

## FACTOR ANALYSIS PROCEDURES REVISITED FROM THE COMPREHENSIVE MODEL WITH UNIQUE FACTORS DECOMPOSED INTO SPECIFIC FACTORS AND ERRORS

KOHEI ADACHI 

OSAKA UNIVERSITY

Factor analysis (FA) procedures can be classified into three types (Adachi in *WIREs Comput Stat* <https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1458>, 2019): latent variable FA (LVFA), matrix decomposition FA (MDFA), and its variant (Stegeman in *Comput Stat Data Anal* 99: 189–203, 2016) named completely decomposed FA (CDFA) through the theorems proved in this paper. We revisit those procedures from the Comprehensive FA (CompFA) model, in which a multivariate observation is decomposed into common factor, specific factor, and error parts. These three parts are separated in MDFA and CDFA, while the specific factor and error parts are not separated, but their sum, called a unique factor, is considered in LVFA. We show that the assumptions in the CompFA model are satisfied by the CDFA solution, but not completely by the MDFA one. Then, how the CompFA model parameters are estimated in the FA procedures is examined. The study shows that all parameters can be recovered well in CDFA, while the sum of the parameters for the specific factor and error parts is approximated by the LVFA estimate of the unique factor parameter and by the MDFA estimate of the specific factor parameter. More detailed results are given through our subdivision of the CompFA model according to whether the error part is uncorrelated among variables or not.

**Key words:** comprehensive factor analysis model, matrix decomposition factor analysis, completely decomposed factor analysis, latent variable factor analysis, Inter-variable error correlations.

Factor analysis (FA) was originally conceived of by Spearman (1904) and developed toward its modern form by Thurstone (1947). FA is performed for a multivariate data set in order to extract two types of mutually uncorrelated factors: common factors and all others. The common factors, whose number is much less than that of observed variables, serve to explain the variation of all variables. On the other hand, the other factors explain the variations in the variables that remain unaccounted for by the common factors. The latter factors are referred to as unique factors in a prevalent FA model, while those factors are called specific factors in another FA model. Those two FA models are introduced in the following two paragraphs.

The prevalent FA model is expressed as

$$\mathbf{x} = \mathbf{\Lambda}\mathbf{f} + \tilde{\mathbf{e}} = \mathbf{\Lambda}\mathbf{f} + \mathbf{\Psi}\mathbf{u} \quad (1)$$

for a  $p$ -variate observed vector  $\mathbf{x}(p \times 1)$ , with  $\tilde{\mathbf{e}} = \mathbf{\Psi}\mathbf{u}$  and the expectations of the elements in  $\mathbf{x}$  being zero (e.g., Bartholomew, Knott, & Moustaki, 2011; Mulaik, 2010; Yanai & Ichikawa, 2007). Here, vectors  $\mathbf{f}(m \times 1)$  and  $\mathbf{u}(p \times 1)$  contain the common and unique factor scores, respectively, with  $m < p$ . The scores in  $\mathbf{f}$  and  $\mathbf{u}$  are treated as random latent variables, while  $\mathbf{\Lambda}(p \times m)$  and  $\mathbf{\Psi}(p \times p)$  are the nonrandom parameter matrices to be estimated. The elements of  $\mathbf{\Lambda}$  are called factor loadings, as they describe how the  $p$  variables in  $\mathbf{x}$  load the  $m$  common factors in  $\mathbf{f}$ . In contrast to  $\mathbf{\Lambda}$  being unconstrained,  $\mathbf{\Psi}$  is restricted to a diagonal matrix. This implies that the  $p$  variables in  $\mathbf{x}$  have one-to-one correspondences to the  $p$  unique factors in  $\mathbf{u}$ : its  $j$ th element

Correspondence should be made to Kohei Adachi, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, Japan.  
Email: [adachi@hus.osaka-u.ac.jp](mailto:adachi@hus.osaka-u.ac.jp)

uniquely affects the  $j$ th one of  $\mathbf{x}$  with the  $j$ th diagonal element of  $\Psi$  being a coefficient. We refer to the FA procedure based on (1) as *latent variable FA (LVFA)* following Adachi (2019), in order to distinguish it from a procedure to be introduced later.

The vector  $\tilde{\mathbf{e}} = \Psi\mathbf{u}$  in (1) is sometimes referred to as an error vector; however, the classic literature such as Harman (1976), Reymont & Jöreskog (1993), and Thurstone (1947) includes the description that  $\tilde{\mathbf{e}} = \Psi\mathbf{u}$  is divided into two vectors as

$$\Psi\mathbf{u} = \Theta\mathbf{s} + \mathbf{e}. \quad (2)$$

Here,  $\mathbf{e}$  (rather than  $\tilde{\mathbf{e}}$ ) is referred to as an error vector, and the elements of  $\mathbf{s}$  ( $p \times 1$ ) are called specific factor scores, with  $\Theta$  ( $p \times p$ ) being diagonal. That is, the FA model is also introduced by incorporating (2) into (1), i.e.,

$$\mathbf{x} = \Lambda\mathbf{f} + \Theta\mathbf{s} + \mathbf{e}. \quad (3)$$

Here,  $\Theta\mathbf{s}$  is found to perform a role similar to the unique factor part  $\Psi\mathbf{u}$  in (1):  $\Theta$  is diagonal as is  $\Psi$ , which implies that the  $j$ th specific factor score in  $\mathbf{s}$  specifically (i.e., uniquely) affects the  $j$ th variable of  $\mathbf{x}$ , with the  $j$ th diagonal element of  $\Theta$  a coefficient. As found in this sentence, the adjective "specific" used for  $\mathbf{s}$  in (3) has the same implication as the "unique" for  $\mathbf{u}$  in (1): "specific" and "unique" merely serve to avoid the confusion between  $\Psi\mathbf{u}$  in (1) and  $\Theta\mathbf{s}$  in (3). However, the error vector  $\mathbf{e}$  is included in (3) in addition to  $\Theta\mathbf{s}$ , while the error vector  $\tilde{\mathbf{e}}$  in (1) equals  $\Psi\mathbf{u}$ . That is, each error in  $\tilde{\mathbf{e}}$  is assumed to uniquely affect the corresponding variable in (1), but  $\Theta\mathbf{s}$  is separated from  $\mathbf{e}$ . In this point, (3) is more generalized or comprehensive than (1). We thus refer to (3) as a *comprehensive FA (CompFA) model*. This model can also be considered compatible with Spearman's (1904) original conception of FA (Anderson & Rubin, 1956, p. 112; Yanai, Shegemasu, Mayekawa, & Ichikawa, 1990, p. 4).

However, the CompFA model has been left out of consideration in recent FA studies, as found in the fact that (3) is not treated in the major introductions to FA published in this century (e.g., Bartholomew, Knott, & Moustaki, 2011; Mulaik, 2010). That can be attributed to the following two points seen in the above classic literature: First, the CompFA model has been very briefly mentioned on only a couple of pages in Harman (1976, pp. 19–20), Reymont and Jöreskog, (1993, pp. 75–76), and Thurstone (1947, pp. 74–75); thus, model (3) was not impressed as fulfilling a certain role in FA. Second, in that literature, (3) has been introduced merely as a model; no procedure is described for estimating parameters based on (3).

However, a recently proposed FA procedure, which is called matrix decomposition FA (MDFA) to distinguish it from LVFA (Adachi, 2019; Adachi & Trendafilov, 2018), can be viewed as a parameter estimation procedure for the CompFA model. In more exactness, MDFA can be modeled as a nonrandom matrix version of (3):

$$\mathbf{X} = \mathbf{F}\Lambda' + \mathbf{S}\Theta + \mathbf{E}. \quad (4)$$

Here,  $\mathbf{X}$  is an  $n$ -observations  $\times$   $p$ -variables column-centered data matrix,  $\mathbf{F}$  is the  $n \times m$  matrix of common factor scores,  $\mathbf{S}$  is the  $n \times p$  matrix of specific factor scores, and  $\mathbf{E}$  is an  $n \times p$  error matrix, with  $\Lambda$  and  $\Theta$  being the same as those in (3). The rows of  $\mathbf{X}$ ,  $\mathbf{F}$ ,  $\mathbf{S}$ , and  $\mathbf{E}$  in (4) correspond to the transposes of  $\mathbf{x}$ ,  $\mathbf{f}$ ,  $\mathbf{s}$ , and  $\mathbf{e}$  in (3), respectively. However,  $\mathbf{F}$  and  $\mathbf{S}$  in (4) are treated as nonrandom parameter matrices, while  $\mathbf{f}$  and  $\mathbf{s}$  in (3) are random variable vectors. To the best of our knowledge, MDFA was first proposed by Professor Henk A. L. Kiers at the University of Groningen in 2001, as described in Sočan (2003, pp. 19–20). Then, the properties of the MDFA

solutions were studied by Adachi and Trendafilov (2018) and Stegeman (2016). The last author has also proposed a restrictive variant of MDFA. The factor vectors  $\mathbf{f}$  and  $\mathbf{u}$  in LVFA model (1) are random, as are  $\mathbf{f}$  and  $\mathbf{s}$  in (3), but  $\mathbf{F}$  and  $\mathbf{S}$  in MDFA model (4) are not random, as described above. This difference is not crucial in this paper, but rather we note that the unique factor part in (1) is decomposed into the specific factor and error parts in (3), whose nonrandom matrix version is (4).

In the above MDFA papers, the elements in the matrix corresponding to  $\mathbf{S}$  have been referred to as unique factor scores, but they must be called specific factor scores according to the terminology in CompFA model (4). Such a confusing reference to  $\mathbf{S}$  in the MDFA papers is due to the fact that the CompFA model was not related to MDFA there. This fact provides a major motivation for this paper: we study the relationships of the CompFA model to MDFA and its restrictive variant. This study includes reformulating those MDFA procedures to elucidate whether their parameter estimates are matched to the CompFA model assumptions for the parameters, with those assumptions to be introduced in Sect. 1. Here, we must inform beforehand that the reformulation (in Sect. 4) would allow Stegeman's (2016) restrictive variant of MDFA to be rephrased as *completely decomposed FA (CDFA)*. We thus call the variant CDFA hereafter. The relationships of LVFA to the CompFA model are also studied in this paper, as model (1) for LVFA is linked to the CompFA model through (2). The goal of our studies is to theoretically and empirically show how LVFA, MDFA, and CDFA behave for CompFA data, i.e., how the parameters in the CompFA model are recovered in the FA procedures, where the CompFA data refer to the data underlaid by the CompFA model. The goal would include showing that all parameters in that model can be recovered fairly well in CDFA but cannot be recovered as well in LVFA and MDFA.

The remaining sections in this paper are organized as follows: In Sect. 1, we specify the CompFA model by introducing its assumptions. Then, LVFA, MDFA, and CDFA are treated in Sects. 2, 3, and 4, respectively; we theoretically discuss how each procedure is related to the CompFA model and behaves for the CompFA data. The discussions for the behaviors are numerically assessed in Sect. 5 and illustrated in Sect. 6. Throughout this paper, we suppose  $n > p > m$ .

A more detailed prospect for the remaining sections will be presented in the final part of the next section, as that presentation is possible only after the CompFA model is fully specified. This specification also includes our subdivision of the CompFA model along whether the errors are assumed to be uncorrelated among variables or not. The subdivision is made, because the behaviors of the FA procedures differ between the CompFA data with uncorrelated errors and those with correlated ones, as discussed in Sects. 2, 3, 4 and shown numerically in Sect. 5.

## 1. Comprehensive Factor Analysis Model

In Sects. 1.1 and 1.2, we review the standard assumptions for CompFA model (3) and its nonrandom matrix version (4), respectively. Those standard assumptions are not involved with the inter-variable correlations of errors. We discuss in Sect. 1.3 that the consideration of the error correlations allows the CompFA model to be subdivided. The prospects for the following sections are given in the final subsection.

### 1.1. Random Version of Standard Assumptions

The CompFA model is expressed as (3), i.e.,  $\mathbf{x} = \mathbf{A}\mathbf{f} + \mathbf{O}\mathbf{s} + \mathbf{e}$ , for the random observation vector  $\mathbf{x}$ , whose expectation  $E[\mathbf{x}]$  is the  $p \times 1$  zero vector  $\mathbf{0}_p$ . We review the assumptions for the expectations and covariances of  $\mathbf{f}$ ,  $\mathbf{s}$ , and  $\mathbf{e}$  in the classic literature (Harman, 1976; Reymnt & Jöreskog, 1993; Thurstone, 1947).

In line with  $E[\mathbf{x}] = \mathbf{0}_p$ , the expectations of the common factor, specific factor, and error vectors are supposed as

$$E[\mathbf{f}] = \mathbf{0}_m, E[\mathbf{s}] = \mathbf{0}_p, E[\mathbf{e}] = \mathbf{0}_p. \quad (5)$$

The covariance matrices for the factor score vectors are assumed to satisfy

$$C[\mathbf{f}, \mathbf{f}] = \mathbf{I}_m, C[\mathbf{s}, \mathbf{s}] = \mathbf{I}_p, C[\mathbf{f}, \mathbf{s}] = {}_m\mathbf{O}_p. \quad (6)$$

Here,  $C[\mathbf{f}, \mathbf{s}] = E[(\mathbf{f} - E[\mathbf{f}])(\mathbf{s} - E[\mathbf{s}])']$  denotes the  $m \times p$  covariance matrix between  $\mathbf{f}$  and  $\mathbf{s}$ ,  $\mathbf{I}_m$  is the  $m \times m$  identity matrix, and  ${}_m\mathbf{O}_p$  expresses the  $m \times p$  zero matrix. The factor score vectors are assumed to be uncorrelated to the error vector:

$$C[\mathbf{f}, \mathbf{e}] = {}_m\mathbf{O}_p, C[\mathbf{s}, \mathbf{e}] = {}_p\mathbf{O}_p. \quad (7)$$

The standard constraints for model (3) consist of (5)–(7).

From (6), the covariance matrix of  $\Lambda\mathbf{f}$  and that of  $\Theta\mathbf{s}$  are found to be  $C[\Lambda\mathbf{f}, \Lambda\mathbf{f}] = \Lambda C[\mathbf{f}, \mathbf{f}] \Lambda' = \Lambda \Lambda'$  and  $C[\Theta\mathbf{s}, \Theta\mathbf{s}] = \Theta C[\mathbf{s}, \mathbf{s}] \Theta' = \Theta^2$ , respectively. Here, the diagonal elements of  $\Theta^2$  are called specific variances, as they stand for the variances of the specific factor part  $\Theta\mathbf{s}$ . Further, (6) and (7) lead to  $C[\Lambda\mathbf{f}, \Theta\mathbf{s}] = C[\Lambda\mathbf{f}, \mathbf{e}] = {}_m\mathbf{O}_p$  and  $C[\Theta\mathbf{s}, \mathbf{e}] = {}_p\mathbf{O}_p$ . Using these results, the inter-variable covariance matrix  $C[\mathbf{x}, \mathbf{x}]$  for (3) is found to be expressed as

$$C[\mathbf{x}, \mathbf{x}] = C[\Lambda\mathbf{f} + \Theta\mathbf{s} + \mathbf{e}, \Lambda\mathbf{f} + \Theta\mathbf{s} + \mathbf{e}] = \Lambda \Lambda' + \Theta^2 + C[\mathbf{e}, \mathbf{e}]. \quad (8)$$

### 1.2. Nonrandom Matrix Version of Standard Assumptions

The nonrandom matrix version of the CompFA model can be expressed as (4), i.e.,  $\mathbf{X} = \mathbf{F}\Lambda' + \mathbf{S}\Theta + \mathbf{E}$  for the  $n \times p$  data matrix  $\mathbf{X}$ . Here,  $\mathbf{X}$  is column-centered with  $\mathbf{1}'_n \mathbf{X} = \mathbf{0}'_p$  and supposed to have full column rank with  $\text{rank}(\mathbf{X}) = p$ , with  $\mathbf{1}_n$  and  $\text{rank}(\mathbf{X})$  denoting the  $n \times 1$  vector of ones and the rank of  $\mathbf{X}$ , respectively. We summarize the assumptions for (4), i.e., the matrix versions of the assumptions in Sect. 1.1.

