PSYCHOMETRIKA—VOL. 85, NO. 3, 555–574
SEPTEMBER 2020
https://doi.org/10.1007/s11336-020-09715-4

FACTOR UNIQUENESS OF THE STRUCTURAL PARAFAC MODEL

PAOLO GIORDANI
SAPIENZA UNIVERSITÀ DI ROMA

ROBERTO ROCCI
SAPIENZA UNIVERSITÀ DI ROMA

GIUSEPPE BOVE
UNIVERSITY OF ROMA TRE

Factor analysis is a well-known method for describing the covariance structure among a set of manifest variables through a limited number of unobserved factors. When the observed variables are collected at various occasions on the same statistical units, the data have a three-way structure and standard factor analysis may fail. To overcome these limitations, three-way models, such as the Parafac model, can be adopted. It is often seen as an extension of principal component analysis able to discover unique latent components. The structural version, i.e., as a reparameterization of the covariance matrix, has been also formulated but rarely investigated. In this article, such a formulation is studied by discussing under what conditions factor uniqueness is preserved. It is shown that, under mild conditions, such a property holds even if the specific factors are assumed to be within-variable, or within-occasion, correlated and the model is modified to become scale invariant.

Key words: three-way factor analysis, maximum likelihood, factor uniqueness property.

1. Introduction

Factor analysis (FA) (Bartholomew, Knott, & Moustaki, 2011) is a well-known method explaining the relationships among a set of manifest variables, observed on a sample of statistical units, in terms of a limited number of latent variables. In FA, data are stored in a matrix, say $\mathbf{X}$, of order $(I \times J)$ being I and J the number of statistical units and variables, respectively. Thus, FA deals with two-way two-mode data, where the modes are the entities of the data matrix, i.e., statistical units and manifest variables, and the ways are the indexes of the elements of $\mathbf{X}$, i.e., $i = 1, \ldots, I$ and $j = 1, \ldots, J$. In many practical situations, it may occur that the scores on the same manifest variables with respect to a sample of statistical units are replicated across K different occasions, e.g., time, locations, conditions, etc. Examples can be found in several domains. In medicine, these are the daily measures of some vital characteristics on a set of patients. In chemistry, samples measured at different emission wavelengths and excitation wavelengths. In social sciences, think about the yearly evaluation of the main cities of a given area in terms of indicators assessing the quality of life. In marketing, ratings on some goods expressed by a sample of consumers. In psychology, the scores for a group of children on a set of personality variables (traits) rated by different judges (methods). In the previous examples, there are three sets of entities (statistical units, manifest variables and occasions), hence three modes. For instance, in the latter one, these are children, traits and methods. The available information is stored in

Correspondence should be made to Paolo Giordani, Department of Statistical Sciences, Sapienza Università di Roma, P.le Aldo Moro, 5, 00185 Rome, Italy. Email: paolo.giordani@uniroma1.it

© 2020 The Author(s)

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

the so-called array, or tensor, usually denoted by $\underline{\mathbf{X}}$ of order $(I \times J \times K)$. Its generic element is $x_{ijk}, i = 1, \ldots, I, j = 1, \ldots, J$ and $k = 1, \ldots, K$, expressing the score of statistical unit i (e.g., child) on manifest variable j (e.g., trait) at occasion k (e.g., method). Therefore, the elements have three indexes and the array three ways. For all of these reasons, data are three-way three-mode. For further details, refer to, e.g., Kiers (2000) and Kroonenberg (2008).

The basic FA model is not adequate to handle three-way three-mode data. The reason is that FA does not take into account properly the multiway multimode structure of the data. In principle, FA could still be applied. Specifically, it is possible to convert $\underline{\mathbf{X}}$ into a matrix by the so-called matricization or unfolding. For instance, one can easily obtain the matrix $\mathbf{X}_{\mathrm{A}}$ of order $(I \times JK)$ by juxtaposing next to each other the frontal slabs of $\underline{\mathbf{X}}$ (statistical unit-mode matricization), where the frontal slabs of $\underline{\mathbf{X}}$ are the standard two-way two-mode matrices $\mathbf{X}_k$ $(k = 1, \ldots, K)$ of order $(I \times J)$ collected at the different occasions. Nonetheless, the FA model fitted to $\mathbf{X}_{\mathrm{A}}$ discovers latent variables ignoring that the same manifest variables are replicated across the occasions. For this reason, the standard FA model has been extended in order to take into account and exploit the increasing complexity of three-way three-mode data.

The most famous three-way three-mode extensions of FA are essentially based on the three-mode factor analysis model (Tucker, 1963; Tucker, 1966), usually named Tucker3, and parallel factor analysis model (Harshman, 1970), usually named Parafac. The latter can be seen as a particular case of the former with a useful property of parameter uniqueness (Kruskal, 1977). Although such extensions were presented as suitable generalizations of FA, they can be considered as extensions of the principal component analysis (PCA) solution to the FA model. As a matter of fact, Carroll & Chang (1970) proposed a canonical decomposition equivalent to Parafac, named Candecomp; as an extension of PCA in multidimensional scaling, Kroonenberg & De Leeuw (1980) named three-mode principal component analysis a method based on the ordinary least squares (OLS) estimation of Tucker3. For these reasons, hereinafter such models will be said to follow a component-based approach. The reader interested in the relationships between FA and PCA may refer to, for instance, Unkel & Trendafilov (2010) and Adachi & Trendafilov (2019).

Some authors revised Tucker3 and Parafac as structural models for the covariance structure of the manifest variables (see, for example, Bentler, Poon, & Lee, 1988 and references therein). In other terms, the data generation process is explicitly specified. The I statistical units are considered independent observations of the same random variable, and the models are reformulated as a suitable reparameterization of the parameters of its distribution. In particular, each model specifies a certain structure for the variables $\times$ occasions covariance matrix where the covariance matrix of the specific factors, i.e., the error term, is explicitly considered (see, for example, Kroonenberg & Oort, 2003).

As far as we know, in the literature, the component-based approach has received much more attention than the structural one. In this paper, after recalling the main features of methods and models characterizing the two approaches, a structural extension of Parafac is considered. It will be shown how its fundamental factor uniqueness property is preserved even when some specific factors are correlated across occasions, or variables, and/or its structure is modified to become scale invariant. The effectiveness of the proposal is illustrated by a real-life example.

2. The Component-Based Approach

The Tucker3 model (Tucker, 1966) summarizes the three-way three-mode tensor $\underline{\mathbf{X}}$ by looking for a limited number of components for the modes. The matrix formulation of the Tucker3 model in terms of the statistical unit-mode matricization $\mathbf{X}_{\mathrm{A}}$ is

$$\mathbf{X}_{\mathrm{A}} = \mathbf{A}\mathbf{G}_{\mathrm{A}}(\mathbf{C} \otimes \mathbf{B})' + \mathbf{E}_{\mathrm{A}}, \tag{1}$$

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

where the symbol '$\otimes$' denotes the Kronecker product of matrices and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ have order $(I \times P)$, $(J \times Q)$, $(K \times R)$, respectively. The Tucker3 model can be seen as a generalization of PCA where the JK columns of $\mathbf{X}_{\mathrm{A}}$, J variables measured at K different occasions, are modeled through the linear combinations of P latent components for the statistical units (the columns of $\mathbf{A}$), with weights $\mathbf{G}_{\mathrm{A}}(\mathbf{C} \otimes \mathbf{B})'$. By exploiting the symmetry of the model with respect to the modes, we derive that P, Q and R are the number of components for the statistical units, the manifest variables and the occasions, respectively, while the elements of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are the scores of the entities of the various modes on the corresponding components. Finally, $\mathbf{G}_{\mathrm{A}}$ is the statistical unit-mode matricization, of order $(P \times QR)$, of the so-called core tensor $\underline{\mathbf{G}}$, of order $(P \times Q \times R)$, expressing the strength of the triple interactions among the components of the three modes. Like PCA, the model parameters are estimated in the OLS sense by minimizing $\|\mathbf{E}_{\mathrm{A}}\|^2$, where $\|\cdot\|$ is the Frobenius norm of matrices. The solutions are not obtained by the singular value decomposition (SVD, Eckart & Young, 1936) of a particular matrix as is for PCA. For this purpose, alternating least squares (ALS) algorithms can be applied. See, for instance, Kroonenberg & De Leeuw (1980). As a possible extension of SVD to the multiway case, we mention the higher-order singular value decomposition (De Lathauwer, De Moor, & Vandewalle, 2000).

