British Journal of Mathematical and Statistical Psychology (2016), 69, 1–19

The British Psychological Society

www.wileyonlinelibrary.com

Multi-set factor analysis by means of Parafac2

Alwin Stegeman* and Tam T. T. Lam

Heymans Institute for Psychological Research, University of Groningen, The Netherlands

We consider multi-set data consisting of N_k observations, $k = 1,\ldots, K$ (e.g., subject scores), on J variables in K different samples. We introduce a factor model for the $J \times J$ covariance matrices $\mathbf{\Sigma}_k$, $k = 1,\ldots, K$, where the common part is modelled by Parafac2 and the unique variances $\mathbf{U}_k$, $k = 1,\ldots, K$, are diagonal. The Parafac2 model implies a common loadings matrix that is rescaled for each k, and a common factor correlation matrix. We estimate the unique variances $\mathbf{U}_k$ by minimum rank factor analysis on Σ_k for each k. The factors can be chosen orthogonal or oblique. We present a novel algorithm to estimate the Parafac2 part and demonstrate its performance in a simulation study. Also, we fit our model to a data set in the literature. Our model is easy to estimate and interpret. The unique variances, the factor correlation matrix and the communalities are guaranteed to be proper, and a percentage of explained common variance can be computed for each k. Also, the Parafac2 part is rotationally unique under mild conditions.

1. Introduction

The goal of this paper is to introduce and demonstrate a novel model and algorithm for exploratory factor analysis of multi-set data. In multi-set data the same variables are observed for several different populations or subpopulations. Under the assumption that parallel proportional latent factors underlie the observed data in each (sub)population, exploratory factor analysis can be used to estimate factor loadings and the strengths of the factors in each (sub)population.

Let $\mathbf{X}_k(N_k \times J)$ be the centred data matrix of the sample from (sub)population k, for $k = 1,\ldots,K$. We measure the same J variables in each sample, where we have N_k observations in sample k. For R underlying factors, our exploratory factor model is of direct Parafac2 form (Harshman, 1972; Kiers, Ten Berge & Bro, 1999):

$$\mathbf{X}_k = \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T + \mathbf{E}_k, \quad k = 1,\ldots,K, \tag{1}$$

where $\mathbf{F}_k$ $(N_k \times R)$ is the matrix of factor scores in sample k, matrix $\mathbf{B}$ $(J \times R)$ is a loading matrix common to all samples, $\mathbf{C}_k$ $(R \times R)$ is a diagonal matrix containing the factor strengths in sample k, and $\mathbf{E}_k$ $(N_k \times J)$ is the unique part of sample k. The common part of sample k is thus modelled as $\mathbf{F}_k\mathbf{C}_k\mathbf{B}^T$. The loading matrix of sample k is $\mathbf{B}_k = \mathbf{BC}_k$ and is congruent to each $\mathbf{B}_l$ for $k \neq l$. We assume a random factor model in which $\mathbf{F}_k$ contains N_k realizations of random variables $F_1,\ldots,F_R$ that have mean zero and variance one.

*Correspondence should be addressed to Alwin Stegeman, Heymans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands (email: a.w.stegeman@rug.nl).

DOI:10.1111/bmsp.12061

Analogously, $\mathbf{E}_k$ contains N_k realizations of random variables $E_1^{(k)}, \ldots, E_J^{(k)}$ that have mean zero. The factor correlation matrix $\mathrm{Corr}(F_1, \ldots, F_R) = \mathbf{\Phi}$ is constant over k, and the covariance matrix $\mathrm{Cov}(E_1^{(k)}, \ldots, E_J^{(k)}) = \mathbf{U}_k$ of the unique part is diagonal. Additionally, the following assumptions on uncorrelatedness hold: $\mathrm{Corr}(F_r, E_j^{(k)}) = 0$ for all r,j,k, and $\mathrm{Corr}(E_i^{(k)}, E_j^{(l)}) = 0$ for all $k \neq l$ and all i, j.

Under the assumptions above, $\mathbf{X}_k$ contains N_k realizations of random variables $X_1^{(k)}, \ldots, X_J^{(k)}$ that satisfy model (1). Then the common part of the covariance matrices $\mathbf{\Sigma}_k = \mathrm{Cov}(X_1^{(k)}, \ldots, X_J^{(k)})$ is of indirect Parafac2 form (Harshman, 1972; Kiers, 1993):

$$\mathbf{\Sigma}_k = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T + \mathbf{U}_k, \quad k = 1, \ldots, K. \tag{2}$$

Matrix $\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k$ can be seen as the factor covariance matrix in sample k, with $\mathbf{C}_k$ containing the factor standard deviations. We refer to (2) as the multi-set Parafac2 factor model, and fit it to the observed covariance matrices (which we henceforth denote by $\mathbf{\Sigma}_k$, in a slight abuse of notation). There is scaling ambiguity between the columns of $\mathbf{B}$ and the diagonals of $\mathbf{C}_k$. Throughout, we assume that the columns of $\mathbf{B}$ have sum of squares unity and the scaling constants have been absorbed in $\mathbf{C}_k$.

We present a novel estimation procedure for the multi-set Parafac2 factor model (2), in which we first estimate the unique variances $\mathbf{U}_k$ by minimum rank factor analysis on each $\mathbf{\Sigma}_k$ separately. This guarantees that $\mathbf{\Sigma}_k - \mathbf{U}_k$ is a covariance matrix (i.e., positive semi-definite) for each k. Next, we estimate $\mathbf{B}$, $\mathbf{C}_k$, and $\mathbf{\Phi}$ by fitting the direct Parafac2 model to $\mathbf{Y}_k$, with $\mathbf{\Sigma}_k - \mathbf{U}_k = \mathbf{Y}_k^T\mathbf{Y}_k$. Since the $\mathbf{\Sigma}_k - \mathbf{U}_k$ are covariance matrices, we can compute the explained common variance (ECV) for each variable and each sample separately. In our estimation procedure, we can specify whether the factors should be orthogonal or oblique. In the case of orthogonal factors, the model for the common part in (2) becomes $\mathbf{B}\mathbf{C}_k^2\mathbf{B}^T$, $k = 1,\ldots,K$, which is also known as the Indscal model (Carroll & Chang, 1970). The factors and loadings in the common part of our model (2) are rotationally unique when $K \geq 4$ (Harshman & Lundy, 1996; Kiers *et al.*, 1999; Ten Berge & Kiers, 1996).

Our exploratory multi-set Parafac2 factor model (2) is the factor analogue of the simultaneous component models SCA-PF2 (for oblique components) and SCA-IND (for orthogonal components) presented in Timmerman and Kiers (2003). The difference between our model (2) and the SCA models is the same as the difference between exploratory common factor analysis (EFA) and principal component analysis (PCA) for one data matrix. The lively discussion of EFA versus PCA has been documented extensively in the literature (see e.g., Costello & Osborne, 2005; Velicer & Jackson, 1990, for an overview). PCA can be favoured because of its computational simplicity and manifest component approach, whereas EFA is computationally more difficult. However, EFA explains the correlation-producing part of the data by latent factors, while PCA is only a data-reduction method. In a simulation study, we compare the estimation accuracy of our multi-set Parafac2 factor model (2) to SCA-PF2 and SCA-IND. However, the choice between a component and factor model is a fundamental one and should not be based solely on estimation accuracy. Whenever measurement error (or a unique part in general) should be taken into account explicitly, a factor-analytic model should be applied. Naturally, this also applies to multi-set data.

Multi-set or multi-group factor analysis originated with Jöreskog (1971) and Sörbom (1974). These models are usually used in a confirmatory way, with chi-squared tests for zero restrictions on the loading matrix $\mathbf{B}$, or equal loadings, unique variances, or factor means per group to assess measurement invariance (see Chen, Sousa & West, 2005, and the references therein). Our exploratory multi-set Parafac2 factor model (2) relates to multi-group confirmatory factor analysis (CFA) models as EFA relates to CFA for a single group.

