Open Research Online

The Open University's repository of research publications and other research outputs

Some Mathematical Properties of the Matrix Decomposition Solution in Factor Analysis

Journal Item

How to cite:

Adachi, Kohei and Trendafilov, Nickolay (2018). Some Mathematical Properties of the Matrix Decomposition Solution in Factor Analysis. *Psychometrika*, 83(2) pp. 407–424.

For guidance on citations see [FAQs](#).

© 2017 The Psychometric Society

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1007/s11336-017-9600-y>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

**SOME MATHEMATICAL PROPERTIES OF THE MATRIX
DECOMPOSITION SOLUTION IN FACTOR ANALYSIS**

Abstract

A new factor analysis (FA) procedure has recently been proposed which can be called matrix decomposition FA (MDFA). All FA model parameters (common and unique factors, loadings, and unique variances) are treated as fixed unknown matrices. Then, the MDFA model simply becomes a specific data matrix decomposition. The MDFA parameters are found by minimizing the discrepancy between the data and the MDFA model. Several algorithms have been developed and some properties have been discussed in the literature (notably by Stegeman, 2016), but, as a whole, MDFA has not been studied fully yet. A number of new properties are discovered in this paper, and some existing ones are derived more explicitly. The properties provided concern the uniqueness of results, covariances among common factors, unique factors, and residuals, and assessment of the degree of indeterminacy of common and unique factor scores. The properties are illustrated using a real data example.

Key words: exploratory factor analysis; model identifiability; covariances between factors and residuals; higher rank approximation; factor indeterminacy

1. Introduction

The main goal of factor analysis (FA) is to explain the variation of p observed variables by two types of mutually uncorrelated latent factors called common and unique. The common factors, whose number m is supposed to be much less than that of observed variables ($m < p$), serve for explaining the variation of all p variables. On the other hand, each unique factor explains specifically the variation of the corresponding variable that remains unaccounted for by the common factors.

The classic FA model can be expressed as

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

for p -variate vector $\mathbf{x}$ with its expectation $E[\mathbf{x}]$ being the $p \times 1$ zero vector $\mathbf{0}_p$. Here, $\mathbf{f}$ ($m \times 1$) and $\mathbf{u}$ ($p \times 1$) are the latent random vectors containing m common factors and p unique factors, respectively (e.g., Harman, 1976; Mulaik, 2010). On the other hand, $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are the unknown fixed parameter matrices: $\mathbf{\Lambda}$ is the $p \times m$ matrix of factor loadings describing the relationships of variables to common factors, and $\mathbf{\Psi}$ is the $p \times p$ diagonal matrix. The squares of the diagonal elements of $\mathbf{\Psi}$ are called unique variances, as they describe to what extents the variations in the variables are explained by the corresponding unique factors. The expectations for $\mathbf{f}$ and $\mathbf{u}$ are assumed to satisfy

$$E[\mathbf{f}] = \mathbf{0}_m, E[\mathbf{u}] = \mathbf{0}_p, E[\mathbf{ff}'] = \mathbf{I}_m, E[\mathbf{uu}'] = \mathbf{I}_p, \text{ and } E[\mathbf{fu}'] = {}_m\mathbf{O}_p \quad (2)$$

with $\mathbf{I}_m$ the $m \times m$ identity matrix and ${}_m\mathbf{O}_p$ the $m \times p$ matrix of zeros. The FA model (1) with (2) leads to the inter-variable covariance matrix for $\mathbf{\Lambda}\mathbf{f} + \mathbf{\Psi}\mathbf{u}$ being $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2$. It is fitted to the empirical covariance matrix $\mathbf{S}_{XX} = n^{-1}\mathbf{X}'\mathbf{X}$ to obtain $\mathbf{\Lambda}$ and $\mathbf{\Psi}^2$. Here, $\mathbf{X}$ is an n -observations $\times$ p -variables column-centered data matrix.

Recently, a very different formulation of FA has been proposed (Sočan, 2003; de Leeuw, 2004). In this formulation, the common and unique factors are also treated as fixed unknown parameters: the $n \times m$ and $n \times p$ matrices $\mathbf{F}$ and $\mathbf{U}$ contain the common and unique factor scores, respectively. The FA model part $\mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi}$ is a fixed matrix and cannot be equalized directly to $\mathbf{X}$. Thus, using $n \times p$ error matrix $\mathbf{E}$, FA is modeled as

$$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi} + \mathbf{E} = \mathbf{Z}\mathbf{B}' + \mathbf{E}. \quad (3)$$

Here, $\mathbf{B} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ is a $p \times (m + p)$ block matrix, and $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ is an $n \times (m + p)$ matrix, which is constrained by the matrix version of (2):

$$\mathbf{1}_n' \mathbf{Z} = \mathbf{0}_{m+p}, \quad (4)$$

$$\frac{1}{n} \mathbf{F}'\mathbf{F} = \mathbf{I}_m, \quad \frac{1}{n} \mathbf{U}'\mathbf{U} = \mathbf{I}_p, \quad \text{and } \mathbf{F}'\mathbf{U} = {}_m\mathbf{O}_p, \quad \text{i.e.,} \quad \frac{1}{n} \mathbf{Z}'\mathbf{Z} = \mathbf{I}_{m+p}, \quad (5)$$

with $\mathbf{1}_n$ the $n \times 1$ vector of ones. The new formulation is thus to minimize the sum of the squared errors

$$f(\mathbf{Z}, \mathbf{B}) = \|\mathbf{E}\|^2 = \|\mathbf{X} - (\mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi})\|^2 = \|\mathbf{X} - \mathbf{Z}\mathbf{B}'\|^2 \quad (6)$$

over $\mathbf{Z}$ and $\mathbf{B}$ subject to (4) and (5). We call this formulation matrix decomposition FA (MDFA), as FA has been regarded as a matrix factorization (or decomposition) problem.

To the best of our knowledge, Professor Henk A. L. Kiers in the University of Groningen firstly proposed MDFA as found in Sočan's (2003, pp. 17-19) dissertation. Independently of it, de Leeuw (2004) has presented MDFA. Then, Unkel and Trendafilov (2010) reconsidered MDFA jointly with other parameter estimation methods in FA. In the above papers, an MDFA algorithm has been described that needs the original data matrix $\mathbf{X}$. On the other hand, Adachi (2012, 2015) has presented another algorithm in which the optimal $\mathbf{B} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ can be obtained from empirical covariance matrix $\mathbf{S}_{XX}$ even if the original data matrix is not available. In Adachi (2012, 2015) and Stegeman (2016), it is empirically found that MDFA solutions are very similar to the classic FA ones.

Two constrained versions of MDFA have been presented. One of them was proposed together with the original MDFA by Henk A. L. Kiers (Sočan, 2003, pp. 19-20). In this version, constraint $n^{-1}\mathbf{X}'\mathbf{U} = \mathbf{D}$ is added to (4) and (5) with $\mathbf{D}$ a diagonal matrix. That constraint is further strengthened as $n^{-1}\mathbf{X}'\mathbf{U} = \mathbf{\Psi}$ in the other version proposed by Stegeman (2016). Although those constrained versions are mentioned again in a later section, this paper focuses on MDFA without the above additional constraints. Various results in Stegeman (2016), in which MDFA is also considered together with the above version, have been **re-derived** in the present paper, or used in a slightly different way.

Though the algorithms for MDFA have been developed and assessed empirically, mathematical properties of the MDFA solution have not been studied fully. To elucidate them is the purpose of this paper. The properties to be studied are summarized as follows:

- [1] Identifiability of the model part $\mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi}$ in (3)
- [2] Properties of the covariances for the resulting residuals in $\mathbf{E}$
- [3] How the common and unique factor scores in $\mathbf{Z}$ are undetermined.

They are discussed in Sections 3 to 5, respectively, after presenting the preliminary results in Section 2. In those sections, a small data set in Table 1, a part of Tanaka and Tarumi's data (1995, p. 125), is used for illustrating mathematical results.

Although Stegeman (2016, Sections 3 and 4.2, Appendix 2) also describes properties of the MDFA solution, some results have been mentioned without giving a full derivation. A

Table 1. Test score data, where 20 observations $\times$ 5 variables scores have been chosen from the original 50×6 data matrix (Tanaka & Tarumi, 1995)

	Japanese	English	Social*	Mathematics	Science
1	64	65	83	69	70
2	54	56	53	40	32
3	80	68	75	74	84
4	71	65	40	41	68
5	63	61	60	56	80
6	47	62	33	57	87
7	42	53	50	38	23
8	54	17	46	58	58
9	57	48	59	26	17
10	54	72	58	55	30
11	67	82	52	50	44
12	71	82	54	67	28
13	53	67	74	75	53
14	90	96	63	87	100
15	71	69	74	76	42
16	61	100	92	53	58
17	61	69	48	63	71
18	87	84	64	65	53
19	77	75	78	37	44
20	57	27	41	54	30

*Social Studies

complete derivation of the results will be given here. As shown above in [1] and [3], this paper shows how $\mathbf{FA}' + \mathbf{U}\Psi = \mathbf{ZB}'$ is identified after discussion on its uniqueness. It also treats the indeterminacy of the matrix $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ containing the unique factors scores together with the common ones.

In this paper, covariance matrices are expressed by $\mathbf{S}$ with the related matrices attached as subscripts. For example, $\mathbf{S}_{XX} = n^{-1}\mathbf{X}'\mathbf{X}$ has already been defined, and $\mathbf{S}_{XZ} = n^{-1}\mathbf{X}'\mathbf{Z} = [\mathbf{S}_{XF}, \mathbf{S}_{XU}]$ stands for the p -variables $\times$ $(m+p)$ -factors covariance matrix, whose blocks $\mathbf{S}_{XF} = n^{-1}\mathbf{X}'\mathbf{F}$ and $\mathbf{S}_{XU} = n^{-1}\mathbf{X}'\mathbf{U}$ are the covariance matrices of variables to common and unique factors, respectively. Throughout the paper, we assume $n > (p+m)$, $p > m$, and

$$\text{rank}(\mathbf{XB}) = p \quad (7)$$

which implies $\text{rank}(\mathbf{X}) = p$, $\text{rank}(\mathbf{B}) = p$ and $\mathbf{BB}^+ = \mathbf{I}_p$ with $\mathbf{B}^+$ the Moore-Penrose inverse of $\mathbf{B}$. In this case (of $\mathbf{B}$ being a full row-rank matrix) we simply have $\mathbf{B}^+ = \mathbf{B}'(\mathbf{BB}')^{-1}$.

