### THEORY AND METHODS



# Exact Exploratory Bi-factor Analysis: A Constraint-Based Optimization Approach

Jiawei Qiao<sup>1</sup>, Yunxiao Chen<sup>2</sup> and Zhiliang Ying<sup>3</sup>

<sup>1</sup>School of Mathematical Sciences, Fudan University, Shanghai, China; <sup>2</sup>Department of Statistics, London School of Economics and Political Science, London, UK; <sup>3</sup>Department of Statistics, Columbia University, New York, NY, USA

Corresponding author: Yunxiao Chen; Email: y.chen186@lse.ac.uk

### Abstract

Bi-factor analysis is a form of confirmatory factor analysis widely used in psychological and educational measurement. The use of a bi-factor model requires specifying an explicit bi-factor structure on the relationship between the observed variables and the group factors. In practice, the bi-factor structure is sometimes unknown, in which case, an exploratory form of bi-factor analysis is needed. Unfortunately, there are few methods for exploratory bi-factor analysis, with the exception of a rotation-based method proposed in Jennrich and Bentler ([2011, Psychometrika 76, pp. 537–549], [2012, Psychometrika 77, pp. 442–454]). However, the rotation method does not yield an exact bi-factor loading structure, even after hard thresholding. In this article, we propose a constraint-based optimization method that learns an exact bi-factor loading structure from data, overcoming the issue with the rotation-based method. The key to the proposed method is a mathematical characterization of the bi-factor loading structure as a set of equality constraints, which allows us to formulate the exploratory bi-factor analysis problem as a constrained optimization problem in a continuous domain and solve the optimization problem with an augmented Lagrangian method. The power of the proposed method is shown via simulation studies and a real data example.

Keywords: augmented Lagrangian method; bi-factor model; exploratory bi-factor analysis; hierarchical factor model

# 1. Introduction

The bi-factor model was originally proposed by Holzinger and Swineford (1937) for linear factor analysis and further extended by Gibbons and Hedeker (1992), Gibbons et al. (2007), Cai et al. (2011), among others, to nonlinear factor analysis settings to account for dichotomous, ordinal, and nominal data. These models assume that the observed variables can be accounted for by (G + 1) factors, with a general factor, onto which all items load directly, and *G* group factors that each is associated with a subset of variables. Such a specification leads to good interpretations in many real-world applications. These models have received wide applications in psychological and educational measurement (see, e.g., Bradlow et al., 1999; Cai et al., 2016; Chen et al., 2012; DeMars, 2006, 2012; Gibbons et al., 2009; Gignac & Watkins, 2013; Jeon et al., 2013; Reise et al., 2007; Rijmen, 2010). However, we note that all these applications of bi-factor analysis are confirmatory in the sense that one needs to pre-specify the number of group factors and the relationship between the observed variables and the group factors. Such prior knowledge may not always be available. In that case, an exploratory form of bi-factor analysis is needed.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/ licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

#### 2 Qiao et al.

Exploratory bi-factor analysis can be seen as a special case of exploratory factor analysis, which dates back to the seminal work of Thurstone (1947) concerning finding a "simple structure" of loadings. Various rotation methods have been proposed for exploratory factor analysis. A short list of relevant works includes Kaiser (1958), McCammon (1966), Jennrich and Sampson (1966), McKeon (1968), Crawford and Ferguson (1970), Yates (1987), Jennrich (2006), Jennrich (2004), Jennrich and Bentler (2011, 2012), and Liu et al. (2023). We refer the readers to Browne (2001) for a review of rotation methods for exploratory factor analysis.

However, standard exploratory factor analysis methods do not apply to the bi-factor analysis setting, and few methods have been developed for exploratory bi-factor analysis. An exception is the seminal work of Jennrich and Bentler (2011, 2012), who proposed a rotation-based method for exploratory bi-factor analysis with orthogonal or oblique factors. However, their approach has some limitations. First, as a common issue with rotation-based methods, their method does not yield many zero loadings, and thus, the resulting loading structure does not have an exact bi-factor structure. Although a post-hoc thresholding procedure (i.e., treating loadings with an absolute value below a threshold as zero) can be applied to obtain a cleaner loading pattern, it does not work well when some variables show relatively large loadings on more than one group factor after the rotation. In fact, one cannot always find a threshold that yields an exact bi-factor structure that each variable loads on one and only one group factor. Second, as noted in Jennrich and Bentler (2012), their method fails completely in the best case where there is a rotation of an initial loading matrix that has an exact bi-factor structure. This failure is due to that their rotation method cannot incorporate the zero constraints on the correlations between the general factor and the group factors.

This article proposes a constrained optimization method for exploratory bi-factor analysis, which overcomes the issues with the rotation-based method. The contribution is four-fold. First, we provide a mathematical characterization of the bi-factor loading structure as a set of nonlinear equality constraints, which allows us to formulate the exploratory bi-factor analysis problem as a constrained optimization problem. In other words, it turns a discrete model selection problem into a continuous optimization problem, which reduces the computational demand in some sense. It is shown that in the aforementioned best case where the rotation method fails, the global solutions to the optimization can perfectly recover the true bi-factor structure. Second, we propose an augmented Lagrangian method (ALM; Bertsekas, 2014) for solving this optimization problem, which is a standard numerical optimization method for solving constrained optimization with robust empirical performance and good theoretical properties. Third, we combine the proposed method with the Bayesian information criterion (BIC; Schwarz, 1978) for selecting the number of group factors. Compared with existing exploratory factor analysis methods for determining the number of factors, our method is tailored to the bi-factor model structure and, thus, tends to be statistically more efficient when the data is indeed generated by a bi-factor model. Finally, we demonstrate that the proposed method can be extended to learning the loading structure of hierarchical factor models (Schmid & Leiman, 1957; Yung et al., 1999), a higherorder extension of the bi-factor model that has received wide applications (see, e.g., Brunner et al., 2012, and the references therein). The bi-factor model can be viewed as a special hierarchical factor model with a two-layer factor structure, with the general factor in one layer and the group factors in the other. Similar to exploratory bi-factor analysis, the proposed method yields exact hierarchical factor loading structures without a need for post-hoc treatments.

The rest of the article is organized as follows. In Section 2, we formulate the exploratory bi-factor analysis problem as a constrained optimization problem and propose an ALM for solving it. We also propose a BIC-based procedure for selecting the number of group factors. Simulation studies and a real data example are presented in Sections 3 and 4, respectively, to evaluate the performance of the proposed method. We conclude with discussions in Section 5. The Appendix in the Supplementary Material includes additional details about the computation, the simulation studies and the real data example, an extension of the proposed method to exploratory hierarchical factor analysis, and proof of the theoretical results.

### 2. Exploratory bi-factor analysis by constrained optimization

### 2.1. Bi-factor model structure and a constrained optimization formulation

For the ease of exposition and simplification of the notation, we focus on the linear bi-factor model, while noting that the constraints that we derive for the bi-factor loading matrix below can be combined with the likelihood function of other bi-factor models (e.g., Cai et al., 2011; Gibbons et al., 2007; Gibbons & Hedeker, 1992) for their exploratory analysis. We focus on the extended bi-factor model, also known as the oblique bi-factor model, as considered in Jennrich and Bentler (2012) and Fang et al. (2021). This model is more general than the standard bi-factor model, in the sense that the latter assumes all the factors to be uncorrelated while the former allows the group factors to be correlated. As established in Fang et al. (2021), this extended bi-factor model is identifiable under mild conditions. We also point out that the proposed method can be easily adapted to the standard bi-factor model.