There are two crucial methodological differences between our multi-set factor model (2) and existing (exploratory or confirmatory) multi-group factor analysis models based on Jöreskog (1971) and Sörbom (1974). First, our model does not require any distributional assumptions such as normality, while deviations from normality should be taken into account when using chi-squared tests to compare nested models. A disadvantage of lacking distributional assumptions is that no significance testing can be used (besides bootstrap-like methods). Second, for common estimation methods (usually maximum likelihood) $\mathbf{\Sigma}_k - \mathbf{U}_k$, $k = 1,\ldots,K$, are not necessarily covariance matrices. This implies that a percentage of ECV cannot be computed, which is essential in determining the fit for the common part of the data for each (sub)population k. As an example, consider the analysis of the classical multi-group data set from Holzinger and Swineford (1939) conducted in Jöreskog (1971) and Sörbom (1974). Here, the matrices $\mathbf{\Sigma}_k - \mathbf{U}_k$ have negative and positive eigenvalues for all k. Hence, they are not covariance matrices.

Our model and estimation procedure are related to the exploratory three-mode factor model of Stegeman and Lam (2014), which is given by $\mathbf{\Sigma} = (\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{U}$. Here, $\mathbf{\Sigma}$ is the full $JK \times JK$ covariance matrix, $\mathbf{B}$ is as above, $\mathbf{C}$ ($K \times R$) contains the diagonals of $\mathbf{C}_k$ as rows, $\odot$ denotes the column-wise Khatri–Rao product, and $\mathbf{U}$ is a $JK \times JK$ diagonal matrix containing the unique variances. Instead of Parafac2, the model for the common part is based on Candecomp/Parafac (Carroll & Chang, 1970; Harshman, 1970). This model can only be fitted to three-mode data, that is, when $N_k = N$ for all k and the observed objects/units are the same in all samples. The relation with our multi-set Parafac2 factor model is that under the three-mode factor model of Stegeman and Lam (2014) the $J \times J$ diagonal blocks $\mathbf{\Sigma}_k$ of $\mathbf{\Sigma}$ satisfy (2).

This paper is organized as follows. In Section 2 we present our algorithm for finding a best solution of the multi-set Parafac2 factor model (2). Also, we show how the percentage of ECV can be obtained for each variable and each sample k, and discuss criteria for model selection. In Section 3 we assess the performance of our estimation procedure in a simulation study, and make a comparison with the SCA models. In Section 4 we apply our multi-set Parafac2 factor model to a data set in the literature. Section 5 contains a discussion of our findings.

2. Multi-set factor analysis by means of Parafac2

Here, we present our algorithm to estimate the multi-set Parafac2 factor model (2), starting with a discussion of Minimum Rank Factor Analysis (MRFA).

2.1. Minimum rank factor analysis

Here we briefly describe the MRFA method for two-mode factor analysis (Ten Berge & Kiers, 1991). Conceptually, the factor model splits up the (centred) observed data $\mathbf{X}$ into a common part $\mathbf{G}$ and a unique part $\mathbf{W}$. Hence, $\mathbf{X} = \mathbf{G} + \mathbf{W}$, with the common and unique parts being uncorrelated ($N^{-1}\mathbf{G}^T\mathbf{W} = \mathbf{O}$) and the unique parts of different variables being uncorrelated ($N^{-1}\mathbf{W}^T\mathbf{W} = \mathbf{U}$ diagonal). It follows that the covariance matrix of the common part is given by $N^{-1}\mathbf{G}^T\mathbf{G} = N^{-1}(\mathbf{X} - \mathbf{W})^T(\mathbf{X} - \mathbf{W}) = \mathbf{\Sigma} - \mathbf{U}$, where $\mathbf{\Sigma} = N^{-1}\mathbf{X}^T\mathbf{X}$ is the data covariance matrix. In MRFA, the common part is approximated by a small number of R factors: $\mathbf{G} \approx \mathbf{F}\,\mathbf{B}^T$, where $\mathbf{F}\,\mathbf{B}^T$ is a best rank-R approximation of $\mathbf{G}$. This implies that $\mathbf{B\Phi B}^T$ is a best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$, and can be computed from the R largest eigenvalues and associated eigenvectors of $\mathbf{\Sigma} - \mathbf{U}$ (Eckart & Young, 1936). Here $\mathbf{\Phi} = N^{-1}\mathbf{F}^T\mathbf{F}$ is the factor correlation matrix. The variance of the common part $\mathbf{G}$ that remains unexplained can be written as $\text{trace}(N^{-1}(\mathbf{G} - \mathbf{FB}^T)^T(\mathbf{G} - \mathbf{FB}^T))$ and is equal to the sum of the $J - R$ smallest eigenvalues of $\mathbf{\Sigma}-\mathbf{U}$.

The MRFA method first finds $\mathbf{U}$ such that $\mathbf{\Sigma}-\mathbf{U}$ is a covariance matrix (i.e., has nonnegative eigenvalues) and the unexplained common variance is minimized. This is done via an iterative algorithm due to Ten Berge and Kiers (1991).[1] Next, $\mathbf{B}$ and $\mathbf{\Phi}$ can be obtained via the eigendecomposition of $\mathbf{\Sigma} - \mathbf{U}$ as described above. Note that the obtained estimate of $\mathbf{U}$ depends on the number R of factors. The advantage of MRFA is that the (un)explained common variance can be computed, which is the ideal measure of model fit in the common factor model. The key constraint that must be imposed for this is that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix (i.e., has non-negative eigenvalues). This is not true in general for other factor analysis methods, such as MINRES (Harman & Jones, 1966) or maximum likelihood (Jöreskog, 1967).

2.2. Estimation procedure for the multi-set Parafac2 factor model

We present the following procedure to estimate our multi-set Parafac2 factor model (2). We need to solve

$$\min_{\substack{\mathbf{U}_k,\, \mathbf{C}_k \text{ diagonal} \\ \mathbf{U}_k \geq 0,\ \mathbf{\Phi} \geq 0,\ \mathbf{\Sigma}_k - \mathbf{U}_k \geq 0}} \sum_{k=1}^{K} \|\mathbf{\Sigma}_k - \mathbf{U}_k - \mathbf{BC}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T\|^2, \tag{3}$$

where ≥ 0 denotes semi-positive definiteness (i.e., having non-negative eigenvalues). The steps of our estimation procedure are as follows:

Step 1. For each $k \in \{1,\ldots,K\}$, use the MRFA method of Ten Berge and Kiers (1991) to estimate the unique variances $\mathbf{U}_k$ corresponding to $\mathbf{\Sigma}_k$.

Step 2. Compute the eigenvalue decomposition or singular value decomposition $(\mathbf{\Sigma}_k - \mathbf{U}_k) = \mathbf{V}_k\mathbf{C}_k\mathbf{V}_k^T$, with $\mathbf{V}_k$ having orthonormal columns, and the diagonal matrix $\mathbf{C}_k$ containing the eigenvalues in decreasing order. Set $\mathbf{Y}_k = \mathbf{C}_k^{1/2}\mathbf{V}_k^T$. Hence, $(\mathbf{\Sigma}_k - \mathbf{U}_k) = \mathbf{Y}_k^T\mathbf{Y}_k$, $k = 1,\ldots K$.

Step 3. Fit the direct Parafac2 model $\mathbf{Y}_k \approx \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T$, $k = 1,\ldots, K$, by means of the alternating least squares (ALS) algorithm of Kiers *et al.* (1999). In this step, we run the ALS algorithm 10 times for random starting values and once for the starting values suggested by Kiers *et al.* (1999) and keep the solution with the highest fit percentage. The matrix $\mathbf{\Phi}$ is defined as $\mathbf{F}_k^T\mathbf{F}_k$.

Note that in step 3 a different scaling is applied to $\mathbf{F}_k$ than in Section 1, where $N_k^{-1}\mathbf{F}_k^T\mathbf{F}_k = \mathbf{\Phi}$.

