CrossMark

Retrieving relevant factors with exploratory SEM and principal-covariate regression: A comparison

Marlies Vervloet[1] · **Wim Van den Noortgate**[1] · **Eva Ceulemans**[1]

Abstract

Behavioral researchers often linearly regress a criterion on multiple predictors, aiming to gain insight into the relations between the criterion and predictors. Obtaining this insight from the ordinary least squares (OLS) regression solution may be troublesome, because OLS regression weights show only the effect of a predictor on top of the effects of other predictors. Moreover, when the number of predictors grows larger, it becomes likely that the predictors will be highly collinear, which makes the regression weights' estimates unstable (i.e., the "bouncing beta" problem). Among other procedures, dimension-reduction-based methods have been proposed for dealing with these problems. These methods yield insight into the data by reducing the predictors to a smaller number of summarizing variables and regressing the criterion on these summarizing variables. Two promising methods are principal-covariate regression (PCovR) and exploratory structural equation modeling (ESEM). Both simultaneously optimize reduction and prediction, but they are based on different frameworks. The resulting solutions have not yet been compared; it is thus unclear what the strengths and weaknesses are of both methods. In this article, we focus on the extents to which PCovR and ESEM are able to extract the factors that truly underlie the predictor scores and can predict a single criterion. The results of two simulation studies showed that for a typical behavioral dataset, ESEM (using the BIC for model selection) in this regard is successful more often than PCovR. Yet, in 93% of the datasets PCovR performed equally well, and in the case of 48 predictors, 100 observations, and large differences in the strengths of the factors, PCovR even outperformed ESEM.

Keywords Principal-covariate regression · Exploratory structural equation modeling · Multicollinearity · Dimension reduction

In the behavioral sciences, researchers often linearly regress a criterion variable $\mathbf{y}$ on J predictor variables $X_1, X_2 \ldots X_J$, aiming to gain insight into the relations between the predictors and the criterion (Johnson, 2000). Obtaining this insight from the ordinary least squares regression (OLS) solution is troublesome, however. First, OLS regression weights show only the additional effect of a predictor on top of the others, and therefore do not reveal their shared effects (Bulteel, Tuerlinckx, Brose, & Ceulemans, 2016). For example, Bulteel et al. (2016) demonstrated for a dataset with 11 depression-related symptoms that

only a small part of the explained variance of each variable was attributable to the unique direct effects. Moreover, when the number of predictors grows larger, it becomes likely that the predictors will be highly collinear, which makes the regression weights unstable as well (i.e., the bouncing beta problem; Kiers & Smilde, 2007).

Several solutions have been proposed for dealing with the "bouncing beta" problem. Most of them can be classified as variable selection methods (such as forward selection and backward elimination; Hocking, 1976), penalty methods (such as ridge regression and Lasso; Hoerl & Kennard, 1970; Tibshirani, 1996), ensembles (such as random forests; Breiman, 2001), and dimension reduction-based methods (Kiers & Smilde, 2007). The first three solutions do not directly shed light on which predictors have similar effects on the criterion. For instance, if a predictor gets a regression weight close to zero when a lasso penalty is added to the OLS regression, this can either mean that the predictor is unrelated to the criterion or that the effect of the predictor coincides to a large extent with that of another predictor. In

✉ Marlies Vervloet
marlies.vervloet@kuleuven.be

Wim Van den Noortgate
wim.vandennoortgate@kuleuven.be

Eva Ceulemans
eva.ceulemans@kuleuven.be

[1] KU Leuven, Leuven, Belgium

contrast, dimension-reduction-based methods do aim to yield insight into the underlying mechanisms, by reducing the predictors to a smaller number of summarizing variables and regressing the criterion on these summarizing variables.

The most popular dimension-reduction-based regression methods are partial least squares (PLS; Wold, Ruhe, Wold, & Dunn, 1984) and principal-component regression (PCR; Jolliffe, 1982), which both extend principal-component analysis (PCA; Pearson, 1901) to the regression context. However, neither of these two methods simultaneously optimizes reduction and prediction, in contrast to principal-covariate regression (PCovR; De Jong & Kiers, 1992). Indeed, PCovR explicitly searches for components that are not only good summarizers of the predictors but also explain the variance of the criterion, whereas PCR focuses exclusively on the former and PLS on the latter. Not surprisingly, PCovR outperformed PCR and PLS in a previous study (Vervloet, Van Deun, Van den Noortgate, & Ceulemans, 2016) that investigated how well the three methods recover the underlying components that are *relevant* for predicting the criterion (i.e., they explain some variance in the criterion), irrespective of their *strength* (i.e., how much variance in the predictors they explain).

Within the factor-analysis-based framework, another method is available that simultaneously combines reduction and prediction: exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009). When ESEM is applied, latent factors are searched for that explain the correlations between the predictors, and these factors are then used to predict the criterion variable. PCovR and ESEM stem from different frameworks and traditions, though, and the resulting solutions have not yet been compared. It is thus unclear what the strengths and weaknesses are of the two methods. It is important to notice that the theoretical statuses of components and factors are not the same (Bollen & Lennox, 1991; Borsboom, Mellenbergh, & Van Heerden, 2003; Coltman, Devinney, Midgley, & Veniak, 2008): Whereas components are assumed to be influenced by the observed variables (*formative relationship*), factors are assumed to cause the scores on the observed variables (*reflective relationship*). Moreover, components are linear combinations of the observed variables, whereas factors are assumed to exist independent of the observed variables (for a detailed comparison of formative and reflective models, see Coltman et al., 2008). Despite of these theoretical differences, it has been established that factor analysis and component analysis often lead to the same conclusions with respect to the dimensions underlying the data (Ogasawara, 2000; Velicer & Jackson, 1990). A more fundamental difference between the two methods is that PCovR has a weighting parameter with which the user can tune the degrees of emphasis on reduction versus prediction, whereas no such parameter is available for ESEM. It

is therefore useful to investigate the extent to which ESEM and PCovR lead to factors that can be similarly interpreted.

When comparing the performance of two methods, it is crucial to specify the evaluation criterion that one is interested in (Doove, Wilderjans, Calcagnì, & Van Mechelen, 2017). In a regression context, a lot of possibilities are conceivable, such as predictive ability or the estimation accuracy of the regression weights of the separate predictors (e.g., Kiers & Smilde, 2007). In line with the research questions mentioned above, we will focus in this article on the extent to which PCovR and ESEM are able to extract factors or components that truly underlie the predictor scores and that predict a single criterion. To this end, simulated data will be used, since we can then know for sure which factors or components were used to generate the data.

The remainder of this article has been structured as follows. In the following two sections, we briefly recapitulate PCovR and ESEM, respectively. Then we put both methods to the test in two simulation studies, focusing on the number and the nature of the estimated factors. Next we further illustrate the performance of ESEM and PCovR, by applying them to a real dataset, and we end with some discussion points and concluding remarks.

Principal-covariate regression

Model

PCovR combines two goals: reduction (Eq. 1) and prediction (Eq. 2). The predictors $X_1, \ldots X_J$ are reduced to R components, which are linear combinations of the predictor scores (i.e., formative relation), and the criterion $\mathbf{y}$ is regressed on those components:

$$\mathbf{X} = \mathbf{T}\mathbf{P_X} + \mathbf{E_X} \tag{1}$$

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

$\mathbf{X}$ is a matrix that horizontally concatenates the J predictors. $\mathbf{P_X}$ and $\mathbf{p_y}$ are, respectively, the loading matrix and the regression weight vector. The component scores are given by $\mathbf{T}$. $\mathbf{E_X}$ and $\mathbf{e_y}$ contain, respectively, the residual $\mathbf{X}$ and $\mathbf{y}$ scores. Because component-based methods model the total variance of a matrix, $\mathbf{E_X}$ refers to the part of the observed scores that is explained by the components that are not retained, and is therefore of rank $R - J$ or lower. Note that the columns of $\mathbf{T}$ do not correlate with the columns of $\mathbf{E_X}$ or with $\mathbf{e_y}$.

We assume that the data are centered around zero. This is necessary for $\mathbf{X}$, to ensure that PCovR models the correlation or covariance structure of the data. Centering of $\mathbf{y}$ is optional, but discards the need for an intercept in Eq. 2, as discussed by Vervloet, Kiers, Van den Noortgate, and Ceulemans (2015).

