Chemometrics and Intelligent Laboratory Systems 151 (2016) 26–33

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Model selection in principal covariates regression

Marlies Vervloet [a,*], Katrijn Van Deun [a,b], Wim Van den Noortgate [a], Eva Ceulemans [a]

[a] KU Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium
[b] Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form 10 November 2015
Accepted 4 December 2015
Available online 12 December 2015

Keywords:
Principal covariates regression
Principal component regression
Partial least squares
Regression
Dimension reduction
Model selection

ABSTRACT

Dimension-reduction based regression methods reduce the predictors to a few components and predict the criterion using these components. When applying such methods, it is often not only important to achieve good prediction of the criterion, but also desirable to gain correct information about the underlying structure of the predictors (i.e., recovery of the underlying components). In contrast to PLS and PCR, PCovR explicitly aims at achieving both goals simultaneously. Moreover, the extent to which both aspects play a role in the construction of the components can be determined flexibly with a weighting parameter. This has as a downside that a dual model selection strategy is needed: selection of the number of components and selection of the weighting parameter value. Therefore, four model selection strategies are examined, and the optimality of the extracted components is studied in comparison to those resulting from PCR and PLS analyses. Based on the results of two simulation studies, we conclude that when the questions of a researcher match the optimality criteria specified in this paper, it is advised to use PCovR rather than PCR or PLS. Moreover, we recommend to use a weighting parameter that puts a lot of emphasis on the reconstruction of the predictor scores as well as to combine the results of a scree test and a cross-validation procedure when deciding on the number of components.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

When regressing a criterion variable $\mathbf{y_d}$ ($N \times 1$) on many predictors, of which the scores are stored in a matrix $\mathbf{X_d}$ ($N \times J$), it often happens that some of the predictors are highly collinear. This complicates interpretation, as each regression weight reveals only the additional contribution of the predictor on top of all other predictors, ignoring shared effects with other predictors. Moreover, (multi)collinearity leads to so-called bouncing betas, implying that small changes in the data can lead to large differences in obtained regression weights [1].

Both these problems are often addressed by using partial least squares (PLS[1] [2]) or principal components regression (PCR [3]). PLS and PCR reduce the predictors to a few linear combinations of them, called components, and regress the criterion on the components rather than the predictors. This simplifies matters, because far less regression weights have to be interpreted and shared effects of predictors that load on the same component will be expressed in these weights. Additionally, the obtained (unrotated) components will be uncorrelated, which handles multicollinearity and further solves shared effects issues.

However, as we will demonstrate in this paper, also PLS and PCR may yield unsatisfactory results, if the aim of the analysis is to fully grasp and gain insight into the mechanisms that link the criterion to the predictors, by retrieving the components that are truly underlying both the predictor and criterion scores. This is often the aim in chemistry, where the components in many cases reflect different chemical compounds that are being mixed [4]. To clarify the concept of true components, consider simulated data. In a first step, true data ($\mathbf{X_t}$ and $\mathbf{y_t}$) are generated in such a way that the scores in $\mathbf{X_t}$ and $\mathbf{y_t}$ are both a weighted sum of the same 'true' components. Next, random error is added to both $\mathbf{X_t}$ and $\mathbf{y_t}$, yielding the data $\mathbf{X_d}$ and $\mathbf{y_d}$ to be analyzed. PLS may fail to yield all true components, because adequate reduction of $\mathbf{X_d}$ is not directly optimized in the analysis. Rather, PLS focuses on minimizing prediction error, in that it looks for linear combinations of $\mathbf{X_d}$ that maximize the covariance with $\mathbf{y_d}$. PCR on the other hand, completely ignores $\mathbf{y_d}$, when extracting the components. Indeed, PCR starts with performing principal component analysis on $\mathbf{X_d}$, and, regresses, in a next step, $\mathbf{y_d}$ on the resulting components. This means that the components that are found may be irrelevant for predicting $\mathbf{y_d}$.

An interesting alternative to solve the regression problems that we started from and, at the same time, recover the true components, is principal covariates regression (PCovR [5]). PCovR simultaneously maximizes the amount of explained predictor variance (further called reduction) and minimizes the amount of prediction error, when extracting the components. To maximally enable users to tune the analysis to their interests, a weighting parameter α is specified that determines the extent to which reduction and prediction are emphasized

* Corresponding author.
E-mail addresses: marlies.vervloet@ppw.kuleuven.be (M. Vervloet), k.vandeun@uvt.nl (K. Van Deun), wim.vandennoortgate@kuleuven-kulak.be (W. Van den Noortgate), eva.ceulemans@ppw.kuleuven.be (E. Ceulemans).

[1] PLS = partial least squares
PCR = principal component regression
PCovR = principal covariates regression.

http://dx.doi.org/10.1016/j.chemolab.2015.12.004
0169-7439/© 2015 Elsevier B.V. All rights reserved.

(with a high α emphasizing reduction and a low α prediction). This flexibility, however, has as a downside that model selection is more intricate. Whereas with PLS and PCR only the number of components, R, needs to be selected, with PCovR, also the α value has to be tuned.

Regarding the tuning of α, we found in a previous simulation study that the α value in most cases only had a limited influence [6]. An exception to these findings were extreme conditions, such as conditions with large amounts of error on $\mathbf{y_d}$ and conditions with almost equal numbers of observations and variables. However, in the latter simulation study, we disentangled the selection of α and R, as R was not varied but put equal to the correct number (i.c., the number of true components simulated). In reality, however, the number of true components is unknown, and it can be expected that the specific α value used will influence the optimal R and vice versa. For instance, think of an example in which predictor data are generated based on two orthogonal components that are equally strong (i.e., explain the same amount of variance in $\mathbf{X_t}$). One of these components is relevant for predicting the criterion and one irrelevant (i.e., regression weight of zero). When two components are extracted, the α value used will hardly matter, as discussed above. When, however, only one component is selected, α has much more influence: When reduction of $\mathbf{X_d}$ is emphasized (i.e., large α), sampling fluctuations in the error will determine whether the relevant component is recovered or the irrelevant one. Emphasizing prediction (i.e., small α) will steer PCovR towards the relevant component.

Hence, although PCovR seems to be the ideal method to uncover the true components that explain the relation between the criterion and the predictors, 'optimal' model selection is not trivial, for two reasons. First, one has to be very clear about what optimal means. In other words, what are the characteristics of an optimal PCovR solution for a specific dataset? A first aim of this paper is to propose such an optimality definition. Second, PCovR model selection implies tuning R as well as α. Therefore, a second aim of this paper is to evaluate the performance of a number of model selection strategies. Moreover, we will investigate whether the results of different model selection approaches can be combined in a smart way to further enhance optimality of the retained components. Finally, we will compare the optimality of PCovR, PLS and PCR solutions for simulated data, to further demonstrate that the different methods depart from different optimality definitions.