Consider a dataset with N observation units from a certain population and J observed variables. The extended bi-factor model assumes that there exists a general factor and G group factors. The group factors are loaded by independent clusters of variables, where each variable belongs to only one cluster. The model further assumes that the population covariance matrix of the observed variables can be decomposed as

$$\Sigma = \Lambda \Phi \Lambda^{\top} + \Psi,$$

where  $\Lambda = (\lambda_{jg})_{J \times (G+1)}$  is the loading matrix,  $\Phi = (\phi_{gg'})_{(G+1) \times (G+1)}$  is the correlation matrix of the factors, which is assumed to be strictly positive definite, and  $\Psi$  is a  $J \times J$  diagonal matrix with diagonal entries  $\psi_1, \ldots, \psi_J$ . Let  $\mathcal{B}_g \subset \{1, \ldots, J\}$  denote the cluster of variables loading on the *g*th group factor. Then the bi-factor model assumption implies that  $\mathcal{B}_g \cap \mathcal{B}_{g'} = \emptyset$ ,  $g \neq g'$ , and  $\cup_{g=1}^G \mathcal{B}_g = \{1, \ldots, J\}$ . The following zero constraints on the loading matrix  $\Lambda$  hold:

$$\lambda_{j,g+1} = 0$$
 for all  $j \notin \mathcal{B}_g$ .

In addition, the correlation matrix  $\Phi$  satisfies  $\phi_{1k} = 0$  for all  $k \neq 1$ , meaning that all the group factors are uncorrelated with the general factor. This constraint on  $\Phi$  is necessary to ensure that the extended bi-factor model is identifiable (Fang et al., 2021), as otherwise, there will be a rotational indeterminacy issue.

Now suppose that the number of group factors *G* is known, while the clusters  $\mathcal{B}_{g,g} = 1, \ldots, G$ , are unknown. Section 2.3 considers the selection of *G* when it is unknown. The bi-factor structure means that for each *j*, there is at most one nonzero element in  $(\lambda_{j,2}, \ldots, \lambda_{j,G+1})^{\top}$ . Consequently, the loading matrix  $\Lambda$  should satisfy the following J(G-1)G/2 constraints:

$$\lambda_{ik}\lambda_{ik'} = 0$$
, for all  $k, k' = 2, \dots, G+1, k \neq k', j = 1, \dots, J.$  (1)

Therefore, the exploratory bi-factor analysis problem can be translated into the following constrained optimization problem:

$$\min_{\Lambda, \Phi, \Psi} \quad l(\Lambda \Phi \Lambda^{\top} + \Psi; S)$$
s.t.  $\lambda_{jk} \lambda_{jk'} = 0$ , for all  $k, k' = 2, \dots, G + 1, k \neq k', j = 1, \dots, J$ ,  
 $\phi_{1k} = 0, k = 2, \dots, G + 1, \Phi$  is correlation matrix,  
and  $\Psi$  is a diagonal matrix,
(2)

where *l* is a loss function and *S* is the sample covariance matrix of observed data. We focus on the case when *l* is the fit function based on the normal likelihood

$$l(\Lambda\Phi\Lambda^{\top} + \Psi; S) = N(\log(\det(\Lambda\Phi\Lambda^{\top} + \Psi)) + \operatorname{tr}(S(\Lambda\Phi\Lambda^{\top} + \Psi)^{-1}) - \log(\det(S)) - J),$$

while noting that this loss function can be replaced by other loss functions for factor analysis (see, e.g., Chen et al., 2023), including the Frobenius norm of  $\Lambda\Phi\Lambda^{T} + \Psi - S$  that is used in the least square

estimator for factor analysis. We can also replace the sample covariance matrix in (2) with the sample correlation matrix, which is equivalent to performing exploratory bi-factor analysis based on variance-standardized variables.

The following theorem shows that the proposed method can perfectly recover the true bi-factor loading structure in the best case when  $S = \Sigma^*$ , where  $\Sigma^*$  is the true covariance matrix of data. Note that the rotation method proposed in Jennrich and Bentler (2012) completely fails in this case. Before giving the statement of the theorem, we introduce some additional notations. For any matrix  $A = (a_{i,j})_{i=1,...,m,j=1,...,n}$ ,  $S_1 \subset \{1,...,m\}$  and  $S_2 \subset \{1,...,n\}$ , let  $A[S_1,S_2] = (a_{i,j})_{i\in S_1,j\in S_2}$  be the submatrix of A consisting of elements that lie in rows belonging to set  $S_1$  and columns belonging to set  $S_2$ . For example, consider a matrix A with more than two rows and three columns. With index sets  $S_1 = \{1,2\}$  and  $S_2 = \{1,3\}$ , the submatrix  $A[S_1,S_2]$  is a two-by-two matrix, taking the form

$$A[\mathcal{S}_1, \mathcal{S}_2] = A[\{1, 2\}, \{1, 3\}] = \begin{pmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{pmatrix}.$$

For any set S, let |S| be the cardinality of S.

Let  $\{\mathcal{B}_g^*, g = 1, ..., G\}$  be the true nonoverlapping clusters of the *J* variables, satisfying for each  $j \in \mathcal{B}_g^*$ ,  $\lambda_{j,g+1}^* \neq 0, g = 1, ..., G$ . Further, let  $\mathcal{H}^* = \{g | \Lambda^*[B_g^*, \{1, 1+g\}]$  has column rank 2} be the set of group factors for which the group loadings are linearly independent of the corresponding common loadings. Let  $\mathcal{D}$  be the set of diagonal matrix with its diagonal entries taking values either 1 or -1, and let  $\mathcal{P}$  be the set of permutation matrix *P* such that each row and column of *P* has exactly one nonzero entry of value 1 and  $P_{11} = 1$ . Each matrix in  $\mathcal{D}$  corresponds to a simultaneous sign flip of certain factors and the corresponding loading parameters. Each matrix in  $\mathcal{P}$  corresponds to a swapping of the group factors. They are introduced to account for the sign-indeterminacy of the G + 1 factors and the label indeterminacy of the group factors, respectively. See Theorem 1 and Remark 1 for more explanations.

Let  $\Lambda^*$ ,  $\Phi^*$ , and  $\Psi^*$  be the true values of the corresponding parameter matrices. The following conditions are sufficient for the identifiability of the bi-factor structure and its parameters.

**Condition 1.** Given  $S = \Sigma^* = \Lambda^* \Phi^* (\Lambda^*)^\top + \Psi^*$ . Suppose that there exists another pair of parameters  $\Lambda, \Phi, \Psi$  satisfy the bi-factor structure constraints, we have  $\Lambda^* \Phi^* (\Lambda^*)^\top = \Lambda \Phi (\Lambda)^\top$  and  $\Psi^* = \Psi$ .

**Condition 2.**  $|\mathcal{H}^*| \ge 2$ . In addition, there exists  $g_1 \in \mathcal{H}^*$  such that  $|\mathcal{B}_{g_1}^*| \ge 3$  and any two rows of  $\Lambda^*[\mathcal{B}_{g_1}^*, \{1, 1+g_1\}]$  are linearly independent.

**Theorem 1.** Suppose that Conditions 1 and 2 hold. For any parameters  $\Lambda, \Phi, \Psi$  that satisfy  $S = \Sigma^* = \Lambda \Phi(\Lambda)^T + \Psi$ , there exist a diagonal sign-flip matrix  $D \in \mathcal{D}$  and a permutation matrix  $P \in \mathcal{P}$  such that  $\Lambda = \Lambda^* PD$  and  $\Phi = DP^T \Phi^* PD$ .

The proof of Theorem 1 is given in Section G.1 of the Supplementary Material.