Estimation

The key feature of PCovR is that the two model aspects (reduction of $\mathbf{X}$ and prediction of $\mathbf{y}$) are optimized simultaneously, which can be seen in the following loss function:

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

where α $(0 < \alpha \leq 1)$ is the prespecified weighting parameter that determines to which degree reduction rather than prediction is emphasized. A closed form solution always exists, given a specific α and R value. This solution is obtained by estimating $\mathbf{T}$ by the first R eigenvectors of matrix $\mathbf{G}$:

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

with $\mathbf{H_X}$ being a projection matrix that projects $\mathbf{y}$ on $\mathbf{X}$. Note that $\mathbf{T}$ is usually rescaled so that the columns have a variance of 1, in order to resolve nonidentification. $\mathbf{P_X}$ and $\mathbf{p_y}$ can then be computed as follows:

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

$$\mathbf{P_y} = \mathbf{T}'\mathbf{y} \tag{6}$$

Estimation usually ends with a rotation procedure, as PCovR solutions with at least two components have rotational freedom. Indeed, premultiplying $\mathbf{P_X}$ and $\mathbf{p_y}$ by any transformation matrix, and postmultiplying $\mathbf{T}$ by the inverse of the transformation matrix, does not change the reconstructed $\mathbf{X}$ scores or the predicted $\mathbf{y}$ scores. This rotational freedom implies that the loading matrix can for example be rotated toward simple structure, in order to obtain a loading matrix that is easier to interpret. Browne (2001) defines a simple structure loading matrix as a matrix with only one nonzero loading per predictor, and with more than R but fewer than J zero-loadings per component. A distinction can be made between rotation criteria that change the correlation between factors (i.e., oblique rotation criteria) or not (i.e., orthogonal rotation criteria). In the R package PCovR (Vervloet et al., 2015), the following rotation criteria are made available: Varimax (Kaiser, 1958), Quartimin (Carroll, 1953), weighted Varimax (Cureton & Mulaik, 1975), oblique (Browne, 1972b) and orthogonal (Browne, 1972a) target rotation, and Promin (Lorenzo-Seva, 1999). However, other rotation criteria could be used as well.

Model selection

PCovR model selection consists of both the selection of the α value and the number of components. In a previous study (Vervloet et al., 2016), the so-called COMBI strategy led to the best results in terms of retrieving the relevant components, which is the performance criterion that we also focus on in this study. This strategy, starts with the computation of the α value that maximizes the likelihood of the data. On the basis of the work of Gurden (n.d.), the following formula can be used:

$$\alpha_{ML} = \frac{\|\mathbf{X_d}\|^2}{\|\mathbf{X_d}\|^2 + \|\mathbf{y_d}\|^2 \dfrac{\sigma_{\mathbf{E_X}}^2}{\sigma_{\mathbf{e_y}}^2}} \tag{7}$$

For selecting the number of components given α_{ML}, two approaches are combined. Firstly, the analysis is run with the number of components varying from 1 to a specified number R_{max}. Subsequently, the model is selected (called the ML_SCR model) that yields the optimal balance between this loss function value and the number of components, by looking after which number of components the decrease in L levels off (i.e., a scree test procedure; Cattell, 1966; Ceulemans & Kiers, 2006). Secondly, we compute the cross-validation fit for the models with 1 to R_{max} components and retain the model with the optimal cross-validation fit (called the ML_RCV model). Note that instead of choosing the model with the highest average cross-validation fit (across several random partitions into folds), we select the most parsimonious model among all models with a cross-validation fit that differs less than 1.5 standard error from the highest average cross-validation fit (Filzmoser, Liebmann, & Varmuza, 2009; Hastie, Tibshirani, & Friedman, 2001).

The final COMBI model contains the components that are present in both the ML_SCR and the ML_RCV model, indicated by high Tucker congruencies (Tucker, Koopman, & Linn, 1969) between components in both models. Moreover, the components from the ML_RCV model that have at least a moderate regression weight—that is, higher than .30 (Cohen, 1988)—are added. For full details of the COMBI strategy, see Vervloet et al. (2016).

Exploratory structural equation modeling

Model

Structural equation models (Kline, 2015) consist of two parts. The measurement part reflects how observed variables are linked to underlying latent factors. In general, the structural part specifies the relations between these latent factors. Usually, structural equation models are confirmatory in that the loading structure of the variables is (partly) specified beforehand. However, using such a confirmatory approach often leads to biased structural model estimates, because typically the loadings are restricted to have zero cross-loadings, leading to inflated factor correlations (Marsh, Liem, Martin, Nagengast, & Morin, 2011; Marsh, Morin, Parker, & Kaur,

2014; Marsh et al., 2009). Asparouhov and Muthén (2009) proposed ESEM as a general framework that integrates confirmatory factor analysis with structural equation modeling and exploratory factor analysis, in which the loadings can be free parameters that are estimated. Often, ESEM is used as a confirmatory tool—for example, for measurement invariance testing—but it has been demonstrated (Marsh et al., 2014) that it is a valuable exploratory tool, as well. Specifically, we focus in this article on how ESEM can be used to model a single criterion[1] that is regressed on latent factors in an exploratory way, which comes very close to a PCovR model. Hence, the same model formulae as in Eqs. 1 and 2 can apply here, but with factors instead of components—thus implying that the relations between the observed and latent variables are reflective. Both the residuals of $\mathbf{X}$ and $\mathbf{y}$ are assumed to be normally distributed with mean 0. $\mathbf{E_X}$ refers to the part of the observed scores that originates from unique variances of the predictors and/or error variance, since factor-based methods only model common variance. Note that ESEM further assumes that the observed data are continuous and multivariate normal.

Estimation

ESEM has no closed-form solution. Hence, an iterative algorithm is necessary for estimating the ESEM parameters. The algorithm that is programmed in Mplus (the only software in which ESEM is available) is based on the gradient projection algorithm (Jennrich, 2001, 2002) and makes use of a maximum likelihood estimator, although other estimators are possible, as well (e.g., weighted least squares estimator or robust alternatives; Marsh et al., 2014). The use of multiple starting values is common practice, to avoid local minima. By default in Mplus, ESEM is performed with 30 random starting values. Note that nonconvergence problems can occur, which is why the default settings imply that the algorithm will be stopped after a maximum of 1,000 iterations. For more details on the estimation, see Asparouhov and Muthén (2009).

Again, the estimation step usually ends with a rotation procedure, in order to simplify the interpretation. In Mplus the following rotation criteria are available, among others: Varimax (Kaiser, 1958), Quartimin (Carroll, 1953), Geomin (Yates, 1987), and target rotation (Browne, 1972a, 1972b).

Model selection

When running ESEM analyses, model selection consists of determining the appropriate number of latent factors. In Mplus, the following information criteria are available that can be used for this task: the Bayesian information criterion (BIC; Schwarz, 1978), the sample-size-adjusted BIC (SABIC;

Sclove, 1987), and the Akaike information criterion (AIC; Akaike, 1974). The BIC value is calculated as follows:

$$BIC = 2\ln(\hat{L}) + k\ln(N). \qquad (8)$$

It can be seen that the BIC consists of a term involving $\hat{L}$, which is the maximized value of the likelihood function of the model, and a penalty term. The k in this penalty term refers to the number of free parameters:

$$k = (J+1)(R+2) - \frac{R^2 - R}{2}, \qquad (9)$$

which is the number of free parameters: the intercepts of $\mathbf{X}$ and $\mathbf{y}$, the loadings, the residual variances of $\mathbf{X}$ and $\mathbf{y}$, and the regression weights, minus the number of correlations between the factors. The number of factors that leads to the lowest BIC value is preferred. Alternatively, model selection can be based on the SABIC or AIC, which are calculated in similar ways, but with different penalty terms:

$$AIC = -2\ln(\hat{L}) + 2*k \qquad (10)$$

and

$$SABIC = -2\ln(\hat{L}) + k\ln\left(\frac{N+2}{24}\right). \qquad (11)$$

From the formulae it can be concluded that the penalty term will usually be highest for BIC and lowest for AIC. Therefore, when comparing the models that are selected by the three different strategies, BIC will often lead to the selection of the least complex model, and AIC to the most complex model. Bulteel, Wilderjans, Tuerlinckx, and Ceulemans (2013) found, for instance, a very low success rate for AIC at estimating the correct number of underlying factors in the context of mixtures of factor analyzers, due to its tendency toward selecting too many factors. BIC had much better performance in their simulation study, but it underestimated the number of factors in difficult conditions. Vrieze (2012) indicates that the BIC is assumed to select the true model as the sample size grows, as long as the true model is under consideration, whereas the AIC selects the model that minimizes the mean squared error of estimation, and can be preferred over BIC if the true model is not among the candidate models being considered.

