Stat Methods Appl
DOI 10.1007/s10260-012-0215-2

Multilevel dimensionality-reduction methods

Pietro Giorgio Lovaglio · Giorgio Vittadini

Accepted: 16 September 2012

Abstract When data sets are multilevel (group nesting or repeated measures), different sources of variations must be identified. In the framework of unsupervised analyses, multilevel simultaneous component analysis (MSCA) has recently been proposed as the most satisfactory option for analyzing multilevel data. MSCA estimates submodels for the different levels in data and thereby separates the “within”-subject and “between”-subject variations in the variables. Following the principles of MSCA and the strategy of decomposing the available data matrix into orthogonal blocks, and taking into account the between- and the within data structures, we generalize, in a multilevel perspective, multivariate models in which a matrix of response variables can be used to guide the projections (formed by responses predicted by explanatory variables or by a limited number of their combinations/composites) into choices of meaningful directions. To this end, the current paper proposes the multilevel version of the multivariate regression model and dimensionality-reduction methods (used to predict responses with fewer linear composites of explanatory variables). The principle findings of the study are that the minimization of the loss functions related to multivariate regression, principal-component regression, reduced-rank regression, and canonical-correlation regression are equivalent to the separate minimization of the sum of two separate loss functions corresponding to the between and within structures, under some constraints. The paper closes with a case study of an application focusing on the relationships between mental health severity and the intensity of care in the Lombardy region mental health system.

P. G. Lovaglio (✉) · G. Vittadini
Department of Quantitative Methods, University Bicocca-Milan, V. Sarca, 202, 20143 Milan, Italy
e-mail: piergiorgio.lovaglio@unimib.it

G. Vittadini
e-mail: giorgio.vittadini@unimib.it

Published online: 27 September 2012

Keywords Dimensionality-reduction methods · Reduced-rank regression · Principal-component regression · Multilevel simultaneous component analysis

1 Introduction

Many data sets generated in the applied sciences are multilevel data sets, in which variation occurs simultaneously on different levels (e.g., variation between individuals and variation in time).

Examples of multilevel-type problems include the monitoring of hospital patients in time, the monitoring of batch processes in process industries, and the economic time-series analysis of multiple countries, economic branches, or companies. Proper statistical analyses must take into account such different types of variation in data. For example, if statistical techniques that give a simplified. lower-dimensional representation of the variation present in data, such as principal-component analysis (PCA), are used for the analysis of multilevel data, the different types of variation in the multilevel data will not be separated, and the obtained principal components will describe a mixture of different types of variation (e.g., the time-dynamic variation and the variation between individuals). Both types of variation in the data are confounded within these methodologies, which seriously hampers the interpretation of the phenomena underlying the variation in the data.

Further, if the multivariate data collected jointly describe the phenomenon of interest, and if the primary aim is to summarize this structure with reference to a dependent variable, several statistical regression models that take into account the multilevel nature of the data have been proposed. The standard methods for clustered-data regression analysis postulate models relating covariates to the response without regard to "between"- and "within"-cluster covariate effects. Implicit in these analyses is the assumption that these effects are identical (Neuhaus and Kalbfleisch 1998).However, the most popular data-analytic approach to multilevel data is multilevel regression modeling (Bryk and Raudenbush 1992; Hox 2002; Reise and Duan 2003; Snijders and Bosker 1999).

In multilevel regression, the relationships among observed explanatory and dependent variables are modeled, thereby taking into account the multilevel structure of the data. Unfortunately, this approach has three main limitations. First, it is based on stringent assumptions about the distribution (and the mutual relationship) of random errors at different levels. Second, in the presence of multilevel data with several dependent variables, multilevel regression cannot be used. Third, when the number of explanatory variables is large, with some of them possibly highly correlated with each other, (multilevel) regression fails. Otherwise, it may be advantageous to predict the responses with fewer linear combinations of the explanatory variables, rather than with the original predictors.

Several methods have been proposed in the literature to fit multiple dependent variables in a multigroup or multilevel perspective, such as multilevel factor analysis (Muthén 1991) or multilevel covariance-structure analysis (Joreskog 1993; Muthén 1994; Goldstein and McDonald 1988). Nonetheless, these methods were developed from the perspective of covariance structure analysis, so they typically assume reflective indicators (latent factors are underlying causal antecedents of manifest indicators, as in the classical factor model) and suffer from the problem of factor score indeterminacy (Schönemann and Steiger 1976).

An alternative approach to examining the multivariate structure of data is component analysis (CA; Meredith and Millsap 1985) which replaces factors by linear combinations of observed variables. In the presence of multiple dependent variables, CA aims to model observed data to predict the responses by a limited number (d) of components, called linear composites (LCs), using a least-squares fitting procedure. The predictions are obtained from a subspace of the space spanned by the explanatory variables. Such methods are referred to as dimensionality-reduction methods (DRMs), which build a sequence of orthogonal linear composites, of which an optimal number will be used for prediction (Burnham et al. 1996).

Specifically, DRMs select a few linear composites (combinations of the original variables) in the predictor space as the new predictor space and then regress the response variables on this reduced predictor space. The most commonly used DRMs for prediction are principal-component regression (PCR; Massey 1965), reduced-rank regression (RRR; Izenman 1975), canonical-correlation regression (CCR; Anderson 1951) and partial least squares (PLS; Wold 1982). Each method involves using a subspace of the $\mathbf{X}$-space as a new reduced predictor space.

Technically, in PCR, the first d principal components of the $\mathbf{X}$ space define the reduced predictor space for regression, whereas for RRR and CCR the choice also depends on the $\mathbf{Y}$ space. RRR amounts to extracting a series of linear composites from the $\mathbf{X}$ set (redundancy variates) in such a way that they are mutually orthogonal and successively explain the maximum variance of the set of $\mathbf{Y}$ variables. Specifically, the objective function for redundancy variates is to maximize the fraction of variation in each $\mathbf{Y}$ variable explained by linear regressions on d redundancy variates. In RRR, the number of components d is restricted to being less than the minimum between the ranks of $\mathbf{X}$ and $\mathbf{Y}$. CCR is based on the technique of canonical-correlation analysis (CCA), with the objective of finding series of composites (canonical variates) from $\mathbf{X}$ and $\mathbf{Y}$ (orthogonal variates in the column spaces of $\mathbf{X}$ and $\mathbf{Y}$, respectively) that are most highly correlated with one another.

Specifically, RRR and CCR obtain estimates of LCs in different ways. In RRR, a subspace of the column space of predictors spanned by the first d-predicted values of $\hat{\mathbf{Y}}$ is used as the predictor space: the least-squares estimates of LCs are given by the first d principal components of the projection of responses $\mathbf{Y}$ onto the column space of predictors $\mathbf{X}$ ($\hat{\mathbf{Y}}$). The RRR composites are the same as those obtained with the method of redundancy analysis (RA; Van den Wollenberg 1977). In CCR, the first d canonical variates of $\mathbf{X}$, spanning a d dimensional subspace of the column space of $\mathbf{X}$, are used as the predictor space for regression. In addition, the weights that define LCs in RRR and CCR are least-squares solutions: in the RRR, solutions minimize the squared errors of prediction using d redundancy variates, whereas in CCR the canonical variables in the $\mathbf{X}$ space are obtained as generalized least-squares solutions to the RRR model (Abraham and Merola 2005).

LCs for RRR and CCR are obtained through the maximization of certain objective functions of the prediction errors, whereas in PCR they are obtained by objective functions that cannot be related to the prediction of the responses. On the other hand, PLS is an algorithmic method with an objective function that cannot be expressed in a closed form, except for the extraction of the first factor (or when $\mathbf{Y}$ is unidimensional). In the context of unsupervised analysis for a single data set, constrained principal-component analysis (CPCA; Takane and Hunter 2001) has been suggested in the literature as a useful method.

CPCA allows the incorporation of external information into PCA of a data matrix. When the rows of a data matrix represent subjects, as external information we may take a vector of a categorical variable indicating the subjects' group membership. CPCA first decomposes the data matrix according to the external information (external analysis), and then applies PCA to decomposed matrices (internal analysis), analyzing the differences among the groups.

Recently, Timmerman (2006) has proposed a method entitled multilevel component analysis (MLCA), in which various component submodels give a summary of the different types of variation present in the data. For ease of interpretation, a constrained version of MLCA, called multilevel simultaneous-component analysis (MSCA), has been proposed, with the objective of approximating the data and explaining the sum of squares (across subjects and time points) of the between-subject and within-subject variations as specifically as possible with composites. Once the static variation has been removed from the data, such composites are useful for describing the static differences between the individuals (the between-individual variation) and the dynamic variation of the individuals (the within-individual variation). Unfortunately, MSCA is restricted to unsupervised analyses and has not been proposed in the context of DRMs, where the primary aim is to obtain a few linear composites of explanatory variables for use as reduced (predictor) space for regression.

In the present paper, we propose the use of MSCA principles (based on the decomposition of observed data matrices in orthogonal blocks, taking into account the between- and within-data structures) to generalize, in a multilevel framework, both the multivariate regression (MR) model and the best known DRMs, namely the principal-component regression model, the reduced-rank regression model, and the canonical-correlation regression model.

In the literature, multilevel versions of DRMs have been proposed by Hwang et al. (2007) and by de Noord and Theobald (2005) in a PLS framework. In the context of experimental design with hierarchical data, they have been proposed by Smilde et al. (2005) and Jansen et al. (2005) and further worked out by Thissen et al. (2009) for PLS regression. We discuss such approaches in the concluding section.

The remainder of the article is organized as follows. Section 2 begins with an explanation of the MSCA method, and Sect. 3 then examines the multivariate regression model in the multilevel perspective. Sections 4–6 concern DRMs, and Sect. 7 discusses rotational freedom. Section 8 presents an application focused on the relationships between mental illness severity and the intensity of the provided care, based on data collected in the Lombardy Region of Italy. The final section provides conclusions.