The remainder of the paper is structured as follows: First, we will recapitulate PCovR analysis. In the next sections, we will put forward an optimality definition for PCovR solutions and discuss four possible model selection strategies. Finally, the design and the results of two simulation studies will be described, and the implications will be discussed.

2. PCovR

2.1. Model

In PCovR, the J predictors are reduced to R components, which are linear combinations of the predictor scores:

$$\mathbf{X_d} = \mathbf{T}\mathbf{P_X} + \mathbf{E_X} = \mathbf{X_d}\mathbf{W}\mathbf{P_X} + \mathbf{E_X}, \tag{1}$$

where $\mathbf{T}$ contains the scores of the N observations on the R components, and $\mathbf{P_X}$ is the loading matrix ($R \times J$) that indicates how strongly each predictor loads on each component. $\mathbf{E_X}$ holds the reconstruction error of the predictors and $\mathbf{W}$ is a $J \times R$ weight matrix. The components do not only summarize the predictors, but are also used to predict the criterion:

$$\mathbf{y_d} = \mathbf{T}\mathbf{p_y} + \mathbf{e_y}, \tag{2}$$

with $\mathbf{p_y}$ being the ($R \times 1$) regression weight vector, and $\mathbf{e_y}$ containing the residual $\mathbf{y_d}$ scores. Note that we assume that $\mathbf{X_d}$ and $\mathbf{y_d}$ are centered around zero [7], so that there is no need for an intercept in both Eqs. (1) and (2), and scaled (i.e., each column has a variance of one), to give each variable an equal weight in the analysis.

Each PCovR solution has rotational freedom. Specifically, premultiplying $\mathbf{P_X}$ and $\mathbf{p_y}$ with a rotation matrix, and postmultiplying $\mathbf{T}$ with the inverse of that rotation matrix, does not alter the reconstructed $\mathbf{X_d}$ scores or the predicted $\mathbf{y_d}$ scores. This rotational freedom implies that the loading matrix can be rotated towards simple structure—defined by Browne [8] as a matrix with only one non-zero loading per variable, and with more than R but fewer than J zero-loadings per component—to enhance the interpretability of the components. Rotation criteria that pursue simple structure are Varimax [9], yielding uncorrelated components, Quartimin [10], which allows for component correlations, and many others.

2.2. Loss function and parameter estimation

Previously, it was highlighted that PCovR performs a simultaneous reduction of $\mathbf{X_d}$ and prediction of $\mathbf{y_d}$. The weighting parameter α determines the degree of emphasis on either of these two objectives. The following loss function is minimized in PCovR analysis:

$$L = \alpha \frac{\|\mathbf{X_d} - \mathbf{T}\mathbf{P_X}\|^2}{\|\mathbf{X_d}\|^2} + (1-\alpha)\frac{\|\mathbf{y_d} - \mathbf{T}\mathbf{p_y}\|^2}{\|\mathbf{y_d}\|^2}, \tag{3}$$

with $\|\mathbf{A}\| = \sum_{i,j} a_{ij}^2$ being the Frobenius matrix norm of $\mathbf{A}$. The α values range between 0 and 1 (corresponding to PCR).

Given α and R, a closed form solution exists for estimating the PCovR parameters. Specifically, $\mathbf{T}$ is set equal to the first R eigenvectors of the matrix

$$\mathbf{G} = \alpha \frac{\mathbf{X_d}\mathbf{X_d}'}{\|\mathbf{X_d}\|^2} + (1-\alpha)\frac{\mathbf{H_X}\mathbf{y_d}\mathbf{y_d}'\mathbf{H_X}}{\|\mathbf{y_d}\|^2}, \tag{4}$$

with $\mathbf{H_X} = \mathbf{X_d}(\mathbf{X_d}'\mathbf{X_d})^{-1}\mathbf{X_d}'$ being the projection matrix[2] which projects $\mathbf{Y_d}$ on $\mathbf{X_d}$. Next, $\mathbf{P_X}$ and $\mathbf{p_y}$ can be computed:

$$\mathbf{P_X} = \mathbf{T}'\mathbf{X_d} \tag{5}$$

$$\mathbf{p_y} = \mathbf{T}'\mathbf{y_d} \tag{6}$$

[5]. Finally, $\mathbf{T}$, $\mathbf{P_X}$, and $\mathbf{p_y}$, are rescaled such that the columns of $\mathbf{T}$ have a variance of 1.

3. Optimality of PCovR solutions

In order to propose and compare model selection strategies for optimizing α and R, we should first clearly define what we conceive as an optimal PCovR solution. Here, an optimal PCovR solution is defined as one that sheds light on the mechanisms that link the criterion to the predictor variables and thus recovers the components that truly underlie both types of variable scores.

Before we can further elaborate this optimality concept, two distinctions have to be introduced. Firstly, components can differ in how relevant they are for predicting the true criterion $\mathbf{y_t}$, meaning that the absolute value of their regression weights can vary in size. Secondly, they can also differ in strength, with strength referring to the amount of variance in the true data $\mathbf{X_t}$ accounted for (see e.g. [11]). Note that a strong component does not necessarily explain a lot of variance in the observed data $\mathbf{X_d}$, however, because of the noise perturbation.

Given these distinctions, our optimality concept can be elucidated in terms of the following criteria. First and obviously, an optimal PCovR

[2] When $\mathbf{X_d}'\mathbf{X_d}$ is not of full rank, the inverse cannot be calculated. In PCovR, this is circumvented by post-multiplying the eigenvectors $\mathbf{X_d}'\mathbf{X_d}$ with the inverse of the (positive) eigenvalues and the transposed eigenvectors: $\mathbf{H_X} = \mathbf{X_d}\mathbf{K}\mathbf{S}^{-1}\mathbf{K}'\mathbf{X_d}'$, where $\mathbf{K}$ denotes the matrix containing the eigenvectors and $\mathbf{S}$ a diagonal matrix holding the eigenvalues.

solution should include all relevant components. Second, irrelevant but true components are not required but allowed, as including such components gives insight into which (parts of the) predictors are not helpful for explaining the criterion. Third, we explicitly do not want to find components that are not truly underlying the predictors. Such components can either be noise-driven or result from merging or splitting true components. In the following paragraphs, we will further elaborate these last three types of components.

3.1. Noise-driven

Noise-driven components are extremely weak and irrelevant, as they explain 0% of the variance of the true data. They are picked up, because they account for some of the error variance in $\mathbf{X_d}$ and $\mathbf{y_d}$.