Simulation studies

To make a solid comparison between PCovR and ESEM with regard to the recovery of the underlying factors, both methods were put to the test in two simulation studies. In the first study, the number of true underlying factors was manipulated as well as their strength, in order to investigate whether the

---

[1] In this case, the latent criterion factor in the measurement model equals the manifest criterion variable.

performance of PCovR and ESEM would be affected by these (latent) data characteristics. Regarding the strength of the factors, Velicer and Fava (1998) showed that the more variables load high on a factor, the more likely it is that this factor will be recovered. Therefore, weak factors can be hypothesized to be difficult to retrieve, especially in conditions with a lower number of predictors. As we stated in the introduction, we also wanted to investigate the influence of the relevance of the factors. Yet, in the case that the number of true factors is varied, it is difficult to come up with a balanced design in which both the strength and the relevance of the factors can be manipulated. For this reason, a second study was performed as well, in which the number of factors was fixed to four. In this way, the datasets in the second study could contain factors with different strength–relevance characteristics. This manipulation would be especially interesting, because in such cases, balancing good reduction and prediction becomes more important.

Next to the number, strength, and relevance of the factors, the number of predictors, the number of observations, the percentage of error variance in **X**, and the percentage of error variance in **y** are other interesting data characteristics to manipulate, because they determine the amount of information available in the data. The amount of error on the predictor block, especially, has been shown to influence the performance of PCovR (Vervloet et al., 2016). It can be hypothesized, however, that ESEM will have fewer difficulties with error on the predictor block, because factor-based methods model common variance, whereas component-based methods model the total variance (Widaman, 1993). The number of observations was not manipulated in the study of Vervloet et al. (2016), but for both factor-based and component-based methods (Guadagnoli & Velicer, 1988), researchers have already found a connection in simulation studies between the strength of the factors and the number of observations that is recommended, with more observations being needed to retrieve weaker factors. Besides, for PCovR, the tuning of the weighting parameter becomes more important in the case of fewer observations (Vervloet, Van Deun, Van den Noortgate, & Ceulemans, 2013).

In both simulation studies we inspected the numbers of factors[2] that were retained by PCovR versus ESEM. Next, we explored the similarities and differences of the techniques in retrieving optimal solutions. The definition of what constitutes an optimal solution was based on the typology of factors introduced by Vervloet et al. (2016). They classified estimated factors as those identical to the true ones, merged factors (i.e., congruent to the sum of true factors), split factors (i.e., congruent to a true factor when summed up), and noise-driven factors. The true factors can further be divided into factors that

are relevant for predicting the criterion, and irrelevant factors. A solution is considered optimal if all the true factors that are relevant for predicting the criterion are revealed. In the second simulation study irrelevant factors were also present, and recovering these factors was unnecessary but not problematic. Retrieving split, merged, or noise-driven factors is not allowed.

To this end, Tucker congruencies were calculated between each (sum of) factor(s) and each (sum of) true factor(s). In line with the work of Lorenzo-Seva and ten Berge (2006), we called factors with a congruence higher than .85 *fairly similar*, and those higher than .95 *equal*. For the detection of merging and splitting, the cutoff value $C = .85$ was used. For deciding whether or not a true factor was missing, we inspected both cutoff values, since this gives us extra information. If a factor had a high congruency with both a true factor and the sum of true factors, only the highest congruency counted. The same held when the sum of two factors, considered as one factor, was found to be congruent with a true factor.

Simulation Study 1

Data generation

In this simulation study, the following data characteristics were manipulated in a full factorial design:

1. The number of predictors J: 30, 60
2. The percentage of error variance in **X**: 25%, 45%, 65%
3. The percentage of error variance in **y**: 25%, 45%, 65%
4. The number of observations N: 100, 200, 500, 1000
5. The number of factors R: 1, 2, 3, 4, 5
6. Strengths of the factors: equal versus unequal strength (see Table 1)

For each of the $2 \times 3 \times 3 \times 4 \times 5 \times 2 = 720$ cells of the factorial design, 200 datasets were constructed, yielding 144,000 datasets in total. The factor score matrix was sampled from a standard normal distribution. Subsequently, the columns were standardized and orthogonalized. The loading matrices were created using only ones and zeros, with each predictor having only one loading of 1 and with the number of 1s per factor varying according to Table 1.[3] The regression weights were set to $\sqrt{1/R}$, such that the sum of squared regression weights equaled 1.

The error matrices of **X** and **y** were drawn from standard normal distributions as well, and were also standardized and

---

[2] For reasons of clarity, both (ESEM) factors and (PCovR) components will be referred to as *factors* in the remainder of this article.

[3] Note that in the case of only one factor underlying the data, no distinction can be made between equal or unequal factors. Hence, both conditions are equivalent. Moreover, for cases in which equal strengths of the factors was not possible ($R = 4$, $J = 30$), the numbers of high-loading variables per factor were set to be as similar as possible.

Table 1 Numbers of high-loading variables per factor in the different strength conditions

Condition	Strength	
	Equal	Unequal
J = 30		
R = 1	30	30
R = 2	15–15	3–27
R = 3	10–10–10	3–10–17
R = 4	8–8–7–7	3–6–9–12
R = 5	6–6–6–6–6	3–5–6–7–9
J = 60		
R = 1	60	60
R = 2	30–30	6–54
R = 3	20–20–20	6–20–34
R = 4	15–15–15–15	6–12–18–24
R = 5	12–12–12–12–12	6–9–12–15–18

orthogonalized. Furthermore, we made sure that the columns of the factor score matrix and the error on **y** were uncorrelated. The error matrices were rescaled to obtain the desired average percentage of error variance specified above (Data Characteristics 2 and 3), in comparison with the total variance per predictor. The exact amount of error per predictor, however, was varied such that one third of the variables contained less error variance and one third contained more (see Vervloet et al., 2016, for more details).

X was created by multiplying the factor score matrix with the loading matrix and adding the error matrix of **X**, according to Eq. 1. The **y** vector was obtained by multiplying the factor score matrix by the regression weight vector and adding the error matrix of **y**, according to Eq. 2. Finally, **X** and **y** were standardized.

Data analysis

All datasets were analyzed with PCovR and ESEM, using from one up to seven factors. Afterward, for both methods and each dataset, we determined the number of factors to retain. For PCovR solutions we applied the COMBI strategy described above, and for ESEM we used AIC, BIC, and SABIC.

Since the estimated factors had rotational freedom, rotating them was necessary to evaluate the recovery of the true factors. Target rotation toward the true factors might sound appealing here, but this was not feasible, because the estimated models could differ in the number of factors from the true models. Moreover, when analyzing real data in an exploratory way, target rotation is not an option, either, because no target structure is available. Thus, we opted for Varimax (Kaiser, 1958), the most often used orthogonal rotation criterion,

which enforces a simple structure of the loading matrices by maximizing the following function:

$$f(\mathbf{P_x}) = \sum_{r=1}^{R} \left[\frac{1}{J} \sum_{j=1}^{J} \left(p_{rj}^2 - \overline{p_r^2} \right)^2 \right]. \tag{12}$$

Evaluation

We started by inspecting the numbers of factors selected by the COMBI strategy for PCovR and by the AIC, BIC, or SABIC information criteria for ESEM. This was not the main research question, but it would already give some important information, since it would indicate that the methods differed in their sensitivity toward specific types of true factors. Moreover, if more factors were selected than the true underlying number of factors, then factors had been included that were noise-driven and/or the true factors had been split into multiple fragments. In the case that fewer factors were selected, the true factors could have been merged and/or not been picked up at all. We could not be sure unless we compared the scores of the observations for the true and estimated factors, which would allow us to determine whether a so-called optimal solution had been found.

Results

Because of the iterative nature of ESEM, it can be expected that convergence will not always be reached. We will first zoom in on whether and when nonconvergence occurred, before examining the numbers of factors that were retrieved and inspecting the optimality of the obtained solutions.

Nonconvergence Nonconvergence occurred often with ESEM, but only when extracting specific numbers of factors. For every dataset, convergence was reached for at least one model, so a solution could always be selected. Moreover, Fig. 1 shows that if a model with a specific number of factors failed to converge, the number of factors being considered was in more than 99% of the cases higher than the number of true factors.

