_g[1]\tilde{\mathbf{a}}_g[3]}{\tilde{\mathbf{a}}_g[3]x}; \\
\tilde{\mathbf{a}}_g^{'}[2] &= \frac{(1-c^2)\tilde{\mathbf{a}}_0[2]\tilde{\mathbf{a}}_0[3] + \tilde{\mathbf{a}}_g[2]\tilde{\mathbf{a}}_g[3]}{\tilde{\mathbf{a}}_g[3]x}; \\
\tilde{\mathbf{a}}_g^{'}[3] &= \tilde{\mathbf{a}}_g[3]x; \\
x &= \frac{((1-c^2)\tilde{\mathbf{a}}_0[1]\tilde{\mathbf{a}}_0[3] + \tilde{\mathbf{a}}_g[1]\tilde{\mathbf{a}}_g[3])((1-c^2)\tilde{\mathbf{a}}_0[2]\tilde{\mathbf{a}}_0[3] + \tilde{\mathbf{a}}_g[2]\tilde{\mathbf{a}}_g[3])}{((1-c^2)\tilde{\mathbf{a}}_0[1]\tilde{\mathbf{a}}_0[2] + \tilde{\mathbf{a}}_g[1]\tilde{\mathbf{a}}_g[2])\tilde{\mathbf{a}}_g[3]^2}.
\end{aligned} \tag{S21}$$

Notice that $\tilde{\mathbf{a}}_g[3] \neq 0$ according to assumption that $|\mathcal{Q}_g| \geq 3$. Hence the above solution will be different from true parameter when $c^2 \neq 1$. When c is close enough to 1, we know that x is positive and $\tilde{\mathbf{a}}_g^{'}$ has the same sign as that of $\tilde{\mathbf{a}}_g$. Thus Case 2.c is not identifiable.

Case 3 Obviously, it is not identifiable by the same reason as stated in Case 3 in the proof of Theorem 1.

Once loading matrix is identifiable, we can immediately identify $\mathbf{d}$ by using (3.2). Thus we conclude the proof. ■

D Proofs for Extended Bifactor Models

For the linear and probit extended bifactor model identifiability, we only provide the complete proof for the linear case (i.e., Theorem 2 and Theorem 3). It should be apparent from Appendix A that the proofs for linear and probit models are very similar.

The following are two support theorems for the proof of Theorems 2 and 3.

Theorem S1. *Under the the linear extended bifactor model with known error variance, if parameters satisfy* $|\mathcal{H}_3| \geq 2$ *and* $|\mathcal{N}| = 0$*, then it is identifiable.*

Theorem S2. *Under the linear extended bifactor model, the model parameter is not identifiable if* $|\mathcal{H}_3| \leq 1$*.*

Before proof of Theorems S1 and S2, we first state the following Lemma S1 which plays an important role in proving the identification of the extended bifactor model.

Lemma S1. *Assume* A *and* B *are both two-column matrices. Let* Σ *and* Σ' *be two by two matrices. Consider the following situations:*

1. *Suppose both* A *and* B *are full-column rank. Thus* $A\Sigma B^T = A\Sigma' B^T$ *implies that* $\Sigma = \Sigma'$.

2. *Suppose* A *is full-column rank and* B *has column-rank 1, i.e.,* $B = \mathbf{b}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}^T$. *Thus* $A\Sigma B^T = A\Sigma' B^T$ *implies that* $\Sigma\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \Sigma'\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$.

3. *Suppose both* A *and* B *are column-rank 1, i.e.,* $A = \mathbf{a}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^T$ *and* $B = \mathbf{b}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}^T$.

$$\textit{Thus } A\Sigma B^T = A\Sigma^{'} B^T \textit{ implies that } \begin{pmatrix} a_1 \\ \\ a_2 \end{pmatrix}^T \Sigma \begin{pmatrix} b_1 \\ \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ \\ a_2 \end{pmatrix}^T \Sigma^{'} \begin{pmatrix} b_1 \\ \\ b_2 \end{pmatrix}.$$

Proof of Theorem S1. We prove this by contradiction. Suppose there exists another set of $\{A^{'}_g\}$ leading to the same distribution. We pick any item pair g_1 and g_2 from set $\mathcal{H}^*$. We know that $\Sigma_{gg} = \Sigma^{'}_{gg}$ for $g = g_1, g_2$, which implies that

$$\bar{A}^{'}_{g_1} = \bar{A}_{g_1} \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix}, \qquad \bar{A}^{'}_{g_2} = \bar{A}_{g_2} \begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ -\sin\theta_2 & \cos\theta_2 \end{pmatrix} \tag{S22}$$

In addition, we know that $\Sigma_{g_1g_2} = \Sigma^{'}_{g_1g_2}$, which implies that

$$\begin{pmatrix} 1 & 0 \\ 0 & \sigma_{g_1g_2} \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma^{'}_{g_1g_2} \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix}. \tag{S23}$$

After simplification, we have that

$$0 = -\cos\theta_1 \sin\theta_2 + \sigma^{'} \sin\theta_1 \cos\theta_2 \tag{S24}$$

$$0 = -\sin\theta_1 \cos\theta_2 + \sigma^{'} \cos\theta_1 \sin\theta_2. \tag{S25}$$

Observe that $\cos\theta_1$ and $\cos\theta_2$ are not equal to zero, otherwise $\cos\theta_1 \cos\theta_2 + \sigma^{'} \sin\theta_1 \sin\theta_2 < 1$. By (S24), we have $\tan\theta_1 = \sigma^{'2} \tan\theta_1$. This implies that $\theta_1 = 0$, or π. It implies that $\bar{A}_{g_1} = \bar{A}^{'}_{g_1}$ and $\sigma_{g_1g_2} = \sigma^{'}_{g_1g_2}$.