2. Preliminary Results

Two types of MDFA algorithms are available: working with either raw data matrix, or covariance/correlation matrix, whichever is available. We briefly recall them here but in a different way, which will facilitate later our theory.

We start with the algorithm working with available data matrix. Let us define the columns and rows of matrices as $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_p]$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p]$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$, $\mathbf{B} =$

$\begin{bmatrix} \mathbf{b}'_1 \\ \mathbf{M} \\ \mathbf{b}'_p \end{bmatrix}$, and $\mathbf{\Lambda} = \begin{bmatrix} \lambda'_1 \\ \mathbf{M} \\ \lambda'_p \end{bmatrix}$. Then, loss function (6) is rewritten as $f(\mathbf{Z}, \mathbf{B}) = \sum_{j=1}^p f_j(\mathbf{Z}, \mathbf{b}_j)$, by

defining the sum of squared errors for variable j as

$$f_j(\mathbf{Z}, \mathbf{b}_j) = \|\mathbf{e}_j\|^2 = \|\mathbf{x}_j - \mathbf{F}\lambda_j - \psi_j \mathbf{u}_j\|^2 = \|\mathbf{x}_j - \mathbf{Z}\mathbf{b}_j\|^2 \quad (8)$$

with ψ_j the j th diagonal element of $\mathbf{\Psi}$.

Theorem 2.1. Let $\mathbf{s}_j^{\text{XZ}'} (1 \times (p+m))$ denote the j th row of $\mathbf{S}_{\text{XZ}} = [\mathbf{s}_1^{\text{XZ}}, \dots, \mathbf{s}_p^{\text{XZ}}]' (p \times (p+m))$. Under the constraint (5), the function (8) can be rewritten as

$$f_j(\mathbf{Z}, \mathbf{b}_j) = \|\mathbf{e}_j\|^2 = \|\mathbf{x}_j - \mathbf{Z}\mathbf{s}_j^{\text{XZ}}\|^2 + n\|\mathbf{s}_j^{\text{XZ}} - \mathbf{b}_j\|^2. \quad (9)$$

Proof. (8) can be rewritten as $\|\mathbf{x}_j - \mathbf{Z}\mathbf{s}_j^{\text{XZ}} + \mathbf{Z}\mathbf{s}_j^{\text{XZ}} - \mathbf{Z}\mathbf{b}_j\|^2 = \|\mathbf{x}_j - \mathbf{Z}\mathbf{s}_j^{\text{XZ}}\|^2 + h + 2c$ with $h = \|\mathbf{Z}\mathbf{s}_j^{\text{XZ}} - \mathbf{Z}\mathbf{b}_j\|^2$ and $c = (\mathbf{x}_j - \mathbf{Z}\mathbf{s}_j^{\text{XZ}})'(\mathbf{Z}\mathbf{s}_j^{\text{XZ}} - \mathbf{Z}\mathbf{b}_j)$. Here, we can expand c and use (5) to have $c = \mathbf{x}_j' \mathbf{Z} \mathbf{s}_j^{\text{XZ}} - \mathbf{x}_j' \mathbf{Z} \mathbf{b}_j - \mathbf{s}_j^{\text{XZ}'} \mathbf{Z}' \mathbf{Z} \mathbf{s}_j^{\text{XZ}} + \mathbf{s}_j^{\text{XZ}'} \mathbf{Z}' \mathbf{Z} \mathbf{b}_j = n \mathbf{s}_j^{\text{XZ}'} \mathbf{s}_j^{\text{XZ}} - n \mathbf{s}_j^{\text{XZ}'} \mathbf{b}_j - n \mathbf{s}_j^{\text{XZ}'} \mathbf{s}_j^{\text{XZ}} + n \mathbf{s}_j^{\text{XZ}'} \mathbf{b}_j = 0$. Further, (5) allows us to rewrite h as $n\|\mathbf{s}_j^{\text{XZ}} - \mathbf{b}_j\|^2$. $\square$

The summation of both sides of (9) over variables $j = 1, \dots, p$ leads to:

$$f(\mathbf{Z}, \mathbf{B}) = \|\mathbf{E}\|^2 = \|\mathbf{X} - \mathbf{Z}\mathbf{S}_{\text{XZ}}'\|^2 + n\|\mathbf{S}_{\text{XZ}} - \mathbf{B}\|^2. \quad (10)$$

It allows us to find that, for given $\mathbf{Z}$, the optimal $\mathbf{B} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ is minimizing $\|\mathbf{S}_{\text{XZ}} - \mathbf{B}\|^2$. Here, it should be kept in mind, that $\mathbf{\Psi} (p \times p)$ is diagonal, and $\mathbf{S}_{\text{XZ}} = [\mathbf{S}_{\text{XF}}, \mathbf{S}_{\text{XU}}]$, with $\mathbf{S}_{\text{XF}}$ being the p (variables) $\times m$ (common factors) covariance matrix and $\mathbf{S}_{\text{XU}}$ - the p (variables) $\times p$ (unique factors) one. This shows that the optimal $\mathbf{B} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ is given by

$$\hat{\mathbf{B}} = [\mathbf{S}_{\text{XF}}, \text{diag}(\mathbf{S}_{\text{XU}})], \text{ i.e., } \hat{\mathbf{\Lambda}} = \mathbf{S}_{\text{XF}} \text{ and } \hat{\mathbf{\Psi}} = \text{diag}(\mathbf{S}_{\text{XU}}), \quad (11)$$

given $\mathbf{Z}$, with $\text{diag}(\mathbf{S}_{\text{XU}})$ denoting the diagonal matrix whose diagonal elements are those of

$\mathbf{S}_{\text{XU}}$ (de Leeuw, 2004; Sočan, 2003). The resulting $\hat{\mathbf{B}} = \begin{bmatrix} \hat{\mathbf{b}}'_1 \\ \mathbf{M} \\ \hat{\mathbf{b}}'_p \end{bmatrix}$ is found to satisfy

$$\mathbf{S}_{\text{XZ}} \hat{\mathbf{B}}' = [\mathbf{S}_{\text{XF}}, \mathbf{S}_{\text{XU}}] \begin{bmatrix} \hat{\mathbf{\Lambda}}' \\ \hat{\mathbf{\Psi}} \end{bmatrix} = \hat{\mathbf{\Lambda}} \hat{\mathbf{\Lambda}}' + \mathbf{S}_{\text{XU}} \hat{\mathbf{\Psi}}, \text{ whose } j\text{th diagonal element is expressed as}$$

$$\mathbf{s}_j^{\text{XZ}'} \hat{\mathbf{b}}_j = \hat{\lambda}'_j \hat{\lambda}_j + \hat{\psi}_j^2. \quad (12)$$

with $\hat{\lambda}'_j$ the j th row of $\hat{\mathbf{\Lambda}}$ and $\hat{\psi}_j$ the j th diagonal element of $\hat{\mathbf{\Psi}}$.

Using (5), we can rewrite (6) as $f(\mathbf{Z}, \mathbf{B}) = \|\mathbf{X}\|^2 + n\|\mathbf{B}\|^2 - 2\text{tr}(\mathbf{X}\mathbf{B})'\mathbf{Z}$, which shows that

the minimization of (6) over $\mathbf{Z}$ for a given $\mathbf{B}$ amounts to maximizing $\text{tr}(\mathbf{XB})'\mathbf{Z}$ subject to (5). The maximization is attained for $\mathbf{Z}$ equaling

$$\hat{\mathbf{Z}} = n^{1/2} \tilde{\mathbf{K}} \tilde{\mathbf{L}}' = n^{1/2} \mathbf{K} \mathbf{L}' + n^{1/2} \mathbf{K}_\perp \mathbf{L}_\perp' = \mathbf{XBL} \Delta^{-1} \mathbf{L}' + n^{1/2} \mathbf{K}_\perp \mathbf{L}_\perp' \quad (13)$$

as found in ten Berge (1983, Theorem 2; 1993, Theorem in pp. 28 - 29). Also see in Stegeman (2016, p. 194). Here, $\tilde{\mathbf{K}} = [\mathbf{K}, \mathbf{K}_\perp]$ ($n \times (p + m)$) and $\tilde{\mathbf{L}} = [\mathbf{L}, \mathbf{L}_\perp]$ ($(p + m) \times (p + m)$) satisfy $\tilde{\mathbf{K}}'\tilde{\mathbf{K}} = \tilde{\mathbf{L}}'\tilde{\mathbf{L}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}' = \mathbf{I}_{p+m}$, and $\mathbf{K}$ ($n \times p$), $\mathbf{L}$ ($(p + m) \times p$) and Δ ($p \times p$) are given through the singular value decomposition (SVD) of $n^{-1/2} \mathbf{XB}$:

$$n^{-1/2} \mathbf{XB} = \mathbf{K} \Delta \mathbf{L}' , \quad (14)$$

with Δ diagonal, $\mathbf{K}_\perp$ of order $n \times m$, and $\mathbf{L}_\perp$ of $(p + m) \times m$. The identity $n^{1/2} \mathbf{K} \mathbf{L}' = \mathbf{XBL} \Delta^{-1} \mathbf{L}'$ in (13) directly follows from (14). The $\mathbf{Z}$ given by (13) is also found to satisfy (4), since of $\mathbf{1}_n' \mathbf{K} = \mathbf{0}_p'$ and $\mathbf{1}_n' \mathbf{K}_\perp = \mathbf{0}_m'$, as shown in Adachi (2012, Appendix 1). In (13) and (14), we can find that $\mathbf{K} \mathbf{L}'$ is uniquely determined, however $\mathbf{K}_\perp \mathbf{L}_\perp'$ and $\hat{\mathbf{Z}}$ are not unique. In spite of this non-uniqueness of factor scores, the alternate iteration of (11) and (13) for $\mathbf{X}$ provides the optimal $\mathbf{B}$ (de Leeuw, 2004; Sočan, 2003).