The versions of (5) and (6) are expressed as

$$\mathbf{1}'_n \mathbf{F} = \mathbf{0}'_m, \mathbf{1}'_n \mathbf{S} = \mathbf{0}'_p, \mathbf{1}'_n \mathbf{E} = \mathbf{0}'_p, \quad (9)$$

$$\frac{1}{n} \mathbf{F}' \mathbf{F} = \mathbf{I}_m, \frac{1}{n} \mathbf{S}' \mathbf{S} = \mathbf{I}_p, \frac{1}{n} \mathbf{F}' \mathbf{S} = {}_m\mathbf{O}_p \text{ or } \mathbf{F}' \mathbf{S} = {}_m\mathbf{O}_p, \quad (10)$$

respectively. The two equations in (7) can be changed into the matrix forms  $n^{-1} \mathbf{F}' \mathbf{E} = {}_m\mathbf{O}_p$  and  $n^{-1} \mathbf{S}' \mathbf{E} = {}_p\mathbf{O}_p$ , which are equivalent to

$$\mathbf{F}' \mathbf{E} = {}_m\mathbf{O}_p, \quad (11)$$

$$\mathbf{S}' \mathbf{E} = {}_p\mathbf{O}_p, \quad (12)$$

respectively. The standard assumptions for model (4) consist of (9)–(12).

The  $p \times p$  inter-variable matrices for  $\mathbf{X}$  and  $\mathbf{E}$  can be expressed as  $\mathbf{C}_{\mathbf{X}\mathbf{X}} = n^{-1}\mathbf{X}'\mathbf{X}$  and  $\mathbf{C}_{\mathbf{E}\mathbf{E}} = n^{-1}\mathbf{E}'\mathbf{E}$ , because of  $\mathbf{1}_n'\mathbf{X} = \mathbf{0}_p'$  and (9). Then, (9)–(12) lead to the nonrandom matrix version of (8):

$$\mathbf{C}_{\mathbf{X}\mathbf{X}} = \frac{1}{n}(\mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Theta} + \mathbf{E})'(\mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Theta} + \mathbf{E}) = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Theta}^2 + \mathbf{C}_{\mathbf{E}\mathbf{E}}, \quad (13)$$

where  $\mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Theta} + \mathbf{E}$  is column-centered because of (9).

### 1.3. Uncorrelated Error and Correlated Error Assumptions

In the classic literature (Harman, 1976; Reymont & Jöreskog 1993; Thurstone, 1947), the elements of  $\mathbf{e}$  in (3) are particularly referred to as measurement errors. If such errors are considered to be uncorrelated among variables, we can add the assumption that the off-diagonal elements of  $C[\mathbf{e}, \mathbf{e}]$  are zeros, i.e.,

$$C[\mathbf{e}, \mathbf{e}] = D[\mathbf{e}, \mathbf{e}] \quad (14)$$

to those in Sect. 1.1, with  $D[\mathbf{e}, \mathbf{e}] = \text{diag}(C[\mathbf{e}, \mathbf{e}])$ . Here,  $\text{diag}(\mathbf{N})$  denotes the diagonal matrix whose diagonal elements are those of a square matrix  $\mathbf{N}$ . In the classic literature, (14) is not explicitly presented, but might be implicitly supposed for the following reasons: In that literature, only latent variable FA (LVFA) is described for estimating parameters, and the constraints considered in LVFA follow from adding (14) to the standard ones in Sect. 1.1, as shown in Sect. 2.2. We can also add the nonrandom matrix version of (14), i.e.,

$$\mathbf{C}_{\mathbf{E}\mathbf{E}} = \mathbf{D}_{\mathbf{E}\mathbf{E}} \quad (15)$$

to those in Sect. 1.2, with  $\mathbf{D}_{\mathbf{E}\mathbf{E}} = \text{diag}(\mathbf{C}_{\mathbf{E}\mathbf{E}})$ .

We can also consider a version of the CompFA model, in which (14) and (15) are not assumed, that is, the errors are allowed to be correlated among variables. In this correlated error version,  $C[\mathbf{e}, \mathbf{e}]$  in (8) and  $\mathbf{C}_{\mathbf{E}\mathbf{E}}$  in (13) are merely supposed to be unconstrained covariance matrices that are nonnegative-definite.

We should note the difference between uncorrelated error constraints (14) and (15), the latter being stronger, as explained next. The rows of the error matrix  $\mathbf{E}$  in  $\mathbf{C}_{\mathbf{E}\mathbf{E}} = n^{-1}\mathbf{E}'\mathbf{E}$  can be considered the realizations of the transpose of  $\mathbf{e}$  in (14). This consideration leads to  $C[\mathbf{e}, \mathbf{e}]$  in (14) equaling  $E[\mathbf{C}_{\mathbf{E}\mathbf{E}}]$ , as detailed in Appendix 1. Thus, we can rewrite (14) as the expectation of  $\mathbf{C}_{\mathbf{E}\mathbf{E}}$  being diagonal: (14) requires the diagonality of  $E[\mathbf{C}_{\mathbf{E}\mathbf{E}}]$ , but not that of  $\mathbf{C}_{\mathbf{E}\mathbf{E}}$  itself. This diagonality is required by (15), in contrast. To emphasize the strength of (15), we call this the *strong* uncorrelated error condition. We can consider that data meeting (15) is hardly encountered, i.e., is unusual. However, in the later sections, such data would be noted: The FA procedures are shown to perfectly fit the data meeting strong condition (15), which is a motivation to study the related properties of the FA procedures. The next theorem gives the foundation for the perfect fit to be shown later.

**Theorem 1.** *If the error matrix  $\mathbf{E}$  in CompFA model (4) satisfies (15), (4) can be rewritten as the error-free model:*

$$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \tilde{\mathbf{S}}\tilde{\mathbf{\Theta}} = \mathbf{F}\mathbf{\Lambda}' + \tilde{\mathbf{S}}(\mathbf{\Theta}^2 + \mathbf{D}_{\mathbf{E}\mathbf{E}})^{1/2}. \quad (16)$$

Here,  $\tilde{\mathbf{\Theta}}^2 = \mathbf{\Theta}^2 + \mathbf{D}_{\mathbf{E}\mathbf{E}}$  and  $\tilde{\mathbf{S}} = (\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1}$  can be regarded as the specific variance and factor score matrices, respectively.

TABLE 1.  
Relationships of the FA procedures to the CompFA model.

Proc.	Model	Note	Specific Factor and Errors	Estimate $\approx$ True
LVFA	$\mathbf{x} = \mathbf{F}\mathbf{\Lambda} + \mathbf{\Psi}\mathbf{u}$	$\mathbf{\Psi}\mathbf{u} = \mathbf{\Theta}\mathbf{s} + \mathbf{e}$	$\mathbf{\Theta}\mathbf{s}$ & $\mathbf{e}$ unseparated	$\mathbf{\Psi}^2 \approx \mathbf{\Theta}^2 + \mathbf{D}_{EE}$ $\mathbf{\Lambda} \approx \mathbf{\underline{\Lambda}}$
MDFA	$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Theta} + \mathbf{E}$	↓	$\text{diag}(\mathbf{S}'\mathbf{E}) = {}_p\mathbf{O}_p$	$\mathbf{\Theta}^2 \approx \mathbf{\underline{\Theta}^2} + \mathbf{D}_{EE}$ $\mathbf{\Lambda} \approx \mathbf{\underline{\Lambda}}$
CDFA	$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Theta} + \mathbf{E}$	Nonrandom version	$\mathbf{S}'\mathbf{E} = {}_p\mathbf{O}_p$	$\mathbf{\Theta}^2 \approx \mathbf{\underline{\Theta}^2}$ $\mathbf{\Lambda} \approx \mathbf{\underline{\Lambda}}$

*Proof.* We can rewrite (4) as (16):  $\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Theta} + \mathbf{E} = \mathbf{F}\mathbf{\Lambda}' + (\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1}\tilde{\mathbf{\Theta}}$ . The matrix  $\tilde{\mathbf{\Theta}}^2 = \mathbf{\Theta}^2 + \mathbf{D}_{EE}$  ( $p \times p$ ) is diagonal and nonnegative, thus regarded as the specific variance matrix. We can show why  $\tilde{\mathbf{S}} = (\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1}$  ( $n \times p$ ) is the specific factor score matrix, as follows: (9)–(12) and (15) lead to  $\mathbf{1}_n'\tilde{\mathbf{S}} = \mathbf{1}_n'(\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1} = \mathbf{0}_p'$ ,

$$\frac{1}{n}\tilde{\mathbf{S}}'\tilde{\mathbf{S}} = \frac{1}{n}\tilde{\mathbf{\Theta}}^{-1}(\mathbf{S}\mathbf{\Theta} + \mathbf{E})'(\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1} = (\mathbf{\Theta}^2 + \mathbf{D}_{EE})^{-1/2}(\mathbf{\Theta}^2 + \mathbf{D}_{EE})(\mathbf{\Theta}^2 + \mathbf{D}_{EE})^{-1/2} = \mathbf{I}_p,$$

and  $\mathbf{F}'\tilde{\mathbf{S}} = \mathbf{F}'(\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1} = {}_m\mathbf{O}_p$ . Thus,  $\tilde{\mathbf{S}} = (\mathbf{S}\mathbf{\Theta} + \mathbf{E})\tilde{\mathbf{\Theta}}^{-1}$  can be substituted into  $\mathbf{S}$  in (9)–(11). Further, (16) can be rewritten as  $\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \tilde{\mathbf{S}}\tilde{\mathbf{\Theta}} + \tilde{\mathbf{E}}$ , with the error matrix  $\tilde{\mathbf{E}}$  being  ${}_n\mathbf{O}_p$ , which implies that those  $\tilde{\mathbf{E}}$  and  $\tilde{\mathbf{S}}$  can be substituted into  $\mathbf{E}$  and  $\mathbf{S}$  in (12), respectively.  $\square$

#### 1.4. Prospects for Relating the CompFA Model to LVFA, MDFA, and CDFA

As the assumptions in the CompFA model have been specified, they can now be used for providing the prospects for the following sections, where the relationships of the CompFA model to LVFA, MDFA, and CDFA will be studied.

Among the relationships, those independent of the uncorrelated and correlated error assumptions (in Sect. 1.3) can be summarized as in Table 1. Its left-hand "Model" and "Note" columns merely present the facts described before Sect. 1: (2) links LVFA model (1) to CompFA model (3), and this matrix version is model (4) underlying MDFA and CDFA.

The right-hand columns in Table 1 present the key points in the relationships to be found. The lower cells in the "Specific Factor & Errors" column show the facts to be found in Sects. 3.2 and 4.2: The MDFA solution does not meet (12) in the CompFA assumptions, but only satisfies its diagonal part  $\text{diag}(\mathbf{S}'\mathbf{E}) = {}_p\mathbf{O}_p$ , but (12) is satisfied by the CDFA solution. On the other hand, the specific factor and error parts ( $\mathbf{\Theta}\mathbf{s}$  and  $\mathbf{e}$ ) are unseparated in LVFA.

The column furthest right in Table 1 shows how  $\mathbf{\Psi}^2$ ,  $\mathbf{\Theta}^2$ , and  $\mathbf{\Lambda}$  are usually estimated in the FA procedures for the CompFA data matrix

$$\mathbf{X} = \mathbf{F}\mathbf{\underline{\Lambda}}' + \mathbf{S}\mathbf{\underline{\Theta}} + \mathbf{E} \quad \text{with} \quad \mathbf{D}_{EE} = \text{diag}(\mathbf{C}_{EE}) = \frac{1}{n}\text{diag}(\mathbf{E}'\mathbf{E}), \quad (17)$$

i.e., the observations underlaid by (4) with  $\mathbf{F}$ ,  $\mathbf{\Lambda}$ ,  $\mathbf{S}$ ,  $\mathbf{\Theta}$ , and  $\mathbf{E}$  set to particular matrices  $\mathbf{F}$ ,  $\mathbf{\underline{\Lambda}}$ ,  $\mathbf{S}$ ,  $\mathbf{\underline{\Theta}}$ , and  $\mathbf{E}$ , respectively. Here, the latter matrices have been underlined for the sake of indicating that particular values are substituted into the elements of those matrices. The final (third) subsections in Sect. 2–4 are particularly concerned with how the true  $\mathbf{\underline{\Theta}}$  and  $\mathbf{D}_{EE}$  in (17) are related to LVFA, MDFA, and CDFA estimates, respectively, as outlined in the following two paragraphs.

In Sects. 2.3 and 3.3, we will discuss that the LVFA estimate of  $\mathbf{\Psi}^2$  approximates  $\mathbf{\underline{\Theta}^2} + \mathbf{D}_{EE}$  and the MDFA one of  $\mathbf{\Theta}^2$  approximates  $\mathbf{\underline{\Theta}^2} + \mathbf{D}_{EE}$ , respectively. The latter MDFA property of  $\mathbf{\Theta}^2 \approx \mathbf{\underline{\Theta}^2} + \mathbf{D}_{EE}$  is restated as  $\mathbf{\Theta}^2$  being contaminated by  $\mathbf{D}_{EE}$ . In Sect. 4.3, CDFA will be shown to

provide the estimate  $\Theta^2 \approx \underline{\Theta}^2$  differently from MDFA, with a discussion of how this difference follows from the CDFA solution satisfying (12), which is not met by the MDFA solution.

In Sects. 2.3, 3.3, and 4.3, we will also discuss that the uncorrelated error assumption (in Sect. 1.3) leads to the facts that are not covered in Table 1. The facts are summarized as follows. Two of the formulas with " $\approx$ " in Table 1 are replaced by equalities only in strong uncorrelated error condition (15):  $\Psi^2 = \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$  in LVFA and  $\Theta^2 = \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$  in MDFA and CDFA, with all procedures fitting data perfectly. However, apart from that strong condition, the CDFA estimate of  $\Theta^2$  can approximate  $\underline{\Theta}^2$ , as described above.

The good recovery of loadings with  $\Lambda \approx \underline{\Lambda}$  in all procedures will be shown numerically in Sect. 5.

## 2. Latent Variable Factor Analysis

In Sect. 2.1, we review the formulation of latent variable FA (LVFA) with the assumptions added to (1). Then, in Sect. 2.2, we show how those LVFA assumptions can follow from the CompFA ones. In Sect. 2.3, we discuss how the estimate of  $\Psi^2$  approximates  $\underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$ , as shown in Table 1.

### 2.1. Formulation

LVFA is modeled as (1), i.e.,  $\mathbf{x} = \Lambda\mathbf{f} + \tilde{\mathbf{e}} = \Lambda\mathbf{f} + \Psi\mathbf{u}$ , where the expectations and covariances for  $\mathbf{f}$  and  $\tilde{\mathbf{e}} = \Psi\mathbf{u}$  are assumed to satisfy

$$E[\mathbf{f}] = \mathbf{0}_m, \quad C[\mathbf{f}, \mathbf{f}] = \mathbf{I}_m, \quad (18)$$

$$E[\Psi\mathbf{u}] = \mathbf{0}_p, \quad C[\mathbf{f}, \Psi\mathbf{u}] = m\mathbf{O}_p, \quad (19)$$

$$C[\Psi\mathbf{u}, \Psi\mathbf{u}] = \Psi^2. \quad (20)$$

(e.g., Bartholomew, Knott, & Moustaki, 2011; Yanai & Ichikawa, 2007). Here, the diagonal elements of  $\Psi^2$  are called unique variances, as they stand for the variances of the unique factor part  $\Psi\mathbf{u}$ .