2 Multilevel component analysis

Let $\mathbf{Y}$ $(\Sigma_i K_i \times q)$ be the total data matrix containing K_i measurement time points for individual $i\,(i = 1, \ldots, I)$ on q variables $(j = 1, \ldots, q)$ centered on the overall mean for variable j across all individuals, where $\Sigma_i K_i$ is the total number of observations. The specification of the MLCA model for data matrix $\mathbf{Y}_i(K_i \times q)$, containing the part of $\mathbf{Y}$ with data from individual $i\,(i = 1, \ldots, I)$, is

$$\mathbf{Y}_i = \mathbf{1}_{Ki} \quad \mathbf{f}'_{ib}\mathbf{B}'_b + \mathbf{F}_{iw}\ \mathbf{B}'_{iw} + \mathbf{E}_i \quad \text{subject to} \quad \Sigma_i\, K_i\, \mathbf{f}_{ib} = 0 \ \text{ and } \mathbf{1}'_{Ki}\mathbf{F}_{iw} = \mathbf{0}', \quad (1)$$

where $\mathbf{1}_{K_i}$ denotes a size K_i column vector of unitary values (K_i is the number of measurement time points for individual i), $\mathbf{f}_{ib}(d_b \times 1)$ contains the between-individual scores of individual i for d_b retained between components, and $\mathbf{B}_b(q \times d_b)$ is the between-individual loading matrix. On the other hand, $\mathbf{F}_{iw}(K_i \times d_{w,i})$ contains the within-individual scores of subject i for d_w retained in components, $\mathbf{B}_{iw}$ $(q \times d_{w,i})$ is the within-subject loading matrix for individual i, and $\mathbf{E}_i(K_i \times q)$ contains the residuals.

Equation (1) shows that the data in matrix $\mathbf{Y}_i$ are reconstructed in the MLCA model by a between-individual part $\mathbf{1}_{Ki}\mathbf{f}'_{ib}\mathbf{B}'_b$ that is equal for all samples belonging to an individual i and a within-individual part $\mathbf{F}_{iw}\ \mathbf{B}'_{iw}$, which is different for each individual in the data. It is then assumed that the between-loading matrix $\mathbf{B}_b$ is invariant over individuals and that the within-loading matrices are time invariant but not invariant over individuals.

The constraint given to $\mathbf{f}_{ib}$ imposes the function of the between-individual scores being able to describe the deviation of each individual from the mean values of each column of $\mathbf{Y}_i$, whereas the matrix with component scores $\mathbf{F}_{iw}$ of individual i is centered column-wise (individual mean-centered). In the general MLCA model of Eq. (1), the within-loading matrices $\mathbf{B}_{iw}$ may vary across individuals, which complicate the interpretation of the different within-loading matrices. To this end, Timmerman, using the approach adopted in simultaneous-component analysis (SCA; Timmerman and Kiers 2003), specifies a constrained version of MLCA in which the within-loading matrix is constrained to be equal across individuals.

Formally, by imposing in Eq. (1) the constraint of equivalent loading matrices for all individuals [$\mathbf{B}_{iw}= \mathbf{B}_w$ where $\mathbf{B}_w$ is $(q \times d_w)$ matrix containing the within-individual loadings for $i = 1, \ldots, I$, and $\mathbf{F}_{iw}$ becomes a $(K_i \times d_w)$ matrix], the model for MSCA is defined. In MSCA, since the same weighted sum of scores on the variables is used on the components from the variables, the interpretation of the between-individual and within-individual components is equal for all individuals. The MSCA model can be fitted to the data using a least-squares method, which minimizes the sum of squares (SSQ) of the loss function f_{MSCA} as follows:

$$f_{\mathrm{MSCA}}\,(\mathbf{f}_{ib}, \mathbf{B}_b, \mathbf{F}_w, \mathbf{B}_w) = \Sigma_i \mathrm{SSQ}[\mathbf{Y}_i - (\mathbf{1}_{Ki}\ \mathbf{f}'_{ib}\ \mathbf{B}'_b + \mathbf{F}_{iw}\ \mathbf{B}'_w)] \quad i = 1, \ldots I \tag{2}$$

subject to $\Sigma_i K_i \mathbf{f}_{ib} = 0$ and $1'_{K_i}\mathbf{F}_{iw} = \mathbf{0}'$, where SSQ($\mathbf{A}$) denotes the sum of squares of the elements of matrix $\mathbf{A}$. The constraints imposed on the between- and within-component scores are sufficient to ensure that the between part and the within part of the model are uniquely separated and can be separately estimated. MSCA has the fundamental property that all within-subject models are orthogonal to the between-individual model—i.e., the components $\mathbf{f}_{\mathrm{ib}}$ and $\mathbf{F}_{\mathrm{iw}}$ are orthogonal within each subject and for all subjects.

MSCA has been successfully applied to different examples of multilevel data, such as analysis of data from a metabolic study of monkeys (Jansen et al. 2005), for analyzing the mood structure across time among individuals diagnosed with Parkinson's disease (Timmerman 2006) and to explore the dynamics of chemical processes (de Noord and Theobald 2005). The MSCA model is also the basis for a range of methods, which differ in the additional constraints that are defined for covariances of the within-individual scores $\mathbf{F}_{\mathrm{iw}}$. The most general MSCA model is the MSCA-P model, which does not impose constraints on the inter-individual variability (variances and covariances) in intra-individual structure (within-component scores).

The result is a two-level MSCA-P model of time-resolved measurements on multiple individuals consisting of two submodels: a PCA model describing the static differences between individuals (the between-individual variations) and a PCA model describing the dynamic variations of the individuals (the within-individual variations). Next, we show how to estimate the MSCA-P model parameters.

2.1 Obtaining the MSCA-P model parameters

The between-individual model and the within-individual model can be determined after a decomposition of matrix $\mathbf{Y}$, where matrix $\mathbf{Y}$ $(\Sigma_{\mathrm{i}} K_i \times q)$ contains the grand mean-centered data. Each matrix $\mathbf{Y}_{\mathrm{i}}$ can be decomposed as

$$\mathbf{Y}_{\mathrm{i}} = \mathbf{Y}_{\mathrm{c,i}} + 1\mathrm{K}_{\mathrm{i}}\mathbf{m}'_{\mathrm{iy}}, \tag{3}$$

where $\mathbf{m}'_{\mathrm{iy}}$ is a $(1xq)$ row vector containing the column means of matrix $\mathbf{Y}_{\mathrm{i}}$, and $\mathbf{Y}_{\mathrm{c,i}}$ is the $(K_i \times q)$ matrix containing only the dynamic within-individual variation belonging to individual i. The I vectors $\mathbf{m}'_{\mathrm{iy}}$ can be concatenated into an $(I \times q)$ matrix $\mathbf{M}_{\mathrm{y}}$, which now contains the non-dynamic differences between the individuals (between-individual variations).

As proven by Timmerman (2006), MSCA-P has the main property of orthogonality of the two parts of the model. In fact, the loss function of Eq. (2) can be expressed by the following equivalent formulation:

$$\Sigma_{\mathrm{i}}\mathrm{SSQ}(\mathbf{1}_{\mathrm{Ki}}\ \mathbf{m}'_{\mathrm{iy}} - \mathbf{1}_{\mathrm{Ki}}\ \mathbf{f}'_{\mathrm{ib}}\mathbf{B}'_{\mathrm{b}}) + \Sigma_{\mathrm{i}}\mathrm{SSQ}\left(\mathbf{Y}_{\mathrm{c,i}} - \mathbf{F}_{\mathrm{iw}}\,\mathbf{B}'_{\mathrm{w}}\right).\quad i = 1,\ldots I \tag{4}$$

Because these two functions (the between and the within parts, respectively) deal with different parameter sets, they can all be minimized separately. Thus, the component scores and loadings can be separately estimated in both structures by using $\mathbf{M}_{\mathrm{y}}$ and $\mathbf{Y}_{\mathrm{c}}$, where $\mathbf{Y}_{\mathrm{c}}$ is a $(\Sigma_{\mathrm{i}} K_i \times q)$ matrix in which all matrices $\mathbf{Y}_{\mathrm{c,i}}$ are vertically concatenated (Timmerman and Kiers 2003; Timmerman 2006).

Specifically, the between-individual model can be determined by performing a PCA on $\mathbf{WM}_y$, where $\mathbf{W}$ denotes an $(I \times I)$ diagonal (weight) matrix with $w_{ii} = \sqrt{K_i}$, taking into account the possible unequal numbers of measurement occasions for individuals. The within-individual model can be calculated by fitting a PCA on $\mathbf{Y}_c$. Hence, the MSCA-P model is a combination of two PCA models describing as much variation as possible on both the between- and within-individual levels.

Furthermore, since each data matrix ($\mathbf{Y}$ and $\mathbf{X}$) is reconstructed following the decomposition of Eq. (3), the sums-of-squares SSQ ($\mathbf{Y}_c$) and SSQ ($\mathbf{WM}_y$) can be used in the MSCA-P framework to determine the magnitudes of the within- and between-individual variations for each block and for each variable, as well as to determine the percentages of total variances taken into account by the retained d_b (between-individuals) and d_w (within-individuals) components. See Timmerman (2006) for full details.

3 Multilevel multivariate regression (MLMR)

As shown, the primary goal of MSCA is to decompose the overall variability of observed variables into two separate components: the between-subject and the within-subject contributions (in a longitudinal perspective), or into that of the between-groups (when subjects are nested in groups). However, this property substantially arises from the fact that, under specific constraints on the between- and within-component scores, the MSCA-P extracts linear components in both structures, which are mutually orthogonal in the same dimension and for all dimensions, since these components are extracted from the mutually orthogonal matrices $\mathbf{M}_y$ and $\mathbf{Y}_c$. Hence, following the strategy of decomposing the data matrix into two orthogonal blocks, taking into account the between and within data structures (Eq. 3), we can generalize, in a multilevel perspective, statistical models in which projections are guided in meaningful directions by a matrix of explanatory variables.

In applications of regression methods, it is typically assumed that between-subject and within-subject comparisons will produce the same estimated effect on the risk factor. Instead, we fit models that allow separate between- and within-cluster covariate effects: separate effects can be estimated in a linear model by partitioning the predictors' matrix ($\mathbf{X}_i$) into between-subject ($\mathbf{1}_{K_i}\mathbf{m}'_{ix}$) and within-subject ($\mathbf{X}_{c,i}$) structures (Scott and Holt 1982; Neuhaus and Kalbfleisch 1998). Specifically, let $\mathbf{X}_i$ and $\mathbf{Y}_i$ be data matrices of full rank, containing K_i measurements on p and q variables (assuming that $p \leq q$), respectively, for subject i (where the overall matrices $\mathbf{X}$ and $\mathbf{Y}$—collecting vertically $\mathbf{X}_i$ and $\mathbf{Y}_i$—are column-wise centered across all individuals).