The applicability of the Tucker3 model may be limited for several reasons. First of all, the interpretability of the solution, in particular of the elements of the core tensor, is usually a very complex issue. Moreover, differently from standard PCA, the solutions are not nested. Finally, the obtained solution is not unique. In fact, it is straightforward to see that we can post-multiply $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ by square non-singular transformation matrices of appropriate order, say $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$, obtaining the new component matrices $\mathbf{A}_{\mathrm{T}} = \mathbf{AP}$, $\mathbf{B}_{\mathrm{T}} = \mathbf{BQ}$ and $\mathbf{C}_{\mathrm{T}} = \mathbf{CR}$. If we compensate such transformations in the core by setting $\mathbf{G}_{\mathrm{AT}} = \mathbf{P}^{-1}\mathbf{G}_{\mathrm{A}}(\mathbf{R}^{-1} \otimes \mathbf{Q}^{-1})'$, we get an equally well-fitting solution because

$$\mathbf{A}_{\mathrm{T}}\mathbf{G}_{\mathrm{AT}}(\mathbf{C}_{\mathrm{T}} \otimes \mathbf{B}_{\mathrm{T}})' = \mathbf{APP}^{-1}\mathbf{G}_{\mathrm{A}}(\mathbf{R}^{-1} \otimes \mathbf{Q}^{-1})'(\mathbf{R} \otimes \mathbf{Q})'(\mathbf{C} \otimes \mathbf{B})' = \mathbf{AG}_{\mathrm{A}}(\mathbf{C} \otimes \mathbf{B})'. \quad (2)$$

To overcome the possible limitations of the Tucker3 model, it may be convenient to consider the Parafac one, independently proposed by Carroll & Chang (1970) and Harshman (1970). The Parafac model can be seen as a constrained version of the Tucker3 one where the same number of components, say S, is sought for all the modes $(P = Q = R = S)$ and the core tensor is equal to the identity tensor ($\underline{\mathbf{G}} = \underline{\mathbf{I}}$, i.e., $g_{pqr} = 1$, if $p = q = r$, and $g_{pqr} = 0$, otherwise). We have

$$\mathbf{X}_{\mathrm{A}} = \mathbf{AI}_{\mathrm{A}}(\mathbf{C} \otimes \mathbf{B})' + \mathbf{E}_{\mathrm{A}} = \mathbf{A}(\mathbf{C} \bullet \mathbf{B})' + \mathbf{E}_{\mathrm{A}}, \quad (3)$$

where the symbol '$\bullet$' denotes the Khatri–Rao product of matrices, i.e., it is $\mathbf{C} \bullet \mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{c}_S \otimes \mathbf{b}_S]$, where $\mathbf{b}_s$ and $\mathbf{c}_s$ are the sth columns of $\mathbf{B}$ and $\mathbf{C}$, respectively $(s = 1, \ldots, S)$. As for the Tucker3 case, the parameter estimates are found in the OLS sense by ALS algorithms.

The most interesting feature of Parafac is that under mild conditions, the solution is unique. This point has been deeply investigated by Kruskal (1977), who has found the following result. Let us denote by k-rank($\mathbf{Z}$) the so-called k-rank of a matrix $\mathbf{Z}$. It is defined as the largest number k such that every subset of k columns of $\mathbf{Z}$ is linearly independent. Moreover, let ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) and ($\mathbf{A}_{\mathrm{T}}$, $\mathbf{B}_{\mathrm{T}}$, $\mathbf{C}_{\mathrm{T}}$) be two Parafac solutions. Kruskal (1977) has shown that if

$$k\text{-rank}(\mathbf{A}) + k\text{-rank}(\mathbf{B}) + k\text{-rank}(\mathbf{C}) \geq 2S + 2 \quad (4)$$

then, by considering (3),

$$\mathbf{A}(\mathbf{C} \bullet \mathbf{B})' = \mathbf{A}_{\mathrm{T}}(\mathbf{C}_{\mathrm{T}} \bullet \mathbf{B}_{\mathrm{T}})' \quad (5)$$

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

implies that there exist a permutation matrix $\mathbf{P}$ and three diagonal matrices $\mathbf{D}_A$, $\mathbf{D}_B$ and $\mathbf{D}_C$, for which $\mathbf{D}_A\mathbf{D}_B\mathbf{D}_C = \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix, such that

$$\mathbf{A}_T = \mathbf{APD}_A, \mathbf{B}_T = \mathbf{BPD}_B, \mathbf{C}_T = \mathbf{CPD}_C. \tag{6}$$

In other words, if (4) holds, then the solution ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) is unique up to scaling and a simultaneous column permutation. Although Kruskal's condition has been extended by some other authors (Jiang & Sidiropoulos, 2004; Stegeman, ten Berge, & De Lathauwer, 2006; Stegeman, 2009; Domanov & De Lathauver, 2013a; Domanov & De Lathauver, 2013b), what follows is based on such a condition because practitioners mainly refer to it in their applications. In particular, without loss of generality, we set the scaling of the factor loading matrices by assuming that $\mathbf{B}$ and $\mathbf{C}$ are column-wise normalized, i.e.,

$$(\mathbf{B}'\mathbf{B}) * \mathbf{I} = (\mathbf{C}'\mathbf{C}) * \mathbf{I} = \mathbf{I}, \tag{7}$$

where '*' denotes the Hadamard product, i.e., element-wise product of matrices. Therefore, (4) and (5) imply that there exists a permutation matrix $\mathbf{P}$ such that

$$\mathbf{A}_T = \mathbf{AP},\ \mathbf{B}_T = \mathbf{BP},\ \mathbf{C}_T = \mathbf{CP}. \tag{8}$$

It is important to note that parameter estimates are not scale invariant. A rescaling of the data does not imply only an analogous rescaling of the estimates. It follows that there exists a way to rescale the data better, or more appropriate, than another, making crucial the choice. This is witnessed by a remarkably large number of papers devoted to the preprocessing step. See, for instance, Harshman & Lundy (1984), Bro & Smilde (2003), Kiers (2006) and references therein. The goal of preprocessing is similar to that for standard PCA. Namely, data are standardized in order to eliminate unwanted differences among the variables. Difficulties arise because it is no longer obvious what is the best strategy to adopt. For instance, a three-way three-mode array can be normalized within one of the three modes or even within a combination of two modes. Different ways of preprocessing the data lead to different solutions, and a wrong preprocessing may lead to unfeasible results (see, e.g., Kiers, 2006). Obviously, a scale-invariant model would be simpler to apply in practice especially for non-expert users because such a crucial decision on the best preprocessing strategy would be no longer necessary.

For a thorough discussion on three-way three-mode analysis according to a component-based approach, see the monographies of Bro (1998), Smilde, Bro, & Geladi (2004) and Kroonenberg (2008) and the reviews of Acar & Yener (2009), Kolda & Bader (2009), Mørup (2011) and Giordani & Kiers (2018).

3. The Structural Approach

3.1. Models

Starting from the original formulations of the models, we can derive what is the covariance structure corresponding to each model. Let us consider the Tucker3 model in (1) by limiting our attention to the ith row of $\mathbf{X}_A$, say $\mathbf{x}'_{Ai}$, pertaining to the ith statistical unit. $\mathbf{x}'_{Ai}$ is the vector of length JK containing the scores of statistical unit i on the J manifest variables during the K occasions. We get

$$\mathbf{x}'_{Ai} = \mathbf{a}'_i\mathbf{G}_A(\mathbf{C} \otimes \mathbf{B})' + \mathbf{e}'_{Ai}, \tag{9}$$

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

where with obvious notation, $\mathbf{e}'_{Ai}$ represents the ith row of $\mathbf{E}_{\mathrm{A}}$ and $\mathbf{a}'_i$ is the ith row of $\mathbf{A}$ containing the component scores for statistical unit i. To simplify the notation, we omit the subscripts 'A' and 'i,' and we rewrite (9) in terms of column vectors, explicitly considering a vector of intercepts

$$\mathbf{x} = \boldsymbol{\mu} + (\mathbf{C} \otimes \mathbf{B})\mathbf{G}'\mathbf{a} + \mathbf{e}. \tag{10}$$

As usual in standard FA, we assume that the common factors $\mathbf{a}$ and the specific factors $\mathbf{e}$ are random with $E(\mathbf{a}) = \mathbf{0}$ and $E(\mathbf{e}) = \mathbf{0}$, without loss of generality because of $\boldsymbol{\mu}$, and $E(\mathbf{a}\mathbf{e}') = \mathbf{0}$. If $E(\mathbf{a}\mathbf{a}') = \boldsymbol{\Phi}$ and $E(\mathbf{e}\mathbf{e}') = \boldsymbol{\Psi}$, then the covariance matrix of $\mathbf{x}$ is given by

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}'\boldsymbol{\Phi}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})' + \boldsymbol{\Psi}. \tag{11}$$

In what follows, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ will be assumed to be positive definite. The generic element of the matrix $\boldsymbol{\Sigma}$ (of order $JK \times JK$), $\sigma_{jk,j'k'}$, holds the covariance between manifest variable j at occasion k and manifest variable j' at occasion k' ($j, j' = 1, \ldots, J; k, k' = 1, \ldots, K$). Bearing in mind the standard FA model, it should be clear that the Tucker3 model is a constrained version of standard FA. If we set $\boldsymbol{\Lambda} = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}'$, then (11) can be rewritten as

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}, \tag{12}$$

which is the oblique FA model, where $\boldsymbol{\Lambda} = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}'$ is the matrix of factor loadings having a particular form taking into account the three-way three-mode structure of the data.