Number of factors Table 2 shows for each strategy under study in which percentage of the datasets the correct number of factors was recovered, or a lower or higher number, or no solution at all.[4]

It can be seen that the PCovR COMBI and the ESEM BIC strategies usually selected a number of factors that was too

---

[4] PCovR yielded a model with no factors in two datasets, because the COMBI strategy did not find a match between the factors of the ML_SCR model and the ML_RCV model, and the ML_RCV did not yield any factors with regression weights higher than the cutoff value. Both datasets had 65% error on the criterion.
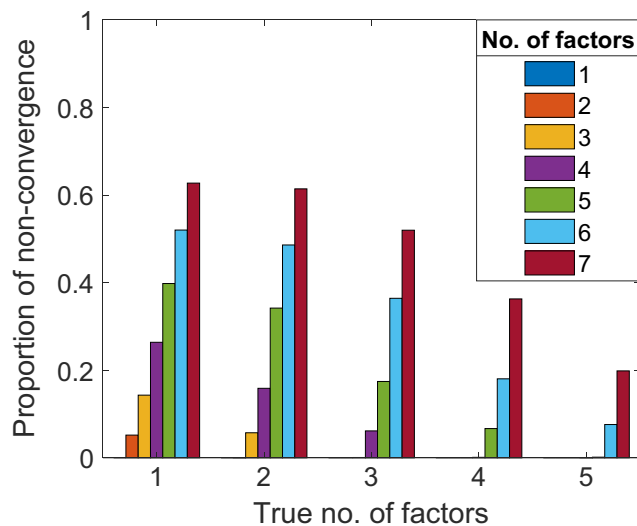
Fig. 1 Proportions of nonconvergence per considered number of factors for each true number of factors.

low when the correct number was not selected, whereas ESEM AIC and SABIC instead selected a model that was too complex. For SABIC, this was the case in 22% of the datasets.

When zooming in on the numbers of factors that were selected per the true value of the number of factors (Fig. 2), it can be concluded that whereas the ESEM AIC and BIC only slightly misspecified the number of factors when the correct number was not selected, PCovR COMBI and ESEM SABIC deviated more from the true number. Even when the true number equaled 1, SABIC sometimes selected up to seven factors.

Optimality of the obtained solutions Looking at the selected number of factors does not, however, provide complete information about how the methods are performing. Indeed, even a model with two merged factors and a noise-driven factor might still be a model with the correct number of factors.

In the last column of Table 3, it can be seen that ESEM BIC led most often to an optimal solution (in 60% of the datasets with $C = .95$). If an optimal solution was not reached, this was usually because a true factor was missing. Only in 4% of the

Table 2 Percentages of datasets for which a lower, correct, or higher number of factors was found than the true number

	Lower	Correct	Higher	No Solution
PCovR				
COMBI	5%	93%	2%	<1%
ESEM				
AIC	<1%	91%	8%	0%
BIC	6%	94%	0%	0%
SABIC	<1%	78%	22%	0%

datasets did BIC yield noise-driven factors, which was a lower percentage than for any of the other strategies. PCovR COMBI and ESEM AIC led to similar percentages of optimal solutions (respectively, 54% and 56%), but COMBI had more missing true factors, whereas AIC had more noise-driven factors, explaining their tendencies toward, respectively, selecting a lower and higher number of factors. PCovR SABIC had the worst performance, with only 47% optimal solutions. The nonoptimal solutions had split (6%) and/or noise-driven factors (25%) and/or true factors that were missing (45%).

ESEM BIC yielded a higher percentage of optimal solutions than did AIC or SABIC for both cutoff values. To focus on the difference in performance between ESEM and PCovR, we therefore will only consider the BIC strategy for ESEM for the remainder of this section.

To explore which conditions are more challenging for ESEM and which are more challenging for PCovR, we performed an ANOVA (including all interaction effects) with the six data characteristics as independent variables and the difference between PCovR and ESEM in proportions of replications for which optimal solutions were found as the dependent variable. However, we left out the conditions with only one true factor, because the strength of the factor could not be manipulated as being "equal" or "unequal", thus leading to an unbalanced design.

Only considering larger effects ($\eta_p^2 > .20$), we found for a cutoff value $C = .95$ effects of the amount of error on $\mathbf{X}$ and the number of predictors J. Moreover, interaction effects were found between the error on $\mathbf{X}$ and the strength of the factors, between J and the error on $\mathbf{X}$, and a three-way interaction with the error on $\mathbf{X}$, J, and the strength of the factors. When using $C = .85$, we only found a large effect of the strength of the factors.

Figure 3 shows that ESEM mainly outperformed PCovR in the case of low error on $\mathbf{X}$, $J = 30$, and unequal strength, when the strict cutoff value $C = .95$ was applied (right panels). With $C = .85$ (left panels), the pattern was different: Differential performance mainly occurred for high levels of error on $\mathbf{X}$ and low numbers of observations. Usually, ESEM outperformed PCovR in those conditions, but it was the other way around with unequal strength and $J = 60$.

Overall, more optimal solutions are found when the data contain fewer underlying factors, more predictors, and less error on $\mathbf{X}$ and $\mathbf{y}$. Also the number of observations plays a big role, since the percentage of optimal solutions is always near 100% with a less strict cutoff value, as long as the data have 500 or more observations. Finally, the strength of the factors is mainly an influence if other complicating data characteristics co-occur, with weaker factors (i.e., in unequal-strength conditions) being more difficult to recover. Note that in many conditions, big differences were found between the results with different cutoff values. Specifically, in many cases the factors that were found were all similar to the true factors
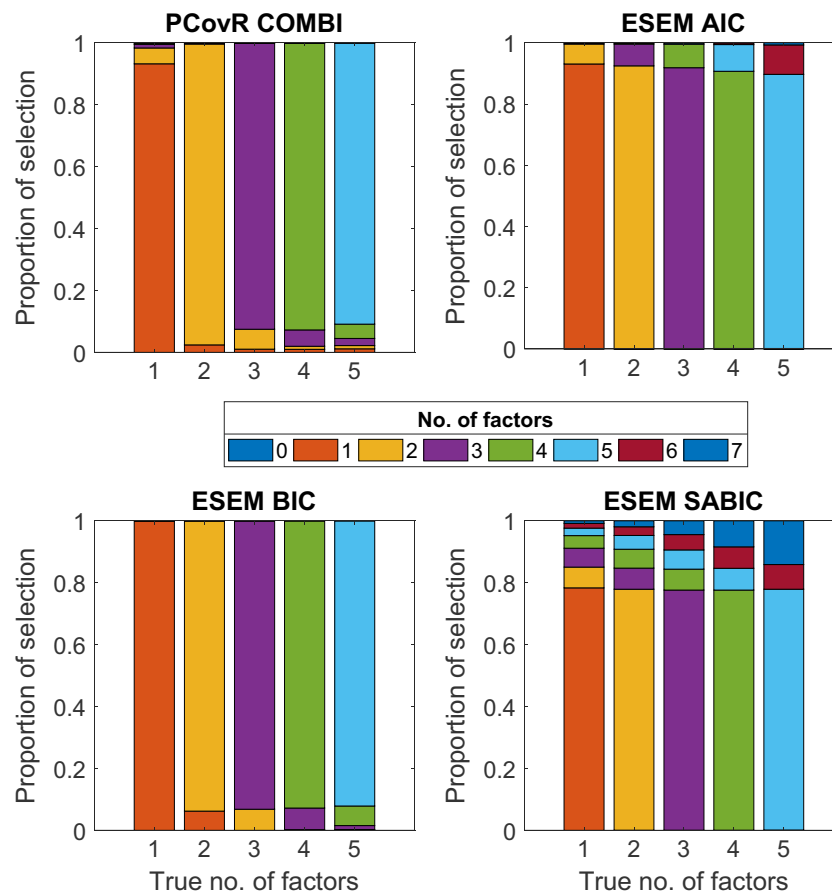
Fig. 2 Numbers of factors selected by the four strategies per the true number of factors. The segments of the bars are ranked such that higher segments indicate a higher number of factors.

(especially with 500 or more observations), but not perfectly identical.

Finally, it should be noted that although differential performance was usually in favor of ESEM (in around 5% of the datasets), the performance for ESEM and PCovR was equally good (or bad) in more than 93% of the datasets (Table 4).

Conclusions Although nonconvergence often occurred during the ESEM analyses, at least one solution could be found for each dataset.

