University of Groningen

PCovR2

Gvaladze, Sopiko; Vervloet, Marlies; Van Deun, Katrijn; Kiers, Henk A. L.; Ceulemans, Eva

Link to publication in University of Groningen/UMCG research database

PCovR2: A flexible principal covariates regression approach to parsimoniously handle multiple criterion variables

Sopiko Gvaladze[1] · Marlies Vervloet[1] · Katrijn Van Deun[2] · Henk A. L. Kiers[3] · Eva Ceulemans[1]

Abstract

Principal covariates regression (PCovR) allows one to deal with the interpretational and technical problems associated with running ordinary regression using many predictor variables. In PCovR, the predictor variables are reduced to a limited number of components, and simultaneously, criterion variables are regressed on these components. By means of a weighting parameter, users can flexibly choose how much they want to emphasize reconstruction and prediction. However, when datasets contain many criterion variables, PCovR users face new interpretational problems, because many regression weights will be obtained and because some criteria might be unrelated to the predictors. We therefore propose PCovR2, which extends PCovR by also reducing the criteria to a few components. These criterion components are predicted based on the predictor components. The PCovR2 weighting parameter can again be flexibly used to focus on the reconstruction of the predictors and criteria, or on filtering out relevant predictor components and predictable criterion components. We compare PCovR2 to two other approaches, based on partial least squares (PLS) and principal components regression (PCR), that also reduce the criteria and are therefore called PLS2 and PCR2. By means of a simulated example, we show that PCovR2 outperforms PLS2 and PCR2 when one aims to recover all relevant predictor components and predictable criterion components. Moreover, we conduct a simulation study to evaluate how well PCovR2, PLS2 and PCR2 succeed in finding (1) all underlying components and (2) the subset of relevant predictor and predictable criterion components. Finally, we illustrate the use of PCovR2 by means of empirical data.

Keywords Principal covariates regression · Dimension reduction · Regression models · Multiple criteria · PLS2

Introduction

Principal covariates regression (PCovR) was proposed by de Jong and Kiers (1992) to deal with the interpretational and technical problems that are encountered when applying ordinary least squares (OLS) regression using a relatively high number of predictor variables that are correlated to some extent. Interpreting the regression weights becomes challenging in such cases, as the weights only indicate the unique contributions of the predictor variables to the prediction problem and, thus, do not reflect possible shared effects (Cohen, Cohen, West, & Aiken, 2003). Moreover, the bouncing beta problem might occur, implying that removing or adding a single observation can already alter the regression weights drastically (Kiers & Smilde, 2007).

In PCovR, the predictor variables are reduced to a limited number of components that are linear combinations of the original predictors. Simultaneously, the criterion variables are regressed on these components. The components are extracted in such a way that they are good summarizers of the

---

Sopiko Gvaladze and Marlies Vervloet Joint first author

✉ Eva Ceulemans
  eva.ceulemans@kuleuven.be

  Sopiko Gvaladze
  sofiagvaladze13@gmail.com

  Marlies Vervloet
  marlies.vervloet@kuleuven.be

  Katrijn Van Deun
  k.vandeun@uvt.nl

  Henk A. L. Kiers
  h.a.l.kiers@rug.nl

[1] KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium

[2] Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

[3] Department of Psychology, University of Groningen, Groningen, The Netherlands

predictor data (i.e., explain a lot of the predictor variance), but also predict the criterion scores well (i.e., explain a lot of the criterion variance). PCovR has at least three attractive features: First, a closed-form solution is available for extracting the components. Second, the user may choose the extent to which either reduction or prediction is emphasized through a weighting parameter that ranges between 0 and 1. When the weighting parameter equals 0, PCovR corresponds to reduced rank regression (RRR, Aldrin, 2002), whereas a value of 1 implies that PCovR boils down to principal components regression (PCR: Jolliffe, 1982). Third, PCovR has been found to outperform popular methods such as partial least squares (PLS; Wold, Ruhe, Wold, & Dunn III, 1984) and PCR (Jolliffe, 1982) in retrieving relevant predictor components (i.e., the subset of predictor components that explain variance in the criterion block; see Vervloet, Van Deun, Van den Noortgate, & Ceulemans, 2016 and Van Deun, Crompvoets, & Ceulemans, 2018). A user-friendly R package is available for performing all the steps of a PCovR analysis (including preprocessing, model selection, rotation, etc.); the use of this package is described in Vervloet, Kiers, Van den Noortgate, and Ceulemans (2015). Note that PCovR differs from so-called coupled matrix factorization methods (Acar, Kolda, & Dunlavy, 2011; Acar, et al., 2014; De Lathauwer & Kofidis, 2017; Wilderjans, Ceulemans, & Van Mechelen, 2009; Wilderjans, Ceulemans, Van Mechelen, & van den Berg, 2011), in that PCovR treats predictors and criteria in an asymmetric way (i.e., reversing the role of the predictors and criteria would yield a different solution), whereas both sets of variables are treated symmetrically and have a similar status in coupled matrix factorization.

Performance of PCovR has mostly been studied in datasets that include one criterion variable only, however. When datasets contain a high number of criterion variables, PCovR users face new interpretational problems, because the number of regression weights will be high. Indeed, each criterion variable will correspond to a separate set of regression weights. Moreover, the criterion variables can be correlated with each other and partly include the same information; this may not be evident from these separate sets of regression weights, however. Finally, it might happen that some of the criteria are not related to the predictors, and thus cannot be predicted well. In line with what has been proposed in the PLS framework (i.e., PLS2, Manne, 1987), this interpretational burden can be lifted by also reducing the criterion variables to a few components and predicting these criterion components rather than the individual criteria. Such an approach facilitates interpretation, as it sheds light on the interrelations among the predictors and among the criteria, and reduces the number of regression weights that have to be interpreted. Herewith, it is important that the number of predictor components may differ from the number of criterion components, if one wants to fully grasp the predictor structure. Indeed, imposing an equality restriction on the number of predictor and criterion components, as is the case in PLS2, is more defendable if one is solely focused on predicting the criteria as well as possible.

In this paper, we therefore propose PCovR2, which extends PCovR by also reducing the criteria to a few components. These criterion components are predicted based on the predictor components, allowing both sets of components to differ in number. In line with PCovR, users can specify how much they want to emphasize reduction and prediction, by means of a weighting parameter. We will demonstrate that due to the weighting parameter, PCovR2 is indeed a very flexible tool. It can accommodate different research goals that researchers might want to pursue. Specifically, using a low weighting value allows one to prioritize retrieving the relevant predictor components and predictable criterion components and thus to focus on the regression problem, whereas with a high weighting value, the predictor and criterion components are formed based on how much variance they can explain in the corresponding data blocks, disregarding the link between the two blocks.

Moreverover, we also give suggestions on how to further decide on the number of predictor components and criterion components once the decision on the weighting value has been taken. Especially, finding the number of relevant predictor components and predictable criterion components is not trivial. Therefore, we put forward a model selection approach to assist users when taking this decision.

We will compare the performance of PCovR2 with two alternative approaches: PLS2 and PCR2. As mentioned above, PLS2 is more restrictive than PCovR2 in that it requires the number of criterion components to equal the number of predictor components (Manne, 1987). We will demonstrate that even though PLS2 has a high predictive power, it does not necessarily reconstruct the underlying predictor and criterion components well. PCR2 is a natural extension of PCR that boils down to performing two separate principal component analyses (PCA; Pearson, 1901), one on the predictors and one on the criterion variables. Next, it predicts the criterion components on the basis of the predictor components in an additional ordinary least squares (OLS) step. We expect that this approach will perform worse than PCovR2, if one aims to focus on extracting the relevant predictor components and predictable criterion components. Indeed, with PCR2, the predictor information is completely ignored when extracting the criterion components and vice versa. As a result, the obtained predictor components may be irrelevant for predicting the criterion components.

The remainder of the paper is structured as follows: First, we discuss the original PCovR method. We then present the new PCovR2 technique and discuss the model selection problem. We also show that PCR2 and PLS2 fail when the research focus lies on reconstructing the underlying relevant predictor and predictable criterion components. Second, we

present a simulation study in which the performance of PCovR2 is evaluated as well as the proposed PCovR2 model selection heuristic for deciding on the number of relevant predictor components and predictable criterion components. In the subsequent section, we analyze data from the 500 Family Study (Schneider & Linda, 2008). Discussion points and concluding remarks are given in the final section.

From PCovR to PCovR2

Data

The data consist of a block of predictors $\mathbf{X}$ ($N \times J$) and a block of criteria $\mathbf{Y}$ ($N \times K$), which are all measured for the same N observations. Preprocessing the data is important, as it will influence the results that are obtained. Specifically, we will assume all variables to be centered (i.e., the mean of each variable is set to zero). As holds for component analysis methods in general, centering the data is necessary in order to model the covariance structure of the data. Scaling is important if differences in variance between variables are considered as arbitrary, since variables with larger variances would otherwise influence the obtained components more. In this paper, we assume that the sum of squares of each variable is set to one, implying that each variable has an equal weight in the analysis.

PCovR

Model

The PCovR model (De Jong & Kiers, 1992) decomposes the predictor matrix $\mathbf{X}$ as follows:

$$\mathbf{X} = \mathbf{X}\mathbf{W_X}\mathbf{P_X}' + \mathbf{E_X} = \mathbf{T_X}\mathbf{P_X}' + \mathbf{E_X}. \quad (1)$$

$\mathbf{T_X} = \mathbf{X}\mathbf{W_X}$ is the $N \times R_X$ component score matrix that holds the scores of the N observations on the R_X components. The sum of squares per column of $\mathbf{T_X}$ is constrained to be equal to 1. $\mathbf{P_X}$ is the $J \times R_X$ loading matrix that contains the loadings of the predictor variables on the components. If the columns of $\mathbf{T_X}$ are mutually orthogonal, these loadings boil down to the correlations among the predictors and the components, since we assume the data to be centered around zero and to be scaled to the sum of squares of one per variable (see 2.1). $\mathbf{E_X}$ holds the residuals for the predictors, and $\mathbf{W_x}$ is a $J \times R_X$ weight matrix. The criteria in $\mathbf{Y}$ are regressed on the R_X components:

$$\mathbf{Y} = \mathbf{T_X}\mathbf{P_Y}' + \mathbf{E_Y}, \quad (2)$$

where the matrix $\mathbf{P_Y}$ ($K \times R_X$) contains the regression weights for each of the K criteria and $\mathbf{E_Y}$ the residuals for $\mathbf{Y}$. As

extensively discussed in Vervloet et al. (2015), PCovR solutions have rotational freedom as multiplying $\mathbf{P_Y}$ and $\mathbf{P_X}$ by a rotation matrix $\mathbf{Z}$ and counter-rotating $\mathbf{T_X}$ by multiplying it with $(\mathbf{Z}')^{-1}$ does not alter the reconstructed $\mathbf{X}$ or $\mathbf{Y}$ scores.

Estimation

In a PCovR analysis, the reduction of predictor variables to components, and the prediction of the criterion variables by those components, is conducted simultaneously. This can clearly be seen in the PCovR loss function L_{PCovR}:

$$L_{PCovR} = \alpha \frac{\|\mathbf{X} - \mathbf{T_X}\mathbf{P_X}'\|^2}{\|\mathbf{X}\|^2} + (1-\alpha)\frac{\|\mathbf{Y} - \mathbf{T_X}\mathbf{P_Y}'\|^2}{\|\mathbf{Y}\|^2}, \quad (3)$$

with α being a weighting parameter, ranging between 0 and 1, that specifies to what degree reduction and prediction are emphasized. As already mentioned in the introduction, when α equals 0, PCovR boils down to RRR (Aldrin, 2002) and, when α equals 1, to PCR (Jolliffe, 1982). High α values will lead to a higher focus on strong components (i.e., components that explain a lot of variance in $\mathbf{X}$), while low α values most likely lead to more relevant components (i.e., components that explain a lot of variance in $\mathbf{Y}$). In practice, balancing both aspects is often desirable, as using a too high α can lead to irrelevant components and using a too low α to uninterpretable ones. How this balance can be achieved is discussed in Section "Model selection".