For completeness we now include a sketch of the direct Parafac2 algorithm of Kiers *et al.* (1999) that is used in step 3 above. First, the Parafac2 model is slightly reformulated as follows. Because of the constraint on $\mathbf{F}_k$, $k = 1,\ldots, K$, there exist an $R\times R$ matrix $\mathbf{F}$ and columnwise orthonormal matrices $\mathbf{P}_1,\ldots,\mathbf{P}_K$ ($N_k \times R$) such that $\mathbf{F}_k = \mathbf{P}_k\mathbf{F}$, $k = 1,\ldots,K$, and $\mathbf{F}^T\mathbf{F} = \mathbf{\Phi}$. Kiers *et al.* (1999) proposed an ALS algorithm that alternately minimizes

[1] This can be obtained from http://www.gmw.rug.nl/~kiers/

$\sum_k \|\mathbf{Y}_k - \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T\|^2$ over $\mathbf{P}_k$ for fixed $\mathbf{F}$, $\mathbf{C}_1, , \mathbf{C}_K$, and $\mathbf{B}$, for $k = 1,\ldots, K$, and over $\mathbf{F}$, $\mathbf{C}_1, \ldots, \mathbf{C}_K$, and $\mathbf{B}$, for fixed $\mathbf{P}_1, \ldots, \mathbf{P}_K$. The two main steps of this procedure are as follows:

Step PF2-1. Minimizing the objective function over $\mathbf{P}_k$ subject to $\mathbf{P}_k^T\mathbf{P}_k = \mathbf{I}_R$ is equivalent to maximizing $\mathrm{tr}\left(\mathbf{F}\mathbf{C}_k\mathbf{B}^T\mathbf{Y}_k^T\mathbf{P}_k\right)$ over $\mathbf{P}_k$. The solution is found by computing the singular value decomposition $\mathbf{S}_k\boldsymbol{\Delta}_k\mathbf{T}_k^T$ of $\mathbf{F}\mathbf{C}_k\mathbf{B}^T\mathbf{Y}_k^T$, and $\mathbf{P}_k = \mathbf{T}_k\mathbf{S}_k^T$ is the optimal solution, $k = 1,\ldots,K$.

Step PF2-2. Since $\mathbf{P}_k$, $k = 1,\ldots,K$, are columnwise orthonormal, the problem of minimizing the objective function over $\mathbf{F}, \mathbf{C}_1, \ldots, \mathbf{C}_K$, and $\mathbf{B}$ reduces to minimizing $\sum_{k=1}^K \|\mathbf{P}_k^T\mathbf{Y}_k - \mathbf{F}\mathbf{C}_k\mathbf{B}^T\|^2$, which is equivalent to fitting the Parafac model with R components to an $R \times J \times K$ array with frontal slices $\mathbf{P}_k^T\mathbf{Y}_k, k = 1,\ldots,K$. Here, Kiers *et al.* (1999) apply one iteration of the Parafac ALS algorithm to update each of $\mathbf{F}, \mathbf{C}_1, \ldots, \mathbf{C}_K$, and $\mathbf{B}$. When orthogonal factors are required, the Parafac ALS iteration is done under the constraint $\mathbf{F}^T\mathbf{F} = \mathbf{I}_R$.

The ALS procedure with steps PF2-1 and PF2-2 guarantees monotonic convergence of the Parafac2 objective function. We stop the ALS algorithm for direct Parafac2 when the relative decrease of the Parafac2 objective function drops below 10^{-7}.

Next, we discuss how to compute percentages of ECV for the solution provided by our algorithm above. Due to the MRFA method of Ten Berge and Kiers (1991), all $\boldsymbol{\Sigma}_k - \mathbf{U}_k$ are covariance matrices (i.e., have non-negative eigenvalues). The percentage of ECV for each sample k can be computed as

$$100 - 100 \cdot \frac{\mathrm{ssq}(\mathbf{Y}_k - \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T)}{\mathrm{ssq}(\mathbf{Y}_k)}, \quad k = 1, \ldots, K, \tag{4}$$

where ssq($\mathbf{Z}$) denotes the sum of squares of matrix $\mathbf{Z}$, and $\mathrm{ssq}(\mathbf{Y}_k) = \mathrm{trace}(\boldsymbol{\Sigma}_k - \mathbf{U}_k)$.

Since the Parafac ALS algorithm is used in step PF2-2, it is possible to compute the ECV due to each factor when the factors are chosen orthogonal (Stegeman & Lam, 2014). This implies that the same is possible for the ECV in (4), where the contribution of factor r is computed as

$$100 - 100 \cdot \frac{\mathrm{ssq}(\mathbf{Y}_k - \mathbf{f}_r^{(k)} c_{kr}\mathbf{b}_r^T)}{\mathrm{ssq}(\mathbf{Y}_k)}, \quad k = 1, \ldots, K, \tag{5}$$

with $\mathbf{f}_r^{(k)}$ and $\mathbf{b}_r$ denoting the rth columns of $\mathbf{F}_k$ and $\mathbf{B}$, respectively.

Denote by $\mathbf{w}_{k,m}$ the mth column of $\mathbf{Y}_k - \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T$. We define the percentage of ECV for each variable j and each sample k as

$$100 - 100 \cdot \frac{\mathrm{ssq}\left(\mathbf{w}_{k,j}\right)}{\left(\boldsymbol{\Sigma}_k - \mathbf{U}_k\right)_{jj}}, \tag{6}$$

where $(\boldsymbol{\Sigma}_k - \mathbf{U}_k)_{jj}$ is the estimated communality of variable j in sample k.

Note that our estimation procedure does not minimize (3) completely, since the estimation of $\mathbf{U}_k$ by MRFA does not take into account the Parafac2 model form for $\boldsymbol{\Sigma}_k - \mathbf{U}_k$. Instead, MRFA assumes a separate R-factor model for each sample k. This is now approximated by the rank-R indirect Parafac2 model for the common part of each sample.

If the unique variances $\mathbf{U}_k$, $k = 1,\ldots,K$, are deleted, then we obtain a new algorithm (steps 2 and 3) for indirect Parafac2 that is simpler and more efficient than the existing indirect Parafac2 algorithm of Kiers (1993). See Lam (2015) for a simulation study comparing our procedure and Kiers (1993) for the indirect Parafac2 model.

2.3. Model selection

Before fitting the multi-set Parafac2 factor model to a data set, the number of factors R must be chosen and whether to use orthogonal or oblique factors. The following guidelines for model selection are based on Timmerman and Kiers (2003) who consider SCA models. The three criteria that play a role in model selection are model fit, stability of the solution, and interpretability of the solution. For the multi-set Parafac2 factor model the ECV can be computed for each sample k, which provides natural measures of model fit. Note, however, that the ECV does not need to increase for each sample (or variable) when the number of factors R is increased. This is because the MRFA estimates of the unique variances $\mathbf{U}_k$ depend on R. Hence, caution is needed when using ECV values to select the number of factors R.

To assess stability of the solution we may use split-half analysis. This involves randomly splitting each sample k into two halves and fitting the model to each half. The mean absolute deviations of the two obtained estimates of $\mathbf{B}$ and $\mathbf{C}_k$ can then be compared by their mean absolute deviation (after taking permutational and reflectional freedom into account; see Section 2.4). Applying this procedure 50 times, say, yields a mean measure of model stability. Note, however, that exploratory factor analysis requires a very large sample size, a clear factor structure, strong factors, and large communalities for good split-half stability (Osborne & Fitzpatrick, 2012). We expect that the same will hold for multi-set factor analysis, including a clear structure in the weights $\mathbf{C}_k$.

The final criterion is interpretability of the solution, that is, of the common loading matrix $\mathbf{B}$, the weights $\mathbf{C}_k$ for each sample k, and the factor correlations in $\mathbf{\Phi}$. This may be a subjective criterion, however. Ideally, our model has a small number of factors but still fits the data well, and yields an interpretable solution that is stable in the split-half analysis. In Section 4 we will apply these criteria of model selection in our analysis of a data set in the literature.