**Remark 1.** We note that without additional information, the best we can achieve is to recover  $\Lambda^*$  and  $\Phi^*$  up to  $\Lambda = \Lambda^* PD$  and  $\Phi = DP^{\top} \Phi^* PD$ , where the permutation matrix *P* and sign-flip matrix *D* are necessary to account for the label and sign indeterminacies of the factor model. In that case,  $\Lambda^* \Phi^* (\Lambda^*)^{\top} = \Lambda \Phi(\Lambda)^{\top}$ , and thus, the model implied covariance matrix is the same. A similar indeterminacy issue also appears in exploratory factor analysis (see, e.g., Remark 1 in Liu et al., 2023).

**Remark 2.** Condition 1 ensures the separation between low rank matrix  $\Lambda^* \Phi^* (\Lambda^*)^\top$  and diagonal matrix  $\Psi^*$ . A sufficient condition for Condition 1 that can be easily checked in practice is given in Condition 3 below, which requires that each group has at least three nonzero group loadings and there exist at least three groups whose group loadings are linearly independent of the corresponding common loadings. We refer to Theorem 5.1 in Anderson and Rubin (1956) and Theorem 2 in Fang et al. (2021) for alternative sufficient conditions of Condition 1.

**Condition 3.**  $|\mathcal{B}_g^*| \ge 3$  for all  $g = 1, \ldots, G$  and  $|\mathcal{H}^*| \ge 3$ .

**Remark 3.** Condition 2 is similar to Condition E3S for Proposition 1 in Fang et al. (2021), where the latter is used to ensure the identifiability of parameters when the bi-factor structure is known. It is a sufficient condition that ensures if there is a pair of  $\Lambda$  and  $\Phi$  that also satisfies the constraints of a bi-factor model and  $\Lambda \Phi \Lambda^{\top} = \Lambda^* \Phi^* (\Lambda^*)^{\top}$ , then there must be a permutation matrix *P* and sign-flip matrix *D* such that  $\Lambda = \Lambda^* PD$  and  $\Phi = DP^{\top} \Phi^* PD$ . This condition first requires the existence of at least two group factors, for each of which the group loadings are linearly independent of the corresponding common loadings. It further requires that there exists a group factor  $g_1$  among these group factors, such that (1)  $g_1$  has at least three nonzero group loadings, and (2) any two-by-two submatrix of  $\Lambda^*$ , whose rows correspond to any two variables loading on  $g_1$  and columns correspond to the common factor and the group factor  $g_1$ , is of full rank. We note that the requirement of Condition 2 is very mild. In fact, the set of parameters not satisfying this condition has zero Lebesgue measure in the parameter space for bi-factor models satisfying that there are at least two group factors  $g_1$  and  $g_2$  such that  $|\mathcal{B}_{g_1}^*| \ge 3$  and  $|\mathcal{B}_{g_2}^*| \ge 2$ .

# 2.2. Proposed ALM

Following the previous discussion, we see that we can perform exploratory bi-factor analysis by solving the optimization problem with some equality constraints and the constraint that  $\Phi$  is a correlation matrix. To deal with the constraints in  $\Phi$ , we consider a reparameterization of  $\Phi$  based on a Cholesky decomposition, where the explicit form of the reparameterization is given in Section A of the Supplementary Material. With slight abuse of notation, we reexpress the covariance matrix as  $\Phi(\gamma)$ , where  $\gamma$  is a G(G-1)/2 dimensional unconstrained parameter vector. In addition, we use  $\psi = (\psi_1, \dots, \psi_J)^{\mathsf{T}}$  to denote the vector of diagonal entries of  $\Psi$  and reexpress the residual covariance matrix as  $\Psi(\psi)$ . Thus, the optimization problem (2) is now simplified as

$$\min_{\Lambda, \gamma, \psi} l(\Lambda \Phi(\gamma) \Lambda^{\top} + \Psi(\psi); S)$$
s.t.  $\lambda_{jk} \lambda_{jk'} = 0$ , for all  $k, k' = 2, \dots, G+1, k \neq k', j = 1, \dots, J$ ,
(3)

which is an equality-constrained optimization problem.

The standard approach for solving such a problem is the ALM (e.g., Bertsekas, 2014). This method aims to find a solution to (3) by solving a sequence of unconstrained optimization problems. Let *t* denote the *t*th unconstrained optimization problem in the ALM. The corresponding objective function, also known as the augmented Lagrangian function, takes the form

$$\min_{\Lambda, \gamma, \psi} l(\Lambda \Phi(\gamma) \Lambda^{\mathsf{T}} + \Psi(\psi); S) + \left( \sum_{j=1}^{J} \sum_{k=2}^{G} \sum_{k'=k+1}^{G+1} \beta_{jkk'}^{(t-1)} \lambda_{jk} \lambda_{jk'} \right) + \frac{1}{2} c^{(t-1)} \left( \sum_{j=1}^{J} \sum_{k=2}^{G} \sum_{k'=k+1}^{G+1} (\lambda_{jk} \lambda_{jk'})^2 \right), \quad (4)$$

where  $c^{(t-1)} > 0$  and  $\beta_{jkk'}^{(t-1)}$ s are auxiliary coefficients of the ALM determined by the initial values when t = 1 and the previous optimization when  $t \ge 2$ . Details of the ALM are given in Algorithm 1, in which the function *h* returns the second-largest value of a vector.

The ALM can be seen as a penalty method for solving constrained optimization problems. It replaces a constrained optimization problem with a series of unconstrained problems. It adds a penalty term, that is, the third term in (4), to the objective to enforce the constraints. The tuning parameter  $c^{(t-1)}$  can be seen as the weight of the penalty term. In fact, as  $c^{(t)}$  goes to infinity while  $\beta_{jkk'}^{(t)}$ s remain bounded, the solution has to converge to one satisfying the equality constraints in (3), as otherwise, the objective function value in (4) will diverge to infinity. However, the ALM is not purely a penalty method in the sense that it also adds the second term in (4) to mimic a Lagrange multiplier (see, e.g., Chapter 12, Nocedal & Wright, 1999), for which  $\beta_{jkk'}^{(t-1)}$ s are the weights. An advantage of the ALM is that, with the inclusion of the Lagrangian term (i.e., the second term), the method is guaranteed to converge to a local solution satisfying the equality constraints without requiring  $c^{(t)}$  to go to infinity. This is important, as when  $c^{(t)}$  is very large, the optimization problem (4) becomes ill-conditioned and thus hard to solve.

The updating rule of  $\beta_{jkk'}^{(t)}$  and  $c^{(t)}$  follows equations (1) and (47) in Chapter 2.2 of Bertsekas (2014). The updating rule for  $\beta_{jkk'}^{(t)}$  follows the first-order optimality condition based on optimizations (3) and (4). The updating rule for  $c^{(t)}$  ensures that it will become sufficiently large, which is necessary to guarantee the solution of the algorithm to converge to the feasible region defined by the zero constraints. On the other hand, it also prevents  $c^{(t)}$  from growing too quickly with which the optimization (4) is ill-conditioned. As shown in Chapter 2.2 of Bertsekas (2014), as long as the sequences  $\{\beta_{jkk'}^{(t)}\}$  remain bounded, the sequence  $\{c^{(t)}\}$  remains bounded. We follow the recommended choices of  $c_{\theta} = 0.25$  and  $c_{\sigma} = 10$  in Bertsekas (2014), while pointing out that the performance of Algorithm 1 is quite robust against the choices of these tuning parameters (see Section C of the Supplementary Material for a sensitivity analysis). The convergence of Algorithm 1 is guaranteed by Proposition 2.7 of Bertsekas (2014).

