British Journal of Mathematical and Statistical Psychology (2011), 64, 277–290
© 2010 The British Psychological Society



The British Psychological Society

www.wileyonlinelibrary.com

Simultaneous analysis of coupled data matrices subject to different amounts of noise

Tom F. Wilderjans*, Eva Ceulemans, Iven Van Mechelen and Robert A. van den Berg

Katholieke Universiteit Leuven, Belgium

In many areas of science, research questions imply the analysis of a set of coupled data blocks, with, for instance, each block being an experimental unit by variable matrix, and the variables being the same in all matrices. To obtain an overall picture of the mechanisms that play a role in the different data matrices, the information in these matrices needs to be integrated. This may be achieved by applying a data-analytic strategy in which a global model is fitted to all data matrices simultaneously, as in some forms of simultaneous component analysis (SCA). Since such a strategy implies that all data entries, regardless the matrix they belong to, contribute equally to the analysis, it may obfuscate the overall picture of the mechanisms underlying the data when the different data matrices are subject to different amounts of noise. One way out is to downweight entries from noisy data matrices in favour of entries from less noisy matrices. Information regarding the amount of noise that is present in each matrix, however, is, in most cases, not available. To deal with these problems, in this paper a novel maximum-likelihood-based simultaneous component analysis method, referred to as MxLSCA, is proposed. Being a stochastic extension of SCA, in MxLSCA the amount of noise in each data matrix is estimated and entries from noisy data matrices are downweighted. Both in an extensive simulation study and in an application to data stemming from cross-cultural emotion psychology, it is shown that the novel MxLSCA strategy outperforms the SCA strategy with respect to disclosing the mechanisms underlying the coupled data.

1. Introduction

As a result of our information society, in many areas of science researchers face a boom in the amount of available information. In particular, more and more situations arise in which different pieces of information, which may originate from different sources, are available about a given set of entities under study (e.g., subjects, firms, or genes).

*Correspondence should be addressed to Tom F. Wilderjans, Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium (e-mail: tom.wilderjans@psy.kuleuven.be).

DOI:10.1348/000711010X513263

These situations imply data that consist of a set of data blocks, one for each source of information. Such data, as a whole, may be referred to as coupled data.

A special case of coupled data is encountered when each data block is a (two-way two-mode) matrix containing the measurements of a set of experimental units on a set of variables, and when either the experimental units or the variables are the same for all matrices. In cross-cultural psychology, coupled data with shared variables are, for example, encountered when different groups of subjects, stemming from different cultures, are administered the same questionnaire on emotional experiences (see Kuppens, Ceulemans, Timmerman, Diener, & Kim-Prieto, 2006). A situation with coupled data with shared experimental units, for example, shows up when for the same set of persons two kinds of information are available, such as cognitive performance and physiological measures. In this paper, we will focus on coupled data matrices that pertain to common variables.

Coupled data contain a wealth of information and this may give rise to a broad range of research questions. One such intriguing question pertains to mechanisms that may underlie the data in each of the coupled matrices. A possible goal then may be to get an overall picture of these underlying mechanisms for the whole set of data matrices at hand. In the cross-cultural psychology case, for example, one may wish to determine the factors that drive the emotional experience of subjects from different cultures.

To obtain an overall picture of underlying mechanisms, the information that is available in the different data matrices needs to be integrated. This may be achieved by fitting a global model to all data matrices simultaneously. Such a fitting will typically imply the optimization of a global loss function, which may consist of different partial loss functions, one for each data matrix.

To obtain an overall picture of the mechanisms underlying a given set of coupled data matrices, the family of simultaneous component models (Timmerman & Kiers, 2003) may be most useful. Members of this family have in common that components are extracted from all data matrices simultaneously. One of the simultaneous component models that is appealing in the case of coupled data matrices that pertain to a set of common variables is the SCA-P model (Kiers & Ten Berge, 1994; Timmerman & Kiers, 2003). In SCA-P, a principal component analysis (PCA) is applied to each data matrix, with the restriction that the component loadings are to be the same for all data matrices. In the cross-cultural psychology example, these components can be assumed to capture the latent emotion dimensions that account for the (manifest) emotional experience of subjects belonging to different cultures.

When analysing coupled data, however, several difficulties may arise which may blur the overall picture of the mechanisms underlying the data. One such difficulty is that the data blocks under study may be subject to noise, for example, because of measurement error. Note that noise may pertain not only to measurement error but also to weak mechanisms that do not explain much variance and thus are of no interest for further study (Ceulemans & Kiers, 2009). For the sake of simplicity, throughout this paper, noise will be conceived as measurement error only. A further, and even more challenging, difficulty is encountered when the amount of noise varies across different data matrices (i.e., noise heterogeneity). In the cross-cultural psychology example, this may be the case when variables with respect to emotional experience are measured more reliably in some cultures than in others. When analysing all data matrices simultaneously, noise heterogeneity may lead to incorrect inferences with respect to the underlying mechanisms in the data. To circumvent this problem, one might consider downweighting the noisy data matrices in favour of the less noisy ones.

When dealing with noisy coupled data, sometimes *a priori* knowledge regarding the amount of noise in each data matrix is available (e.g., reliability information stemming from preceding analyses). In general, however, such (*a priori*) knowledge is not present; this implies that the amount of noise in each matrix needs to be determined (i.e., estimated) during the data-analytic process.