2.4. Parafac2 uniqueness properties

An indirect Parafac2 solution $(\mathbf{B}, \mathbf{C}_1, \ldots, \mathbf{C}_K, \mathbf{\Phi})$ is called 'essentially unique' if, for any other solution $(\bar{\mathbf{B}}, \bar{\mathbf{C}}_1, \ldots, \bar{\mathbf{C}}_K, \bar{\mathbf{\Phi}})$ with $\mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T = \bar{\mathbf{B}}\bar{\mathbf{C}}_k\bar{\mathbf{\Phi}}\bar{\mathbf{C}}_k\bar{\mathbf{B}}^T$, $k = 1,\ldots,K$, there exist a permutation matrix $\mathbf{P}$ and diagonal scaling matrices $\mathbf{T}_a, \mathbf{T}_d, \mathbf{T}_\Phi$ with $\mathbf{T}_a\mathbf{T}_d\mathbf{T}_\Phi = \mathbf{I}_R$ such that (Kiers *et al.*, 1999):

(i) $\bar{\mathbf{B}} = \mathbf{B}\mathbf{P}\mathbf{T}_a$,
(ii) $\bar{\mathbf{C}}_k = \lambda_k\mathbf{P}^T\mathbf{C}_k\mathbf{P}\mathbf{T}_d$, where $\lambda_k = \pm 1$, for $k = 1,\ldots, K$,
(iii) $\bar{\mathbf{\Phi}} = \mathbf{T}_\Phi\mathbf{P}^T\mathbf{\Phi}\mathbf{P}\mathbf{T}_\Phi$.

Throughout, we scale Parafac2 solutions such that $\mathbf{\Phi}$ has a diagonal of 1s, and $\mathbf{B}$ has columns with sum of squares unity. For the case of orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_R$) the indirect Parafac2 model is $\mathbf{B}\mathbf{C}_k^2\mathbf{B}^T$, and each entry of $\mathbf{C}_k$ is identified up to sign only. Indirect Parafac2 uniqueness is equivalent to direct Parafac2 uniqueness under mild conditions (Kiers *et al.*, 1999). The sign indeterminacy via λ_k in (ii) may influence interpretation of the solution of the direct Parafac2 model (Helwig, 2013). As mentioned in Section 1, a Parafac2 solution generally is essentially unique for $K \geq 4$ (Harshman & Lundy, 1996; Kiers *et al.*, 1999; Ten Berge & Kiers, 1996).

3. Simulations

3.1. Comparing the multi-set factor and component models

Here we assess the performance of the estimation procedure in Section 2.2, and compare it to the SCA-PF2 and SCA-IND models. We consider four scenarios: Either a multi-set Parafac2 factor model or an SCA model is true in the population, and either a multi-set Parafac2 factor model or an SCA model is estimated. For the population models, we consider the following true parameter values. We set $J = 6$, $K = 5$, and $R = 2$. The true matrices $\mathbf{B}$,$\mathbf{C}$, and $\mathbf{\Phi}$ are

$$\mathbf{B}_1 = \begin{pmatrix} 0.80 & 0.10 \\ 0.10 & 0.83 \\ 0.83 & 0.10 \\ 0.10 & 0.79 \\ 0.83 & 0.10 \\ 0.10 & 0.82 \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} 0.65 & 0.10 \\ 0.10 & 0.63 \\ 0.61 & 0.10 \\ 0.10 & 0.70 \\ 0.64 & 0.10 \\ 0.10 & 0.62 \end{pmatrix},$$

$$\mathbf{C}_1 = \begin{pmatrix} 1.00 & 0.80 \\ 0.80 & 1.20 \\ 1.19 & 0.81 \\ 0.81 & 1.18 \\ 1.20 & 0.79 \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} 1.00 & 0.50 \\ 0.50 & 1.20 \\ 1.19 & 0.51 \\ 0.51 & 1.18 \\ 1.20 & 0.59 \end{pmatrix},$$

and

$$\mathbf{\Phi}_1 = \begin{pmatrix} 1 & -0.40 \\ -0.40 & 1 \end{pmatrix},$$

where $\mathbf{C}$ contains the diagonals of $\mathbf{C}_k$, $k = 1,\ldots,5$, as rows. Hence, for $\mathbf{C}$ we use true matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, and for $\mathbf{B}$ we use true matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. We consider both orthogonal factors/components ($\mathbf{\Phi} = \mathbf{I}_2$) and oblique factors/components ($\mathbf{\Phi} = \mathbf{\Phi}_1$).

After the true $\mathbf{B}$,$\mathbf{C}$,$\mathbf{\Phi}$ are chosen, the population covariance matrices are $\mathbf{\Sigma}_k = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T$, $k = 1,\ldots,K$, when SCA is true, and $\mathbf{\Sigma}_k = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T + \mathbf{U}_k$, $k = 1,\ldots,K$, when the multi-set Parafac2 factor model is true. In the latter case, the unique variances $\mathbf{U}_k$ are determined such that the $\mathbf{\Sigma}_k$ have 1s on their diagonals. For the sample sizes, we take $N_1 = 100$, $N_2 = 200$, $N_3 = 100$, $N_4 = 300$, and $N_5 = 500$. The data are generated as

$$\mathbf{X}^{(k)}_{(N_k \times J)} = \mathbf{Z}^{(k)}_{(N_k \times J)}(\mathbf{\Sigma}_k)^{1/2}, \quad k = 1,\ldots,K, \tag{7}$$

where $\mathbf{Z}^{(k)}_{(N_k \times J)}$ has random entries from the standard normal distribution, and $\mathbf{\Sigma}_k$ is as above. For each choice of true model, we generate 100 data sets as in (7) and fit the multi-set Parafac2 factor model to the sample covariance matrix, and the SCA model to the generated data themselves.

We compare the true values of $\mathbf{B}$ and $\mathbf{C}$ to their estimates by means of congruence coefficients for each column of $\mathbf{B}$ and $\mathbf{C}$. For two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, the congruence coefficient is given by (Tucker, 1951)

$$\frac{\mathbf{h}_1^T\mathbf{h}_2}{\sqrt{\mathbf{h}_1^T\mathbf{h}_1}\sqrt{\mathbf{h}_2^T\mathbf{h}_2}}.$$

Taking into account the permutational and reflectional freedom (see Section 2.4), we take the maximum of the absolute values of the congruence coefficients between estimated columns and one true column as a recovery measure. In Tables 1–4 we report the mean and standard deviation of the congruence coefficients of the columns of $\mathbf{B}$ and $\mathbf{C}$ for each case. When the multi-set Parafac2 factor model is the true model, the average communalities are also given for each k and each case. We estimate each true model using both orthogonal and oblique factors. For orthogonal estimation, each entry of $\mathbf{C}_k$ is unique up to sign only (see Section 2.4). Therefore, we take the absolute value of the estimate of $\mathbf{C}$ when using orthogonal estimation. For oblique estimation, the estimated factor correlations are often small. Also here we take the absolute value of the estimate of $\mathbf{C}$. This yields larger congruence coefficients between the true $\mathbf{C}$ and its estimate.

For oblique estimation, the number of cases with diverging components in the Candecomp/Parafac ALS step PF2-2 is also reported. We define two components as diverging if their congruence coefficient is smaller than -0.90. If diverging components occur, then a best-fitting Candecomp/Parafac model probably does not exist (Krijnen, Dijkstra & Stegeman, 2008; Stegeman, 2012). Cases of diverging components are not included in the computation of the congruence coefficients.

The recovery of the true loadings is very good in general, and somewhat better for $\mathbf{C}$ than for $\mathbf{B}$ in most cases. As possible explanations we propose the fact that $\mathbf{B}$ is larger than $\mathbf{C}$, and $\mathbf{B}$ is linked to all $K = 5$ samples simultaneously, while each row of $\mathbf{C}$ is linked to one sample only. This could make the estimation of $\mathbf{C}$ more robust compared to $\mathbf{B}$. The influence of the size of the communalities is as expected, with smaller communalities resulting in poorer recovery due to a worse signal-to-noise ratio (where the unique part is also considered as noise). In general, orthogonal estimation yields better recovery results than oblique estimation, also when the true model has oblique factors/components. As mentioned above, in that case the estimated factor/component correlation is often very small, which makes orthogonal estimation more suitable.