Several papers available in the literature investigate the model in (11). Interested readers may refer to Bloxom (1968), Bentler & Lee (1978), Bentler & Lee (1979), Lee & Fong (1983), Bloxom (1984), Bentler, Poon, & Lee (1988), Kroonenberg & Oort (2003) and references therein. Other multilinear decompositions with Kronecker structured covariance matrices are presented and investigated by, for instance, Gerard & Hoff (2015) and Hoff (2016).

The Parafac model can be derived from (10) setting $\mathbf{G} = \mathbf{G}_{\mathrm{A}} = \mathbf{I}_{\mathrm{A}}$ as in (3). We have

$$\mathbf{x} = \boldsymbol{\mu} + (\mathbf{C} \bullet \mathbf{B})\mathbf{a} + \mathbf{e}. \tag{13}$$

This shows that even the Parafac model is the same as the classical FA model, where the observed variables are expressed as a linear combination of a limited number of common factors ($\mathbf{a}$) and specific factors ($\mathbf{e}$). In order to take into account the three-way three-mode structure of the data, the loadings are constrained to be equal to $\mathbf{C} \bullet \mathbf{B}$. Similarly, the structural version of Parafac can be derived from (11) imposing the restriction $\mathbf{G} = \mathbf{G}_{\mathrm{A}} = \mathbf{I}_{\mathrm{A}}$ obtaining

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = (\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}. \tag{14}$$

Comparing (10) with (13), we see that the restriction $\mathbf{G} = \mathbf{G}_{\mathrm{A}} = \mathbf{I}_{\mathrm{A}}$ greatly simplifies model interpretation. This aspect can be highlighted by focusing on a single occasion, say k. Let $\mathbf{x}_k$ be the vector of the J manifest variables measured at occasion k, with obvious notation, the Tucker3 and Parafac assume the form

$$\mathbf{x}_k = \boldsymbol{\mu}_k + \mathbf{B}\left(\sum\nolimits_r c_{kr}\mathbf{G}_r\right)'\mathbf{a} + \mathbf{e}_k, \tag{15}$$

$$\mathbf{x}_k = \boldsymbol{\mu}_k + \mathbf{B}\mathrm{diag}(\mathbf{c}_k)\mathbf{a} + \mathbf{e}_k, \tag{16}$$

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

respectively, where diag($\mathbf{z}$) is the diagonal matrix with the elements of $\mathbf{z}$ on the main diagonal. Under the Tucker3 model, $\mathbf{B}$ may be interpreted as the matrix of factor loadings for the variables on the common factors $(\boldsymbol{\Sigma}_r\, c_{kr}\mathbf{G}_r)'\mathbf{a}$ varying the covariance structure across the occasions. Even under the Parafac, the factor loadings for the variables are the same ($\mathbf{B}$) in each occasion, but the common factors diag($\mathbf{c}_k$)$\mathbf{a}$ vary only in variance. By exploiting the symmetry of the models with respect to the modes, we may interpret $\mathbf{C}$ as the factor loading matrix for the occasions.

In the following section, we reconsider the structural Parafac model by analyzing whether the constraints $\boldsymbol{\Lambda} = \mathbf{C} \bullet \mathbf{B}$ affect the parameter identifiability under different covariance structure of the specific factors.

3.2. Scale Invariance

The practical applicability of FA is favored by the property of scale invariance. As is well known, the factorial structure of the standard FA model is not destroyed by variable rescaling. For example, if $\mathbf{x}$ is rescaled as $\mathbf{x}_{\mathrm{L}} = \mathbf{L}\mathbf{x}$ where $\mathbf{L}$ is a diagonal matrix, then the covariance matrix of $\mathbf{x}_{\mathrm{L}}$, $\boldsymbol{\Sigma}_{\mathrm{L}}$, is equal to

$$\boldsymbol{\Sigma}_{\mathrm{L}} = V(\mathbf{L}\mathbf{x}) = \mathbf{L}\mathrm{V}(\mathbf{x})\mathbf{L} = \mathbf{L}(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi})\mathbf{L} = (\mathbf{L}\boldsymbol{\Lambda})\boldsymbol{\Phi}(\mathbf{L}\boldsymbol{\Lambda})' + \mathbf{L}\boldsymbol{\Psi}\mathbf{L} = \boldsymbol{\Lambda}_{\mathrm{L}}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{\mathrm{L}}' + \boldsymbol{\Psi}_{\mathrm{L}}, \tag{17}$$

with $\boldsymbol{\Lambda}_{\mathrm{L}} = \mathbf{L}\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}_{\mathrm{L}} = \mathbf{L}\boldsymbol{\Psi}\mathbf{L}$. Hence, the factor loadings are the same but expressed in the new units of measurements. This guarantees that when the model is estimated by a scale-invariant method, like maximum likelihood, a rescaling of the data implies an analogous rescaling of the estimates. If $\mathbf{L}$ contains standard deviations, $\boldsymbol{\Sigma}_{\mathrm{L}}$ is the correlation matrix of $\mathbf{x}$ and standard FA applied to covariance or correlation matrices leads to the same conclusions. This is not true for PCA. The components extracted by using $\boldsymbol{\Sigma}_{\mathrm{L}}$ or $\boldsymbol{\Sigma}$ differ, and therefore, a crucial question to answer before applying PCA is how to preprocess the data.

In the structural case, the Parafac model introduced in (13), and the resulting covariance matrix given in (14), is generally not scale invariant in the sense given by Browne (1982). In fact, if $\mathbf{x}$ in (13) is rescaled as $\mathbf{x}_{\mathrm{L}} = \mathbf{L}\mathbf{x}$, then

$$\boldsymbol{\Sigma}_{\mathrm{L}} = \mathbf{L}[(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}]\mathbf{L} = [\mathbf{L}(\mathbf{C} \bullet \mathbf{B})]\boldsymbol{\Phi}[\mathbf{L}(\mathbf{C} \bullet \mathbf{B})]' + \boldsymbol{\Psi}_{\mathrm{L}}; \tag{18}$$

it follows that, in general, the structural Parafac model in (14) is not scale invariant because the equation

$$\mathbf{L}(\mathbf{C} \bullet \mathbf{B}) = \mathbf{C}_{\mathrm{L}} \bullet \mathbf{B}_{\mathrm{L}} \tag{19}$$

does not always have a solution. In other terms, the matrix $\mathbf{L}$ usually destroys the Khatri–Rao structure of the factor loadings unless $\mathbf{L}$ has a particular structure.

Bearing in mind that the same variables are collected at different occasions, it may be reasonable to perform the same scaling of the variables in the various occasions and the same scaling of the occasions across the different variables. This is equivalent to say that

$$\mathbf{L} = (\mathbf{L}_{\mathrm{O}} \otimes \mathbf{L}_{\mathrm{V}}), \tag{20}$$

where $\mathbf{L}_{\mathrm{V}} = \mathrm{diag}([l_{\mathrm{V}1}, \ldots, l_{\mathrm{V}j}, \ldots, l_{\mathrm{V}J}]')$ and $\mathbf{L}_{\mathrm{O}} = \mathrm{diag}([l_{\mathrm{O}1}, \ldots, l_{\mathrm{O}k}, \ldots, l_{\mathrm{O}K}]')$. The generic element $l_{\mathrm{V}j}$ gives the scaling of variable j ($j = 1, \ldots, J$). Similarly, $l_{\mathrm{O}k}$ represents

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

the scaling of occasion k ($k = 1, \ldots, K$). It follows that x_{jk}, i.e., variable j at occasion k, is rescaled as $l_{Vj} l_{Ok} x_{jk}$. In this particular case, by substituting (20) into (19), we get

$$\mathbf{L}(\mathbf{C} \bullet \mathbf{B}) = (\mathbf{L}_O \otimes \mathbf{L}_V)(\mathbf{C} \bullet \mathbf{B}) = (\mathbf{L}_O\mathbf{C} \bullet \mathbf{L}_V\mathbf{B}) = \mathbf{C}_L \bullet \mathbf{B}_L, \tag{21}$$

with $\mathbf{B}_L = \mathbf{L}_V\mathbf{B}$ and $\mathbf{C}_L = \mathbf{L}_O\mathbf{C}$. Therefore, the scaling is absorbed by the Khatri–Rao product and the particular factor loading structure is not destroyed.