Take any g not in $\mathcal{H}_3$ and g_1 in $\mathcal{H}_3$, we know that $\bar{A}_g$ can be represented as $\mathbf{a}_g[\mathcal{B}_g](c_1, c_2)$. It is easy to see that $A^{'}_g = \mathbf{a}_g[\mathcal{B}_g](c^{'}_1, c^{'}_2)$ with $c^{'2}_1 + c^{'2}_2 = c^2_1 + c^2$. Compare Σ_{g_1g} and $\Sigma^{'}_{g_1g}$, we have that

$$\begin{pmatrix} 1 & 0 \\ 0 & \sigma_{g_1g_2} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^{'}_{g_1g_2} \end{pmatrix} \begin{pmatrix} c^{'}_1 \\ c^{'}_2 \end{pmatrix}. \tag{S26}$$

This implies that $c^{'}_1 = c_1$, $c^{'}_2 = c_2$ and $\sigma^{'}_{g_1g} = \sigma_{g_1g}$.

Furthermore, if $G > |\mathcal{H}_3|$, we take any testlet pair g_1 and g_2 not in $\mathcal{H}_3$. By $\Sigma_{g_1g_2} = \Sigma_{g_1g_2}$, we have

$$\begin{pmatrix} c_{g_1 1} & c_{g_1 2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{g_1 g_2} \end{pmatrix} \begin{pmatrix} c_{g_2 1} \\ c_{g_2 2} \end{pmatrix} = \begin{pmatrix} c_{g_1 1} & c_{g_1 2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma^{'}_{g_1 g_2} \end{pmatrix} \begin{pmatrix} c_{g_2 1} \\ c_{g_2 2} \end{pmatrix}, \tag{S27}$$

which implies that $\sigma^{'}_{g_1 g_2} = \sigma_{g_1 g_2}$. Hence, all parameters are identifiable. This concludes our proof. ■

Proof of Theorem S2. For simplicity, we suppose $|\mathcal{H}_3| = 1$ and $g_1 \in \mathcal{H}_3$. For any $g \neq g_1$, it must hold that

$$\begin{pmatrix} 1 & 0 \\ 0 & \sigma^{'}_{gg_1} \end{pmatrix} \begin{pmatrix} \cos\theta_g & \sin\theta_g \\ -\sin\theta_g & \cos\theta_g \end{pmatrix} \begin{pmatrix} c_{g1} \\ c_{g2} \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{gg_1} \end{pmatrix} \begin{pmatrix} c_{g1} \\ c_{g2} \end{pmatrix} \tag{S28}$$

according to Lemma S1, where $\bar{A}_g$ has the form of $\mathbf{a}_g[\mathcal{B}_g](c_{g1}, c_{g2})$. Here c_2 is a non-zero constant. By simplification, we then have

$$\begin{pmatrix} c_{g1}\cos\theta_g + c_{g2}\sin\theta_g \\ \sigma^{'}_{gg_1}(-c_{g1}\sin\theta_g + c_{g2}\cos\theta_g) \end{pmatrix} = \begin{pmatrix} c_{g1}\cos\theta_1 + c_{g2}\sigma_{gg_1}\sin\theta_1 \\ (-c_{g1}\sin\theta_1 + c_{g2}\sigma_{gg_1}\cos\theta_1) \end{pmatrix}. \tag{S29}$$

Equation (S29) can be viewed as a function of θ_1, θ_g and $\sigma^{'}_{gg_1}$. Clearly, it admits the solution $\theta_g = 0, \theta_1 = 0, \sigma^{'}_{gg_1} = \sigma_{gg_1}$.

We perturb θ_1 a bit at $\theta_1 = 0$ locally, i.e., $\theta_1 = \delta$. Here we can always choose δ such that it keeps the sign of the first item in $\mathcal{B}_g$ to be positive. Otherwise, we can choose $-\delta$. Thus, by the implicit function theorem, (S29) admits the solution $\theta^{'}_g = \theta_g(\delta)$ and $\sigma^{'}_{gg_1} = \sigma_{gg_1}(\delta)$, since the determinant of gradient does not vanish at $\theta_g = 0, \sigma^{'}_{gg_1} = \sigma_{gg_1}$ when $c_{g2} \neq 0$. In addition, we know that $\theta_g(\delta) \to 0$ as $\delta \to 0$. Then it keeps the sign of first non-zero item in $\mathcal{B}_g$ since both $a_0[j_g]$ and $a_g[j_g] > 0$ when δ is close enough to 0. (Here j_g is the first non-zero item in $\mathcal{B}_g$.) Furthermore, it admits that $\sigma^{'}_{gg'} = \sigma_{gg'}(\delta)$ for any item pair $g, g^{'} \neq g_1$. Let δ go to 0, then $\sigma^{'}_{gg'}$s uniformly go to $\sigma_{gg'}$ since the number of parameters is finite. By eigenvalue perturbation theory, there exists a δ such that that $\Sigma^{'}$ is still positive definite. This guarantees that $\Sigma^{'}$ is still a covariance matrix.

By the above displays, the model is not identifiable since we have constructed another set of parameters leading to the same distribution. By the same technique, the model is not identifiable when $|\mathcal{H}_3| = 0$. Thus we conclude the proof. ■

Now we are ready to prove the main results.

Proof of Theorem 2. Proof under Condition *E2S*: The first step is to show that the covariance is identifiable. We prove this by checking the Condition C0 of Theorem 4. It is also suffices to check the Condition C1 of Theorem S3.