In this paper, a new data-analytic strategy is proposed to deal with the difficulties mentioned above in the case of coupled data matrices that pertain to common variables. In particular, a novel SCA-P strategy, called maximum-likelihood simultaneous component analysis (MxLSCA-P), is introduced. In this strategy, the amount of noise in each data matrix is estimated and the coupled data matrices are analysed in such a way that matrices that include a large amount of noise are downweighted. The performance of the new (MxL)SCA-P strategy will further be evaluated and compared with that of the original SCA-P method in terms of recovering the underlying mechanisms of a coupled data set.

The remainder of this paper is organized in five main sections. In Section 2 the SCA-P strategy will be briefly recapitulated, and in Section 3 the new MxLSCA-P strategy will be introduced. For both strategies, the model will be discussed together with the loss function that is to be optimized and with an algorithm to estimate the model parameters. Throughout, we will limit ourselves to the case of two coupled matrices. Generalizations to cases with more than two coupled data matrices are straightforward, however. In the following two sections, the two strategies will be compared with respect to their capability to reveal true underlying data mechanisms. In Section 4 this will be done by means of an extensive simulation study, and in Section 5 by means of comparative analyses of an empirical coupled data set from the field of cross-cultural (emotion) psychology. Section 6 contains some concluding remarks.

2. Simultaneous component analysis (SCA-P)

The SCA-P model approximates a first $I \times K$ real-valued data matrix $\mathbf{D}^1$ with the scores of I experimental units on K variables, and a second $J \times K$ real-valued data matrix $\mathbf{D}^2$ with the scores of J (other) experimental units on the same K variables, by an $I \times K$ model matrix $\mathbf{M}^1$ and a $J \times K$ model matrix $\mathbf{M}^2$, respectively:

$$\begin{aligned} d^1_{ik} &= m^1_{ik} + e^1_{ik}, \\ d^2_{jk} &= m^2_{jk} + e^2_{jk}. \end{aligned} \tag{1}$$

The model matrices $\mathbf{M}^1$ and $\mathbf{M}^2$ can further be represented by two PCA models with S components, with the component loadings for the variables restricted to being identical:

$$\begin{aligned} m^1_{ik} &= \sum_{s=1}^{S} a_{1_{(is)}} b_{ks}, \\ m^2_{jk} &= \sum_{s=1}^{S} a_{2_{(js)}} b_{ks}, \end{aligned} \tag{2}$$

where $a_{1_{(is)}}$ and $a_{2_{(js)}}$ denote the entries of the $I \times S$ component score matrix $\mathbf{A}_1$ and the $J \times S$ component score matrix $\mathbf{A}_2$, and b_{ks} the entries of the $K \times S$ component loading matrix $\mathbf{B}$ that is shared by the two model matrices. Note that the component matrices $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{B}$ are unique up to a joint permutation, scaling, reflection, and rotation of the components (Smilde, Bro, & Geladi, 2004).

The aim of a SCA-P analysis with S components of a real-valued coupled data set ($\mathbf{D}^1$, $\mathbf{D}^2$) is to estimate a real-valued model matrix couple ($\mathbf{M}^1$, $\mathbf{M}^2$) such that the value of the least squares loss function

$$f = \sum_{i=1}^{I}\sum_{k=1}^{K}(d_{ik}^1 - m_{ik}^1)^2 + \sum_{j=1}^{J}\sum_{k=1}^{K}(d_{jk}^2 - m_{jk}^2)^2 \tag{3}$$

is minimized, and such that $\mathbf{M}^1$ and $\mathbf{M}^2$ can be represented by model (2) with S components. Prior to the actual analysis, one might consider preprocessing the coupled data matrices by centring and/or scaling them. Because we are interested in the mechanisms that underlie each of the data matrices, differences between matrices in the overall means of the variables should be removed. To achieve this, we centre (i.e., zero-mean) the variables for each data matrix. Further, in order to remove arbitrary scale differences between the variables, we scale the data by setting the sum of squares of each variable (computed over all experimental units in both data matrices) equal to the total number of experimental units in both matrices, $I+J$ (see Timmerman & Kiers, 2003). As such, in the analysis the importance of all variables, considered for the whole of the data matrices under study, is equalized, whereas for each variable possible differences in variance between the data matrices are preserved.

To estimate the parameters of the SCA-P model, a PCA is performed on $[\mathbf{D}^1; \mathbf{D}^2]$, with the semicolon denoting matrix concatenation (Kiers, 2000) and the resulting matrix having $(I+J)$ rows and K columns.

3. A novel strategy: Maximum-likelihood simultaneous component analysis

The (global) loss function in (3) is an unweighted sum of partial loss functions for the corresponding data matrices. This implies that all data entries, regardless of the data matrix they belong to, contribute equally to the analysis. When, however, data matrices are subject to noise heterogeneity, this may be troublesome since it may lead to incorrect inferences with respect to the component scores and loadings. To overcome this problem, we propose maximum-likelihood simultaneous component analysis (MxLSCA-P), a stochastic extension of SCA-P. In particular, for the noise entries e_{ik}^1 and e_{jk}^2 in (1) it is assumed that $e_{ik}^1 \sim N(0,\sigma_1^2)$ and $e_{jk}^2 \sim N(0,\sigma_2^2)$. In order to deal with possible noise heterogeneity, the variances σ_1^2 and σ_2^2, which capture the amount of noise in the two data matrices, are allowed to differ.