We remark on the stopping criterion in the implementation of Algorithm 1. We monitor the convergence of the algorithm based on two criteria: (1) the change in parameter values in two consecutive steps, measured by

$$\left(\|\Lambda^{(t)}-\Lambda^{(t-1)}\|_{F}^{2}+\|\boldsymbol{\gamma}^{(t)}-\boldsymbol{\gamma}^{(t-1)}\|^{2}+\|\boldsymbol{\psi}^{(t)}-\boldsymbol{\psi}^{(t-1)}\|^{2}\right)^{1/2}/\sqrt{J(G+2)+G(G-1)/2},$$

# Algorithm 1 Augmented Lagrangian Method for Exact Exploratory Bi-factor Analysis

**Input:** Initial value  $\Lambda^{(0)}$ ,  $\gamma^{(0)}$  and  $\psi^{(0)}$ , initial Lagrangian parameters  $\beta_{j,k,k'}^{(0)}$  for j = 1, ..., J, k = 2, ..., Gand k' = k + 1, ..., G + 1, initial penalty coefficient  $c^{(0)} > 0$ , constants  $c_{\theta} \in (0,1)$  and  $c_{\sigma} > 1$ , tolerances  $\delta_1, \delta_2 > 0$ .

1: while 
$$t = 1, 2, ... dc$$

2: Solve the following problem:

$$\Lambda^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\psi}^{(t)} = \arg\min_{\Lambda, \boldsymbol{\gamma}, \boldsymbol{\psi}} l(\Lambda \Phi(\boldsymbol{\gamma}) \Lambda^{\top} + \Psi(\boldsymbol{\psi}); S) + \left( \sum_{j=1}^{J} \sum_{k=2}^{G} \sum_{k'=k+1}^{G+1} \beta_{jkk'}^{(t-1)} \lambda_{jk} \lambda_{jk'} \right) + \frac{1}{2} c^{(t-1)} \left( \sum_{j=1}^{J} \sum_{k=2}^{G} \sum_{k'=k+1}^{G+1} (\lambda_{jk} \lambda_{jk'})^2 \right).$$

3: Update  $\beta_{ikk'}^{(t)}$  and  $c^{(t)}$  according to equations (5) and (6)

$$\beta_{jkk'}^{(t)} = \beta_{jkk'}^{(t-1)} + c^{(t-1)} \lambda_{jk}^{(t)} \lambda_{jk'}^{(t)},$$
(5)

and

$$c^{(t)} = \begin{cases} c_{\sigma} c^{(t-1)} & \text{if } \left( \sum_{j=1}^{J} \sum_{k=2}^{G} \sum_{k'=k+1}^{G+1} (\lambda_{jk}^{(t)} \lambda_{jk'}^{(t)})^2 \right)^{1/2} > c_{\theta} \left( \sum_{j=1}^{J} \sum_{k=2}^{G} \sum_{k'=k+1}^{G+1} (\lambda_{jk}^{(t-1)} \lambda_{jk'}^{(t-1)})^2 \right)^{1/2}; \\ c^{(t-1)} & \text{otherwise.} \end{cases}$$

$$\tag{6}$$

4: 
$$if \left( \|\Lambda^{(t)} - \Lambda^{(t-1)}\|_{F}^{2} + \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|^{2} + \|\boldsymbol{\psi}^{(t)} - \boldsymbol{\psi}^{(t-1)}\|^{2} \right)^{1/2} / \sqrt{J(G+2) + G(G-1)/2} < \delta_{1},$$
5: and  $\max_{j \in \{1,...,J\}} h(|\lambda_{j2}^{(t)}|,...,|\lambda_{j,G+1}^{(t)}|) < \delta_{2}$  then
6: Break
7: end if
8: end while
Output:  $\Lambda^{(t)}, \boldsymbol{y}^{(t)}, \boldsymbol{\psi}^{(t)}.$ 

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix and  $\|\cdot\|$  denotes the standard Euclidian norm, and (2) the distance between the estimate and the space of bi-factor loading matrices measured by

$$\max_{j \in \{1,...,J\}} h(|\lambda_{j2}^{(t)}|,...,|\lambda_{j,G+1}^{(t)}|)$$

We stop the algorithm when both criteria are below their pre-specified thresholds,  $\delta_1$  and  $\delta_2$ , respectively. The first criterion is a standard criterion for monitoring parameter convergence. This criterion being sufficiently small suggests the convergence of the algorithm. The second criterion is used to ensure that the solution is sufficiently close to the feasible set of optimization defined by the equality constraints. This criterion being below  $\delta_2$  means that for each variable *j*, there can only be one group loading whose absolute value is above the threshold  $\delta_2$ , and all the rest have absolute values below the threshold. Based on this, we can obtain an estimate of the bi-factor structure. More specifically, let *T* be the last iteration number. Then the estimated bi-factor model structure is given by

$$\widehat{\mathcal{B}}_g = \{j : |\lambda_{j,g+1}^{(T)}| > \delta_2\}.$$

By our choice of the stopping criterion, the resulting  $\widehat{\mathcal{B}}_g$ ,  $g = 1, \ldots, G$ , gives a partition of all the variables, and thus, the bi-factor structure is satisfied. For simulation studies in Section 3, we choose  $\delta_1 = \delta_2 = 10^{-2}$ . For real data analysis in Section 4, we choose  $\delta_1 = \delta_2 = 10^{-4}$  to get a more accurate and reliable result.

The optimization problem (4) is non-convex and can get stuck in a local minimum. Thus, we recommend running the proposed algorithm multiple times with random starting points and choosing the solution with the smallest objective function value. The algorithm can also suffer from slow convergence, especially when the penalty term becomes large. When the algorithm does not converge within  $T_{\text{max}}$  iterations, we suggest using the estimated parameters at the  $T_{\text{max}}$ th iteration as the initial parameters and restarting the optimization until a good proportion of them converge. In the simulation study in Section 3, the estimated parameters obtained using 50 random starting points are close to the global minimum in most cases in the simulation study. For the real data example in Section 4, 200 random starting points are used to ensure a reliable result. We set  $T_{\text{max}} = 100$  in all of our numerical studies.

# 2.3. Selecting the number of group factors

In Sections 2.1 and 2.2, the number of group factors *G* is treated as known. In practice, we can select its value based on the BIC (Schwarz, 1978). Let  $l_G$  denote the minimum loss function value in (2) when the number of group factors is *G*. As  $l_G$  differs from twice the negative log-likelihood of the bi-factor model with *G* group factors by a constant, and the numbers of nonzero parameters in  $\Lambda$  and  $\Psi$  do not depend on *G*, it is not difficult to see that the BIC of the bi-factor model with *G* group factors differs from  $l_G + ((G-1)G\log(N))/2$  by a constant. Note that (G-1)G/2 is the number of free parameters in the correlation matrix  $\Phi$ . Thus, we write

$$BIC_G = l_G + ((G-1)Glog(N))/2.$$

In practice, we choose the number of group factors *G* from a candidate set  $\mathcal{G}$ . For each value of  $G \in \mathcal{G}$ , we run the ALM described in Section 2.2 to obtain the value of  $l_G$ . We then compute BIC<sub>G</sub> and choose  $\widehat{G}$  with the smallest BIC value, that is,

$$\widehat{G} = \underset{G \in \mathcal{G}}{\operatorname{arg\,min}} \operatorname{BIC}_G.$$

### 3. Simulation study

### 3.1. Study I

In this study, we compare the proposed method with the oblique bi-factor rotation in Jennrich and Bentler (2012) regarding their performance in the estimation accuracy of parameters and the recovery of the bi-factor structure. We consider two different settings for the data generation mechanism: (1) the observed data are generated from an exact bi-factor model, and (2) the observed data are generated from an approximate bi-factor model, where the loading matrix is generated by adding small perturbations to an exact bi-factor loading matrix.