Replacing in the i-th block the decomposition of the predictor matrix $\mathbf{X}_i$ in the between and within predictors' scores (as in Eq. 3), we fit the MR model of the form

$$\mathbf{Y}_i = \mathbf{1}_{Ki}\,\mathbf{m}'_{ix}\mathbf{A}_b + \mathbf{X}_c, i\mathbf{A}_w + \mathbf{E}_i, \quad i = 1, \dots I \tag{5}$$

where the regression coefficients are partitioned into two separate blocks to allow different effects of predictors on responses. The coefficients of the $p \times q$ matrix $\mathbf{A}_b$indicate the effects of the covariates (which are supposed to be invariant over subjects) based on a comparison between subjects (i.e., the coefficient in column j and row s measures the difference in the j-th response variable between subjects whose average s-th covariate value differs by 1 unit).

The coefficients of the $p \times q$ matrix $\mathbf{A}_{\mathrm{w}}$ measure (for all blocks) the effects of the covariates based on comparisons within subjects (i.e., measuring the difference in the responses between time points with covariate values that differ by 1 unit within a subject). Finally, $\mathbf{E}_{\mathrm{i}}$ is the error matrix. In order to obtain two separate MR models (the between and within models), we must explicitly demonstrate that the loss function related to Eq. (5) can be partitioned into two orthogonal components that refer to both structures. Applying the decomposition (3) to matrices $\mathbf{Y}_{\mathrm{i}}$, the loss function related to Eq. (5) becomes

$$\begin{aligned} &\Sigma_{\mathrm{i}}\mathrm{SSQ}(\mathbf{Y}_{\mathrm{c,i}} + \mathbf{1}_{\mathrm{Ki}}\,\mathbf{m}'_{\mathrm{iy}} - \mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{ix}}\mathbf{A}_{\mathrm{b}} - \mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}}) \\ &\quad = \mathrm{Tr}\Sigma_{\mathrm{i}}[\mathbf{Y}'_{\mathrm{c,i}}\mathbf{Y}_{\mathrm{c,i}} + 2\mathbf{Y}'_{\mathrm{c,i}}\mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{iy}} - 2\mathbf{A}'_{\mathrm{b}}\mathbf{m}_{\mathrm{ix}}\mathbf{1}'_{\mathrm{Ki}}\mathbf{Y}_{\mathrm{ci}} - 2\mathbf{Y}'_{\mathrm{c,i}}\mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}} \\ &\qquad + \mathbf{m}_{\mathrm{iy}}\mathbf{1}'_{\mathrm{Ki}}\mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{iy}} - 2\mathbf{m}_{\mathrm{iy}}\mathbf{1}'_{\mathrm{Ki}}\mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{ix}}\mathbf{A}_{\mathrm{b}} - 2\mathbf{m}_{\mathrm{iy}}\mathbf{1}'_{\mathrm{Ki}}\mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}} \\ &\qquad + 2\mathbf{A}'_{\mathrm{b}}\mathbf{m}_{\mathrm{ix}}\mathbf{1}'_{\mathrm{Ki}}\mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}} + \mathbf{A}'_{\mathrm{b}}\mathbf{m}_{\mathrm{ix}}\,\mathbf{1}'_{\mathrm{Ki}}\,\mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{ix}}\mathbf{A}_{\mathrm{b}} + \mathbf{A}'_{\mathrm{w}}\mathbf{X}'_{\mathrm{c,i}}\mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}}]. \end{aligned} \tag{6}$$

Due to the orthogonality of some terms ($\mathbf{Y}'_{\mathrm{c,i}}1_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{iy}} = \mathbf{0}$), and because$\mathbf{Y}_{\mathrm{c,i}}$ is centered over individual i ($\mathbf{1}'_{\mathrm{K_i}}\mathbf{Y}_{\mathrm{c,i}} = \mathbf{0}'$), the second and the third terms of Eq. (6) vanish. Furthermore, under the constraint that each column of the fitted matrix $\mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}}$ (least-squares projections in the within space) is individual mean-centered ($1'_{\mathrm{Ki}}\mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}} = \mathbf{0}'$), the seventh and eighth terms of Eq. (6) vanish. This implies that minimizing Eq. (6) is equivalent to minimizing the sum of the following two functions:

$$\Sigma_{\mathrm{i}}\mathrm{SSQ}(\mathbf{Y}_{\mathrm{c,i}} - \mathbf{X}_{\mathrm{c,i}}\mathbf{A}_{\mathrm{w}}) + \mathrm{SSQ}(\mathbf{W}[\mathbf{M}_{\mathrm{y}} - \mathbf{M}_{\mathrm{x}}\mathbf{A}_{\mathrm{b}}]), \tag{7}$$

where the second term of Eq. (7) arises because $\mathbf{1}_{\mathbf{K}_{\mathrm{i}}}\mathbf{m}'_{\mathrm{ix}}$ and $\mathbf{1}_{\mathbf{K}_{\mathrm{i}}}\mathbf{m}'_{\mathrm{iy}}$ arecomposed of $\mathrm{K_i}$ constant terms in the i-th block and thus only single rows $\mathbf{m}'_{\mathrm{ix}}$ and $\mathbf{m}'_{\mathrm{iy}}$ (concatenated vertically in matrices $\mathbf{M}_{\mathrm{y}}$ and $\mathbf{M}_{\mathrm{x}}$) contribute to the sums of squares.

Because these two functions deal with different parameter sets, the between and the within parts, they can be minimized separately. Consequently, the between and the within MR can be separately estimated: in the within structure, by regressing $\mathbf{Y}_{\mathrm{c}}$ on $\mathbf{X}_{\mathrm{c}}$ (obtained by vertically concatenating the I blocks $\mathbf{Y}_{\mathrm{c,i}}$ and $\mathbf{X}_{\mathrm{c,I}}$, respectively), and in the between structure by regressing $\mathbf{WM}_{\mathrm{y}}$ on $\mathbf{WM}_{\mathrm{x}}$. Furthermore, the total sums-of-squares can be separated into a part of between-individual sums-of-squares and a part of within-individual sums-of-squares, and these terms can be used to determine the magnitudes of the within- and between-individual fit. As in MR, the percentage of explained variation can be used as a measure of fit of the model to the data. Because the two-level MR model consists of two independent submodels, three different criteria of fit can be defined: the percentage (1) of the explained variation of the within-individual submodel, (2) of the between-individual submodel, and of the entire MR model.

4 Multilevel principal-component regression (MLPCR)

As mentioned previously, DRMs aim to predict the responses with fewer linear composites of the explanatory variables. Let $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_d) = \mathbf{X}(\mathbf{b}_1, \ldots, \mathbf{b}_d) = \mathbf{XB}$ be the LCs of the predictor space to be determined subject to some criterion (e.g., $\mathbf{T}$ may contain principal components in the PCA framework, redundancy variates in RRR, or canonical variates in CCA), where the vectors $\mathbf{b}_j$ contain unknown coefficients. Now, let us consider the linear regression model in the d linear composites $\mathbf{t}_j$ $(j = 1, \ldots, d;\ 1 \leq d \leq p)$:

$$\mathbf{Y} = \mathbf{TA} + \mathbf{E}, \tag{8}$$

where $\mathbf{A}$ is a $(d \times q)$ matrix of regression coefficients and $\mathbf{E}$ represents the residuals. The linear composites $\mathbf{T}$ used in PCR are the least-squares solutions for the model

$$\mathbf{X} = \mathbf{TP} + \mathbf{G}, \tag{9}$$

where $\mathbf{P}$ is the matrix of loadings and $\mathbf{G}$ is the matrix of residuals. Since the first d ordinary principal components of $\mathbf{X}$ minimize SSQ($\mathbf{X} - \mathbf{TP}$), PCR regresses the responses $\mathbf{Y}$ onto the first d principal components of the explanatory variables.

By means of MSCA-P, extracting d_b and d_w principal components in the between and the within structures from matrices $\mathbf{M}_x$ and $\mathbf{X}_c$, respectively (typically, $d_b = d_w$ for a suitable comparison of the contributions of variables in the reduced spaces of similar dimensionality) using Eq. (5), the multilevel principal-component regression (MLPCR) model in the i-th block provides

$$\mathbf{Y}_i = \mathbf{1}_{K_i}\, \mathbf{t}'_{ib}\mathbf{A}_b + \mathbf{T}_{iw}\mathbf{A}_w + \mathbf{E}_i \tag{10}$$

subject to $\Sigma_i K_i\, \mathbf{t}_{ib} = \mathbf{0}$ and $\mathbf{1}'_{K_i}\, \mathbf{T}_{iw} = \mathbf{0}'$ (MSCA-P constraints).

Applying the decomposition (3) to matrices $\mathbf{Y}_i$, the loss function related to Eq. (10) is expressed as

$$f_{\mathrm{MLPCR}}(\mathbf{A}_b, \mathbf{A}_w) = \mathrm{Tr}[\Sigma_i(\mathbf{Y}_{c,i} + \mathbf{1}_{Ki}\mathbf{m}'_{iy} - \mathbf{1}_{Ki}\mathbf{t}'_{ib}\mathbf{A}_b - \mathbf{T}_{iw}\mathbf{A}_w)]. \tag{11}$$

In the “Appendix”, we show that, under the constraint $\mathbf{1}'_{Ki}\mathbf{T}_{iw}\,\mathbf{A}_w = \mathbf{0}'$ (the scores of least-squares projections $\mathbf{T}_{iw}\mathbf{A}_w$ in the within space are individual mean-centered for each dependent variable), the minimization of f_{MLPCR} is equivalent to the minimization of the sum of two separate PCR loss functions corresponding to the between and within structures. Specifically, minimizing (11) is equivalent to minimizing

$$\Sigma_i \mathrm{SSQ}\,(\mathbf{Y}_{c,i} - \mathbf{T}_{iw}\mathbf{A}_w) + \mathrm{SSQ}(\mathbf{W}[\mathbf{M}_y - \mathbf{T}_b\mathbf{A}_b]), \tag{12}$$

where $\mathbf{T}_b$ is an $(I \text{ x } d_b)$ matrix, of which the rows contain the between-individual scores $\mathbf{t}'_{ib}$ and $\mathbf{A}_b$, which is the matrix of regression coefficients. Because these two functions deal with different parameter sets, the between and the within parts can be minimized separately. Further, the PCR linear-composite scores related to the between and within structures become mutually orthogonal.