By the requirement that $|\mathcal{Q}_g| \geq 3$, we denote these three items in testlet g as $j_{g,1}, j_{g,2}$ and $j_{g,3}$. From the requirement that $|\mathcal{H}_2| \geq 1$, we can assume $g_1 \in \mathcal{H}_2$ and $\bar{A}_{g_1}[\mathcal{B}_{g_1},:]$, $\bar{A}_{g_1}[\mathcal{B}_{g_2},:]$ are full-column rank with $\mathcal{B}_{g_1,1} = \{j_{g_1,1}, j_{g_1,2}\}$ and $\mathcal{B}_{g_1,2} = \{j_{g_1,3}, j_{g_1,4}\}$. By the requirement of $|\mathcal{H}_3| \geq 2$, we know there exists $g_2 \in \mathcal{H}_3$ such that $g_2 \neq g_1$. Then we can assume $\bar{A}_{g_2}[j_{g_2,2},:]$ and $\bar{A}_{g_2}[j_{g_2,3},:]$ are linearly independent for items $j_{g_2,2}$ and $j_{g_2,3}$.

We then can construct a partition $\mathcal{B}_1 \cup \mathcal{B}_2$ satisfying that $\mathcal{B}_1 = \{j_{g_1,1}, j_{g_1,2}, j_{g_2,1}, \ldots, j_{g,1}; g \neq g_1, g \neq g_2\}$, $\mathcal{B}_2 = \{j_{g_1,3}, j_{g_1,4}, j_{g_2,2}, j_{g_2,3}, \ldots, j_{g,2}, j_{g,3}; g \neq g_1, g \neq g_2\}$. It is easy to check that $A[\mathcal{B}_1,:]$ has full column rank. Let $\mathcal{B}_{2a} = \{j_{g_1,3}, j_{g_1,4}, j_{g_2,2}, \ldots, j_{g,2};; g \neq g_1, g \neq g_2\}$. It is also easy to check that $A[\mathcal{B}_{2a},:]$ has full column rank and $A[\mathcal{B}_2 - \{j\},:]$ has full column rank for $\forall j \in \mathcal{B}_{2a}$. Thus Condition C1 is satisfied.

Hence the problem is reduced to the linear case with known variance. We only need to check the condition that $|\mathcal{N}| = 0$ according to Theorem S1. If not, there is a g such that $\mathbf{a}_g = \mathbf{0}$. It contradicts with $|\mathcal{Q}_g| \geq 3$. This concludes the proof.

Proof under Condition *E1S*: The first step is still to show that the item covariance matrix is identifiable. Again, we prove this by checking the Condition C1.

By the requirement that $|\mathcal{Q}_g| \geq 3$ for each testlet g, we take any three items in testlet g and denote them as $j_{g,1}, j_{g,2}$ and $j_{g,3}$. By the requirement of $|\mathcal{H}_3| \geq 3$, we can assume $g_1, g_2, g_3 \in \mathcal{H}_3$ and assume $\bar{A}_{g_1}[\mathcal{B}_{g_1,a},:]$, $\bar{A}_{g_2}[\mathcal{B}_{g_2,a},:]$, $\bar{A}_{g_3}[\mathcal{B}_{g_3,a},:]$ have full-column rank with $\mathcal{B}_{g_1,a} = \{j_{g_1,1}, j_{g_1,2}\}$, $\mathcal{B}_{g_2,a} = \{j_{g_2,1}, j_{g_2,2}\}$ and $\mathcal{B}_{g_3,a} = \{j_{g_3,1}, j_{g_3,2}\}$.

In the following, we need to verify that $\{1, \ldots, J\} - j$ can be partitioned into two item sets $\mathcal{B}_{1,j}$ and $\mathcal{B}_{2,j}$ for each item j such that $A[\mathcal{B}_{1,j},:]$ and $A[\mathcal{B}_{2,j},:]$ satisfy Condition C1.

If j belongs to testlet g ($g \neq g_1, g_2, g_3$), we can assume $j = j_{g,1}$ without loss of generality. Then we can set

$$\mathcal{B}_{1,j} = \{j_{g_1,1}, j_{g_1,2}, j_{g_2,3}, j_{g_3,3}, \ldots, j_{g,2}, \ldots; g \neq g_1, g_2, g_3\},$$

and

$$\mathcal{B}_{2,j} = \{j_{g_1,3}, j_{g_2,1}, j_{g_2,2}, j_{g_3,1}, j_{g_3,2}, \ldots, j_{g,3}, \ldots; g \neq g_1, g_2, g_3\}.$$

If j belongs to testlet g ($g \in \{g_1, g_2, g_3\}$), we can assume $j = j_{g_1,1}$ without loss of generality. Then we can set

$$\mathcal{B}_{1,j} = \{j_{g_1,2}, j_{g_2,1}, j_{g_2,2}, j_{g_3,3}, \ldots, j_{g,2}, \ldots; g \neq g_1, g_2, g_3\},$$

and

$$\mathcal{B}_{2,j} = \{j_{g_1,3}, j_{g_2,3}, j_{g_3,1}, j_{g_3,2}, \ldots, j_{g,3}, \ldots; g \neq g_1, g_2, g_3\}.$$

Then by Condition C1 we know that the covariance matrix is identifiable. Obviously $|\mathcal{N}| = 0$ and $|\mathcal{H}_3| \geq 2$ hold. Hence we conclude the proof. ■

Proof of Theorem 3. To prove the necessity that $|\mathcal{H}_3| \geq 2$, we want to show that we can always construct another set of parameters leading to the same distribution for any Θ satisfies $\mathcal{H}_1 \leq 1$.