The oblique bi-factor rotation method first estimates the loading matrix  $\overline{\Lambda}$  under the exploratory factor analysis setting by the optimization problem

$$\widehat{\Lambda}, \widehat{\psi} = \underset{\Lambda \in \mathbb{R}^{J \times (G+1)}, \psi \in \mathbb{R}^{J}}{\operatorname{argmin}} l(\Lambda \Lambda^{\mathsf{T}} + \Psi(\psi); S).$$
(7)

We restrict  $\lambda_{ij} = 0$  for i = 2, ..., G and j = i + 1, ..., G + 1 to avoid the rotational indeterminacy of  $\widehat{\Lambda}$ , as suggested in Anderson and Rubin (1956). Then the rotated solutions  $\widehat{\Lambda}^{oblq}$  and  $\widehat{\Phi}^{oblq}$  are obtained by finding a rotation matrix that solves the optimization problem for oblique bi-factor rotation (Jennrich & Bentler, 2012). The implementation in the R package GPArotation (Bernaards & Jennrich, 2005) is used for solving this optimization problem, which is based on a gradient projection algorithm. The optimization problem for rotation is also non-convex and thus may converge to local solutions. For a fair comparison, we also use 50 random starting points for the initial rotation matrix, which is the same as the number of random starting points that are used when running Algorithm 1.

We first examine the accuracy in estimating the loading matrix. We calculate the mean squared error (MSE) for  $\widehat{\Lambda}$ , after adjusting for the label and sign indeterminacy as considered in Theorem 1 and further discussed in Remark 1. More specifically, let  $\mathcal{P}$  and  $\mathcal{D}$  be the sets of permutation and sign flip matrices, respectively, as defined in Theorem 1. We define the MSE for  $\widehat{\Lambda}$  as

$$MSE_{\widehat{\Lambda}} = \min_{P \in \mathcal{P}, D \in \mathcal{D}} \|\widehat{\Lambda} - \Lambda^* PD\|_F^2 / (J(1+G)).$$

Note that when data are generated from an approximate bi-factor model,  $\Lambda^*$  does not have an exact bi-factor structure. This MSE is calculated for the loading matrix estimates from both methods.

To compare the two methods in terms of their performance in recovering the bi-factor structure, we derive a sparse loading structure from the rotated solution by hard thresholding, a procedure also performed in Jennrich and Bentler (2012) for examining structure recovery. We let

$$\widehat{\mathcal{B}}_{g}^{oblq} = \{j : |\lambda_{j,g+1}^{oblq}| > \delta\},\$$

for g = 1, ..., G and some hard thresholding parameter  $\delta > 0$ . In the analysis below, we consider three choices of hard thresholding parameter  $\delta \in \{0.1, 0.2, 0.3\}$ . We note that  $\widehat{\mathcal{B}}_{g}^{oblq}$ , g = 1, ..., G, may not yield an exact bi-factor structure as it is not guaranteed to return only one nonzero group loading parameter for each variable.

Let  $\{\mathcal{B}_{g}^{*}, g = 1, ..., G\}$  be the true nonoverlapping clusters of the *J* variables, and let  $\{\mathcal{B}_{g}, g = 1, ..., G\}$  be their estimates, either from the proposed method or the rotation method. When data are generated from an approximate bi-factor model, the true group clusters  $\{\mathcal{B}_{g}^{*}, g = 1, ..., G\}$  are based on the corresponding bi-factor loading matrix before the perturbation. As the group factors can only be recovered up to label swapping, as Theorem 1 suggests, we measure the matching between the true and estimated structure up to a permutation of the factor labels. Specifically, the following two evaluation criteria are considered:

• Exact Match Criterion (EMC):  $\max_{\sigma \in \widetilde{\mathcal{P}}} \prod_{g=1}^{G} \mathbf{1}(\mathcal{B}_{\sigma(g)} = \mathcal{B}_{g}^{*})$ , which equals 1 when the bi-factor structure is correctly learned and 0 otherwise. Here,  $(\sigma(1), \ldots, \sigma(G))$  is a permutation of  $1, \ldots, G$ , and  $\widetilde{\mathcal{P}}$  is the set of all such permutations.

• Average Correctness Criterion (ACC):  $\max_{\sigma \in \widetilde{\mathcal{P}}} \sum_{g=1}^{G} (|\mathcal{B}_{\sigma(g)} \cap \mathcal{B}_{g}^{*}| + |\mathcal{B}_{\sigma(g)}^{C} \cap \mathcal{B}_{g}^{*C}|)/(JG)$ , which is the proportion of correctly identified zero and nonzero group loadings. Here, for any set  $\mathcal{B}, \mathcal{B}^{C} = \{1, \dots, J\} \setminus \mathcal{B}$  is the complement of set  $\mathcal{B}$ .

Here, the EMC measures the perfect recovery of the true bi-factor structure, while the ACC can be viewed as a smooth version of EMC that measures the level of partial recovery. EMC = 1 when ACC = 1 and, EMC = 0 when ACC < 1. A larger value of ACC indicates a higher level of partial recovery of the true bi-factor structure. More specifically, for a given permutation  $\sigma \in \widetilde{\mathcal{P}}$ , the quantity  $|\mathcal{B}_{\sigma(g)} \cap \mathcal{B}_g^*| + |\mathcal{B}_{\sigma(g)}^C \cap \mathcal{B}_g^*|$ 

To examine the recovery of the bi-factor structure, we consider  $(J,G) \in \{(15,3),(30,5)\}$  and  $N \in \{500,2,000\}$ . These choices, combined with the two settings for the data generation mechanism, lead to eight simulation settings. For each setting, we let  $\mathcal{B}_g^* = \{g,g+G,\ldots,g+G(J/G-1)\}$  for  $g = 1,\ldots,G$ ,  $\Psi^* = \mathbb{I}_{J\times J}$ , and  $\Phi^* = \Phi^*(\gamma^*)$  follow the reparameterization in Section 2.2, where the entries of  $\gamma^*$  are i.i.d., following a Uniform(-0.5,0.5) distribution. Under the settings where data are generated from an exact bi-factor model, we generate the true loading matrix  $\Lambda^*$  by

$$\lambda_{jk}^{*} = \begin{cases} u_{jk}, \text{ if } k = 1, \\ 0, \text{ if } k > 1, j \notin \mathcal{B}_{k-1}^{*}, \\ (1 - 2x_{jk})u_{jk}, \text{ if } k > 1, j \in \mathcal{B}_{k-1}^{*}, \end{cases}$$
(8)

for j = 1, ..., J and k = 1, ..., G+1. In (8),  $u_{jk}$ s are i.i.d., following a Uniform(0.2,1) distribution, and  $x_{jk}$ s are i.i.d., following a Bernoulli(0.5) distribution. Under the settings where data are generated from an approximate bi-factor model, we generate the true loading matrix  $\Lambda^*$  by

$$\lambda_{jk}^{*} = \begin{cases} u_{jk}, \text{ if } k = 1, \\ (1 - 2x_{jk})w_{jk}, \text{ if } k > 1, j \notin \mathcal{B}_{k-1}^{*}, \\ (1 - 2x_{jk})u_{jk}, \text{ if } k > 1, j \in \mathcal{B}_{k-1}^{*}, \end{cases}$$
(9)

for j = 1, ..., J and k = 1, ..., G+1. Here,  $u_{jk}$ s and  $x_{jk}$ s are generated in the same way as those in the exact bi-factor model, and  $w_{jk}$  are i.i.d., following a Uniform(0,0.1) distribution. In (9), the nonzero values of  $\lambda_{jk}^*$  when k > 1 and  $j \notin \mathcal{B}_{k-1}^*$  represent the perturbation of  $\Lambda^*$  from an exact bi-factor structure.