When the multi-set Parafac2 factor model (PF2F) is the true model (in Tables 1 and 2), estimation by SCA results in better recovery of $\mathbf{B}$ for oblique estimation. On the other hand, estimation by PF2F results in better recovery of $\mathbf{C}$. When SCA is the true model (in Tables 3 and 4), recovery of $\mathbf{C}$ is nearly identical for both estimation methods. Recovery of $\mathbf{B}$ is also nearly identical but better for SCA in some cases of oblique estimation.

When PF2F is used as the estimation method (Tables 1 and 4), the recovery of $\mathbf{C}_1$ is equal for both true models but is slightly better for $\mathbf{C}_2$ when SCA is the true model. The recovery of $\mathbf{B}$ is better when SCA is the true model. A possible explanation for this may be the following. When SCA is the true model, estimation by PF2F results in capturing part of the sampling error in the unique variances. Since the sampling error is white noise, estimation by PF2F may yield more robust results compared to PF2F being the true model. In the latter case the true unique variances are unequal which may imply a bigger challenge for the PF2F algorithm. When SCA is used as the estimation method (in Tables 2 and 3), recovery of $\mathbf{C}$ is better when SCA is the true model while recovery of $\mathbf{B}$ is better for oblique estimation when PF2F is the true model. This is surprising and we do not have an explanation for this.

Table 1. Recovery results for the multi-set Parafac2 factor model with $R = 2$ (true and estimated model)

C	**B**	**Φ**	Communality	Estimation	Congr. coeff. **B**		Congr. coeff. **C**		Div.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56, 0.70, 0.70, 0.69, 0.70	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56, 0.70, 0.70, 0.69, 0.70	Oblique	0.98 (0.05)	0.98 (0.05)	1.00 (0.00)	1.00 (0.00)	3
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34, 0.44, 0.43, 0.44, 0.43	Orth.	1.00 (0.00)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34, 0.44, 0.43, 0.44, 0.43	Oblique	0.97 (0.06)	0.96 (0.07)	1.00 (0.00)	1.00 (0.00)	2
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50, 0.64, 0.64, 0.63, 0.64	Orth.	0.98 (0.01)	0.98 (0.01)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50, 0.64, 0.64, 0.63, 0.64	Oblique	0.96 (0.06)	0.95 (0.07)	1.00 (0.00)	1.00 (0.00)	6
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30, 0.40, 0.38, 0.39, 0.38	Orth.	0.97 (0.02)	0.98 (0.01)	1.00 (0.01)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30, 0.39, 0.38, 0.39, 0.38	Oblique	0.96 (0.06)	0.96 (0.05)	1.00 (0.01)	1.00 (0.01)	1
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43, 0.57, 0.57, 0.56, 0.61	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43, 0.57, 0.57, 0.56, 0.61	Oblique	0.95 (0.08)	0.95 (0.07)	0.99 (0.01)	0.99 (0.01)	2
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26, 0.36, 0.35, 0.36, 0.37	Orth.	0.99 (0.01)	0.99 (0.01)	0.99 (0.00)	0.99 (0.01)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26, 0.36, 0.35, 0.36, 0.37	Oblique	0.97 (0.06)	0.97 (0.05)	0.99 (0.01)	0.98 (0.01)	4
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39, 0.53, 0.53, 0.52, 0.56	Orth.	0.99 (0.01)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39, 0.53, 0.53, 0.52, 0.57	Oblique	0.96 (0.07)	0.96 (0.06)	0.99 (0.01)	0.99 (0.01)	4
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23, 0.33, 0.32, 0.32, 0.34	Orth.	0.98 (0.01)	0.98 (0.01)	0.99 (0.00)	0.99 (0.01)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23, 0.33, 0.32, 0.32, 0.34	Oblique	0.96 (0.06)	0.97 (0.04)	0.99 (0.01)	0.99 (0.01)	1

Table 2. Recovery results for the multi-set Parafac2 factor model as true model and SCA as estimated model, and $R = 2$

C	B	Φ	Communality	Estimation	Congr. coeff. B		Congr. coeff. C		Div.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56, 0.70, 0.70, 0.69, 0.70	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56, 0.70, 0.70, 0.69, 0.70	Oblique	0.98 (0.05)	0.98 (0.04)	1.00 (0.00)	0.99 (0.00)	0
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34, 0.44, 0.43, 0.44, 0.43	Orth.	0.99 (0.01)	0.99 (0.01)	0.99 (0.00)	0.99 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34, 0.44, 0.43, 0.44, 0.43	Oblique	0.98 (0.06)	0.98 (0.05)	0.99 (0.00)	0.99 (0.00)	0
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50, 0.64, 0.64, 0.63, 0.64	Orth.	0.98 (0.01)	0.98 (0.01)	1.00 (0.00)	0.99 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50, 0.64, 0.64, 0.63, 0.64	Oblique	0.98 (0.03)	0.97 (0.04)	0.99 (0.00)	0.99 (0.00)	0
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30, 0.40, 0.38, 0.39, 0.38	Orth.	0.98 (0.02)	0.98 (0.01)	0.99 (0.00)	0.99 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30, 0.39, 0.38, 0.39, 0.38	Oblique	0.95 (0.07)	0.97 (0.03)	0.99 (0.00)	0.99 (0.00)	0
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43, 0.57, 0.57, 0.56, 0.61	Orth.	1.00 (0.00)	1.00 (0.00)	0.98 (0.00)	0.98 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43, 0.57, 0.57, 0.56, 0.61	Oblique	0.98 (0.06)	0.97 (0.07)	0.98 (0.01)	0.98 (0.01)	0
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26, 0.36, 0.35, 0.36, 0.37	Orth.	1.00 (0.00)	0.99 (0.01)	0.98 (0.00)	0.97 (0.01)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26, 0.36, 0.35, 0.36, 0.37	Oblique	0.98 (0.06)	0.97 (0.06)	0.97 (0.01)	0.97 (0.01)	0
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39, 0.53, 0.53, 0.52, 0.56	Orth.	0.99 (0.01)	0.99 (0.01)	0.98 (0.00)	0.98 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39, 0.53, 0.53, 0.52, 0.57	Oblique	0.97 (0.06)	0.97 (0.06)	0.98 (0.01)	0.97 (0.01)	0
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23, 0.33, 0.32, 0.32, 0.34	Orth.	0.98 (0.01)	0.99 (0.01)	0.98 (0.01)	0.97 (0.01)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23, 0.33, 0.32, 0.32, 0.34	Oblique	0.97 (0.05)	0.97 (0.06)	0.97 (0.01)	0.97 (0.01)	0

Table 3. Recovery results for the SCA model with $R = 2$ (true and estimated model)

C	**B**	**Φ**	Estimation	Congr. coeff. **B**		Congr. coeff. **C**		Div.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	Oblique	0.98 (0.05)	0.99 (0.04)	1.00 (0.00)	1.00 (0.00)	2
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	Oblique	0.99 (0.03)	0.99 (0.04)	1.00 (0.00)	1.00 (0.00)	6
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Orth.	0.98 (0.01)	0.98 (0.01)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Oblique	0.97 (0.03)	0.96 (0.05)	1.00 (0.00)	1.00 (0.00)	5
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Orth.	0.98 (0.01)	0.98 (0.01)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Oblique	0.96 (0.05)	0.96 (0.06)	1.00 (0.00)	1.00 (0.00)	10
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	Oblique	0.96 (0.09)	0.97 (0.07)	0.99 (0.01)	0.99 (0.02)	5
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	Oblique	0.97 (0.08)	0.98 (0.05)	1.00 (0.01)	0.99 (0.01)	5
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Orth.	0.99 (0.00)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Oblique	0.97 (0.06)	0.96 (0.05)	1.00 (0.01)	0.99 (0.01)	2
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Orth.	0.98 (0.00)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Oblique	0.96 (0.06)	0.97 (0.06)	1.00 (0.01)	0.99 (0.01)	6