Under this case, we keep $\boldsymbol{\lambda}$ fixed and only consider to construct another loading matrix A. It is easy to compute that the covariance between items j_1 and j_2 from group g is

$$\sigma_{j_1 j_2} = \mathbf{a}_0[j_1]\mathbf{a}_0[j_2] + \mathbf{a}_g[j_1]\mathbf{a}_g[j_2]; \tag{S30}$$

the covariance between items j_1 and j_2 from groups g_1 and g_2 is

$$\sigma_{j_1 j_2} = \mathbf{a}_0[j_1]\mathbf{a}_0[j_2] + \mathbf{a}_{g_1}[j_1]\mathbf{a}_{g_2}[j_2]\sigma_{g_1 g_2}. \tag{S31}$$

We can write them in matrix form which becomes

$$\Sigma_{gg} = \bar{A}_g \bar{A}_g^T \text{ and } \Sigma_{g_1 g_2} = \bar{A}_{g_1} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{g_1 g_2} \end{pmatrix} \bar{A}_{g_2}^T. \tag{S32}$$

Then, by Theorem S2, we have that $|\mathcal{H}_3| \geq 2$.

To prove the necessity that $|\mathcal{Q}_g| \geq 2$ for all $g \in \{1, \ldots, G\}$, we construct another set of parameters leading to the same distribution. Take any g such that $|\mathcal{Q}_g| \leq 1$. Without loss of generality, we can assume $|\mathcal{Q}_g| = 1$ and assume item j satisfies $\mathbf{a}_g[j] \neq 0$. Thus, we can construct another set of parameters such that $\mathbf{a}_g^{'}[j] = x \cdot a_g[j]$, $\Sigma_G^{'}[g, g^{'}] = \Sigma_G[g, g^{'}]/x$ for $g^{'} \neq g$ and keep other parameters fixed. It can be verified that this set of parameters works.

Furthermore, by the same argument in the proof of Theorem 1, we must have $|\mathcal{Q}_g| \geq 3$ for all g satisfying $\Sigma_G[g, -g] = \mathbf{0}$, ■

Proof of Proposition 1. We still use proof by contradiction. The first step is to show that the main factor loadings are identifiable. By assumption, we must have that

$$\bar{A}_{g_1}^{'} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_G^{'}[g_1, g_2] \end{pmatrix} \bar{A}_{g_2}^{'T} = \bar{A}_{g_1} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_G[g_1, g_2] \end{pmatrix} \bar{A}_{g_2}^T. \tag{S33}$$

Since $g_1, g_2 \in \mathcal{H}_3$, we know $\bar{A}_{g_1}, \bar{A}_{g_2}$ are full column rank matrices. There must exist full rank matrices R_1 and R_2 such that $\bar{A}_{g_1}^{'} = \bar{A}_{g_1} \cdot R_1$ and $\bar{A}_{g_2}^{'} = \bar{A}_{g_2} \cdot R_2$.

Next, again by assumption, we know that

$$\bar{A}_{g_1}^{'}[-j, :]\bar{A}_{g_1}^{'}[j, :]^T = \bar{A}_{g_1}[-j, :]\bar{A}_{g_1}[j, :]^T; \quad j \in \mathcal{Q}_{g_1}. \tag{S34}$$

This implies that $R_1 R_1^T A_{g_1}[j, :]^T = A_{g_1}[j, :]^T$ for $j \in \mathcal{Q}_{g_1}$, as the Kruskal rank of $\bar{A}_{g_1}$ is 2. It must hold that $R_1 R_1^T = I$. We then parameterize R_1 as

$$R_1 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{S35}$$

By (S33), we further have

$$R_2 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma'} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma \end{pmatrix}. \tag{S36}$$

Again by assumption, we know that

$$\bar{A}'_{g_2}[-j,:]\bar{A}'_{g_2}[j,:]^T = \bar{A}_{g_2}[-j,:]\bar{A}_{g_2}[j,:]^T; \quad j \in \mathcal{Q}_{g_2}. \tag{S37}$$

This implies that determinant of $R_2 R_2^T - I$ is zero. By straightfoward calculation,

$$\begin{aligned} 0 &= \det(R_2 R_2^T - I) \\ &= \det(\begin{pmatrix} \sin^2\theta \cdot (\frac{1}{\sigma'^2} - 1) & \sigma\cos\theta\sin\theta \cdot (1 - \frac{1}{\sigma'^2}) \\ \sigma\cos\theta\sin\theta \cdot (1 - \frac{1}{\sigma'^2}) & \frac{\sigma^2}{\sigma'^2}\cos^2\theta + \sigma^2\sin^2\theta - 1 \end{pmatrix}) \\ &= \sin^2\theta \cdot (\frac{1}{\sigma'^2} - 1) \cdot (\sigma^2 - 1). \end{aligned}$$

Hence, $\theta = 0$ and $R_1 = I$. This implies that main factor loading $\mathbf{a}_{g_1}$ is identifiable. By assumption, it holds that

$$\bar{A}_{g_1} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_G[g_1, g]' \end{pmatrix} \bar{A}_g'^T = \bar{A}_{g_1} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_G[g_1, g] \end{pmatrix} \bar{A}_g^T \tag{S38}$$

for any $g \neq g_1$. We must have that $\mathbf{a}_g' = \mathbf{a}_g$, since $\bar{A}_{g_1}$ has full column rank. Hence all main factor loadings are identifiable.

Next, we prove the identifiability of testlet-specific loadings. By assumption we have that

$$\mathbf{a}_g'[-j] \cdot \mathbf{a}_g'[j] = \mathbf{a}_g[-j] \cdot \mathbf{a}_g[j] \tag{S39}$$

for any j and g. If $|\mathcal{Q}_g| \geq 3$, we have that $\mathbf{a}_g = \mathbf{a}_g'$ according to proof in Theorem 1. If

$|\mathcal{Q}_g| = 2$, we know that $\Sigma_G[g, -g]$ is not zero-vector. Hence, we must have that $\mathbf{a}_g^{'} = x \cdot \mathbf{a}_g$ for some positive x by comparing the cross variance. Finally, by (S39), we have $x^2 = 1$. Thus $x \equiv 1$. Then $\mathbf{a}_g$ is identifiable. We conclude the proof. ■

To end this section, we provide a variant of Theorem 5.1 in Anderson and Rubin (1956) to item factor models with probit link, which provides a basis to the identifiability results for the probit extended bifactor and two-tier models. Based on this result, there will be no difference between the proofs of linear bifactor models and probit bifactor models.