However, in general, the structural Parafac model is not scale invariant and a researcher may be interested in knowing how to modify its structure to derive a scale-invariant version. Following Lee & Fong (1983), the Parafac formulation in (13) can be modified by incorporating a positive definite diagonal matrix $\mathbf{D}$ in the model able to absorb scale changes. We get

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{D}[(\mathbf{C} \bullet \mathbf{B})\mathbf{a} + \mathbf{e}]. \tag{22}$$

From (22), the covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \mathbf{D}[(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}]\mathbf{D}. \tag{23}$$

In this way, the Parafac model becomes scale invariant. In fact, when $\mathbf{x}$ is rescaled as $\mathbf{x}_L = \mathbf{L}\mathbf{x}$, (22) becomes

$$\mathbf{x}_L = \mathbf{L}\mathbf{x} = \mathbf{L}\{\boldsymbol{\mu} + \mathbf{D}[(\mathbf{C} \bullet \mathbf{B})\mathbf{a} + \mathbf{e}]\} = \boldsymbol{\mu}_L + \mathbf{D}_L[(\mathbf{C} \bullet \mathbf{B})\mathbf{a} + \mathbf{e}], \tag{24}$$

with $\boldsymbol{\mu}_L = \mathbf{L}\boldsymbol{\mu}$ and $\mathbf{D}_L = \mathbf{L}\mathbf{D}$. The covariance matrix of $\mathbf{x}_L$ is

$$\boldsymbol{\Sigma}_L = \mathbf{D}_L[(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}]\mathbf{D}_L. \tag{25}$$

Hence, the model is not affected by $\mathbf{L}$ because the scaling is absorbed in the diagonal matrix $\mathbf{D}_L$. It is important to note that if we impose a particular structure on $\mathbf{D}$, then the same should be on $\mathbf{D}_L = \mathbf{L}\mathbf{D}$ for every possible $\mathbf{L}$. It follows that we cannot impose any particular structure on $\mathbf{D}$. In order to identify the diagonal matrix $\mathbf{D}$, either the constraint $((\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}) * \mathbf{I} = \mathbf{I}$ (Lee & Fong, 1983) or $\boldsymbol{\Psi} = \mathbf{I}$ (Bentler, Poon, & Lee, 1988) is imposed. The two formulations are not equivalent. The former allows us to interpret the model as a reparameterization of the correlation matrix, and the latter has an interpretation less clear but avoids complex constraints on the estimates of the parameter matrices except for $\boldsymbol{\Psi}$.

4. Factor Uniqueness of the Structural Parafac Model

A fundamental property of the Parafac model is its factor uniqueness. In this section, we start showing that Parafac maintains this property even in the structural formulation when, as in the standard FA model in (12), the specific factors are assumed to be uncorrelated, i.e., the matrix $\boldsymbol{\Psi}$ is diagonal. However, in three-way three-mode FA, we note that the uncorrelation of the specific factors might be a too restrictive assumption. For instance, it might be reasonable to assume that $\boldsymbol{\Psi}$ is block diagonal, i.e., the specific factors of the different variables are correlated within the same occasion

$$\boldsymbol{\Psi} = \text{blockdiag}(\boldsymbol{\Psi}_{11}, \ldots, \boldsymbol{\Psi}_{kk}, \ldots, \boldsymbol{\Psi}_{KK}), \tag{26}$$

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

where blockdiag$(\mathbf{Z}_1, \ldots, \mathbf{Z}_K)$ denotes the block-diagonal matrix with diagonal elements equal to the square matrices $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$ and $\mathbf{\Psi}_{kk}$ denotes the covariance matrix of order $(J \times J)$ for the specific factors at occasion k, $k = 1, \ldots, K$. The Parafac covariance model in (14) with the correlation structure of the specific factors given in (26) represents a more realistic model able to fit reasonably well in many practical situations.

To interpret/understand correlations between specific factors, we note that they are allowed only between specific factors in the same occasion. In other words, the idea is that correlations between variables measured in different occasions are due only to common factors, while correlations between variables in the same occasion can also be due to other factors (i.e., the specific factors in that occasion) that are not present in other occasions and therefore out of our interest. An accurate estimation of the 'true'/interesting common factors should consider explicitly the presence of such factors that are in 'common' only with the variables measured in the same occasion. From a statistical point of view, such covariances are only nuisance parameters. In the literature, the assumption of correlated specific factors is not new. See, e.g., Browne (1984a) and Kroonenberg & Oort (2003).

In the following, we investigate under what conditions factor uniqueness holds even with a block-diagonal structure for $\mathbf{\Psi}$. Finally, we address the scale invariance problem studying the factor uniqueness property of the scale-invariant version of Parafac in (22–23).

It is important to note that what follows can be extended to the case where the specific factors of the same variable are correlated across the different occasions. Such an extension can be easily obtained by exploiting the symmetry of the model with respect to variables and occasions.

4.1. $\mathbf{\Psi}$ Diagonal

Let us start by showing the factor uniqueness property of the structural Parafac when $\mathbf{\Psi}$ is diagonal. Note that disjoint submatrices are submatrices having no entry in common.

Result 1 Suppose that the following hold:

$$\text{(a)} \quad (\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \mathbf{\Psi} = (\mathbf{C}_\mathrm{T} \bullet \mathbf{B}_\mathrm{T})\,\mathbf{\Phi}_\mathrm{T}(\mathbf{C}_\mathrm{T} \bullet \mathbf{B}_\mathrm{T})' + \mathbf{\Psi}_\mathrm{T}, \tag{27}$$

$$\text{(b)} \quad \text{k-rank}(\mathbf{B}) + \text{k-rank}(\mathbf{C}) \geq \mathrm{S} + 2, \tag{28}$$

(c) if any row of $\mathbf{C} \bullet \mathbf{B}$ is deleted, there remain two disjoint submatrices of rank S,

where $\mathbf{\Psi}$ and $\mathbf{\Psi}_\mathrm{T}$ are diagonal matrices and all the remaining ones have S columns.

Under the scaling set in (7), there exists a permutation matrix $\mathbf{P}$ such that

$$\mathbf{B}_\mathrm{T} = \mathbf{BP}, \mathbf{C}_\mathrm{T} = \mathbf{CP}, \mathbf{\Phi}_\mathrm{T} = \mathbf{P\Phi P}. \tag{29}$$

Proof Let us rewrite equality (a) in terms of standard FA models with uncorrelated common factors

$$\mathbf{\Sigma} = \left[(\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}^{1/2}\right]\left[(\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}^{1/2}\right]' + \mathbf{\Psi} = \left[(\mathbf{C}_\mathrm{T} \bullet \mathbf{B}_\mathrm{T})\mathbf{\Phi}_\mathrm{T}^{1/2}\right]\left[(\mathbf{C}_\mathrm{T} \bullet \mathbf{B}_\mathrm{T})\mathbf{\Phi}_\mathrm{T}^{1/2}\right]' + \mathbf{\Psi}_\mathrm{T}. \tag{30}$$

Condition (c) implies that if any row of $(\mathbf{C} \bullet \mathbf{B})\,\mathbf{\Phi}^{1/2}$ is deleted, there remain two disjoint submatrices of rank S since rank$(\mathbf{\Phi})$ = rank$(\mathbf{\Phi}^{1/2}) = S$. From Theorem 5.1 of Anderson & Rubin

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

(1956), we know that when this is true, there exists a column-wise orthonormal matrix $\mathbf{T}$ such that

$$(\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}^{1/2} = (\mathbf{C}_{\mathrm{T}} \bullet \mathbf{B}_{\mathrm{T}})\mathbf{\Phi}_{\mathrm{T}}^{1/2}\mathbf{T}. \tag{31}$$

Inequality (b) implies

$$k\text{-rank}\left(\mathbf{\Phi}^{1/2}\right) + k\text{-rank}(\mathbf{B}) + k\text{-rank}(\mathbf{C}) \geq 2S + 2, \tag{32}$$

indicating that Kruskal's condition is met and the two decompositions in (31) are equal up to a column permutation under constraints (7). Let us indicate with $\mathbf{P}$ such a permutation matrix, we have

$$\mathbf{B}_{\mathrm{T}} = \mathbf{BP},\ \mathbf{C}_{\mathrm{T}} = \mathbf{CP},\ \mathbf{\Phi}_{\mathrm{T}}^{1/2} = \mathbf{T}'\mathbf{\Phi}^{1/2}\mathbf{P} \Rightarrow\ \mathbf{\Phi}_{\mathrm{T}} = \mathbf{P}'\mathbf{\Phi}\mathbf{P}. \tag{33}$$

□

Finally, we note that, in the previous proposition, Kruskal's condition can be replaced with any other sufficient condition available in the literature.