3.2. Merging

A merged component is a component that is strongly congruent to the weighted sum of two true components. PCovR yields merged components because merging decreases model complexity while maintaining the variance accounted for in $\mathbf{y_t}$. Indeed, the following formula, where $\mathbf{t}_r$ denotes the scores on the r-th component and p_r the associated regression weight:

$$\hat{\mathbf{y}} = \mathbf{t}_1 p_1 + \mathbf{t}_2 p_2 = \left(\mathbf{t}_1 + \frac{p_2}{p_1}\mathbf{t}_2\right) p_1 = \mathbf{t_m} p_1, \tag{7}$$

shows that for each dataset components can be merged without altering the estimated criterion scores.[3] However, the variance accounted for in $\mathbf{X_t}$ (and thus also in $\mathbf{X_d}$) will decrease. Whether merging will occur or not will thus depend on (1) the degree to which the model selection strategy favors less complex models, (2) how much variance accounted for in $\mathbf{X_t}$ is emphasized in the analysis (i.e., the α value), and (3) the degree to which the variance accounted for in $\mathbf{X_d}$ is decreased by merging (which depends, among others, on the strength of the components and the error on the data).

3.3. Splitting

Splitting is the opposite phenomenon: one truly underlying component is split into two different components during estimation. Splitting may for instance be encountered if one or a few of the predictors that load on a specific true component contain much more error than the other predictors. In such cases, the first set of predictors may form a separate component, which has almost no additional predictive value and therefore receives a low regression weight. The latter set of predictors constitute a second component, which gets a stronger regression weight. As the noisy variables are discarded, this form of splitting results in better prediction of $\mathbf{y_d}$.

4. Model selection strategies

In this section we propose four model selection strategies (see Table 1) for selecting a PCovR solution, building on those introduced by Vervloet, Kiers, Van den Noortgate, and Ceulemans [7]. These strategies result from combining proposals on how to determine α and R, conditional upon one another.

4.1. Selecting the number of components given α

In dimension reduction based regression approaches, R is almost always determined by means of a scree test or using cross-validation. When performing a scree test [12], one inspects to which extent model fit can be increased by extracting additional components and selects the R value after which this fit increase levels off. To do this in the PCovR case, one calculates for several R values, the minimal loss function value (Eq. (3)). Next, the R value that maximizes the scree ratio sr is retained:

$$sr_R = \frac{L_{R-1} - L_R}{L_R - L_{R+1}}. \tag{8}$$

As noted by Wilderjans, Ceulemans, and Meers [13], the results of this scree test can be flawed in case the L difference for two subsequent R values is extremely small. Specifically, the sr_R value of the first of these two R values will become very high and therefore this R value will be selected. To deal with this flaw, Wilderjans, Ceulemans, and Meers [13] recommend to add an extra step, which is to exclude all models for which the fit is less than 1% better than the fit of a less complex model. For a similar approach, see Timmerman and Kiers [14]. Note that a scree test balances fit and complexity, and will therefore only select the "major components" (i.e., the components that explain the most variance [15]).

To determine R by means of K-fold cross-validation [16], the data are split in K roughly equal-sized parts, called folds. Next, one conducts, for every R value that is considered, K analyses, each time discarding a different data part and fitting the model to the other data parts. Given the estimates that result from the k-th analysis ($k = 1..K$), reconstructed $\mathbf{y}_k^{CV}$ scores are computed for the discarded observations:

$$\mathbf{y}_k^{CV} = \mathbf{X}_k \mathbf{W}_{\neg k} \mathbf{p}_{\mathbf{y},\neg k}, \tag{9}$$

where $\mathbf{X}_k$ contains the predictor scores of the k-th data part, and $\mathbf{W}_{\neg k}$ and $\mathbf{p}_{\mathbf{y},\neg k}$ indicate the $\mathbf{W}$ and $\mathbf{p_y}$ matrix of the analysis after deleting the k-th data part. Finally, concatenating all $\mathbf{y}_k^{CV}$scores in a $\mathbf{y}^{CV}$ vector, the cross-validation fit is computed:

$$Q^2_{\mathbf{y}} = 1 - \frac{\left\|\mathbf{y_d} - \mathbf{y}^{CV}\right\|^2}{\left\|\mathbf{y_d}\right\|^2}. \tag{10}$$

The R value which yields the model with the highest cross-validation fit will be selected. To decrease the influence of outliers in the data, this procedure can be repeated several times using a different random partitioning of the data into K parts and averaging the resulting cross-validation fits [17]. In that case, the R value which yields the model with the highest average cross-validation fit is selected. In contrast to the scree test, this cross-validation approach does not focus on the major components only, but will select the minimal number of components that is needed to reconstruct the criterion [15] as well as possible, although some of them might explain only little variance. Therefore, this approach might yield a rather high number of components. Therefore, to take estimation uncertainty into account, it is further recommended to use the one standard error rule[4] [18]. This rule implies that among all models that have a cross-validation fit that differs less than one standard error from the highest average cross-validation fit, the most parsimonious model is selected. Note that the standard error is computed by

Table 1
Overview of the four model selection strategies.

		Selection of α	
		Maximum likelihood	Cross-validation
Selection of R	Scree test	ML_SCR	SCR_ACV
	Cross-validation	ML_RCV	CV

[3] Of course the component matrices still need to be rescaled such that the columns of $\mathbf{T}$ have a variance of 1 again.

[4] Another strategy for finding more parsimonious models is to select a model with a mean squared cross-validation error smaller than the pre-specified significance limit of a chi square statistic [22]. However, when using a significance level of .05, this strategy led in our first simulation study in most cases to the selection of only 1 component; therefore, we did not consider this strategy further.

dividing the standard deviation of the $Q^2_{\mathbf{y}}$-values of the solution with the highest average cross-validation fit (across different random partitions of the data) by the square root of the number of partitions considered. Some authors argue to be even more conservative and use 1.5 times the standard error as cut-off [19]. Note that in case of preprocessed data (i.e., centered, scaled), it is important to re-preprocess the data inside every loop of the cross-validation procedure [20]. Otherwise, the discarded data parts can still bias the model estimates.

4.2. Selecting the weighting parameter value given R

To tune α, given R, most authors propose to apply cross-validation [5,21]. Note that a scree test cannot be used, since varying α does not affect the complexity of the model.