Theorem S3. *Under the general factor model with probit link as in* (3.1)*, assume the following holds:*

C1 There exists a partition of the items $\{1, \ldots, J\} = \mathcal{B}_1 \cup \mathcal{B}_2$ such that (1) $A[\mathcal{B}_1, :]$ is full-column rank; (2) there exists a subset of $\mathcal{B}_2$, $\mathcal{B}_{2a}$, satisfying that $A[\mathcal{B}_{2a}, :]$ is full-column rank and $A[\mathcal{B}_2 - \{j\}, :]$ is full-column rank for $\forall j \in \mathcal{B}_{2a}$.

Then, the covariance matrix $A\Sigma A^T$ is identifiable.

The result is closely related to Kruskal rank. That is, Condition C1 is satisfied when $\mathcal{B}_2$ contains a subset of $\mathcal{B}_{2a}$ of $K+1$ items and $A[\mathcal{B}_{2a}, :]^T$ has Kruskal rank K. Thus, Theorem S3 requires $2K+1$ or more items. The condition is very weak, especially in terms of number of required items. Notice that there are JK loading parameters and $J(J-1)/2$ restrictions. Then $J(J-1)/2 \geq JK$ if and only if $J \geq 2K+1$. In fact, we can prove that $2K+1$ is minimal possible number of items for model identifiability for $K = 1$ and 2.

For the probit bifactor model, due to its special structure, the conditions for identifiability can be less restrictive than what Theorem S3 requires for general probit factor models: Specifically, Theorem S3 requires at least $3K+1$ items for meeting Condition C1, which is slightly stronger than the optimal conditions for the bifactor model as stated in Theorem 6.

Proof of Theorem S3. By comparing the correlation between $\mathcal{B}_1$ and $\mathcal{B}_2$, we have that

$$\tilde{A}[\mathcal{B}_1, :]\Sigma\tilde{A}[\mathcal{B}_2, :]^T = \tilde{A}^{'}[\mathcal{B}_1, :]\Sigma^{'}\tilde{A}^{'}[\mathcal{B}_2, :]^T. \tag{S40}$$

Since both $\tilde{A}[\mathcal{B}_1, :]$ and $\tilde{A}[\mathcal{B}_2, :]$ have full column rank, we can assume $\tilde{A}^{'}[\mathcal{B}_1, :] = \tilde{A}[\mathcal{B}_1, :]R_1$ and $\tilde{A}^{'}[\mathcal{B}_2, :] = \tilde{A}[\mathcal{B}_2, :]R_2$. Thus (S40) becomes

$$\Sigma = R_1\Sigma^{'}R_2^T. \tag{S41}$$

By comparing the correlation within $\mathcal{B}_2$, we have that

$$\tilde{A}[\mathcal{B}_2 - \{j\}, :]\Sigma\tilde{A}[j, :]^T = \tilde{A}^{'}[\mathcal{B}_2 - \{j\}, :]\Sigma^{'}\tilde{A}^{'}[j, :]^T \quad \forall j \in \mathcal{B}_{2a}. \tag{S42}$$

Notice that $\tilde{A}[\mathcal{B}_2 - \{j\}, :]$ has full rank K as $A[\mathcal{B}_2 - \{j\}, :]$ does. (S42) is reduced to

$$\Sigma\tilde{A}[j, :]^T = R_2\Sigma^{'}R_2^T\tilde{A}[j, :]^T \quad \forall j \in \mathcal{B}_{2a}, \tag{S43}$$

that is $(\Sigma - R_2\Sigma^{'}R_2^T)\tilde{A}[j, :]^T = \mathbf{0}$ for all $j \in \mathcal{B}_{2a}$. This implies that the kernel of the linear map $\Sigma - R_2\Sigma^{'}R_2^T$ is the whole vector space. Thus $\Sigma \equiv R_2\Sigma^{'}R_2^T$ which further implies $R_1 = R_2$. Therefore,

$$\tilde{A}\Sigma\tilde{A}^T = \tilde{A}^{'}\Sigma^{'}\tilde{(A^{'})}^T. \tag{S44}$$

We conclude that the covariance matrix is identifiable by comparing the diagonals of the above equation. ■

E Proofs for Two-Tier Models

Following similar rationale as for the extended bifactor models, it suffices to prove the following three theorems.

Theorem S4. *Under a linear two-tier model, suppose $\boldsymbol{\lambda}$ is known. Then the parameter is identifiable if it satisfies T1S.*

Theorem S5. *Under a linear two-tier model, suppose $\boldsymbol{\lambda}$ is known. Then the parameter is identifiable if it satisfies T2S.*

Theorem S6. *Under a linear two-tier model, suppose $\boldsymbol{\lambda}$ is known. Then the parameter is identifiable if it satisfies T3S.*

Proof of Theorem S4. We prove this by using method of contradiction. We adopt notation $\Sigma_{\mathcal{H}_1,\mathcal{H}_2}$ to denote the item covariance across testlets from $\mathcal{H}_1$ and $\mathcal{H}_2$. Since $|\mathcal{H}_3| \geq 3$, we then take three Testlets g_1, g_2 and g_3 such that $g_1, g_2, g_3 \in \mathcal{H}_3$. Note that $\Sigma'_{gg'} = \Sigma_{gg'}$ by comparing the item covariance across different testlets. We then know that $A'[\mathcal{B}_g, 1:L] = A_g[\mathcal{B}_g, 1:L]R_g$ ($g = g_1, g_2, g_3$), where R_g is a L by L full rank matrix. We also know that $R'_g \Sigma'_L R_g = \Sigma_L$. This indicates that $R_{g_1} = R_{g_2} = R_{g_3} = R$ and $R\Sigma'_L R^T = \Sigma_L$. It further implies

$$A[\mathcal{B}_{\mathcal{H}_3}, 1:L]R\Sigma'_L R^T A[\mathcal{B}_{\mathcal{H}_3}, 1:L]^T = A[\mathcal{B}_{\mathcal{H}_3}, 1:L]\Sigma_L A[\mathcal{B}_{\mathcal{H}_3}, 1:L]^T \tag{S45}$$

Since $A[\mathcal{B}_{\mathcal{H}_3}, 1:L]$ contains an identity, we then have $R = I$ and $\Sigma'_L = \Sigma_L$.