4.2. $\mathbf{\Psi}$ Block Diagonal

When $\mathbf{\Psi}$ is block diagonal, the conditions of Anderson & Rubin (1956) cannot be longer applied as it assumes a diagonal structure for $\mathbf{\Psi}$. To prove the factor uniqueness, the following proposition is considered.

Proposition (Browne, 1980) *Let us consider the decomposition $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$, where $\mathbf{\Lambda}$ is partitioned as $\mathbf{\Lambda} = [\mathbf{\Lambda}_1', \ldots, \mathbf{\Lambda}_k', \ldots, \mathbf{\Lambda}_K']'$ and $\mathbf{\Psi}$ has the corresponding block-diagonal structure $\mathbf{\Psi} = blockdiag(\mathbf{\Psi}_{11}, \ldots, \mathbf{\Psi}_{kk}, \ldots, \mathbf{\Psi}_{KK})$. The identification of $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ up to post-multiplication by an orthogonal matrix $\mathbf{T}$ holds if at least three of the submatrices $\mathbf{\Lambda}_k$ are of full column rank.*

The above proposition has been formulated in the context of the FA model for multiple batteries of tests. The matrices $\mathbf{\Lambda}_k$ and $\mathbf{\Psi}_{kk}$ refer to the kth battery. The factor loading matrix is obtained by juxtaposing the factor loading matrices of the various batteries, and the specific factors are uncorrelated between the batteries, but correlated within the batteries. The proposition can be applied in order to prove the factor uniqueness of the structural Parafac model when the specific factors are correlated within the occasions. A sufficient condition for the factor uniqueness of the Parafac solution is given in Result 2.

Result 2 Suppose that the following hold:

$$\text{(a)} \quad (\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \mathbf{\Psi} = (\mathbf{C}_{\mathrm{T}} \bullet \mathbf{B}_{\mathrm{T}})\mathbf{\Phi}_{\mathrm{T}}(\mathbf{C}_{\mathrm{T}} \bullet \mathbf{B}_{\mathrm{T}})' + \mathbf{\Psi}_{\mathrm{T}}, \tag{34}$$

$$\text{(b)} \quad k\text{-rank}(\mathbf{B}) + k\text{-rank}(\mathbf{C}) \geq S + 2, \tag{35}$$

$$\text{(c)} \quad \text{rank}(\mathbf{B}\text{diag}(\mathbf{c}_k)) = S \text{ for at least three occasions}, \tag{36}$$

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

where $\boldsymbol{\Psi} = \text{blockdiag}(\boldsymbol{\Psi}_{11}, \ldots, \boldsymbol{\Psi}_{kk}, \ldots, \boldsymbol{\Psi}_{KK})$ is block diagonal with blocks defined by the occasions, $\boldsymbol{\Psi}_{\text{T}}$ has the same structure, all the remaining matrices have S columns, and $\mathbf{c}_k$ is the vector containing the elements of the kth row of $\mathbf{C}$.

Under the scaling set in (7), there exists a permutation matrix $\mathbf{P}$ such that

$$\mathbf{B}_{\text{T}} = \mathbf{BP}, \mathbf{C}_{\text{T}} = \mathbf{CP}, \boldsymbol{\Phi}_{\text{T}} = \mathbf{P}\boldsymbol{\Phi}\mathbf{P}. \tag{37}$$

Proof Let us rewrite equality (a) in terms of standard FA models with uncorrelated common factors

$$\boldsymbol{\Sigma} = \left[(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}^{1/2}\right]\left[(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}^{1/2}\right]' + \boldsymbol{\Psi} = \left[(\mathbf{C}_{\text{T}} \bullet \mathbf{B}_{\text{T}})\boldsymbol{\Phi}_{\text{T}}^{1/2}\right]\left[(\mathbf{C}_{\text{T}} \bullet \mathbf{B}_{\text{T}})\boldsymbol{\Phi}_{\text{T}}^{1/2}\right]' + \boldsymbol{\Psi}_{\text{T}}, \tag{38}$$

where the matrix of factor loadings $(\mathbf{C} \bullet \mathbf{B})\,\boldsymbol{\Phi}^{1/2}$, if partitioned according to the blocks of $\boldsymbol{\Psi}$, takes the form $(\mathbf{C} \bullet \mathbf{B})\,\boldsymbol{\Phi}^{1/2} = [(\mathbf{B}\text{diag}(\mathbf{c}_1)\boldsymbol{\Phi}^{1/2})', \ldots, (\mathbf{B}\text{diag}(\mathbf{c}_k)\boldsymbol{\Phi}^{1/2})', \ldots, (\mathbf{B}\text{diag}(\mathbf{c}_K)\boldsymbol{\Phi}^{1/2})']'$. Condition (c) implies $\text{rank}(\mathbf{B}\text{diag}(\mathbf{c}_k)\boldsymbol{\Phi}^{1/2}) = S$ for at least three occasions since $\text{rank}(\boldsymbol{\Phi}) = \text{rank}(\boldsymbol{\Phi}^{1/2}) = S$. This allows us to use Browne's proposition in order to show that (38) implies the existence of a column-wise orthonormal matrix $\mathbf{T}$ such that

$$(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}^{1/2} = (\mathbf{C}_{\text{T}} \bullet \mathbf{B}_{\text{T}})\boldsymbol{\Phi}_{\text{T}}^{1/2}\mathbf{T}. \tag{39}$$

Inequality (b) implies

$$k\text{-rank}\left(\boldsymbol{\Phi}^{1/2}\right) + k\text{-rank}(\mathbf{B}) + k\text{-rank}(\mathbf{C}) \geq 2S + 2, \tag{40}$$

indicating that Kruskal's condition is met and the two decompositions in (39) are equal up to a column permutation under constraints (7). Let us indicate with $\mathbf{P}$ such a permutation matrix, we have

$$\mathbf{B}_{\text{T}} = \mathbf{BP}, \mathbf{C}_{\text{T}} = \mathbf{CP}, \boldsymbol{\Phi}_{\text{T}}^{1/2} = \mathbf{T}'\boldsymbol{\Phi}^{1/2}\mathbf{P} \Rightarrow \boldsymbol{\Phi}_{\text{T}} = \mathbf{P}'\boldsymbol{\Phi}\mathbf{P}. \tag{41}$$

□

4.3. Scale Invariance

It remains to prove that the scale-invariant versions of the structural Parafac model in (25), with constraint (a) $((\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}) * \mathbf{I} = \mathbf{I}$ (Lee & Fong, 1983) or (b) $\boldsymbol{\Psi} * \mathbf{I} = \mathbf{I}$ (Bentler, Poon, & Lee, 1988), satisfy the factor uniqueness property. Note that the latter constraint is modified with respect to the original version to also handle the case where $\boldsymbol{\Psi}$ is block diagonal. The following results hold for either of the two cases where $\boldsymbol{\Psi}$ is diagonal or block diagonal.

Let us consider the equality

$$\mathbf{D}[(\mathbf{C} \bullet \mathbf{B})\boldsymbol{\Phi}(\mathbf{C} \bullet \mathbf{B})' + \boldsymbol{\Psi}]\mathbf{D} = \mathbf{D}_{\text{T}}[(\mathbf{C}_{\text{T}} \bullet \mathbf{B}_{\text{T}})\boldsymbol{\Phi}_{\text{T}}(\mathbf{C}_{\text{T}} \bullet \mathbf{B}_{\text{T}})' + \boldsymbol{\Psi}_{\text{T}}]\mathbf{D}_{\text{T}} \tag{42}$$

In order to show that the previous results hold even in this case, it is sufficient to show that the use of either constraint (a) or (b) implies $\mathbf{D} = \mathbf{D}_{\text{T}}$.

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

In case (a), by focusing the attention on the main diagonal elements of the covariance matrices on the left and right of equality (42), we deduce that

$$\mathbf{DD} = \mathbf{D}_{\mathrm{T}}\mathbf{D}_{\mathrm{T}} \Rightarrow \mathbf{D} = \mathbf{D}_{\mathrm{T}} \tag{43}$$

because $\mathbf{D}$ and $\mathbf{D}_{\mathrm{T}}$ are required to be positive definite.