Given a specific α value and number of components R_X, a closed-form solution exists for the PCovR parameters. Specifically, $\mathbf{T_X}$ is set equal to the first R_X eigenvectors of the matrix

$$\mathbf{D} = \alpha \frac{\mathbf{X}\mathbf{X}'}{\|\mathbf{X}\|^2} + (1-\alpha)\frac{\mathbf{H_X}\mathbf{Y}\mathbf{Y}'\mathbf{H_X}}{\|\mathbf{Y}\|^2}, \quad (4)$$

in which $\mathbf{H_X} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}'$ (see Vervloet et al., 2016, for more information). The loading matrix and regression weights can then be calculated, respectively, as $\mathbf{P_X}' = \mathbf{T_X}'\mathbf{X}$ and $\mathbf{P_Y}' = \mathbf{T_X}'\mathbf{Y}$.

Model selection

Performing model selection for PCovR is complicated, as one should determine both the number of components and the weighting parameter value in Eq. 3. Moreover, researchers may have different opinions on which components they want to retrieve. Some may aim to retrieve all important predictor components, irrespective of their relevance. Others may take the same stance as Vervloet et al. (2016) and consider a PCovR solution to be appropriate if all components that are relevant for predicting the criterion data, i.e., have a non-negligible regression weight, are recovered, irrespective

of the amount of predictor variance they explain. Different selection criteria can be used, such as cross-validation, scree tests, and combinations thereof. One could, for instance, first select the number of components by performing a scree test on the predictor data and afterwards use cross-validation to tune the weighting parameter. This approach allows one to focus on the number of strong components, i.e., the subset of components that explain a lot of variance (Ceulemans & Kiers, 2009). Yet, these strong components are not necessarily relevant for prediction purposes. Alternatively, one could simultaneously optimize the number of components and the weighting parameter using K-fold cross-validation (Breiman, Friedman, Olshen, & Stone, 1984). This might be cumbersome, however, because of the large number of possible combinations. Moreover, the study by Vervloet et al. (2016) has shown that cross-validation tends to extract too many components. Importantly, the results reported by Vervloet et al. (2013, 2016) all indicate that using a low α is a good idea, if one is mainly interested in finding the relevant components. Indeed, a low α will penalize extracting components that do not help in reconstructing the criterion data.

PCovR2

Model

In PCovR2, we still assume the columns of the predictor matrix $\mathbf{X}$ and the criterion matrix $\mathbf{Y}$ to be centered and scaled to a unit sum of squares. The predictor matrix $\mathbf{X}$ is decomposed according to Eq. 1. The criterion matrix $\mathbf{Y}$ is modeled as follows:

$$\mathbf{Y} = \mathbf{Y}\mathbf{W_Y}\mathbf{P_Y}' + \mathbf{E_Y} = \mathbf{T_Y}\mathbf{P_Y}' + \mathbf{E_Y}. \tag{5}$$

with $\mathbf{T_Y}$ holding the $N \times R_Y$ scores of the N observations on the R_Y criterion components and $\mathbf{P_Y}$ the $K \times R_Y$ loadings of the criterion variables on the criterion components. Imposing $\mathbf{T_Y}'\mathbf{T_Y} = \mathbf{I}$, these loadings reflect the correlations among the criteria and the criterion components. $\mathbf{E_Y}$ holds the residuals for the criteria, and $\mathbf{W_Y}$ is a $K \times R_Y$ weight matrix. Finally, $\mathbf{T_Y}$ is regressed upon $\mathbf{T_X}$:

$$\mathbf{T_Y} = \mathbf{T_X}\mathbf{B}' + \mathbf{E_{T_Y}}, \tag{6}$$

where $\mathbf{B}$ is an $R_Y \times R_X$ matrix that holds the associated regression weights, and $\mathbf{E_{T_Y}}$ contains the corresponding residuals. Because the columns of $\mathbf{T_X}$ and $\mathbf{T_Y}$ have a sum of squares of one, the regression weights can be directly compared in size.

PCovR2 solutions have rotational freedom, in that without losing fit, the predictor components and criterion components can be rotated independently from each other to facilitate interpretation. Indeed, multiplying $\mathbf{P_X}$ and $\mathbf{P_Y}$ by the rotation matrices $\mathbf{Z_X}$ and $\mathbf{Z_Y}$ and counter-rotating $\mathbf{T_X}$ and $\mathbf{T_Y}$, by multiplying them with $(\mathbf{Z_X}')^{-1}$ and $(\mathbf{Z_Y}')^{-1}$, does not alter the reconstructed $\mathbf{X}$ or $\mathbf{Y}$ scores. For instance, both sets of

components can be rotated using a simple structure rotation criterion such as varimax. Obviously, such rotations would imply that the component scores $\mathbf{T_X}$ and $\mathbf{T_Y}$ as well as the regression weights $\mathbf{B}$ need to be recalculated to take the rotation into account. When applying orthogonal rotations, the following formula can be used:

$$\mathbf{B}_{rotated} = \mathbf{Z_Y}'\mathbf{B}\mathbf{Z_X} \tag{7}$$

Estimation

The key feature of PCovR2 is that the three sub-models—reduction of the predictor block, reduction of the criterion block, and prediction of the criterion components based on the predictor components—are fitted simultaneously. This leads to the following loss function to be minimized:

$$
\begin{aligned}
L_{PCovR2} &= \alpha \frac{\|\mathbf{X}-\mathbf{X}\mathbf{W_X}\mathbf{P_X}'\|^2}{\|\mathbf{X}\|^2} + \alpha \frac{\|\mathbf{Y}-\mathbf{Y}\mathbf{W_Y}\mathbf{P_Y}'\|^2}{\|\mathbf{Y}\|^2} + (1-\alpha)\frac{\|\mathbf{Y}\mathbf{W_Y}-\mathbf{X}\mathbf{W_X}\mathbf{B}'\|^2}{\|\mathbf{Y}\mathbf{W_Y}\|^2} \\
&= \alpha\left(\frac{\|\mathbf{X}-\mathbf{T_X}\mathbf{P_X}'\|^2}{\|\mathbf{X}\|^2} + \frac{\|\mathbf{Y}-\mathbf{T_Y}\mathbf{P_Y}'\|^2}{\|\mathbf{Y}\|^2}\right) + (1-\alpha)\frac{\|\mathbf{T_Y}-\mathbf{T_X}\mathbf{B}'\|^2}{R_Y},
\end{aligned}
\tag{8}
$$

where R_Y refers to the number of criterion components with $0 < \alpha \le 1$. Again, α is a tuning parameter that allows one to put more emphasis on the prediction aspect (α close to zero) or the reduction aspect (α close to one). Note that setting α to 1 corresponds to a PCR2 approach. Moreover, setting α to 0 will lead to a method that focuses entirely on finding strongly related predictor and criterion components, but totally ignores how much variance of the predictor and criterion matrix is explained by these components. The latter is not desirable for our purposes as we want to recover the components that underlie both predictor and criterion matrices. Therefore, we do not recommend setting α to 0.

Given specific α, R_X, and R_Y values, we minimize (8) under the constraint that $\mathbf{T_X}'\mathbf{T_X}$ and $\mathbf{T_Y}'\mathbf{T_Y}$ equal an identity matrix. To this end, we rely on an alternating lower[1] squares algorithm since no closed-form solution is readily available. The PCovR2 algorithm alternates between two steps that we discuss in detail below: 1) conditional estimation of $\mathbf{T_X}$, $\mathbf{P_X}$, and $\mathbf{B}$ given fixed values for $\mathbf{T_Y}$ and $\mathbf{P_Y}$, and 2) conditional estimation of $\mathbf{T_Y}$ and $\mathbf{P_Y}$ given fixed values for $\mathbf{T_X}$, $\mathbf{P_X}$, and $\mathbf{B}$. Note that step 2 consists of running an iterative procedure. Both steps are repeated until convergence of the loss function (i.e., $L_{previous} - L_{current} < \varepsilon$, where ε is a pre-specified small number—we used 10^{-6}). Because of the alternating nature of the algorithm, we advise the use of multiple starts, to avoid ending up in a local minimum.

a) *Step 1*

---

[1] We use the term alternating lower squares rather than alternating least squares, because step 2 consists of an iterative procedure.

Given current estimates of $\mathbf{T_Y}$ and $\mathbf{P_Y}$, we can update $\mathbf{T_X}$, $\mathbf{P_X}$, and $\mathbf{B}$ by optimizing

$$L_{step}1 = \alpha \frac{\|\mathbf{X}-\mathbf{T_X P_X'}\|^2}{\|\mathbf{X}\|^2} + (1-\alpha)\frac{\|\mathbf{T_Y}-\mathbf{T_X B'}\|^2}{\|\mathbf{T_Y}\|^2}. \qquad (9)$$

In this optimization problem, a standard PCovR problem can be recognized, with $\mathbf{X}$ as the predictor matrix and $\mathbf{T_Y}$ as the criterion matrix. Hence, this step has a closed-form solution, which boils down to setting $\mathbf{T_X}$ equal to the first R_X eigenvectors of matrix $\mathbf{D}$:

$$\mathbf{D} = \alpha\frac{\mathbf{XX'}}{\|\mathbf{X}\|^2} + (1-\alpha)\frac{\mathbf{H_X T_Y T_Y' H_X}}{\|\mathbf{T_Y}\|^2}, \qquad (10)$$

with $\mathbf{H_X} = \mathbf{X(X'X)^{-1}X'}$. $\mathbf{P_X}$ and $\mathbf{B}$ can then easily be computed by means of regression steps (see Section "Estimation"), as $\mathbf{P_X} = \mathbf{X'T_X}$ and $\mathbf{B} = \mathbf{T_y'T_X}$.

b) *Step 2*

Given current estimates of $\mathbf{T_X}$, $\mathbf{P_X}$, and $\mathbf{B}$, the non-constant part of the PCovR2 loss function can be re-expressed as a function of $\mathbf{T_Y}$ and $\mathbf{P_Y}$ (and later of $\mathbf{T_Y}$ alone, when the optimal $\mathbf{P_Y}$ is substituted by $\mathbf{Y'T_Y}$) as follows:

$$\begin{aligned} L_{Step2} &= \alpha\frac{\|\mathbf{Y}-\mathbf{T_Y P_Y'}\|^2}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{T_X B'}\|^2 \\ &= \alpha\frac{\|\mathbf{Y}-\mathbf{T_Y P_Y'}\|^2}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2 \\ &= -\alpha\frac{tr[\mathbf{Y'T_Y T_Y'Y}]}{\|\mathbf{Y}\|^2} - \frac{2(1-\alpha)}{R_Y}tr(\mathbf{U'T_Y}) + c \\ &= -\alpha\frac{tr[\mathbf{Y'YW_Y W_Y'Y'Y}]}{\|\mathbf{Y}\|^2} - \frac{2(1-\alpha)}{R_Y}tr(\mathbf{U'YW_Y}) + c \\ &= -\alpha\frac{tr[\mathbf{V'Y'YV}]}{\|\mathbf{Y}\|^2} - \frac{2(1-\alpha)}{R_Y}tr\left(\mathbf{U'Y(Y'Y)^{-1/2}V}\right) + c \\ &= tr[\mathbf{V'CV}] + tr[\mathbf{AV}] + c. \end{aligned} \qquad (11)$$

with $\mathbf{U} = \mathbf{T_X B'}$, $\mathbf{V} = \mathbf{(Y'Y)^{1/2}W_Y}$, $\mathbf{A} = -\frac{2(1-\alpha)}{R_Y}\mathbf{U'Y(Y'Y)^{-1/2}}$, , and c being a constant (see appendix for the detailed derivation of Eq. 11). This is a function that can be minimized through iterative majorization (Kiers, 2002). In each step of this iterative majorization procedure, the parameters are updated by solving a simple problem for which a closed-form solution is available. Here, the problem boils down to updating $\mathbf{V}$ subject to $\mathbf{T_Y'T_Y} = \mathbf{W_Y'Y'YW_Y} = \mathbf{V'V} = \mathbf{I}$. Specifically, for updating $\mathbf{V}$ (i.e.,$\mathbf{V}^{update}$), the current estimates of $\mathbf{V}$ (i.e., $\mathbf{V}^{current}$) are used to construct the matrix $\mathbf{G}$:

$$\mathbf{G} = \mathbf{A} + 2\mathbf{V}^{current'}\mathbf{C}-2\lambda\mathbf{V}^{current'}, \qquad (12)$$

with λ being a value higher or equal to the highest eigenvalue of $\mathbf{C}$. Next, we compute a singular value decomposition of this matrix $-\mathbf{G}$ as:

$$svd(-\mathbf{G}) = \mathbf{PDQ'}. \qquad (13)$$

$\mathbf{V}^{update}$ can then be calculated as

$$\mathbf{V}^{update} = \mathbf{QP'}. \qquad (14)$$

This majorization procedure within step 2 is repeated until $L_{Step2, \ previous} - L_{Step2, \ current} < \varepsilon$. Next, $\mathbf{T_Y}$ and $\mathbf{P_Y}$ are computed based on the final estimate of $\mathbf{V}$. Specifically, $\mathbf{T_Y} = \mathbf{Y(Y'Y)^{-1/2}V}$ and $\mathbf{P_Y} = \mathbf{Y'T_Y}$.