To estimate the MxLSCA-P parameters the following (log)likelihood function is optimized:

$$\begin{aligned} l = {} & \frac{IK+JK}{2}\log(2\pi) + IK\log(\sigma_1) + JK\log(\sigma_2) \\ & + \frac{1}{2\sigma_1^2}\sum_{i=1}^{I}\sum_{k=1}^{K}(d_{ik}^1 - m_{ik}^1)^2 + \frac{1}{2\sigma_2^2}\sum_{j=1}^{J}\sum_{k=1}^{K}(d_{jk}^2 - m_{jk}^2)^2. \end{aligned} \tag{4}$$

Note that (4) implies that entries belonging to noisy data matrices are downweighted in favour of entries from less noisy matrices, since the weight for each data matrix is inversely related to the estimated amount of noise (i.e., σ_1^2 or σ_2^2) in that matrix.

For the actual estimation, an alternating least squares algorithm is used in which the following two sets of parameters are alternatingly re-estimated conditionally upon each other until no improvement in the loss function is observed: (1) the entries of the component matrices $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{B}$ and (2) the noise parameters σ_1 and σ_2. To update the entries of the component matrices $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{B}$, a weighted SCA (see Kiers & Ten Berge, 1994; Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen, 2009) is performed on $[\mathbf{D}^1; \mathbf{D}^2]$, with $\sqrt{1/2\hat{\sigma}_1^2}$ and $\sqrt{1/2\hat{\sigma}_2^2}$ being the weight for $\mathbf{D}^1$ and $\mathbf{D}^2$, respectively. The noise parameters, in turn, are updated as follows:

$$\hat{\sigma}_1 = \sqrt{\frac{\sum_{i=1}^{I}\sum_{k=1}^{K}(d_{ik}^1 - m_{ik}^1)^2}{IK}}, \qquad \hat{\sigma}_2 = \sqrt{\frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(d_{jk}^2 - m_{jk}^2)^2}{JK}}. \tag{5}$$

Because the MxLSCA-P algorithm, unlike the SCA-P algorithm, may end up in a local instead of a global minimum, it is advisable to apply a multi-start procedure. In such a procedure, the algorithm is run a pre-specified number of times, each time from different initial parameter estimates; subsequently, the solution with the best loss function value is retained.

4. Simulation study

4.1. Problem

In this section, the performance of the novel MxLSCA-P strategy is compared to the performance of the SCA-P strategy with respect to uncovering the common component loadings (i.e., $\mathbf{B}$). For this purpose, three types of real-valued $I \times K$ and $J \times K$ matrix couples must be distinguished: (1) the (known) true matrix couple ($\mathbf{T}^1$, $\mathbf{T}^2$) that can be decomposed into matrices $\mathbf{A}_1^{(T)}$, $\mathbf{A}_2^{(T)}$, and $\mathbf{B}^{(T)}$ with S components; (2) the data matrix couple ($\mathbf{D}^1$, $\mathbf{D}^2$) that is obtained by adding noise to ($\mathbf{T}^1$, $\mathbf{T}^2$); and (3) the model matrix couple ($\mathbf{M}^1$, $\mathbf{M}^2$) that is obtained by applying SCA-P or MxLSCA-P with S components to the data matrix couple ($\mathbf{D}^1$, $\mathbf{D}^2$) and that can be decomposed into $\mathbf{A}_1^{(M)}$, $\mathbf{A}_2^{(M)}$, and $\mathbf{B}^{(M)}$. For each strategy, it will be determined to what extent the true component loadings are recovered by comparing the component loadings as obtained from the algorithm (i.e., $\mathbf{B}^{(M)}$) to the true component loadings (i.e., $\mathbf{B}^{(T)}$).

First, the design and the procedure of the simulation study will be outlined. Then, the simulation results with respect to the recovery performance at the level of the common component loadings will be presented.

4.2. Design and procedure

Four factors were systematically manipulated in a completely randomized factorial design, with all factors considered random. The first two factors, which pertain to the sizes of the data matrices, are:

(a) the *size of the common mode*, K, manipulated at three levels, $K = 200$, $K = 100$, $K = 50$;

(b) the *array/total ratio*, r, of the size of $\mathbf{T}^1$ ($\mathbf{D}^1$, $\mathbf{M}^1$) to the total size $\mathbf{T}^1 + \mathbf{T}^2$ ($\mathbf{D}^1 + \mathbf{D}^2$, $\mathbf{M}^1 + \mathbf{M}^2$, respectively),

$$r = \frac{I \times K}{(I \times K) + (J \times K)}, \tag{6}$$

manipulated at five levels, .10, .30, .50, .70, .90.

Table 1 presents the sizes of $\mathbf{T}^1$ and $\mathbf{T}^2$ that result from an orthogonal combination of the factors K and r. Note that the total size of $\mathbf{T}^1 + \mathbf{T}^2$ ($\mathbf{D}^1 + \mathbf{D}^2$, $\mathbf{M}^1 + \mathbf{M}^2$) was kept constant at a value of 20,000 entries.