Note that Eq. (12) expresses two separate principal-component regression models: in both structures, regression parameters can be separately estimated in the within structure by regressing $\mathbf{Y}_c$ on $\mathbf{T}_w$ (obtained by concatenating the I blocks $\mathbf{T}_{iw}$) and in the between structure by regressing $\mathbf{WM}_y$ on $\mathbf{WT}_b$.

5 Multilevel reduced-rank regression (MLRRR)

The MLPCR model described above does not take into account the correlation existing between $\mathbf{Y}$-variables, since to extract LCs it does not use any of the information available in the relationships among the variables in the response space. Other DRM methods involve using a subspace of the $\mathbf{X}$-space similar to that used in PCR, but we can extend our choice of $\mathbf{X}$-subspace to methods that allow for association with $\mathbf{Y}$-space. Now, let us consider the linear regression model in d linear composites $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_d)$ within the RRR framework, which is

$$\mathbf{Y} = \mathbf{TA} + \mathbf{E} = \mathbf{X}\Gamma_{[d]} + \mathbf{E}, \tag{13}$$

where $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_d) = \mathbf{X}(\mathbf{b}_1, \ldots, \mathbf{b}_d) = \mathbf{XB}$ and $\boldsymbol{\Gamma}_{[d]} = \mathbf{BA}$ is a p x q matrix of rank $d(1 \le d \le \mathrm{p})$, meaning that the maximum number of LCs in RRR is the minimum between rank ($\mathbf{X}$) and rank ($\mathbf{Y}$).This is known as the reduced-rank regression (RRR) model, in which the corresponding loss function is the residual sum of squares, which is SSQ($\mathbf{Y} - \mathbf{TA}$). In RRR, the resulting LCs $\mathbf{t}_j$ (redundancy variates) are the ordered principal components extracted by $\hat{\mathbf{Y}}'\,\hat{\mathbf{Y}}$, and thus they are contained in the subspace spanned by the least-squares solutions $\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\Gamma}^*_{[d]}$, where $\boldsymbol{\Gamma}^*$is the estimation of $\boldsymbol{\Gamma}$ (Izenman 1975; Merola and Abraham 2001).

The optimal choice of rank d will usually be unknown and is considered to be a meta-parameter of the method. One of the weak points of RRR is its instability when the predictors are strongly correlated or the number of objects is smaller than the number of predictors. Under these circumstances, and more generally when $n < p$, the least-squares criterion for estimating $\boldsymbol{\Gamma}_{[d]}$ cannot be applied because the $p \times p$ covariance matrix $\mathbf{X}'\mathbf{X}$ (which has a rank of at most n) is singular and $\boldsymbol{\Gamma}_{[d]}$ is said to be under-determined. Under these circumstances, the use of a generalized inverse (i.e., Moore–Penrose) of $\mathbf{X}$ is suggested (Davies and Tso 1982).

The same problem plagues CCR since in CCA, when $n < p$ or $n < q$, the covariance matrix of $\mathbf{X}$ with itself and $\mathbf{Y}$ with itself ($\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$, respectively) are ill-conditioned. The condition placed on the data to guarantee that $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ will be invertible is $n \ge p + q + 1$. Instead, since PCA works in situations where there are many more variables than observations ($n < p$), PCR is not prone to the problem of under-determined regression, unless a very extreme situation occurs—i.e., when the number of observations is smaller than the number of retained principal components ($n < d$). As for the multilevel version of RRR (Eq. 13), after having decomposed matrix $\mathbf{X}$ as in Eq. (3) we specified $\boldsymbol{\Gamma}$(dropping the index d for simplicity) parameters (as in Eq. 5) in the between ($\boldsymbol{\Gamma}_b$) and the within ($\boldsymbol{\Gamma}_w$) structures, which leads to the associated loss function

$$f_{\mathrm{MLRRR}}(\boldsymbol{\Gamma}_{\mathrm{b}}, \boldsymbol{\Gamma}_{\mathrm{w}}) = \mathrm{Tr}[\Sigma_{\mathrm{i}}(\mathbf{Y}_{\mathrm{c,i}} + \mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{iy}} - \mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{ix}}\boldsymbol{\Gamma}_{\mathrm{b}} - \mathbf{X}_{\mathrm{c,i}}\boldsymbol{\Gamma}_{\mathrm{w}})]. \tag{14}$$

In this case, if we return to the MLPCR case and proceed with similar arguments, we obtain the result that the loss function (14) is equivalent to minimizing the sum of the following two functions:

$$\Sigma_{\mathrm{i}}\mathrm{SSQ}\,(\mathbf{Y}_{\mathrm{c,i}} - \mathbf{X}_{\mathrm{c,i}}\boldsymbol{\Gamma}_{\mathrm{w}}) + \mathrm{SSQ}(\mathbf{W}[\mathbf{M}_{\mathrm{y}} - \mathbf{M}_{\mathrm{x}}\boldsymbol{\Gamma}_{\mathrm{b}}]) \tag{15}$$

under the constraint $\mathbf{1}'_{K_i}\mathbf{X}_{\mathrm{c,i}}\boldsymbol{\Gamma}_w = \mathbf{0}'$ (i.e., least-squares projections of LC scores in the within structure are individual mean centered for each dependent variable). Note that this constraint is equivalent to restricting the LC scores in the within structure $\mathbf{T}_{\mathrm{c,i}}$ to being individual mean centered ($\mathbf{1}_{\mathrm{K_i}}{}'\mathbf{T}_{\mathrm{c,i}} = \mathbf{0}'$). Hence, the within and between parts can be minimized separately.

With regard to projection matrices, let $\hat{\mathbf{Y}}_{\mathrm{m}} = \mathbf{1}_{\mathrm{Ki}}\mathbf{m}'_{\mathrm{ix}}\boldsymbol{\Gamma}^*_b$ be the matrix of fitted responses on the matrix of predictors in the between structure, and let $\hat{\mathbf{Y}}_{\mathrm{c}} = \mathbf{X}_{\mathrm{c}}\boldsymbol{\Gamma}^*_{\mathrm{w}}$ be the matrix of fitted responses on the matrix of predictors in the within structure. Since, they are combinations of mutually orthogonal matrices, the resulting RRR linear composites in the between and within structures are mutually orthogonal and are the eigenvectors corresponding to the largest eigenvalues of the matrices $\hat{\mathbf{Y}}_{\mathrm{m}}\hat{\mathbf{Y}}'_{\mathrm{m}}$ and $\hat{\mathbf{Y}}_{\mathrm{c}}\hat{\mathbf{Y}}'_{\mathrm{c}}$, respectively.

6 Multilevel canonical-correlation regression

Another useful multivariate DRM that uses the information available in the relationships among the variables in the response space is canonical-correlation regression (CCR; Anderson 1951). Canonical variates in the **X** space are based on the technique of CCA. Here, the original objective is to find vectors **XB** and **YG** (canonical variates) in the column spaces of **X** and **Y**, respectively, that are most highly correlated with each other. The first d canonical variates **XB** (= **T**), orthogonal in the column space of **X**, are used as the predictor space for regression on dependent block **Y**, resulting in CCR.

An appealing aspect of CCA is its intuitive geometric interpretation (Anderson 2003): maximizing the correlation (i.e., the cosine) between the canonical variates can be interpreted as minimizing the angle between **XB** and **YG**, which in turn is equivalent to minimizing the distance for canonical variates of equal length (SSQ [$\mathbf{YG} - \mathbf{XB}$]). This equivalent formulation has been extensively used in the context of nonlinear canonical-correlation analysis (see van der Burg et al. 1994).

Since CCA is the primary step of CCR, we illustrate first that CCA can be solved in a multilevel perspective, resulting in the multilevel canonical-correlation analysis (MLCCA). In addition, the CCR model is obtained as shown in Sect. 4 because it is formally equivalent to the MLPCR model, with the only difference being that in CCR the canonical variates instead of the principal components (as in MLPCR) are used as the predictor space for regression. Specifically, let $\mathbf{X}_{\mathrm{i}}$ and $\mathbf{Y}_{\mathrm{i}}$ be data matrices containing $\mathrm{K_i}$ measurements on p and q variables, respectively, for subject i (with **X**- and **Y**-centered matrices on the overall mean across all individuals), and let $\mathbf{G}_{\mathrm{i}}$ and

$\mathbf{B}_\mathrm{i}$ be the matrices containing (canonical) coefficients that define canonical variates for block i.

Following the decompositions of matrices and weights in both structures—Eqs. (3) and (5)—the minimization of the loss function Σ_i SSQ $[(\mathbf{Y}_\mathrm{i}\,\mathbf{G}_\mathrm{i} - \mathbf{X}_\mathrm{i}\,\mathbf{B}_\mathrm{i})]$ can be divided into the separate minimizations of the within terms ($\mathbf{G}_\mathrm{w}$ and $\mathbf{B}_\mathrm{w}$ for $\mathbf{Y}_{\mathrm{c,i}}$ and $\mathbf{X}_{\mathrm{c,i}}$, respectively) and the between terms ($\mathbf{G}_\mathrm{b}$ and $\mathbf{B}_\mathrm{b}$ for $\mathbf{M}_\mathrm{y}$ and $\mathbf{M}_\mathrm{x}$, respectively) as follows:

$$f_{\mathrm{MLCCA}}(\mathbf{G}_\mathrm{w}, \mathbf{B}_\mathrm{w}, \mathbf{G}_\mathrm{b}, \mathbf{B}_\mathrm{b}) = \Sigma_\mathrm{i}\mathrm{SSQ}[\mathbf{Y}_{\mathrm{c,i}}\mathbf{G}_\mathrm{w} - \mathbf{X}_{\mathrm{c,i}}\mathbf{B}_\mathrm{w}] + \mathrm{SSQ}[\mathbf{W}(\mathbf{M}_\mathrm{y}\mathbf{G}_\mathrm{b} - \mathbf{M}_\mathrm{x}\mathbf{B}_\mathrm{b})] \tag{16}$$

under the constraint $\mathbf{1}'_{\mathrm{Ki}}\mathbf{X}_{\mathrm{c,i}}\mathbf{B}_\mathrm{w} = \mathbf{0}'$—i.e., that scores of the $\mathbf{X}$-canonical variates in the within structure are individual mean centered for each column. Hence, in the between structure, CCA can be obtained by applying the CCA model to matrices $\mathbf{M}_\mathrm{x}$ and $\mathbf{M}_\mathrm{y}$, which contain, as rows, the mean vectors $\mathbf{m}_{\mathrm{ix}}$ and $\mathbf{m}_{\mathrm{iy}}$ for the i-th subject. The within-subject CCA can be estimated by applying CCA to matrices $\mathbf{X}_\mathrm{c}$ and $\mathbf{Y}_\mathrm{c}$, which contain the concatenated, centered matrices $\mathbf{X}_{\mathrm{c,i}}$ and $\mathbf{Y}_{\mathrm{c,i}}$, respectively.