For each setting, we first generate  $\Lambda^*$  and  $\Phi^*$  once and use them to generate 100 datasets. The true parameter values for these simulations are given in Section B of the Supplementary Material. The results about the estimation of the loading matrix are shown in Table 1. When data are generated from an exact bi-factor model, the proposed method outperforms the rotation method in terms of the MSE of the estimated loading matrix, as shown in Table 1(a). When data are generated from an approximate bi-factor model, as shown in Table 1(b), the proposed method is slightly better under the small-sample settings when N = 500 but slightly worse under the large-sample settings when N = 2,000. The disadvantage of the proposed method under the large-sample settings is due to the bias brought by model misspecification. That is, the data generation model is not an exact bi-factor model, while the proposed method restricts its estimates in the space of exact bi-factor models.

The results about the recovery of the bi-factor structure based on the EMC and ACC metrics are shown in Tables 2 and 3, respectively. For the rotation method, the threshold  $\delta = 0.2$  yields the best results among the three threshold choices under all the simulation settings and for both performance

<b>Table 1.</b> Simulation results of the MSE of $\overline{\Lambda}$ estimated by the
proposed ALM method and the exploratory bi-factor rotation
nethod

( <i>J</i> , <i>G</i> )	Ν	ALM	Rotation				
(a) The exact bi-factor model cases.							
(15,3)	500	$2.10\times10^{-3}$	$3.62 \times 10^{-3}$				
	2,000	$0.54 \times 10^{-3}$	$0.92 \times 10^{-3}$				
(30,5)	500	$1.36 \times 10^{-3}$	$4.94 \times 10^{-3}$				
	2,000	$0.30 \times 10^{-3}$	$1.15 \times 10^{-3}$				
(b) The approximate bi-factor model cases.							
(15,3)	500	$4.74 \times 10^{-3}$	$5.92 \times 10^{-3}$				
	2,000	$3.06 \times 10^{-3}$	$2.40 \times 10^{-3}$				
(30,5)	500	$3.75 \times 10^{-3}$	$3.88 \times 10^{-3}$				
	2,000	$2.63 \times 10^{-3}$	$1.25 \times 10^{-3}$				

**Table 2.** Simulation results of the EMC of the proposed ALM method and the exploratory bi-factor rotation method with three choices of hard thresholding parameter  $\delta$ 

( <i>J</i> , <i>G</i> )	Ν	ALM	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$			
(a) The exact bi-factor model cases.								
(15,3)	500	1.00	1.00 0.18 0.90		0.28			
	2,000	1.00	0.99	1.00	0.50			
(30,5)	500	0.85	0.00	0.15	0.00			
	2,000	1.00	0.55	0.68	0.00			
(b) The approximate bi-factor model cases.								
(15,3)	500	0.99	0.00	0.62	0.33			
	2,000	1.00	0.02	0.94	0.67			
(30,5)	500	0.86	0.00	0.23	0.00			
	2,000	1.00	0.00	0.82	0.00			

metrics. However, even the results of the rotation method under this choice of threshold are not as good as those from the proposed method, especially when we look at the EMC metric. For example, when J = 30, G = 5, and N = 500, the rotation method with  $\delta = 0.2$  can only correctly recover the entire bifactor structure 15 times among 100 simulations, while the proposed method can correctly recover it 85 times.

### 3.2. Study II

In this study, we examine the selection of the number of factors by BIC<sub>*G*</sub> in Section 2.3. We compare it with selecting the number of factors under the exploratory factor analysis model without assuming a bi-factor structure. For the proposed method, we set the candidate set  $\mathcal{G} = \{G^* - 1, G^*, G^* + 1\}$ , where  $G^*$  is the true number of group factors. For exploratory factor analysis, we also use the BIC for determining the number of factors, which is defined as

$$BIC_{K}^{e} = l_{K}^{e} + (JK - K(K - 1)/2)\log(N),$$

Table 3. Simulation results of the ACC of the proposed ALM method and the exploratory bi-factor rotation method with three choices of hard thresholding parameter  $\delta$ 

(J,G)	Ν	ALM	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$			
(a) The exact bi-factor model cases.								
(15,3)	500	1.000	0.966	0.998	0.976			
	2,000	1.000	0.999	1.000	0.987			
(30,5)	500	0.998	0.892	0.987	0.973			
	2,000	1.000	0.996	0.998	0.980			
(b) The approximate bi-factor model cases.								
(15,3)	500	0.999	0.864	0.989	0.978			
	2,000	1.000	0.928	0.999	0.991			
(30,5)	500	0.998	0.848	0.989	0.978			
	2,000	1.000	0.927	0.999	0.980			

 Table 4. Simulation results of the selection of the number of factors by BIC

		AI	_M	Exploratory		
$(J,G^{^{\star}})$	Ν	Ĝ	SC	Ĝ	SC	
(15,3)	500	3	1	3	1	
	2,000	3	1	3	1	
(30,5)	500	5.02	0.98	4	0	
	2,000	5	1	4.99	0.99	

where *K* is the number of factors in the exploratory factor analysis model, and  $l_K^e = l(\widehat{\Lambda}\widehat{\Lambda}^\top + \Psi(\widehat{\psi}); S)$  with  $\widehat{\Lambda}$  and  $\widehat{\psi}$  from (7). As the number of factors in the exploratory factor analysis model equals the number of group factors plus one, we choose *K* from the candidate set  $\mathcal{K} = \{G+1 : G \in \mathcal{G}\}$ . Let  $\widehat{K} = \arg\min_{K \in \mathcal{K}} BIC_K^e$ . Then the estimate of *G* by exploratory factor analysis is  $\widehat{G} = \widehat{K} - 1$ . The selection accuracy is evaluated by the selection correctness (SC) criterion, defined as  $\mathbf{1}(\widehat{G} = G^*)$ , where  $\widehat{G}$  is obtained using the proposed method in Section 2.3 or under the exploratory factor analysis model described above.

We conduct simulations under four settings, with  $(J,G^*) \in \{(15,3),(30,5)\}$  and  $N \in \{500,2,000\}$ and the data generation models being the same exact bi-factor models in Study I. For each setting, 100 independent simulations are performed. The results are given in Table 4, where the column indicated by  $\tilde{G}$  reports the average value of  $\hat{G}$ . We see that both methods can select the number of factors reasonably well, with their accuracy being 100% when  $G^* = 3$  for both sample sizes. When  $G^* = 5$  and the sample size N = 2,000, the proposed method achieves an accuracy of 100%, and the exploratory factor analysis method achieves an accuracy of 99%. This is not surprising, given that the BIC has asymptotic consistency in selecting the number of factors under both models. It is worth noting that, when  $G^* = 5$  and for the smaller sample size N = 500, which is the most challenging setting, the proposed method achieves an accuracy of 98%, while that of the exploratory factor analysis method is zero. More precisely, the exploratory factor analysis method selects G = 4 in all the replications. It suggests that the proposed method has an advantage in smaller sample settings. This result is expected, as the exploratory factor analysis method doesn't utilize the information about the bi-factor structure. Consequently, it overestimates the number of parameters, which leads to a larger penalty term and, subsequently, a tendency to under-select G.

# 4. Real data analysis

In this section, we apply the exact exploratory bi-factor analysis to a personality assessment dataset based on the International Personality Item Pool (IPIP) NEO 120 personality inventory (Johnson, 2014).<sup>1</sup> We investigate the structure of the Extraversion scale based on a sample of 1,107 U.K. male participants aged between 25 and 30 years. This scale consists of 24 items, which are designed to measure six facets of Extraversion, including Friendliness (E1), Gregariousness (E2), Assertiveness (E3), Activity Level (E4), Excitement-Seeking (E5), and Cheerfulness (E6) (see Section E of the Supplementary Material for the details). All the items are on a 1-5 Likert scale, and the reversely worded items have been reversely scored so that a larger score always means a higher level of extraversion. There is no missing data. Detailed descriptions of the items can be found in the Section E of the Supplementary Material.