In the majority of the datasets, ESEM (with the AIC, BIC, and SABIC strategies) and PCovR (with the COMBI strategy) selected models that corresponded with the true number of underlying factors. In the case that this number was misspecified, AIC and SABIC had a tendency to select a too-complex model (consisting of noise-driven and/or split factors), whereas BIC and COMBI instead underestimated the number of factors, sometimes yielding models with merged factors. All strategies mainly had difficulties recovering the weakest factors in the unequal-strength conditions if a

Table 3 Percentages of datasets in which the four strategies produced split, merged, or noise-driven factors, or failed to recover a true factor, and the percentages of optimal solutions ($C = .85/C = .95$)

Model Selection Strategy	Splitting	Merging	Noise	Missing (Similar/Equal)	Optimal (Similar/Equal)
PCovR					
COMBI	1%	2%	8%	13%/46%	87%/54%
ESEM					
AIC	2%	0%	13%	8%/41%	85%/56%
BIC	0%	1%	4%	9%/40%	91%/60%
SABIC	6%	0%	25%	13%/45%	73%/47%
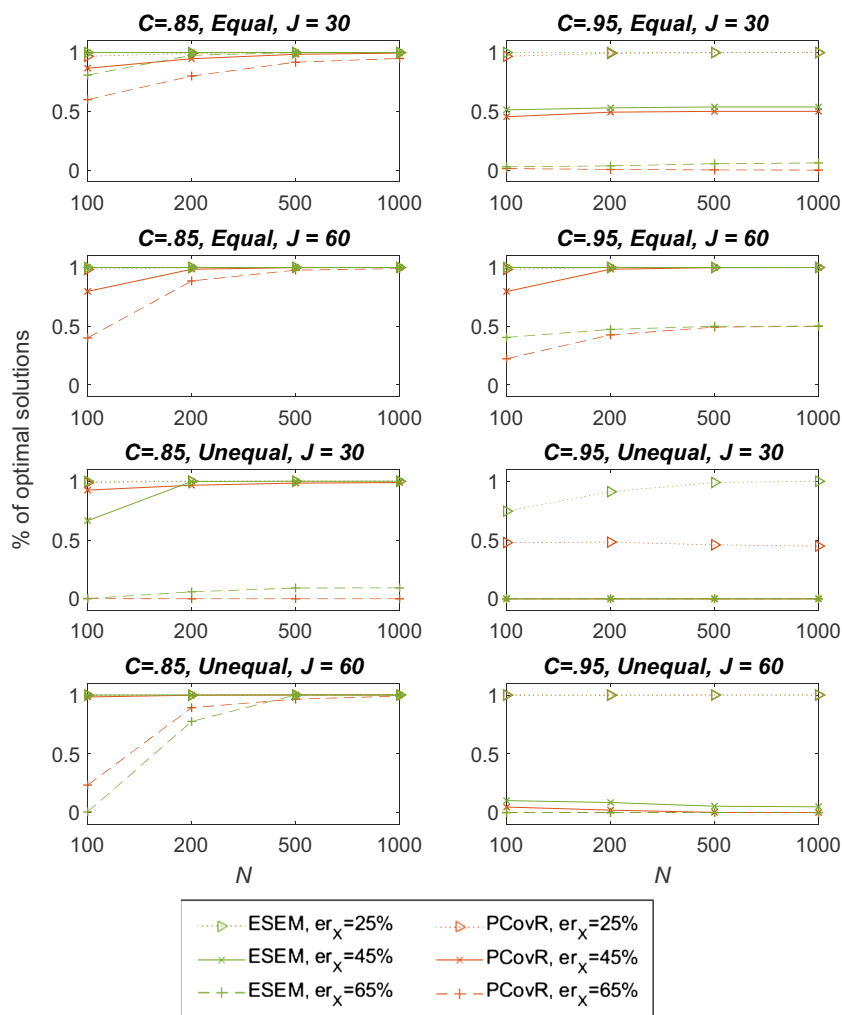
Fig. 3 Effects of the number of observations, the error on **X**, the strength of factors, and the number of predictors on the percentage of optimal solutions for PCovR (orange) and ESEM (green), for cutoff values $C = .85$ and $.95$.

strict cutoff value was used, leading to missing factors in at least 40% of the datasets.

In general, ESEM BIC had the highest percentage of optimal solutions, whereas the lowest performance was found with ESEM SABIC (25% of the SABIC models had noise-driven factors). In more than 93% of the datasets, however, the performance of ESEM BIC was equally as good (or bad) as the performance of PCovR COMBI. In the remainder of the datasets, ESEM BIC usually outperformed PCovR COMBI,

but the specific conditions in which this occurs depended on the cutoff value used.

Simulation Study 2

Data generation, analyses, and evaluation

In the second simulation study, we fixed the number of factors at four. The following data characteristics were manipulated in a full factorial design:

1. The number of predictors J: 24, 48
2. The percentage of error variance in **X**: 25%, 45%, 65%
3. The percentage of error variance in **y**: 25%, 45%, 65%
4. The number of observations N: 100, 200
5. The strength of the factors, expressed in terms of the percentage of variance in **X** accounted for by the four factors: 13%–13%–38%–38%, 17%–17%–33%–33%, 21%–21%–29%–29%, 25%–25%–25%–25% (see Table 5)

Table 4 Proportions of datasets in which ESEM and PCovR yielded optimal solutions

	ESEM	
PCovR	Optimal	Not Optimal
Optimal	85%/54%	1%/<1%
Not optimal	5%/ 6%	8%/40%

The cutoff values used were, respectively $C = .85/C = .95$.

Table 5 Numbers of high-loading variables in the different strength conditions

| Condition	Factor			
	1	2	3	4
J = 24				
13%–13%–38%–38%	3	3	9	9
17%–17%–33%–33%	4	4	8	8
21%–21%–29%–29%	5	5	7	7
25%–25%–25%–25%	6	6	6	6
J = 48				
13%–13%–38%–38%	6	6	18	18
17%–17%–33%–33%	8	8	16	16
21%–21%–29%–29%	10	10	14	14
25%–25%–25%–25%	12	12	12	12

For each cell of the factorial design, 2×3×3×2×4 = 144 in total, 200 datasets were constructed, using a data generation procedure similar to that in Simulation Study 1. The regression weights of the four factors, however, were, respectively, set to .60, 0, .80, and 0. Combined with the strength manipulation (Data Characteristic 5), this resulted in datasets with four factors that were, respectively, weak but relevant, weak and irrelevant, strong and relevant, and strong but irrelevant.

The 28,800 datasets were analyzed with the same strategies described for the first simulation study. The evaluation procedure, however, was slightly adapted: Following Vervloet et al. (2016), irrelevant but true factors were not required for considering a solution to be optimal, although the recovery of these factors was allowed. The recovery of irrelevant factors can give insight into the (parts of the) predictors that are not helpful for explaining the criterion.

Results

Nonconvergence Also in this simulation study, ESEM nonconvergence occurred only for specific numbers of factors, with a pattern similar to that in the first study (Fig. 4). In 21 datasets, however, no convergence was reached for a model with three factors (i.e., a lower number than the number of true factors), and in 104 datasets even the model with four factors (i.e., the number of true factors) failed to converge. The latter datasets usually had only 24 predictors, 65% of error on **X**, and large differences in the strengths of the factors (13%–13%–38%–38%).

Number of factors Figure 5 shows the numbers of factors selected by the different model selection strategies. It can be seen that in the majority of datasets, four factors were selected by each model selection strategy, corresponding to the number of truly underlying factors. The selection of a lower number of
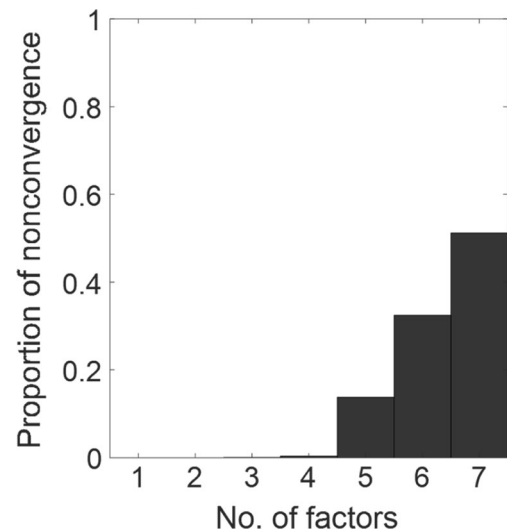
Fig. 4 Proportions of nonconvergence for each considered number of factors.

factors mainly occurred with PCovR COMBI (20% of the datasets) and ESEM BIC (10%), as was the case in Simulation Study 1. Selection of a lower number, however, was not necessarily problematic, given our definition of optimality and the presence of irrelevant factors underlying the datasets. It is alarming, though, that both ESEM AIC and SABIC selected more than four factors in, respectively, 12% and 43% of the datasets, since the data were generated on the basis of four factors only.