Table 4. Recovery results for the SCA model as true model and the multi-set Parafac2 factor model as estimated model, and $R = 2$

C	**B**	**Φ**	Estimation	Congr. coeff. **B**		Congr. coeff. **C**		Div.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	Orth.	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	Oblique	0.98 (0.04)	0.99 (0.03)	1.00 (0.00)	1.00 (0.00)	10
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	Oblique	0.98 (0.05)	0.99 (0.04)	1.00 (0.00)	1.00 (0.00)	6
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Orth.	0.98 (0.01)	0.98 (0.01)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Oblique	0.97 (0.04)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	16
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Orth.	0.98 (0.01)	0.98 (0.01)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Oblique	0.97 (0.03)	0.97 (0.04)	1.00 (0.00)	1.00 (0.00)	10
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	Oblique	0.96 (0.08)	0.97 (0.05)	0.99 (0.01)	0.99 (0.01)	6
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	Orth.	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	Oblique	0.97 (0.07)	0.98 (0.05)	1.00 (0.01)	0.99 (0.01)	7
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Orth.	0.99 (0.00)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	Oblique	0.97 (0.05)	0.96 (0.05)	1.00 (0.01)	0.99 (0.01)	5
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Orth.	0.99 (0.00)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	–
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	Oblique	0.97 (0.04)	0.97 (0.05)	1.00 (0.01)	1.00 (0.01)	6

Simulation studies suggest that diverging components are more likely to occur in data with more white noise (Stegeman, 2012). In the simulations this corresponds to SCA being the true model (in Tables 3 and 4), which indeed features more cases of diverging components. Diverging components are not encountered in Table 2, when PF2F is the true model and SCA the estimated model.

In terms of computation time, the PF2F algorithm is much faster than the SCA algorithm since the latter is applied to much bigger data matrices ($N_k \times 6$ vs. 6×6 for PF2F). When the true and estimated models are the same, computation time for SCA is about 2.5 times longer on average for orthogonal estimation and about 9.9 times longer on average for oblique estimation.

To sum up, the recovery results of SCA and PF2F are quite similar in general, with SCA being slightly more robust on average. However, we do not see this as a recommendation to use SCA. The choice between a component and factor model is a fundamental one and should not be based solely on estimation accuracy.

3.2. Simulations for the multi-set Parafac2 factor model with $R = 3$

Here we consider the case of $R = 3$ factors, with true and estimated model equal to PF2F. We take $J = 6$, $K = 5$, N_k as above, and true matrices

$$\mathbf{B} = \begin{pmatrix} 0.90 & 0.10 & 0.40 \\ 0.10 & 0.41 & 0.89 \\ 0.93 & 0.40 & 0.10 \\ 0.40 & 0.91 & 0.10 \\ 0.10 & 0.90 & 0.42 \\ 0.41 & 0.10 & 0.90 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1.00 & 0.80 & 0.45 \\ 0.80 & 0.39 & 1.00 \\ 0.41 & 1.00 & 0.79 \\ 1.00 & 0.40 & 0.81 \\ 0.43 & 0.80 & 1.00 \end{pmatrix},$$

$$\mathbf{\Phi}_1 = \begin{pmatrix} 1 & -0.40 & -0.30 \\ -0.40 & 1 & 0.30 \\ -0.30 & 0.30 & 1 \end{pmatrix}.$$

We use convergence criterion 10^{-9} in the direct Parafac2 algorithm, which yields slightly better recovery results than 10^{-7} for oblique estimation. The recovery results can be found in Table 5.

As for $R = 2$, recovery is better for orthogonal estimation, and recovery is better for $\mathbf{C}$ than for $\mathbf{B}$. Although recovery is slightly worse than for $R = 2$, the results are still acceptable. Interestingly, the results for oblique estimation can be improved by using 100 Parafac ALS iterations in step PF2-2 instead of just one such iteration. For this modified direct Parafac2 algorithm the recovery results for $\mathbf{B}$ are 0.96 (0.04), 0.95 (0.06), and 0.97 (0.03) for $\mathbf{\Phi} = \mathbf{I}_3$ and oblique estimation, and 0.96 (0.02), 0.92 (0.12), and 0.96 (0.05) for $\mathbf{\Phi} = \mathbf{\Phi}_1$ and oblique estimation.

4. Application of the multi-set Parafac2 factor model

We analyse real multi-set data from Meijer, Egberink, Emons and Sijtsma (2008) concerning the Self-Perception Profile for Children (SPPC). The SPPC is used to investigate the judgement of children between 8 and 12 years of age about their own functioning in several specific domains and their global self-worth. The SPPC consists of six subscales each consisting of six items scored on a 4-point scale. Five of the six subscales represent specific domains of self-concept: Scholastic competence (SC), social

Table 5. Recovery results for the multi-set Parafac2 factor model with $R = 3$ (true and estimated model)

$\boldsymbol{\Phi}$	Communality	Estimation		Congr. coeff. **B**			Congr. coeff. **C**		Div.
$\mathbf{I}_3$	0.62, 0.59, 0.59, 0.60, 0.60	Orth.	0.99 (0.00)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	0.99 (0.00)	1.00 (0.00)	–
$\mathbf{I}_3$	0.62, 0.59, 0.59, 0.60, 0.60	Oblique	0.95 (0.10)	0.95 (0.10)	0.96 (0.07)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	15
$\boldsymbol{\Phi}_1$	0.50, 0.51, 0.58, 0.50, 0.59	Orth.	0.95 (0.03)	0.98 (0.02)	0.99 (0.02)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	–
$\boldsymbol{\Phi}_1$	0.50, 0.51, 0.58, 0.50, 0.59	Oblique	0.93 (0.10)	0.91 (0.12)	0.95 (0.08)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	14

acceptance (SA), athletic competence (AC), physical appearance (PA), and behavioural conduct (BC). The sixth scale measures global self-worth (GS), which is a more general concept.

Meijer *et al.* (2008) showed that there are differences in the item response theory model fit for children between age 8 and 9 and for children between age 10 and 13, and between boys and girls. These differences may be due to young children finding the questions too difficult or not yet having a differentiated self-concept. Also, girls may have a more differentiated self-concept than boys. Therefore, we divide this sample into $K = 4$ groups: Young girls (YoGi), with $N_1 = 147$ and age 8–9 years; young boys (YoBo), with $N_2 = 119$ and age 8–9 years; old girls (OlGi), with $N_3 = 196$ and age 10–13 years; and old boys (OlBo), with $N_4 = 149$ and age 10–13 years.

We use the sum-scores on the six subscales as observed variables. The four 6×6 covariance matrices are given in the Appendix. We apply our multi-set Parafac2 factor model (2) to this data set. First, we select an appropriate number of factors R and make a choice between using orthogonal and oblique factors. Table 6 shows the ECV for each group k for $R = 2$, $R = 3$, and $R = 4$. From $R = 2$ to $R = 3$ the ECV increases from 84% on average to 93% on average, with not much difference between orthogonal and oblique estimation. For $R = 4$ the ECV is differently distributed over the four groups and there is not much improvement overall compared to $R = 3$. As mentioned in Section 2.3, ECV does not need to increase for all groups when R is increased. This is indeed not the case when going from $R = 3$ to $R = 4$, although the sum of the ECV values is nearly the same. Based on the above, we consider the $R = 2$ and $R = 3$ solutions only.