For any $g \notin \mathcal{H}_3$, we know that $\Sigma'_{\mathcal{H}_3,g} = \Sigma_{\mathcal{H}_3,g}$, indicating that $A'[\mathcal{B}_g, 1:L] = A[\mathcal{B}_g, 1:L]$. Lastly, we then have $\mathbf{a}'_g = \mathbf{a}_g$ for all g by comparing the item covariance within testlets. ■

Proof of Theorem S5. We still use method of contradiction to prove this theorem. Suppose there exists another set of parameter leading to the same distribution. Since $|\mathcal{H}_3| \geq 2$ and $|\mathcal{H}_4| \geq 1$, we then take two Testlets g_1 and g_2 such that $g_1 \in \mathcal{H}_3 \cap \mathcal{H}_4$ and $g_2 \in \mathcal{H}_3$. Notice that $\Sigma'_{g_1g_2} = \Sigma_{g_1g_2}$. We then know that $A'[\mathcal{B}_g, 1:L] = A[\mathcal{B}_g, 1:L]R_g$ ($g = g_1, g_2$), where R_g is a L by L full rank matrix. Recall $\bar{A}_{g_1} = (A[\mathcal{B}_{g_1}, 1:L], \mathbf{a}_{g_1}[\mathcal{B}_{g_1}])$, we then know $\bar{A}'_{g_1} = \bar{A}_{g_1}\begin{pmatrix} R_{g_1} & c \\ 0 & d \end{pmatrix}$. Furthermore, we know that $\Sigma'_{g_1g_1} = \Sigma_{g_1g_1}$. This gives us that

$$c \cdot d = 0 \text{ and } d^2 = 1, \tag{S46}$$

which implies that $\mathbf{a}_{g_1}^{'}(\mathbf{a}_{g_1}^{'})^T = \mathbf{a}_{g_1}\mathbf{a}_{g_1}^T$. We then also have $R_{g_1}\Sigma_L^{'}R_{g_1}^T = \Sigma_L$. Since $R_{g_1}\Sigma_L^{'}R_{g_2}^T = \Sigma_L$, it implies that $R_{g_1} = R_{g_2}$. Then $R_g\Sigma_L^{'}R_g^T = \Sigma_L$ for any $g \in \mathcal{H}_3$. It implies that

$$A[\mathcal{B}_{\mathcal{H}_3}, 1:L]R\Sigma_L^{'}R^T A[\mathcal{B}_{\mathcal{H}_3}, 1:L]^T = A[\mathcal{B}_{\mathcal{H}_3}, 1:L]\Sigma_L A[\mathcal{B}_{\mathcal{H}_3}, 1:L]^T \tag{S47}$$

Since $A[\mathcal{B}_{\mathcal{H}_3}, 1:L]$ contains an identity, we then have $R = I$ and $\Sigma_L^{'} = \Sigma$.

Furthermore, for any $g \notin \mathcal{H}_3$, we know that $\Sigma_{\mathcal{H}_3,g}^{'} = \Sigma_{\mathcal{H}_3,g}$, indicating that $A^{'}[\mathcal{B}_g, 1:L] = A[\mathcal{B}_g, 1:L]$. Lastly, we have $\mathbf{a}_g^{'} = \pm\mathbf{a}_g$ by using the fact that $\Sigma_{gg}^{'} = \Sigma_{gg}$ for all g. ■

Proof of Theorem S6. By the same strategy, suppose the model is not identifiable. There exists another set of parameters $A^{'}$ and $\Sigma^{'}$ leading to the same model. Hence, $A^{'}[\mathcal{B}_{\mathcal{G}_1}, 1:L]\Sigma_L^{'}A^{'}[\mathcal{B}_{\mathcal{G}_2}, 1:L]^T = A[\mathcal{B}_{\mathcal{G}_1}, 1:L]\Sigma_L A[\mathcal{B}_{\mathcal{G}_2}, 1:L]^T$. Since both $A[\mathcal{B}_{\mathcal{G}_1}, 1:L]$ and $A[\mathcal{B}_{\mathcal{G}_2}, 1:L]$ are full column rank matrices according to Condition *T3S*-(b). This implies that $A^{'}[\mathcal{B}_{\mathcal{G}_1}, 1:L]$ spans the same subspace as $A[\mathcal{B}_{\mathcal{G}_1}, 1:L]$ does. In other words, $A^{'}[\mathcal{B}_{\mathcal{G}_1}, 1:L]$ has the form of $A[\mathcal{B}_{\mathcal{G}_1}, 1:L] \cdot D$ for some D being a L by L full rank matrix.