Case (b) is more complex to show because now the constraint applies only on the covariance matrix of the specific factors and it becomes effective in the identification only if the decomposition of the total covariance is unique. To this end, we rewrite (42) in terms of FA models. Namely, setting $\mathbf{\Lambda} = \mathbf{D}(\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}^{1/2}$ and $\mathbf{Z} = \mathbf{D}\mathbf{\Psi}\mathbf{D}$, with obvious notation, we get

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{Z} = \mathbf{\Lambda}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}' + \mathbf{Z}_{\mathrm{T}}. \tag{44}$$

If the decomposition of the total covariance into variability due to the common factors and variability due to the specific factors is unique, then it must be $\mathbf{Z} = \mathbf{Z}_{\mathrm{T}}$, i.e., $\mathbf{D}\mathbf{\Psi}\mathbf{D} = \mathbf{D}_{\mathrm{T}}\mathbf{\Psi}_{\mathrm{T}}\mathbf{D}_{\mathrm{T}}$; hence, $\mathbf{D} = \mathbf{D}_{\mathrm{T}}$ because $\mathbf{\Psi} * \mathbf{I} = \mathbf{\Psi}_{\mathrm{T}} * \mathbf{I} = \mathbf{I}$. It remains to discuss under what conditions the total covariance decomposition is unique. In particular, we are going to show that this is obtained when condition (c) in either Result 1 or 2 is satisfied.

In the case of a diagonal $\mathbf{\Psi}$, when condition (c) of Result 1 is met, if any row of $\mathbf{\Lambda} = \mathbf{D}(\mathbf{C} \bullet \mathbf{B})\mathbf{\Phi}^{1/2}$ is deleted, there remain two disjoint submatrices of rank S. This allows us to apply Theorem 5.1 of Anderson & Rubin (1956) obtaining the required factor uniqueness.

In the case of a block diagonal $\mathbf{\Psi}$, when condition (c) of Result 2 is met, at least three submatrices of $\mathbf{D}(\mathbf{C} \bullet \mathbf{B})\,\mathbf{\Phi}^{1/2} = [(\mathbf{D}_1\mathbf{B}\mathrm{diag}(\mathbf{c}_1)\,\mathbf{\Phi}^{1/2})', \ldots, (\mathbf{D}_k\mathbf{B}\mathrm{diag}(\mathbf{c}_k)\,\mathbf{\Phi}^{1/2})', \ldots, (\mathbf{D}_K\mathbf{B}\mathrm{diag}(\mathbf{c}_K)\,\mathbf{\Phi}^{1/2})']'$, where $\mathbf{D} = \mathrm{blockdiag}(\mathbf{D}_1, \ldots, \mathbf{D}_k, \ldots, \mathbf{D}_K)$, are of full column rank. This allows us to apply Browne's proposition obtaining the required factor uniqueness.

5. Estimation of the Structural Parafac Model

The parameter estimation of the structural Parafac model can be done in several ways. Ordinary least squares (OLS), generalized least squares (GLS), maximum likelihood (ML) approaches can be applied. They differ in the discrepancy function to be minimized between the sample covariance matrix and the one induced by the Parafac model. If observations are independent and identically distributed as a multinormal distribution, the estimate is obtained in such a way to minimize

$$f(\theta) = \mathrm{tr}\{[\mathbf{S} - \mathbf{\Sigma}(\theta)]\mathbf{W}^{-1}\}^2/2 \tag{45}$$

where $\theta = \{\mathbf{B}, \mathbf{C}, \mathbf{\Phi}, \mathbf{\Psi}\}$ denotes the set of parameters, $\mathbf{S}$ is the sample covariance matrix of order $(JK \times JK)$, $\mathbf{\Sigma}(\theta)$ is the estimated covariance matrix according to the Parafac model, and $\mathbf{W}$ is a weight matrix. If $\mathbf{W} = \mathbf{I}$ in (45), then the OLS estimate is found. The OLS approach is, however, not recommended because it is not scale invariant regardless the model properties. If we set $\mathbf{W} = \mathbf{S}$, then the so-called best GLS estimate is found, while the ML estimate is almost certainly obtained setting $\mathbf{W} = \mathbf{\Sigma}(\theta)$ when I is large enough (Browne, 1974). The 'best' GLS and ML estimators are asymptotically equivalent and therefore share the same asymptotic properties. The most relevant difference is that the GLS approach does not exactly reproduce the sample variances of the manifest variables (see, e.g., Bentler & Lee, 1979). If the multivariate normality assumption is violated, the asymptotically distribution-free (ADF) estimation method (Browne,

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

1984b) can be adopted. The ADF method minimizes the discrepancy function in (45) with a particular choice of $\mathbf{W}$ involving the sample fourth-order moments about the mean (see, e.g., Lee, 2007, p. 45). A Bayesian strategy can also be adopted (see, e.g., Hoff, 2016).

A closed-form solution to the minimization of (45) does not exist. However, Bentler, Poon, & Lee (1988) show that well-known software packages designed for structural equation models (for a review, see Narayanan, 2012) can be adapted for solving three-way three-mode FA estimation problems such as the Parafac one in (45).

6. Related Models

The structural Parafac model is estimated by using a scale-invariant method of estimation, i.e., GLS or ML, and it is not the only attempt to go beyond the component Parafac and its OLS estimation. In the component approach, Bro, Sidiropoulos, & Smilde (2002) address the estimation problem of various models for which OLS fitting procedures exist, such as the Parafac one, when errors are no longer homoscedastic and uncorrelated. They minimize a discrepancy function like

$$g(\theta) = \text{tr}\{[\mathbf{X}_A - \mathbf{Y}_A(\theta)]\mathbf{W}^{-1}\}^2 \tag{46}$$

where $\theta = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ and $\mathbf{Y}_A(\theta) = \mathbf{A}(\mathbf{C} \bullet \mathbf{B})'$. Following the same framework, Vega-Montoto & Wentzell (2003) propose some ALS algorithms for estimating the parameters of the component Parafac model assuming different error structures. Further theoretical and practical results can also be found in two companion papers (Vega-Montoto, Gu, & Wentzell, 2005; Vega-Montoto & Wentzell, 2006).

In between the component and structural approaches, we found the proposal of Harshman (1972), named Parafac2, modeling the within-occasion covariance matrices as, with obvious notation,

$$\mathbf{S}_k = \mathbf{\Sigma}_k(\theta) + \mathbf{E}_k, \tag{47}$$

$$\mathbf{\Sigma}_k(\theta) = \mathbf{B}\text{diag}(\mathbf{c}_k)\mathbf{\Phi}\text{diag}(\mathbf{c}_k)\mathbf{B}', \tag{48}$$

where $\mathbf{S}_k$ is the sample covariance matrix of the manifest variables at occasion k and $\mathbf{E}_k$ is an error term. The model is estimated by OLS, i.e., by minimizing the loss

$$h(\theta) = \sum\nolimits_k \text{tr}\{[\mathbf{S}_k - \mathbf{\Sigma}_k(\theta)]\mathbf{W}^{-1}\}^2, \tag{49}$$

with $\mathbf{W} = \mathbf{I}$. It is important to note that (48) differs from (14) because in the former, covariances among different occasions and among specific factors are not considered. Mayekawa (1987) extends the Parafac2 model by introducing the variances of the specific factors and providing the ML estimation.

About the structural approach, it is worth mentioning the work of Stegeman & Lam (2014), where an estimation procedure, based on minimum rank FA, for the structural Parafac model in (14) is proposed under the assumption of uncorrelated specific factors. In the proposal, properties of the estimators are not studied and their standard errors are not derived. Moreover, the adequacy of the model is not evaluated by goodness-of-fit tests.

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

TABLE 1.
Bentler & McClain (1976) correlation data.

	Peer report (P)				Teacher rating (T)				Self-rating (S)			
	E	A	I	M	E	A	I	M	E	A	I	M
P												
E	1											
A	−0.38	1										
I	0.42	−0.21	1									
M	−0.25	0.54	−0.54	1								
T												
E	0.64	−0.15	0.26	−0.05	1							
A	−0.29	0.66	−0.19	0.44	−0.25	1						
I	0.38	−0.09	0.56	−0.19	0.59	−0.14	1					
M	−0.22	0.51	−0.33	0.66	0.06	0.62	−0.05	1				
S												
E	0.45	−0.05	0.12	0.10	0.50	−0.05	0.36	0.17	1			
A	0.04	0.38	−0.03	0.14	0.08	0.30	0.09	0.16	0.02	1		
I	0.33	−0.13	0.35	−0.18	0.41	−0.14	0.45	−0.13	0.43	0.16	1	
M	−0.21	0.37	−0.44	0.58	−0.01	0.41	−0.10	0.62	0.06	0.04	−0.37	1

7. Application

The type (b) scale-invariant structural Parafac model is applied to the data introduced by Bentler & McClain (1976). The data refer to the correlation matrix observed on a sample of $I = 68$ fifth-grade children with respect to $J = 4$ personality variables (E: Extraversion; A: Test anxiety; I: Impulsivity; M: Academic achievement motivation) measured by Peer report (P), Teacher rating (T) and Self-rating (S), hence $K = 3$. They are reported in Table 1. The data in Table 1 represent an example of multitrait multimethod matrix. The traits are the personality variables, and the methods refer to the judge ratings.