LVFA assumptions (18)–(20) imply that inter-variable covariance matrix  $C[\mathbf{x}, \mathbf{x}]$  for (1) is expressed as

$$C[\mathbf{x}, \mathbf{x}] = C[\Lambda\mathbf{f} + \Psi\mathbf{u}, \Lambda\mathbf{f} + \Psi\mathbf{u}] = \Lambda\Lambda' + \Psi^2. \quad (21)$$

Thus,  $\Lambda$  and  $\Psi^2$  can be estimated so that model-based (21) approximates its data-based counterpart  $\mathbf{C}_{XX} = n^{-1}\mathbf{X}'\mathbf{X}$ . This estimation can be attained by minimizing

$$f^{\text{LS}}(\Lambda, \Psi) = \|\mathbf{C}_{XX} - C[\mathbf{x}, \mathbf{x}]\|^2 = \|\mathbf{C}_{XX} - (\Lambda\Lambda' + \Psi^2)\|^2 \quad (22)$$

over  $\Lambda$  and  $\Psi^2$  (Harman & Jones, 1966), with  $\|\mathbf{M}\|^2 = \text{tr}\mathbf{M}'\mathbf{M}$  denoting the squared Frobenius norm of a matrix  $\mathbf{M}$ . Another estimation procedure is to minimize the function

$$\begin{aligned} f^{\text{NL}}(\Lambda, \Psi) &= \text{tr}\mathbf{C}_{XX}C[\mathbf{x}, \mathbf{x}]^{-1} - \log|\mathbf{C}_{XX}C[\mathbf{x}, \mathbf{x}]^{-1}| - p \\ &= \text{tr}\mathbf{C}_{XX}(\Lambda\Lambda' + \Psi^2)^{-1} - \log|\mathbf{C}_{XX}(\Lambda\Lambda' + \Psi^2)^{-1}| - p, \end{aligned} \quad (23)$$

which is the negative of the log likelihood following from (21) and the additional normality assumptions for  $\mathbf{f}$  and  $\mathbf{u}$  (e.g., Bartholomew, Knott, & Moustaki, 2011; Yanai & Ichikawa, 2007).

## 2.2. Relationship to the CompFA Model

LVFA model (1) and assumptions (18)–(20) do not include the specific factor and error parts, which differs from CompFA model (3) and its assumptions. However, (1) is linked to (3) through (2). Further, it is shown next how LVFA constraints (18)–(20) follow from the CompFA counterparts:

**Theorem 2.** *Under (1)–(3), LVFA assumptions (18)–(20) follow from CompFA ones (5)–(7) and uncorrelated error assumption (14).*

*Proof.* Obviously, (18) follows, since its two equations appear in (5) and (6). Next,  $E[\Psi\mathbf{u}] = \mathbf{0}_p$  and  $C[\mathbf{f}, \Psi\mathbf{u}] = {}_m\mathbf{O}_p$  in (19) follow from (5)–(7) under (2), since (2) and (5) imply  $E[\Psi\mathbf{u}] = \Theta E[\mathbf{s}] + E[\mathbf{e}] = \mathbf{0}_p$ , while (2), (6), and (7) lead to  $C[\mathbf{f}, \Psi\mathbf{u}] = C[\mathbf{f}, \Theta\mathbf{s} + \mathbf{e}] = C[\mathbf{f}, \mathbf{s}]\Theta' + C[\mathbf{f}, \mathbf{e}] = {}_m\mathbf{O}_p$ . The remaining task is to show how (20) follows. Its left side is rewritten, using (2), (6), and (7), as  $C[\Psi\mathbf{u}, \Psi\mathbf{u}] = C[\Theta\mathbf{s} + \mathbf{e}, \Theta\mathbf{s} + \mathbf{e}] = \Theta^2 + C[\mathbf{e}, \mathbf{e}]$ . For this to equal the right side of (20), i.e., nonnegative diagonal  $\Psi^2$ ,  $C[\mathbf{e}, \mathbf{e}]$  must equal  $\Psi^2 - \Theta^2$  which is also nonnegative diagonal. This holds true for (14), i.e.,  $\Psi^2 - \Theta^2 = D[\mathbf{e}, \mathbf{e}]$ .  $\square$

This theorem shows that the LVFA assumptions follow from adding uncorrelated error assumption (14) to the standard CompFA ones in Sect. 1.1. However, LVFA can be performed for CompFA data (17), independently of whether its errors satisfy (14) or not. In the next subsection, how LVFA behaves for data (17) is discussed.

## 2.3. Behaviors for CompFA Data

CompFA data matrix (17) leads to the covariance matrix

$$\mathbf{C}_{XX} = \underline{\Lambda} \underline{\Lambda}' + \underline{\Theta}^2 + \underline{\mathbf{C}}_{EE}, \quad (24)$$

i.e., (13) with  $\underline{\Lambda}$ ,  $\underline{\Theta}^2$ , and  $\underline{\mathbf{C}}_{EE}$  set to particular matrices  $\underline{\Lambda}$ ,  $\underline{\Theta}^2$ , and  $\underline{\mathbf{C}}_{EE} = n^{-1}\mathbf{E}'\mathbf{E}$ , respectively. Let us consider the LVFA solution for (24).

Loss functions (22) and (23) are known to be minimized for

$$\Psi^2 = \text{diag}(\mathbf{C}_{XX} - \underline{\Lambda} \underline{\Lambda}') \quad (25)$$

for a given  $\underline{\Lambda}$  (e.g., Mulaik, 2010, (8.47), (8.80)). Using (24) in (25), this can be rewritten as  $\Psi^2 = \text{diag}(\underline{\Lambda} \underline{\Lambda}' + \underline{\Theta}^2 + \underline{\mathbf{C}}_{EE} - \underline{\Lambda} \underline{\Lambda}') = \text{diag}(\underline{\Lambda} \underline{\Lambda}' - \underline{\Lambda} \underline{\Lambda}') + \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$ . This implies

$$\Psi^2 \approx \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE} \quad (26)$$

for  $\underline{\Lambda} \underline{\Lambda}' \approx \underline{\Lambda} \underline{\Lambda}'$ ; if estimated  $\underline{\Lambda} \underline{\Lambda}'$  approximates the true counterpart  $\underline{\Lambda} \underline{\Lambda}'$ , each of the diagonal elements in  $\Psi^2$ , i.e., a unique variance, can be the estimate of the sum of the corresponding true specific and error variances, though these two cannot be estimated separately.

The next corollary, which follows from Theorem 1, shows that formula (26) with " $\approx$ " is replaced by the equality  $\Psi^2 = \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$ , if (24) satisfies strong uncorrelated error condition (15):

**Corollary 1.** *For (15), (24) is restricted to  $\mathbf{C}_{XX} = \underline{\Lambda} \underline{\Lambda}' + \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$ . For this matrix, LVFA loss functions (22) and (23) can attain their lower limit zero with the solution of  $\Psi^2$  given by  $\Psi^2 = \underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$  and that of  $\underline{\Lambda}$  satisfying  $\underline{\Lambda} \underline{\Lambda}' = \underline{\Lambda} \underline{\Lambda}'$ .*



*Proof.* The matrix difference  $\mathbf{C}_{XX} - (\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2)$  in (22) and the product  $\mathbf{C}_{XX}(\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2)^{-1}$  in (23) are rewritten, using  $\mathbf{C}_{XX} = \underline{\mathbf{A}}\underline{\mathbf{A}}' + \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$ , as  $\underline{\mathbf{A}}\underline{\mathbf{A}}' + \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE} - (\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2)$  and  $(\underline{\mathbf{A}}\underline{\mathbf{A}}' + \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE})(\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2)^{-1}$ , respectively. The former difference can be  ${}_p\mathbf{O}_p$ , and the latter product can be  $\mathbf{I}_p$ , which allows (22) and (23) to attain zeros, for  $\mathbf{\Lambda}\mathbf{\Lambda}' = \underline{\mathbf{A}}\underline{\mathbf{A}}'$  and  $\mathbf{\Psi}^2 = \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$ .  $\square$

The result in this corollary is referred to in Sect. 3.3.

### 3. Matrix Decomposition Factor Analysis

In Sect. 3.1, we review the original formulation of matrix decomposition FA (MDFA). Then, in Sects. 3.2 and 3.3, we discuss two properties of the MDFA solution shown in Table 1. One property is the solution's satisfying  $\text{diag}(\mathbf{S}'\mathbf{E}) = {}_p\mathbf{O}_p$  but not (12), which implies that MDFA can be reformulated with constraints that are less restrictive than the CompFA assumptions, as discussed in Sect. 3.2. The other property  $\mathbf{\Theta}^2 \approx \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$  is suggested by a theorem to be presented in Sect. 3.3.

#### 3.1. Original Formulation

For the data matrix  $\mathbf{X}$  with  $\mathbf{1}_n'\mathbf{X} = \mathbf{0}'_p$ , MDFA is formulated as minimizing the least squares function for (4), i.e.,

$$f(\mathbf{F}, \mathbf{\Lambda}, \mathbf{S}, \mathbf{\Theta}) = \frac{1}{n} \|\mathbf{E}\|^2 = \frac{1}{n} \|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{S}\mathbf{\Theta}\|^2, \quad (27)$$

over  $\mathbf{F}$ ,  $\mathbf{\Lambda}$ ,  $\mathbf{S}$ , and  $\mathbf{\Theta}$  subject to constraints (9) and (10). In the MDFA literature (e.g., Adachi & Trendafilov, 2018; Sočan, 2003; Stegeman, 2016),  $\mathbf{1}_n'\mathbf{E} = \mathbf{0}'_p$  in (9) has not been described, but this is trivial, since  $\mathbf{1}_n'\mathbf{E} = \mathbf{0}'_p$  follows from the other equations in (9), (4), and  $\mathbf{1}_n'\mathbf{X} = \mathbf{0}'_p$ .

One difference in (27) from the LVFA loss functions, besides the former being based on (4), is that (22) and (23) in LVFA do not include the factor scores, while (27) includes those scores as  $\mathbf{F}$  and  $\mathbf{S}$  to be estimated. However, infinite solutions of  $\mathbf{F}$  and  $\mathbf{S}$  exist that minimize (27); the optimal  $\mathbf{F}$  and  $\mathbf{S}$  are not unique (Adachi & Trendafilov, 2018, Sect. 4). Thus, only the solutions of  $\mathbf{\Lambda}$  and  $\mathbf{\Theta}$  are interpreted in MDFA, as are those of  $\mathbf{\Lambda}$  and  $\mathbf{\Psi}$  in LVFA.

Though (27) is defined using data matrix  $\mathbf{X}$ , the MDFA solution of  $\mathbf{\Lambda}$  and  $\mathbf{\Theta}$  can be obtained only if the covariance matrix  $\mathbf{C}_{XX} = n^{-1}\mathbf{X}'\mathbf{X}$  is available, i.e., even if the original  $\mathbf{X}$  is unavailable (Adachi, 2012, 2020). This can be captured in the fact that (27) can be expanded as  $n^{-1}\text{tr}(\mathbf{X}'\mathbf{X} + \mathbf{\Lambda}'\mathbf{F}'\mathbf{F}\mathbf{\Lambda} + \mathbf{\Theta}'\mathbf{S}'\mathbf{S}\mathbf{\Theta}) - 2n^{-1}\text{tr}(\mathbf{X}'\mathbf{F}\mathbf{\Lambda}' + \mathbf{X}'\mathbf{S}\mathbf{\Theta}) - \mathbf{\Lambda}'\mathbf{F}'\mathbf{S}\mathbf{\Theta}$  and simplified using (10) as  $\text{tr}\mathbf{C}_{XX} + \text{tr}\mathbf{\Lambda}\mathbf{\Lambda}' + \text{tr}\mathbf{\Theta}^2 - 2\text{tr}\mathbf{C}_{XF}\mathbf{\Lambda}' - 2\text{tr}\mathbf{C}_{XS}\mathbf{\Theta}$ , in which  $\mathbf{X}$  does not appear. Here,  $\mathbf{C}_{XF} = n^{-1}\mathbf{X}'\mathbf{F}$  and  $\mathbf{C}_{XS} = n^{-1}\mathbf{X}'\mathbf{S}$  contain the covariances of variables to factors and are uniquely determined for given  $\mathbf{\Lambda}$  and  $\mathbf{\Theta}$  (Adachi & Trendafilov, 2018, p. 411). On the other hand, the optimal  $\mathbf{\Lambda}$  and  $\mathbf{\Theta}$  can be obtained for given  $\mathbf{C}_{XF}$  and  $\mathbf{C}_{XS}$  (Adachi & Trendafilov, 2018, p. 410). Thus, the optimal updates of the block matrices  $[\mathbf{C}_{XF}, \mathbf{C}_{XS}]$  and  $[\mathbf{\Lambda}, \mathbf{\Theta}]$  are iterated alternately to provide the solution of  $[\mathbf{\Lambda}, \mathbf{\Theta}]$  in Adachi's (2012, 2020) MDFA algorithm.

#### 3.2. Reformulation from Properties of the Solution

Besides (9) and (10) imposed as constraints in MDFA, (11) and (12) are included in the standard CompFA assumptions (Sect. 1.2). Thus, we must note whether (11) and (12) are satisfied by the MDFA solution, i.e., the parameter estimates and  $\mathbf{E}$  resulting in the minimization of (27)

under (9) and (10). Adachi and Trendafilov (2018, Theorem 4.1) show that the solution satisfies (11) and the diagonal part of (12), i.e.,

$$\text{diag}(\mathbf{S}'\mathbf{E}) = {}_p\mathbf{O}_p, \quad (28)$$

but does not meet the off-diagonal part of (12) with  $\mathbf{S}'\mathbf{E} - \text{diag}(\mathbf{S}'\mathbf{E}) \neq {}_p\mathbf{O}_p$  in general. That is, the MDFA solution satisfies standard CompFA assumptions (9)–(11) but does not meet (12).

The MDFA solution satisfying (11) and (28) implies that these equations can be included in the constraints. That is, MDFA can be reformulated as minimizing (27) over  $\mathbf{F}$ ,  $\mathbf{\Lambda}$ ,  $\mathbf{S}$ , and  $\mathbf{\Theta}$  subject to constraints (9)–(11) and (28). Here, (28) being only the diagonal part of (12) implies that the MDFA formulation is less restrictive than the CompFA model, in that (12) is relaxed as (28). In contrast, the theorems presented in Sect. 4.2 show that the CDFA solutions satisfy CompFA constraints (9)–(12) completely.

### 3.3. Behaviors for CompFA Data

In this section, we consider how MDFA behaves for CompFA data matrix (17). The next corollary, which follows from Theorem 1, shows the MDFA solution in strong uncorrelated error condition (15).

**Corollary 2.** For data matrix (17) satisfying  $\underline{\mathbf{C}}_{EE} = \underline{\mathbf{D}}_{EE}$ , i.e., (15) given by setting  $\mathbf{E}$  to  $\underline{\mathbf{E}}$ , MDFA loss function (27) can attain its lower limit zero for the solution satisfying  $\mathbf{F}\mathbf{\Lambda}' = \underline{\mathbf{F}}\underline{\mathbf{\Lambda}}'$ ,  $\mathbf{\Theta} = \underline{\tilde{\Theta}}$ ,  $\mathbf{S} = \underline{\tilde{S}}$  with  $\underline{\tilde{\Theta}} = (\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE})^{1/2}$  and  $\underline{\tilde{S}} = (\underline{\mathbf{S}}\underline{\mathbf{\Theta}} + \underline{\mathbf{E}})\underline{\tilde{\Theta}}^{-1}$ .