As discussed by Vervloet, Van Deun, Van den Noortgate, and Ceulemans [6], a maximum likelihood-based approach for specifying α can also be used. Assuming that the error on the predictor block ($\mathbf{E_X}$) and the error on the criterion block ($\mathbf{e_y}$) is drawn from normal distributions with mean 0 and variances of, respectively, ${\sigma_{\mathbf{E_X}}}^2$ and ${\sigma_{\mathbf{e_y}}}^2$, the α value that will maximize the likelihood of the data given the model is equal to:

$$\alpha_{ML} = \frac{\|\mathbf{X_d}\|^2}{\|\mathbf{X_d}\|^2 + \|\mathbf{y_d}\|^2 \frac{\sigma^2_{\mathbf{E_X}}}{\sigma^2_{\mathbf{e_y}}}}. \tag{11}$$

This optimal α value can be approximated by replacing the variances ${\sigma_{\mathbf{E_x}}}^2$ and ${\sigma_{\mathbf{e_y}}}^2$ by estimates. Although procedures for estimating these variances have been proposed by Vervloet, Kiers, Van den Noortgate, and Ceulemans [7], in this paper, we will assume these variances to be equal. Given that we only use standardized data in this paper, one criterion variable, and at least 24 predictors (see Section 5.1.1), α_{ML} will always approximate 1. This means that using the maximum-likelihood based approach, reducing $\mathbf{X}$ will be emphasized more in the analysis than predicting $\mathbf{y}$. However, it should be noted that analyses with α values that approximate 1 still may strongly outperform analyses with $\alpha = 1$ (i.e., PCR [6]).

4.3. Four model selection strategies

Crossing the maximum-likelihood and cross-validation approaches for specifying α, with the scree test and cross-validation procedure for determining R, we obtain the following four model selection strategies (see Table 1):

1) *Maximum likelihood specification of α—scree test for R* (ML_SCR). The first strategy is a sequential one. First, one computes the maximum likelihood α value, α_{ML}. Next, a scree test is conducted to decide on R, given α_{ML}.
2) *Scree test for R—cross-validation of α* (SCR_ACV). This strategy starts with the selection of R, which is done using the scree test procedure. Note that a starting value for α has to be specified as without it no PCovR computations can be performed. To this end, we will use α_{ML} in this paper. The final α value is selected afterwards, through a cross-validation procedure. In this step, only the R value that was selected in the scree test step is used.
3) *Maximum likelihood specification of α—cross-validation of R* (ML_RCV). This strategy also starts with computing the maximum likelihood α, α_{ML}. Next, cross-validation is applied to determine the optimal R value given α_{ML}.
4) *Simultaneous cross-validation of α and R* (CV). Finally, although computationally demanding, one can simultaneously tune α and R through cross-validation.

As scree-based strategies will aim for the strong components and cross-validation approaches for all components needed to reconstruct the criterion as well as possible, we can put forward some hypotheses about the number and nature (intact, merging, splitting, noise-driven) of the components that will be retrieved by the separate heuristics. Specifically, we expect that ML_SCR will recover the strong components, as it uses the scree test, and that the maximum likelihood approach will in the first place reduce the $\mathbf{X}$. Similarly, SCR_ACV will start with selecting the strong components, but, given the cross-validation step that follows, the nature of these components can change in order to capture $\mathbf{y}$ as well as possible.

ML_RCV puts a lot of emphasis on both $\mathbf{X}$ through the maximum likelihood approach and $\mathbf{y}$ through the cross-validation. Thus, we expect that a high number of components will be selected, in an attempt to recover all strong and all relevant components. Also the simultaneous cross-validation strategy CV is expected to lead to a high number of components, because cross-validation generally puts model fit above model parsimony (i.e., in that $\mathbf{y}$ should be reconstructed as well as possible).

5. Simulation studies

In order to evaluate these hypotheses on the differential performances of the four PCovR model selection strategies, with respect to the optimality criteria that were defined in Section 3, we conducted two simulation studies. In a first simulation study, the focus lies on the manipulation of the relevance and strength of the components, as our hypotheses indicate that this factor will probably strongly influence the differential performance of the strategies. Other data characteristics that are manipulated in the first simulation study, such as the error proportions, and the number of predictors, are included because they are likely to interact with the relevance and strength of the components [6], making model selection even more difficult. In general, the data of the first simulation study can be seen as relatively simple data, that were concordant with the assumptions made by the different model selection strategies (i.e., error samples drawn from a normal distribution). In the second simulation study, however, the data characteristics are more complex, but also more realistic from the perspective of empirical data. Specifically, the true components can be correlated, which makes it more difficult to disentangle the distinct components, and which creates extra collinearity of the predictors. Also conditions with skewed error matrices were added, as this violates the normality assumption underlying the maximum likelihood approach. Moreover, the data size was manipulated, to include conditions with more predictors than observations. In both simulation studies, a comparison will be made with the performance of PCR and PLS model selection strategies. As was stated in the introduction, PLS optimizes prediction, but may fail to reveal the true predictor structure, yielding merged, split and noise-driven components. In contrast PCR focuses only on reduction of $\mathbf{X}$ and thus might not recover weak but relevant components.

5.1. Simulation study 1

5.1.1. Data creation

In the simulation study, we fixed the number of observations N to 200, and the number of components R to 4. The following data characteristics were manipulated in a full factorial design:

1. The number of predictors J: 24, 48
2. The average proportion of error on $\mathbf{X_d}$: 5%, 25%, 45%
3. The proportion of error on $\mathbf{y_d}$: 5%, 25%, 45%
4. The regression weights $\mathbf{p}_{\mathbf{y}_t}$ of the relevant components: equal, small difference, large difference (see Table 3)
5. Strength of the components, expressed in terms of the proportion of variance in $\mathbf{X}_t$ accounted for: 4 and 46%, 8 and 42%, 13 and 38%, 17 and 33%, 21 and 29%, 25 and 25% (see Table 2)

Using this design the amount of available information about the true components was varied (factor 1, 2, 3), as well as their relevance and weakness (factor 4, 5), maximally challenging the model selection

Table 2
Number of high loading variables and percentage of variance in X_t explained per component (between parentheses) in different strength conditions.