By Condition *T3S*-(a) that $span(A[\mathcal{B}_{\mathcal{G}_1}, 1:L]) \cap span(A^{'}[\mathcal{B}_{\mathcal{G}_1}, L+\mathcal{G}_1]) = \mathbf{0}$ and by assumption that $A^{'}[\mathcal{B}_{\mathcal{G}_1}, :]\Sigma^{'}A^{'}[\mathcal{B}_{\mathcal{G}_1}, :]^T = A[\mathcal{B}_{\mathcal{G}_1}, :]\Sigma A[\mathcal{B}_{\mathcal{G}_1}, :]^T$, we then know $A^{'}[\mathcal{B}_{\mathcal{G}_1}, (1:L)+\mathcal{G}_1] = A[\mathcal{B}_{\mathcal{G}_1}, (1:L)+\mathcal{G}_1] \cdot R$ where R is $L+G_1$ by $L+G_1$ matrix (Here $G_1 = |\mathcal{G}_1|$). Furthermore, R should has the form of $\begin{pmatrix} D & X_1 \\ \mathbf{0} & X_2 \end{pmatrix}$. As a result, we then have

$$\begin{pmatrix} D & X_1 \\ \mathbf{0} & X_2 \end{pmatrix} \begin{pmatrix} \Sigma_L^{'} & 0 \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} D & \mathbf{0} \\ X_1^T & X_2^T \end{pmatrix} = \begin{pmatrix} \Sigma_L & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}, \tag{S48}$$

which implies that $X_1X_2^T = \mathbf{0}$, $X_2X_2^T = I$, $D\Sigma_L^{'}D + X_1X_1^T = \Sigma_L$. Therefore, it gives $X_1 = \mathbf{0}$, $X_2 = I$ and $D\Sigma_L^{'}D^T = \Sigma_L$. This further gives us that $A[:, 1:L]D\Sigma_L^{'}D^T A[:, 1:L]^T = A[:, 1:L]\Sigma_L A[:, 1:L]^T$. Since $A_{:,1:L}$ contains an identity, we then have D is an identity matrix. Thus, we have $A[:, 1:L]^{'} = A[:, 1:L]$ which further implies $A^{'} = A$. We thus conclude the proof. ■

F Proof of Theorem 10 in Section 4

Proof of Theorem 10. Let A_G be $(\mathbf{a}_1, \ldots, \mathbf{a}_G)$. We prove the results by considering the following two cases.

Case 1: $\mathbf{a}_0$ is not in the range of A_G. Under this case, we have that $\mathbf{a}_0, \ldots, \mathbf{a}_G$ are linearly independent. If not, there exists $c_0, \ldots, c_G$ such that $\sum_{g=0}^{G} c_g \mathbf{a}_g = \mathbf{0}$. Then we have $\sum_{g=1}^{G} c_g \mathbf{a}_g = \mathbf{0}$ by assumption that a_0 is not in the range of A_0. Since $\mathbf{a}_g$ has non-zero loadings on gth testlet, $\mathbf{a}_1, \ldots, \mathbf{a}_G$ are linearly independent. Thus we have $c_0 = \cdots = c_G = 0$.

If there is another model $\mathcal{P}^{'}(A^{'}, d, \Sigma_G^{'}, \rho^{'})$ that implies the same distribution, we must have

$$A^{'} = A \begin{pmatrix} x & \mathbf{0}^T \\ \mathbf{y} & \Lambda \end{pmatrix} \tag{S49}$$

since A has full column rank. Furthermore, we must have

$$\begin{pmatrix} x & \mathbf{0}^T \\ \mathbf{y} & \Lambda \end{pmatrix} \begin{pmatrix} 1 & (\boldsymbol{\rho}^{'})^T \\ \boldsymbol{\rho}^{'} & \Sigma_G^{'} \end{pmatrix} \begin{pmatrix} x & \mathbf{y}^T \\ \mathbf{0} & \Lambda \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\rho}^T \\ \boldsymbol{\rho} & \Sigma_G \end{pmatrix}. \tag{S50}$$

This gives us that $x^2 = 1$, $x(\mathbf{y} + \Lambda\boldsymbol{\rho}^{'}) = \boldsymbol{\rho}$ and $\mathbf{y}\mathbf{y}^T + \Lambda\boldsymbol{\rho}^{'}\mathbf{y}^T + \mathbf{y}(\boldsymbol{\rho}^{'})^T\Lambda + \Lambda\Sigma_G^{'}\Lambda = \Sigma_G$. After simplification, we have

$$\begin{aligned} (\mathbf{y} + \Lambda\boldsymbol{\rho}^{'})(\mathbf{y} + \Lambda\boldsymbol{\rho}^{'})^T - (\Lambda\boldsymbol{\rho}^{'})(\Lambda\boldsymbol{\rho}^{'})^T + \Lambda\Sigma_G^{'}\Lambda &= \Sigma_G \\ \boldsymbol{\rho}\boldsymbol{\rho}^T - (\Lambda\boldsymbol{\rho}^{'})(\Lambda\boldsymbol{\rho}^{'})^T + \Lambda\Sigma_G^{'}\Lambda = \Sigma_G \\ \boldsymbol{\rho}\boldsymbol{\rho}^T - (\tilde{\boldsymbol{\rho}})(\tilde{\boldsymbol{\rho}})^T + \Lambda\Sigma_G^{'}\Lambda = \Sigma_G, \end{aligned}$$

where $\tilde{\boldsymbol{\rho}} = \Lambda\boldsymbol{\rho}^{'}$.

We choose $\tilde{\boldsymbol{\rho}}$ to be arbitrary close to $\boldsymbol{\rho}$, set $\Lambda[g,g] = \sqrt{(\Sigma_G + (\tilde{\boldsymbol{\rho}})(\tilde{\boldsymbol{\rho}})^T - \boldsymbol{\rho}\boldsymbol{\rho}^T)[g,g]}$ and set $\Sigma_G^{'} = \Lambda^{-1}(\Sigma_G + (\tilde{\boldsymbol{\rho}})(\tilde{\boldsymbol{\rho}})^T - \boldsymbol{\rho}\boldsymbol{\rho}^T)\Lambda^{-1}$. Hence we find another set of parameters which leads to the same distribution. Thus the model is not identifiable.