*Proof.* Theorem 1 shows that (17) can be rewritten into the error-free form  $\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \tilde{\mathbf{S}}\tilde{\mathbf{\Theta}}$  for  $\underline{\mathbf{C}}_{EE} = \underline{\mathbf{D}}_{EE}$ . Substituting  $\mathbf{X} = \underline{\mathbf{F}}\underline{\mathbf{\Lambda}}' + \underline{\tilde{S}}\underline{\tilde{\Theta}}$  in (27), it is rewritten as  $n^{-1} \left\| \underline{\mathbf{F}}\underline{\mathbf{\Lambda}}' + \underline{\tilde{S}}\underline{\tilde{\Theta}} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{S}\mathbf{\Theta} \right\|^2$ . This can attain the lower limit zero for  $\mathbf{F}\mathbf{\Lambda}' = \underline{\mathbf{F}}\underline{\mathbf{\Lambda}}'$ ,  $\mathbf{S} = \underline{\tilde{S}}$ , and  $\mathbf{\Theta} = \underline{\tilde{\Theta}} = (\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE})^{1/2}$ .  $\square$

By comparing Corollaries 1 and 2, we can find that both LVFA and MDFA perfectly fit the CompFA data with strong condition (15), then  $\mathbf{\Psi}^2 = \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$  in LVFA, but  $\mathbf{\Theta}^2 = \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$  in MDFA; its estimates of the specific variances are not their true values, and they are contaminated by the error variances in  $\underline{\mathbf{D}}_{EE}$ . This shows an undesirable property of MDFA.

Though the data with (15) are unusual, as mentioned in Sect. 1.3, whether the above contamination can occur for usual CompFA data (17) that are not restricted by (15) is to be considered. For this consideration, we reparameterize the error matrix in (17) as

$$\mathbf{E} = \mathbf{G}\mathbf{\Gamma}' \quad \text{with} \quad \frac{1}{n}\mathbf{G}'\mathbf{G} = \mathbf{I}_p, \quad (29)$$

and  $\mathbf{\Gamma}$  being  $p \times p$ . The next theorem suggests that the MDFA estimate of  $\mathbf{\Theta}^2$  can be contaminated by  $\underline{\mathbf{D}}_{EE}$ , i.e.,  $\mathbf{\Theta}^2$  can be close to  $\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$ :

**Theorem 3.** Consider the case where MDFA is performed for the data matrix specified as (17) with (29) and  $\text{rank}(\mathbf{E}) = \text{rank}(\mathbf{\Theta}) = p$ . Let two matrices be defined as

$$\bar{\mathbf{\Theta}} = (\underline{\mathbf{\Theta}}^2 + \mathbf{\Lambda}^2)^{1/2} \quad \text{and} \quad \bar{\mathbf{S}} = (\underline{\mathbf{S}}\underline{\mathbf{\Theta}} + \mathbf{G}\mathbf{\Lambda})\bar{\mathbf{\Theta}}^{-1}, \quad (30)$$

with  $\mathbf{\Lambda}$  a  $p \times p$  diagonal matrix. Then, the matrices  $\bar{\mathbf{\Theta}}$  and  $\bar{\mathbf{S}}$  in (30) can be substituted into  $\mathbf{\Theta}$  and  $\mathbf{S}$  in function (27), respectively, and  $\bar{\mathbf{S}}$  can be substituted into  $\mathbf{S}$  in (9) and (10). Moreover, if

$\mathbf{\Lambda} = \text{diag}(\mathbf{\Gamma})$  and  $\mathbf{F}$  satisfies  $\mathbf{F}'\bar{\mathbf{S}} = \mathbf{F}'\bar{\mathbf{E}} = {}_m\mathbf{O}_p$ , then  $\bar{\mathbf{S}}$  in (30) can be substituted into  $\mathbf{S}$  in (28), and we can use  $\mathbf{\Theta} = \bar{\mathbf{\Theta}}$ ,  $\mathbf{S} = \bar{\mathbf{S}}$ , and (17) in (27) to rewrite this function as

$$f(\mathbf{F}, \mathbf{\Lambda}, \bar{\mathbf{S}}, \bar{\mathbf{\Theta}}) = \frac{1}{n} \|\mathbf{F}\bar{\mathbf{\Lambda}}' - \mathbf{F}\bar{\mathbf{\Lambda}}'\|^2 + \text{tr}\bar{\mathbf{D}}_{\text{EE}} - \text{tr}\bar{\mathbf{\Lambda}}^2. \quad (31)$$

*Proof.* See Appendix 2.

Theorem 3 suggests that the MDFA estimate of  $\mathbf{\Theta}^2$  can be contaminated by  $\bar{\mathbf{D}}_{\text{EE}}$  with

$$\mathbf{\Theta}^2 \approx \underline{\mathbf{\Theta}}^2 + \bar{\mathbf{D}}_{\text{EE}}, \quad (32)$$

as explained next. The theorem allows us to consider that  $\bar{\mathbf{\Theta}}$  and  $\bar{\mathbf{S}}$  in (30) can be the MDFA estimates of  $\mathbf{\Theta}$  and  $\mathbf{S}$ , respectively. By substituting those  $\bar{\mathbf{\Theta}}$  and  $\bar{\mathbf{S}}$  into  $\mathbf{\Theta}$  and  $\mathbf{S}$  in the loss function (27) to be minimized, it can be rewritten as (31). This value can be small for  $\mathbf{F}\bar{\mathbf{\Lambda}}' \approx \underline{\mathbf{F}}\bar{\mathbf{\Lambda}}'$  and  $\bar{\mathbf{\Lambda}}^2 \approx \bar{\mathbf{D}}_{\text{EE}}$ . This use in (30) and  $\mathbf{\Theta} = \bar{\mathbf{\Theta}}$  lead to (32). It will be confirmed in Sect. 5 that (32) actually arises.

#### 4. Completely Decomposed Factor Analysis

In Sect. 4.1, we review Stegeman's (2016) restrictive variant of MDFA that has been rephrased as completely decomposed FA (CDFA) through the theorems present in Sect. 4.2. They also allow CDFA to be reformulated so that it is perfectly matched with the CompFA model. In Sect. 4.3, we argue how the CDFA estimate of  $\mathbf{\Theta}^2$  approximates  $\underline{\mathbf{\Theta}}^2$  as shown in Table 1.

##### 4.1. Stegeman's Factor Analysis Procedure

Stegeman's (2016) original formulation of CDFA is to add the constraint

$$\mathbf{S}'(\mathbf{X} - \mathbf{S}\mathbf{\Theta}) = {}_p\mathbf{O}_p \quad (33)$$

to MDFA formulated as in Sect. 3.1: In CDFA, (27) is minimized over  $\mathbf{F}$ ,  $\mathbf{\Lambda}$ ,  $\mathbf{S}$ , and  $\mathbf{\Theta}$  subject to (9), (10), and (33), for the data matrix  $\mathbf{X}$  with  $\mathbf{1}'_n\mathbf{X} = \mathbf{0}'_p$ .

As described in Stegeman (2016, p. 196), the CDFA solution can be obtained through the following three sequential steps: First, the optimal  $\mathbf{\Theta}$  subject to (33) can be obtained by performing ten Berge and Kiers' (2001) minimum rank factor analysis (MRFA) for  $\mathbf{C}_{\text{XX}} = n^{-1}\mathbf{X}'\mathbf{X}$ . Then, the resulting  $\mathbf{\Theta}$  provides the solution of  $\mathbf{S}$ . Finally, loss function (27), whose  $\mathbf{\Theta}$  and  $\mathbf{S}$  are fixed to the ones resulting so far, is minimized over  $\mathbf{F}$  and  $\mathbf{\Lambda}$  subject to (9) and (10). This minimization is attained for

$$\mathbf{F} = \sqrt{n}\mathbf{V}_m = \sqrt{n}(\mathbf{X} - \mathbf{S}\mathbf{\Theta})\mathbf{W}_m\mathbf{\Omega}_m^{-1} \text{ and } \mathbf{\Lambda} = \frac{1}{\sqrt{n}}\mathbf{W}_m\mathbf{\Omega}_m \quad (34)$$

through the singular value decomposition (SVD) of  $\mathbf{X} - \mathbf{S}\mathbf{\Theta}$  defined as

$$\mathbf{X} - \mathbf{S}\mathbf{\Theta} = \mathbf{V}\mathbf{\Omega}\mathbf{W}'. \quad (35)$$

Here,  $\mathbf{V}'\mathbf{V} = \mathbf{W}'\mathbf{W} = \mathbf{I}_q$ , and  $\mathbf{\Omega}$  is the  $q \times q$  diagonal matrix whose diagonal elements are arranged in decreasing order, with  $q = \text{rank}(\mathbf{X} - \mathbf{S}\mathbf{\Theta})$  and  $q \geq m$  supposed. The matrices  $\mathbf{V}_m$

$(n \times m)$  and  $\mathbf{W}_m$  ( $p \times m$ ) in (34) contain the first  $m$  columns of  $\mathbf{V}$  and  $\mathbf{W}$ , respectively, with  $\mathbf{\Omega}_m$  the upper-left  $m \times m$  diagonal block of  $\mathbf{\Omega}$ .

The optimal  $\mathbf{F}$  and  $\mathbf{\Lambda}$  in (34) are found to satisfy

$$\frac{1}{n} \mathbf{X}'\mathbf{F} = \mathbf{\Lambda}, \quad (36)$$

since we can use (10), (34), (35), and  $\mathbf{V}'\mathbf{V}_m = [\mathbf{I}_{m,m} \mathbf{O}_{q-m}]'$  (from  $\mathbf{V}'\mathbf{V} = \mathbf{I}_q$ ) to derive (36) as

$$\frac{1}{n} \mathbf{X}'\mathbf{F} = \frac{1}{n} \mathbf{X}'\mathbf{F} - \frac{1}{n} \mathbf{\Theta}'\mathbf{S}'\mathbf{F} = \frac{1}{n} (\mathbf{X} - \mathbf{S}\mathbf{\Theta})'\mathbf{F} = \frac{1}{n} \mathbf{W}\mathbf{\Omega}\mathbf{V}' (\sqrt{n}\mathbf{V}_m) = \frac{1}{\sqrt{n}} \mathbf{W}_m \mathbf{\Omega}_m = \mathbf{\Lambda}.$$

As in MDFA, the optimal  $\mathbf{F}$  and  $\mathbf{S}$  cannot be uniquely determined, but their estimation can be skipped to obtain the optimal  $\mathbf{\Lambda}$  and  $\mathbf{\Theta}$ , only if the covariance matrix  $\mathbf{C}_{\mathbf{X}\mathbf{X}}$  is available. This is shown using the fact that (10) and (33) lead to  $\mathbf{X}'\mathbf{S} = n\mathbf{\Theta}^2$ . This, (33), and (35) imply that the SVD of  $(\mathbf{X} - \mathbf{S}\mathbf{\Theta})'(\mathbf{X} - \mathbf{S}\mathbf{\Theta}) = \mathbf{X}'\mathbf{X} - n\mathbf{\Theta}^2 = n(\mathbf{C}_{\mathbf{X}\mathbf{X}} - \mathbf{\Theta}^2)$  can be defined as  $n(\mathbf{C}_{\mathbf{X}\mathbf{X}} - \mathbf{\Theta}^2) = \mathbf{W}\mathbf{\Omega}^2\mathbf{W}'$  with  $\mathbf{\Theta}^2$  given by MRFA for  $\mathbf{C}_{\mathbf{X}\mathbf{X}}$ . This SVD can provide  $\mathbf{\Lambda}$  with (34).

#### 4.2. Reformulation Matched to the CompFA Model

The CDFA solution, i.e., the parameter estimates resulting in the minimization of (27) under (9), (10), and (33), satisfies (36). This implies that (36) can be included in the constraints: CDFA can also be formulated as minimizing (27) subject to (9), (10), (33), and (36). Here, constraints (33) and (36) are proved to be equivalent to (12) and (11), respectively, under (4) and (10), in the next two theorems.

**Theorem 4.** Under (4) and (10), (33) is equivalent to (12), i.e.,  $\mathbf{S}'\mathbf{E} = {}_m\mathbf{O}_p$ .

*Proof.* First, (33)  $\rightarrow$  (12) is proved as follows: Using (4), we can rewrite (33) as  $\mathbf{S}'(\mathbf{F}\mathbf{\Lambda}' + \mathbf{E}) = {}_p\mathbf{O}_p$ , which implies (12), i.e.,  $\mathbf{S}'\mathbf{E} = {}_m\mathbf{O}_p$ , from (10). Next, (12)  $\rightarrow$  (33) is proved as follows: We can use (4) and (10) to rewrite (12) as  $\mathbf{S}'(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{S}\mathbf{\Theta}) = \mathbf{S}'(\mathbf{X} - \mathbf{S}\mathbf{\Theta}) = {}_p\mathbf{O}_p$ , i.e., (33).  $\square$

**Theorem 5.** Under (4) and (10), (36) is equivalent to (11), i.e.,  $\mathbf{F}'\mathbf{E} = {}_m\mathbf{O}_p$ .

*Proof.* First, let us prove (36)  $\rightarrow$  (11). The former is rewritten as  $\mathbf{F}'\mathbf{X} = n\mathbf{\Lambda}'$ . Using this, (4), and (10), we have  $\mathbf{F}'\mathbf{E} = \mathbf{F}'(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{S}\mathbf{\Theta}) = \mathbf{F}'\mathbf{X} - \mathbf{F}'\mathbf{F}\mathbf{\Lambda}' = n\mathbf{\Lambda}' - n\mathbf{\Lambda}' = {}_m\mathbf{O}_p$ , i.e., (11). Next, (11)  $\rightarrow$  (36) is proved as follows: We can use (4) and (10) to rewrite (11) as  $\mathbf{F}'(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{S}\mathbf{\Theta}) = \mathbf{F}'\mathbf{X} - \mathbf{F}'\mathbf{F}\mathbf{\Lambda}' = \mathbf{F}'\mathbf{X} - n\mathbf{\Lambda}' = {}_m\mathbf{O}_p$ , which leads to (36).  $\square$

These theorems and (10) show that  $\mathbf{F}'\mathbf{S} = {}_m\mathbf{O}_p$ ,  $\mathbf{F}'\mathbf{E} = {}_m\mathbf{O}_p$ , and  $\mathbf{S}'\mathbf{E} = {}_p\mathbf{O}_p$  in (10)–(12) hold true in the CDFA solution. These three equations imply that the common factors, specific factors, and errors are mutually decomposed completely, thus the name CDFA. This name is considered more suitable than another name, *constrained uniqueness FA*, used in Adachi (2019), as the CompFA model has not been considered in the latter naming. Further, the theorems show that Stegeman's (2016) procedure in Sect. 4.1 can be reformulated as minimizing least squares function (27) for model (4) subject to its constraints (9)–(12). Thus, CDFA is perfectly matched to the CompFA model and its standard constraints in Sect. 2.2. Here, let us remember that MDFA can be reformulated as minimizing (27) subject to (9)–(11) and (28) (Sect. 3.2); this (28) is strengthened into (12) in CDFA.