Model selection

Defining the problem If model selection was already complicated for PCovR (see 2.2.3), it obviously is even more challenging for PCovR2, because one now needs to decide on R_X, R_Y, as well as on α. The choice regarding the R_X, R_Y, and α values will again depend on the research goal. Indeed, as indicated in the introduction, some researchers aim to retrieve all the predictor and criterion components, regardless if the predictor components are relevant or not (i.e., explain variance of the criterion variables) and regardless if the criterion components are predictable or not (i.e., capture variance that is related to the predictors). Some take a narrower regression perspective, implying that they are only interested in retrieving the relevant predictor and predictable criterion components. On top of that, they often want to interpret the structure of the predictors and criteria, and thus want the retrieved relevant predictor and predictable criterion components to be an accurate representation of the predictor and criterion scores.

To illustrate these two different scenarios further, let us consider a simulated toy example. The dataset was simulated with three predictor components and with two criterion components. The associated loading matrices and regression weight matrix are shown in Table 1. The true number of relevant predictor and predictable criterion components can be derived from the simulated regression weights matrix. In particular, we see that the first criterion component is predictable (i.e., has at least one non-negligible regression weight), and the first two predictor components are relevant (i.e., have at least one non-negligible regression weight), whereas the second criterion component is not predictable, and the third predictor component is irrelevant, as the associated regression weights are all equal to zero.

To demonstrate how the reconstruction of the respective $\mathbf{X}$ and $\mathbf{Y}$ blocks and the prediction of the criterion components, $\mathbf{T_Y}$, is affected by the weighting value across different R_X and R_Y values, we analyzed the data multiple times, where we varied R_X from 1 to 4 and R_Y from 1 to 3, and let α increase from .001 to 1, in steps of 001. Afterwards, we computed how

Table 1 The loadings and regression weights used to simulate a toy example, next to the PCovR2 loadings and regression weights for the solutions with R_X=2, R_Y=1 and R_X=3, R_Y=1 components

| | | Values used to simulate the data | | | | | Results of PCovR2 with R_X=2 and R_Y=1 | | | Results of PCovR2 with R_X=3 and R_Y=1 | | | |
| | | P_X | | | P_Y | | P_X | | P_Y | P_X | | | P_Y |
		pc1	pc2	pc3	cc1	cc2	pc1	pc2	cc1	pc1	pc2	pc3	cc1
Loadings		**1**	0	0	**1**	0	**.60**	.02	**.91**	**.88**	.03	-.06	**.91**
		1	0	0	**1**	0	**.70**	.04	**.88**	**.89**	.05	.07	**.88**
		1	0	0	**1**	0	**.62**	−.14	**.86**	**.91**	−.12	−.07	**.86**
		0	**1**	0	0	**1**	**.61**	.06	.08	−.01	.05	**.91**	.08
		0	**1**	0	0	**1**	**.52**	−.05	.05	−.11	−.06	**.88**	.05
		0	**1**	0	0	**1**	**.58**	.06	−.05	.01	.05	**.85**	−.05
		0	**1**	0	0	**1**	**.61**	−.03	.04	.00	−.04	**.89**	.04
		0	**1**	0	0	**1**	**.61**	.02	.02	.02	.01	**.87**	.02
		0	**1**	0	0	**1**	**.62**	.04	−.02	.02	.03	**.89**	−.02
		0	0	**1**	0	**1**	−.01	**.88**	.04	−.03	**.88**	.02	.04
		0	0	**1**	0	**1**	.03	**.86**	−.04	.05	**.86**	.00	−.04
		0	0	**1**	0	**1**	−.03	**.88**	−.01	−.03	**.88**	−.03	−.01
		0	0	**1**	0	**1**	.05	**.89**	.05	.01	**.89**	.06	.05
		0	0	**1**	0	**1**	.09	**.87**	.05	.03	**.87**	.09	.05
		0	0	**1**	0	**1**	.00	**.91**	.05	−.06	**.91**	.05	.05
		0	0	**1**	0	**1**	.02	**.90**	.05	−.03	**.90**	.05	.05
		0	0	**1**	0	**1**	.02	**.87**	−.01	.08	**.87**	−.06	−.01
		0	0	**1**	0	**1**	−.06	**.88**	−.02	−.05	**.88**	−.04	−.02
		0	0	**1**	0	**1**	−.04	**.87**	.01	−.05	**.87**	−.02	.01
		0	0	**1**	0	**1**	.00	**.86**	.07	.03	**.86**	−.04	.07
		0	0	**1**	0	**1**	−.04	**.88**	.00	−.04	**.88**	−.02	.00
		0	0	**1**	0	**1**	.03	**.85**	.02	.01	**.85**	.03	.02
		0	0	**1**	0	**1**	.02	**.89**	.02	.02	**.89**	.00	.02
		0	0	**1**	0	**1**	−.04	**.87**	.02	−.06	**.87**	.00	.02
Regression weights	cc1	.71	.71	**0**			**.95**	−.02		**.70**	.02	**−.65**	
	cc2	0	0	0									

To obtain these solutions we set α to .05. The simulated regression weights show that this example includes 3 predictor components and 2 criterion components in total, but that only two of the three predictor components are relevant and only one of the criterion components is predictable. The loadings and regression weights of these relevant and predictable components are printed in bold. pc1 up to pc3 denote the first up to third predictor component and similarly cc1 and cc2 denote the first and the second criterion component

well **X** and **Y** were reconstructed [i.e., first part of the loss function (8)] and how well the criterion components could be predicted (i.e., second part of the loss function). We plotted the obtained values in Fig. 1, where each panel represents a different R_X and R_Y combination. The x-axis of each plot pertains to the α-value and the y-axis to the loss function parts.

We first turn to the goal of retrieving all predictor and criterion components, disregarding their relevance or predictability. From Fig. 1, it can be concluded that increasing the number of R_X and R_Y clearly improves the reconstruction of the **X** and **Y** blocks, respectively, with this improvement being more pronounced for high α values (i.e., larger than .5; this is confirmed by the simulation results in Section "Recovery of the number of relevant predictor and the number of predictable criterion components as well as recovery of all the simulated components"). As could be expected, this improvement levels off once R_X = 3 and R_Y = 2, in that going further up than or to the right of the R_X = 3 and R_Y = 2 plot hardly changes the picture any further, so this suggests that one should not take R_X and R_Y higher than these values. The latter could be further verified by running a scree test (Cattell, 1966; Ceulemans & Kiers, 2006; Wilderjans, Ceulemans, & Meers, 2013). Thus, Fig. 1 might be useful to decide on the total numbers of predictor and criterion components.

When it comes to the regression problem and deciding on the number of relevant predictor components and predictable criterion components only, Fig. 1 shows that using a low α value (i.e., lower than .2) yields clearly better prediction
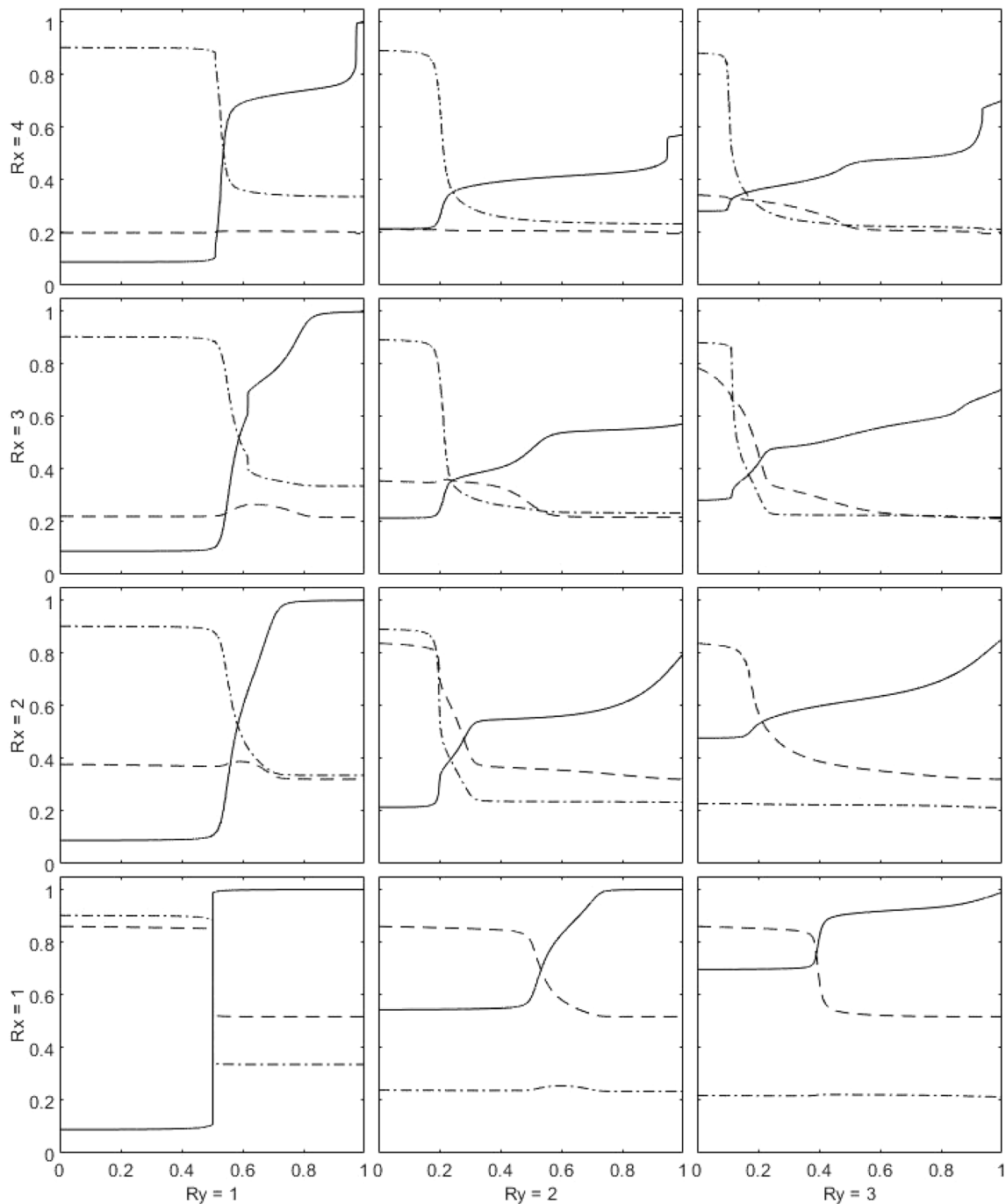
Fig. 1 Values of the reconstruction and prediction parts of the PCovR2 loss function (shown on the Y-axis) for the toy data as a function of the value of the weighting parameter α (shown on the X-axis). The panels pertain to 1 to 4 predictor components (rows) and 1 to 3 criterion components (columns). The dashed line '–', dash-dot line '-.' and solid line '-' denote the reconstruction error of X and Y, and the prediction error, respectively

results and is therefore clearly indicated (this will again be confirmed by the simulation study). Moreover, we see a specific prediction trend for α values that are very close to 0: it makes no sense to take more predictor components than there are criterion components, as it would not further reduce the prediction loss. Note that the models with $R_Y = 1$ and with one to four predictor components, for example, all show quite

similar prediction error values for low α values. Analogously, if $R_Y = 2$, taking more than two predictor components does not lead to a decreased prediction loss, and generally, if R_x is larger than R_Y, this does not lead to reduced prediction loss compared to what is found when $R_x = R_Y$.