The oblique solution for $R = 2$ is nearly the same as the orthogonal solution for $R = 2$, with the estimated factor correlation being only $-.09$. Hence, for $R = 2$ we prefer orthogonal factors because of parsimony. For $R = 3$ and oblique factors we obtain a factor correlation of .79. This solution is discarded due to lack of interpretability. The two remaining options are $R = 2$ and $R = 3$, both with orthogonal factors. Split-half analyses of these solutions do not produce good results (with mean absolute deviation for two estimates of $\mathbf{B}$ being 0.20 for $R = 2$) due to the relatively small sample size for each group k. Below, we present the $R = 2$ solution and briefly discuss the $R = 3$ solution.

The results for $R = 2$ orthogonal factors are as follows:

$$\mathbf{B} = \begin{array}{c} \begin{pmatrix} 0.22 & \mathbf{0.48} \\ 0.34 & \mathbf{0.58} \\ 0.22 & \mathbf{0.62} \\ \mathbf{0.58} & -0.04 \\ 0.30 & 0.21 \\ \mathbf{0.61} & 0.01 \end{pmatrix} \end{array} \begin{array}{l} \text{SC} \\ \text{SA} \\ \text{AC} \\ \text{PA} \\ \text{BC} \\ \text{GS} \end{array}, \quad \mathbf{C} = \begin{pmatrix} 6.19 & 2.97 \\ 4.54 & 5.15 \\ 6.77 & 2.84 \\ 4.26 & 2.48 \end{pmatrix} \begin{array}{l} \text{YoGi} \\ \text{YoBo} \\ \text{OlGi} \\ \text{OlBo} \end{array}, \tag{8}$$

where $\mathbf{B}$ has column sums of squares equal to 1, and the loadings whose absolute values are larger than or equal to 0.4 are in bold font.

The unique variances $\mathbf{U}_k$ are given in Table 7. A unique variance of zero is a boundary solution. This may also occur for other models and estimation methods (Bentler & Lee, 1979), or when Heywood cases are suppressed. The zero and small unique variances for GS are not surprising, since this scale may be considered as a summary of the other five scales. As such, it has no specific part and a very large common part.

The ECV percentages for each variable j, each group k, and due to each factor are given in Table 8. The solution (8) with $R = 2$ orthogonal factors can be interpreted as follows. Factor 1 is a strong general factor with highest loadings for PA and GS and is stronger for

Table 6. Percentages of explained common variance for sample k for the SPPC data set for models with different numbers of factors, and orthogonal or oblique factors

	$R = 2$		$R = 3$		$R = 4$	
	Orth.	Oblique	Orth.	Oblique	Orth.	Oblique
YoGi	88.5	88.7	95.4	95.4	97.9	97.5
YoBo	83.1	83.4	93.4	93.3	85.8	83.2
OlGi	85.0	85.1	94.8	94.8	93.1	96.7
OlBo	80.3	80.3	88.3	89.0	88.5	84.6

Table 7. Unique variances for variable j and sample k for the SPPC data set and $R = 2$ factors

	SC	SA	AC	PA	BC	GS
YoGi	10.48	5.11	9.14	6.69	7.37	0.67
YoBo	7.64	8.37	3.03	7.07	7.69	0
OlGi	10.22	9.54	0	6.52	4.41	0
OlBo	8.12	0	8.47	6.24	8.39	0

the girls than for the boys. Hence, factor 1 captures differences in the variability of PA and GS judgements between boys and girls. The weaker factor 2 is a combination of SC, SA, and AC, and is much stronger for young boys than for the other groups. The young boys apparently show more variability on these scales than the other groups. This may be due to their lack of a coherent self-perception. The percentages of ECVs for group OlBo are rather low for the sum-scores SC and AC. Compared to other groups, the percentage of ECV of YoBo is smallest for factor 1 but largest for factor 2.

Next, we briefly discuss the solution with $R = 3$ orthogonal factors. The ECV is above 70% for every variable j and group k, which is somewhat better than for $R = 2$ (Table 8). The obtained loadings and weights are

$$\mathbf{B} = \begin{array}{c} \begin{pmatrix} 0.30 & 0.06 & \mathbf{0.61} \\ 0.30 & 0.35 & \mathbf{0.45} \\ -0.05 & \mathbf{0.66} & \mathbf{0.57} \\ \mathbf{0.53} & \mathbf{0.49} & -0.13 \\ 0.38 & 0.10 & 0.30 \\ \mathbf{0.63} & \mathbf{0.44} & -0.03 \end{pmatrix} \end{array} \begin{array}{l} \text{SC} \\ \text{SA} \\ \text{AC} \\ \text{PA} \\ \text{BC} \\ \text{GS} \end{array}, \quad \mathbf{C} = \begin{pmatrix} 4.96 & 4.86 & 3.30 \\ 4.15 & 3.22 & 5.28 \\ 4.86 & 5.99 & 2.01 \\ 4.04 & 2.20 & 2.00 \end{pmatrix} \begin{array}{l} \text{YoGi} \\ \text{YoBo} \\ \text{OlGi} \\ \text{OlBo} \end{array}. \tag{9}$$

The first two factors are strongest in terms of ECV and are similar to each other, with the largest differences being the loadings on AC. These factors have larger weights for girls

Table 8. Percentages of explained common variance for variable j and sample k, and due to each factor for the SPPC data set and $R = 2$ orthogonal factors

	SC	SA	AC	PA	BC	GS	Factor 1	Factor 2
YoGi	80.6	82.7	81.7	94.2	72.3	97.1	68.3	20.3
YoBo	80.1	62.7	85.4	92.5	73.0	94.7	49.9	33.2
OlGi	75.4	90.4	74.9	94.2	61.6	96.0	67.4	17.6
OlBo	54.2	77.1	51.5	97.9	70.8	96.7	59.2	21.1

than for boys and they resemble the strong factor in the $R = 2$ solution (8). The third factor resembles the weaker factor in the $R = 2$ solution. In terms of interpretability, the $R = 2$ solution (8) is clearer. This comes at the cost of lower ECV for some variables and groups.

Finally, we present the solution for SCA-IND with $R = 2$. We fit the SCA-IND model to the centred SPPC sumscores for each group. We obtain

$$\mathbf{B} = \begin{pmatrix} 0.15 & \mathbf{0.73} \\ 0.36 & \mathbf{0.46} \\ 0.25 & \mathbf{0.42} \\ \mathbf{0.66} & -0.16 \\ 0.30 & 0.23 \\ \mathbf{0.51} & 0.07 \end{pmatrix} \begin{matrix} \text{SC} \\ \text{SA} \\ \text{AC} \\ \text{PA} \\ \text{BC} \\ \text{GS} \end{matrix}, \qquad \mathbf{C} = \begin{pmatrix} 6.15 & 4.30 \\ 5.36 & 4.94 \\ 6.42 & 3.96 \\ 5.15 & 3.76 \end{pmatrix} \begin{matrix} \text{YoGi} \\ \text{YoBo} \\ \text{OlGi} \\ \text{OlBo} \end{matrix}. \tag{10}$$

The SCA-IND solution (10) is similar to (8), but the factor structure is less clear. Compared to (8) the differences between large and small loadings in **B** are smaller for SA, AC, BC, and GS. Also, the differences in weights are smaller. The larger loadings for SC and PA are due to these variables having the largest variances in the four groups (Appendix). In the multi-set factor model this is mediated by the unique variances. This illustrates the benefit of using the multi-set factor model in which most non-systematic variation is captured by the unique variances, while in the SCA model this variation influences the estimated loadings and weights.

5. Discussion

In this paper, we have presented an exploratory multi-set factor model with common covariance part of indirect Parafac2 form. To estimate our multi-set Parafac2 factor model we use MRFA to obtain the unique variances $\mathbf{U}_k$ and a new indirect Parafac2 algorithm to estimate the common loading matrix $\mathbf{B}$, factor strengths $\mathbf{C}_k$ for each sample k, and the factor correlation matrix $\mathbf{\Phi}$. The matrices $\mathbf{\Sigma}_k - \mathbf{U}_k$ are guaranteed to be covariance matrices due to the MRFA algorithm. Therefore, percentages of ECV can be computed for each sample k, and for each variable in each sample k. For other factor methods of multi-set data analysis, such as multi-group exploratory or confirmatory factor analysis, it is not guaranteed that such ECVs can be computed.