Table 1. The sizes of $\mathbf{T}^1$ ($I \times K$) and $\mathbf{T}^2$ ($J \times K$) that are obtained by an orthogonal combination of the size of the common mode, K, and array/total ratio, r

K	r	Size of $\mathbf{T}^1$ ($I \times K$)	Size of $\mathbf{T}^2$ ($J \times K$)
50	.10	(40 × 50)	(360 × 50)
50	.30	(120 × 50)	(280 × 50)
50	.50	(200 × 50)	(200 × 50)
50	.70	(280 × 50)	(120 × 50)
50	.90	(360 × 50)	(40 × 50)
100	.10	(20 × 100)	(180 × 100)
100	.30	(60 × 100)	(140 × 100)
100	.50	(100 × 100)	(100 × 100)
100	.70	(140 × 100)	(60 × 100)
100	.90	(180 × 100)	(20 × 100)
200	.10	(10 × 200)	(90 × 200)
200	.30	(30 × 200)	(70 × 200)
200	.50	(50 × 200)	(50 × 200)
200	.70	(70 × 200)	(30 × 200)
200	.90	(90 × 200)	(10 × 200)

The other two factors, which pertain to the amount of noise in the data matrices, are:

(c) the *total amount of noise*, ε^{tot}, in the data, defined as $\varepsilon_1 + \varepsilon_2$, with ε_1 and ε_2 being defined as the (expected) ratio of the sum of squares of the noise (i.e., the discrepancies between the data and the truth) to the sum of squares of the data in $\mathbf{D}^1$ and $\mathbf{D}^2$, respectively,

$$\varepsilon_1 = E\left[\frac{\sum_{i=1}^{I}\sum_{k=1}^{K}(d_{ik}^1 - t_{ik}^1)^2}{\sum_{i=1}^{I}\sum_{k=1}^{K}(d_{ik}^1)^2}\right], \qquad \varepsilon_2 = E\left[\frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(d_{jk}^2 - t_{jk}^2)^2}{\sum_{j=1}^{J}\sum_{k=1}^{K}(d_{jk}^2)^2}\right], \tag{7}$$

with t_{ik}^1 and t_{jk}^2 denoting the entries of $\mathbf{T}^1$ and $\mathbf{T}^2$, respectively, and $E[.]$ the expected value. This factor was manipulated at two levels, .40 and .80;

(d) the *relative amount of noise*, ε^{rel}, for $\mathbf{D}^1$ to the total amount of noise ε^{tot}, defined as $\varepsilon_1/\varepsilon^{\text{tot}}$. This factor was manipulated at five levels, 0, .25, .50, .75, 1. Note that $\varepsilon^{\text{rel}} = 1$ ($\varepsilon^{\text{rel}} = 0$) implies that only $\mathbf{D}^1$ ($\mathbf{D}^2$) is subject to noise.

Table 2. The values of ε_1 and ε_2 resulting from an orthogonal combination of the total amount of noise, ε^{tot}, and relative amount of noise, ε^{rel}

ε^{tot}	ε^{rel}	ε_1	ε_2
.40	0	0	.40
.40	.25	.10	.30
.40	.50	.20	.20
.40	.75	.30	.10
.40	1	.40	0
.80	0	0	.80
.80	.25	.20	.60
.80	.50	.40	.40
.80	.75	.60	.20
.80	1	.80	0

In Table 2, the values of ε_1 and ε_2 are displayed that are obtained by orthogonally combining the factors ε^{tot} and ε^{rel}.

For each combination of the levels of the four manipulated factors, true component matrices $\mathbf{A}_1^{(T)}$, $\mathbf{A}_2^{(T)}$, and $\mathbf{B}^{(T)}$ with four components were constructed by independently drawing entries from a standard normal distribution. (Note that only true component matrices with four components were constructed because in a pilot study it was demonstrated that the performance of the (MxL)SCA-P strategy is hardly influenced by the true number of components S of the (MxL)SCA-P model.) Next, the true matrix couple ($\mathbf{T}^1$, $\mathbf{T}^2$) was computed by combining the true component matrices $\mathbf{A}_1^{(T)}$, $\mathbf{A}_2^{(T)}$, and $\mathbf{B}^{(T)}$ according to the (MxL)SCA-P decomposition rule in (2). Further, for each true matrix couple ($\mathbf{T}^1$, $\mathbf{T}^2$), a data matrix couple ($\mathbf{D}^1$, $\mathbf{D}^2$) was constructed by adding an $I \times K$ noise matrix $\mathbf{E}^1$ to $\mathbf{T}^1$ and a $J \times K$ noise matrix $\mathbf{E}^2$ to $\mathbf{T}^2$. The matrices $\mathbf{E}^1$ and $\mathbf{E}^2$ were created by drawing independently entries from a normal distribution with mean zero and variance chosen in such a way that the values for ε_1 and ε_2, as defined in (7), are equal to the specified design values (i.e., 0, .10, .20, .30, .40, .60, and .80).