To understand how picking low α values affects the reconstruction part of the model given these prediction findings, we

inspect the associated loading and regression matrices from the ($R_X = 2$, $R_Y = 1$, $\alpha = .05$) analysis in Table 1. Given that the simulated data contain two relevant predictor components and one predictable criterion component, we might a priori have expected good recovery of these components. Yet, the prediction findings in the previous paragraphs already suggest that this might not be the case. Indeed, Table 1 shows that the relevant predictor components are not well recovered but are rather merged into one component[2]. This merged component has a very high regression weight (.95), while the second predictor component has a low regression weight (.02) and corresponds to the irrelevant component. Further scrutinizing the results, we see that in order to properly recover both relevant predictor components, we actually need to extract more predictor components. Indeed, the ($R_X = 3$, $R_Y = 1$, $\alpha = .05$) PCovR2 solution adequately recovers the structure of the two relevant predictor components. Note that the ($R_X = 3$, $R_Y = 2$, $\alpha = .05$) analysis again yields merged predictor components; this makes sense as Fig. 1 shows that for this (R_X, R_Y) combination, we need a higher α value to properly reconstruct **X**. Overall, these findings indicate that model selection in terms of retrieving the relevant predictor and predictable criterion components is not an easy matter, and it cannot be fully solved by scrutinizing a plot like Fig. 1. To further supplement the decision process, and also to make automatic selection possible, in the next section, we propose a two-step model selection heuristic that users can apply to get a better grip on this problem for a particular data set.

A two-step model selection heuristic to retrieve all relevant and predictable components To retrieve all relevant and predictable components, we developed a two-step approach that can be automated. To run this approach, one estimates multiple PCovR2 models with R_X and R_Y varying from 1 up to a pre-defined number. In the first step, we decide on the number of predictable criterion components, R_Y. In the second step, we assess the number of relevant predictor components, R_X. We fixed the α-value to .05, since Fig. 1 points towards using a small value in case one wants to focus on the relevant and predictable components. Note that using other small α values would lead to the same conclusions.

Step 1: *Finding the number of predictable criterion components*

To decide on the number of predictable criterion components, we assess for each criterion component how well it is reconstructed by the predictor components. To this end, we calculate for each criterion component the sum of the

corresponding squared entries of $\mathbf{E_{T_Y}}$ (see Eq. 6), yielding a criterion component-specific prediction error. Panel a of Fig. 2 shows the obtained prediction errors for the criterion components of the toy example, revealing a lot of variation in predictability within and across solutions. At this point, we need a rule to decide when we consider a criterion component to be predictable. We propose to compare the prediction errors of each criterion component to a threshold value T_{Ry} that is derived from the data. To obtain this threshold, we make a histogram of all the prediction errors; panel b of Fig. 2 shows the histogram for the toy example. We observe a lot of variation in the component-specific prediction errors that seem to consist of five groups of values (one of which is small). The gaps in prediction errors between the five groups yield four possible thresholds corresponding to the highest values of the first four groups. To restrict oneself to the most predictable components, we suggest being strict and to set the threshold T_{Ry} at the highest value of the first group, which in this case equals .10. We see that none of the rows of the table in panel a of Fig. 2, and thus none of the 25 considered PCovR solutions, contains more than one value smaller than .10, and hence we can conclude that, according to our rule, none of the solutions contain more than one predictable criterion component. Therefore, we decide that these data contain only one predictable criterion component. Note that if we set the threshold in a less strict way, by using, for example, the highest prediction error in the second group of the histogram, we would have some rows that contain two, three, or even four predictable criterion components, which would obviously be too high for this dataset.

Step 2: *Finding the number of relevant predictor components*

Once we have chosen the number of predictable criterion components, we assess the number of relevant predictor components. To do so, we examine the regression weights of the different solutions with the selected number of criterion components. These solutions vary with respect to the value of R_X. For instance, panel c of Fig. 2 shows these regression weights for the toy example, with the number of predictor components R_X varying from one to five. To decide which regression weights indicate a relevant component, we again make use of a threshold T_{Rx}. This threshold will be computed based on the threshold T_{Ry} from the first step. In case only one predictor component is needed to explain the predictable criterion component, the corresponding regression weight would amount to $\sqrt{1-T}$. As multiple relevant predictor components might be needed however to reconstruct the predictable criterion component, we set the threshold lower, and we somewhat arbitrarily choose to set it at $T_{Rx} = \sqrt{1-T}/2$. Note that computing T_{Rx} in this way allows one to automate and evaluate the model selection method in a simulation study (see later).

---

[2] Pilot simulations showed that when we estimate the PCovR2 model setting R_Y equal to the number of predictable criterion components, we always recover these components well, without merging or other problems occurring.

R_Y	R_X	sum of squared prediction errors of each of the extracted criterion components (cc)				
		cc1	cc2	cc3	cc4	cc5
1	1	.09				
	2	.09				
	3	.09				
	4	.09				
	5	.09				
2	1	.09	1.00			
	2	.34	.09			
	3	.09	.34			
	4	.09	.34			
	5	.34	.09			
3	1	1.00	.09	.99		
	2	1.00	.09	.34		
	3	.34	.09	.42		
	4	.09	.41	.34		
	5	.09	.41	.34		
4	1	.99	.10	1.00	1.00	
	2	.09	1.00	.68	.66	
	3	1.00	.09	.42	.34	
	4	.40	.43	.09	.41	
	5	.09	.38	.43	.42	
5	1	1.00	1.00	1.00	.10	.99
	2	1.00	.10	1.00	.70	.63
	3	1.00	.10	.69	.42	.65
	4	.09	.43	.98	.43	.38
	5	.42	.54	.44	.09	.38

a)

b)
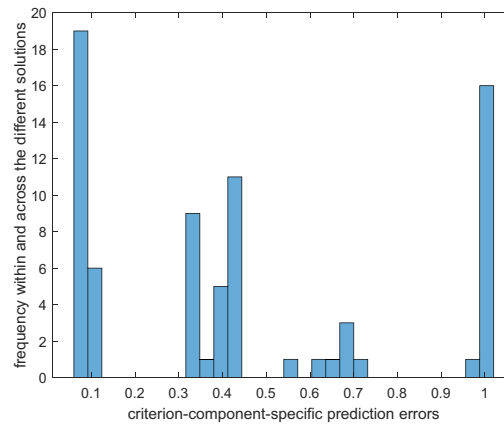
R_X	Regression weights					# Relev pc
1	.95					1
2	.95	.02				1
3	.70	.02	-.65			2
4	.70	.02	.65	.03		2
5	.70	.02	.65	.02	.09	2

c)

Fig. 2 (**a**) The sum of squared prediction errors for each criterion component across 25 PCovR2 models, estimated with R_X and R_Y varying from one up to five. cc1 up to cc5 indicates the first up to the fifth criterion component within the considered solution. Predictable criterion components are marked in bold. (**b**) Histogram presenting the distribution of all the sum of squared prediction errors reported in Fig. 2a. (**c**) The regression weights that are obtained when R_Y is fixed to 1 and R_X varies from 1 to 5

When applying the method to real data, one could again make a histogram of the regression weights and evaluate different options. For the toy example, T_{Rx} equals .47, which implies that some solutions in the regression weights table (panel c of the Fig. 2) include one relevant predictor only and some two. We propose selecting the most parsimonious solution with the highest number of relevant predictor components, as in such a solution, relevant components will probably not have been merged and the number of irrelevant predictor components remains low. So this means that for the toy data, we would select the solution with three predictor components, but would communicate that only two of these three components are relevant. The fact that an irrelevant component shows up in the solution is a consequence of the fact that our aim is not only to get good predictions, but also good reconstruction of **X**.

PCR2 and PLS2 fail in parsimoniously recovering the relevant predictor and predictable criterion components

Using the toy example in Table 1, the previous section demonstrated how relevant predictor and criterion components can

be found in a parsimonious way with PCovR2. We now investigate how PCR2 and PLS2 handle the problem, using the same data.

PCR2 PCR2 performs the reduction and regression steps in a sequential way, starting with the reduction step. Therefore, we expect that when the used R_X and R_Y are lower than the underlying ones and the relevant/predictable components are weaker than the irrelevant/unpredictable ones, PCR2 will extract the irrelevant/unpredictable components rather than the relevant/predictable ones. Remember that PCR2 corresponds to setting $\alpha = 1$, as confirmed by Fig. 1, indeed, noting the bad prediction score for $R_X = 2$ and $R_X = 1$ solution. It becomes clear that we need to set R_X to three and R_Y to two to extract the relevant predictable and predictable criterion components (see Table 2).

PLS2 Like PCovR2, PLS2 performs the prediction and reduction steps simultaneously. However, PLS2 restricts R_X to be equal to R_Y and implies a one-to-one correspondence between the respective predictor and criterion components. Therefore,

Table 2 The loadings and the (criterion) component-specific sum of squared prediction errors for the PCR2 solution with R_X=3 and R_Y=2 and the PLS2 solutions with 1, 2 and 3 components

| | PCR2 with R_X=3 and R_Y=2 | | | | | PLS2 with R_X=R_Y=1 | | PLS2 with R_X=R_Y=2 | | | | PLS2 with R_X=R_Y=3 | | | | | |
| | P_X | | | P_Y | | P_X | P_Y | P_X | | P_Y | | P_X | | | P_Y | | |
	pc1	pc2	pc3	cc1	cc2	pc1	cc1	pc1	pc2	cc1	cc2	pc1	pc2	pc3	cc1	cc2	cc3
Loadings	.03	−.06	**−.89**	−.03	**−.91**	−.35	**−.63**	−.23	**−.51**	**−.72**	−.22	.04	.06	**.7**	**.72**	.21	−.15
	.05	.08	**−.91**	.03	**−.89**	**−.49**	**−.63**	−.34	**−.62**	**−.67**	−.07	.1	.06	**.89**	**.73**	.18	.1
	−.12	−.06	**−.92**	−.07	**−.9**	−.35	**−.62**	−.24	**−.45**	**−.63**	.06	.01	−.1	**.79**	**.68**	.05	.12
	.05	**.9**	.03	**−.86**	−.05	**−.79**	−.06	**−.83**	.08	.06	**.43**	**.89**	.05	−.03	−.14	**−.58**	−.14
	−.06	**.89**	.09	**−.86**	−.08	**−.75**	−.05	**−.8**	.13	.08	**.44**	**.83**	−.08	.02	−.17	**−.62**	−.17
	.05	**.86**	−.03	**−.86**	.05	**−.78**	.06	**−.8**	−.02	.1	.13	**.81**	.04	.12	−.18	−.29	−.23
	−.04	**.88**	.01	**−.89**	−.05	**−.78**	0	**−.83**	.11	.11	**.4**	**.86**	−.05	.02	−.14	**−.45**	.03
	.01	**.87**	−.02	**−.88**	−.05	**−.79**	−.01	**−.83**	.09	.09	.38	**.88**	.01	−.01	−.15	**−.48**	−.06
	.03	**.9**	−.04	**−.85**	−.01	**−.82**	0	**−.85**	.01	.13	**.47**	**.87**	.02	.1	−.21	**−.62**	−.12
	.88	.01	.05	**−.9**	−.05	−.04	−.01	.11	**−.6**	.09	.33	.02	**.88**	−.07	−.15	**−.45**	−.1
	.86	−.01	−.03	**−.85**	−.01	−.06	.02	.11	**−.68**	.07	.16	−.02	**.86**	.06	−.15	−.33	−.24
	.88	−.04	.04	**−.89**	.04	0	.05	.13	**−.55**	.18	**.47**	.02	**.89**	−.16	−.21	**−.52**	.06
	.89	.06	−.02	**−.85**	−.04	−.12	0	.04	**−.66**	.06	.22	.06	**.89**	.02	−.14	−.37	−.2
	.87	.08	−.03	**−.87**	−.08	−.14	−.03	.01	**−.63**	.11	**.51**	.1	**.88**	−.01	−.16	**−.6**	−.01
	.91	.05	.05	**−.87**	−.09	−.08	−.02	.09	**−.69**	.14	**.6**	.01	**.9**	.03	−.2	**−.7**	.01
	.9	.07	−.01	**−.87**	−.04	−.12	0	.06	**−.71**	.11	.38	.03	**.88**	.09	−.15	**−.46**	−.01
	.87	−.07	−.06	**−.87**	.01	−.02	.04	.14	**−.66**	.12	.27	−.04	**.88**	.01	−.18	−.38	−.1
	.88	−.03	.03	**−.88**	.07	−.01	.02	.16	**−.69**	.12	.36	−.08	**.86**	.06	−.21	**−.52**	−.16
	.87	−.03	.07	**−.87**	−.03	0	.02	.16	**−.63**	.12	.36	−.06	**.86**	−.02	−.15	**−.41**	.02
	.86	−.03	−.04	**−.88**	−.04	−.04	−.04	.13	**−.7**	.07	.39	−.06	**.85**	.1	−.12	**−.49**	−.06
	.88	−.01	.01	**−.88**	−.01	−.04	.02	.13	**−.69**	.13	.38	−.05	**.87**	.06	−.16	**−.43**	.02
	.85	.03	.01	**−.87**	−.03	−.07	.01	.09	**−.66**	.12	**.41**	0	.85	.05	−.18	**−.52**	−.06
	.89	−.01	−.01	**−.87**	−.02	−.05	.03	.11	**−.65**	.1	.24	.01	**.89**	−.02	−.15	−.34	−.09
	.87	0	.05	**−.87**	−.05	−.03	.01	.11	**−.58**	.15	**.53**	.02	**.88**	−.1	−.18	**−.58**	.08
cc–specific prediction errors				1.00	.15		.22			.21	.95				.18	.9	.93