Now, assume that the data matrix $\mathbf{X}$ is not available. Adachi (2012) has developed an algorithm, which finds the optimal $\mathbf{B} = [\mathbf{A}, \mathbf{\Psi}]$ if only empirical covariance matrix $\mathbf{S}_{XX}$ is available. It is based on (11) and the following three results. First, SVD (14) leads to the eigenvalue decomposition (EVD) of $\mathbf{B}' \mathbf{S}_{XX} \mathbf{B}$:

$$\mathbf{B}' \mathbf{S}_{XX} \mathbf{B} = \mathbf{L} \Delta^2 \mathbf{L}' . \quad (15)$$

Second, the covariance matrix between the variables and the optimal factor scores is given by

$$\mathbf{S}_{X\hat{\mathbf{Z}}} = \mathbf{B}'^+ \mathbf{L} \Delta \mathbf{L}' . \quad (16)$$

This follows from the use of $\mathbf{B} \mathbf{B}'^+ = \mathbf{I}_p$ in the equation $n^{-1/2} \mathbf{B}' \mathbf{X}' \hat{\mathbf{Z}} = n^{1/2} \mathbf{L} \Delta \mathbf{L}'$ derived by the transpose of (14) post-multiplied by (13). Third, the loss function (6) value is expressed as

$$f(\hat{\mathbf{Z}}, \hat{\mathbf{B}}) = \|\hat{\mathbf{E}}\|^2 = n \text{tr}(\mathbf{S}_{XX} - \hat{\mathbf{A}} \hat{\mathbf{A}}' - \hat{\mathbf{\Psi}}^2) . \quad (17)$$

This follows from the use of (5) and (11) in (6), which expands as $\text{tr} \mathbf{X}' \mathbf{X} + n \text{tr} \mathbf{A} \mathbf{A}' + n \text{tr} \mathbf{\Psi}^2 - 2 \text{tr} \mathbf{X}' \mathbf{F} \mathbf{A}' - 2 \text{tr} \mathbf{X}' \mathbf{U} \mathbf{\Psi}$. Equations (15), (16), and (17) show that we can perform the update (11) and assess the loss function value for checking convergence, only if $\mathbf{S}_{XX}$ is available. Thus, Adachi's (2012) MDFA algorithm can be described as follows:

- Step 1. Initialize $\mathbf{B} = [\mathbf{A}, \mathbf{\Psi}]$
- Step 2. Perform EVD (15) to obtain $\mathbf{S}_{X\hat{\mathbf{Z}}}$ with (16)
- Step 3. Update $\mathbf{B}$ with (11)
- Step 4. Finish if convergence is reached; otherwise, back to Step 2.

Table 2. Loadings and unique variances obtained with MDFA, LS-FA, and ML-FA for the standard scores of the data in Table 1, with the residuals (Res) in MDFA

	MDFA				LS-FA			ML-FA		
	Λ	ψ^2	Res		Λ	ψ^2		Λ	ψ^2	
Japa	0.38	0.60	0.50	0.001	0.38	0.60	0.50	0.37	0.61	0.50
Eng	0.21	0.76	0.37	0.002	0.19	0.77	0.37	0.21	0.76	0.38
Soc	0.03	0.65	0.58	0.002	0.03	0.64	0.59	0.02	0.65	0.58
Math	0.59	0.34	0.53	0.003	0.58	0.35	0.54	0.58	0.34	0.55
Sci	0.89	0.10	0.19	0.001	0.91	0.10	0.17	0.90	0.11	0.17

This algorithm is equivalent to the one using $\mathbf{X}$ (de Leeuw, 2004; Sočan, 2003) in a sense that it finds the same solution for $\mathbf{B} = [\mathbf{A}, \mathbf{\Psi}]$. The former algorithm is used for the numerical examples in the paper; see Adachi (2012) for details.

Table 2 shows the MDFA solution for the standardized version of the raw data in Table 1, together with the solutions of the least squares FA (LS-FA) (Harman & Jones, 1966) and the maximum likelihood FA (ML-FA) (Rubin & Thayer, 1982) which are based on the classic FA model (1) with (2). The loading matrices in the table have been rotated by the varimax method. We can find that the loadings and unique variances are almost equivalent among the three FA solutions. Only the panel for MDFA includes a column for the residuals, which is given by

$$\frac{1}{n} \|\hat{\mathbf{e}}_j\|^2 = s_{jj}^{xx} - \hat{\mathbf{b}}_j' \hat{\mathbf{b}}_j = s_{jj}^{xx} - \hat{\lambda}_j' \hat{\lambda}_j - \hat{\psi}_j^2. \quad (18)$$

It follows after substitution of (11) and (12) in (8), which expands as $n(s_{jj}^{xx} - 2\mathbf{s}_j^{xz'}\mathbf{b}_j + \mathbf{b}_j'\mathbf{b}_j)$.

The residuals are not presented in the panels for the classic LS- and ML-FA, since their estimates of $\mathbf{\Psi}^2$ given by

$$\mathbf{\Psi}_C^2 = \text{diag}(\mathbf{S}_{xx} - \mathbf{\Lambda}_C \mathbf{\Lambda}_C') \quad (19)$$

and the residual corresponding to (18) is always zero, with $\mathbf{\Lambda}_C$ the estimate of $\mathbf{\Lambda}$ in the classic LS- or ML-FA (Harman, 1976; Mulaik, 2010). The residuals for the MDFA solution are found to be close to zero, which shows that almost all of the variations in $\mathbf{X}$ are explained by the model part $\mathbf{ZB}'$.

3. Identifiability of Model Part

In this section, we discuss the uniqueness of the model part $\mathbf{ZB}' = \mathbf{FA}' + \mathbf{U\Psi}$ in (3) to show its identifiability. This has also been mentioned by Stegeman (2016, Appendix 2), but without a derivation. We also demonstrate numerically the identification of the model part and the resulting residual values in $\hat{\mathbf{E}}$.

The following lemma is useful to show identifiability:

Lemma 3.1. The covariance matrix (16) between variables and factors can be rewritten as

$$\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} = \mathbf{S}_{\mathbf{X}\mathbf{X}}\mathbf{B}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}'. \quad (20)$$

Proof. Left multiplication of (13) by $n^{-1}\mathbf{X}'$ leads to $\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} = \mathbf{S}_{\mathbf{X}\mathbf{X}}\mathbf{B}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}' + n^{1/2}\mathbf{X}'\mathbf{K}_{\perp}\mathbf{L}_{\perp}'$. To prove (20), we need to show that $\mathbf{X}'\mathbf{K}_{\perp}\mathbf{L}_{\perp}'$ vanishes. To see this, we note that (14) implies:

$$\mathbf{X} = n^{1/2}\mathbf{K}\mathbf{\Delta}\mathbf{L}'\mathbf{B}^+, \quad (21)$$

and that by construction $\mathbf{K}'\mathbf{K}_{\perp} = {}_p\mathbf{O}_m$. $\square$

It also shows that (20) may be substituted for (16) in Step 2 (Section 2).

Apparently, the loading matrix $\mathbf{L}$ in $\mathbf{B}$ has rotational indeterminacy as seen by $\mathbf{F}\mathbf{T}\mathbf{T}'\mathbf{L}' = \mathbf{F}\mathbf{L}'$ for any $m \times m$ orthonormal matrix $\mathbf{T}$. In order to consider the uniqueness of $\mathbf{B}$ regardless of the rotational indeterminacy, we assume $\mathbf{Z}\mathbf{B}' = \tilde{\mathbf{Z}}\tilde{\mathbf{B}}'$ for some $\tilde{\mathbf{B}} = [\tilde{\mathbf{L}}, \tilde{\mathbf{\Psi}}]$ differing from $\mathbf{B}$. Then, $(\mathbf{Z}\mathbf{B}')'(\mathbf{Z}\mathbf{B}') = (\tilde{\mathbf{Z}}\tilde{\mathbf{B}}')'(\tilde{\mathbf{Z}}\tilde{\mathbf{B}}')$ in which we can use (5) to have $\mathbf{B}\mathbf{B}' = \tilde{\mathbf{B}}\tilde{\mathbf{B}}'$, or equivalently,

$$\mathbf{L}\mathbf{L}' + \mathbf{\Psi}^2 = \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}}^2. \quad (22)$$

For this equation, the following assertion has already been proven (Anderson & Rubin, 1956; Ihara & Kano, 1986): A sufficient condition for (22) to hold true and $[\mathbf{L}, \mathbf{\Psi}] = [\tilde{\mathbf{L}}, \tilde{\mathbf{\Psi}}]$, is that $\mathbf{L}$ should possess two disjoint submatrices of rank m , after deleting any one of its rows. This is also a sufficient condition for the uniqueness of $\mathbf{B}$ in MDFA.