### 4.3. Behaviors for CompFA Data

In this section, we consider how CDFA behaves for CompFA data (17). At first, the following corollary shows the CDFA solution in strong uncorrelated error condition (15):

**Corollary 3.** *CDFA can be substituted for MDFA in Corollary 2.*

*Proof.* This is the same as the proof for Corollary 2, since (27) is also the CDFA loss function.  $\square$

This corollary shows that the CDFA estimate of  $\Theta^2$  (specific variances) is contaminated by  $\underline{\mathbf{D}}_{EE}$  (error variances) for the CompFA data with (15), as is the MDFA estimate. However, we can argue that the contamination is less likely to occur in CDFA than in MDFA for the usual CompFA data that are not restricted by (15). This argument follows from the fact that constraint (28), i.e.,  $\text{diag}(\mathbf{S}'\mathbf{E}) = {}_p\mathbf{O}_p$ , in MDFA is strengthened as (12), i.e.,  $\mathbf{S}'\mathbf{E} = {}_p\mathbf{O}_p$ , in CDFA, as explained in the following paragraph.

In Sect. 3.3, we discussed that the MDFA estimate of  $\Theta^2$  can be contaminated as (32), which follows from (30) with  $\bar{\mathbf{S}} = (\underline{\mathbf{S}}\underline{\Theta} + \underline{\mathbf{G}}\underline{\Delta})\bar{\Theta}^{-1}$ . This matrix and  $\mathbf{E}$  lead to  $\bar{\mathbf{S}}'\mathbf{E} = n\bar{\Theta}^{-1}(\underline{\Delta}\underline{\Gamma}' - \underline{\Delta}^2)$  as shown by (A6) in Appendix 2. Here,  $\underline{\Delta}$  is a  $p \times p$  diagonal matrix and  $\underline{\Gamma}$  is defined as in (29). In CDFA, the above equation for  $\bar{\mathbf{S}}'\mathbf{E}$  must be substituted into  $\mathbf{S}'\mathbf{E}$  in (12) as

$$\bar{\mathbf{S}}'\mathbf{E} = n\bar{\Theta}^{-1}(\underline{\Delta}\underline{\Gamma}' - \underline{\Delta}^2) = {}_p\mathbf{O}_p, \quad (37)$$

since (12) is included in the CDFA constraints as described in Sect. 4.2. The equivalence of (37) to strong uncorrelated error condition (15) is shown next:

**Theorem 6.** *Substituting  $\underline{\mathbf{E}}$  in (17) into  $\mathbf{E}$  in (15), it is rewritten as  $\underline{\mathbf{C}}_{EE} = \underline{\mathbf{D}}_{EE}$ . This is equivalent to (37).*

*Proof.* The last identity in (37) holds if and only if  $\underline{\Gamma}$  is a diagonal matrix:  $\underline{\Gamma} = \underline{\Delta}$ . This equivalence to  $\underline{\mathbf{C}}_{EE} = \underline{\mathbf{D}}_{EE}$  is shown using (29) rewritten as  $\underline{\mathbf{C}}_{EE} = n^{-1}\mathbf{E}'\mathbf{E} = \underline{\Gamma}\underline{\Gamma}'$ . That is,  $\underline{\Gamma} = \underline{\Delta}$  implies that  $\underline{\mathbf{C}}_{EE} = \underline{\Gamma}\underline{\Gamma}'$  is also diagonal:  $\underline{\mathbf{C}}_{EE} = \underline{\mathbf{D}}_{EE}$ . On the other hand,  $\underline{\mathbf{C}}_{EE} = \underline{\mathbf{D}}_{EE}$ , i.e.,  $\underline{\mathbf{C}}_{EE} = \underline{\Gamma}\underline{\Gamma}'$  being diagonal, implies that  $\underline{\Gamma}$  is diagonal:  $\underline{\Gamma} = \underline{\Delta}$ .  $\square$

This theorem shows that contamination (32) is less likely to occur in CDFA for the data that do not satisfy (15). On the other hand, even for such data, (32) can occur in MDFA, where  $\mathbf{S}'\mathbf{E}$  may not be  ${}_p\mathbf{O}_p$ , but only  $\text{diag}(\mathbf{S}'\mathbf{E}) = {}_m\mathbf{O}_p$  is required. These arguments are empirically supported in the next section.

## 5. Simulation Study

We assess the performance of the FA procedures for the CompFA data in a simulation study. Its purposes and the data types to be simulated are detailed in Sect. 5.1. Data analysis and assessment procedures are described in Sect. 5.2, and the results are reported in Sects. 5.3–5.5.

### 5.1. Purposes and Data Synthesis Procedures

In this study, the FA procedures are carried out for the CompFA data synthesized with the true loading matrix  $\underline{\mathbf{A}}$  and specific variance matrix  $\underline{\Theta}^2$ . The major purpose of this study is to numerically assess the following hypotheses:

- [H<sub>1</sub>] The CDFA estimate of  $\Theta^2$  approximates the true  $\underline{\Theta}^2$  better than the MDFA estimate.  
 [H<sub>2</sub>] The MDFA estimate of  $\Theta^2$  approximates  $\underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$ , i.e., is contaminated by  $\underline{\mathbf{D}}_{EE}$ , with  $\underline{\mathbf{D}}_{EE}$  the diagonal matrix including the true error variances.  
 [H<sub>3</sub>] The LVFA estimate of unique variance matrix  $\Psi^2$  approximates  $\underline{\Theta}^2 + \underline{\mathbf{D}}_{EE}$ .  
 [H<sub>4</sub>] The estimates of  $\Lambda$  in all procedures approximate the true  $\underline{\Lambda}$ .

Here, [H<sub>1</sub>] follows from the discussion in Sect. 4.3, and hypotheses [H<sub>2</sub>] and [H<sub>3</sub>] are equivalent to (32) and (26), respectively. [H<sub>4</sub>] has not been discussed, but rather has been presupposed for the discussions in Sects. 2–4. The estimates in the hypotheses are obtained, given the covariance matrix  $\mathbf{C}_{XX} = n^{-1}\mathbf{X}'\mathbf{X}$  for the CompFA data. Thus, we describe below how the true  $\underline{\Lambda}$  and  $\underline{\Theta}^2$  are set and how they lead to  $\mathbf{C}_{XX}$ .

Let  $U(\alpha, \beta)$  denote the uniform distribution for the interval  $[\alpha, \beta]$ . Each element of  $\underline{\Lambda}$  and each diagonal one of  $\underline{\Theta}^2$  are drawn from  $U(-1, 1)$  and  $U(0.1, 0.8)$ , respectively, subject to  $\text{rank}(\underline{\Lambda}) = m$ . From the resulting  $\underline{\Lambda}$  and  $\underline{\Theta}^2$ , we generate 12 types of  $\mathbf{C}_{XX}$  with  $\text{rank}(\mathbf{C}_{XX})$  constrained to be  $p$ . Here the 12 ( $= 3 \times 2 \times 2$ ) types are defined by combining the three levels of error correlations, two levels of error magnitudes, and two versions of the CompFA model, which are explained in the following paragraphs.

The two versions of the model correspond to its nonrandom (N) version (4) and random (R) version (3). The covariance matrix for the N version is given by (24) for (17) with  $\underline{\mathbf{C}}_{EE} = n^{-1} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i'$ :

$$\mathbf{C}_{XX} = \underline{\Lambda} \underline{\Lambda}' + \underline{\Theta}^2 + \underline{\mathbf{C}}_{EE} = \underline{\Lambda} \underline{\Lambda}' + \underline{\Theta}^2 + \frac{1}{n} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i'. \quad (38)$$

Here, error vectors  $\mathbf{e}_i$  ( $i = 1, \dots, n$ ) are chosen as  $\mathbf{e}_i = \alpha \boldsymbol{\varepsilon}_i$ , with  $\boldsymbol{\varepsilon}_i$  drawn from  $N_p(\mathbf{0}_p, \Phi)$ , i.e., the  $p$ -variate normal distribution whose mean vector is  $\mathbf{0}_p$  and covariance matrix is  $\Phi(p \times p)$ . How  $\alpha$  and  $\Phi$  are defined is described later. The matrix for the R version is given by

$$\mathbf{C}_{XX} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \frac{1}{n} \sum_{i=1}^n (\underline{\Lambda} \mathbf{f}_i + \underline{\Theta} \mathbf{s}_i + \mathbf{e}_i)(\underline{\Lambda} \mathbf{f}_i + \underline{\Theta} \mathbf{s}_i + \mathbf{e}_i)'. \quad (39)$$

This follows from (3), whose random vectors are followed by the observation-number subscript  $i$ . The factor score vectors in (39) are sampled with  $[\mathbf{f}_i', \mathbf{s}_i']' \sim N_{m+p}(\mathbf{0}_{m+p}, \mathbf{I}_{m+p})$ .

The three levels of error correlations can be referred to as no, low, and high levels ( $\mathbf{C}_N$ ,  $\mathbf{C}_L$ , and  $\mathbf{C}_H$ ), while the two levels of error magnitudes can be called low and high levels ( $\mathbf{E}_L$  and  $\mathbf{E}_H$ ), respectively. Here, C, E, N, L, and H in the parentheses are abbreviations of correlation, error, no, low, and high, respectively. At the  $\mathbf{C}_N$  level,  $\Phi$  is set to the diagonal matrix  $\mathbf{D}_R$  whose diagonal elements are drawn from  $U(0.1, 0.8)$ . At the  $\mathbf{C}_L$  and  $\mathbf{C}_H$  levels,  $\Phi$  is set to  $\mathbf{D}_R^{1/2} \mathbf{R} \mathbf{D}_R^{1/2}$ , with  $\mathbf{R} = (r_{jk})$  the  $p \times p$  symmetric nonnegative-definite matrix whose elements are chosen as  $r_{jj} = 1$  and  $r_{jk} = \tau_{jk} \tilde{r}_{jk}$  for  $j \neq k$ . Here,  $\tau_{jk}$  is randomly set to 1 or  $-1$ , and  $\tilde{r}_{jk}$  is drawn from  $N_1(0.2\rho, 0.05^2\rho^2)$  subject to  $-1 < \tilde{r}_{jk} < 1$ , with  $\rho = 1$  for  $\mathbf{C}_L$  and  $\rho = 2$  for  $\mathbf{C}_H$ . The  $\alpha$  value is set so that  $\text{tr}\Phi/\text{tr}\Lambda\Lambda' = 0.1$  and  $0.2$  for the  $\mathbf{E}_L$  and  $\mathbf{E}_H$  levels, respectively. We should notice that  $\Phi$  is diagonal at the  $\mathbf{C}_N$  level, but  $\underline{\mathbf{C}}_{EE} = n^{-1} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i'$  obtained with  $\mathbf{e}_i \sim N_p(\mathbf{0}_p, \Phi)$  is not necessarily diagonal; the  $\mathbf{C}_N$  level does not exactly match (15). In Appendix 3, it is reported that a simulation study for  $\mathbf{C}_{XX}$  satisfying (15) numerically demonstrated the facts in Corollaries 1–3.

The procedures in the last three paragraphs were replicated 500 times with the settings  $n = 200$ ,  $p = 12$ , and  $m = 3$ . Thus, we had 6000 ( $= 12$  types  $\times$  500 times)  $\mathbf{C}_{XX}$ .

## 5.2. Data Analysis and Assessment Procedures

For each of the 6000  $\mathbf{C}_{XX}$ , we performed MDFA, CDFA, and the two types of LVFA, i.e., the least squares LVFA (LS-LVFA) and maximum likelihood LVFA (ML-LVFA) for minimizing (22) and (23), respectively. Their algorithms are described in Appendix 4. In every procedure,  $\mathbf{\Lambda}$  has rotational indeterminacy; thus, it was rotated by the orthogonal Procrustes method (e.g., Adachi, 2020, p. 206), to optimally approximate the true counterpart  $\underline{\mathbf{\Lambda}}$  in a least squares sense. This approximation allows  $\mathbf{\Lambda}$  to be comparable to  $\underline{\mathbf{\Lambda}}$ ; thus, the Procrustes method has been typically used in previous simulation studies for FA (e.g., Adachi, 2013, Appendix D; Stegeman, 2016, Sect. 4.2).

The similarities between estimated parameters and their true values can be assessed with the smallness of a mean absolute difference (MAD). The MAD for loadings is defined as  $\text{MAD}(\underline{\mathbf{\Lambda}}) = \|\mathbf{\Lambda} - \underline{\mathbf{\Lambda}}\|_1 / (pm)$ , where  $\|\mathbf{\Lambda} - \underline{\mathbf{\Lambda}}\|_1$  denotes the  $L_1$  norm of  $\mathbf{\Lambda} - \underline{\mathbf{\Lambda}}$ , i.e., the sum of the absolute values of the elements in  $\mathbf{\Lambda} - \underline{\mathbf{\Lambda}}$ . The denominator  $pm$  in this definition is replaced by  $p$  in the following two MAD for  $p \times p$  diagonal matrices:

$$\text{MAD}(\underline{\mathbf{\Theta}}^2) = \begin{cases} \frac{1}{p} \|\underline{\mathbf{\Psi}}^2 - \underline{\mathbf{\Theta}}^2\|_1 & \text{for LVFA} \\ \frac{1}{p} \|\underline{\mathbf{\Theta}}^2 - \underline{\mathbf{\Theta}}^2\|_1 & \text{for MDFA and CDFA} \end{cases}, \quad (40)$$

$$\text{MAD}(\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}) = \begin{cases} \frac{1}{p} \|\underline{\mathbf{\Psi}}^2 - (\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE})\|_1 & \text{for LVFA} \\ \frac{1}{p} \|\underline{\mathbf{\Theta}}^2 - (\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE})\|_1 & \text{for MDFA and CDFA} \end{cases} \quad (41)$$

with  $\underline{\mathbf{D}}_{EE} = \text{diag}(\underline{\mathbf{C}}_{EE}) = \text{diag}(n^{-1} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}'_i)$ . Here, MAD in (40) and (41) for LVFA differs from those for MDFA and CDFA, as  $\underline{\mathbf{\Psi}}^2$  corresponds to  $\underline{\mathbf{\Theta}}^2$  in LVFA. We obtain (41) in addition to (40), as both MAD are necessary for examining  $[\mathbf{H}_2]$  and  $[\mathbf{H}_3]$  in Sect. 5.1.

Let  $\{\mathbf{\Lambda}_{LS}, \mathbf{\Psi}_{LS}\}$  and  $\{\mathbf{\Lambda}_{ML}, \mathbf{\Psi}_{ML}\}$  denote the LS- and ML-LVFA estimates of  $\{\mathbf{\Lambda}, \mathbf{\Psi}\}$ , respectively. The broad equivalence between  $\{\mathbf{\Lambda}_{LS}, \mathbf{\Psi}_{LS}\}$  and  $\{\mathbf{\Lambda}_{ML}, \mathbf{\Psi}_{ML}\}$  was found with the averages (standard deviations) of  $\|\mathbf{\Lambda}_{LS} - \mathbf{\Lambda}_{ML}\|_1 / (pm)$  and  $\|\mathbf{\Psi}_{LS}^2 - \mathbf{\Psi}_{ML}^2\|_1 / p$  over the solutions for 5935 ( $= 6000 - 65$ )  $\mathbf{C}_{XX}$  being 0.010 (0.011) and 0.022 (0.019), respectively. Here, we have removed the 65  $\mathbf{C}_{XX}$  for which LS-LVFA provided improper solutions with negative unique variances. As described in Appendix 4, such improper solutions show an undesirable property of LS-LVFA. These results allowed us to only consider the ML-LVFA solutions; thus, we refer to only ML-LVFA as LVFA, hereafter.