Loadings with absolute values higher than .40 are printed in bold

when the number of underlying relevant predictor components differs from the number of predictable criterion components, we expect PLS2 to miss out on some of these components or to merge them. The PLS2 loadings in Table 2, obtained using one, two, or three components, support this expectation. The solution with one component recovers the predictable criterion component, but merges the two relevant predictor components. When we extract one more component, the weakest relevant predictor component is merged with the irrelevant one. In the solution with three components, all predictor components are recovered properly, but the criterion component that corresponds to the weakest relevant component is not meaningful, as can also be derived from the (criterion) component-specific sums of squared prediction errors (see Table 2).

Simulation study

In this section, we report the results of a simulation study in which we evaluate which criterion and predictor components are recovered by the PCovR2 algorithm and how this is affected by the α value. Second, we evaluate the performance of the proposed model selection heuristic.

In line with the PCovR simulation studies of Vervloet et al. (2013, 2016, 2018), we manipulated the *relevance* and the *strength* of the predictor component. Additionally, we also varied the *predictability* and *strength* of the criterion components. The relevance of the predictor components is defined as follows. Predictor components that are relevant are useful in modeling part of the criterion variance (i.e., they have at least one large regression weight), whereas irrelevant predictor

components are not, and have regression weights of zero. The predictability of the criterion components is also determined by the regression weights. Whereas irrelevant criterion components cannot be predicted based on the predictor components and thus correspond with regression weights of zero, predictable criterion components are associated with at least one larger regression weight. The strength of a predictor or criterion component is defined by the percentage of variance that it explains in the predictor or criterion data, respectively. We evaluate both the recovery of the relevant predictor and predictable criterion components and the recovery of all the simulated predictor and criterion components. We expect that when it comes to recovery of all the components, higher α values will do better, while if we are interested in relevant predictor and predictable criterion components, lower α values are going to yield better results. Moreover, we expect the merging issue that was described in Section "Defining the problem" to pop up in some conditions. We expect that, in those conditions, recovery of the relevant predictor and predictable criterion components will benefit from applying the proposed heuristic.

Data generation

All constructed datasets contained 100 observations of 24 predictor variables and 24 criterion variables. These data were generated according to $\mathbf{X} = \mathbf{T_X P_X}' + \mathbf{E_X}$ and $\mathbf{Y} = \mathbf{T_Y P_Y}' + \mathbf{E_Y}$, while taking into account the predictive relation between $\mathbf{T_X}$ and $\mathbf{T_Y}$ $\mathbf{T_Y} = \mathbf{T_X B}' + \mathbf{E_{T_X}}$. The number of criterion components R_Y was set to two throughout this study, while the number of predictor components R_X equaled either two or three. We will now describe how all underlying matrices were simulated.

The elements of the component matrices $\mathbf{T_X}$ and $\mathbf{T_Y}$ were randomly drawn from a standard normal distribution, taking into account the relevance and predictability of the predictor and criterion components. This was done as follows: the component score matrices $\mathbf{T_X}$ and $\mathbf{T_Y}$ were sampled together from a multivariate normal distribution. The diagonal entries of the (R_X + R_Y) by (R_X + R_Y) covariance matrix of this multivariate distribution were set to one, implying that this matrix is a correlation matrix, and the off-diagonal entries equal zero if they pertain to pairs of predictor components or to pairs of criterion components. The off-diagonal entries that refer to a combination of a predictor component and a criterion component are set to the values of the regression weight matrix $\mathbf{B}$, which imposes the predictive relation $\mathbf{T_Y} = \mathbf{T_X B}' + \mathbf{E_{T_Y}}$ without having to generate $\mathbf{E_{T_Y}}$. Afterwards, in order to discard deviations from orthogonality and unit variance restrictions that are due to sampling, $\mathbf{T_X}$ and $\mathbf{T_Y}$ were orthogonalized and standardized.

The loading matrices $\mathbf{P_X}$ and $\mathbf{P_Y}$ were created using only ones and zeros, with each variable (i.e., predictor or criterion) having only one loading of 1 and with a different number of 1s per component (see Table 1 for an example). The loading matrices thus have simple structure. Loading matrices with two components had, respectively, three and 21 highly loading variables; loading matrices with three components had, respectively, three, six, and 15 highly loading variables. Hence, the first component was always weak and the last component always strong, referring to the amount of variance in the variable scores accounted for by those components; the component that consists of six variables can also be considered weak.

The regression weight matrix $\mathbf{B}$ ($R_Y \times R_X$) was constructed according to Table 3 with the size of the non-zero regression weight(s) determined according to factor 2. This manipulation implies that in some conditions, the non-zero regression weights amount to .995, indicating that the predictable criterion components can be almost perfectly predicted based on the relevant predictor components. In other conditions, the regression weights amount to .86, implying that only part of the scores on the predictable criterion components can be captured by means of the relevant predictor components. We also manipulated the location of the non-zero regression weights (factor 3), yielding four types of regression weight matrices. In types 1 and 3, the relevant predictor components are weak, and the irrelevant ones are strong. In types 2 and 4, the situation is reversed for $R_X = 2$; for $R_X = 3$, one weak predictor component is relevant, as well as the strong one. Types 1 and 3 differ with respect to the strength of the predictable criterion component: In type 1 the predictable criterion component is weak, while in type 3 it is strong. The same holds for type 2 and type 4.

The error matrices $\mathbf{E_X}$ and $\mathbf{E_Y}$ were drawn from a standard normal distribution as well, and were also standardized and orthogonalized. The error matrices were rescaled to obtain 25% of error variance (i.e., the predictor and criterion components explain 75% of the predictor and criterion variance, respectively).

Next, $\mathbf{X}$ and $\mathbf{Y}$ were created by multiplying $\mathbf{T_X}$ and $\mathbf{T_Y}$ with $\mathbf{P_X}$ and $\mathbf{P_Y}$, respectively, and by adding $\mathbf{E_X}$ and $\mathbf{E_Y}$. Finally, $\mathbf{X}$ and $\mathbf{Y}$ were normalized so that each predictor and criterion had a sum of squares of one.

Summarizing the above setup, the following data characteristics were manipulated in a full factorial design:

1. The number of predictor components: 2, 3
2. The size of the non-zero regression weights: .995, .86
3. The strength of the relevant predictor and predictable criterion components: four types of regression weight matrices for each level of factor 1 (see Table 3)

For each of the $2 \times 2 \times 4 = 16$ cells of the factorial design, 25 datasets were constructed, yielding 400 datasets in total.

Table 3 Four types of regression weight matrices **B** (factor 3), with either 2 or 3 predictor components (factor 1), and with "x" depicting a regression weight of .995 or .86 (factor 2)

R_X	Strength of criterion component	Type 1		Type 2		Type 3		Type 4	
		Weak Pred. Comp.	Strong Pred. Comp.	Weak Pred. Comp.	Strong Pred. Comp.	Weak Pred. Comp.	Strong Pred. Comp.	Weak Pred. Comp.	Strong Pred. Comp.
2	Weak	x	0	0	x	0	0	0	0
	Strong	0	0	0	0	x	0	0	x
3	Weak	x x	0	0 x	x	0 0	0	0 0	0
	Strong	0 0	0	0 0	0	x x	0	0 x	x

Data analysis

To investigate the recovery of the components and how it depends on α, we first analyzed each data set six times. Specifically, the R_X and R_Y values were set to the number of relevant predictor and the predictable criterion components, on the one hand, and to all the simulated predictor and criterion components, on the other hand. Moreover, in each of both settings, we used three α values: . 1[3], . 5, and .9 Note that when we talk about the numbers of relevant predictor components and predictable criterion components, we refer to the number of components which are associated with at least one non-zero regression weight according to the simulation settings in Table 3.

Next, we applied the model selection heuristic to all datasets. Here, we let R_X and R_Y vary from one to five, yielding 25 solutions. We set the α value to .05 (see Section "Model selection"), as we wanted to retrieve the relevant predictor and predictable criterion components regardless of their strength.

To avoid obtaining a local minimum, we always applied a multi-start procedure with 25 starts[4], and set $\varepsilon = 10^{-6}$ when specifying the convergence criterion. For each solution, we varimax-rotated both the predictor loadings and the criterion loadings, and compensated for these rotations in the component scores and the regression weights (see Section "Model"). Afterwards, we applied the heuristic described in Sect. 2.5.4 to the obtained 25 solutions, to select the relevant predictor and predictable criterion components.

Results

Recovery of the number of relevant predictor and the number of predictable criterion components as well as recovery of all the simulated components

PCovR2 In order to investigate the recovery of the relevant predictor and predictable criterion components, we computed the Tucker congruence value (Korth & Tucker, 1975) between the estimated component loadings by the PCoVR2 model with R_X and R_Y fixed to the number of simulated relevant predictor and predictable criterion components (i.e., $R_X = 2$ or 3 and $R_Y = 2$), respectively, and the true relevant predictor and predictable criterion components (i.e., $R_X = 2$ and $R_Y = 1$). When the number of components exceeded one, we calculated the congruence per component and averaged the obtained values.

Table 4 shows the mean recovery when we extract the numbers of relevant predictor components and predictable criterion components (i.e., taking $R_X = 1$ or 2 based on the level of factor 1, and $R_Y = 1$). Table 5 shows the recovery results when extracting all the simulated components (i.e., setting $R_X = 2$ or 3 based on the level of factor 1 and $R_Y = 2$) across multiple α values. As expected, when it comes to recovering the numbers of all simulated predictor and criterion components (hence not just the relevant ones), the higher the α value, the better. In fact, using an α value of .9, always yields almost perfect congruence values of .99, whereas these values range between .67 and .92 for an α value of .1.

This is not the case when only extracting the numbers of relevant predictor and predictable criterion components, however. For example, in the first row of Table 4, the recovery of the criterion components goes down from .97 to .01 when we increase α from .1 to .9. In this simulation cell, the data sets were simulated with weak but predictable and strong but non-predictable criterion components. When we model these data sets with one criterion component, the algorithm extracts the predictable criterion component if α = .1, while it selects the strong but non-predictable criterion component with higher α values. Moreover, there are some simulation cells for which,

---

[3] Note that next to. 1, we also examined other low alpha values: . 04 and . 05. The results were very similar, thus we only report the results for. 1.

[4] In a pilot study, we performed 400 starts, but the first 25 starts always contained at least one start that yielded the best encountered loss function value across the 400 starts. Therefore, we reduced computational effort by decreasing the number of random starts to 25.