Using a candidate set  $\mathcal{G} = \{2, ..., 12\}$ , the BIC procedure given in Section 2.3 selects seven group factors. The estimated loading matrix is given in Table 5, and the estimated factor correlation matrix is given below. The estimated model fits the data well, as implied by the commonly used fit statistics, including RMSEA = 0.044, SRMR = 0.031, CFI = 0.965, and TLI = 0.953. We point out that the estimated bi-factor structure does not meet Condition 3, one of the sufficient conditions for Theorem 1. However, as shown in Section G.2 of the Supplementary Material, with some additional mild assumptions, this structure and its parameters are still identifiable.

We now examine the estimated model. We first notice that the loadings on the general factor are all positive. Consequently, this factor can be easily interpreted as the general extraversion factor. The seven group factors are closely related to the six aspects of extraversion. Specifically, we interpret the group factors G1, G3, G4, and G5 as the Friendliness, Cheerfulness, Assertiveness, and Activity Level factors, respectively, as the items loading on them highly overlap with the items that are used to define the corresponding aspects. In particular, the items loading on G3 and G5 are exactly those that define the Cheerfulness and Activity Level aspects, respectively. The items loading on G1 include all the items that define the Friendliness aspect and an additional item "7. Prefer to be alone," a negatively worded item that is used to define the Gregariousness aspect. This additional item aligns well with the Friendliness that define the Assertiveness aspect and an additional item "6. Talk to a lot of different people at parties," which is used to define the Gregariousness aspect. This additional item aligns with the Assertiveness dimension in that talking to many different people at parties typically requires sufficient confidence, a key element of Assertiveness.

$$\widehat{\Phi} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.24 & 0.54 & 0.37 & 0.16 & 0.51 & 0.08 \\ 0 & -0.24 & 1 & -0.04 & 0.05 & -0.09 & -0.01 & 0.51 \\ 0 & 0.54 & -0.04 & 1 & 0.30 & 0.28 & 0.20 & 0.25 \\ 0 & 0.37 & 0.05 & 0.30 & 1 & 0.38 & 0.15 & 0.29 \\ 0 & 0.16 & -0.09 & 0.28 & 0.38 & 1 & 0.11 & 0.22 \\ 0 & 0.51 & -0.01 & 0.20 & 0.15 & 0.11 & 1 & 0.10 \\ 0 & 0.08 & 0.51 & 0.25 & 0.29 & 0.22 & 0.10 & 1 \end{pmatrix}$$

The group factors G2 and G7 may be viewed as two subdimensions of the Excitement-Seeking aspect, as each of them is loaded with two items that define the Excitement-Seeking aspect. Specifically, G2 is loaded with the items "19. Enjoy being reckless" and "20. Act wild and crazy," while G7 is loaded with the items "17. Love excitement" and "18. Seek adventure." We believe that G2 emphasizes the thrill of the

<sup>&</sup>lt;sup>1</sup>The data are downloaded from https://osf.io/tbmh5/.

Items	Sign	General	G1	G2	G3	G4	G5	G6	G7
1	+E1	0.85	0.26	0	0	0	0	0	0
2	+E1	0.73	0.48	0	0	0	0	0	0
3	-E1	0.74	0.57	0	0	0	0	0	0
4	-E1	0.68	0.58	0	0	0	0	0	0
5	+E2	0.94	0	0	0	0	0	0.26	0
6	+E2	1.01	0	0	0	0.17	0	0	0
7	-E2	0.53	0.52	0	0	0	0	0	0
8	-E2	0.67	0	0	0	0	0	1.06	0
9	+E3	0.37	0	0	0	0.86	0	0	0
10	+E3	0.38	0	0	0	0.81	0	0	0
11	+E3	0.28	0	0	0	0.74	0	0	0
12	-E3	0.39	0	0	0	0.75	0	0	0
13	+E4	0.20	0	0	0	0	0.81	0	0
14	+E4	0.40	0	0	0	0	0.82	0	0
15	+E4	0.40	0	0	0	0	0.60	0	0
16	-E4	0.04	0	0	0	0	0.47	0	0
17	+E5	0.46	0	0	0	0	0	0	0.46
18	+E5	0.47	0	0	0	0	0	0	0.71
19	+E5	0.35	0	0.86	0	0	0	0	0
20	+E5	0.56	0	0.71	0	0	0	0	0
21	+E6	0.59	0	0	0.42	0	0	0	0
22	+E6	0.64	0	0	0.48	0	0	0	0
23	+E6	0.46	0	0	0.76	0	0	0	0
24	+E6	0.41	0	0	0.74	0	0	0	0

**Table 5.** Estimated bi-factor loading matrix  $\widehat{\Lambda}$  with seven group factors

moment of excitement and the disregard for potential consequences, while G7 emphasizes the pursuit of meaningful and fulfilling experiences. Therefore, we interpret G2 as the Reckless Excitement-Seeking factor, while interpret G7 as the Meaningful Excitement-Seeking factor. Finally, G6 is loaded with two items, "5. Love large parties" and "8. Avoid crowds," where item 8 is reversely worded. Both items are used to define the Gregariousness aspect. Compared with items 6 and 7, which are also used to define the Gregariousness aspect but now load on two different group factors, these two items may better reflect the essence of Gregariousness—the tendency to enjoy the company of others. We thus interpret G6 as the Gregariousness factor. We also notice that most correlations between the group factors are positive, except for some of the correlations with G2. Specifically, we see that G2 (Reckless Excitement-Seeking) has a moderate negative correlation with G1 (Friendliness) while a reasonably high correlation with G7 (Meaningful Excitement-Seeking).

We have also applied the bi-factor rotation method of Jennrich and Bentler (2012) to the same data, which gives a solution with seven group factors. The resulting bi-factor structure is similar to that given by the proposed method, except that the rotation solution does not seem to contain a clear Friendliness factor (see Section F of the Supplementary Material for further details).

### 5. Discussions

This article proposes a constraint-based optimization method for exploratory bi-factor analysis. This method turns the problem of exploratory bi-factor analysis into an equality-constrained optimization problem in a continuous domain and solves this optimization problem by an ALM. Compared with the rotation method of Jennrich and Bentler (2011, 2012), the proposed method can learn an exact loading structure without a post-hoc treatment step. In the simulation studies, the ALM method achieves higher estimation accuracy when data are generated from an exact bi-factor model. In addition, it has a higher chance of recovering the true bi-factor structure than the rotation method, whether data are generated from an exact or approximate bi-factor model. Moreover, the ALM method correctly estimates the number of the group factors in most of the simulation replications. In the real data analysis concerning an Extraversion personality scale, the ALM method yields a bi-factor structure with seven group factors. The identified group factors are psychologically interpretable.

An innovation of current research is turning a model selection problem, which is combinatory by nature, into a continuous optimization problem. This avoids a computationally intensive search procedure for fitting many possible models and comparing their fits, noting that the number of possible models grows exponentially with *J*. We admit that this continuous optimization formulation also has a limitation. The space for the bi-factor loading matrix characterized by the nonlinear equality constraints in (1) is highly non-convex, and thus, the ALM may sometimes converge to a local minimum. To alleviate this issue, we suggest running the ALM with multiple random starting points and then choosing the solution with the smallest objective function value. Based on our simulation results, using 50 starting points seems sufficient to converge to somewhere close to the true parameters up to a label swapping of the group factors and a sign indeterminacy of loadings in almost all replications under the settings considered in the simulation study.