Thus, not only are the extracted canonical variates orthogonal within their space of reference (each canonical variate is orthogonal to all other variates derived from the same set of data) but, after obtaining group mean-centered canonical variates in the within structure of the predictor space, all within canonical variates are orthogonal to the between canonical variates. Therefore, we found that MSCA principles can be used to generalize, in a multilevel perspective, other multivariate methods such as CCA, in which projections are guided into directions by two sets of symmetrical variables.

In the final step, the CCR model is extended in the multilevel framework using the findings of Sect. 4 (e.g., to estimate regression parameters $\mathbf{A}_\mathrm{w}$, $\mathbf{A}_\mathrm{b}$), in which the coefficients defining the LCs of the predictors ($\mathbf{T}_{\mathrm{iw}}$ for the within structure and $\mathbf{t}_{\mathrm{ib}}$ for the between structure, as in Eq. 10) are given by the first d coefficients of the canonical variates in the $\mathbf{X}$ space ($\mathbf{B}_\mathrm{w}$ and $\mathbf{B}_\mathrm{b}$ of Eq. 16 for the within and between structures, respectively).

7 Rotational freedom

As known, both submodels (the between-loading and within-loading matrices) of MSCA-P have rotational freedom, meaning that each of those matrices can be (orthogonally and obliquely) rotated using a non-singular matrix, provided that this rotation is compensated for in the accompanying composite scores matrix, without altering the model's fit. The same holds in classic PCA for loading matrices that define principal components. The freedom of transformation can be used in various ways (de Jong and Kiers 1992). As a first option, it is required that principal components are orthogonal and standardized. In MSCA-P, Timmerman (2006) suggested fixing at 1 the variances of the between- and within-component scores for all individuals. A second option is to transform the estimated loading matrices. Because the interpretation of the solution is normally based on the loadings, a typical strategy is the rotation of the estimated loading matrices. Usually, this will be a rotation to a simple structure—for example, using varimax rotation (Kaiser 1958).

The multilevel extension of the principal-component regression (MLPCR) model obviously has the same rotational freedom since MLPCR admits as predictors the composite scores of MSCA-P. Hence, in the MLPCR the rotational freedom of composite loadings may be used in both perspectives. As far as CCA is concerned, theoretically, orthogonal rotation of canonical variates can be considered as an aid to increase the interpretability of canonical results since rotation does not change the sums of the squared canonical correlation coefficients, although it does lead to a simpler structure. However, any rotation will destroy one key feature of the CCA analysis—i.e., that successive pairs of composites combinations have maximum correlations. Moreover, rotation introduces correlations among succeeding canonical variates. Instead, since the canonical correlations do not depend on the scaling of canonical coefficients, canonical variates in each block are typically restricted to being uncorrelated, with unit variances. Hence, in MLCCR we use these constraints for $\mathbf{X}$ and $\mathbf{Y}$ canonical variates in both structures.

In the multilevel DRMs, there is even then a remaining indeterminacy involving the regression matrices $\mathbf{A}_b$ and $\mathbf{A}_w$. Specifically, each DRM can be viewed as a regression problem with the aim of estimating the (deficient rank d) matrix of regression $\boldsymbol{\Gamma}$ between $\mathbf{Y}$ and $\mathbf{X}$, which was modified in a new regression problem that specifies, as a new predictor space for the response space $\mathbf{Y}$, a series of d linear composites ($\mathbf{T} = \mathbf{XB}$), where $\mathbf{A}$ denotes the regression matrix of $\mathbf{Y}$ onto the $\mathbf{T}$ space. Hence, a simultaneous rotation of both parameter $\mathbf{B}$ and $\mathbf{A}$ by an arbitrary orthogonal matrix is possible without changing the coefficient matrix $\boldsymbol{\Gamma} = \mathbf{BA}$ and violating the specific constraints, such as orthonormal redundancy variates, canonical variates, and principal components. However, MLPCR and MLCCR are two-step strategies in which regression matrices $\mathbf{A}_b$ and $\mathbf{A}_w$ are estimated once the orthonormal linear composites $\mathbf{T}_b$ and $\mathbf{T}_w$ are obtained in a previous step by MSCA-P and MLCCA, respectively. Thus, $\mathbf{A}_b$ and $\mathbf{A}_w$ have unique solutions in both structures, since in the regression step the predictors are orthogonal components. This is not the case for multilevel reduced-rank regression (MLRRR), since reduced-rank regression simultaneously estimates both the components' loadings and the regression matrices.

Apart from situations where precise restrictions are needed—which are typically resolved by generalized procrustes analysis (see Gower 1975)—in more general cases this non-uniqueness is not explicitly resolved in the literature. However, this limitation, which may be mitigated by enforcing specific constraints, is less problematic if estimation of the coefficients is required for predictive rather than interpretive purposes. Hence, in MLRRR redundancy variates are themselves restricted to being orthogonal with unit variances in both structures.

Rotational freedom has large implications for estimating empirical standard errors within bootstrap strategies (Timmerman et al. 2009). In the Multilevel DRMs, even if the parameters of interest are the between and within regression matrices ($\mathbf{A}_b$ and $\mathbf{A}_w$, respectively), which have unique solutions in each rotated bootstrap sample, rotational freedom must be significantly mitigated in order to make the loading matrices of bootstrap solutions comparable. In fact, non-uniqueness, after simple structure rotation (e.g., varimax) of the rotated bootstrap-loading matrices, also depends both on the sign of the loadings and on the ordering of the rotated components, which are arbitrary. Procrustes procedures (which also take care of reflection) are typically used to optimally rotate the bootstrap solutions toward a target, normally the composite loadings found in the entire sample. This makes comparable the bootstrap loading matrices and thus the regression matrices estimated by DRMs.

8 Application

The purpose of this section is to show the potential of the two multilevel analysis techniques—MLPCR and MLRRR—utilizing an example from psychiatric research. Both methods are applied within the longitudinal research study "HoNOS 4," which was conducted in mental health departments in the Lombardy Region of Italy during 2009, and which was aimed at measuring the effectiveness and the quality of care of Lombard psychiatric health services (Erlicher and Lora 2002). The main focus of the present application is to evaluate the relationship between the severity of mental disease and the intensity of psychiatric care (intensity of treatment) at the patient level. The study recruited 1,624 patients who were in contact with Lombard mental health departments (MHDs). Mental disease ratings were completed at baseline (November 2008) and longitudinally in three periods during the year 2009: March (1st assessment), July (2nd assessment), and November (3rd assessment). Overall, 1,361 patients received at least one measurement during the study period (2009), and 741 patients were evaluated three times (three assessments). Only the latter sample of patients was included in the analysis.

The severity of mental disease was measured by the Health of the Nation Outcome Scale (HoNOS), a mental-health-status outcome scale developed by the U.K. Royal College of Psychiatrists' Research Unit (Wing et al. 1998) and later validated in Italy (Lora et al. 2001; Lovaglio and Monzani 2012).

HoNOS is a clinician-completed instrument consisting of 12 items to be scored on five-point scales, from 0 (no problem) to 4 (severe/very severe problem), covering clinical and psychosocial problems (Table 1).

HoNOS describes patients' current severity, not only in clinical and behavioral areas but also in the psychosocial aspect as well. Specifically, the 12 items are intended to cover four areas of mental health (subscales): Behaviour (H13), Impairment (H45), Symptoms (H68), and Social Functioning (H912). Intensity of care was measured, for each psychiatric patient, as the total number of contacts with MHDs, during each assessment. Patients have contacts with MHDs for different types of treatments.

To this end, in Italy the model of psychiatric care is no longer centered on mental hospitals (hospitalizations occur only when particular events, such as violence or aggression, arise) but rather on a community-integrated network of mental health facilities (community centers, day-care facilities, psychiatric wards in general hospitals) located in the catchment area and coordinated by the MHD. Specifically, intensity of care indicators were measured for each four-month period as the total number of visits for clinical treatment (CLIN_visits, meaning treatments in community centers by psychiatrists or psychologists), within-the-community treatment (COMM_visits, meaning, in addition to clinical treatment, interventions by other professionals such as

Table 1 HoNOS items, subscales, and scores

HoNOS items	Subscales	
H1 overactive, aggressive, disruptive		
H2 non-accidental self-injury	H1–H3 behaviour subscale	
H3 problem drinking or drug taking	H4–H5 impairment subscale	
H4 cognitive problems	H6–H8 symptom subscale	
H5 physical illness or disability problems	H9–H12 social functioning subscale	
H6 problems with hallucinations and delusions		
H7 problems with depressed mood	Score	Description
H8 other mental and behavioural problems	0	No problem
H9 problems with relationships	1	Minor problem requiring no action
H10 problems with activities of daily living	2	Mild problem but definitely present
H11 problems with living conditions	3	Moderately severe problem
H12 problems with occupation and activities	4	Severe to very severe problem

nurses, social workers, and rehabilitation therapists), and within-day-care treatment (DayC_visits, meaning treatments in day-care centers). Further, two additional indicators were measured for each four-month period: the number of hospitalizations in a psychiatric ward (Charge_Hosp), and the length of stay (Length_Hosp). The data set described was analyzed at two levels by the examination of the relationship between intensity of care and mental disease, separating between-patient (static structure) and within-patient variations (dynamic structure).

To apply MLPCR and MLRRR, longitudinal data (three time points) were organized into two blocks: the dependent block (Y-block) that collected scores on HoNOS subscales (Behaviour, Impairment, Symptoms, and Social Functioning) and the predictor block (X-block), which included intensity-of-care measures. The latter block also included six other (time-invariant) patient characteristics: the HoNOS subscales at baseline (Behaviour0, Impairment0, Symptoms0, and Social Functioning0), the patient's age, and the duration of the contact with an MHD in years, both measured at baseline.