Now, we have the MAD values for the three matrices ( $\mathbf{M}$ ),  $\underline{\mathbf{\Lambda}}$ ,  $\underline{\mathbf{\Theta}}^2$ , and  $\underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}$ , which were obtained from each solution of the three FA procedures (P) performed for the 6000 ( $= 3 \times 2 \times 2 \times 500$ )  $\mathbf{C}_{XX}$ . Here, these  $\mathbf{C}_{XX}$  are classified by combining the three levels of error correlations (C), two levels of error magnitudes (E), two versions (V) of the model, and 500 replications (R). In order to assess the hypotheses in Sect. 5.1, we performed analysis of variance of the randomized block design (ANOVA-RBD) (e.g., Kirk, 2013) for the MAD values. Here, P, M, C, E, and V were treated as treatments and R was treated as a block factor; the factorial design can be expressed as  $P \times M \times C \times E \times V$  with R consisting of 500 blocks, with the sets of the levels in P, M, C, E, and V being {LVFA, MDFA, CDFA},  $\{\underline{\mathbf{\Lambda}}, \underline{\mathbf{\Theta}}^2, \underline{\mathbf{\Theta}}^2 + \underline{\mathbf{D}}_{EE}\}$ ,  $\{E_L, E_H\}$ , and  $\{N, R\}$ , respectively.

The above ANOVA-RBD provided the  $F$  values for main and interaction effects in Table 2. We do not use those values for statistical hypothesis testing, since this is senseless for our simulated data, whose sample size is too large to reject null hypotheses. But rather, we consider the  $F$  values as standing for the sizes of the effects, as the number of the levels in each treatment is restricted to two or three; thus, it makes sense to compare the  $F$  values. We regard the effects with  $F \geq 24,000$  as substantial enough, though the lower limit of 24,000 is a benchmark threshold chosen because 24,000 is far greater than the largest one ( $= 14,491.0$ ) among the  $F$ -values less than 24,000. Thus, the five effects boldfaced in Table 2 are regarded as substantial.

TABLE 2.  
*F* values resulting in ANOVA-RBD with the substantial effects boldfaced whose *F* values > 24,000.

Effect	Source and <i>F</i> value					
Main	P	<b>M</b>	C	<b>E</b>	<b>V</b>	Block
	444.5	<b>195361.0</b>	14491.0	<b>107830.5</b>	<b>26992.8</b>	36.4
Two-way interaction	<b>P × M</b>	P × C	P × E	P × V	M × C	<b>M × E</b>
	<b>42997.4</b>	67.7	213.6	952.1	5845.3	<b>24629.0</b>
	M × V	C × E	C × V	E × V		
Three-way interaction	8393.9	1938.9	235.3	341.9		
	P × M × C	P × M × E	P × M × V	P × C × E	P × C × V	P × E × V
	2051.2	4591.3	400.7	16.3	27.7	77.7
	M × C × E	M × C × V	M × E × V	C × E × V		
Higher-order interactions	1090.1	908.0	102.3	4.8		
	P × M × C × E	P × M × C × V	P × M × E × V	P × C × E × V		
	291.1	277.6	28.6	1.3		
	M × C × E × V	P × M × C × E × V				
	32.2	7.8				

TABLE 3.  
Averages of MAD for substantial effects with those for main effects shown by boldfaced italic letters.

	(A) Procedure × Matrix			(B) Matrix × Error			(C) Version	
	$\underline{\Lambda}$	$\underline{\Theta}^2$	$\underline{\Theta}^2 + \underline{D}_{EE}$		$E_L$	$E_H$	<i>Nonrandom</i>	<i>0.099</i>
LVFA	0.045	0.228	0.074	$\underline{\Lambda}$	0.037	0.053	<i>Random</i>	<i>0.131</i>
MDFA	0.044	0.217	0.073	$\underline{\Theta}^2$	0.132	0.249		
C DFA	0.046	0.126	0.182	$\underline{\Theta}^2 + \underline{D}_{EE}$	0.082	0.137		
<i>Matrix</i>	<i>0.045</i>	<i>0.191</i>	<i>0.110</i>	<i>Error</i>	<i>0.084</i>	<i>0.146</i>		

### 5.3. Averages for Substantial Effects

The averages of MAD for the levels associated with the substantial effects are presented in Table 3. The averages for the main effect of V in Panel (C) show that the MAD for R version (39) is greater than that for N one (38); this can be attributable to the fact that (39) has higher randomness. The averages for the other main effects in the bottom rows of Panels (A) and (B) can be interpreted from the interactions considered next.

The averages for the  $P \times M$  levels in Panel (A) show that  $[H_1]$ – $[H_4]$  in Sect. 5.1 were supported:  $[H_1]$  is supported by the CDFA average 0.126 for (40) being less than the MDFA counterpart 0.217,  $[H_2]$  and  $[H_3]$  are supported by the result that the LVFA and MDFA averages for (41) are less than those for (40), and  $[H_4]$  is supported by the averages of  $MAD(\underline{\Lambda})$  being small and almost equivalent among the procedures.

Panel (B) shows that every MAD increases with the change from  $E_L$  to  $E_H$ , but the increments in (40) and (41) values are much larger than those in  $MAD(\underline{\Lambda})$ . This difference in the increments can be interpreted by taking into account the related three-way interaction, as explained in the next subsection.



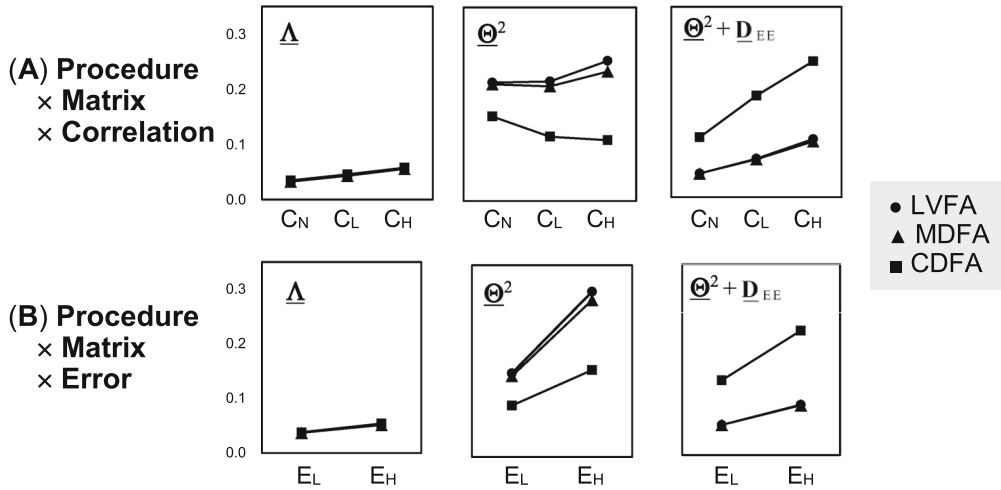


FIGURE 1.

Averages of MAD for three-way interactions (A) and (B), which facilitate the interpretation of the substantial two-way interactions.

5.4. Three-way Interactions Related to the Substantial Two-way Interactions

Figure 1 shows the averages associated with the  $P \times M \times C$  and  $P \times M \times E$  interactions, whose  $F$  values are the largest among those for the three-way interactions (Table 2). Although those values are not substantial, we note Fig. 1 for exploring the mechanisms that underlie the interactions treated in Sect. 5.3.

Panel (A) in Fig. 1 shows that the changes of  $C_N$  to  $C_L$  and  $C_L$  to  $C_H$  increase MAD ( $\underline{\Lambda}$ ) for all procedures, but decrease the (40) value for CDFA. This result can be explained by the similarity of the matrices  $\underline{C}_{EE} = n^{-1} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}'_i$  at  $C_N$  to (15) and the deviation of  $\underline{C}_{EE}$  at  $C_H$  from (15). Here, (15) leads to the solution of  $\underline{\Lambda} = \underline{\Lambda}$  and  $\underline{\Psi}^2 = \underline{\Theta}^2 = \underline{\Theta}^2 + \underline{D}_{EE}$ , as shown in Corollaries 1–3. However, for  $\underline{C}_{EE}$  deviating from (15), the CDFA estimate of  $\underline{\Theta}^2$  can approximate  $\underline{\Theta}^2$ , as discussed with Theorem 6. Thus, the deviation of  $\underline{C}_{EE}$  from (15) increases MAD( $\underline{\Lambda}$ ), but decreases (40) only in CDFA. The above explanation can also be used for the increase in the CDFA average of (41) with the deviation from (15). On the other hand, the deviation is not found to decrease the LVFA and MDFA values of (40). This is congruous with (26) and (32). Further, the LVFA and MDFA averages of (40) rather increase with the change from  $C_L$  to  $C_H$ . These increases may be correlated to those in MAD( $\underline{\Lambda}$ ) from  $C_L$  to  $C_H$ , as  $\underline{\Lambda}$  is jointly estimated with  $\underline{\Theta}^2$  and  $\underline{\Psi}^2$  in (40). The LVFA and MDFA averages of (41) being smaller than those of (40) at all  $C$  levels is also congruous with (26) and (32).

Panel (B) shows that the increments in MAD with the change from  $E_L$  to  $E_H$  differ among the procedures and matrices. For explaining the differences, we must consider error magnitudes (EM) and  $\underline{D}_{EE}$  effects, where the EM effect stands for the errors with greater EM disturbing parameter estimation more deeply, and the  $\underline{D}_{EE}$  effect refers to the fact that the diagonal elements of  $\underline{D}_{EE}$  for  $E_H$  have greater values than those for  $E_L$ . We can consider that the increase in MAD( $\underline{\Lambda}$ ) for every procedure follows only from the EM effect, while those in the LVFA and MDFA values of (40) are affected by the EM plus  $\underline{D}_{EE}$  effects, because of (26) and (32). The increments in (40) being greater than their CDFA counterpart can be explained with Theorem 6, which shows that (32) is less likely to occur in CDFA for  $\underline{C}_{EE}$  deviating from (15). On the other hand, (41) values and their increments for MDFA and LVFA are found to be smaller than their (40) counterparts. This result can be attributable to the fact that (41) is affected only by the EM effect in LVFA and

MDFA, as the use of their properties (26) and (32) in (41) allows us to find that  $\text{MAD}(\underline{\Theta}^2 + \underline{\mathbf{D}}_{EE})$  can approximate zero, but this approximation can be deteriorated by the EM effect.

### 5.5. Additional Results

The results considered so far show the advantage of CDFA over the two other procedures. Which of those two is better may be answered from the following result: The panels for  $\underline{\Theta}^2$  in Fig. 1 show that the MDFA estimate of  $\underline{\Theta}^2$  is slightly closer to  $\underline{\Theta}^2$  than the LVFA estimate of  $\underline{\Psi}^2$  on average. The generality of this relationship was found: The  $\Delta$  value, which is defined as the LVFA value of (40) minus its MDFA value, was positive for the 5819 pairs of LVFA–MDFA solutions among the 6000 for all  $\mathbf{C}_{XX}$ . Further, we performed ANOVA-RBD for  $\Delta$  with C, E, and V treated as treatments, and R treated as blocks. Among all resulting  $F$  values, the one for the main effect of C (3651.7) and the value for the effect of E (3238.4) were the largest, with the other  $F$  values less than 940. However, the two effects for the largest  $F$  values are not considered substantial, with the averages at levels  $C_N$ ,  $C_L$ ,  $C_H$ ,  $E_L$ , and  $E_H$  being 0.003, 0.008, 0.019, 0.006, and 0.014, respectively; even the largest of those cannot be regarded as substantial. In conclusion, the MDFA estimate of  $\underline{\Theta}^2$  is significantly closer to  $\underline{\Theta}^2$  than the LVFA estimate of  $\underline{\Psi}^2$ , but not substantially closer.

## 6. Real Data Illustration

The results in the previous sections can be summarized as follows: In CDFA whose formulation is exactly matched to the CompFA model, its parameters (factor loadings and specific variances) are well recovered for the CompFA data. In MDFA and LVFA, the loadings can be recovered as well as in CDFA, but the specific variances cannot be recovered, so their MDFA estimates are contaminated by the error variances, and the LVFA estimates of the unique variances approximate the sums of the specific and error variances. These results suggest that CDFA should be used for the CompFA data, particularly for the purpose of estimating the specific variances. Therefore, in this section, we use two real data sets to illustrate how the CDFA estimates should be interpreted, and we show the merits of using CDFA through comparisons among CDFA, MDFA, and LVFA solutions, on the supposition that the data sets are underlaid by the CompFA model. Here, LVFA is restricted to ML-LVFA, as its solutions were almost equivalent to LS-LVFA.

One of the data sets is that from Yanai and Ichikawa (2007) for personality test scores of  $n = 200$  students for  $p = 12$  items. The other data set, which we obtained from Izenman's (2008) website, is known as Holzinger and Swineford's (1939) 24 psychological tests data and contains intelligence test scores of  $n = 301$  participants for  $p = 24$  items. We performed the FA procedures for the correlation matrices for the personality and intelligence test scores, with  $m$  set at 3 and 5, respectively. In every procedure,  $\mathbf{\Lambda}$  was rotated by the varimax method (Kaiser, 1958) which is typically used in FA for real data. Tables 4 and 5 present the resulting  $\mathbf{\Lambda} = [\lambda_1, \dots, \lambda_p]$ ,  $\|\lambda_1\|^2, \dots, \|\lambda_p\|^2$ ,  $\psi_j, \theta_j$ , and  $ev_j$ , with the last three terms being the  $j$ th diagonal elements of  $\underline{\Psi}^2$ ,  $\underline{\Theta}^2$ , and  $\mathbf{C}_{EE}$ , respectively, and  $ev$  an abbreviation of error variance. In the tables, the sub-/superscripts of L, M, and C, which stand for LVFA, MDFA, and CDFA, respectively, have been attached to  $\mathbf{\Lambda}$ ,  $\lambda_j$ ,  $\theta_j$ , and  $ev_j$ , for distinguishing the solutions for different procedures. We use the above notation with sub-/superscripts in this section.

Tables 4 and 5 show that the loadings resulting in all procedures are mutually similar and lead to the identical interpretation of common factors. Although this result does not show the merits of using CDFA, the merits are shown by the other results, as described in the following paragraphs. There, an important role is fulfilled by the fact that

$$v_j = \|\lambda_j^C\|^2 + \Theta_j^{C2} + ev_j^C = \|\lambda_j^M\|^2 + \Theta_j^{M2} + ev_j^M = \|\lambda_j^L\|^2 + \Psi_j^2 \quad (42)$$

holds for  $v_j = n^{-1}\|\mathbf{X}\|^2$ , which denotes the variance for variable  $j$  and is also the  $j$ th diagonal element of  $\mathbf{C}_{XX}$ . Here, the first identity follows from the fact that the CDFA solution meets (9)–(12) and thus (13), the last identity follows from (25), and the second identity is derived as follows: (4), (10), and (11) imply  $\text{diag}(\mathbf{C}_{XX}) = \text{diag}(\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Theta}^2 + \mathbf{C}_{EE} + n^{-1}\mathbf{\Theta}\mathbf{S}'\mathbf{E} + n^{-1}\mathbf{E}'\mathbf{S}\mathbf{\Theta})$  with  $\text{diag}(\mathbf{E}'\mathbf{S}\mathbf{\Theta}) = \text{diag}(\mathbf{\Theta}\mathbf{S}'\mathbf{E}) = \mathbf{\Theta}\text{diag}(\mathbf{S}'\mathbf{E}) = {}_p\mathbf{O}_p$  from (28). Each of the terms  $\|\lambda_j\|^2$  (with superscripts) in (42) can be called a *common variance* (or communality) by abbreviating the variance of the common factor part affecting variable  $j$ , as found from  $\|\lambda_j\|^2$  equaling  $n^{-1}\|\mathbf{F}\lambda_j\|^2$  and the variance of  $\lambda_j'\mathbf{f}$  under (6) and (10). When  $\mathbf{C}_{XX}$  is a correlation matrix as in our case,  $v_j = 1$ ; thus,  $\|\lambda_j\|^2$ ,  $\theta_j^2$ , and  $ev_j$  (with superscripts) in (42) stand for the proportions of the common, specific, and error variances in the variance of variable  $j$ , respectively, with  $\psi_j^2$  the proportion of the unique variance.