Theorem 3.1. If the solution $\hat{\mathbf{B}}$ is unique, then the resulting model part $\hat{\mathbf{Z}}\hat{\mathbf{B}}' = \hat{\mathbf{F}}\hat{\mathbf{L}}' + \hat{\mathbf{U}}\hat{\mathbf{\Psi}}$ is also unique, which is given by

$$\hat{\mathbf{Z}}\hat{\mathbf{B}}' = \sqrt{n}\mathbf{K}\mathbf{L}'\hat{\mathbf{B}}' = \mathbf{X}\hat{\mathbf{B}}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}'\hat{\mathbf{B}}' = \mathbf{X}\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}\hat{\mathbf{B}}' = \mathbf{X}\hat{\mathbf{B}}\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}'\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}. \quad (23)$$

Proof. Pre-multiplying both sides of (14) by $n^{1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ leads to $\hat{\mathbf{B}} = n^{1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{\Delta}\mathbf{L}'$, which implies $\hat{\mathbf{B}}\mathbf{L}_{\perp} = {}_p\mathbf{O}_m$ since of $\mathbf{L}'\mathbf{L}_{\perp} = {}_p\mathbf{O}_m$ following from $\tilde{\mathbf{L}}'\tilde{\mathbf{L}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}' = \mathbf{I}_{p+m}$ with $\tilde{\mathbf{L}} = [\mathbf{L}, \mathbf{L}_{\perp}]$. The equalities $\hat{\mathbf{B}}\mathbf{L}_{\perp} = {}_p\mathbf{O}_m$ and (13) lead to $\hat{\mathbf{Z}}\hat{\mathbf{B}}' = (n^{1/2}\mathbf{K}\mathbf{L}' + n^{1/2}\mathbf{K}_{\perp}\mathbf{L}_{\perp}')\hat{\mathbf{B}}' = n^{1/2}\mathbf{K}\mathbf{L}'\hat{\mathbf{B}}'$. It is uniquely determined, since the matrix $\mathbf{K}\mathbf{L}'$ given by the SVD in (14) is unique. The remaining three terms are derived from $n^{1/2}\mathbf{K}\mathbf{L}'\hat{\mathbf{B}}'$ as follows. The first one $\mathbf{X}\hat{\mathbf{B}}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}'\hat{\mathbf{B}}'$ follows after substitution of $n^{1/2}\mathbf{K}\mathbf{L}' = \mathbf{X}\hat{\mathbf{B}}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}'$. The next equality $\mathbf{X}\hat{\mathbf{B}}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}'\hat{\mathbf{B}}' = \mathbf{X}\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}\hat{\mathbf{B}}'$ follows from (20) which leads to $\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}\hat{\mathbf{B}}' = \hat{\mathbf{B}}\mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}_{\perp}'\hat{\mathbf{B}}'$. As this matrix is symmetric, $\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}\hat{\mathbf{B}}' = \hat{\mathbf{B}}\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}'\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}$, and thus, the last identity holds. $\square$

Table 3. Model part and residuals resulting in MDFA.

Obs.	$\hat{\mathbf{ZB}}' : \text{Model Part}$					$\hat{\mathbf{E}} : \text{Residuals}$				
	Japa	Eng	Soc	Math	Sci	Japa	Eng	Soc	Math	Sci
1	0.044	-0.038	1.508	0.715	0.743	-0.048	-0.008	0.018	0.076	-0.029
2	-0.805	-0.525	-0.439	-1.118	-0.946	-0.003	0.017	-0.012	-0.010	0.005
3	1.315	0.094	1.017	1.048	1.362	-0.033	0.014	-0.018	0.074	-0.037
4	0.554	-0.125	-1.229	-1.025	0.621	0.005	0.079	-0.079	-0.038	0.007
5	-0.055	-0.289	0.036	-0.090	1.159	-0.030	0.038	-0.026	0.020	-0.009
6	-1.402	-0.193	-1.792	0.096	1.402	0.032	-0.007	0.022	-0.099	0.053
7	-1.779	-0.652	-0.671	-1.225	-1.355	0.007	-0.011	0.021	-0.036	0.022
8	-0.782	-2.496	-0.916	-0.020	0.232	-0.026	-0.014	0.003	0.083	-0.041
9	-0.530	-0.982	0.003	-2.096	-1.570	-0.036	0.063	-0.059	0.040	-0.024
10	-0.837	0.357	-0.167	-0.087	-1.056	0.030	-0.044	0.046	-0.049	0.028
11	0.204	0.816	-0.506	-0.399	-0.447	0.033	0.010	-0.012	-0.067	0.028
12	0.491	0.893	-0.431	0.723	-1.143	0.068	-0.067	0.046	-0.064	0.028
13	-0.888	0.137	0.847	1.175	-0.033	0.000	-0.080	0.086	0.014	0.007
14	2.054	1.558	0.207	2.013	2.012	0.032	-0.013	0.001	-0.030	0.010
15	0.549	0.225	0.883	1.215	-0.489	0.010	-0.066	0.050	0.039	-0.016
16	-0.203	1.720	2.117	-0.266	0.180	-0.042	0.030	0.002	-0.002	0.012
17	-0.271	0.175	-0.798	0.445	0.732	0.026	-0.016	0.017	-0.051	0.026
18	1.824	0.920	0.302	0.526	-0.016	0.020	0.009	-0.028	0.000	-0.010
19	1.097	0.368	1.288	-1.389	-0.381	-0.057	0.100	-0.092	0.062	-0.037
20	-0.580	-1.961	-1.256	-0.239	-1.006	0.013	-0.036	0.013	0.037	-0.023

Theorem 3.1 also implies the identifiability of the residuals in $\hat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{ZB}}'$. The model part (23) and the residuals are illustrated in Table 3. They present the values corresponding to the MDFA solution in Table 2. The sum of $\hat{\mathbf{ZB}}'$ and $\hat{\mathbf{E}}$ in Table 3 gives the standard scores for the raw data in Table 1. Comparing $\hat{\mathbf{ZB}}'$ and $\hat{\mathbf{E}}$ in Table 3, we see that the residuals are close to zero. This implies that $\hat{\mathbf{ZB}}'$ well approximates $\mathbf{X}$. The total sum of squared residuals divided by n is $n^{-1}\|\hat{\mathbf{E}}\|^2 = 0.0017$, while $\text{tr}\mathbf{S}_{\text{XX}} = n^{-1}\|\mathbf{X}\|^2 = 5$. Thus, we have $n^{-1}\|\hat{\mathbf{E}}\|^2/\text{tr}\mathbf{S}_{\text{XX}} = 0.0003$, which shows that almost all variation in $\mathbf{X}$ is explained by the model part $\mathbf{ZB}' = \mathbf{FA}' + \mathbf{U\Psi}$ in (3).

In the classic FA model (1), the normality assumptions for factors $\mathbf{f}$ and $\mathbf{u}$ lead to the χ^2 -based statistics used for assessing whether the resulting solution is acceptable (Anderson & Rubin, 1956; Bartlett, 1950). Such statistics cannot be derived in a similar manner in MDFA with factor scores treated as fixed parameters. However, as found in the last paragraph, the proportion $\eta = n^{-1}\|\hat{\mathbf{E}}\|^2/\text{tr}\mathbf{S}_{\text{XX}}$ can be used as a standardized badness-of-fit index for the MDFA solution, with $0 \leq \eta \leq 1$. Equivalently, $1-\eta = \text{tr}(\hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}^2)/\text{tr}\mathbf{S}_{\text{XX}}$ serves as a goodness-of-fit (GOF) index with $0 \leq 1-\eta \leq 1$, where we have used (17).

The matrix $\hat{\mathbf{E}}$ in Table 3 demonstrates the strength of MDFA: the residual standing for the badness of fit of the model can be obtained for each observation. This is not possible in the classic FA. In place of $\hat{\mathbf{E}}$, the discrepancy of the empirical covariance matrix from the model-based one,

$$\mathbf{S}_{\text{res}} = \mathbf{S}_{\text{XX}} - (\mathbf{\Lambda}_c\mathbf{\Lambda}_c' + \mathbf{\Psi}_c^2), \quad (24)$$

is obtained as a residual matrix in the classic FA, with its estimates of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ expressed as $\mathbf{\Lambda}_c$ and $\mathbf{\Psi}_c$. Table 4 shows the residual matrices (24) obtained for the data in Table 1 together

Table 4. Residuals for covariances

		Japa	Eng	Soc	Math	Sci
MDFA	Japa	0.001				
	Eng	0.017	0.002		Sym.	
	Soc	-0.035	0.005	0.002		
	Math	0.020	-0.051	0.052	0.003	
	Sci	-0.012	0.016	-0.014	0.001	0.001
LS-FA	Japa	0.000				
	Eng	0.020	0.000		Sym.	
	Soc	-0.031	0.006	0.000		
	Math	0.018	-0.052	0.045	0.000	
	Sci	-0.015	0.025	-0.017	0.002	0.000
ML-FA	Japa	0.000				
	Eng	0.015	0.000		Sym.	
	Soc	-0.037	0.007	0.000		
	Math	0.028	-0.050	0.058	0.000	
	Sci	-0.007	0.008	-0.007	0.001	0.000

with the corresponding one for MFDA. In the table, we can find that the values of the off-diagonal elements are similar between MDFA and the classic FA. But, a crucial difference is in the diagonal elements. They are always zeros in the classic FA thanks to (19), while the diagonal elements for MDFA are given by (18) and are ≥ 0 . This implies that in the classic FA, the badness of fit cannot be assessed for variables, but it can be assessed between variables. On the other hand, in MDFA, this is possible for both variables and between variables.

From the above discussions we conclude that MDFA is more useful than the classic FA if one wishes to know the badness of fit of the FA model for each variable and for each observation. If a variable with substantially large residual is found in a MDFA solution, it is reasonable to consider performing FA again with that variable removed. Further, the residual for each observation can be used for detecting an aberrant observation. For example, if a too great positive or small negative residual is detected in a solution for test data analysis as in Table 3, this can indicate that the corresponding examinee's score is exceeding or below her/his ability for special reasons.

Equation (19) implies $\text{tr}\Psi_C^2 = \text{tr}(\mathbf{S}_{XX} - \Lambda_C\Lambda_C')$, while the MDFA loss function value (17) ≥ 0 leads to $\text{tr}\hat{\Psi}^2 \leq \text{tr}(\mathbf{S}_{XX} - \hat{\Lambda}\hat{\Lambda}')$. By comparing the two formulas, we can consider that the MDFA tends to give smaller unique variances than the classic FA. This tendency has been found empirically in Adachi (2015). However, as $\text{tr}\Psi_C^2 = \text{tr}(\mathbf{S}_{XX} - \Lambda_C\Lambda_C')$ and $\text{tr}\hat{\Psi}^2 \leq \text{tr}(\mathbf{S}_{XX} - \hat{\Lambda}\hat{\Lambda}')$ involve traces with $\Lambda_C \neq \hat{\Lambda}$ in general, $\Psi_C^2 \geq \hat{\Psi}^2$ does not always hold: exceptions can occur, as shown by the unique variance for "Science" in Table 2.