Table 4 Mean recovery of the relevant predictor components and predictable criterion components with PLS, PCR2, and PCovR2. For PCovR2, we used three different α values. In all these analyses, the number of extracted predictor and criterion components is set to the number of relevant predictor components and predictable criterion components. The "X" denotes the cases for which the number of relevant predictor components differs from the number of predictable criterion components, implying that PLS cannot be applied

Conditions			PLS		PCovR2						PCR2	
					α = .1		α =.5		α = .9			
NofPred	RegWght	Type	P_X	P_Y	P_X	P_Y	P_X	P_Y	P_X	P_Y	P_X	P_Y
2	.995	1	.80	.98	.97	.97	.01	.01	.01	.01	.01	.01
3	.995	1	X	X	.41	.98	.43	.86	.50	.01	.50	.01
2	.86	1	.79	.97	.94	.95	.01	.01	.01	.01	.01	.01
3	.86	1	X	X	.41	.97	.43	.67	.50	.01	.50	.01
2	.995	2	1.00	.99	1.00	.98	1.00	.96	1.00	.01	1.00	.01
3	.995	2	X	X	1.00	.98	1.00	.91	1.00	.01	1.00	.01
2	.86	2	1.00	.98	1.00	.97	1.00	.85	1.00	.01	1.00	.01
3	.86	2	X	X	.99	.97	.99	.81	1.00	.01	1.00	.01
2	.995	3	.95	1.00	.98	1.00	.97	1.00	.01	1.00	.01	1.00
3	.995	3	X	X	.42	1.00	.43	1.00	.49	1.00	.51	1.00
2	.86	3	.93	1.00	.97	1.00	.87	1.00	.01	1.00	.01	1.00
3	.86	3	X	X	.42	1.00	.43	1.00	.49	1.00	.50	1.00
2	.995	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	.995	4	X	X	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	.86	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	.86	4	X	X	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 5 Mean recovery of all the simulated components with PLS, PCR2 and PCovR2. For PCovR2 we used three different α values. In all these analyses the number of extracted predictor components and criterion components are set to the numbers of simulated ones (see factor 1)

Conditions			PLS		PCovR2						PCR2	
					α = .1		α =.5		α = .9			
NofPred	RegWght	Type	P_X	P_Y	P_X	P_Y	P_X	P_Y	P_X	P_Y	P_X	$\mathbf{P_Y}$
2	.995	1	.89	.63	.89	.88	.99	.99	.99	.99	.99	.99
3	.995	1	X	X	.75	.90	.94	.99	.99	.99	.99	.99
2	.86	1	.89	.66	.86	.83	.99	.99	.99	.99	.99	.99
3	.86	1	X	X	.80	.89	.94	.99	.99	.99	.99	.99
2	.995	2	.87	.99	.72	.88	.79	.99	.99	.99	.99	.99
3	.995	2	X	X	.82	.89	.89	.99	.99	.99	.99	.99
2	.86	2	.85	.98	.70	.91	.79	.99	.99	.99	.99	.99
3	.86	2	X	X	.79	.87	.85	.99	.99	.99	.99	.99
2	.995	3	.97	.56	.90	.71	.99	.77	.99	.99	.99	.99
3	.995	3	X	X	.83	.71	.94	.75	.99	.99	.99	.99
2	.86	3	.96	.57	.92	.75	.99	.79	.99	.99	.99	.99
3	.86	3	X	X	.79	.72	.94	.80	.99	.99	.99	.99
2	.995	4	.98	.86	.75	.72	.86	.81	.99	.99	.99	.99
3	.995	4	X	X	.83	.67	.90	.76	.99	.99	.99	.99
2	.86	4	.92	.66	.69	.73	.77	.83	.99	.99	.99	.99
3	.86	4	X	X	.81	.71	.87	.82	.99	.99	.99	.99

The 'X' denotes the cases where the simulated number of predictor components differs from the number of simulated criterion components, implying that PLS cannot be applied

Table 6 The number of datasets for which the correct number of relevant predictor and predictable criterion components were found, as a function of the number of predictor components, the size of the non-zero regression weight (RegWght), and type of regression weights matrix

Conditions			Frequency of finding the correct number of components	Mean recovery of the loadings	
R_X	RegWght	Type		P_X	P_Y
2	.995	1	25	.97	.97
3	.995	1	25	.99	.98
2	.86	1	18	.95 | .98	.96 | .94
3	.86	1	25	.99	.97
2	.995	2	21	1 | .99	.99| .98
3	.995	2	25	1	.98
2	.86	2	18	1 | 1	.97 | .96
3	.86	2	23	.99 | 1	.97 | .98
2	.995	3	25	.98	1
3	.995	3	25	.99	1
2	.86	3	19	.97 | .98	1 | 1
3	.86	3	24	.99 | .99	1 | 1
2	.995	4	20	1 | .99	1 | 1
3	.995	4	25	1	1
2	.86	4	22	1 | .99	1 | 1
3	.86	4	25	1	1

The last two columns present the mean recovery of the relevant predictor and predictable criterion loadings, in case of correct and incorrect model selection (separated by a vertical line)

regardless of α, the recovery is never good enough. For example, in the second row of the table, the recovery of the predictor loadings is never higher than .5. This is due to the merging issue, demonstrated in Section "Model selection", which occurs when the number of relevant predictor components differs from the number of predictable criterion components. For such scenarios, the model selection heuristic, proposed in Section "Model selection", might yield better results.

PLS2 and PCR2 We also examined the performance of PLS2 and PCR2, by investigating how well these methods recovered the relevant predictor and predictable criterion components on the one hand and all the simulated predictor and criterion components on the other hand. As discussed earlier in this paper, PLS2 imposes an equality restriction on the number of predictor and criterion components. Therefore, we only applied PLS2 in the conditions where the number of simulated components equalled two. Like PCovR2, PCR2 does not restrict the numbers of predictor and criterion components to be the same. Hence, PCR2 could be applied in all conditions.

The results are presented in Tables 4 and 5. Like PCovR2, PLS2 seems to be affected by the strength of the relevant predictor and predictable criterion components, when

focusing on these components only. Indeed, as can be seen in Table 4, PLS2 performs worst, when the relevant predictor and predictable criterion components are both weak (i.e., with type 1 regression weights, see Table 3). Moreover, the recovery of all simulated components (Table 5) is worse than that of the relevant and predictable ones (Table 4). The latter is not surprising as PLS2 is designed to find predictor and criterion components that are related to one another. If we compare the PLS2 results to the PCovR2 ones, we see in Table 4 that with $\alpha = .1$, PCovR2 performs sligtly better than PLS2. Table 5 shows that when targeting all the simulated components, PCovR2 analyses with higher α values recover the components better than PLS2. The PCR2 results in both tables are very similar to the ones of PCovR2 with $\alpha = .9$. This makes sense, since PCR2 equals PCovR2 with $\alpha = 1$. Overall, this comparison of PLS2, PCR2, and PCovR2 under two different scenarios (focusing on the relevant predictor and predictable criterion components only or on all the components) demonstrates the benefits of the PCovR2 weighting parameter.

Model selection

First, we checked whether the model selection heuristic yielded the correct *number* of relevant predictor components, which amounts to one when $R_X = 2$ and to two when $R_X = 3$, and the number of predictable criterion components, which always equaled one. Table 6 shows the number of datasets for which the model selection heuristic found the correct number of relevant predictor and predictable criterion components. Specifically, the heuristic yielded a correct result for 365 out of the 400 datasets. Performance is mostly affected by the number of predictor components and the size of the non-zero regression weights, respectively, as 32 and 26 of the 35 mistakes occurred when the number of predictor components equaled two and when the non-zero regression weight amounted to .86. The effect of the strength of the components is less outspoken. Further analyzing the 35 mistakes, the first step of the model selection rule yielded an incorrect result, which is a too high number of predictable criterion components, for 24 of them. This implies that for the remaining 16 cases, model selection went wrong when assessing the number of relevant predictor components. Again, all mistakes pertained to selecting too many components.

Next, we again investigated the recovery of the relevant and predictable components by computing the Tucker congruence value between the estimated component loadings and the true relevant predictor and predictable criterion ones. When too many components were extracted, we computed the congruence value for the components that resembled the true ones the most. The last two columns of Table 5 show the recovery of the predictor and criterion loadings. Per cell of the design, the number before the vertical line indicates mean recovery for the datasets for which the model selection heuristic returned

Table 7 Varimax rotated loading matrices and regression weights of three interesting PCovR2 solutions; loadings and regression weights above .40 are shown in bold

Father variables	$R_X=1, R_Y=1, \alpha=.05$	$R_X=2, R_Y=1, \alpha=.05$		$R_X=2, R_Y=2, \alpha=.05$	
	FathComp1	FathComp1	FathComp2	FathComp2	FathComp1
Relationship with partners	-29	.22	**.51**	.11	**.60**
Argue with partners	.05	**-.07**	**.81**	-.05	.36
Child's bright future	**.81**	**.82**	.03	**.86**	.02
Activities with the child	.02	.01	.07	-.01	.09
Feelings about parenting	.36	.32	.35	.35	.12
Communication with the child	.26	.24	.11	.02	**.74**
Argue (aggressively) with the child	.45	.36	**.64**	.38	.30
Confidence about oneself	.15	.05	**.71**	.14	.12
Child variables	ChiComp1	ChiComp1		ChiComp1	ChiComp2
Self-confidence/esteem	**.72**	**.71**		**.47**	**.65**
Academic performance	**.79**	**.79**		**.87**	.07
Social life and extracurricular activities	.07	.07		.02	.16
Importance of friendship	.23	.23		-.04	**.59**
Self-image	.13	.12		-.09	**.51**
Happiness	.26	.25		-.02	**.61**
Confidence about the future	**.50**	**.50**		.18	**.73**
Regression weights	FathComp1	FathComp1	FathComp2	FathComp1	FathComp2
ChiComp1	**.46**	**.46**	.05	**.44**	-.01
ChiComp2				.10	.33

the correct numbers of relevant predictor components and predictable criterion components. The number after the line corresponds to the datasets for which the heuristic returned an incorrect result. The high Tucker congruence values for the datasets, for which the model selection heuristic returned an incorrect result, show that even for these datasets, the relevant predictor and predictable criterion components were adequately recovered. The average Tucker congruence values in Table 5 are higher or equal to .95, which, according to Lorenzo-Seva & ten Berge (2006), indicates that the components compared can be considered as equal. The average Tucker congruence seems to be affected by the strength of the relevant and predictable components: In type 1, both the relevant predictor and predictable criterion components are weak, yielding worst recovery. In contrast, in type 4, the predictable criterion component and one of the relevant predictor components are strong, yielding about perfect recovery.

Conclusion

The simulation study showed that if one is interested in recovering all the predictor and criterion components, regardless if they are relevant or predictable, one best uses PCovR2 with a high α value or PCR2. In contrast, lower α values lead to a better recovery of the subset of relevant predictor and predictable criterion components. In case the number of relevant predictor components equals the number of predictable criterion components, PLS2 yielded similar results. However, when the latter does not hold, PLS2 cannot be applied, and naively applying PCovR2 with the true number of relevant predictor and predictable criterion components did not yield good enough results. The latter is due to the merging issue described in Section "Defining the problem". For such cases, the model selection heuristic, proposed in Section "A two-step model selection heuristic to retrieve all relevant and predictable components", was useful. Indeed, using this heuristic, the mean recovery of the relevant predictor and predictable criterion loadings outperformed that of analyses with the number of relevant predictor and predictable criterion components.