Since such time-invariant predictors only referred to the between structure, they did not contribute to the within analysis. Thus, the results of the within analyses were obtained net of contribution of age, contact duration, and the HoNOS subscales at baseline, which are typical confounding factors in psychiatric research. Moreover, in the MLPCR analysis, although the focus of application is essentially exploratory, we considered some inferential aspects, by implementing non-parametric bootstrap strategies for estimating empirical standard errors on the individual parameters, as suggested by Timmerman et al. (2009). Given that in the application patients were considered random, whereas measurement occasions were fixed (by the study design), we considered, as a resampling scheme, the so called "multi-observation" case, which implies resampling only the level-2 units (patients), thereby keeping all level-1 units (temporal measurements) associated with the selected level-2 units. Finally, in the MLPCR

Table 2 Percentages of the within- and between-individual variations for each block and time-variant variable

Y-block	Variance within (%)	Variance between (%)	X-block	Variance within (%)	Variance between (%)
Behaviour	58.7	41.3	CLIN_visits	58.6	41.4
Impairment	49.0	51.0	COMM_visits	60.9	39.1
Symptom	56.3	43.7	DayC_visits	69.9	30.1
Social functioning	52.6	47.4	Charge_Hosp	15.3	84.7
			Length_Hosp	17.8	82.2
Overall Y-block	54.1	45.9	Overall X-block	40.2	59.8

analysis, orthogonal Procrustes analysis was used to rotate the bootstrap solutions toward the (between and within) composite loadings found by MSCA-P in the entire sample.

8.1 Results

Looking at severity levels, the HoNOS scores decreased as the level of illness severity anticipated within treatments decreased; at the last assessment, the highest mean HoNOS score and the coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, was obtained for acute wards (mean = 12.5, CV = 10.5) and day-care treatment (mean = 11.3, CV = 11.4), followed by community treatment (mean = 9.4, CV = 7.5) and clinical treatment (mean = 7.4, CV = 8.5). As the first step of the analysis, we determined (using Eq. 3) the magnitudes of the within- and between-individual variations for each block and time-variant variable. Results are presented in Table 2.

Table 2 illustrates that the variations in the data for intensity of care (X-block) and mental severity (Y-block) were nearly balanced between the two types of variation, although in different directions. In the dependent block (**Y**), the variation in severity of illness revealed higher intra-individual or dynamic variation (54 %); this was particularly marked for the Behavior subscale (nearly 60 % of total variability was within individuals). In the predictor block (**X**), 60 % of the intensity-of-care variation was between individuals, essentially due to the marked between-patient differences regarding the number of hospitalizations and the length of stay (for both variables, more than 80 % of total variability was between individuals). Hence, both observed blocks varied between structures: intensity-of-care indicators (**X**) vary greatly between patients and to a lesser extent over time, whereas mental-severity indicators (**Y**) vary in the opposite direction. This justified further multilevel DRM analyses.

In the MLPCR and MLRRR analysis, we extracted linear composites from predictors in the between and within structures. For both DRMs, using a scree graph, three composites were chosen for the between-individual (**B0–B1–B2**) and within-individual (**W1–W2–W3**) models. Such a labeling strategy facilitated the comparison of LCs between both structures. In fact, as we will show later, because the

Table 3 Variance explained by the between-subject (B0–B2) and within-subject (W1–W3) LCs of predictors in the multilevel principal-component regression (MLPCR) and multilevel reduced-rank regression (MLRRR)

LC	%Expl.X	%Cum.X	%Expl.Y	%Cum.Y	LC	%Expl.X	%Cum.X	%Expl.Y	%Cum.Y
MLPCR between structure					*MLPCR within structure*				
B0	29.7	29.7	21.6	21.6	W1	29.5	29.5	9.0	9.0
B1	19.4	49.1	12.5	34.0	W2	20.8	50.3	12.7	21.7
B2	14.7	63.8	6.2	40.2	W3	13.1	63.4	11.2	32.9
MLRRR between structure					*MLRRR within structure*				
B0	22.5	22.5	30.0	30.0	W1	18.5	18.5	16.8	16.8
B1	12.6	35.1	13.9	43.9	W2	9.9	28.4	12.2	29.0
B2	9.8	45.0	9.8	53.7	W3	14.5	42.9	9.0	38.0

first LC in the between-structure (**B0**) did not contribute to the within analysis (involving time-invariant predictors), it has no counterpart in the within structure. Table 3 illustrates the results of the MLPCR in principal-axes position (upper) and MLRRR (lower). Specifically, for the between-subject and within-subject parts of the models, the percentages of predictors' variance (*%Expl.X*) and responses' variance (*%Expl.Y*, plus their cumulative percentages *%Cum.X* and *%Cum.Y*, respectively), which were explained by the retained LCs, were given. Note that in the MLPCR (upper part of Table 3), the *%Expl.X* refers to the variances of the **X** block explained by the between-subject (B0–B2) and within-subject LCs (**W1–W3**)—i.e., by the MSCA-P components.

Analyzing first MLPCR, the first three (X-block) LCs accounted for nearly two-thirds of the variance of their indicators (**X** variation) for both structures (64 % between individuals and 63 % within individuals). With regard to the responses' variation (*%Cum.Y*), the LCs of the predictors in the between structure (**B0–B2**) explained 40 % of the between variation in mental severity (the first, **B0,** explains 21 %), whereas the three within LCs (**W1–W3**) explained 33 % of the dynamic variation in mental severity. This means that, after we checked for the effects of time-invariant predictors (age, contact duration, and HoNOS subscales at baseline), the components capturing differences in intensity of care between patients explained a significant portion (40 %) of the between-patients variability in mental severity. In contrast, LCs capturing differences of care intensity over time explained a smaller portion (33 %) of the longitudinal dynamic variation of mental severity.

As expected, in the MLRRR model (lower part of Table 3) the first three between and within LCs of predictors explained a smaller portion of **X** and a larger portion of **Y** than did the corresponding MLPCR components. In particular, the first three RRR composites of the **X** space explained more than 40 % of the **X** variation in both structures (45 % between and 43 % within variation), whereas they do a better job of explaining response variability, especially in the between structure. The percentage of between-individuals mental-severity variation explained by the first three between composites (**B0–B2**) is 54 % (the first, **B0**, explains 30 %), whereas only 38 % is explained by the within composites (**W1–W3**; the first component explains 17 %).

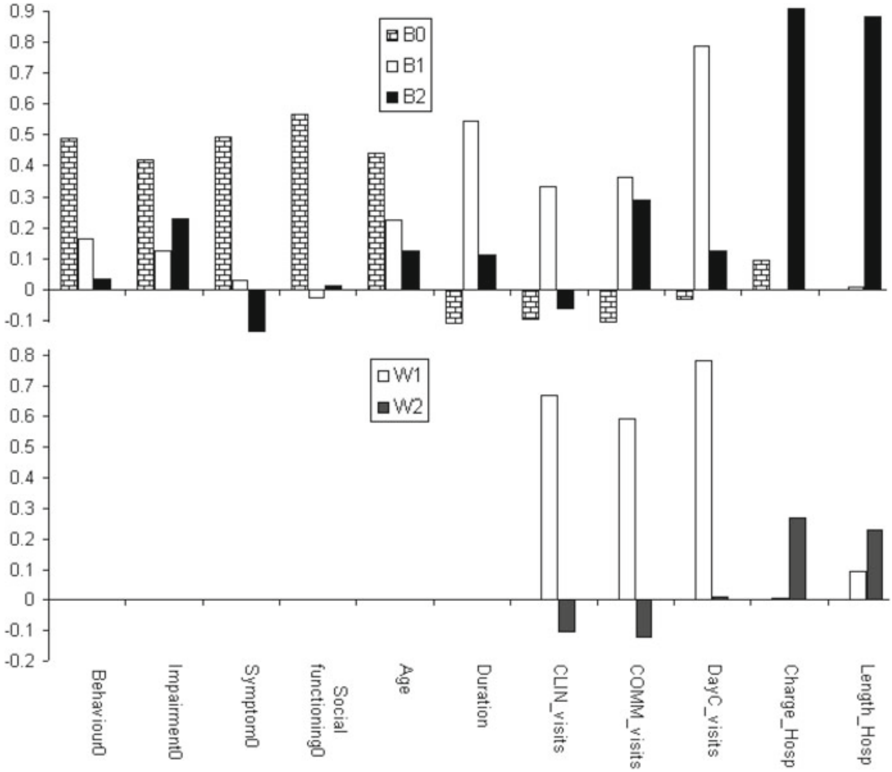

Fig. 1 Between- and within-subject loadings in the MLPCR model

Hence, also for MLRRR, the first three composites explain a larger portion of intra-individual **Y** variance (54 %) than of the inter-individual **Y** variability (38 %), meaning that differences in intensity of care between patients explain more than half of the variation related to differences in mental severity between patients, whereas longitudinal dynamics of care intensity poorly affect dynamic variations of mental severity. Further, for both DRM models, the percentages of **X** variance explained by retained LCs are very similar among the between and within structures (45 and 43 %, respectively, for MLRRR; 64 and 63 %, respectively, for MLPCR). Analyzing the MLPCR model in depth, the interpretation of the **X** composites is, as usual, based on the loadings or weights. Since loadings are more influenced by the **X** variance than by weights (most are influenced by the **X**–**Y** covariances), the former are more appropriate in an MLPCR framework.

In Fig. 1, the loadings for the first three LCs (even in the principal-axes position) in the between structure (**B0**–**B1**–**B2**) and for the first two LCs in the within structure (**W1**–**W2**) are depicted. They appear to be easily interpretable composites. The first between composite (**B0**) can be interpreted as *baseline mental severity*, the second (**B1**) as *intensity of community care*, and the third (**B2**) as *intensity of hospital care*. The within composites **W1** and **W2** reflect the same interpretation as **B1** and **B2**,

Table 4 Robust t-statistics (bootstrapped standard errors) for A_b (between) and A_w (within) in the multilevel principal-component regression model

Source	Predictors	Dependent variables			
		Behaviour	Impairment	Symptom	Social functioning
Between	B0	7.41***	6.68***	11.79***	10.31***
	B1	−4.01***	−1.87*	−4.20***	−1.96**
	B2	9.24***	0.69	6.01***	1.05
Within	W1	3.12***	4.60***	5.41***	4.77***
	W2	4.37***	−1.12	0.37	−0.29

*** Significant at the 0.01 level, ** Significant at the 0.05 level, * Significant at the 0.1 level

although with some differences. To this end, the loadings of the between- and within-individual models can be compared in order to identify compounds that are important for describing the variations between and within the individuals.