Let us note the CDFA solution in Table 4 with keeping (42) in mind. For example, the results  $\|\lambda_1^C\|^2 = 0.44$ ,  $\Theta_1^{C2} = 0.39$ , and  $ev_1^C = 0.176$  for the variable *extraversion* show that 44 percent of the variations in *extraversion* are explained by the common factors, 39 percent are accounted for by the factor specific to *extraversion*, and 17.6 percent of the variations remain unexplained by common and specific factors. The comparison of  $\Theta_j^{C2}$  across  $j (= 1, \dots, p)$  shows that  $\Theta_3^{C2} = 0.45$  for *empathy* is the largest and *empathy* is affected most by the corresponding specific factor among all variables. On the other hand,  $\Theta_2^{C2} = 0.18$  for *activity* is the smallest among all  $\Theta_j^{C2}$ , but  $\|\lambda_2^C\|^2 = 0.70$  is the largest among all  $\|\lambda_j^C\|^2$ , implying that *activity* is affected least by the corresponding specific factor, but is explained best by the common factors, among all variables.

MDFA and LVFA solutions can be interpreted in a parallel manner, except that the term "specific" is replaced by "unique" and an error variance is not obtained in LVFA. However, we can find in Tables 4 and 5 that  $\Theta_j^{M2}$  and  $\Psi_j^2$  are much greater than  $\Theta_j^{C2}$  and rather close to  $\Theta_j^{C2} + ev_j^C$  for almost all variables. This finding is congruous with (26) and (32). For example,  $\Theta_1^{C2} = 0.60$  and  $\Psi_1^2 = 0.61 > \Theta_1^{C2} = 0.39$  for the variable *extraversion* in Table 4. We can consider that  $\Theta_1^{C2} = 0.60$  and  $\Psi_1^2 = 0.61$  are contaminated by the true error variance for *extraversion* and greater than its true specific variance.

We can also find that the error variance  $ev_j^M$  resulting in MDFA is far smaller than its CDFA counterpart  $ev_j^C$  for every variable. This result follows from the fact that MDFA is less restrictive than CDFA; (27) value  $n^{-1}\|\mathbf{E}\|^2 = \text{tr}\mathbf{C}_{EE} = \sum_{j=1}^p ev_j$  in MDFA cannot be greater than that in CDFA, which suggests that  $ev_j^M < ev_j^C$  tends to occur. This property does not imply goodness of the MDFA solution but is rather congruous with its undesirable property shown by (32); comparing this with  $ev_j^M = v_j - \|\lambda_j^M\|^2 - \Theta_j^{M2}$  following from (42) allows us to find that  $\Theta_j^{M2}$  being larger than its true value decreases  $ev_j^M$ .

## 7. Conclusion

In this paper, latent variable factor analysis (LVFA), matrix decomposition factor analysis (MDFA), and its variant from Stegeman (2016) were revisited from the comprehensive FA (CompFA) model. The variant of MDFA was reformulated to be called CDFA and exactly underlaid by the CompFA model. On the other hand, MDFA was reformulated as the procedure with (12) in the CompFA model assumptions relaxed as (28). We also showed how the model for LVFA can be related to the CompFA model.

TABLE 4.  
Solutions for personality test data.

Variable	LVFA			MDFA			CDFA							
	$\Lambda_L = [\lambda_1^L \dots \lambda_{12}^L]'$	$\ \lambda_j^L\ ^2$	$\psi_j^2$	$\Lambda_M = [\lambda_1^M \dots \lambda_{12}^M]'$	$\ \lambda_j^M\ ^2$	$\theta_j^{M2}$	$ev_j^M$	$\Lambda_C = [\lambda_1^C \dots \lambda_{12}^C]'$	$\ \lambda_j^C\ ^2$	$\theta_j^{C2}$	$ev_j^C$			
<i>Extraversion</i>	-0.32	0.48	0.24	0.39	0.61	0.25	0.39	0.46	0.25	0.50	0.25	0.44	0.39	0.176
<i>Activity</i>	-0.25	0.65	0.42	0.66	0.33	0.64	0.43	0.66	0.33	0.67	0.43	0.70	0.18	0.118
<i>Empathy</i>	-0.02	0.03	0.58	0.34	0.66	-0.03	0.01	0.60	0.62	0.00	0.63	0.40	0.45	0.151
<i>Novelty</i>	-0.05	0.62	0.04	0.39	0.62	-0.06	0.61	0.05	0.38	0.65	0.04	0.43	0.44	0.141
<i>Durability</i>	-0.07	0.04	0.68	0.47	0.53	-0.07	0.04	0.67	0.46	0.05	0.68	0.47	0.38	0.151
<i>Regularity</i>	0.06	0.17	0.71	0.54	0.46	0.07	0.17	0.71	0.54	0.18	0.76	0.62	0.23	0.152
<i>Self-revelation</i>	0.16	0.62	0.03	0.41	0.59	0.14	0.62	0.04	0.41	0.16	0.66	0.46	0.42	0.122
<i>Aggressiveness</i>	0.39	0.54	-0.14	0.46	0.54	0.39	0.57	-0.15	0.50	0.41	0.56	0.51	0.41	0.084
<i>Uncooperativeness</i>	0.45	0.16	-0.17	0.26	0.74	0.46	0.18	-0.19	0.28	0.51	0.20	0.35	0.39	0.259
<i>Inferiority feeling</i>	0.62	-0.30	-0.17	0.50	0.49	0.63	-0.29	-0.17	0.51	0.63	-0.30	0.52	0.37	0.111
<i>Nervousness</i>	0.71	0.00	0.22	0.55	0.45	0.70	0.01	0.23	0.54	0.76	-0.02	0.65	0.18	0.169
<i>Depression</i>	0.82	0.03	-0.07	0.68	0.32	0.82	0.04	-0.06	0.68	0.82	0.03	0.68	0.22	0.103

TABLE 5.  
Solutions for intelligence test data.

Variable	LVFA				MDFA				CDEFA														
	$\Lambda_L = \{\lambda_1^L, \dots, \lambda_{24}^L\}$	$\ \lambda_j^L\ ^2$	$\psi_j^2$	$\Lambda_M = \{\lambda_1^M, \dots, \lambda_{24}^M\}$	$\ \lambda_j^M\ ^2$	$\theta_j^{M2}$	$ev_j^M$	$\Lambda_C = \{\lambda_1^C, \dots, \lambda_{24}^C\}$	$\ \lambda_j^C\ ^2$	$\theta_j^{C2}$	$ev_j^C$												
<i>Spatial 1</i>	0.31	0.63	0.10	0.12	0.01	0.52	0.49	0.31	0.10	0.12	0.03	0.52	0.48	0.006	0.30	0.65	0.10	0.13	0.12	0.55	0.35	0.095	
<i>Spatial 2</i>	0.14	0.48	-0.06	0.01	0.12	0.27	0.73	0.14	0.48	-0.06	0.00	0.14	0.27	0.73	0.002	0.13	0.46	-0.07	0.01	0.22	0.28	0.65	0.075
<i>Spatial 3</i>	0.19	0.43	0.06	-0.03	0.00	0.23	0.77	0.19	0.44	0.07	-0.03	0.01	0.24	0.76	0.005	0.19	0.49	0.06	-0.03	0.05	0.28	0.54	0.177
<i>Spatial 4</i>	0.06	0.63	0.09	0.14	0.09	0.44	0.56	0.06	0.64	0.09	0.14	0.10	0.45	0.55	0.002	0.05	0.61	0.08	0.15	0.20	0.44	0.49	0.059
<i>Verbal 1</i>	0.82	0.06	0.12	-0.02	0.04	0.69	0.30	0.83	0.06	0.12	-0.02	0.04	0.71	0.29	0.008	0.84	0.07	0.11	-0.02	0.06	0.73	0.16	0.106
<i>Verbal 2</i>	0.80	0.10	0.09	0.12	0.06	0.68	0.32	0.80	0.10	0.09	0.12	0.07	0.68	0.32	0.007	0.80	0.10	0.08	0.12	0.11	0.68	0.23	0.087
<i>Verbal 3</i>	0.88	0.05	0.06	0.01	-0.02	0.78	0.22	0.89	0.04	0.06	0.01	0.00	0.80	0.20	0.007	0.91	0.04	0.06	0.00	0.04	0.83	0.07	0.104
<i>Verbal 4</i>	0.71	0.16	0.10	0.10	0.07	0.55	0.44	0.71	0.15	0.10	0.10	0.10	0.56	0.44	0.007	0.72	0.14	0.10	0.10	0.15	0.58	0.29	0.139
<i>Verbal 5</i>	0.82	0.15	0.06	0.08	0.11	0.72	0.28	0.82	0.15	0.05	0.08	0.12	0.72	0.28	0.007	0.83	0.14	0.05	0.08	0.16	0.74	0.15	0.114
<i>Speed 1</i>	0.08	-0.08	0.79	0.08	0.14	0.66	0.34	0.08	-0.09	0.78	0.08	0.14	0.65	0.34	0.010	0.07	-0.12	0.80	0.07	0.14	0.68	0.23	0.098
<i>Speed 2</i>	0.29	0.18	0.57	0.24	-0.09	0.51	0.50	0.30	0.18	0.57	0.23	-0.04	0.50	0.49	0.009	0.30	0.17	0.60	0.23	0.02	0.53	0.31	0.155
<i>Speed 3</i>	0.09	0.23	0.62	0.00	0.04	0.45	0.56	0.09	0.22	0.62	0.00	0.06	0.44	0.55	0.005	0.08	0.22	0.65	0.00	0.11	0.49	0.37	0.137
<i>Speed 4</i>	0.17	0.45	0.51	0.06	-0.11	0.51	0.50	0.17	0.44	0.51	0.05	-0.06	0.49	0.51	0.007	0.16	0.46	0.54	0.04	0.03	0.53	0.32	0.142
<i>Memory 1</i>	0.15	0.05	0.04	0.64	0.08	0.44	0.56	0.15	0.04	0.04	0.63	0.11	0.43	0.56	0.007	0.15	0.00	0.02	0.69	0.14	0.52	0.31	0.179
<i>Memory 2</i>	-0.03	0.16	0.05	0.56	0.08	0.35	0.66	-0.03	0.17	0.04	0.57	0.08	0.36	0.63	0.010	-0.05	0.17	0.02	0.63	0.09	0.44	0.39	0.178
<i>Memory 3</i>	0.17	0.37	0.11	0.45	0.18	0.41	0.58	0.17	0.36	0.11	0.44	0.22	0.41	0.58	0.006	0.16	0.32	0.10	0.44	0.30	0.42	0.48	0.105
<i>Memory 4</i>	0.06	0.03	0.34	0.54	-0.03	0.41	0.60	0.06	0.03	0.34	0.54	-0.02	0.41	0.58	0.007	0.06	0.02	0.35	0.57	-0.01	0.45	0.39	0.153
<i>Memory 5</i>	0.12	0.17	0.22	0.42	0.01	0.27	0.73	0.12	0.17	0.23	0.44	0.02	0.29	0.70	0.012	0.12	0.17	0.24	0.49	0.02	0.34	0.43	0.226
<i>Memory 6</i>	0.24	0.16	0.11	0.31	0.22	0.24	0.76	0.24	0.13	0.12	0.30	0.26	0.25	0.74	0.013	0.23	0.06	0.12	0.31	0.35	0.29	0.51	0.200
<i>Math.1</i>	0.34	0.39	-0.01	0.21	0.31	0.41	0.60	0.33	0.38	-0.01	0.21	0.32	0.40	0.60	0.004	0.31	0.33	-0.03	0.21	0.40	0.41	0.49	0.101
<i>Math.2</i>	0.29	0.35	0.35	0.10	0.39	0.49	0.52	0.28	0.34	0.34	0.10	0.42	0.50	0.50	0.007	0.24	0.26	0.33	0.08	0.55	0.54	0.32	0.133
<i>Math.3</i>	0.49	0.36	0.06	0.11	0.24	0.44	0.55	0.49	0.35	0.06	0.11	0.27	0.45	0.55	0.009	0.48	0.29	0.05	0.10	0.38	0.47	0.38	0.155
<i>Math.4</i>	0.42	0.47	0.16	0.13	0.36	0.57	0.43	0.41	0.46	0.15	0.12	0.38	0.56	0.43	0.008	0.39	0.40	0.13	0.11	0.51	0.60	0.23	0.160
<i>Math.5</i>	0.42	0.12	0.36	0.19	0.42	0.53	0.47	0.41	0.09	0.35	0.18	0.47	0.55	0.45	0.005	0.40	0.01	0.34	0.17	0.51	0.56	0.35	0.085

A goal of the revisit was to show how LVFA, MDFA, and CDFA behave for the CompFA data based on the CompFA model. Except for the unusual case where the data satisfy strong condition (15), the following results were theoretically and numerically found: The CDFA estimates of the specific variances can approximate their true values, but the MDFA estimates are contaminated by the error variances, and the LVFA estimates of the unique variances approximate the sum of the true specific and error variances. It was also shown numerically that the factor loadings can be recovered well in all three procedures. On the supposition that the data to be analyzed by FA are underlain by the CompFA model, the above results have the practical implications described in the following paragraph.

When only factor loadings are of interest, LVFA, MDFA, and CDFA are equally useful. However, if the specific variances are also interesting, CDFA is to be used. The LVFA estimate of unique variances and the MDFA estimates of specific variances must be considered as larger than the true specific variances.

However, a problem remains for showing the above implications to FA users, who are interested in only the factor loadings, not the specific variances. This problem would be dealt with, if psychometricians could enlighten the users about the importance of the specific variances. For this enlightenment, CompFA model (4) can be used as follows: By removing the specific factor part by setting  $\Theta = \rho \mathbf{O}_p$  in (4), this model and corresponding least squares function (27) are rewritten as  $\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{E}$  and  $\|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}'\|^2$ , respectively. Minimizing this function gives the formulation of principal component analysis (PCA) as approximating  $\mathbf{X}$  by reduced rank matrix  $\mathbf{F}\mathbf{\Lambda}'$  (Eckart & Young, 1936). This fact demonstrates that FA can be distinguished from PCA simply by the fact that the former has specific factor part  $\mathbf{S}\Theta$ . That is, the significance of using FA (rather than PCA) is in obtaining  $\mathbf{S}\Theta$ , which convinces the users of how it is necessary to interpret the specific variances in  $\Theta^2$  together with loading matrix  $\mathbf{\Lambda}$  in the FA solution.

Finally, we should remember that MDFA can be regarded as a procedure for a relaxed variant of the CompFA model with (12) replaced by (28). To study such a relaxed CompFA model is beyond the scope of the present study; thus, it remains for future approaches.

### Acknowledgments

Funding was provided by the Japan Society for the Promotion of Science (Grant No. (C)-18K11191). The author thanks the Associate Editor and the anonymous reviewers for their useful comments and suggestions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Appendix 1. Expectation of the Error Covariance Matrix

If the  $n$  rows of  $\mathbf{E}$  in (4) are filled with the realizations of the transpose of  $\mathbf{e}$  in (3), the expectation of  $\mathbf{C}_{XX} = n^{-1}\mathbf{E}'\mathbf{E}$  is rewritten as

$$E[\mathbf{C}_{XX}] = E\left[\frac{1}{n}\mathbf{E}'\mathbf{E}\right] = E\left[\frac{1}{n}\sum_{i=1}^n \mathbf{e}\mathbf{e}'\right] = E[\mathbf{e}\mathbf{e}'] = E[(\mathbf{e} - E[\mathbf{e}])(\mathbf{e} - E[\mathbf{e}])'] = C[\mathbf{e}, \mathbf{e}].$$

Here, we have used (5).