		$R = 4$			
		1	*2*	*3*	*4*
$J = 24$	4% vs. 46%	1 (4%)	1 (4%)	11 (46%)	11 (46%)
	8% vs. 42%	2 (8%)	2 (8%)	10 (42%)	10 (42%)
	13% vs. 38%	3 (13%)	3 (13%)	9 (38%)	9 (38%)
	17% vs. 33%	4 (17%)	4 (17%)	8 (33%)	8 (33%)
	21% vs. 29%	5 (21%)	5 (21%)	7 (29%)	7 (29%)
	25% vs. 25%	6 (25%)	6 (25%)	6 (25%)	6 (25%)
$J = 48$	4% vs. 46%	2 (4%)	2 (4%)	22 (46%)	22 (46%)
	8% vs. 42%	4 (8%)	4 (8%)	20 (42%)	20 (42%)
	13% vs. 38%	6 (13%)	6 (13%)	18 (38%)	18 (38%)
	17% vs. 33%	8 (17%)	8 (17%)	16 (33%)	16 (33%)
	21% vs. 29%	10 (21%)	10 (21%)	14 (29%)	14 (29%)
	25% vs. 25%	12 (25%)	12 (25%)	12 (25%)	12 (25%)

strategies under study. For each cell of the factorial design, $2 \times 3 \times 3 \times 3 \times 6 = 324$ in total, 25 datasets were constructed as follows:

$$\mathbf{X_d} = \mathbf{T_t}\mathbf{P}_{\mathbf{X_t}} + \mathbf{E_X} \tag{12}$$

$$\mathbf{y_d} = \mathbf{T_t}\mathbf{p}_{\mathbf{y_t}} + \mathbf{e_y}, \tag{13}$$

where the component score matrix $\mathbf{T}_t$ was generated by sampling from a standard normal distribution. Subsequently, we standardized and orthogonalized the columns of $\mathbf{T}_t$. The loading matrix $\mathbf{P}_{\mathbf{X_t}}$ was created in such a way that every variable loads perfectly (i.e., loading of one) on one component only and has a zero-loading on all other components, but varying the number of variables per component according to Table 2 (factors 1 and 5). The regression weights $\mathbf{p}_{\mathbf{y_t}}$ were set to the values that are shown in Table 3 (factor 4). Combining these loading matrices and regression weights results in datasets that have four components that are, respectively, weak but relevant, weak and irrelevant, strong and relevant, and strong but irrelevant. The error matrices $\mathbf{E_x}$ and $\mathbf{e_y}$ were sampled from a standard normal distribution and their columns were standardized and orthogonalized. The correlations between the columns of $\mathbf{T}_t$ and $\mathbf{e_y}$ were also fixed to zero. Subsequently, the error matrices, the loadings, and the regression weights were rescaled to obtain predictor and criterion variables that contain the correct proportion of error variance (factors 2 and 3) when computed across all variables. Herewith, the exact amount of error variance differed among predictors, but corresponded on average with the values listed above (factor 2). Finally, the resulting matrices $\mathbf{X_d}$ and $\mathbf{y_d}$ were standardized. This data generation process resulted in 8100 datasets.

5.1.2. Analyses

All datasets were analyzed with PCovR, PLS and PCR, using one up to nine components. In the PCovR case, the four model selection strategies that were introduced in Section 4.3 (ML_SCR, SCR_ACV, ML_RCV and CV) were applied. When α is tuned through cross-validation, we considered 20 possible values evenly spread between .05 and 1. In the PLS and PCR cases, we assessed the number of components R by means of cross-validation as well as by means of the scree test. To this end, the L measures for PLS pertain to $\mathbf{y_d}$, and for PCR to $\mathbf{X_d}$. All cross-validation analyses were performed using 10 folds and 10 replications, and considering highest average cross-validation fit, as well as the 1 and 1.5 standard error rules. All scree tests included the correction step proposed by Wilderjans, Ceulemans, and Meers [13] that was described in the previous section.

An issue that is important to ensure fair comparison, is how the rotational freedom of the obtained solutions is dealt with. Target rotations towards the true components are not an option, because the obtained solutions can have a different number of components. To maximally enforce simple structure, we applied Varimax rotation [9] to each loading matrix, which maximizes the following criterion:

$$f(\mathbf{P}_{\mathbf{x_m}}) = \sum_{r=1}^{R}\left[\frac{1}{J}\sum_{j=1}^{J}\left(p_{rj}^2 - \overline{p_r^2}\right)^2\right]. \tag{14}$$

After rotating the loading matrices, the component score matrix and the regression weights were counter rotated.

5.1.3. Evaluation

To evaluate the performance of the different model selection strategies in terms of the optimality criteria that were specified in Section 3, we determined the nature of the retrieved components. To this end, we calculated Tucker congruencies [23] between the retrieved components and the true ones. More specifically, we distinguish four types of retrieved components:

1) intact components: these components have a congruence value of at least .85 with a true component
2) merged components: these components have a congruence value of at least .85 with the (weighted) sum of two true components
3) split components: these components have a congruence value of at least .85 with a true component, when summed
4) noise-driven components: retrieved components for which none of the above criteria applies

When a component is highly congruent with both a single true component and a sum of true components, the highest congruence determines whether it is called intact or merged. Similarly, a true component can be retrieved either intact or split. Solutions are marked as optimal when they contain two intact components that are congruent with the true relevant components, but no noise-driven, merged, or split components. Taking the cut-off values of Lorenzo-Seva and ten Berge [24] into account, a further distinction is made between solutions that have intact components that are fairly similar ($phi^{intact} > .85$) with the underlying relevant components, and solutions that have intact components that are equal ($phi^{intact} > .95$) to the underlying relevant components. Thus, solutions can be optimal in a very strict (equal components) or less strict way (fairly similar components).

Apart from evaluating the specified optimality criteria, we also assess the predictive ability of the different methods and strategies. To this end, we generated a test dataset for each original dataset, using the same data generating process. This way, the mean absolute deviations could not only be calculated for the $\mathbf{y}$ scores of the original dataset ($MAD_{\mathbf{y}}$), but also for the $\mathbf{y}$ scores that resulted from applying the same model (i.e., the same $\mathbf{W}$ and $\mathbf{p_y}$) to the test set ($MAD_{\mathbf{y_{test}}}$).

5.1.4. Results

Table 4 presents the results for all PCovR, PLS and PCR model selection strategies under consideration. The two last columns report the proportion of datasets for which an optimal solution is selected, according to the criteria mentioned in Section 5.1.3. Columns 4 to 8 give insight into the prevalence of the different types of retrieved components (intact and relevant, noise-driven, merged, split).

5.1.4.1. PCovR. SCR_ACV and ML_SCR perform best, with optimal solutions in 75% of the cases, followed by ML_RCV with 67% of optimal solutions, and CV with 64%. When zooming in on the specific problems associated with each PCovR strategy, it is striking that CV and ML_RCV

Table 3
The regression weights of the true datasets in the three different conditions.

		$R = 4$			
		1	*2*	*3*	*4*
$\mathbf{p}_{\mathbf{y_t}}$	Equal	0.71	0	0.71	0
	Small difference	0.60	0	0.80	0
	Large difference	0.44	0	0.90	0

Table 4
Percentage of datasets that contain particular types of retrieved components, and percentage of datasets for which the different methods and model selection strategies yield optimal solutions, in a very strict (equality) and a less strict way (similarity).