This research leads to several new directions for exploratory analysis of factor models with structure constraints on the loading matrix. First, as pointed out earlier, the proposed approach can be easily adapted to nonlinear bi-factor models for dichotomous, ordinal, and nominal data. Under the confirmatory setting, these models are typically estimated by maximizing the marginal log-likelihood function or other objective functions (e.g., a composite likelihood). Under the exploratory setting, one only needs to maximize the same objective function subject to the same bi-factor constraints in (1), for which the ALM adapts naturally. It is worth noting that, however, the marginal likelihood of the nonlinear bi-factor models typically involves multidimensional integrals with respect to the factors, and they do not have an analytic form. Consequently, solving the Lagrangian augmented objective functions using the standard expectation-maximization (EM) algorithm (Bock & Aitkin, 1981; Dempster et al., 1977) can be computationally intensive. One possible solution is to use a stochastic approximation method (Zhang & Chen, 2022). These methods avoid the high computational cost of numerical integrals in the EM algorithm by constructing stochastic gradients of the marginal log-likelihood through Markov chain Monte Carlo sampling.

Second, the proposed constraints can also be combined with exploratory factor analysis techniques to learn a bi-factor structure in two steps. Suppose an initial loading matrix estimate  $\widehat{\Lambda}$  has been obtained under the constraint that the factors are orthogonal (i.e.,  $\Phi$  is an identity matrix). It may be obtained by a standard exploratory factor analysis method. In that case, we can find a bi-factor structure that best approximates  $\widehat{\Lambda}$  (up to a rotation) by minimizing  $\|\Lambda \Phi(\boldsymbol{\gamma})\Lambda^{\top} - \widehat{\Lambda}(\widehat{\Lambda})^{\top}\|_{F}$  with respect to  $\Lambda$  and  $\boldsymbol{\gamma}$  under the constraints in (1). This optimization can again be solved by an ALM.

Third, as we demonstrate in Section D of the Supplementary Material, the set of constraints in (1) can be extended to characterize the loading structure of a hierarchical factor model (Schmid & Leiman, 1957; Yung et al., 1999), which can be used to learn a hierarchical factor structure. This exploratory hierarchical factor analysis may allow researchers to learn more refined and interpretable latent structures from psychometric data. However, one should note that exploratory hierarchical factor analysis is more complex than exploratory bi-factor analysis, as the factor hierarchy in the former can be much more complex than the two-layer hierarchy in the latter. The learning algorithm in Section D of the Supplementary Material requires the factor hierarchy to be known (see, e.g., Panel (b) of Figure D.1

in the Supplementary Material). The problem becomes more challenging when the factor hierarchy is unknown, in which case we need to learn both the factor hierarchy and the loading pattern of the variables. We leave this problem for future investigation.

Finally, we point out that the proposed method always returns an estimated bi-factor model, whether it fits the data or not. The simulation study in Section 3 shows that the proposed method has robust performance when data are generated by an approximate bi-factor model. However, under more general settings, it remains to test the goodness-of-fit of the estimated model to decide whether a bi-factor model suits the data. If the bi-factor model does not fit the data well, we may consider a more flexible factor model. For example, we may apply the bi-factor rotation method or a rotation method for traditional exploratory factor analysis to allow for more cross-loadings. Alternatively, we may learn approximate bi-factor models in an exploratory manner by relaxing the equality constraints in (2) with inequality constraints in the form of  $|\lambda_{ik}\lambda_{ik'}| \leq \varepsilon$ , for all j = 1, ..., J, and  $k, k' = 2, ..., G+1, k \neq k'$ , where  $\varepsilon$  is a tuning parameter that controls the level of approximation to a bi-factor model. A larger value of  $\varepsilon$  leads to a more flexible model space and, thus, a more satisfactory fit, while a smaller value of  $\varepsilon$ leads to a better approximation to a bi-factor model that may have better interpretability. In this sense,  $\varepsilon$  provides a trade-off between model goodness-of-fit and bi-factor interoperability. This inequalityconstrained optimization may be solved using an interior-point method, which can incorporate the inequality constraints through suitable barrier functions (e.g., log-barriers). We leave this idea for future investigation.

Data availability statement. Code implementing the proposed method is open-source and publicly available at https://github.com/EmetSelch97/Eebf.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2025.17.

Funding statement. No funding was received for conducting this study.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

### References

- Anderson, T., & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 111. University of California Press.
- Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65(5), 676–696.
- Bertsekas, D. P. (2014). Constrained optimization and Lagrange multiplier methods. Academic Press.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443–459.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. Psychometrika, 64, 153-168.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150.
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80(4), 796–846.
- Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item response theory. *Annual Review of Statistics and Its Application*, 3(1), 297–321.
- Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16(3), 221.
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, 80(1), 219–251.
- Chen, Y., Moustaki, I., & Zhang, S. (2023). On the estimation of structural equation models with latent variables. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 145–162). Guilford Publications.
- Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, 35(3), 321–332.
- DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, 43(2), 145–168.

DeMars, C. E. (2012). Confirming testlet effects. Applied Psychological Measurement, 36(2), 104-121.

- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal* of the Royal Statistical Society: Series B (Methodological), 39(1), 1–22.
- Fang, G., Guo, J., Xu, X., Ying, Z., & Zhang, S. (2021). Identifiability of bifactor models. Statistica Sinica, 31, 2309–2330.
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31(1), 4–19.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. Psychometrika, 57(3), 423-436.
- Gibbons, R. D., Rush, A. J., & Immekus, J. C. (2009). On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. *Journal of Psychiatric Research*, 43(4), 401–410.
- Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, 48(5), 639–662.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. Psychometrika, 2(1), 41-54.
- Jennrich, R. I. (2004). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika*, 69, 257–273.
- Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, 71, 173–191.
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. Psychometrika, 76, 537-549.
- Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. Psychometrika, 77(3), 442-454.

Jennrich, R. I., & Sampson, P. (1966). Rotation for simple loadings. Psychometrika, 31(3), 313-323.

- Jeon, M., Rijmen, F., & Rabe-Hesketh, S. (2013). Modeling differential item functioning using a generalization of the multiplegroup bifactor model. *Journal of Educational and Behavioral Statistics*, 38(1), 32–60.
- Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78–89.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23(3), 187-200.
- Liu, X., Wallin, G., Chen, Y., & Moustaki, I. (2023). Rotation to sparse loadings using *l<sup>p</sup>* losses and related inference problems. *Psychometrika*, 88(2), 527–553.
- McCammon, R. B. (1966). Principal component analysis and its application in large-scale correlation studies. *The Journal of Geology*, 74(5, Part 2), 721–733.
- McKeon, J. (1968). Rotation for maximum association between factors and tests. Unpublished manuscript, Biometric Laboratory, George Washington University.
- Nocedal, J., & Wright, S. J. (1999). Numerical optimization. Springer.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19–31.
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47(3), 361–372.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. Psychometrika, 22(1), 53-61.
- Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461-464.
- Thurstone, L. L. (1947). Multiple factor analysis. University of Chicago Press.
- Yates, A. (1987). Multivariate exploratory data analysis: A perspective on exploratory factor analysis. State University of New York Press.
- Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64, 113–128.
- Zhang, S., & Chen, Y. (2022). Computation for latent variable model estimation: A unified stochastic proximal framework. *Psychometrika*, 87(4), 1473–1502.

Cite this article: Qiao, J., Chen, Y. and Ying, Z. (2025). Exact Exploratory Bi-factor Analysis: A Constraint-Based Optimization Approach. *Psychometrika*, 1–16. https://doi.org/10.1017/psy.2025.17