For each cell of the design, the data generation procedure was repeated 50 times, resulting in 50 (replications) $\times$ 5 (array/total ratio) $\times$ 3 (size of the common mode) $\times$ 2 (total amount of noise) $\times$ 5 (relative amount of noise) = 7,500 different true and data matrix couples. Subsequently, after preprocessing the data (see Section 2), one SCA-P and one MxLSCA-P with the true number of components (i.e., four) were applied to each data matrix couple ($\mathbf{D}^1$, $\mathbf{D}^2$). To deal with local minima, a multi-start procedure was implemented for the MxLSCA-P strategy (see Section 3). In particular, for each data set ($\mathbf{D}^1$, $\mathbf{D}^2$), first the MxLSCA-P algorithm was performed three times, each time with different rational initial values for σ_1 and σ_2[1]. Next, the algorithm was performed 300 times with smart-random starting values (by perturbing each rational start 100 times) and 100 times with random starting values (for a discussion of different types of starting

[1]Rational starting values for σ_1 and σ_2 were obtained by performing (1) a separate PCA on $\mathbf{D}^1$ and $\mathbf{D}^2$, (2) a SCA-P on ($\mathbf{D}^1$, $\mathbf{D}^2$), and (3) a weighted SCA on [$\mathbf{D}^1$; $\mathbf{D}^2$] with the weight for each data matrix being the inverse of the total number of entries of the corresponding matrix, implying that all data matrices contribute equally to the analysis. Smart-random starting values were generated by adding randomly sampled numbers to these rational starting values, with these numbers being drawn from a $U(-\delta,\delta)$ distribution, δ being equal to the corresponding rational starting value divided by 2. Random starting values were obtained by independently sampling values from a $U(0,1)$ distribution.

procedures, see Ceulemans, Van Mechelen, & Leenen, 2007). From the 403 resulting solutions, the solution with the best value of the loss function in (4) was retained.

4.3. Results

To study the recovery performance of both data-analytic strategies at the level of the common component loadings **B**, the *goodness-of-recovery* statistic

$$\mathrm{GOR}_{\mathbf{B}} = \frac{\sum_{s=1}^{S} \left|\phi\left(\mathbf{b}_s^{(T)}, \mathbf{b}_s^{(M)}\right)\right|}{S}, \tag{8}$$

was used, with ϕ being the Tucker phi-coefficient (Tucker, 1951), $\mathbf{b}_s^{(T)}$ and $\mathbf{b}_s^{(M)}$ the sth component of the true loading matrix $\mathbf{B}^{(T)}$ and the loading matrix $\mathbf{B}^{(M)}$ resulting from the data analysis respectively, and S the number of components. To account for the rotational freedom of the components (see Section 2), each obtained component loading matrix $\mathbf{B}^{(M)}$ was (obliquely) rotated towards the corresponding true component loading matrix $\mathbf{B}^{(T)}$. The $\mathrm{GOR}_{\mathbf{B}}$ statistic takes values between 0 (no recovery at all) and 1 (perfect recovery). Note that the Tucker phi-coefficient is invariant under rescaling of the components and that reflectional freedom is dealt with by taking the absolute value of this coefficient.

When comparing the mean $\mathrm{GOR}_{\mathbf{B}}$ value for both strategies, it turned out that the MxLSCA-P strategy ($M = .974$) yields a better recovery than the SCA-P strategy ($M = .965$). This implies that a strategy in which entries belonging to noisy data matrices are downweighted in favour of entries from less noisy matrices, outperforms a strategy in which all entries contribute equally to the analysis.

To get an insight into how the relative performance of both strategies varies with the manipulated characteristics of the data, the following statistic was computed:

$$\mathrm{GOR}_{\mathbf{B}}^{\mathrm{diff}} = \mathrm{GOR}_{\mathbf{B}}^{\mathrm{MxLSCA\text{-}P}} - \mathrm{GOR}_{\mathbf{B}}^{\mathrm{SCA\text{-}P}}. \tag{9}$$

It turned out that for 90% of the data sets the MxLSCA-P strategy outperforms the SCA-P strategy (i.e., a $\mathrm{GOR}_{\mathbf{B}}^{\mathrm{diff}}$ value greater than zero), while the opposite is true in 10% of the cases (the mean $\mathrm{GOR}_{\mathbf{B}}^{\mathrm{diff}}$ value is .0086, with 95% confidence interval from .0083 to .0090). Further, an analysis of variance with $\mathrm{GOR}_{\mathbf{B}}^{\mathrm{diff}}$ as the dependent variable revealed that all main and interaction effects of the manipulated data characteristics are significant at the $\alpha = .05$ level. When only taking effects into account with an intra class correlation $\hat{\rho}_I$ (Haggard, 1958; Kirk, 1982) greater than or equal to .10, it further appears that the MxLSCA-P strategy especially outperforms the SCA-P strategy when the (total) amount of noise in the data is large ($\hat{\rho}_I = .18$) and when one data matrix is subject to more noise than the other ($\hat{\rho}_I = .13$). Both main effects, however, are qualified by a two-way relative amount of noise by total amount of noise interaction ($\hat{\rho}_I = .22$) and a three-way total amount of noise by array/total ratio by relative amount of noise interaction ($\hat{\rho}_I = .14$). In particular, MxLSCA-P especially outperforms SCA-P when the larger data matrix is most noise-prone and this pattern is more pronounced when the data contain more noise in general.