Optimality of the obtained solutions Table 6 shows a pattern very similar to that in Table 3 from the first simulation study. ESEM BIC led most often to an optimal solution (45% of the datasets, in the case of *C* = .95), followed by PCovR COMBI (43%) and ESEM AIC (41%).
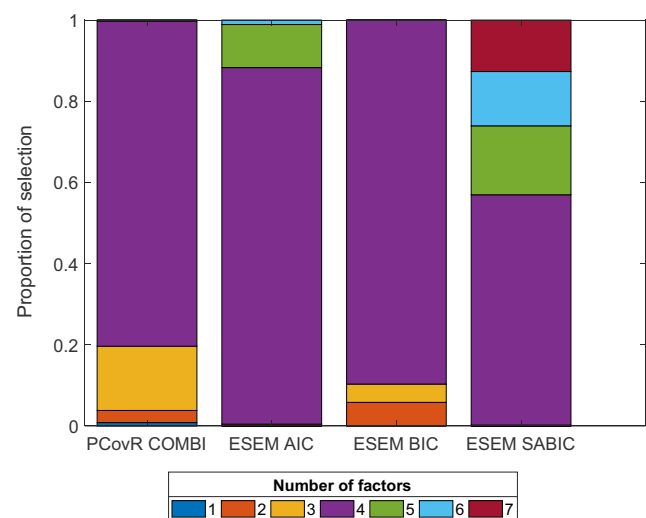
Fig. 5 Proportions of datasets in which specific numbers of factors were selected.

Table 6 Percentages of datasets in which the four strategies produced split, merged, or noise-driven factors or failed to recover a true factor, and the percentages of optimal solutions ($C = .85/C = .95$)

Strategy	Splitting	Merging	Noise	Relevant missing (Similar/Equal)	Optimal (Similar/Equal)
PCovR					
COMBI	2%	5%	10%	13%/57%	84%/43%
ESEM					
AIC	3%	0%	22%	8%/55%	77%/41%
BIC	0%	2%	8%	9%/55%	86%/45%
SABIC	10%	0%	49%	10%/55%	49%/25%

Also in this simulation study, ESEM BIC outperformed AIC and SABIC, so we can focus on this strategy for the comparison between ESEM and PCovR. Again, an ANOVA was performed, for both $C = .85$ and $C = .95$, with the data characteristics as independent variables and difference scores for the percentages of optimality of ESEM and PCovR as the dependent variable. The effects that were found can be seen in Fig. 6.

As long as the error level on **X** was below 65%, the relevant factors were always recovered to some extent (i.e., with $C = .85$). In the case of 65% error, optimal solutions were only found in the easier conditions (200
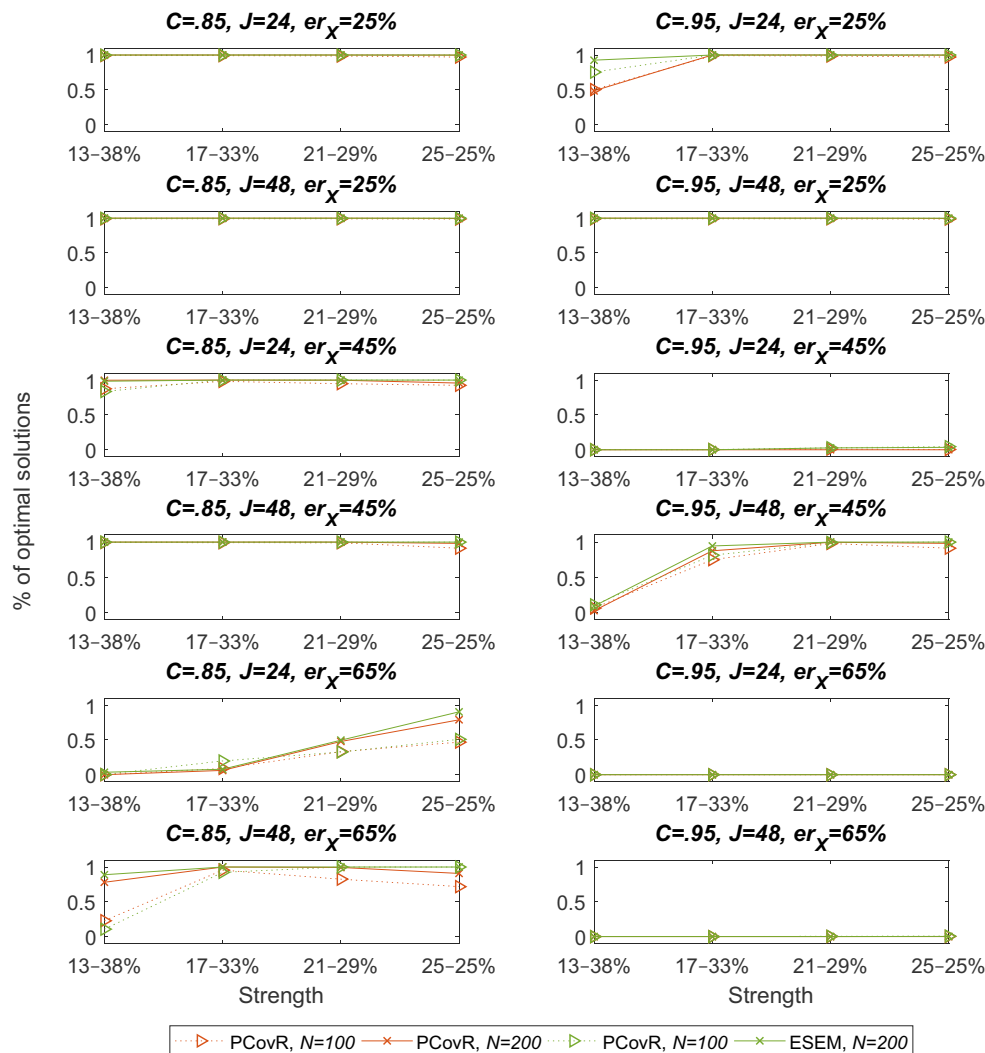
Fig. 6 Effects of the strength of the factors, the error on **X**, the number of observations, and the number of predictors on the percentage of optimal solutions per model selection strategy, for cutoff values $C = .85$ and $.95$.

Table 7 Proportions of datasets in which ESEM and PCovR yield optimal solutions

PCovR	ESEM	
	Optimal	Not optimal
Optimal	82%/43%	2%/<1%
Not optimal	4%/3%	12%/54%

The cutoff values used were, respectively, $C = .85/ C = .95$.

observations, factors with equal strength, 48 predictors . . .). Mainly in these conditions, differential performance can be seen, usually in favor of ESEM. In the case of 48 predictors, 100 observations, and large differences in the strengths of the factors, PCovR outperformed ESEM slightly, but still performs poorly (less than 50% optimal solutions).

When the cutoff $C = .95$ was used, it turns out that in the case of 65% error, no optimal solutions occurred at all. In the case of 45% error on **X**, optimal solutions were found only with 48 predictors and not too large differences in the strengths of the factors, whereas in the case of 25% error, mostly optimal solutions were found for both PCovR and ESEM, except with 24 predictors and large differences in strength. Again, differential performance is mostly seen in those specific conditions.

Given the results in Fig. 6, it is not a surprise that ESEM and PCovR performed equally well in at least 94% of the datasets (Table 7). In the remaining datasets, ESEM usually outperformed PCovR.

Conclusions The general patterns of the first simulation study were replicated in this study. ESEM BIC and PCovR COMBI again showed a tendency to select fewer than the true number of underlying factors. However, in this study, selecting fewer factors was not necessarily problematic, given the presence of irrelevant factors in the datasets. ESEM AIC and SABIC instead selected overly complex models, leading even to 49% of models with noise-driven factors in the case of SABIC. Overall, BIC had the best performance, although PCovR had a slight advantage in a few datasets.

Application

In this section, we reanalyze an empirical dataset collected by Van Mechelen and De Boeck (1990). The dataset contains the scores of 30 Belgian psychiatric patients on 23 psychiatric symptoms and four psychiatric disorders (*toxicomania*, *schizophrenia*, *depression*, and *anxiety disorder*). The patients were examined by 15 psychiatrists, who each provided a binary score for each patient on each symptom and disorder, reflecting its presence (1) or absence (0); we summed these 15 scores. With our PCovR and ESEM analyses, we aimed to unravel whether the presence of subsets of the 23 psychiatric symptoms could predict how many psychiatrists would diagnose someone with *depression*. However, we will also briefly discuss the findings obtained for *toxicomania*, *schizophrenia*, and *anxiety disorder*.