## Appendix 2. Proof for Theorem 3.

In the proof, we use the fact that  $\underline{\mathbf{F}}$ ,  $\underline{\mathbf{S}}$ , and  $\underline{\mathbf{E}}$  in (17) can be substituted into  $\mathbf{F}$ ,  $\mathbf{S}$ , and  $\mathbf{E}$  in (9)–(12), respectively:

$$\begin{aligned} \mathbf{1}'_n \underline{\mathbf{F}} &= \mathbf{0}'_m, \mathbf{1}'_n \underline{\mathbf{S}} = \mathbf{0}'_p, \mathbf{1}'_n \underline{\mathbf{E}} = \mathbf{0}'_p, \frac{1}{n} \underline{\mathbf{F}}' \underline{\mathbf{F}} = \mathbf{I}_m, \frac{1}{n} \underline{\mathbf{S}}' \underline{\mathbf{S}} = \mathbf{I}_p, \\ \underline{\mathbf{F}}' \underline{\mathbf{S}} &= {}_m \mathbf{O}_p, \underline{\mathbf{F}}' \underline{\mathbf{E}} = {}_m \mathbf{O}_p, \underline{\mathbf{S}}' \underline{\mathbf{E}} = {}_p \mathbf{O}_p. \end{aligned} \quad (\text{A1})$$

We also use that fact that (29) and  $\text{rank}(\underline{\mathbf{E}}) = p$  imply the nonsingularity of  $\underline{\mathbf{\Gamma}}'$  and  $\underline{\mathbf{G}} = \underline{\mathbf{E}} \underline{\mathbf{\Gamma}}'^{-1}$ . This equation and (A1) imply

$$\mathbf{1}'_n \underline{\mathbf{G}} = \mathbf{0}'_p, \underline{\mathbf{F}}' \underline{\mathbf{G}} = {}_m \mathbf{O}_p, \text{ and } \underline{\mathbf{S}}' \underline{\mathbf{G}} = {}_p \mathbf{O}_p. \quad (\text{A2})$$

At first,  $\bar{\mathbf{\Theta}}$  and  $\bar{\mathbf{S}}$  in (30) can be substituted into  $\mathbf{\Theta}$  and  $\mathbf{S}$  in (27), respectively, since  $\bar{\mathbf{\Theta}} = (\mathbf{\Theta}^2 + \mathbf{\Lambda}^2)^{1/2}$  is a  $p \times p$  diagonal matrix as is  $\mathbf{\Theta}$ , and  $\bar{\mathbf{S}} = (\underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{G}} \mathbf{\Lambda}) \bar{\mathbf{\Theta}}^{-1}$  is  $n \times p$  as is  $\mathbf{S}$ . The substitution of  $\bar{\mathbf{S}} = (\underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{G}} \mathbf{\Lambda}) \bar{\mathbf{\Theta}}^{-1}$  into  $\mathbf{S}$  in  $\mathbf{1}'_n \mathbf{S}$  and  $n^{-1} \mathbf{S}' \mathbf{S}$  leads to  $\mathbf{1}'_n \mathbf{S} = \mathbf{0}'_p$  in (9) and  $n^{-1} \mathbf{S}' \mathbf{S} = \mathbf{I}_p$  in (10) as follows:  $\mathbf{1}'_n \mathbf{S} = (\mathbf{1}'_n \underline{\mathbf{S}} \mathbf{\Theta} + \mathbf{1}'_n \underline{\mathbf{G}} \mathbf{\Lambda}) \bar{\mathbf{\Theta}}^{-1} = \mathbf{0}'_p$  and

$$\frac{1}{n} \mathbf{S}' \mathbf{S} = \frac{1}{n} \bar{\mathbf{\Theta}}^{-1} (\underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{G}} \mathbf{\Lambda})' (\underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{G}} \mathbf{\Lambda}) \bar{\mathbf{\Theta}}^{-1} = (\mathbf{\Theta}^2 + \mathbf{\Lambda}^2)^{-1/2} (\mathbf{\Theta}^2 + \mathbf{\Lambda}^2) (\mathbf{\Theta}^2 + \mathbf{\Lambda}^2)^{-1/2} = \mathbf{I}_p.$$

Here, we have used (A1), (A2), (29), (30), and  $\text{rank}(\mathbf{\Theta}) = p$ .

Through the following five steps in this paragraph, we show that  $\bar{\mathbf{S}}$  in (30) can be substituted into  $\mathbf{S}$  in (28) for  $\mathbf{\Lambda} = \text{diag}(\underline{\mathbf{\Gamma}})$  and  $\mathbf{F}$  satisfying  $\mathbf{F}' \bar{\mathbf{S}} = {}_m \mathbf{O}_p$ . First,  $\mathbf{E}$  in (28) refers to that in model (4), which shows that  $\mathbf{E}$  in (28) is rewritten as  $\mathbf{E} = \mathbf{X} - \mathbf{F} \mathbf{\Lambda}' - \mathbf{S} \mathbf{\Theta}$ . Since  $\mathbf{X}$  is specified as (17), this use in  $\mathbf{E} = \mathbf{X} - \mathbf{F} \mathbf{\Lambda}' - \mathbf{S} \mathbf{\Theta}$  allows it to be further rewritten as

$$\mathbf{E} = (\underline{\mathbf{F}} \underline{\mathbf{\Lambda}}' + \underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{E}}) - \mathbf{F} \mathbf{\Lambda}' - \mathbf{S} \mathbf{\Theta}. \quad (\text{A3})$$

Second, the substitution of  $\bar{\mathbf{S}} \bar{\mathbf{\Theta}} = \underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{G}} \mathbf{\Lambda}$  following from (30) into  $\mathbf{S} \mathbf{\Theta}$  in (A3) gives

$$\mathbf{E} = \underline{\mathbf{F}} \underline{\mathbf{\Lambda}}' + \underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{E}} - \mathbf{F} \mathbf{\Lambda}' - (\underline{\mathbf{S}} \mathbf{\Theta} + \underline{\mathbf{G}} \mathbf{\Lambda}) = \underline{\mathbf{F}} \underline{\mathbf{\Lambda}}' - \mathbf{F} \mathbf{\Lambda}' + \underline{\mathbf{E}} - \underline{\mathbf{G}} \mathbf{\Lambda}. \quad (\text{A4})$$

Third, the transpose of  $\bar{\mathbf{S}}$  in (30) post-multiplied by (A4) can be rewritten as

$$\bar{\mathbf{S}}'\mathbf{E} = \bar{\mathbf{S}}'(\mathbf{F}\mathbf{\Lambda}' + \mathbf{E} - \mathbf{G}\mathbf{\Delta}) = \bar{\mathbf{\Theta}}^{-1}(\mathbf{S}\mathbf{\Theta} + \mathbf{G}\mathbf{\Delta})'(\mathbf{F}\mathbf{\Lambda}' + \mathbf{E} - \mathbf{G}\mathbf{\Delta}), \quad (\text{A5})$$

where the first identity follows from  $\mathbf{F}'\bar{\mathbf{S}} = {}_m\mathbf{O}_p$ . Fourth, we can use (A1), (A2), and (29) to further rewrite (A5) as

$$\begin{aligned} \bar{\mathbf{S}}'\mathbf{E} &= \bar{\mathbf{\Theta}}^{-1} \mathbf{\Delta}\mathbf{G}'(\mathbf{F}\mathbf{\Lambda}' + \mathbf{E} - \mathbf{G}\mathbf{\Delta}) \\ &= \bar{\mathbf{\Theta}}^{-1}(\mathbf{\Delta}\mathbf{G}'\mathbf{E} - \mathbf{\Delta}\mathbf{G}'\mathbf{G}\mathbf{\Delta}) \\ &= \bar{\mathbf{\Theta}}^{-1}(\mathbf{\Delta}\mathbf{G}'\mathbf{G}\mathbf{\Gamma}' - \mathbf{\Delta}\mathbf{G}'\mathbf{G}\mathbf{\Delta}) \\ &= n\bar{\mathbf{\Theta}}^{-1}(\mathbf{\Delta}\mathbf{\Gamma}' - \mathbf{\Delta}^2). \end{aligned} \quad (\text{A6})$$

Finally, the substitution of (A6) into  $\mathbf{S}'\mathbf{E}$  in (28) leads to  $\text{diag}(\bar{\mathbf{S}}'\mathbf{E}) = n\text{diag}(\bar{\mathbf{\Theta}}^{-1}(\mathbf{\Delta}\mathbf{\Gamma}' - \mathbf{\Delta}^2))n\bar{\mathbf{\Theta}}^{-1}\text{diag}(\mathbf{\Delta}\mathbf{\Gamma}' - \mathbf{\Delta}^2) = n\bar{\mathbf{\Theta}}^{-1}\{\mathbf{\Delta}\text{diag}(\mathbf{\Gamma}) - \mathbf{\Delta}^2\} = {}_p\mathbf{O}_p$  for  $\mathbf{\Delta} = \text{diag}(\mathbf{\Gamma})$ .

The remaining task is to show that substituting  $\mathbf{\Theta} = \bar{\mathbf{\Theta}}$ ,  $\mathbf{S} = \bar{\mathbf{S}}$  and (17) in (27) allows this function to be rewritten as (31), for  $\mathbf{\Delta} = \text{diag}(\mathbf{\Gamma})$  and  $\mathbf{F}$  satisfying  $\mathbf{F}'\mathbf{E} = {}_m\mathbf{O}_p$ . Function (27) after the substitution is given by the use of (A4) in (27):  $f(\mathbf{F}, \mathbf{\Lambda}, \bar{\mathbf{S}}, \bar{\mathbf{\Theta}}) = n^{-1}\|\mathbf{F}\mathbf{\Lambda}' - \mathbf{F}\mathbf{\Lambda}' + \mathbf{E} - \mathbf{G}\mathbf{\Delta}\|^2$ . Further, this can be decomposed as

$$f(\mathbf{F}, \mathbf{\Lambda}, \bar{\mathbf{S}}, \bar{\mathbf{\Theta}}) = \frac{1}{n}\|\mathbf{F}\mathbf{\Lambda}' - \mathbf{F}\mathbf{\Lambda}'\|^2 + \frac{1}{n}\|\mathbf{E} - \mathbf{G}\mathbf{\Delta}\|^2, \quad (\text{A7})$$

since (A1), (A2),  $\mathbf{F}'\mathbf{E} = {}_m\mathbf{O}_p$ , and the equation  $\mathbf{F}'\mathbf{G} = {}_m\mathbf{O}_p$  following from  $\mathbf{F}'\mathbf{E} = {}_m\mathbf{O}_p$  and  $\mathbf{G} = \mathbf{E}\mathbf{\Gamma}'^{-1}$ . The last term in (A7) can be rewritten as  $n^{-1}\|\mathbf{E} - \mathbf{G}\mathbf{\Delta}\|^2 = \text{tr}\mathbf{D}_{\text{EE}} + \text{tr}\mathbf{\Delta}^2 - 2n^{-1}\text{tr}\mathbf{E}'\mathbf{G}\mathbf{\Delta} = \text{tr}\mathbf{D}_{\text{EE}} + \text{tr}\mathbf{\Delta}^2 - 2\text{tr}\mathbf{\Gamma}\mathbf{\Delta} = \text{tr}\mathbf{D}_{\text{EE}} + \text{tr}\mathbf{\Delta}^2 - 2\text{tr}\text{diag}(\mathbf{\Gamma})\mathbf{\Delta} = \text{tr}\mathbf{D}_{\text{EE}} - \text{tr}\mathbf{\Delta}^2$ , where we have used  $n^{-1}\text{tr}\mathbf{E}'\mathbf{E} = \text{tr}\mathbf{D}_{\text{EE}}$ ,  $\mathbf{\Delta} = \text{diag}(\mathbf{\Gamma})$ , (29), and this implying  $\mathbf{E}'\mathbf{G} = n\mathbf{\Gamma}$ . This completes the proof for Theorem 3.

### Appendix 3. Demonstration for Corollaries

We performed a simulation study. Its procedures are the same as those in Sects. 5.1 and 6.2, except that simulated data satisfy (15):  $\mathbf{C}_{\text{XX}}$  are restricted to (38) with its  $\mathbf{C}_{\text{EE}}$  replaced by the diagonal matrix  $\mathbf{\Phi}$  for the  $\text{C}_N$  level. For all of the resulting 1000  $\mathbf{C}_{\text{XX}}$  ( $= 500$  replications  $\times 2$  levels of  $\text{E}_L$  and  $\text{E}_H$ ), every procedure provided the solutions with the loss function value  $= 0$ ,  $\text{MAD}(\mathbf{\Lambda}) < 0.0006$ , and the (41) value  $< 0.0015$ .

### Appendix 4. Algorithms

Here, we describe the FA algorithms used in this paper. Harman and Jones' (1966) MINRES and Adachi's (2013) version of Rubin and Thayer's (1982) EM algorithm are used in LS- and ML-LVFA, respectively. Adachi's (2012, 2020) algorithm is used for MDFA. Bertisimas, Cohenhaver, and Mazumder's (2017) algorithm is used for the first MRFA step in CDFA, and Stegeman's (2016) algorithm is used for the remaining steps in CDFA. The above LS-LVFA algorithm can give an improper solution including a negative unique variance; such a solution cannot be given by the algorithms for the other procedures (Adachi, 2013, 2019).



## References

- Adachi, K. (2012). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis. *Journal of the Japanese Society of Computational Statistics*, 25, 25–38.
- Adachi, K. (2013). Factor analysis with EM algorithm never gives improper solutions when sample covariance and initial parameter matrices are proper. *Psychometrika*, 78, 380–394.
- Adachi, K. (2019). Factor analysis: Latent variable, matrix decomposition, and constrained uniqueness formulations. *WIREs Computational Statistics*, <https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1458>
- Adachi, K. (2020). *Matrix-based introduction to multivariate data analysis* (2nd ed.). Singapore: Springer.
- Adachi, K., & Trendafilov, N. T. (2018). Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, 83, 407–424.
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In: Neyman J. (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, vol. 5. (pp. 111–150). Berkeley, CA: University of California Press.
- Bartholomew, D., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). Chichester: West Sussex: Wiley.
- Bertsimas, D., Copenhaver, M. S., & Mazumder, R. (2017). Certifiably optimal low rank factor analysis. *Journal of Machine Learning Research*, 18, 1–53.
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218.
- Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: The University of Chicago Press.
- Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika*, 31, 351–369.
- Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (p. 48). Supplementary Educational Monographs No. 48: University of Chicago.
- Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. New York: Springer.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200.
- Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Pacific Grove: Sage.
- Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton: CRC Press.
- Reyment, R., & Jöreskog, K. G. (1993). *Applied factor analysis in the natural sciences*. Cambridge: Cambridge University Press.
- Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, 47, 69–76.
- Sočan, G. (2003). *The incremental value of minimum rank factor analysis*. PhD Thesis, University of Groningen.
- Spearman, C. (1904). “General Intelligence”, objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Stegeman, A. (2016). A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts. *Computational Statistics and Data Analysis*, 99, 189–203.
- ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. *Psychometrika*, 56, 309–315.
- Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. University of Chicago Press.
- Yanai, H., & Ichikawa, M. (2007). Factor analysis. In: C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 257–296). Amsterdam: Elsevier.
- Yanai, H., Shigemasa, K., Mayekawa, S., & Ichikawa, M. (1990). *Factor analysis: Its theory and methods*. Tokyo: Asakura Shoten. (in Japanese).

*Manuscript Received: 26 JAN 2021*

*Final Version Received: 26 OCT 2021*

*Published Online Date: 1 FEB 2022*