5. Application to empirical data

In the previous section, artificial data were used to determine which data-analytic strategy best discloses the true mechanisms underlying coupled data. In this section,

the performance of the SCA-P and MxLSCA-P strategies will be compared by means of empirical coupled data from the field of cross-cultural emotion psychology. The data set that will be used was collected by Diener and collaborators in the context of a larger study with respect to well-being (i.e., the International College Survey 2001). Here, we will focus on the part of the study pertaining to self-reported emotional experience. In particular, the data that will be analysed were collected by asking 9,300 participants (all of them college students) from 48 different nations to complete a paper-and-pencil questionnaire in which they had to indicate on a nine-point scale (1 = not at all, 9 = all the time) how often they had experienced each emotion from a set of 14 emotions (including love, pride, stress, and anger) in the past week (for a detailed description of the data set, see Kuppens *et al.*, 2006).

One of the major goals of cross-cultural emotion psychology is to unveil common underlying mechanisms that simultaneously account for intra-cultural individual differences in emotional experience in different cultures. Emotion psychologists often try to capture such mechanisms in terms of underlying dimensions on which all emotion labels may be positioned (see Russell, 1980; Watson, 2000). A key challenge then consists of disclosing these underlying dimensions and determining how they account for intra-cultural individual differences in emotional experience.

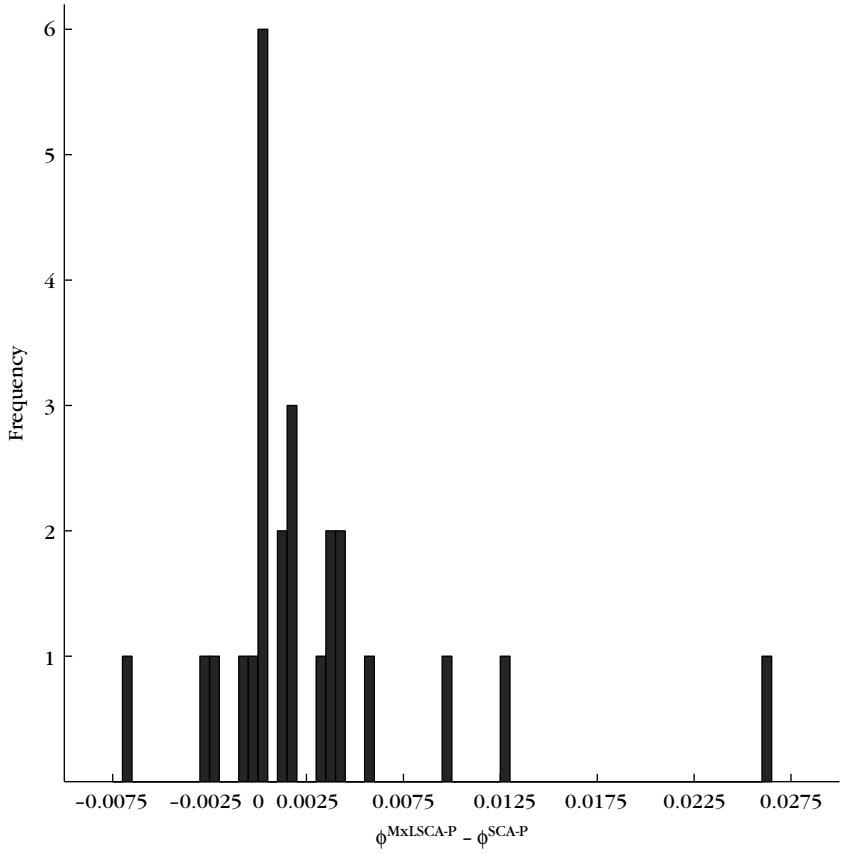


Figure 1. Histogram of differences in Tucker phi-coefficient ϕ between MxLSCA-P and SCA-P for all pairs of nations.

To determine which strategy best unveils the true dimensions underlying the data, these true dimensions should be known, which obviously is not the case. As a learned guess, however, we considered for this purpose the components that were obtained by applying a SCA-P to the complete data set. To compare the recovery performance of the SCA-P and MxLSCA-P strategies, subsequently both strategies were applied to pairs of nations. In particular, we wanted to focus on pairs of nations that differ considerably in amount of noise. This amount, however, was unknown too. As a way out, for each nation we computed how much variance in the data was explained by the 'true' SCA-P components as obtained from the analysis of the full data set; next, the amount of unexplained variance in each nation-specific data matrix was used as a learned guess for the amount of noise in that matrix (leaving aside variance accounted for by culture-specific components). The nations then were ranked in terms of amount of noise and the five best-fitting nations and five worst-fitting nations were selected. Subsequently, 25 pairs were constructed, consisting of a best-fitting and a worst-fitting nation. On each of these pairs, a SCA-P and an MxLSCA-P were performed, after appropriate preprocessing (see Section 2). Finally, the solution as resulting from these analyses was compared with the set of 'true' SCA-P components, making use of the Tucker phi-coefficient ϕ.

In Figure 1, a histogram of the differences in ϕ between MxLSCA-P and SCA-P for all pairs of nations is shown. It can be seen that for the majority of the pairs MxLSCA-P outperformed SCA-P, while the opposite was observed for a small number of pairs only; for a third group of pairs both strategies recovered the reference solution about equally well.