Illustrative application

In this section, we analyze survey data of 192 fathers and their child, taken from "The 500 Family Study" (Schneider & Linda, 2008). For each family, the father filled in eight questionnaires, regarding their relationships, feelings, etc., and the child seven questionnaires (see Table 7). Like Gu and Van Deun (2018), we calculated a single score for each questionnaire, by summing the scores on the questions within the questionnaire. We thus obtained an **X** matrix with dimensions 192 × 8 pertaining to the father data, and a **Y** matrix with dimensions 192 × 7 concerning the child data. This means

R_Y	R_X	sum of squared prediction errors of each of the extracted criterion components (cc)				
		cc1	cc2	cc3	cc4	cc5
1	1	**.79**				
	2	**.79**				
	3	**.79**				
	4	**.79**				
	5	**.79**				
2	1	**.79**	1			
	2	.88	.81			
	3	.88	.81			
	4	.88	.81			
	5	.81	.88			
3	1	1	**.79**	1		
	2	.89	.98	.82		
	3	.81	.91	.90		
	4	.89	.91	.82		
	5	.89	.91	.82		
4	1	.83	1	.96	1	
	2	.89	1	.83	.98	
	3	.95	.93	.90	.83	
	4	.84	.89	.95	.90	
	5	.84	.90	.94	.89	
5	1	1	1	.89	.93	.97
	2	.93	.98	.92	1	.86
	3	.86	.97	.99	.91	.89
	4	.91	.96	.98	.85	.89
	5	.97	.85	.88	.96	.91

a)

b)
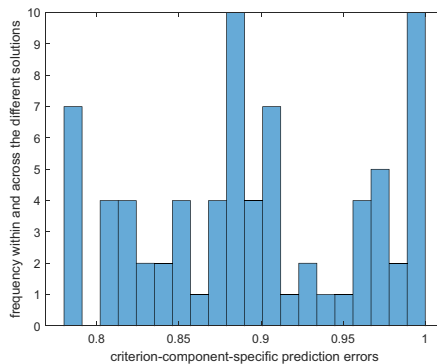
R_X	Regression weights					# Relev. pc
1	**.46**					1
2	**.46**	.05				1
3	**.46**	.05	.04			1
4	**.44**	.09	.04	.09		1
5	**.43**	-.07	.02	-.08	-.12	1

c)

Fig. 3 (**a**) The sum of squared prediction errors for each criterion component across 25 PCovR2 models, estimated with R_X and R_Y varying from one up to five for the illustrative application. cc1 up to cc5 indicates the first up to the fifth criterion component within the considered solution. Predictable criterion components are marked in bold. (**b**) Histogram presenting the distribution of all the sum of squared prediction errors reported in Fig. 3a. **c**) The regression weights that are obtained when R_Y is fixed to 1 and R_X varies from 1 to 5

that, in our analyses, we examine the extent to which the child questionnaires can be predicted by the father ones. After preprocessing (centering the data and setting the sum of squares per questionnaire to 1), the data were analyzed with PCovR2.

In order to decide on the number of relevant predictor and predictable criterion components, we applied the model selection rule that we described in Section "Model selection". First, letting R_X and R_Y vary from 1 to 5, we fitted 25 PCovR2 models. Again, we fixed the α value to .05, because the simulation study confirmed that low α values are preferable when one wants to focus on the regression problem. To decide on the number of predictable criterion components, we considered the criterion component-specific prediction errors in panel a of Fig. 3, and the associated histogram (panel b). Based on the grouping structure in the histogram, we put the threshold T_{Ry} for distinguishing predictable and unpredictable components to .79. Given that each considered solution contains no more than one predictable criterion component, we conclude that this dataset contains one predictable criterion component and set R_Y to one.

Panel c of Fig. 3 shows the regression weights of the PCovR2 solutions with one criterion component and the number of predictor components varying from one to five. Based on the formula $T_{Rx} = \sqrt{(1-T_{Ry})/2}$, we compute the threshold for the regression weights T_{Rx} as $\sqrt{(1-.79)/2} = .23$. Comparing the regression weights to this threshold, we see that each considered solution contains only one relevant predictor component. Thus, we conclude that this dataset has one relevant predictor and one predictable criterion component and retain the $R_X = 1$ and $R_Y = 1$ solution.

However, attentive readers might note that the heuristic includes some arbitrary decisions. For example, panel b of Fig. 3 is not very clear, and if we would have used a different number of bins, we might have ended up with a less strict threshold and possibly a higher number of predictable criterion components. Therefore, we supplement Fig. 3 with Fig. 4. This figure plots how well **X** and **Y** are reconstructed and how well $\mathbf{T_Y}$ is predicted, as a function of the α value used and the number of extracted
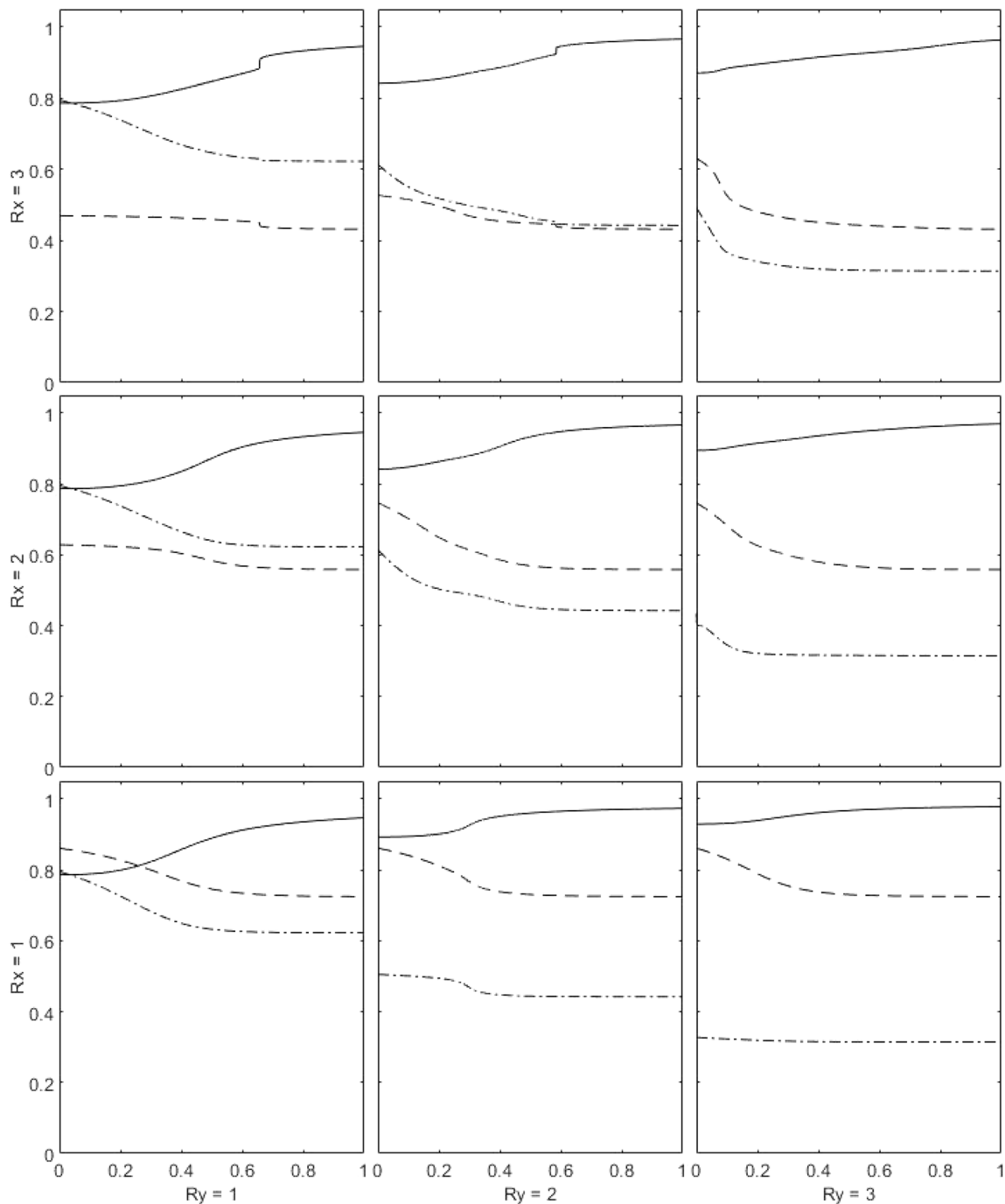
Fig. 4 Values of the reconstruction (dashed) and prediction (solid) parts of the PCovR2 loss function (Y-*axis) for the illustrative application as a function of the value of the weighting parameter* α *(X-axis). The panels pertain to 1 to 3 predictor components (rows) and 1 to 3 criterion* components (columns). The dashed line '–', dash-dot line '-.' and solid line '-' denote the reconstruction error of X and Y, and the prediction error, respectively

predictor and criterion components. We see that next to ($R_X = 1$, $R_Y = 1$, $\alpha = .05$), the ($R_X = 2$, $R_Y = 1$, $\alpha = .05$) as well as ($R_X = 2$, $R_Y = 2$, $\alpha = .05$) might also be interesting solutions. Regarding ($R_X = 2$, $R_Y = 1$), the reconstruction of **X** clearly benefits from the added predictor component. Similarly, for ($R_X = 2$, $R_Y = 2$), adding one more criterion component also improves the reconstruction of **Y,** while the prediction curve does not worsen as

much. Note that increasing the number of criterion components will generally lower the average predictability of those components, because the first criterion component extracted will always be the most predictable one.

Table 6 shows these three solutions. The father's first component is highly similar across the solutions and can be labeled as *optimism about their child's future*. The second father

component differs across the solutions, however, and is a strong but irrelevant component in the ($R_X = 2$, $R_Y = 1$, $\alpha = .05$) solution, and a weaker but relevant one in the ($R_X = 2$, $R_Y = 2$, $\alpha = .05$) solution, where it seems to be related to relationship and communication quality. The first child component is also very stable across the solutions and pertains to *self-esteem and academic performance and confidence about their future.* The regression weights show that regardless of which solution we examine, the father's *optimism about their child's future* is moderately related to the child's *self-esteem and academic performance and confidence about their future.* The second criterion component in the ($R_X = 2$, $R_Y = 2$, $\alpha = .05$) solution encompasses most non-performance-related aspects. This component is moderately related to the relationship and communication quality component of the father.

Discussion

In this paper, we proposed PCovR2, an extension of PCovR, in which both the predictor variables and the criterion variables are reduced to components. Simultaneously, the criterion components are regressed on the predictor components. The PCovR2 weighting parameter can be flexibly used to focus on the reconstruction of the predictors and the criteria, or on filtering out the "relevant" predictor components and "predictable" criterion components. Model selection is challenging, however, especially in the latter case: on the one hand, we want to find predictor components that can predict the criterion components and criterion components that can be predicted by those predictor components. On the other hand, we want these components to be an accurate representation of the underlying components. To aid researchers in this decision, we proposed inspecting a matrix plot of outcomes of the different loss parts for different model choices, as well as a model selection heuristic. The matrix plot of outcomes also yields further information on the range of suitable α values. In case one is only interested in the relevant predictor and predicable criterion components, one should select α values for which the prediction part of the loss is low. In case one wants to extract all the components, the reconstruction parts should be observed. Inspection of such plots shows that the range of α values that are low enough in the first case or high enough in the second case might vary from dataset to dataset. Yet, as a rule of thumb, .05 and .9 will often work well as low and high α values.

We performed a simulation study to investigate the performance of PCovR2 and the proposed model selection heuristic for filtering out the relevant predictor components and the predictable criterion components. The results indeed show that high α values yield good performance, when one wants to extract all predictor and criterion components, whereas low α values have to be preferred when one aims to find the relevant predictor and predictable criterion components. Applying the model selection heuristic further improved the results in the conditions where the number of relevant predictor components exceeded the number of predictable criterion components. It should be noted that the simulated conditions were not easy, as in some conditions, the relevant predictor and predictable criterion components were very weak. Note that even though the latter conditions might be considered extreme, we included them to evaluate and showcase the benefits of PCovR2.

Comparing PCovR2 with competitor methods, we have shown that when focusing on the subset of relevant predictor components and predictable criterion components, PLS2 also predicts the criterion variables well, but does not adequately recover the underlying predictor components, if more than one predictor component is needed to reconstruct a criterion component. In case the number of relevant predictor components equals the number of predictable criterion components, PLS2 and PCovR2 with a low α value had a very similar and mostly good recovery performance, however. Performance of PCR2 is worse in these analyses because it focuses on the strength of components rather than on their relevance/predictability. Yet, when we focused on retrieving all simulated components, PCR2 and PCovR2 with a high α value yield equally excellent results, whereas PLS2 did worse.

PCovR2 has some disadvantages, of which an important one pertains to model selection. Using the matrix plot involves many computations, and it may not be all that easy to compare the subplots, and find a useful compromise across the loss parts. The proposed model selection heuristic might still seem quite complicated, as it involves two steps to find the correct number of relevant predictor and predictable criterion components, and it is based on rather arbitrary thresholds. Another disadvantage is that one cannot simply fix R_X and R_Y to the numbers of relevant predictor and predictable criterion components. Indeed, as we showed for our toy data set, one often needs to extract additional components to model the relevant and predictable components correctly.