The number of day-care visits is an important predictor that adequately describes the variations between and within individuals (loadings > 0.70) in both structures. On the other hand, both indicators related to acute wards had high loading only in the between model (loadings < 0.30 in the within structure), meaning that the number of hospitalizations highly varied between the patients but was relatively constant over time for each patient. In contrast, clinical and community intensity of care (CLIN and COMM visits) encompassed high loadings only for the within-individual model (loadings ≥ 0.60), meaning that the number of outpatient visits varied in time, although the average values were similar among patients. Hence, monitoring clinical, community, and day-care visits over time appears to be the most important factor and is predictive of longitudinal variation in mental severity, after removal of severity at baseline, patients' age, and contact duration.

With regard to inferential aspects, we focus on the regression parameters in the MLPCR ($\mathbf{A}_b$ and $\mathbf{A}_w$ of Eq. 10), which indicate the relationships between the retained (exogenous) composites and each dependent variable (Behavior, Impairment, Symptom, and Social Functioning) for each structure. In this perspective, standard errors of regression parameters were computed by 500 bootstrap replications, after rotation (using orthogonal Procrustes rotation) of the loadings of the bootstrap replications toward the MLPCR loading matrix found in the entire sample (see Fig. 1). Specifically, Table 4 presents the robust t-statistic (ratio of the estimated parameter to the bootstrapped standard error) and the related significance of the ($\mathbf{A}_b$ and $\mathbf{A}_w$) parameters.

In the between-individual model, apart from the strong effect of baseline mental severity (**B0**) on all subscales, only for the Behaviour and Symptom subscales are the differences in intensity of community care (**B1**) and of hospital care (**B2**) significantly associated with the variation of between-patient mental severity. Moreover, the last two components affect both HoNOS subscales in opposite directions: increasing levels of community care are associated with decreasing mental disease (to the same extent for both subscales), whereas the patients with frequent and increasing lengths of stay in psychiatric wards are the most severe, especially for the Behavior subscale.

The within part of Table 4 (including only **W1** and **W2**) describes how components affect the dynamic variation of mental severity over time—controlled for mental severity at baseline—and of other time-invariant predictors. The dynamic variation of community care (**W1**), is positively and significantly associated with all subscales, meaning that longitudinal increases in community care result in a worsening of the mental state during the time period, whereas increases in hospital care over time (**W2**) resulted in a decline in mental disease only in the Behavior subscale.

If we look at Table 4, given that components have the same interpretations in both structures, we can compare the signs and significance of components in the between and within structures. In particular, the components that represent *intensity of community care* (**W1** and **B1**) affect the dependent variables in opposite directions, meaning that the averages (over all individuals) of the HoNOS subscales decrease as the averages of community care increase. In contrast, in a longitudinal perspective mental severity becomes worse over time for patients with larger dynamic variation in intensity of community care.

9 Discussion

In the present paper, following the principles of MSCA and taking into account both the between and within data structures, we generalize, in a multilevel perspective, the best-known DRMs, in which a set of response variables is predicted by a limited number of components that have been extracted by the set of observed predictors. Starting with (column-wise mean) centered data matrices **X** and **Y**, we first decomposed data matrices in two orthogonal blocks to obtain the between and within sources of variability; second, we fitted standard DRMs separately in the between and within structures. Specifically, under specific constraints, we obtained two separate models in which the dependent variables in each structure are regressed on LCs of predictors in the corresponding structure. For MLPCR and MLCCR, LC scores (principal components and canonical variates, respectively) were obtained by separate MSCA-P and MLCCA models in the between and within structures. For MLRRR, LC scores were redundancy variates, which were obtained by two separate RRR models in each structure. Hence, response variables were used within a multilevel framework to guide the projections in meaningful directions.

The main result of the paper is that the minimization of the loss functions related to MLPCR, MLCCR, and MLRRR is equivalent, under suitable constraints, to the separate minimization of the sum of two separate loss functions corresponding to the between and the within models. Therefore, all within composites are orthogonal in the same dimension and for all dimensions to the between composites. In this way, although individuals have different composite scores, the constraint of equivalent loading matrices for all individuals leads to improvement of the interpretation of composites since the same constituting definition is used as that of the composites from the variables. Hence composites measure the same concept for all individuals.

For the proposed methodologies, the models provide information on both the explained **Y** variation and the **X** variation (that for MLPCR is the same as that which can be obtained from classical MSCA-P). The regression parameters allow us to disentangle the effects of covariates on the between model (the effect of average levels of a predictor on the average levels of responses) and on the within model (the effect of the dynamic variation of a predictor on the dynamic variations of responses). The different interpretations of predictors on the two structures represent a very attractive tool in hierarchical contexts (Timmerman M., personal communication, March 19, 2010).

In the empirical application, a number of interpretable composites are obtained, which increase the interpretation of the complex and multifaceted relationships existing between mental severity and intensity of care in both the static and dynamic perspectives. The presented DRM models start with data matrices column-wise centered across all individuals. With respect to variances, it may be useful to consider an appropriate scaling for each variable prior to the analysis. In the MLPCR, the variance of each variable over all measurement occasions and individuals may be set to 1 to equalize the influence of each variable on the final solution. In the CCA, although canonical correlations are invariant under linear transformation of the data, often a canonical analysis with standardized variables is used for better interpretation of canonical weights and canonical variates.

With regard to RRR, the **X** variables may be standardized. The amount of explained variation, as well as the fitted values of **Y**, remains unchanged by centering or standardization of the variables in **X**. This is not a necessary condition for a valid RRR analysis, but removing the scale effects of the explanatory variables turns the regression coefficients into standard regression coefficients, which are comparable to one another. Instead, RRR estimation under the least-squares criterion is sensitive to the choice of scaling of the response variables. Since the loss function (14) attaches equal weight to the lack of fit in each response, it may be wise to use an appropriate scaling prior to the analysis (i.e. standardizing each response variable to have unit variance) when the scales of the various response variables are not comparable.

In the literature, other methods to extend DRMs in a multilevel perspective have been proposed. The most general, called the multilevel generalized structured component analysis (MGSCA; Hwang et al. 2007), was proposed in the context of path analysis, which allows complex relationships to exist between endogenous and exogenous (latent and observed) variables. MGSCA extends generalized structured component analysis (Hwang and Takane 2004), allowing loadings (relating latent variables to observed variables) and path coefficients of latent variables—both specified as random effects—to vary across higher-level units (groups). However, MGSCA, which estimates very complex systems of structural relationships by an Alternating Least-Squares estimation algorithm, is too sophisticated for the main purpose of the present paper. In this end, the principal attractive features of DRMs are an unambiguous interpretation of linear composites (principal components, canonical variates, redundancy variates, etc.) and their estimation in a closed analytical form. For this reason, the implementation of the presented techniques does not require specialized software, so they can be applied without efforts.

In this perspective, the authors of MGSCA suggest as further research a more comprehensive investigation of the relationships between MGSCA and extant multivariate techniques such as "multilevel principal-component and multilevel canonical correlation" (Hwang et al. 2007).

Another explorative methodological tool, called the ANOVA simultaneous component analysis (ASCA; Smilde et al. 2005; Jansen et al. 2005), was recently proposed in the context of experimental design and has some similarities with the approaches presented in this paper. For example, similar to our proposal, the main goal of ASCA, which generalizes analysis of variance (ANOVA) for complex and highly structured multivariate data, is to find linear composites in order to approximate and separate the total variation of overall data in orthogonal and independent factors. Since ANOVA is capable of doing this, the authors of the ASCA approach approximate such ANOVA orthogonal factors (principal effects, interactions, and residual individual variations) by means of a series of linear composites (specifically, principal components), using the same parameterization that occurs in the SCA—the within structure of MSCA-P, with the loading matrices restricted to being equal for all individuals.

In ASCA, the (effects of) experimental factors are directly specified as parameters of the model (e.g., time), whereas in MSCA-P they constitute nesting levels to structure the data matrix. In this perspective, MSCA-P can be viewed as a special case of ASCA for nested design. Despite the advantages of ASCA, mainly for experimental design, it has two limitations in the context of a DRM framework: first, ASCA covers situations in which the predictor block is composed only of qualitative variables; second, the methodology to obtain linear composites in ASCA is limited to principal components (PCA framework).

To conclude, some methodological aspects remain in need of improvement. The first deals with missing data bias (particularly when the data are missing in a non-ignorable manner), which is a major problem in longitudinal studies in which attrition is inevitable over time. In this application, only patients having complete assessments were included in the analysis. Further studies are needed to evaluate the missing-data mechanism and to incorporate this source of uncertainty into the proposed approaches.

Second, the presented methodologies, which generalize principles of MSCA-P, do not make any assumptions about similarities between the time-dynamic variations of different individuals. This means that for each individual the correlation between the within-individual components can differ. Also, the variation described by each composite for each individual can be different. Future research hopefully focus on generalizing multilevel DRMs, which incorporate assumptions about the relationships between the within-individual variations of different individuals as the constrained versions of MSCA-P, such as the MSCA-PF2, IND, and ECP models (see Timmerman 2006).

Appendix: Proof of orthogonality of the between and within parts in the MLPCR model

It will be proved that, to fit the MLPCR model to observed data, the function f_{MLPCR} in (11), subject to $\mathbf{1}'_{\mathrm{K_i}}\, \mathbf{T}_{\mathrm{iw}}\mathbf{A}_{\mathrm{w}} = \mathbf{0}'$, can be separated in mutually orthogonal between and within parts, and thus that they can be resolved separately.