4. Covariances Among Residuals, Factors, and Observations

Since the data matrix $\mathbf{X}$ is column-centered, as well as $\hat{\mathbf{Z}} = [\hat{\mathbf{F}}, \hat{\mathbf{U}}]$ as shown in (4), $\hat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{Z}}\hat{\mathbf{B}}'$ is also column-centered. Thus, $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{Z}}} = n^{-1}\hat{\mathbf{E}}'\hat{\mathbf{Z}}$, $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{E}}} = n^{-1}\hat{\mathbf{E}}'\hat{\mathbf{E}}$, and $\mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} = n^{-1}\mathbf{X}\hat{\mathbf{Z}}$ are

Table 5. Inter-variables covariances of residuals

	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_1$	0.0011				
$\mathbf{e}_2$	-0.0009	0.0021		<i>Sym.</i>	
$\mathbf{e}_3$	0.0006	-0.0019	0.0018		
$\mathbf{e}_4$	-0.0013	0.0004	-0.0005	0.0027	
$\mathbf{e}_5$	0.0006	-0.0004	0.0004	-0.0013	0.0007

covariance matrices. In this section, we consider properties of those matrices and their relationships to the resulting value of loss function (6).

From the residual matrix $\hat{\mathbf{E}}$ identified as in Table 3, we can obtain the covariance matrix $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{E}}} = n^{-1}\hat{\mathbf{E}}'\hat{\mathbf{E}}$ in Table 5, which shows that the residuals are correlated among variables. This fact is illustrated in Figure 1, where two different variables $\mathbf{x}_j$ and $\mathbf{x}_k$ in $\mathbf{X}$ are depicted ($j \neq k$), with common factor matrix ($\hat{\mathbf{F}}$), unique factor vectors ($\hat{\mathbf{u}}_j$ and $\hat{\mathbf{u}}_k$), and residuals ($\hat{\mathbf{e}}_j$ and $\hat{\mathbf{e}}_k$). The correlations of residuals are expressed as $\hat{\mathbf{e}}_j$ being linked with $\hat{\mathbf{e}}_k$ in the figure. The next theorem concerns the covariances of residuals to factors:

Theorem 4.1. The $p \times m$ covariance matrix $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{F}}} = n^{-1}\hat{\mathbf{E}}'\hat{\mathbf{F}}$ between residuals and common factors is filled with zeros:

$$\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{F}}} = {}_p\mathbf{O}_m . \quad (25)$$

The diagonal elements of the $p \times p$ covariance matrix $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{U}}} = n^{-1}\hat{\mathbf{E}}'\hat{\mathbf{U}} = (s_{jk}^{\hat{\mathbf{E}}\hat{\mathbf{U}}})$ between residuals and unique factors are zeros, i.e.,

$$s_{jj}^{\hat{\mathbf{E}}\hat{\mathbf{U}}} = 0 , \quad (26)$$

while the off-diagonal ones equal to the corresponding elements of $\mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}}$:

$$s_{jk}^{\hat{\mathbf{E}}\hat{\mathbf{U}}} = s_{jk}^{\mathbf{X}\hat{\mathbf{U}}} \quad \text{for } k \neq j \quad (27)$$

with $\mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}}$ ($p \times p$) containing the covariances of observed variables to unique factors.

Proof. Using $\hat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{Z}}\hat{\mathbf{B}}'$, (5), and (11), we have $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{Z}}} = n^{-1}(\mathbf{X} - \hat{\mathbf{Z}}\hat{\mathbf{B}}')'\hat{\mathbf{Z}} = \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} - \hat{\mathbf{B}} =$

$[\mathbf{S}_{\mathbf{X}\hat{\mathbf{F}}}, \mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}}] - [\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}] = [{}_p\mathbf{O}_m, \text{Offd}(\mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}})]$ with $\text{Offd}(\mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}})$ ($p \times p$) denoting the matrix $\mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}}$

whose diagonal elements are replaced by zeros. $\square$

Figure 1. Path diagram for variables $\mathbf{x}_j$ and $\mathbf{x}_k$, where one-headed arrows indicate how those factors and residuals affect the variables, while the two-headed arrows indicate linked entities covarying mutually.

The result (25) and (26) show a very useful feature of the MDFA factors: the common factors in $\hat{\mathbf{F}}$ are uncorrelated with the residuals, and each unique factor in $\hat{\mathbf{U}}$ is uncorrelated with the residual for the corresponding variable. These facts may allow us to consider the residuals as unsystematic errors, which are supposed to behave independently of common and unique factors in Spearman's (1904) original conception of FA with the term "specific factors" used in place of "unique factors".

But, the Spearman's FA concept differs from (27): it shows that the residual for a certain variable correlates with the unique factors for the other variables, since $\text{Offd}(\mathbf{S}_{\mathbf{x}\hat{\mathbf{U}}})$ are not zero in general. This is demonstrated in Table 6, which was obtained with (16) for the standardized version of Table 1. Those findings are illustrated in Figure 1. There, $\hat{\mathbf{e}}_j$ is linked with $\hat{\mathbf{u}}_k$ ($k \neq j$), though $\hat{\mathbf{e}}_j$ is not linked with $\hat{\mathbf{F}}$ and $\hat{\mathbf{u}}_j$. The link $\hat{\mathbf{e}}_j$ with $\hat{\mathbf{u}}_k$ implies that variable $\mathbf{x}_j$ is correlated with $\hat{\mathbf{u}}_k$ ($k \neq j$): each variable is correlated with the unique factors for the other variables.

The property that $\mathbf{x}_j$ and $\hat{\mathbf{u}}_k$ ($k \neq j$) are correlated deviates from the concept of unique factors that they uniquely and exclusively contribute to the corresponding variables. Such a deviating property is not possessed in the constrained versions of MDFA mentioned in

Table 6. Covariances matrix $\mathbf{S}_{\mathbf{x}\hat{\mathbf{U}}}$ between variables and unique factors

	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$\mathbf{u}_4$	$\mathbf{u}_5$
Japa	0.7045	0.0126	-0.0248	0.0153	-0.0077
Eng	0.0146	0.6100	0.0052	-0.0415	0.0124
Soc	-0.0230	0.0042	0.7614	0.0345	-0.0082
Math	0.0147	-0.0347	0.0359	0.7307	0.0013
Sci	-0.0125	0.0175	-0.0144	0.0021	0.4344

Section 1. That is, $s_{jk}^{\text{X}\hat{\mathbf{U}}}$ is exactly zero for $k \neq j$ in the Kiers' version with the constraint $n^{-1}\mathbf{X}'\mathbf{U}$ being diagonal (Sočan, 2003), and in the Stegeman's (2016) one with $n^{-1}\mathbf{X}'\mathbf{U} = \mathbf{\Psi}$. However, how much the MDFA solution deviates from the unique factor concept can be assessed empirically by comparing the off-diagonal elements of $\mathbf{S}_{\text{X}\hat{\mathbf{U}}}$ with its diagonal ones, which are also those of $\hat{\mathbf{\Psi}}$ (11). If the absolute values of the off-diagonal elements are **much** smaller than those of the diagonal ones, we can conclude that the deviation is not substantial. For example, this is the case with the results presented in Table 6. **However, if substantial deviation was found with considerably large absolute values of the off-diagonal elements, we could reasonably consider using the constrained versions in place of MDFA.**

The amount of the deviation is shown to be proportional to the squared sum of residuals in Theorem 4.2 appearing later. For its preparation, we present the following two lemmas, which are also interesting by themselves:

Lemma 4.1. If $\hat{\mathbf{B}}$ ($p \times (p+m)$) is of full row-rank with $\hat{\mathbf{B}}\hat{\mathbf{B}}^+ = \mathbf{I}_p$, empirical covariance matrix $\mathbf{S}_{\text{XX}}$ equals the product of the p -variables $\times$ $(p+m)$ -factors covariance matrix updated as (16) and its transpose:

$$\mathbf{S}_{\text{XX}} = \mathbf{S}_{\text{X}\hat{\mathbf{Z}}}\mathbf{S}_{\text{X}\hat{\mathbf{Z}}}^', \text{ or equivalently, } s_{jk}^{\text{XX}} = \mathbf{s}_j^{\text{X}\hat{\mathbf{Z}}'}\mathbf{s}_k^{\text{X}\hat{\mathbf{Z}}} . \quad (28)$$

Proof. Making use of (16) we have $\mathbf{S}_{\text{X}\hat{\mathbf{Z}}}\mathbf{S}_{\text{X}\hat{\mathbf{Z}}}^' = \hat{\mathbf{B}}^+\mathbf{L}\mathbf{\Delta}^2\mathbf{L}'\hat{\mathbf{B}}$. Following (15), the right-hand side can be further rewritten as $\hat{\mathbf{B}}^+\hat{\mathbf{B}}'\mathbf{S}_{\text{XX}}\hat{\mathbf{B}}\hat{\mathbf{B}}^+$, which is exactly $\mathbf{S}_{\text{XX}}$ as $\hat{\mathbf{B}}\hat{\mathbf{B}}^+ = \mathbf{I}_p$. $\square$

Lemma 4.2. Let $\hat{\mathbf{B}}$ be of full row-rank. When (13) is substituted into $\mathbf{Z}$ in (9), the first term in the right-hand side of (9) vanishes:

$$\|\mathbf{x}_j - \hat{\mathbf{Z}}\mathbf{s}_j^{\text{X}\hat{\mathbf{Z}}}\|^2 = 0, \quad \text{i.e., } f_j(\hat{\mathbf{Z}}, \mathbf{b}_j) = n\|\mathbf{s}_j^{\text{X}\hat{\mathbf{Z}}} - \mathbf{b}_j\|^2 \quad (29)$$

Proof. The identity $\|\mathbf{x}_j - \hat{\mathbf{Z}}\mathbf{s}_j^{\text{X}\hat{\mathbf{Z}}}\|^2 = n s_{jj}^{\text{XX}} - n\mathbf{s}_j^{\text{X}\hat{\mathbf{Z}}'}\mathbf{s}_j^{\text{X}\hat{\mathbf{Z}}}$ follows from (5), which, by (28), is zero.