Preprocessing consisted of centering the data (i.e., subtracting the variable means from the original scores) and scaling the scores on each variable to a variance of 1 (see the Model section for principal-covariate regression). The preprocessed data were analyzed with PCovR and ESEM (see the Appendix for an exemplary Mplus input file). Model selection was conducted in the same way as in the simulation studies, using the COMBI strategy for PCovR and the AIC, BIC, and SABIC criteria for ESEM. Again, the numbers of factors that were considered ranged from one to seven. The factors of all models were Varimax-rotated.

The four resulting models differed regarding the number of factors that was selected. PCovR COMBI yielded a model with one factor. Table 8 shows that the lowest AIC, BIC, and SABIC values were found for models with, respectively, five, two, and five factors. This result is consistent with the simulation study finding that AIC and SABIC tend to select more complex models than BIC and PCovR COMBI. It is interesting, though, that ESEM BIC yielded a model with more factors than PCovR COMBI, since ESEM BIC never selected a too-complex model in the simulation studies, whereas this problem sometimes occurred for PCovR COMBI. Note that ESEM did not reach convergence for models with three, six, or seven factors. Although it may seem peculiar that the three-factor model could not converge, whereas the four-factor model did, increasing the maximal

Table 8 AIC, BIC, and SABIC values per number of extracted factors for the ESEM analyses when predicting the criterion *depression*

Model selection procedure	No. of extracted factors						
	1	2	3	4	5	6	7
AIC	2,064	1,965	/	1,935	**1,928**	/	/
BIC	2,165	**2,098**	/	2,128	2,149	/	/
SABIC	1,941	1,802	/	1,698	**1,657**	/	/

The models with the lowest values (in bold) were selected.

Table 9 Loading matrix and regression weights of the psychiatric dataset of Van Mechelen and de Boeck (1990) after analysis with PCovR COMBI and ESEM BIC

	PCovR COMBI	ESEM BIC	
Variable	Factor 1	Factor 1	Factor 2
Disorganized_speech	**− .36**	**− .38**	**.54**
Agitation	.01	− .01	.04
Hallucinations	**− .40**	**− .38**	− .06
Inappropriate	**− .35**	**− .37**	**.60**
Desorientation	.17	.15	.10
Depression	**.97**	**.98**	.07
Fear	.31	**.35**	− .14
Suicide	**.69**	**.70**	.02
Somatic_concern	.14	.08	.14
Narcotics	− .04	− .04	− .30
Antisocial	**− .37**	− .34	− .03
Retardation	.13	.07	**.73**
Social_isolation	.33	.30	**.74**
Routine	.05	.03	**.75**
Alcohol	− .30	− .30	− .03
Negativism	− .27	− .26	.17
Denial	**− .41**	**− .45**	.23
Grandeur	**− .36**	− .34	− .04
Suspicion	− .04	− .05	− .03
Intellectual_obstruction	.03	− .05	.34
Impulse_control	− .08	− .07	.17
Social_leveling	.01	− .05	**1.00**
Occupational_dysfunction	.23	.17	**.63**
Depression	**.99**	**.97**	.06

Loadings with absolute values higher than .35 are shown in bold.

number of iterations to 50,000 did not help. A possible reason for this nonconvergence is the negative residual variance that Mplus yields for the predictor *antisocial* for the three-factor model. Because the AIC and SABIC models consisted of mostly irrelevant factors for understanding depression (only one regression weight had an absolute value higher than .20), and because both strategies have been shown to be less successful in the simulation studies, we will focus on the results obtained with PCovR COMBI and ESEM BIC in the remainder of this section.

Table 9 shows the loadings and the regression weights of both models. It is striking that the first ESEM factor is almost identical to the one PCovR factor (with a Tucker congruence of .99). Moreover, this factor is the only factor that has a high regression weight and that therefore can be considered to be a relevant factor for predicting the diagnosis of depression. For this reason, the PCovR and ESEM models can be seen as equivalent, which corresponds with the finding in the simulation studies that both methods mostly succeeded in retrieving equally optimal solutions. When inspecting the predictors that

load highly on the relevant factor, mainly variables that indicate depressive symptoms have positive loadings, and variables that indicate inappropriate behavior have negative loadings. The second ESEM BIC factor can be labeled *Social Impairment*. Note that several predictors (e.g., agitation and disorientation) have close-to-zero loadings, implying that they are considered to be unimportant for predicting the criterion according to the models.

When using *anxiety disorder* and *schizophrenia* as the criterion variable, the results are very similar, whereas the factor structure for *toxicomania* looks completely different. For both *anxiety disorder* and *schizophrenia*, again only one factor was found by PCovR COMBI, which was also picked up by the ESEM models along with a few irrelevant factors. Interestingly, some variables consistently load high on this single factor across the analyses with different criteria (e.g., disorganized speech, hallucinations, inappropriate behavior, depression, suicide, and denial), suggesting that the three syndromes are partly characterized by a common symptom skeleton. This finding is in line with the high comorbidities of anxiety disorder and depression (Hirschfeld, 2001) and schizophrenia and depression (Buckley, Miller, Lehrer, & Castle, 2009). Yet, for other symptoms the loadings varied across the analyses, indicating that the presence or absence of these symptoms was more tightly linked to one specific syndrome (e.g., the absence of antisocial and grandeur for *depression*). The latter finding illustrates that PCovR and ESEM both extract factors that explain variance in both the predictor and criterion blocks. Also, the loadings on the relevant factor were almost identical across both analyses, which is in line with previous research that has shown that PCA and factor analysis often yield similar interpretations for empirical data (Ogasawara, 2000; Velicer & Jackson, 1990).

Discussion

In this final section, we reflect on the results of the simulation studies and the illustrative application. Subsequently we propose some directions for future research, and we end with a brief conclusion.

Summary and reflections

In this article we have compared ESEM and PCovR, which are, respectively, a factor- and a component-based regression method. Both methods can shed light on the shared effects of the predictors on the criterion, by capturing the main information in a limited number of dimensions (called either *factors* or *components*). ESEM as well as PCovR simultaneously optimizes reduction and prediction, and these methods are therefore closely related. However, both methods stem from different traditions, and their performance had not yet been

compared. Therefore, we performed two simulation studies in which we examined the numbers of factors retained by each method, and when and how often optimal solutions were found.

While inspecting the number of factors, we noticed in both simulation studies a tendency for AIC and SABIC to select an overly complex model, because of splitting one or more true underlying factors and/or including noise-driven factors. Bulteel, Wilderjans, Tuerlinckx, and Ceulemans (2013) pointed out that this tendency of AIC has already be shown in the context of many other methods when the sample size is large. Indeed, for large sample sizes, the AIC assumption that each observation contains new information about the underlying model becomes unrealistic.

ESEM BIC performed best in both simulation studies: It selected the true number of underlying factors most often and yielded the highest proportion of optimal solutions. Moreover, when using a strict cutoff score for indicating congruence with the true factors, BIC was never outperformed by AIC and SABIC, and was outperformed only in less than 1% of the datasets by PCovR. Most of the datasets in which PCovR outperformed ESEM BIC had 48 predictors, 100 observations, and large differences in the strength of the factors. Note, however, that PCovR had outperformed ESEM in more than 20% of the datasets in a pilot study with 48 predictors and only 50 observations.

Although ESEM (when using BIC to decide on the number of factors) in general has better performance than PCovR for typical behavioral datasets, it should be noted that in 93% and 94% of the datasets in, respectively, the first and second simulation studies, ESEM BIC and PCovR COMBI performed equally well. It is therefore not surprising that the real-data application led to equivalent ESEM and PCovR models. Indeed, both methods seemed successful at finding factors that explain variance in both the predictor and criterion blocks, whereas other dimension reduction methods only focus on prediction (e.g., PLS) or only on reduction (e.g., PCR). ESEM and PCovR will often be able to extract factors that are both interpretable and relevant for the criterion. The results for the illustrative application demonstrate that the criterion under consideration indeed alters the extracted factors.