The simulation study shows that our relatively simple estimation procedure for the multi-set Parafac2 factor model performs very well in retrieving underlying factors when the data are randomly sampled with true covariance matrices $\mathbf{\Sigma}_k - \mathbf{U}_k$ satisfying the indirect Parafac2 model. The recovery is better when we use orthogonal factors.

In the simulation study we also compared the performance of our multi-set Parafac2 factor model to the corresponding component SCA models. It was found that estimation accuracy is similar for the multi-set component and factor models, with the SCA models being slightly more robust in general. However, the choice between a component and factor model is a fundamental one and should not be based solely on estimation accuracy. The distinction between common and unique parts is a key property of factor models, while component models are basically used for data reduction (Costello & Osborne, 2005).

The results of the application of our multi-set Parafac2 factor model confirm the considerations of Meijer *et al.* (2008) on differences between young children and old children, and between girls and boys when they judge their own functioning in several specific domains and their global self-worth. Our results show that girls have higher variability in their judgement on physical appearance and global self-worth than boys. Also, young boys have higher variability in their judgements on scholastic competence, social acceptance, and athletic competence. The solution of the corresponding SCA model shows less difference between small and large loadings and weights, as a result of fitting the model to the observed data and not only to the systematic (common) part of the data. This shows the value of a factor model over a component model in practice.

Acknowledgement

The authors would like to thank an anonymous reviewer for detailed comments that have helped to improve the manuscript considerably. This research was supported by The Netherlands Organisation for Scientific Research (NWO), VIDI grant 452-08-001.

References

Bentler, P. M., & Lee, S.-Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, *32*, 87–104. doi:10.1111/j.2044-8317.1979.tb00754.x

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. doi:10.1007/BF02310791

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*, 471–492. doi:10.1207/s15328007 sem1203_7

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, *10*(7). Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218. doi:10.1007/BF02288367

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (MINRES). *Psychometrika*, *31*, 351–368. doi: 10.1007/BF02289468

Harshman, R. A. (1970). Foundations of the Parafac proceduce: Models and conditions for an 'explanatory' multimodal 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*, 30–44.

Harshman, R. A., & Lundy, M. E. (1996). Uniqueness proof for a family of model sharing features of Tucker's three-mode factor analysis and Parafac/Candecomp. *Psychometrika*, *61*, 133–154. doi:10.1007/BF02296963

Helwig, N. E. (2013). The special sign indeterminacy of the direct-fitting Parafac2 model: Some implications, cautions, and recommendations for simultaneous component analysis. *Psychometrika*, *78*, 725–739. doi:10.1007/S11336-013-9331-7

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (Supplementary Educational Monograph No. 48). Chicago, IL: University of Chicago Press.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*, 443–482. doi:10.1007/BF02289658

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426. doi:10.1007/BF02291366

Kiers, H. A. L. (1993). An alternating least squares algorithm for PARAFAC2 and three-way Dedicom. *Computational Statistics & Data Analysis*, *16*, 103–118. doi:10.1016/0167-9473(93)90247-Q

Kiers, H. A. L., Ten Berge, J. M. F., & Bro, R. (1999). Parafac2 – Part I. A direct fitting algorithm for the Parafac2 model. *Journal of Chemometrics*, *13*, 275–294. doi:10.1002/(SICI)1099-128X(199905/08)13:3/4<275::AID-CEM543>3.0.CO;2-B

Krijnen, W. P., Dijkstra, T. K., & Stegeman, A. (2008). On the non-existence of optimal solutions and the occurrence of 'degeneracy' in the Candecomp/Parafac model. *Psychometrika*, *73*, 431–439. doi:10.1007/s11336-008-9056-1

Lam, T. T. T. (2015). *Some new methods for three-mode factor analysis and multi-set factor analysis* (PhD thesis). University of Groningen, The Netherlands.

Meijer, R. R., Egberink, I. J. L., Emons, W. H. M., & Sijtsma, K.(2008). Detection and validation of unscalable item score patterns using item response theory: An illustration with Harter's self-perception profile for children. *Journal of Personality Assessment*, *90*, 227–238. doi:10.1080/00223890701884921

Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation*, *17*(15). Retrieved from http://pareonline.net/getvn.asp?v=17&n=15

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229–239. doi:10.1111/j.2044-8317.1974.tb00543.x

Stegeman, A. (2012). Candecomp/Parafac – from diverging components to a decomposition in block terms. *SIAM Journal on Matrix Analysis and Applications*, *33*, 291–316. doi:10.1137/110825327

Stegeman, A., & Lam, T. T. T. (2014). Three-mode factor analysis by means of Candecomp/Parafac. *Psychometrika*, *79*, 426–443. doi:10.1007/S11336-013-9359-8

Ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309–315. doi:10.1007/BF02294464

Ten Berge, J. M. F., & Kiers, H. A. L. (1996). Some uniqueness results for Parafac2. *Psychometrika*, *61*, 123–132. doi:10.1007/BF02296962

Timmerman, M., & Kiers, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, *68*, 105–121. doi:10.1007/BF02296656

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies*(Personnel Research Section Report No.984). Department of the Army, Washington, DC.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure (with comments and reply). *Multivariate Behavioral Research*, *25*, 1–114. doi:10.1207/s15327906mbr2501_1

Received 13 May 2014; revised version received 1 September 2015

Appendix

The covariance matrices for the four groups in the SPPC data set of Meijer *et al.* (2008) are as follows:

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 14.38 & 4.97 & 3.02 & 2.20 & 3.29 & 4.31 \\ 4.97 & 14.90 & 5.40 & 4.40 & 3.35 & 6.60 \\ 3.02 & 5.40 & 13.50 & 4.44 & 3.26 & 4.66 \\ 2.20 & 4.40 & 4.44 & 20.11 & 5.29 & 12.18 \\ 3.29 & 3.35 & 3.26 & 5.29 & 11.80 & 6.11 \\ 4.31 & 6.60 & 4.66 & 12.18 & 6.11 & 13.91 \end{pmatrix},$$

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 17.68 & 5.77 & 6.89 & 2.27 & 5.94 & 4.38 \\ 5.77 & 15.17 & 6.24 & 4.68 & 4.24 & 6.20 \\ 6.89 & 6.24 & 12.51 & 2.46 & 3.52 & 2.56 \\ 2.27 & 4.68 & 2.46 & 14.80 & 3.16 & 8.11 \\ 5.94 & 4.24 & 3.52 & 3.16 & 12.31 & 4.36 \\ 4.38 & 6.20 & 2.56 & 8.11 & 4.36 & 10.52 \end{pmatrix},$$

$$\mathbf{\Sigma}_3 = \begin{pmatrix} 13.82 & 3.88 & 4.37 & 3.91 & 3.59 & 4.52 \\ 3.88 & 15.23 & 6.15 & 7.72 & 4.58 & 7.70 \\ 4.37 & 6.15 & 13.36 & 7.30 & 2.60 & 5.91 \\ 3.91 & 7.72 & 7.30 & 19.62 & 5.64 & 12.49 \\ 3.59 & 4.58 & 2.60 & 5.64 & 9.65 & 6.09 \\ 4.52 & 7.70 & 5.91 & 12.49 & 6.09 & 12.95 \end{pmatrix},$$

$$\mathbf{\Sigma}_4 = \begin{pmatrix} 12.64 & 4.45 & 0.77 & 3.46 & 3.73 & 2.79 \\ 4.45 & 12.42 & 4.72 & 4.66 & 3.42 & 4.68 \\ 0.77 & 4.72 & 10.78 & 2.13 & 0.81 & 2.37 \\ 3.46 & 4.66 & 2.13 & 13.61 & 4.62 & 7.51 \\ 3.73 & 3.42 & 0.81 & 4.62 & 12.33 & 4.45 \\ 2.79 & 4.68 & 2.37 & 7.51 & 4.45 & 8.10 \end{pmatrix}.$$