Minimizing (11) is equal to minimizing

$$f_{\text{MLPCR}}(\mathbf{A}_{\text{b}}, \mathbf{A}_{\text{w}}) = \text{Tr}[\Sigma_{\text{i}}(\mathbf{Y}_{\text{c,i}} + \mathbf{1}_{Ki}\mathbf{m}'_{\text{iy}} - \mathbf{1}_{Ki}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}} - \mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}})],$$

which is equal to

$$\begin{aligned}\text{Tr }\Sigma_{\text{i}}[&\mathbf{Y}'_{\text{c,i}}\ \mathbf{Y}_{\text{c,i}} + 2\mathbf{Y}'_{\text{c,i}}\ \mathbf{1}_{\text{Ki}}\ \mathbf{m}'_{\text{iy}}\ - 2\mathbf{Y}'_{\text{c,i}}\ \mathbf{1}_{\text{Ki}}\ \mathbf{t}'_{\text{ib}}\ \mathbf{A}_{\text{b}} - 2\mathbf{Y}'_{\text{c,i}}\ \mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}}\\ &+ \mathbf{m}_{\text{iy}}\mathbf{1}'_{\text{Ki}}\mathbf{1}_{\text{Ki}}\mathbf{m}'_{\text{iy}} - 2\mathbf{m}_{\text{iy}}\mathbf{1}'_{\text{Ki}}\mathbf{1}_{\text{Ki}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}} - 2\mathbf{m}_{\text{iy}}\mathbf{1}'_{\text{Ki}}\mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}}\\ &+ \mathbf{A}'_{\text{b}}\mathbf{t}_{\text{ib}}\mathbf{1}'_{\text{Ki}}\mathbf{1}_{\text{Ki}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}} + 2\mathbf{A}'_{\text{b}}\mathbf{t}_{\text{ib}}\mathbf{1}\text{K}'_{\text{i}}\mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}} + \mathbf{A}'_{\text{w}}\mathbf{T}'_{\text{iw}}\mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}}].\end{aligned} \tag{A1}$$

By noting that $\mathbf{Y}_{\text{c,i}}$ is orthogonal to $1_{\text{K}_\text{i}}\ \mathbf{m}'_{\text{iy}}$ and that $\mathbf{Y}'_{\text{c,i}}\mathbf{1}_{\text{K}_\text{i}} = \mathbf{0}$, under the constraint $\mathbf{1}'_{K_i}\ \mathbf{T}_{iw}\mathbf{A}_w = \mathbf{0}'$, Eq. (A1) can be simplified to:

$$\begin{aligned}\text{Tr}\Sigma_{\text{i}}[&\mathbf{Y}'_{\text{c,i}}\mathbf{Y}_{\text{c,i}} - 2\mathbf{Y}'_{\text{c,i}}\mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}} + \mathbf{m}_{\text{iy}}1'_{\text{Ki}}1_{\text{Ki}}\mathbf{m}'_{\text{iy}} - 2\mathbf{m}'_{\text{iy}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}}\\ &+ \mathbf{A}'_{\text{b}}\mathbf{t}_{\text{ib}}\mathbf{1}_{\text{Ki}}\mathbf{1}'_{\text{Ki}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}} + \mathbf{A}'_{\text{w}}\mathbf{T}'_{\text{iw}}\mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}}].\end{aligned} \tag{A2}$$

Thus, minimizing (A2) is equivalent to minimizing the sum of the following two functions:

$$\begin{aligned}&\text{Tr}\Sigma_{\text{i}}[\mathbf{Y}'_{\text{c,i}}\mathbf{Y}_{\text{c,i}} - 2\mathbf{Y}'_{\text{c,i}}\mathbf{A}'_{\text{w}}\mathbf{T}_{\text{iw}} + \mathbf{A}'_{\text{w}}\mathbf{T}'_{\text{iw}}\mathbf{T}_{\text{iw}}\ \mathbf{A}_{\text{w}}] = \Sigma_{\text{i}}\text{SSQ}\ (\mathbf{Y}_{\text{c,i}} - \mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}})\\ &\text{Tr}\Sigma_{\text{i}}[K_{\text{i}}\mathbf{m}_{\text{iy}}\mathbf{m}'_{\text{iy}} - \mathbf{1}_{\text{Ki}}2\mathbf{m}_{\text{iy}}\mathbf{1}'_{\text{Ki}}\mathbf{1}_{\text{Ki}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}} + \mathbf{A}'_{\text{b}}\mathbf{t}_{\text{ib}}\mathbf{1}'_{\text{Ki}}\mathbf{1}_{\text{Ki}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}}]\\ &\quad = \Sigma_{\text{i}}\text{SSQ}(\mathbf{1}_{\text{Ki}}\mathbf{m}'_{\text{iy}} - \mathbf{1}_{\text{Ki}}\mathbf{t}'_{\text{ib}}\mathbf{A}_{\text{b}}).\end{aligned}$$

Because these two functions deal with different parameter sets, the between and the within parts can be minimized separately. Thus, expression (11) is equivalent to:

$$\Sigma_{\text{i}}\text{SSQ}(\mathbf{Y}_{\text{c,i}} - \mathbf{T}_{\text{iw}}\mathbf{A}_{\text{w}}) + \text{SSQ}(\mathbf{W}[\mathbf{M}_{\text{y}} - \mathbf{T}_{\text{b}}\mathbf{A}_{\text{b}}]). \tag{A3}$$

References

Abraham B, Merola G (2005) Dimensionality reduction approach to multivariate prediction. Comput Stat Data Anal 48:5–16

Anderson TW (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann Math Stat 22:327–351

Anderson TW (2003) An introduction to multivariate statistical analysis. Wiley-Interscience, Hoboken, NJ

Bryk AS, Raudenbush SW (1992) Hierarchical linear models, applications and data analysis methods. Sage, Newbury Park, CA

Burnham AJ, Viveros R, MacGregor JF (1996) Frameworks for latent variable multivariate regression. J Chemom 10:31–45

Davies PT, Tso MK (1982) Procedures for reduced-rank regression. Appl Stat 3:244–255

de Jong S, Kiers HA (1992) Principal covariates regression: part I theory. Chemom Intell Lab Syst 14: 155–164

de Noord OE, Theobald EH (2005) Multilevel component analysis and multilevel PLS of chemical process data. J Chemom 19:301–307

Erlicher A, Lora A (2002) Pattern di trattamento e costi dei dipartimenti di salute mentale della regione Lombardia. Progetto di ricerca HoNOS 2. Pensiero Scientifico Editore, Milan, Italy
Goldstein H, McDonald RP (1988) A general model for the analysis of multilevel data. Psychometrika 53:455–468
Gower JC (1975) Generalized procrustes analysis. Psychometrika 40:33–51
Hox J (2002) Multilevel analysis. Erlbaum, Mahwah, NJ
Hwang H, Takane Y (2004) Generalized structured component analysis. Psychometrika 69:81–99
Hwang H, Takane Y, Malhotra N (2007) Multilevel generalized structured component analysis. Behaviormetrika 34:95–109
Izenman AJ (1975) Reduced-rank regression for the multivariate bilinear model. J Multivar Anal 5:248–264
Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Smilde AK (2005) Multilevel component analysis of time-resolved metabolic fingerprinting data. Analytica Chimica Acta 530:173–183
Jansen JJ, Hoefsloot HCJ, Westerhuis JA, van der Greef J, Timmerman ME, Smilde AK (2005) ASCA: analysis of multivariate data obtained from an experimental design. J Chemom 19:469–481
Joreskog KG (1993) Testing structural equation models. In: Bollen KA, Long SJ (eds) Testing structural equation models. Sage Publications, London, pp 294–316
Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. Psychometrika 23:187–200
Lora A, Bai G, Bianchi S, Bolongaro G, Civenti G, Erlicher A et al (2001) La versione italiana della HoNOS (Health of the Nation Outcome Scales), una scala per la valutazione della gravità e dell'esito nei servizi di salute mentale. Epidemiologia e Psichiatria Sociale 10:198–212
Lovaglio PG, Monzani E (2012, in press) Health of the nation outcome scales evaluation in a community setting population. Qual Life Res. doi:10.1007/s11136-011-0071-9
Massey WF (1965) Principal components regression in exploratory statistical research. J Am Stat Assoc 60:234–246
Meredith W, Millsap RE (1985) On component analysis. Psychometrika 50:495–507
Merola GM, Abraham B (2001) Dimensionality reduction approach to multivariate prediction. Can J Stat 29:191–200
Muthén BO (1991) Multilevel factor analysis of class and student achievement components. J Educ Meas 28:338–354
Muthén BO (1994) Multilevel covariance structure analysis. Sociol Methods Res 22:376–398
Neuhaus JM, Kalbfleisch JD (1998) Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics 54:638–645
Reise SP, Duan N (2003) Multilevel modelling. Methodological advances, issues, and applications. Erlbaum, Mahwah, NJ
Schönemann PH, Steiger JH (1976) Regression component analysis. Br J Math Stat Psychol 29:175–189
Scott AJ, Holt D (1982) The effect of two-stage sampling on ordinary least-squares methods. J Am Stat Assoc 77:848–854
Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers RJ, van der Greef J, Timmerman ME (2005) ANOVA simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics 21:3043–3048
Snijders TAB, Bosker RJ (1999) Multilevel analysis: an introduction to basic and advanced multilevel modelling. Sage, London
Takane Y, Hunter MA (2001) Constrained principal component analysis: a comprehensive theory. Appl Algebra Eng Commun Comput 12:391–419
Thissen U, Wopereis S, van den Berg SAA, Bobeldijk I, Kleemann R, Kooistra T et al (2009) Improving the analysis of designed studies by combining statistical modelling with study design information. BMC Bioinformatics 10:52
Timmerman ME (2006) Multilevel component analysis. Br J Math Stat Psychol 59:301–320
Timmerman ME, Kiers HAL, Smilde AK, Ceulemans E, Stouten J (2009) Bootstrap confidence intervals in multilevel simultaneous component analysis. Br J Math Stat Psychol 62:299–318
Timmerman ME, Kiers HL (2003) Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences. Psychometrika 86: 105–122
Van den Wollenberg R (1977) Redundancy analysis: an alternative for canonical correlation analysis. Psychometrica 42:207–219
van der Burg E, de Leeuw J, Dijksterhuis G (1994) Nonlinear canonical correlation with k sets of variables. Comput Stat Data Anal 18:141–163

Wing J, Curtis RH, Beevor AS, Park BG, Hadden S, Burns A (1998) Health of the nation outcome scales (HoNOS): research and development. Br J Psychiatry 172:11–18

Wold H (1982) Soft modeling: the basic design and some extensions. In: Joreskog KG, Wold H (eds) System under indirect observation: causality, structure, prediction. North-Holland, Amsterdam, pp 1–54