Method	Strategy	Cross-validation rule	Percentage of datasets that contain particular types of retrieved components					Percentage of datasets for which an optimal solution is obtained	
			All relevant (similar, intact)	All relevant (equal, intact)	Merged	Split	Noise-driven	Optimal solution (similarity)	Optimal solution (equality)
PCovR	CV	Highest fit	98%	81%	2%	1%	24%	74%	64%
		1 SE	95%	78%	5%	0%	15%	80%	67%
		1.5 SE	93%	76%	7%	0%	12%	81%	68%
	ML_SCR		81%	75%	0%	0%	0%	81%	75%
	SCR_ACV		81%	75%	19%	0%	0%	81%	75%
	ML_RCV	Highest fit	99%	82%	1%	2%	25%	74%	67%
		1 SE	99%	82%	1%	1%	18%	81%	72%
		1.5 SE	99%	82%	1%	1%	16%	83%	73%
	COMBI		99%	82%	0%	0%	0%	99%	82%
PCR	SCREE		78%	71%	0%	0%	0%	78%	71%
	CV	Highest fit	99%	81%	0%	3%	42%	58%	52%
		1 SE	99%	81%	0%	2%	32%	68%	61%
		1.5 SE	99%	81%	0%	2%	29%	71%	63%
PLS	SCREE		45%	36%	54%	0%	10%	37%	29%
	CV	Highest fit	75%	60%	29%	2%	16%	57%	45%
		1 SE	74%	59%	29%	1%	14%	60%	47%
		1.5 SE	74%	58%	29%	1%	13%	60%	48%

lead to noise-driven components in, respectively, 24% and 25% of the cases. As this is most likely a consequence of retaining too many components, it was inspected whether using a standard error rule could diminish this problem. Indeed, the proportion of noise-driven components decreased to, respectively, 12% and 16% when applying the 1.5 standard error rule, clearly improving the overall performance of the two strategies. In turn, SCR_ACV is associated with merging (19%). ML_SCR does not yield non-intact components, but tends to ignore weaker components, and thus is only able to recover *both* relevant components in 75% of the datasets.

As the weaknesses and strengths of the four strategies are quite complementary, we tried to compensate weaknesses by combining different strategies. Based on the above results, it makes most sense to combine ML_SCR and ML_RCV, because ML_SCR always finds intact components, and ML_RCV is most successful in recovering all relevant components. Specifically, we propose the following two-step combination strategy (COMBI). First, all ML_SCR components that are (at least) fairly similar to ML_RCV components (obtained with the 1,5 SE rule) are selected. Next, out of the remaining ML_RCV components, the ones with an absolute regression weight higher than .30 are selected as well.[5] COMBI outperforms all other methods: model selection is optimal in 82% of the datasets (99% when using the less strict optimality criterion), with selection of noise-driven, merged or split components occurring in less than 0.5% of the datasets only.

5.1.4.2. PLS and PCR. PCR performs reasonably well when a scree test is used (71%), but is only able to find an optimal solution in 52% of the datasets when cross-validation is used. The results for PLS are worse: scree test based model selection leads to only 29% of optimal solutions, and cross-validation to 45%.

The weak performance of PLS in this simulation study can be related to the fact that merging occurs in up to 54% of the datasets (depending on the model selection strategy that is used). PCR on the other hand falls short when cross-validation is used to determine the number of components. In 42% of the datasets, the model that is found, contains noise-driven components. Using a standard error rule clearly improves the performance, but even with a 1.5 standard error rule there are noise-driven components in 29% of the datasets. Using a scree test seems the better option when using PCR, as no noise-driven, merged, or split components were obtained with this strategy.

5.1.4.3. Predictive ability. Next to assessing which types of components result from applying the different methods and associated model selection strategies, we also checked the predictive ability of these components. Although there were large between-condition differences in the $MAD_{\mathbf{y}}$ values, the results were very similar across the strategies for both the original data and the test data. The only exceptions are PCR SCR and PCovR ML_SCR, which yield higher $MAD_{\mathbf{y}}$ values (see Fig. 1). Hence, strategies which emphasize reduction of $\mathbf{X}$ and use a scree test to determine the number of components, turn out to have a worse predictive ability. This makes sense, as in these cases the scree test focuses on the reduction of $\mathbf{X}$, largely ignoring the prediction of $\mathbf{y}$. What may be surprising, is that strategies that balance prediction and reduction, such as PCovR SCR_ACV, are able to attain a similar predictive ability as strategies that focus on prediction only, such as PLS CV, although the nature of the components that are found can be entirely different (see Table 4).

5.2. Simulation study 2

5.2.1. Data creation, analyses and evaluation

In the second simulation study, we fixed the number of components R again to 4. The strength of the components was also fixed: the weak components each explained 4% of the variance in $\mathbf{X}_t$, while the strong components explained 46% (see Table 2). The true regression weights of the relevant components had a small difference (see Table 3). The following data characteristics were manipulated in a full factorial design:

1. The average proportion of error on $\mathbf{X_d}$: 5%, 25%, 45%
2. The proportion of error on $\mathbf{y_d}$: 5%, 25%, 45%
3. The skewness of the error on $\mathbf{X_d}$: 0, 1
4. The skewness of the error on $\mathbf{y_d}$: 0, 1
5. The correlation between the components of $\mathbf{T_t}$: 0, 0.5
6. The size of matrix $\mathbf{X_t}$: 200 (N) × 48 (J), 50 (N) × 100 (J)

For each cell of the factorial design, $3 \times 3 \times 2 \times 2 \times 2 \times 2 = 144$ in total, 25 datasets were constructed using a similar data generation procedure to simulation study 1. The residuals were not sampled from a normal distribution, but from a beta distribution[6] with the specified

[5] .30 was selected here because it is Cohen's [25] cut-off value for a regression coefficient of medium size. This looks like a good decision because this cut-off value is high enough to avoid noise-driven, merged, or split components (see Table 4), but low enough to recover as much components as possible.

[6] The matlab function `pearsrnd` was used to this end: `pearsrnd(0,1,skewness,3,N,1)`.