Case 2: $\mathbf{a}_0$ is in the range of A_G. We construct another model $\mathcal{P}(A^{'}, d^{'}, \Sigma_G^{'}, \boldsymbol{\rho}^{'})$ such that $A_G^{'} = A_G$, $\Sigma_G^{'} = \Sigma_G$. To determine the value of $\mathbf{a}_0^{'}$ and $\boldsymbol{\rho}^{'}$, we use the following equations.

$$\begin{aligned}
(\mathbf{a}_0, A_G)\begin{pmatrix}1 & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \Sigma_G\end{pmatrix}(\mathbf{a}_0, A_G)^T &= (\mathbf{a}_0^{'}, A_G)\begin{pmatrix}1 & \boldsymbol{\rho}^{'}\\ \boldsymbol{\rho}^{'} & \Sigma_G\end{pmatrix}(\mathbf{a}_0^{'}, A_G)^T\\
\mathbf{a}_0\mathbf{a}_0^T + A_G\boldsymbol{\rho}\mathbf{a}_0^T + \mathbf{a}_0\boldsymbol{\rho}^T A_G^T &= (\mathbf{a}_0^{'})(\mathbf{a}_0^{'})^T + A_G\boldsymbol{\rho}^{'}(\mathbf{a}_0^{'})^T + (\mathbf{a}_0^{'})(\boldsymbol{\rho}^{'})^T A_G^T\\
(\mathbf{a}_0 + A_G\boldsymbol{\rho})(\mathbf{a}_0 + A_G\boldsymbol{\rho})^T - (A_G\boldsymbol{\rho})(A_G\boldsymbol{\rho})^T &= (\mathbf{a}_0^{'} + A_G\boldsymbol{\rho}^{'})(\mathbf{a}_0^{'} + A_G\boldsymbol{\rho}^{'})^T - (A_G\boldsymbol{\rho}^{'})(A_G\boldsymbol{\rho}^{'})^T\\
\mathbf{x}\mathbf{x}^T - \mathbf{y}\mathbf{y}^T &= (\mathbf{x}^{'})(\mathbf{x}^{'})^T - (\mathbf{y}^{'})(\mathbf{y}^{'})^T
\end{aligned}$$

where $\mathbf{x} = \mathbf{a}_0 + A_G\boldsymbol{\rho}$, $\mathbf{y} = A_G\boldsymbol{\rho}$ and $\mathbf{x}^{'}$, $\mathbf{y}^{'}$ are defined correspondingly. It is easy to check that $\mathbf{x}^{'} = \sqrt{1+c^2}\mathbf{x} + c\mathbf{y}$, $\mathbf{y}^{'} = c\mathbf{x} + \sqrt{1+c^2}\mathbf{y}$ satisfies the above equations. This give us

$$\mathbf{a}_0^{'} + A_G\boldsymbol{\rho}^{'} = \sqrt{1+c^2}(\mathbf{a}_0 + A_G\boldsymbol{\rho}) + cA_G\boldsymbol{\rho} \tag{S51}$$

and

$$A_G\boldsymbol{\rho}^{'} = c(\mathbf{a}_0 + A_G\boldsymbol{\rho}) + \sqrt{1+c^2}A_G\boldsymbol{\rho} \tag{S52}$$

By assumption that $\mathbf{a}_0$ is in the range of A_G, there exists $\mathbf{b}_0$ such that $\mathbf{a}_0 = A_G\mathbf{b}_0$. Equation (S52) becomes $A_G\boldsymbol{\rho}^{'} = A_G(c\mathbf{b}_0 + (c + \sqrt{1+c^2})\boldsymbol{\rho})$. This implies $\boldsymbol{\rho}^{'} = c\mathbf{b}_0 + (c + \sqrt{1+c^2})\boldsymbol{\rho}$. Plug this into (S51), we have

$$\mathbf{a}_0^{'} + A_G(c\mathbf{b}_0 + (c + \sqrt{1+c^2}\boldsymbol{\rho})) = \sqrt{1+c^2}(\mathbf{a}_0 + A_G\boldsymbol{\rho}) + cA_G\boldsymbol{\rho}.$$

Thus $\mathbf{a}_0^{'} = \sqrt{1+c^2}(\mathbf{a}_0 + A_G\boldsymbol{\rho}) + cA_G\boldsymbol{\rho} - A_G(c\mathbf{b}_0 + (c + \sqrt{1+c^2})\boldsymbol{\rho}) = (\sqrt{1+c^2} - c)\mathbf{a}_0$. In summary, we constructed a different set of parameters which has the same distribution as the true model.

Hence we conclude the proof. ■

References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using gibbs sampling. *Journal of educational statistics*, *17*(3), 251–269.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third berkeley symposium on mathematical statistics and probability* (Vol. 5, pp. 111–150).

Bollen, K. A. (1989). Structural equations with latent variables wiley. *New York*.

Celeux, G. (1985). The sem algorithm: a probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational statistics quarterly*, *2*, 73–82.

Diebolt, J., & Ip, E. H. (1996). Stochastic em: method and application. In *Markov chain monte carlo in practice* (pp. 259–273). Springer.

Fox, J.-P. (2003). Stochastic em for estimating the parameters of a multilevel irt model. *British Journal of Mathematical and Statistical Psychology*, *56*(1), 65–81.

Green, S., & Yang, Y. (2018). Empirical underidentification with the bifactor model: A case study. *Educational and Psychological Measurement*, *78*(5), 717–736.

Ip, E. H. (2002). On single versus multiple imputation for a class of stochastic algorithms estimating max. *Computational Statistics*, *17*, 517–524.

Zhang, S., Chen, Y., & Liu, Y. (2020). An improved stochastic em algorithm for large-scale full-information item factor analysis. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 44–71.