Several scale-invariant structural Parafac models are estimated from the correlation data in Table 1. They are obtained by varying the number of factors (either $S = 2$ or $S = 3$), the covariance structure of the common factors (either orthogonal or oblique) and the covariance matrix of the specific factors (i) diagonal, the specific factors are uncorrelated; (ii) block diagonal, the specific factors are correlated only within the methods; (iii) banded, the specific factors are correlated only within the traits). The scale invariance is obtained by adding the matrix $\mathbf{D}$ as in (23) and the constraint $\boldsymbol{\Psi} * \mathbf{I} = \mathbf{I}$ for identification purposes. Estimation is addressed by ML assuming that the vectors $\mathbf{x}_i$, $i = 1, \ldots, I$, are independent and identically distributed as a multivariate normal. The analysis is performed by using the procedure CALIS for structural equation models available in the SAS software. After some experimentations, we found that the best strategy to limit the risk of non-convergence or convergence to a local optimum is to adopt the Newton–Raphson algorithm and operate as follows. For each model with $\boldsymbol{\Psi}$ diagonal, the best-fitting solution is determined by considering 100 runs of the Newton–Raphson algorithm. For every run, the starting point is the OLS solution, i.e., the minimizer of (45) with $\mathbf{W} = \mathbf{I}$, randomly initialized. This choice limits the risk of attaining local optima. When $S = 2$, 82 (orthogonal common factors) and 83 (oblique common factors) times (out of 100), the global optimum is found, whilst when $S = 3$, such numbers decrease to 60 and 46, respectively. Such calculations are made by considering the first four decimal digits. The solution corresponding to the minimum loss function value is

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

TABLE 2.
Fit, AIC, BIC and number of parameters for different Parafac models applied to the Bentler & McClain (1976) correlation data.

Number of factors	Structure of common factors	Covariance matrix of specific factors	χ^2 (p value)	AIC	BIC	Number of parameters
$S = 2$	Orthogonal	Diagonal	1.85 (< 0.0001)	171.78	225.05	24
$S = 2$	Oblique	Diagonal	1.81 (< 0.0001)	171.11	226.60	25
$S = 2$	Orthogonal	Banded	1.07 (0.0031)	143.43	223.33	36
$S = 2$	Oblique	Banded	1.04 (0.0034)	143.71	225.83	37
$S = 2$	Orthogonal	Block diag.	1.12 (0.0001)	158.94	252.16	42
$S = 2$	Oblique	Block diag.	1.09 (0.0002)	159.04	254.48	43
$S = 3$	Orthogonal	Diagonal	1.12 (0.0080)	134.76	**201.34**	30
$S = 3$	Oblique	Diagonal	1.06 (0.0079)	137.05	210.29	33
$S = 3$	Orthogonal	Banded	**0.74 (0.0639)**	**133.70**	226.92	42
$S = 3$	Oblique	Banded	**0.67 (0.0766)**	135.20	235.08	45
$S = 3$	Orthogonal	Block diag.	**0.64 (0.0625)**	138.68	245.22	48
$S = 3$	Oblique	Block diag.	**0.58 (0.0615)**	141.15	254.34	51

considered the best-fitting solution. Such best-fitting solutions are then used as starting points of the corresponding models with $\mathbf{\Psi}$ block diagonal or banded. We observed that alternative starting points led to local optima.

In the scientific community, there is no agreement on what is the best way to select a model. The suggestion is to use several different methods and choose the model on which there is the maximum agreement. Following this, we consider the χ^2 test of overall fit (Browne, 1982), AIC (Akaike, 1974) and BIC (Schwarz, 1978), which are widely used for model selection in this domain. See, for instance, Akaike (1987), Bartholomew, Knott, & Moustaki (2011) and, in the three-way three-mode context, Kroonenberg & Oort (2003). The χ^2 test refers to the null hypothesis

$$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}(\theta),$$

against the general alternative

H_1: $\mathbf{\Sigma}$ is any positive definite matrix different from $\mathbf{\Sigma}(\theta)$.

The summary of the twelve models considered in terms of p-value of the χ^2 test, AIC, BIC and number of parameters is reported in Table 2. By inspecting Table 2, we conclude that there is an agreement toward the use of $S = 3$ factors. In particular, at the level $\alpha = 0.05$, the only four models for which the null hypothesis of the χ^2 test is not rejected are the ones where $\mathbf{\Psi}$ is non-diagonal and the common factors are either orthogonal or oblique. The first model is also the best one according to the AIC criterion. We thus investigate the solution of such a model.

First of all, we report the estimates of the diagonal scaling matrix $\mathbf{D}$ in Table 3. All these estimates appear to be significant. The estimated factor loadings for the traits and the methods are given in Tables 4 and 5. Note that, without loss of fit, in the estimation process the elements of the first rows of the factor loading matrices for the traits and the methods are constrained to 1. Factor 1 is related to the duality between Impulsivity (with positive sign) versus Academic achievement motivation (with negative sign) mainly measured by Peer report. Factor 2 depends on Test anxiety and Academic achievement motivation (both with negative sign) with respect to

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

TABLE 3.
Diagonal elements of the scaling matrix **D**. (Standard errors are within parentheses.)

Extraversion (E)			Test anxiety (A)			Impulsivity (I)			Acad. achiev. motiv. (M)		
P	T	S	P	T	S	P	T	S	P	T	S
0.81 (0.09)	0.44 (0.12)	0.72 (0.12)	0.79 (0.08)	0.65 (0.10)	0.97 (0.09)	0.72 (0.09)	0.72 (0.08)	0.84 (0.08)	0.43 (0.13)	0.43 (0.12)	0.71 (0.10)

TABLE 4.
Estimated factor loading matrix for the traits. (Standard errors are within parentheses.)

	Factor 1	Factor 2	Factor 3
Extraversion (E)	1	1	1
Test anxiety (A)	− 1.04 (0.56)	− 2.07 (1.26)	− 0.04 (0.09)
Impulsivity (I)	2.38 (1.10)	0.36 (0.34)	0.44 (0.21)
Acad. achiev. motiv. (M)	− 4.45 (2.56)	− 2.92 (2.03)	0.32 (0.20)

TABLE 5.
Estimated factor loading matrix for the methods. (Standard errors are within parentheses.)

	Factor 1	Factor 2	Factor 3
Peer report (P)	1	1	1
Teacher rating (T)	0.50 (0.14)	1.70 (0.40)	3.56 (0.93)
Self-rating (S)	0.44 (0.11)	0.39 (0.15)	1.78 (0.56)

TABLE 6.
Diagonal elements of the square root of the estimated covariance matrix for the common factors $\boldsymbol{\Phi}^{1/2}$. (Standard errors are within parentheses.)

Factor 1	Factor 2	Factor 3
− 0.40 (0.16)	0.33 (0.19)	0.56 (0.21)

the traits and Teacher rating and Peer report (both with positive sign) with respect to the methods. Finally, Factor 3 is essentially related to the measurements by Teacher rating for Extraversion. (All such factor loadings are positive.) The square root of the estimated covariance matrix for the common factors, $\boldsymbol{\Phi}^{1/2}$, and of the banded covariance matrix for the unique factors, $\boldsymbol{\Psi}^{1/2}$, are given in Tables 6 and 7, respectively. From Table 7, we can see that several off-diagonal elements are rather different from zero. The banded structure of $\boldsymbol{\Psi}$ implies differences in the estimated common factors in comparison with the model where $\boldsymbol{\Psi}$ is assumed to be diagonal. To clarify this point, we briefly compare the previously described solution with the corresponding one where $\boldsymbol{\Psi}$ is diagonal. Although the null hypothesis of the χ^2 test is rejected, such a model is characterized by the lowest BIC. The interpretation is slightly modified due to the loading of Impulsivity which remarkably increases (3.66). The most relevant differences for Factor 2 involve Academic achievement motivation (loading from − 2.92 to − 1.35) and Teacher rating (from 1.70 to 1.26). With respect to Factor 3, we observe differences in the methods. In detail, the loadings of Teacher rating and Self-rating reduce to 2.13 and 1.03, respectively. Obviously, these differences are related to the fact that assuming $\boldsymbol{\Psi}$ to be diagonal leads to a model such that the common factors fit some amount of error, i.e., the correlations between the specific factors within the traits. Finally, by looking at the standard errors reported in Tables 3, 4, 5, 6 and 7, we note that in some cases, especially for the variable factor loadings, some estimates are not significantly different from zero. This is quite obvious when 42 parameters are estimated by using only 68 observations.