$\square$

Using the lemmas, the next theorem is given:

Theorem 4.2. Let $\hat{\mathbf{B}}$ be of full row-rank. Then, the variance of residuals $n^{-1}\|\hat{\mathbf{e}}_j\|^2$ for variable j , given in (18), is rewritten as

$$\frac{1}{n} \|\hat{\mathbf{e}}_j\|^2 = \sum_{k \neq j}^p |s_{jk}^{\mathbf{x}\hat{\mathbf{U}}}|^2 = \sum_{k=1}^p |s_{jk}^{\hat{\mathbf{E}}\hat{\mathbf{U}}}|^2 \quad (30)$$

with $s_{jk}^{\mathbf{x}\hat{\mathbf{U}}}$ the (j, k) element of $\mathbf{S}_{\mathbf{x}\hat{\mathbf{U}}}$ and $s_{jk}^{\hat{\mathbf{E}}\hat{\mathbf{U}}}$ that of $\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{U}}}$. That is, $n^{-1} \|\hat{\mathbf{e}}_j\|^2$ equals the sum of the squared covariances of variable j to the unique factors for the other variables, and $n^{-1} \|\hat{\mathbf{e}}_j\|^2$ also equals the sum of the squared covariances of the residual for variable j to unique factors.

Proof. Using $\mathbf{s}_j^{\mathbf{x}\hat{\mathbf{F}}'}$ and $\mathbf{h}_j'$ for the j th rows of $\mathbf{S}_{\mathbf{x}\hat{\mathbf{F}}}$ and $\mathbf{I}_p$, the j th row of $\hat{\mathbf{B}}$ in (11) can be rewritten as $\hat{\mathbf{b}}_j' = [\mathbf{s}_j^{\mathbf{x}\hat{\mathbf{F}}'}, \hat{\psi}_j \mathbf{h}_j']$. Substituting it into (29) divided by n leads to $\|\mathbf{s}_j^{\mathbf{x}\hat{\mathbf{U}}'} - \hat{\psi}_j \mathbf{h}_j'\|^2 = \sum_{k \neq j} |s_{jk}^{\mathbf{x}\hat{\mathbf{U}}}|^2 = \sum_k |s_{jk}^{\hat{\mathbf{E}}\hat{\mathbf{U}}}|^2$, where $\mathbf{s}_j^{\mathbf{x}\hat{\mathbf{U}}'}$ is the j th row of $\mathbf{S}_{\mathbf{x}\hat{\mathbf{U}}}$ and (27) has been used. $\square$

The equalities in (30) can be verified in the example: for the j th row in Table 6, we can obtain $\sum_{k \neq j} |s_{jk}^{\mathbf{x}\hat{\mathbf{U}}}|^2$ to find it equaling the j th element of the column “Res” in Table 2. By comparing (30) with (17), we can find

$$f(\hat{\mathbf{Z}}, \hat{\mathbf{B}}) = \|\hat{\mathbf{E}}\|^2 = n \|\text{Offd}(\mathbf{S}_{\mathbf{x}\hat{\mathbf{U}}})\|^2 = n \|\mathbf{S}_{\hat{\mathbf{E}}\hat{\mathbf{U}}}\|^2 = n \text{tr}(\mathbf{S}_{\mathbf{xx}} - \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' - \hat{\mathbf{\Psi}}^2). \quad (31)$$

Here, $n \|\text{Offd}(\mathbf{S}_{\mathbf{x}\hat{\mathbf{U}}})\|^2 = n \sum_{k \neq j} |s_{jk}^{\mathbf{x}\hat{\mathbf{U}}}|^2$ can be viewed as the amount of the deviation from the unique factor concept discussed already. That is, the minimization of the MDFA loss function (6) means reduction of the deviation from the concept that the unique factors uniquely and exclusively contribute to the corresponding variables.

5. Factor Indeterminacy and Higher Rank Approximation

Stegeman (2016, Sections 3 and 4.2) has discussed how the factor scores in $\mathbf{F}$ and $\mathbf{U}$ are undetermined, but in the **remainder** of his paper, he has focused on the degree of indeterminacy of $\mathbf{F}$ only. In this section, we consider the indeterminacy of both $\mathbf{F}$ and $\mathbf{U}$.

The estimation of $\mathbf{Z}$ can be viewed as a higher rank approximation problem. It is because the MDFA loss function (6) can be rewritten as

$$f(\mathbf{Z} | \mathbf{B}) = \|\mathbf{Z} - \mathbf{XB}\|^2 + c \quad (32)$$

with $c = \|\mathbf{X}\|^2 + n\|\mathbf{B}\|^2 - \|\mathbf{XB}\|^2 - n(p+m)$ irrelevant to $\mathbf{Z}$ (Adachi, 2015). The minimization of (32) over $\mathbf{Z}$ for a given $\mathbf{B}$ amounts to approximating $\mathbf{XB}$ by higher rank matrix $\mathbf{Z}$. Here, $\text{rank}(\mathbf{Z}) =$

$m+p$, which follows from (5) and $n > m + p$, is greater than (7).

The optimal $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ is given in (13) as a sum $\hat{\mathbf{Z}} = \mathbf{Z}^* + \mathbf{Z}_\perp^*$, where $\mathbf{Z}^*$ is a unique part

$$\mathbf{Z}^* = [\mathbf{F}^*, \mathbf{U}^*] = n^{1/2} \mathbf{K} \mathbf{L}' = \mathbf{X} \hat{\mathbf{B}} \mathbf{L} \Delta^{-1} \mathbf{L}' = \mathbf{X} \mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} = \mathbf{X} \mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1} [\mathbf{S}_{\mathbf{X}\hat{\mathbf{F}}}, \mathbf{S}_{\mathbf{X}\hat{\mathbf{U}}}] \quad (33)$$

and $\mathbf{Z}_\perp^*$ is an undetermined part

$$\mathbf{Z}_\perp^* = n^{1/2} \mathbf{K}_\perp \mathbf{L}_\perp'. \quad (34)$$

In (33) the fact is used that (20) is rewritten as $\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} = \hat{\mathbf{B}} \mathbf{L} \Delta^{-1} \mathbf{L}'$. By comparing (33) with (11), we find that the unique part for common factor scores is given by $\mathbf{F}^* (n \times m) = \mathbf{X} \mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{S}_{\mathbf{X}\hat{\mathbf{F}}} = \mathbf{X} \mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1} \hat{\mathbf{A}}$. This result was also shown by Stegeman (2016, Lemma 3.1). He pointed out that this is identical to the Thurstone's (1935) estimate of $\mathbf{F}$.

To discuss how $\mathbf{Z}$ is undetermined, we present two results. First, (33) can be rewritten as

$$\mathbf{Z}^* = \mathbf{X} (n^{-1} \mathbf{X}' \mathbf{X})^{-1} (n^{-1} \mathbf{X}' \hat{\mathbf{Z}}) = \Pi_{\mathbf{X}} \hat{\mathbf{Z}}. \quad (35)$$

with $\Pi_{\mathbf{X}} = \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$ be the orthogonal projection matrix onto $\text{Sp}(\mathbf{X})$, the subspace spanned by the columns of the data matrix $\mathbf{X}$. That is, the unique part $\mathbf{Z}^*$ (33) is the projection of $\hat{\mathbf{Z}} = [\hat{\mathbf{F}}, \hat{\mathbf{U}}]$ onto $\text{Sp}(\mathbf{X})$. On the other hand, the following theorem involves the projection of $\mathbf{X}$ onto $\hat{\mathbf{Z}}$:

Theorem 5.1. It holds that

$$\mathbf{X} = \hat{\mathbf{Z}} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}' = \Pi_{\hat{\mathbf{Z}}} \mathbf{X} \quad (36)$$

with $\Pi_{\hat{\mathbf{Z}}} = \hat{\mathbf{Z}} (\hat{\mathbf{Z}}' \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}'$ the projector onto $\text{Sp}(\hat{\mathbf{Z}})$.

Proof. Using (5) and (28), we have $n^{-1} \|\mathbf{X} - \hat{\mathbf{Z}} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}'\|^2 = \text{tr} \mathbf{S}_{\mathbf{X}\mathbf{X}} - \text{tr} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}' = 0$, which implies

$$\mathbf{X} = \hat{\mathbf{Z}} \mathbf{S}_{\mathbf{X}\hat{\mathbf{Z}}}' = \hat{\mathbf{Z}} \{n^{-1} \hat{\mathbf{Z}}' \mathbf{X}\} = \hat{\mathbf{Z}} \{(\hat{\mathbf{Z}}' \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}' \mathbf{X}\}. \quad \square$$

This theorem implies that $\text{Sp}(\mathbf{X})$ is included in $\text{Sp}(\hat{\mathbf{Z}}) = \text{Sp}([\hat{\mathbf{F}}, \hat{\mathbf{U}}])$.

The above facts and the equation

$$h = \|\hat{\mathbf{Z}} - \mathbf{Z}^*\|^2 = \|n^{1/2} \mathbf{K}_\perp \mathbf{L}_\perp'\|^2 = n \text{tr} \mathbf{L}_\perp' \mathbf{L}_\perp = nm \quad (37)$$

derived in Adachi (2012) can give a geometric (cone) representation drawn in Figure 2, where the matrices are depicted as arrows. This figure shows that the optimal common-unique factor score matrix $\hat{\mathbf{Z}}$ forms the cone whose central axis is (33) with the distance of $\hat{\mathbf{Z}}$ to (33) being the constant $(nm)^{1/2}$ as found in (37). In other words, **any $\mathbf{Z}$ is the optimal that exists on** the circumference of the circle whose center is $\mathbf{Z}^*$ and radius is $(nm)^{1/2}$. There, the covariance

Figure 2. Cone formed by common-unique factor scores

matrix $\mathbf{S}_{\hat{\mathbf{Z}}}$ corresponds to the angle between $\hat{\mathbf{Z}}$ and $\mathbf{Z}^*$ which is constant, because $\mathbf{S}_{\hat{\mathbf{Z}}}$ can be uniquely determined as proved in the next theorem:

Theorem 5.2. If the solution $\hat{\mathbf{B}}$ is unique, then the covariance matrix $\mathbf{S}_{\hat{\mathbf{Z}}}$ between p variables and $m+p$ factors is uniquely determined

Proof. Unique $\mathbf{B}$ allows the right-hand side of (15) to be uniquely determined, which also makes (16) unique.