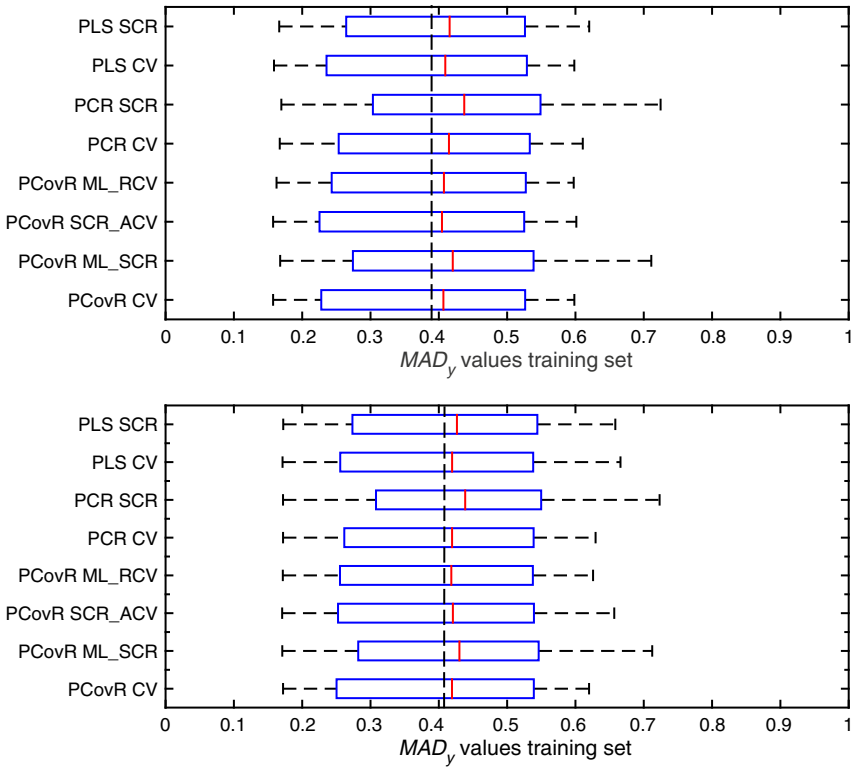


Fig. 1. Boxplots of the MAD_y values for each strategy, in both the training dataset (upper box) as in the test dataset (lower box). The overall means are indicated with a vertical line.

skewness. If the error matrix had a skewness s (in terms of the second and third central moment:

$$s = \frac{m_3}{m_2^{3/2}}) \tag{15}$$

with an absolute deviation greater than .1 from the specified skewness, another one was sampled.

The analyses of the datasets and the evaluation thereof occurred in the same way as in simulation study 1, with the exception of the rotation criterion that was used. For datasets with correlated components, Quartimin [10] rather than Varimax was used. Quartimin is an oblique rotation criterion that minimizes the sum across variables and across component pairs of the cross-products of the squared loadings:

$$f(\mathbf{P_x}) = \sum_{j=1}^{J} \sum_{r_1=1}^{R} \sum_{r_2 \neq r_1}^{R} p_{jr_1}^2 p_{jr_2}^2. \tag{16}$$

5.2.2. Results

In Table 5 the percentages of datasets for which optimal solutions are found are indicated for the different methods and model selection strategies (similar to the last column of Table 4). A distinction is made between some of the conditions: the correlation between the scores of $\mathbf{T_t}$, and the size of the predictor block ($N \times J$). The skewness of the error matrices only had a very small effect, and is therefore not included in this table as a separate factor. The same holds for the error proportions. Although they influence the exact percentages, they never influence which strategies perform better than others.

From Table 5 it can be seen that for both PCR and PLS, the optimal model selection strategy depends on the data characteristics, although differences are sometimes small. However, this is not the case for PCovR, where the combination strategy that was proposed in Section 5.1.4.1 outperforms all other strategies in every condition, as well as the PCR and PLS strategies.

Note that the lower optimality percentages in the first column (in comparison with the percentages in Table 4) are due to the fact that the datasets all had large differences in weakness and strength of the components in this simulation study. Correlation between the components causes even worse results overall, and there is an additional interaction effect between size of the correlation and data size.

6. Discussion

Dimension-reduction based regression methods involve two modeling aspects: reducing the predictors and predicting the criterion. In many cases it can be interesting to optimize both aspects

Table 5
Percentage of datasets for which the different methods and model selection strategies yield optimal solutions, in the most strict sense (equality). The percentages are shown for the different conditions of Factor 5 (correlation between components) and 6 (data size), with the most optimal result for each method printed in bold.

			200×48		50×100	
Method	Strategy	Cross-validation rule	$r_T = 0$	$r_T = 0.5$	$r_T = 0$	$r_T = 0.5$
PCovR	CV	Highest fit	48%	26%	38%	19%
		1 SE	54%	28%	46%	22%
		1.5 SE	55%	29%	49%	23%
	ML_SCR		33%	32%	20%	14%
	SCR_ACV		33%	32%	20%	14%
	ML_RCV	Highest fit	56%	31%	48%	26%
		1 SE	59%	32%	53%	26%
		1.5 SE	59%	32%	54%	26%
	COMBI		**64%**	**44%**	**61%**	**29%**
PCR	SCREE		33%	**33%**	33%	**23%**
	CV	Highest fit	40%	23%	33%	18%
		1 SE	47%	26%	44%	22%
		1.5 SE	**50%**	27%	**46%**	23%
PLS	SCREE		42%	17%	28%	7%
	CV	Highest fit	49%	**33%**	36%	21%
		1 SE	**49%**	33%	37%	21%
		1.5 SE	49%	33%	**38%**	**21%**

simultaneously, to gain insight into the mechanisms that link the predictors and the criterion. A method that is specifically designed for this aim is PCovR. While the more popular methods PLS and PCR only focus on, respectively, prediction and reduction, PCovR emphasizes both. PCovR includes a weighting parameter α that indicates to which extent reduction and prediction are optimized. This increases the flexibility of the method, but creates a complicated model selection problem.

Firstly, we clarified what we conceive as an optimal solution. Specifically, an optimal solution was defined as a solution that included all components that are relevant for predicting the criterion, but contains no components that do not truly underlie the predictors (i.e., no noise-driven, merged or split components).

Secondly, we studied how PCovR model selection should be performed in order to find solutions that maximally adhere to our optimality definition. Therefore, the model selection strategies (ML_SCR, ML_RCV, SCR_ACV, CV) that were proposed in a previous paper were evaluated in a thorough simulation study. Moreover, to show that given our optimality definition, PCovR is the method of choice, we compared the optimality of the PCovR solutions with the optimality of the PLS and PCR solutions.

It was clear that the four PCovR model selection strategies were quite complementary. ML_SCR was associated with very high α values, and therefore favored strong components. Because ML_SCR makes use of a scree test, which looks for "major factors" only, ML_SCR always finds intact components but often ignores weaker relevant components. CV and ML_RCV on the other hand were able to detect weaker components as well, but had the opposite problem: often the number of components that was selected was too high, yielding solutions with noise-driven components. This can be explained because rather than focusing on the "major factors", cross-validation tries to find the minimal number of components needed to completely describe the observed data. The prevalence of noise-driven components decreased when a standard error rule was added, but remained problematic even with a 1,5 standard error rule. SCR_ACV turned out to be a special case. Starting with a scree test, SCR_ACV favored strong components as well, thus selecting a number of components that corresponded with the number of true strong components. Then, cross-validation was used, creating a dilemma for PCovR: how to explain enough variance in $\mathbf{y}$ given the limited number of components that was selected? In the first simulation study, this led in 19% of the datasets to the selection of merged components.