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

TABLE 7.
Square root of the estimated covariance matrix for the unique factors $\Psi^{1/2}$. (Standard errors are within parentheses.)

	P				*T*				*D*			
	E	*A*	*I*	*M*	*E*	*A*	*I*	*M*	*E*	*A*	*I*	*M*
P												
E	0.71 (0.23)				0.66 (0.26)				0.24 (0.19)			
A		0.86 (0.07)				0.36 (0.14)				0.35 (0.11)		
I			0.87 (0.06)				0.48 (0.11)				0.06 (0.17)	
M				0.96 (0.19)				−0.15 (0.59)				−0.24 (0.47)
T												
E					0.94 (0.22)				−0.35 (0.58)			
A						0.97 (0.04)				0.26 (0.14)		
I							0.99 (0.02)				0.12 (0.15)	
M								0.98 (0.05)				0.17 (0.28)
S												
E									1			
A										1		
I											1	
M												1

8. Conclusion and Final Remarks

The structural Parafac model has been discussed in this paper. The focus has been on its major property: uniqueness of factor loading matrices. We have shown that, under mild conditions, such a property, proved only for the component version of the model, holds. In particular, we have shown that it holds even if the specific factors are assumed to be within-variable or within-occasion correlated and the model is modified to become scale invariant. In this respect, a consideration is in order. Our uniqueness proofs are based on Kruskal's sufficient condition (1977) formulated for the component version of the model. However, this could be substituted with any other sufficient condition that exploits the particular features of the structural model. As an example, in this context, Kruskal's condition is always applied by having one of the three matrices square and of full rank, i.e., $\mathbf{\Phi}^{1/2}$. Conditions taking into account such a feature (e.g., Jiang & Sidiropoulos, 2004) probably would lead to more powerful results. Furthermore, we have focused our attention to the structural Parafac model with unconstrained matrices **B** and **C**. However, especially in the component-based version, such matrices and **A** can be constrained (e.g., column-wise orthonormality, linear dependency, nonnegativity, unimodality, symmetry). In general, imposing constraints implies replacing Kruskal's condition with more relaxed uniqueness conditions. For instance, in case of linear dependency, see Stegeman & Lam (2012). Our results might be improved by using such relaxed conditions.

Acknowledgments

Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

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.

References

Acar, E., & Yener, B. (2009). Unsupervised multiway data analysis: A literature survey. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 6–20.

Adachi, K., & Trendafilov, N. T. (2019). Some inequalities contrasting principal component and factor analyses solutions. *Japanese Journal of Statistics and Data Science*, *2*, 31–47.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third berkeley symposium on mathematical statistics and probability, Volume 5: Contributions to econometrics, industrial research, and psychometry* (pp. 111–150). California: University of California.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). Chichester: Wiley.

Bentler, P. M., & Lee, S.-Y. (1978). Statistical aspects of a three-mode factor analysis model. *Psychometrika*, *43*, 343–352.

Bentler, P. M., & Lee, S.-Y. (1979). A statistical development of three-mode factor analysis model. *British Journal of Mathematical and Statistical Psychology*, *32*, 87–104.

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

Bentler, P. M., & McClain, J. (1976). A multitrait-multimethod analysis of reflection-impulsivity. *Child Development*, *47*, 218–226.
Bentler, P. M., Poon, W.-Y., & Lee, S.-Y. (1988). Generalized multimode latent variable models: Implementation by standard programs. *Computational Statistics and Data Analysis*, *6*, 107–118.
Bloxom, B. (1968). A note on invariance in three-mode factor analysis. *Psychometrika*, *33*, 347–350.
Bloxom, B. (1984). Tucker's three-mode factor analysis. In H. G. Law, C. W. Snyder, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 104–121). New York: Praeger.
Bro, R. (1998). *Multi-way analysis in the food industry, models algorithms and applications*. Amsterdam: University of Amsterdam.
Bro, R., Sidiropoulos, N. D., & Smilde, A. K. (2002). Maximum likelihood fitting using ordinary least squares algorithms. *Journal of Chemometrics*, *16*, 387–400.
Bro, P., & Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, *17*, 16–33.
Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1–24.
Browne, M. W. (1980). Factor analysis of multiple batteries by maximum likelihood. *British Journal of Mathematical and Statistical Psychology*, *33*, 184–199.
Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.
Browne, M. W. (1984a). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, *37*, 1–21.
Browne, M. W. (1984b). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.
Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckart–Young decomposition. *Psychometrika*, *35*, 283–319.
De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, *21*, 1253–1278.
Domanov, I., & De Lathauver, L. (2013a). On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part I: Basic results and uniqueness of one factor matrix. *SIAM Journal on Matrix Analysis and Applications*, *34*, 855–875.
Domanov, I., & De Lathauver, L. (2013b). On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Uniqueness of the overall decomposition. *SIAM Journal on Matrix Analysis and Applications*, *34*, 876–903.
Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218.
Gerard, D., & Hoff, P. D. (2015). Equivariant minimax dominators of the MLE in the array normal model. *Journal of Multivariate Analysis*, *137*, 32–49.
Giordani, P., & Kiers, H. A. L. (2018). A review of tensor-based methods and their application to hospital care data. *Statistics in Medicine*, *37*, 137–156.
Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodel factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.
Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, *22*, 33–44.
Harshman, R. A., & Lundy, M. E. (1984). Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216–284). New York: Praeger.
Hoff, P. D. (2016). Equivariant and scale-free Tucker decomposition models. *Bayesian Analysis*, *11*, 627–648.
Jiang, T., & Sidiropoulos, N. D. (2004). Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints. *IEEE Transactions on Signal Processing*, *52*, 2625–2636.
Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, *14*, 105–122.
Kiers, H. A. L. (2006). Properties of and algorithms for fitting three-way component models with offset terms. *Psychometrika*, *71*, 231–256.
Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*, 455–500.
Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*, 69–97.
Kroonenberg, P. M., & Oort, F. J. (2003). Three-mode analysis of multi-mode covariance matrices. *British Journal of Mathematical and Statistical Psychology*, *56*, 305–336.
Kroonenberg, P. M. (2008). *Applied multiway data analysis*. Hoboken, NJ: Wiley.
Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decomposition, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, *18*, 95–138.
Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester: Wiley.
Lee, S.-Y., & Fong, W. (1983). A scale invariant model for three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, *36*, 217–223.
Mayekawa, S. (1987). Maximum likelihood solution to the PARAFAC model. *Behaviormetrika*, *21*, 45–63.
Mørup, M. (2011). Applications of tensor (multi-way array) factorizations and decompositions in data mining. *WIREs Data Mining and Knowledge Discovery*, *1*, 24–40.
Narayanan, A. (2012). A review of eight software packages for structural equation modeling. *The American Statistician*, *66*, 129–138.

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
Smilde, A. K., Bro, R., & Geladi, P. (2004). *Multi-way analysis: Applications in the chemical sciences*. Chichester: Wiley.
Stegeman, A. (2009). On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode. *Linear Algebra and Its Applications*, *431*, 211–227.
Stegeman, A., & Lam, T. T. T. (2012). Improved uniqueness conditions for canonical tensor decompositions with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, *33*, 1250–1271.
Stegeman, A., & Lam, T. T. T. (2014). Three-mode factor analysis by means of Candecomp/Parafac. *Psychometrika*, *79*, 426–443.
Stegeman, A., ten Berge, J. M. F., & De Lathauwer, L. (2006). Sufficient conditions for uniqueness in Candecomp/Parafac and Indscal with random component matrices. *Psychometrika*, *71*, 219–229.
Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 122–137). Madison: University of Wisconsin Press.
Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.
Unkel, S., & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, *78*, 363–382.
Vega-Montoto, L., Gu, H., & Wentzell, P. D. (2005). Mathematical improvements to maximum likelihood parallel factor analysis: Theory and simulations. *Journal of Chemometrics*, *19*, 216–235.
Vega-Montoto, L., & Wentzell, P. D. (2003). Maximum likelihood parallel factor analysis (MLPARAFAC). *Journal of Chemometrics*, *17*, 237–253.
Vega-Montoto, L., & Wentzell, P. D. (2006). Approaching the direct exponential curve resolution algorithm from a maximum likelihood perspective. *Analytica Chimica Acta*, *556*, 383–399.

Manuscript Received: 28 JUN 2019
Published Online Date: 16 AUG 2020

https://doi.org/10.1007/s11336-020-09715-4 Published online by Cambridge University Press