Mulaik (1976) has also presented a cone similar to the one in Figure 2. However, his cone is based on the classic model (1) and only illustrates the common factor indeterminacy, i.e., that $\mathbf{f}$ forms a circle whose center is at $\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{x}$ with $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2$. The cone presented in Figure 2 illustrates the indeterminacy of both common and unique factors. Another difference is that the radius of our cone is constant $(nm)^{1/2}$, while the radius $\|\mathbf{f} - \mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{x}\|$ of the Mulaik's one varies randomly, since $\mathbf{x}$ and $\mathbf{f}$ in (1) are random variables.

Let $\tilde{\mathbf{Z}} = \mathbf{Z}^* + \tilde{\mathbf{Z}}_{\perp}^*$ be another $n \times (m+p)$ factor score matrix which differs from $\hat{\mathbf{Z}} = \mathbf{Z}^* + \mathbf{Z}_{\perp}^*$, by its undetermined part $\tilde{\mathbf{Z}}_{\perp}^*$. Figure 2 shows that the squared norm $\|\hat{\mathbf{Z}} - \tilde{\mathbf{Z}}\|^2 = \|\mathbf{Z}_{\perp}^* - \tilde{\mathbf{Z}}_{\perp}^*\|^2$ is maximal when $\tilde{\mathbf{Z}}_{\perp}^* = -\mathbf{Z}_{\perp}^*$:

$$\|\hat{\mathbf{Z}} - \tilde{\mathbf{Z}}\|^2 \leq \|2\mathbf{Z}_{\perp}^*\|^2 = 4\|\mathbf{Z}_{\perp}^*\|^2 = 4nm. \quad (38)$$

It can be viewed as a common-unique factor version of the Stegeman's (2016, Lemma 3.1, (ii)) result concerning the norm for the common factor score matrix $\mathbf{F}$. Another difference from his lemma is that the upper bound is given by a constant $4nm$ as in (38), due to (37). The inequality (38) implies that the element-wise averaged squared difference satisfies

$$\frac{1}{n(m+p)} \|\hat{\mathbf{Z}} - \tilde{\mathbf{Z}}\|^2 \leq \frac{4nm}{n(m+p)} = \frac{4}{1+p/m}. \quad (39)$$

Its upper limit provides an indicator for common-unique factor indeterminacy. It shows that

Table 7. Minimum and maximum of common and unique factor scores

Obs.	Common Factors						Unique Factors									
	f_1		f_2		u_1		u_2		u_3		u_4		u_5			
	min	max	min	max	min	max	min	max	min	max	min	max	min	max	min	max
1	-0.6	1.9	-1.1	1.8	-1.8	0.7	-2.6	1.1	0.4	2.9	-0.7	1.4	-2.3	2.7		
2	-2.4	0.4	-2.1	1.1	-1.6	1.1	-1.9	2.0	-1.5	1.2	-1.5	0.7	-2.7	2.7		
3	0.0	2.4	-0.9	2.1	-0.4	2.0	-2.6	0.9	-0.4	2.1	-0.8	1.1	-1.9	2.9		
4	-0.7	2.0	-2.1	1.0	-0.3	3.9	-1.6	1.9	-2.5	1.8	-2.5	2.1	-1.8	2.8		
5	-0.5	2.2	-3.1	1.2	-1.6	2.4	-2.3	1.8	-1.0	1.6	-1.8	0.9	-1.6	3.2		
6	-0.5	2.2	-3.7	0.0	-2.6	-0.2	-0.7	2.5	-2.4	0.1	-1.1	0.8	-1.0	3.2		
7	-2.5	-0.1	-2.1	0.8	-2.4	0.5	-1.5	2.7	-1.5	1.5	-1.2	0.8	-2.7	2.1		
8	-0.8	3.4	-3.6	-0.5	-1.0	3.0	-3.6	-0.3	-0.7	1.8	-0.7	1.3	-2.0	2.5		
9	-2.8	-0.4	-2.1	1.1	-0.8	3.0	-2.4	1.3	-0.9	2.4	-2.3	-0.4	-2.7	2.2		
10	-2.2	0.7	-1.3	1.7	-2.2	0.6	-1.1	2.5	-1.6	0.9	-0.5	2.5	-3.2	1.9		
11	-2.2	0.8	-1.7	1.9	-1.1	2.0	-0.7	2.9	-2.3	0.4	-1.4	0.7	-2.7	4.6		
12	-2.0	0.4	-0.6	2.3	-0.7	2.0	-1.1	3.3	-2.5	0.0	0.3	4.5	-3.6	1.2		
13	-1.1	1.4	-1.2	1.7	-2.6	2.6	-1.8	4.0	-0.3	2.2	0.4	2.4	-2.8	2.0		
14	0.8	4.6	-0.2	2.4	-0.1	4.5	-1.0	2.1	-1.8	1.7	-0.3	5.1	-2.0	2.8		
15	-1.5	1.1	-0.9	2.3	-0.9	2.7	-2.2	2.0	-0.8	3.1	0.6	2.5	-3.5	1.8		
16	-1.2	0.8	0.1	2.7	-2.4	-0.1	-0.3	2.7	-1.1	2.7	-1.7	3.0	-1.5	2.7		
17	-0.9	2.0	-1.9	1.3	-1.7	0.9	-1.5	3.2	-2.2	0.5	-0.9	1.2	-2.4	2.8		
18	-1.3	1.4	-0.5	2.6	0.3	3.5	-1.8	2.0	-1.9	0.8	-0.9	1.1	-3.2	3.7		
19	-2.1	0.7	-1.6	2.1	0.1	2.3	-1.7	2.0	-0.1	5.3	-2.6	-0.9	-2.4	2.5		
20	-1.8	0.4	-2.8	0.1	-1.7	1.8	-3.0	0.5	-1.7	0.8	-0.1	1.8	-2.7	1.7		

the indeterminacy is reduced as the number of variables (p) increases and that of factors (m) decreases, which is similar to the Guttman's (1955) results.

If the indeterminacy of factor scores can be assessed for each observation, it may be more intuitively appealing rather than presenting a statistic defined for a set of observations as in (38) or (39). For such an assessment, Stegeman's (2016, Section 4.4) procedure is useful, in which the undetermined parts of factor scores generated randomly are summed to the unique part so as to synthesize a number of possible factor scores for each observation. Though Stegeman (2016, Section 6) used the procedure for assessing the indeterminacy of $\mathbf{F}$ in his constrained version with $n^{-1}\mathbf{X}'\mathbf{U} = \mathbf{\Psi}$, we use it for $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ in MDFA. Table 7 shows the full ranges of the scores in the resulting 3000 factor score matrices for the standardized version of the data in Table 1, with the procedure detailed in Appendix 1. To evaluate the sufficiency of the number of matrices (3000) being used, (38) can be utilized by maximizing $\|\hat{\mathbf{Z}} - \tilde{\mathbf{Z}}\|^2$ over the pairs of 3000 matrices. The maximum is 158.168, which is satisfactorily close to the upper limit $4nm = 4 \times 20 \times 2 = 160$.

In Table 7, the (1, 1) and (1, 2) elements (−0.6 and 1.9) indicate the minimum and maximum of the (1, 1) elements in the generated matrices. It allows us to be certain that observations having same signs for both limits are characterized by the corresponding factors. For example, it is convincing that the 14th observation has a positive score for the first common factor. However, such observations are few: there is a lot of fluctuation in the factor scores. But, their possible values **were limited** now, with the non-overlapping intervals between observations showing their mutual distinctions. It should be noticed that those factor scores are essentially different from the scores considered in the classic FA. There, the factor

scores are random as in (1), and the corresponding loss function is irrelevant to the scores. Thus, if necessary, once after Λ and Ψ^2 are optimally estimated, those solutions and data matrix $\mathbf{X}$ are used to provide the approximations of the common factor scores for each observation, with its unique factor scores of no interest. On the other hand, MDFA's factor scores in $\hat{\mathbf{Z}} = [\hat{\mathbf{F}}, \hat{\mathbf{U}}]$, though not unique, are optimal in that they minimize (6) (under constraints). That is, the scores within the ranges in Table 7 are the optimal, though they fluctuate.

6. Discussion

We investigated several properties of the FA solutions in its matrix decomposition formulation, i.e. when FA is modeled as $\mathbf{X} = \mathbf{F}\Lambda' + \mathbf{U}\Psi + \mathbf{E}$ for column-centered data matrix $\mathbf{X}$, with $\mathbf{F}$ (common factors), $\mathbf{U}$ (unique factors), Λ (loadings), and Ψ (including the square roots of unique variance) treated as fixed parameter matrices.

The main results are summarized as follows:

- [1] The model part $\mathbf{F}\Lambda' + \mathbf{U}\Psi$ and the residuals $\mathbf{E}$ can be uniquely determined, so that we can assess the goodness of fit of the model for each observation, although the factor score matrix $[\mathbf{F}, \mathbf{U}]$ is not unique.
- [2] Common factors are uncorrelated with residuals, and each unique factor is uncorrelated with the residual for the corresponding variable.
- [3] The unique factor for variable j covaries with the other variables $k \neq j$, which deviates from the concept of unique factors. The squares of the covariances are summed to express the amount of the deviation, which is proportional to the resulting value of the MDFA loss function.
- [4] The indeterminacy of common and unique factors can be illustrated as in Figure 2: the optimal $[\mathbf{F}, \mathbf{U}]$ can be viewed as forming a cone whose central axis is the projection of the column space of $[\mathbf{F}, \mathbf{U}]$ onto that of $\mathbf{X}$.