From Table 3, in which for each pair the estimated MxLSCA-P noise parameters σ_1 and σ_2 are displayed, it appears that the superior performance of MxLSCA-P over SCA-P can be explained by the former strategy downweighting in the analysis the most noise-prone data matrix in favour of the least noise-prone one. One can conclude that, in general, the MxLSCA-P strategy better discloses the 'true' mechanisms underlying the data.

Table 3. Estimated MxLSCA-P noise parameters for the least (σ_{least}) and most (σ_{most}) noise-prone data matrix for the 25 pairs of the cross-cultural emotion data set

Pair	σ_{least}	σ_{most}	Pair	σ_{least}	σ_{most}
1	.67	.88	14	.68	.80
2	.63	.88	15	.53	.80
3	.66	.88	16	.64	.83
4	.68	.82	17	.61	.82
5	.52	.81	18	.64	.83
6	.63	.86	19	.68	.80
7	.60	.84	20	.52	.79
8	.63	.85	21	.67	.83
9	.67	.82	22	.63	.82
10	.52	.81	23	.67	.81
11	.67	.86	24	.69	.76
12	.63	.85	25	.53	.74
13	.66	.85			

6. Discussion

When different coupled data blocks are subject to different amounts of noise, a standard simultaneous component analysis strategy (SCA-P), in which all data entries contribute

equally to the analysis, may lead to incorrect inferences with respect to the mechanisms underlying the data. In order to deal with this problem, in this paper a novel maximum-likelihood-based simultaneous component analysis strategy (MxLSCA-P) was proposed, in which data entries belonging to noisy matrices are downweighted. Both in a simulation study and in an application to empirical data, MxLSCA-P has been shown to outperform SCA-P, especially in the case of data sets subject to a large overall amount of noise and with the largest data block being relatively more noise-prone.

To understand why MxLSCA-P outperforms SCA-P, it should be observed that the (deterministic) SCA-P model is equivalent to a stochastic version of model (1) in which the noise entries are assumed to be normally distributed with mean zero and homoscedastic variance (i.e., $e^1_{ik} \sim N(0,\sigma)$ and $e^2_{jk} \sim N(0,\sigma)$). As a consequence, when coupled data matrices are subject to different amounts of noise (which often occurs in psychology) SCA-P, unlike MxLSCA-P, is a misspecified model, resulting in MxLSCA-P outperforming SCA-P in disclosing the underlying true structure. Otherwise, one may note that MxLSCA-P especially outperforms SCA-P in those conditions where the difference in noise between both data matrices is large, because in that case the misspecification of SCA-P becomes worse.

In the remainder of this section, first the relations of MxLSCA-P to other models will be discussed. Then, generalizations of MxLSCA-P will be outlined.

6.1. Relations to other models

MxLSCA-P is a novel model, but has interesting relationships with the family of multigroup common factor models (Jöreskog, 1971; Sörbom, 1974), which belongs to the broader collection of multigroup structural equation models (Muthén, 1989; for a discussion of the general differences between component models, such as MxLSCA-P, and factor analysis models we refer to Jolliffe, 2002; Widaman, 1993, 2007; Timmerman, Ceulemans, Lichtwarck-Aschoff, & Vansteelandt, 2009). The closest counterpart of MxLSCA-P in this family can be obtained by imposing absence of offset terms (i.e., intercepts) and metric or weak factorial invariance, which implies that the loadings on each variable are equal across data blocks (Widaman & Reise, 1997; Hessen, Dolan, & Wicherts, 2006).[2] Both this model and MxLSCA-P rely on a decomposition of the data as in (2). Further, in both models it is assumed that the residuals e^1_{ik} and e^2_{jk} are independently sampled from a normal distribution with zero mean and a variance that is allowed to differ across data matrices. The two models, however, differ in that in the multigroup common factor model the noise variance is allowed to vary across columns (i.e., variables) within a particular data matrix, whereas in MxLSCA-P this variance is assumed to be constant within each matrix; as such, MxLSCA-P models the noise in a more parsimonious way than the multigroup common factor model. Further, in the multigroup common factor model the factor scores are assumed to follow a normal distribution, whereas such an assumption is not included in MxLSCA-P. Next, in multigroup common factor methods, unlike in MxLSCA-P where the component scores are obtained during the analysis, the factor scores are estimated in a second phase only, after the actual model estimation (i.e., factor scores are not considered as model parameters), and are often not a primary target

[2]Note that this also boils down to a multigroup common factor model with (a special case of) strong factorial invariance (Meredith, 1993). In particular, strong factorial invariance is obtained when loadings and offset terms (which are zero here) are equal across groups.

of interest. Finally, multigroup common factor models are often used in a confirmatory way, whereas MxLSCA-P is more suited to reveal, in an exploratory way, the structure that (a large number of) coupled data matrices have in common. From the above, it can be concluded that MxLSCA-P combines features of both a component model (i.e. components being linear combinations of the original variables) and a common factor model (i.e., explicit modelling of the noise, but in a more parsimonious way).