Therefore, when the questions of a researcher match with the optimality criteria specified in this paper, a combination of PCovR model selection strategies seems the best option. When combining ML_SCR and ML_RCV through our COMBI strategy, the obtained components reflect true components in more than 99% of the datasets. Moreover, in 82% of the datasets, both underlying relevant components are recovered almost perfectly. It is virtually impossible to further improve this performance, as the other 18% of datasets are datasets in which there is 45% of error on $\mathbf{X_d}$ and in which the weak relevant component explains less than 20% of variance in $\mathbf{X_t}$, and is therefore masked. In the second simulation study, data characteristics were more difficult, with large differences in weakness and strength of the components, skewed error, and correlated components. Depending on the exact conditions, PCovR_COMBI yielded optimal solutions for 29% to 64% of the datasets, with the worst results in conditions with correlated components and more predictors than observations. However, PCovR_COMBI was still able to outperform the other PCovR strategies in all those conditions.

In general, the PCovR solutions clearly outperformed the PCR and PLS solutions. It may seem peculiar that PCovR strategies that favor very high α values, such as ML_SCR, still outperform PCR, but this effect corresponds with the finding from a previous study that a very high α value and an α value of exactly 1, can still yield quite different results [6]. Especially PLS yielded remarkably less optimal solutions than PCovR. A closer inspection tells us that PLS solutions are often prone to merging. This is not so strange, as merging does not necessarily hamper prediction of the criterion, which is what the boxplots of Fig. 1 also showed. Although there were a lot of differences between the nature of components yielded by the different strategies and methods, their predictive ability only had small differences. Thus, for researchers who have other optimality criteria in mind, such as accurate prediction of $\mathbf{y}$, PLS might still be a suitable method.

It can be concluded that when researchers want to apply dimension-reduction based regression methods on their data, they should reflect on what kind of conclusions they want to draw from the results, before selecting a specific method. Indeed, the modeling method used largely influences what the nature will be of the components that are found, and should therefore be a deliberate choice.

Acknowledgments

The research leading to the results reported in this paper was supported in part by the Research Fund of KU Leuven (GOA/15/003), and by the Interuniversity Attraction Poles program financed by the Belgian government (IAP/P7/06). For the simulations we used the infrastructure of the VSC—Flemish Supercomputer Center, funded by the Hercules foundation and the Flemish Government—department EWI.

References

[1] H.A.L. Kiers, A.K. Smilde, A comparison of various methods for multivariate regression with highly collinear variables, Stat. Method. Appl. 16 (2)) (2007) 193–228.
[2] S. Wold, A. Ruhe, H. Wold, W.J. Dunn III, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, SIAM J. Sci. Stat. Comput. 4 (3) (1984) 735–743.
[3] I.T. Jolliffe, A note on the use of principal components in regression, J. R. Stat. Soc.: Ser. C: Appl. Stat. 31 (3) (1982) 300–303.
[4] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58.
[5] S. De Jong, H.A.L. Kiers, Principal covariates regression: part I. Theory, Chemom. Intell. Lab. Syst. 14 (1–3) (1992) 155–164.
[6] M. Vervloet, K. Van Deun, W. Van den Noortgate, E. Ceulemans, On the selection of the weighting parameter value in principal covariates regression, Chemom. Intell. Lab. Syst. 123 (2013) 36–43.
[7] M. Vervloet, H.A.L. Kiers, W. Van den Noortgate, E. Ceulemans, PCovR: an R package for principal covariates regression, J. Stat. Softw. 65 (8) (2015) 1–14.
[8] M.W. Browne, An overview of analytic rotation in exploratory factor analysis, Multivar. Behav. Res. 36 (1) (2001) 111–150.
[9] H.F. Kaiser, The varimax criterion for analytic rotation in factor analysis, Psychometrika 23 (3) (1958) 187–200.
[10] J.B. Carroll, An analytical solution for approximating simple structure in factor analysis, Psychometrika 18 (1) (1953) 23–38.
[11] E. Ceulemans, H.A.L. Kiers, Discriminating between strong and weak structures in three-mode principal component analysis, Br. J. Math. Stat. Psychol. 62 (2009) 601–620.
[12] J. Niesing, Simultaneous Component and Factor Analysis Methods for two or More Groups: A Comparative Study, Leiden, DSWO Press, 1997.
[13] T.F. Wilderjans, E. Ceulemans, K. Meers, CHull: a generic convex hull based model selection method. Behav. Res. Methods 45 (1) (2013) 1–15.
[14] M.E. Timmerman, H.A.L. Kiers, Three-mode principal component analysis: choosing the numbers of components and sensitivity to local optima, Br. J. Math. Stat. Psychol. 53 (2000) 1–16.
[15] M.E. Timmerman, U. Lorenzo-Seva, E. Ceulemans, The Number of factors problem, in: P. Irwing, T. Booth, D.J. Hughes (Eds.), The Wiley-Blackwell Handbook of Psychometric Testing, Wiley-Blackwell, Oxford, 2015 (in press).
[16] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Brooks/Cole Publishing, Monterey, 1984.
[17] U.M. Braga-Neto, E.R. Dougherty, Is cross-validation valid for small-sample microarray classification, Bioinformatics 20 (3)) (2004) 374–380.
[18] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer-Verlag, New York, 2001.
[19] P. Filzmoser, B. Liebmann, K. Varmuza, Repeated double cross validation, J. Chemom. 23 (2009) 160–171.
[20] Smilde, A., Bro, R., & Geladi, P. (2004). Multi-way Analysis with Applications in the Chemical Sciences. Chichester: John Wiley & Sons.
[21] C. Heij, P.J. Groenen, D. Van Dijk, Forecast comparison of principal component regression and principal covariate regression, Comput. Stat. Data Anal. (2007) 3612–3625.
[22] U. Indahl, A twist to partial least squares regression, J. Chemom. 19 (2005) 32–44.
[23] L.R. Tucker, R.F. Koopman, R.L. Linn, Evaluation of factor analytic research procedures by means of simulated correlation matrices, Psychometrika 34 (1969) 421–459.
[24] U. Lorenzo-Seva, J.M. ten Berge, Tucker's congruence coefficient as a meaningful index of factor similarity. Meth. Eur. J. Res. Meth. Behav. Soc. Sci. 2 (2) (2006) 57–64.
[25] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, second ed. Erlbaum, Hillsdale, NJ, 1988.