Result [3] implies that MDFA reduces the deviation from the unique factor concept, although the deviation cannot be completely avoided. This can be achieved by Kiers' version of MDFA (Sočan, 2003) and Stegeman's (2016) one with the deviation constrained to be null. This is advantageous to the constrained versions, which can be used in place of MDFA if this could not reduce the deviation satisfactorily, as mentioned in Section 4. However, MDFA has a merit of Result [1], i.e., the model part $\mathbf{F}\Lambda' + \mathbf{U}\Psi = \mathbf{ZB}'$ in (3) can be identified, and thus the residual $\mathbf{E} = \mathbf{X} - \mathbf{ZB}'$ is also identifiable, which is useful for assessing the badness of fit of the FA model for each observation, as illustrated in Table 3. In Stegeman's (2016) version, the model part is not uniquely determined, as described in Stegeman (2016, Appendix 2). For Kiers' constrained MDFA (Sočan, 2003), properties have not yet been investigated

sufficiently: what we know is restricted to the simulation study reported in Sočan (2003) and the description in Stegeman (2016, Appendix 3). Properties of the Kiers' **constrained** version remain to be studied, together with further comparisons between MDFA and its constrained versions.

Result [4] means that the optimal factor scores are restricted as depicted in Figure 2, i.e., the terminuses of the arrows expressing $\hat{\mathbf{Z}}$ must be located on the circumference of the circle in the figure. This suggests that we may be able to exploit the factor indeterminacy, as the rotational indeterminacy of the loading matrix is exploited to provide interpretable rotated loadings (e.g., Mulaik, 2010). That is, it may be possible to develop a procedure, in which an interpretable $\hat{\mathbf{Z}}$ existing somewhere on the circumference is found once after one of the optimal $\hat{\mathbf{Z}}$ is obtained by MDFA.

In contrast to the classic FA model, where the factors are random variables, all model parameters/unknowns in MDFA are treated as fixed matrices. As a consequence, no probabilistic aspects of the MDFA parameters are considered. The benefit of this approach is that it makes it possible to derive and clarify a number of new useful matrix-algebraic properties of the FA solutions. As the elements of $\mathbf{E}$ in the MDFA model (3) are random and depend on the particular sample, probabilistic properties of MDFA can still be investigated, which remains for future studies.

As pointed out by an anonymous reviewer, applications should be considered in which FA with ordinal variables is performed. A potential approach to such applications is found in Makino (2015), where MDFA is extended so as to deal with categorical variables. In this extended version, the formulation of nonlinear principal component analysis (Gifi, 1990) is incorporated into MDFA by rewriting the data matrix $\mathbf{X}$ in (6) using two types of matrices. One of them consists of the binary data matrices indicating the correspondence of observations to categories, while the other contains the scores to be assigned to the categories. Then, (6) is minimized over $\mathbf{Z}$, $\mathbf{B}$, and the category scores subject to certain constraints. Here, it should be noted that ordinal variables are the categorical ones with ordinal restriction. Thus, if the restriction is imposed into the categorical scores as made in Gifi (1990, Chapter 4), Makino's (2015) approach could be applied for ordinal variables.

Appendix

Here, we describe the procedure for generating (34) to have $\hat{\mathbf{Z}} = \mathbf{Z}^* + \mathbf{Z}_\perp^*$, i.e., (13), for the matrix $\mathbf{Z}^*$ (33) given after Steps 1 to 4 in Section 2.

Let $\mathbf{J}_1 = \mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'$, $\mathbf{J}_K = \mathbf{I}_n - \mathbf{K}\mathbf{K}'$, and $\mathbf{J}_L = \mathbf{I}_{m+p} - \mathbf{L}\mathbf{L}'$. Using those matrices with

arbitrary $n \times m$ matrix $\mathbf{G}$ and $(m+p) \times m$ matrix $\mathbf{H}$, we consider the following SVDs:

$$\mathbf{J}_1 \mathbf{J}_k \mathbf{G} = \Gamma_G \Theta_G \Xi_G' \quad \text{and} \quad \mathbf{J}_L \mathbf{H} = \Gamma_H \Theta_H \Xi_H' \quad (\text{A1})$$

with Θ_G and Θ_H being diagonal and

$$\Gamma_G' \Gamma_G = \Gamma_H' \Gamma_H = \Xi_G' \Xi_G = \Xi_H' \Xi_H = \mathbf{I}_m. \quad (\text{A2})$$

Those SVDs allow us to set $\mathbf{K}_\perp = \Gamma_G$ and $\mathbf{L}_\perp = \Gamma_H$ to generate (34), which is summed to (33) so that (13) is given by

$$\hat{\mathbf{Z}} = n^{1/2} \tilde{\mathbf{K}} \tilde{\mathbf{L}}' = n^{1/2} \mathbf{K} \mathbf{L}' + n^{1/2} \Gamma_G \Gamma_H' = \mathbf{X} \mathbf{S}_{XX}^{-1} \mathbf{S}_{XZ} + n^{1/2} \Gamma_G \Gamma_H'. \quad (\text{A3})$$

It is because (A1) leads to $\Gamma_G = \mathbf{J}_1 \mathbf{J}_k \mathbf{G} \Xi_G \Theta_G^{-1}$ and $\Gamma_H = \mathbf{J}_L \mathbf{H} \Xi_H \Theta_H^{-1}$, which imply that

$$\mathbf{1}_n' \Gamma_G = \mathbf{0}_m, \quad \mathbf{L}' \Gamma_H = {}_p \mathbf{O}_m, \quad \text{and} \quad \mathbf{K}' \Gamma_G = {}_p \mathbf{O}_m. \quad (\text{A4})$$

Here, the last identity is derived as $\mathbf{K}' \Gamma_G = \mathbf{K}' \mathbf{J}_1 \mathbf{J}_k \mathbf{G} \Xi_G \Theta_G^{-1} = \mathbf{K}' \mathbf{J}_k \mathbf{G} \Xi_G \Theta_G^{-1} = {}_p \mathbf{O}_m$ using $\mathbf{J}_1 \mathbf{K} = \mathbf{K}$ which follows from $\mathbf{1}_n' \mathbf{K} = \mathbf{0}_p'$ (Adachi, 2012). The equations $\mathbf{K}' \mathbf{K} = \mathbf{L}' \mathbf{L} = \mathbf{I}_p$ following from SVD (14) and the above (A2), (A4), and $\mathbf{1}_n' \mathbf{K} = \mathbf{0}_p'$ allow us to find that $\tilde{\mathbf{K}} = [\mathbf{K}, \mathbf{K}_\perp] = [\mathbf{K}, \Gamma_G]$ and $\tilde{\mathbf{L}}' = [\mathbf{L}, \mathbf{L}_\perp] = [\mathbf{L}, \Gamma_H]$ satisfy $\tilde{\mathbf{K}} \tilde{\mathbf{K}}' = \tilde{\mathbf{L}}' \tilde{\mathbf{L}} = \tilde{\mathbf{L}} \tilde{\mathbf{L}}' = \mathbf{I}_{p+m}$ and $\mathbf{1}_n' \hat{\mathbf{Z}} = \mathbf{1}_n' [\mathbf{K}, \Gamma_G] [\mathbf{L}, \Gamma_H]' = \mathbf{0}_{p+m}$: (A3) is the optimal with satisfying (4) and (5).

Thus, we can obtain a factor score matrix $\mathbf{Z}$ with the following procedure:

- Step 1. Generate $\mathbf{G}$ and $\mathbf{H}$ randomly
- Step 2. Perform SVDs in (A1)
- Step 3. Obtain $\hat{\mathbf{Z}}$ with (A3)
- Step 4. Replace $\mathbf{F}$ in $\hat{\mathbf{Z}} = [\hat{\mathbf{F}}, \hat{\mathbf{U}}]$ by rotated $\hat{\mathbf{F}} \mathbf{T}$.

Here, each element of $\mathbf{G}$ and $\mathbf{H}$ is sampled randomly from the uniform distribution ranging from -1 to 1 , and the matrix $\mathbf{T}$ is the orthonormal matrix for the varimax rotation giving the MDFA loadings in Table 2. The replication of the steps gives a number of $\hat{\mathbf{Z}}$.

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. (2015). A matrix-intensive approach to factor analysis. *Journal of the Japan Statistical Society, Japanese Issue*, **44**, 363-382 (in Japanese).

- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, vol. 5. (pp. 111-150). Berkeley, CA: University of California Press.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology (Statistics Section)*, 3, 77-85.
- de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In *Recent developments of structural equation models: Theory and applications*, Eds. by K. van Montfort, J. Oud and A. Satorra, 121-134. Dordrecht: Kluwer Academic Publishers.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.
- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *British Journal of Statistical Psychology*, 8, 65-81.
- Harman, H. H. and Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika*, 31, 351-369.
- Harman, H.H. (1976). *Modern factor analysis (3rd Edition)*. Chicago: The University of Chicago Press.
- Ihara, M. & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, 51, 563-566.
- Makino, N. (2015). Generalized data-fitting factor analysis with multiple quantification of categorical variables. *Computational Statistics*, 30, 279-292.
- Mulaik, S. A. (1976). Comments on "The measurement of factorial indeterminacy". *Psychometrika*, 41, 249-262.
- Mulaik, S. A. (2010). *Foundations of factor analysis*, 2nd Edition. Boca Raton: CRC Press.
- Rubin, D. B. and 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: 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.
- Tanaka, Y., & Tarumi, T. (1995). *Handbook for statistical analysis: Multivariate analysis (windows version)*. Tokyo: Kyoritsu-Shuppan (in Japanese).
- ten Berge, J. M. F. (1983). A generalization of Kristof's theorem on the trace of certain matrix products. *Psychometrika*, 48, 519-523.
- ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis*. Leiden, The Netherlands: DSWO Press.
- Thurstone, L. L. (1935). *The vectors of mind*. University of Chicago Press: Chicago.

Unkel, S. and Trendafilov, N. T. (2010a). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, 78, 363-382.