With respect to the empirical implications of the differences between component analysis and common factor analysis, it is well known that when data generated under a common factor model are analysed with a component model (e.g., PCA), this leads to a poor recovery of the population parameters (Widaman, 1993). In reality, however, a common factor model will never hold perfectly for data at hand, implying that in some empirical applications component models may better estimate these underlying parameters (Timmerman *et al.*, 2009). In the literature, many attempts have been undertaken to compare component models (e.g., PCA) to common factor models with respect to the accuracy in reproducing population loadings, resulting in some authors stating that loadings obtained with PCA are biased (e.g., Snook & Gorsuch, 1989; Widaman, 1993), while others argue that both models often yield an identical representation of the data (e.g., Velicer & Jackson, 1990; Velicer & Fava, 1998; for an overview of the discussion, see Widaman, 2007). For the empirical comparison of MxLSCA-P and the multigroup common factor model, a similar pattern of results may be expected. A possible advantage of MxLSCA-P, however, may be that this model, as opposed to the multigroup common factor model, does not assume normality of the scores. As such, the former model is more flexible in this respect than the latter and MxLSCA-P may be expected to outperform the multigroup common factor model in recovering the scores when these are not normally distributed.

6.2. Generalizations of MxLSCA-P

In the stochastic part of the MxLSCA-P model, it was assumed that the noise was unstructured, implying the noise within a data matrix being homoscedastic and uncorrelated (i.e., independently and identically distributed). In practice, however, this is a rather restrictive assumption. First of all, noise may be heteroscedastic (i.e., variables within a matrix may be subject to different amounts of noise), because, for example, some variables in a data matrix may be measured more reliably than others. In some applications, noise heteroscedasticity may also occur at the level of the entries. A special case of the latter is encountered when entries that take a larger value are subject to a larger amount of noise. Besides heteroscedasticity, the noise in a data matrix may also be correlated. This may, for example, occur when for the same set of variables individuals are measured at different time points, with as a possible consequence dependence (i.e., a correlation) between the noise on subsequent measurements of a given individual. To account for heteroscedastic and/or correlated noise within the context of MxLSCA-P one may adapt the stochastic part of the MxLSCA-P model by specifying a heterogeneous and/or correlated noise pattern. One way to obtain this is by generalizing the model proposed by Wentzell, Andrews, Hamilton, Faber, and Kowalski (1997) to the case of coupled data matrices.

Acknowledgements

The first author is a Research Assistant of the Fund for Scientific Research – Flanders (Belgium). The research reported in this paper was partially supported by the Research Council of K. U.

Leuven (GOA/2005/04 and EF/2005/07, 'SymBioSys') and by IWT Flanders (IWT/060045/SBO Bioframe). We thank Ed Diener (and collaborators) and Peter Kuppens for providing us with the data set used in this paper.

References

Ceulemans, E., & Kiers, H. A. L. (2009). Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical and Statistical Psychology*, *62*, 601–620. doi:10.1348/000711008X369474

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2007). The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika*, *72*, 377–391. doi:10.1007/s11336-007-9000-9

Haggard, E. A. (1958). *Intraclass correlation and the analysis of variance*. New York: Dryden.

Hessen, D. J., Dolan, C. V., & Wicherts, J. M. (2006). The multigroup common factor model with minimal uniqueness constraints and the power to detect uniform bias. *Applied Psychological Measurement*, *30*, 233–246. doi:10.1177/0146621605279760

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag.

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

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, *14*, 105–122. doi:10.1002/1099-128X(200005/06)14:3<105::AID-CEM582>3.0.CO;2-I

Kiers, H. A. L., & Ten Berge, J. M. F. (1994). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, *47*, 109–126.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks/Cole.

Kuppens, P., Ceulemans, E., Timmerman, M. E., Diener, E., & Kim-Prieto, C. (2006). Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology*, *37*, 491–515. doi:10.1177/0022022106290474

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525–543. doi:10.1007/BF02294825

Muthén, B. O. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika*, *54*, 557–585. doi:10.1007/BF02296397

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1151–1178.

Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way analysis with applications in the chemical sciences*. Chichester: Wiley.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*, 148–154. doi:10.1037/0033-2909.106.1.148

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.

Timmerman, M. E., Ceulemans, E., Lichtwarck-Aschoff, A., & Vansteelandt, K. (2009). Multilevel component analysis for studying intra-individual variability and inter-individual differences. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudhary (Eds.), *Dynamic process methodology in the social and developmental sciences* (pp. 291–318). New York: Springer.

Timmerman, M. E., & 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 Rapport # 984). Washington, DC: Department of the Army.

Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. A. L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, *10*, 246. doi:10.1186/1471-2105-10-246

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231–251. doi:10.1037/1082-989X.3.2.231

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*, 1–28. doi:10.1207/s15327906mbr2501_1

Watson, D. (2000). *Mood and temperament*. New York: Guilford.

Wentzell, P. D., Andrews, D. T., Hamilton, D. C., Faber, K., & Kowalski, B. R. (1997). Maximum likelihood principal component analysis. *Journal of Chemometrics*, *11*, 339–366. doi:10.1002/(SICI)1099-128X(199707)11:4<339::AID-CEM476>3.0.CO;2-L

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Different bias in representing model parameters. *Multivariate Behavioral Research*, *28*, 263–311. doi:10.1207/s15327906mbr2803_1

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100. Historical developments and future directions* (pp. 177–203). Mahwah, NJ: Erlbaum.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Received 4 September 2009; revised version received 10 May 2010