Future research might further investigate model selection. In pilot simulations, we tried to tune the α value through cross-validation, for instance. The tricky part is that it is not so obvious which criterion should be optimized and evaluated in the cross-validation. In the pilot simulations, we optimized the predictability of the criterion scores based on the predictor scores. Yet, this criterion does not explicitly take the recovery of the predictor and criterion structure into account. An interesting alternative might be to use stability selection (Meinshausen & Bühlmann, 2010), although, like cross-validation, this procedure is computationally intensive.

PCovR2 holds promise for dyadic studies, as illustrated in the previous section. This research area has seen a significant increase in the last decade and many researchers are involved in the development of dedicated methods (e.g., Bodner et al., 2018; Butler, 2011; Gates & Liu, 2016). Often the major questions steering the analysis of such data focus on the inter-relations between both dyadic partners, which is precisely

what PCovR2 aims to unravel. Since many dyadic studies repeatedly measure the same dyad across time, it would be interesting to study how PCovR2 can be extended to take serial dependencies between observations into account. The recent surge of psychometric research on vector autoregressive modeling can be inspiring in this regard (e.g., Bulteel, Tuerlinckx, Brose, & Ceulemans, 2018; Epskamp et al., 2018; Hamaker et al., 2018). Moreover, so far, PCovR2 is an asymmetric method, implying that the analysis focuses on the effect that one partner has on the other and not vice versa. Further developments are needed that extend the regression weights and the PCovR2 loss function and algorithm, so as to take bi-directional influences into account.

To conclude, we presented PCovR2, a new method for multivariate regression problems with a relatively high number of predictors and criteria, and compared it to PLS2 and PCR2. The proposed PCovR2 algorithm yields comparable results to PCR2 when trying to retrieve all components in the data and comparable results to PLS2 when focusing on finding back the relevant predictor components and predictable criterion components and in case the numbers of the latter components are equal. Yet, unlike PLS2, PCovR2 can also handle settings in which these numbers are not the same, although model selection is challenging.

Appendix

The extended derivation of L_{Step2} (Eq. 11)

Given current estimates of $\mathbf{T_X}$, $\mathbf{P_X}$, and $\mathbf{B}$, the non-constant part (i.e., we will drop the constant part throughout) of the PCovR2 loss function can be re-expressed as

$$L_{Step2} = \alpha \frac{\|\mathbf{Y}-\mathbf{T_Y}\mathbf{P_Y}'\|^2}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{T_X}\mathbf{B}'\|^2$$

Let us define $\mathbf{U}=\mathbf{T_X}\mathbf{B}'$ and replace $\mathbf{P_Y}'$ with $\mathbf{T_Y}'\mathbf{Y}$. The latter results from $\mathbf{Y}=\mathbf{T_Y}\mathbf{P_Y}'$. If we left-multiply both sides of the equation with $\mathbf{T_Y}'$, it follows that $\mathbf{T_Y}'\mathbf{Y}=\mathbf{T_Y}'\mathbf{T_Y}\mathbf{P_Y}'=\mathbf{P_Y}'$, due to the constraint $\mathbf{T_Y}'\mathbf{T_Y}=\mathbf{I}$.

Thus, L_{Step2} can be rewritten in the following way:

$$L_{Step2} = \alpha \frac{\|\mathbf{Y}-\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\|^2}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2 = \alpha \frac{tr\left[(\mathbf{Y}-\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y})'(\mathbf{Y}-\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y})\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2$$

$$= \alpha \frac{tr\left[(\mathbf{Y}'-\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}')'(\mathbf{Y}-\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y})\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2$$

$$= \alpha \frac{tr\left[\mathbf{Y}'\mathbf{Y}-\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}-\mathbf{T_Y}\mathbf{T_Y}\mathbf{Y}+\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2$$

$$= \alpha \frac{tr\left[\mathbf{Y}'\mathbf{Y}-\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}-\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}+\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2 = \alpha \frac{tr\left[\mathbf{Y}'\mathbf{Y}-\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2$$

$$= \alpha \frac{tr\left[\mathbf{Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} - \alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2 = \text{constant} - \alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\|\mathbf{T_Y}-\mathbf{U}\|^2$$

$$= -\alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}tr\left[(\mathbf{T_Y}-\mathbf{U})'(\mathbf{T_Y}-\mathbf{U})\right] = -\alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}tr[(\mathbf{T_Y}'-\mathbf{U}')(\mathbf{T_Y}-\mathbf{U})]$$

$$= -\alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}tr[\mathbf{T_Y}'\mathbf{T_Y}-\mathbf{T_Y}'\mathbf{U}-\mathbf{U}'\mathbf{T_Y}+\mathbf{U}'\mathbf{U}] - \alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2}$$

$$+ \frac{(1-\alpha)}{R_Y}(tr[I]-tr[\mathbf{T_Y}'\mathbf{U}]-tr[\mathbf{U}'\mathbf{T_Y}]+tr[\mathbf{U}'\mathbf{U}])$$

$$= -\alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} + \frac{(1-\alpha)}{R_Y}\left(\text{constant}-2tr\left[\mathbf{U}'^{\mathbf{T_Y}}\right]+\text{constant}\right) = -\alpha \frac{tr\left[\mathbf{Y}'\mathbf{T_Y}\mathbf{T_Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2} - \frac{2(1-\alpha)}{R_Y}tr[\mathbf{U}'\mathbf{T_Y}]$$

Let us define $\mathbf{V} = (\mathbf{Y}'\mathbf{Y})^{1/2}\mathbf{W_Y}$. This implies that $\mathbf{W_Y} = (\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{V}$ and $\mathbf{T_Y} = \mathbf{YW_Y} = \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{V}$. Now, L_{Step2} can be rewritten in the following way:

$$L_{Step}2 = -\alpha \frac{tr\left[\mathbf{Y}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{VV}'(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{Y}'\mathbf{Y}\right]}{\|\mathbf{Y}\|^2}$$
$$- \frac{2(1-\alpha)}{R_Y} tr\left[\mathbf{U}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{V}\right]$$

Here we will use the following properties of the trace function

$$tr[\mathbf{AB}] = tr[\mathbf{AB}]' = tr[\mathbf{B}'\mathbf{A}'] = tr[\mathbf{A}'\mathbf{B}']$$

$$L_{Step2} = -\alpha \frac{tr\left[\mathbf{V}'(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{Y}'\mathbf{YY}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{V}\right]}{\|\mathbf{Y}\|^2}$$
$$- \frac{2(1-\alpha)}{R_Y} tr\left[\mathbf{U}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{V}\right] = -\alpha \frac{tr[\mathbf{V}'\mathbf{Y}'\mathbf{YV}]}{\|\mathbf{Y}\|^2}$$
$$- \frac{2(1-\alpha)}{R_Y} tr\left[\mathbf{U}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{V}\right]$$

Let us define $A = -\frac{2(1-\alpha)}{R_Y}\mathbf{U}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1/2}$ and $C = \frac{-\alpha\mathbf{Y}'\mathbf{Y}}{\|\mathbf{Y}\|^2}$. Finally, we have:

$$L_{Step2} = tr[\mathbf{V}'\mathbf{CV}] + tr[\mathbf{AV}]$$

References

Acar, E., Kolda, T., & Dunlavy, D. (2011). All-at-once optimization for coupled matrix and tensor factorizations. *Mining and Learning with Graphs (MLG'11).* arXiv preprintarXiv:1105.3422

Acar, E., Papalexakis, E. E., Gürdeniz, G., Rasmussen, M. A., L. A., Nilsson, M., & Bro, R. (2014). Structure-revealing data fusion. *BMC bioinformatics, 15*(1), 239.

Aldrin, M. (2002). Reduced-rank regression, in: A.H. El-Shaarawi, W.W. Piegorsch (Eds.). In *Encyclopedia of Environmetrics* (pp. 1724–1728). Chichester: John Wiley & Sons.

Bodner, N., Kuppens, P., Allen, N. B., Sheeber, L. B., & Ceulemans, E. (2018). Affective family interactions and their associations with adolescent depression: A dynamic network approach. *Development and psychopathology, 30*(4), 1459-1473.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Monterey: Brooks/Cole Publishing.

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2018). Improved insight into and prediction of network dynamics by combining VAR and dimension reduction. *Multivariate behavioral research, 53*(6), 853-875.

Butler, E. A. (2011). Temporal interpersonal emotion systems: The "TIES" that form relationships. *Personality and Social Psychology Review, 15*(4), 367-393.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behav Res., 1*(2), 245-276.

Ceulemans, E., & Kiers, H. A. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British journal of mathematical and statistical psychology, 59*(1), 133-150.

Ceulemans, E., & Kiers, H. A. (2009). Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical and Statistical Psychology, 62*(3), 601-620.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.).* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

De Jong, S., & Kiers, H. A. (1992). Principal covariates regression: Part I. Theory. *Chemometrics and Intelligent Laboratory Systems, 14*(1-3), 155-164.

De Lathauwer, L., & Kofidis, E. (2017). Coupled matrix-tensor factorizations — The case of partially shared factors. *2017 51st Asilomar Conference on Signals, Systems, and Computers* (pp. 711-715). Pacific Grove, CA: IEEE.

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A. M., Riese, H., & Cramer, A. O. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science, 6*(3), 416-427.

Gates, K. M., & Liu, S. (2016). Methods for quantifying patterns of dynamic interactions in dyads. *Assessment, 23*(4), 459-471.

Gu, Z., & Van Deun, K. (2018). RegularizedSCA: Regularized simultaneous component analysis of multiblock data in R. *Behavior research methods, 51*(5), 2268-2289.

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research, 53*(6), 820-841.

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Applied Statistics, 31*(3), 300-303.

Kiers, H. A. (2002). Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. *Computational Statistics & Data Analysis, 41*(1), 157-170.

Kiers, H. A., & Smilde, A. K. (2007). A comparison of various methods for multivariate regression with highly collinear variables. *Statistical Methods and Applications, 16*(2), 193-228.

Korth, B., & Tucker, L. R. (1975). The distribution of chance congruence coefficients from simulated data. *Psychometrika, 40*(3), 361-372.

Lorenzo-Seva, U., & ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57-64.

Manne, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems, 2*, 187-197.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(4), 417-473.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*(11), 559-572.

Schneider, B., & Linda, W. (2008). The 500 family study [1998-2000: United states]. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research.* https://doi.org/10.3886/ICPSR04549.v1

Van Deun, K., Crompvoets, E. A., & Ceulemans, E. (2018). Obtaining insights from high-dimensional data: Sparse principal covariates regression. *BMC bioinformatics, 19*(1), 104.

Vervloet, M., Kiers, H. A., Van den Noortgate, W., & Ceulemans, E. (2015). PCovR: An R package for principal covariates regression. *Journal of Statistical Software, 65*(8), 1-14.

Vervloet, M., Van den Noortgate, W., & Ceulemans, E. (2018). Retrieving relevant factors with exploratory SEM and principal-

covariate regression: A comparison. *Behavior research methods,* *50*(4), 1430-1445.

Vervloet, M., Van Deun, K., Van den Noortgate, W., & Ceulemans, E. (2013). On the selection of the weighting parameter value in Principal Covariates Regression. *Chemometrics and Intelligent Laboratory Systems, 123*, 36-43.

Vervloet, M., Van Deun, K., Van den Noortgate, W., & Ceulemans, E. (2016). Model selection in principal covariates regression. *Chemometrics and Intelligent Laboratory Systems, 151*, 26-33.

Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull-based model selection method. *Behavior research methods, 45*(1), 1-15.

Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (2009). Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics & Data Analysis, 53*(4), 1086-1098.

Wilderjans, T. F., Ceulemans, E., Van Mechelen, I., & van den Berg, R. A. (2011). Simultaneous analysis of coupled data matrices subject to different amounts of noise. *British Journal of Mathematical and Statistical Psychology, 64*(2), 277-290.

Wold, S., Ruhe, A., Wold, H., & Dunn III, W. J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal of Statistics and Computations, 5*(3), 735-743.

The data and materials can be requested from the first author, and the study was not preregistered.