Because of the similar performance of ESEM and PCovR, other aspects can be taken into account when choosing between these two methods as an applied researcher. Firstly, although the convergence problems of ESEM (within a reasonable number of iterations) may at first sight seem a limitation, on the basis of our results, nonconvergence may be interpreted as a signal that the requested number of factors is too high, and could therefore prevent researchers from selecting an overly complex model. Secondly, PCovR always has a closed-form solution and therefore will never yield a local minimum. ESEM relies on an iterative estimation procedure, and thus may end in a local minimum. Thirdly, an R

package is available for conducting PCovR, implying that it can be run for free on any computer, whereas ESEM can only be performed with Mplus, which is a commercial software package. The R package PCovR not only provides the actual PCovR algorithm but also assists with the other steps of the analysis: model selection, preprocessing, and so forth. Several options are built in, but since R is code-based software, the code can easily be adapted. Fourthly, ESEM, on the other hand, comes with fit statistics, standard errors, p values for the loadings, and so forth, which is an important advantage when one wants to make inferences about which loadings differ significantly from zero. Finally, we have focused on single-block data until now, but both PCovR and ESEM have extensions for analyzing multiblock data (i.e., data in which the observations are nested in higher levels), called *principal-covariate clusterwise regression* (PCCR; Wilderjans, Vande Gaer, Kiers, Van Mechelen, & Ceulemans, 2017) and *multigroup ESEM*, respectively. Because the two extensions have different focuses, these extensions can be considered complementary, and it depends on a researcher's questions which extension will be more appropriate. Specifically, multigroup ESEM can be used to test whether factor loadings significantly differ across different subgroups of observations, whereas PCCR can be used to infer which subgroups differ with respect to the relevance of the underlying factors.

Future directions

Although PCovR rarely outperformed ESEM in the presented simulation studies, we expect that PCovR might outperform ESEM in datasets that are less typical for the field of behavioral sciences. For example, several studies have claimed that multiple observations are needed per variable when applying factor-based methods (for an overview, see Velicer & Fava, 1998). Thus, PCovR can be expected to outperform ESEM if the number of predictors approaches the number of observations. MacCallum, Widaman, Preacher, and Hong (2001) demonstrated that low observations-to-variables ratios are especially problematic for factor-based methods in the case of weak factors. Indeed, although ESEM was able to find a model with an interpretation equivalent to that of the model found by PCovR for our application that had an observations-to-variables ratio of only 30:23, PCovR performed better than ESEM in a pilot study with 50 observations and 48 predictors, when some of the relevant factors explained less than 10% of the variance in $\mathbf{X}$.

It would also be useful to compare the performance of PCovR and ESEM when analyzing data that are not normally distributed. Because PCovR estimation does not rely on the assumption of normality, we expect PCovR to outperform ESEM in such cases if the default estimator (in Mplus) for ESEM is used, but other ESEM estimation procedures are available that attempt to deal with violations of normality.
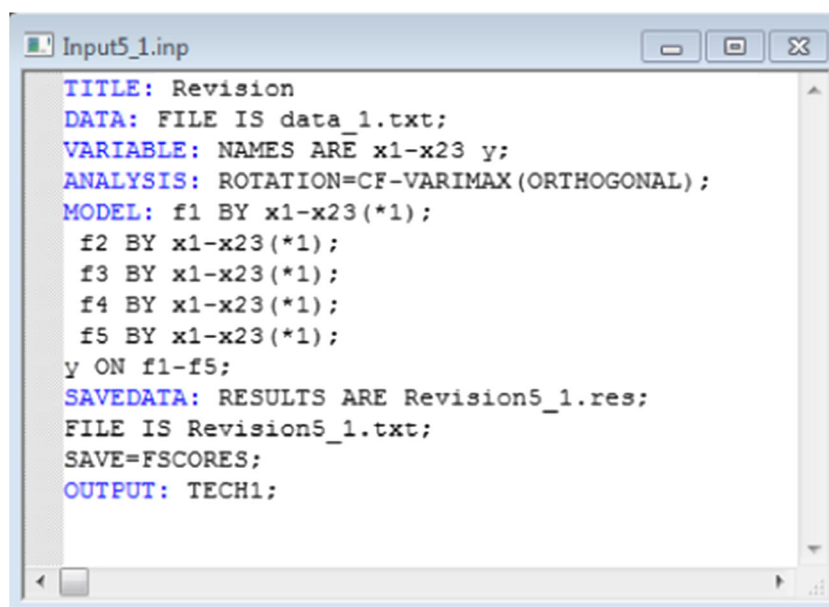
Conclusion

We conclude that although ESEM is usually used as a confirmatory tool, we have shown that it is very valuable for exploratory research, as well. It can compete with state-of-the-art exploratory component-based methods for typical behavioral datasets.

Appendix: Example of Mplus input file

For the ESEM analysis with five factors, the following Mplus input file was used for the real-data application in the Application section:

Fig. 7 Mplus input file of real-data application.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *AC-19*, 716–723. doi: https://doi.org/10.1109/TAC.1974.1100705

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305–314. doi:https://doi.org/10.1037/0033-2909.110.2.305

Borsboom, G., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Browne, M. W. (1972a). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 207–212.

Browne, M. W. (1972b). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 115–120.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150.

Buckley, P. F., Miller, B. J., Lehrer, D. S., & Castle, D. J. (2009). Psychiatric comorbidities and schizophrenia. *Schizophrenia Bulletin*, *35*, 383–402. doi:https://doi.org/10.1093/schbul/sbn135

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Using raw VAR regression coefficients to build networks can be misleading. *Multivariate Behavioral Research*, *51*, 330–344. doi:https://doi.org/10.1080/00273171.2016.1150151

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., & Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods*, *45*, 782–791. doi: https://doi.org/10.3758/s13428-012-0293-y

Carroll, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, *18*, 23–38.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. doi:https://doi.org/10.1207/s15327906mbr0102_10

Ceulemans, E., & Kiers, H. A. L. (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*, 133–150. doi:https://doi.org/10.1348/000711005X64817

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coltman, T., Devinney, T. M., Midgley, D. F., & Veniak, S. (2008). Formative versus reflective measurement. *Journal of Business Research*, *61*, 1250–1262.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika*, *40*, 183–195.

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

Doove, L. L., Wilderjans, T. F., Calcagnì, A., & Van Mechelen, I. (2017). Deriving optimal data-analytic regimes from benchmarking studies. *Computational Statistics and Data Analysis*, *107*, 81–91.

Filzmoser, P., Liebmann, B., & Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics*, *23*, 160–171.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*, 265–275. doi:https://doi.org/10.1037/0033-2909.103.2.265

Gurden, S. (n.d.). *Multiway covariates regression: Discussion facilitator.* Unpublished manuscript.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction.* New York, NY: Springer.

Hirschfeld, R. M. (2001). The comorbidity of major depression and anxiety disorders: Recognition and management in primary care. *Primary Care Companion to the Journal of Clinical Psychiatry*, *3*, 244–254.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, *32*, 1–50.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55–67.

Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, *66*, 289–306. doi:https://doi.org/10.1007/BF02294840

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, *67*, 7–19. doi:https://doi.org/10.1007/BF02294706

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1–19.

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

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200. doi:https://doi.org/10.1007/BF02289233

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

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347–365. doi:https://doi.org/10.1207/S15327906MBR3403_3

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*, 57–64. doi:https://doi.org/10.1027/1614-2241.2.2.57

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611–637.

Marsh, H. W., Liem, G. A., Martin, A. J., Nagengast, B., & Morin, A. J. (2011). Methodological-measurement fruitfulness of Exploratory Structural Equation Modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, *29*, 322–346.

Marsh, H. W., Morin, J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluation of university teaching. *Structural Equation Modeling*, *16*, 439–476.

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika*, *65*, 167–185. doi:https://doi.org/10.1007/BF02294372

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. doi:https://doi.org/10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343.

Tibshirani, R. (1996). Regression, shrinking and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, *58*, 267–288.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421–459.

Van Mechelen, I., & De Boeck, P. (1990). Projection of a binary criterion into a model of hierarchical classes. *Psychometrika*, *55*, 677–694.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231–251. https://doi.org/10.1037/1082-989X.3.2.231, see https://www.researchgate.net/publication/232509045_Effects_of_Variable_and_Subject_Sampling_on_Factor_Pattern_Recovery

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research*, *25*, 97–114.

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*, 1–14.

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.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods*, *17*, 228–243.

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

Wilderjans, T. F., Vande Gaer, E., Kiers, H. A., Van Mechelen, I., & Ceulemans, E. (2017). Principal covariates clusterwise regression (PCCR): Accounting for multicollinearity and population heterogeneity in hierarchically organized data. *Psychometrika*, *82*, 86–111.

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*, 735–743.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis.* New York, NY: SUNY Press.