British Journal of Mathematical and Statistical Psychology

Bootstrap confidence intervals for principal covariates regression

--Manuscript Draft--

Article Type:	Article
Manuscript Number:	BJMSP.19.0085R2
Full Title:	Bootstrap confidence intervals for principal covariates regression
Corresponding Author:	Paolo Giordani Universita degli Studi di Roma La Sapienza Rome, Rome ITALY
Corresponding Author E-Mail:	paolo.giordani@uniroma1.it
Order of Authors (with Contributor Roles):	Paolo Giordani (Writing – original draft; Writing – review & editing)
	Henk A.L. Kiers (Writing – original draft; Writing – review & editing)
Funding Information:	
Keywords:	Principal Covariate Regression; confidence intervals; Bootstrap.
Manuscript Classifications:	Bootstrap Inference; Confidence Intervals; Multivariate statistics; Principal Component Analysis
Abstract:	Principal Covariate Regression (PCOVR) is a method for regressing a set of criterion variables with respect to a set of predictor variables when the latter ones are many and/or collinear. This is done by extracting a limited number of components that simultaneously synthetize the predictor variables and predict the criterion ones. So far, no procedure has been offered for estimating statistical uncertainties of the obtained PCOVR parameter estimates. The present article shows how this goal can be achieved, conditionally on the model specification, by means of the bootstrap approach. Four strategies for estimating bootstrap confidence intervals (CIs) are derived and their statistical behavior in terms of coverage is assessed by means of a simulation experiment. Such strategies are distinguished by the use of the varimax and quartimin procedures and by the use of Procrustes rotations of bootstrap solutions towards the sample solution. In general, the four strategies showed appropriate statistical behavior with coverage tending to the desired level for increasing sample sizes. The main exception involved strategies based on the quartimin procedure in cases characterized by complex underlying structures of the components. The appropriateness of the statistical behavior was higher when the proper number of components was extracted.
Additional Information:	
Question	**Response**
If you have any potentially competing interests to declare, please enter them in the box below. If you have no interests to declare, please enter 'none'.	None
Please declare any prior use or publication of any part of this article. For example: overlap with a thesis, posting on a pre-print server, using a shared data set as part of a long-term project, overlap with a conference publication, an extension of earlier work. Any overlap not declared and later discovered can result in the manuscript being withdrawn from consideration.	No prior use or publication of any part of this article.

Please specify the word count of your manuscript (excluding the abstract, tables, figures and references).	about 10,000
Do you or your co-authors use Twitter for academic purposes? (If you do and your paper is accepted, we will 'mention' you when we share a link to the online article.)	no

Dear Associate Editor,

we thank you for all the suggestions, comments and remarks. The points you raised and the corrections we made are discussed below (your points are in a different font).

Thank you for submitting a revised version of your manuscript entitled "Bootstrap confidence intervals for principal covariates regression" to the British Journal of Mathematical and Statistical Psychology. I now have received comments from two reviewers with expertise in the area and have read your paper myself. The reviewers' comments are appended below.

As you can see, both reviewers consider the manuscript to have improved following the changes you applied, but both reviewers feel further improvement is necessary before the manuscript is ready to be published in the British Journal of Mathematical and Statistical Psychology. Therefore, I again invite you to attend to the matters identified by the reviewers, address the issues raised and when appropriate changes have been made, submit the revised manuscript. On receiving this third version of the manuscript, I shall send it to the same reviewers to obtain their opinion as whether it is suitable for publication in the British Journal of Mathematical and Statistical Psychology. So please give sufficient attention to the issues raised and the modifications made.

As before, I would like to take this opportunity to thank you for considering BJMSP as a possible outlet for your work, and I hope that you will continue to do so in future.

We have revised the manuscript following all the matters identified by the Reviewers as you can see in our point-to-point replies to the two Reviewers. For easy comparison of versions, we made all changes using the Word option "track changes" (including slight modifications that we found desirable ourselves, without being requested by the Reviewers).

Dear Reviewer 1,

we thank you for all the suggestions, comments and remarks. The points you raised and the corrections we made are discussed below (your points are in a different font). Note that, for easy comparison of versions, in the revised manuscript we made all changes using the Word option "track changes" (including slight modifications that we found desirable ourselves, without being requested by the Reviewers).

Reviewer #1: First I would like to thank the Authors for heavily revising and substantially improving the paper! There are still many issues mainly regarding how they write. Below I give a couple of examples but the manuscript need to go through to make sure the text is good enough.

We have carefully reconsidered the entire manuscript by conducting our own check in addition to your comments and those made by the other Reviewer.

1."If so, X is (close to) rank deficient implying that the estimates of the regression coefficients grow as well as the variances of the corresponding estimators." Wording, "grow" implies that something happens which is not the case as X is fixed. Rewrite stating that if X becomes closer to be rank deficient then...

We rewrote it by, in particular, replacing the term "grow" by "tend to be large".

2.I have problem understanding the following: "The primary goal of RRR is the prediction of Y by looking for the components of X such that the prediction error of Y is minimized, even if they may synthesize X in a poor way. We must note that such predictor components may not explain the criterion variables reasonably well. This occurs when the variables in X have very limited predictive power." The problem I have is that if X has limited predictive power then it doesn't matter which method you use. So, how is this a drawback of RRR and implies an advantage of PCOVR?

To avoid ambiguities, we decided to remove this sentence so that the objective of PCOVR is seen as a compromise between those of PCR and RRR.

3."When the multi-normality assumption is violated, as is often the case, analytic results are no longer available." I fully understand what you try to say here but there is a huge literature on misspecification and how this affects parameter estimates and standard errors, see e.g. White (1982) for consistency conditions etc. There are two main reasons for using the bootstrap. The first is that there is no need to specify the distributions needed to calculate the proper version of the sandwich estimator for correct inference, and that the bootstrap distribution often is a better approximation to the small sample distribution than the standard asymptotic approximation.

At the beginning of Section 3.1, we mentioned the issue of misspecification and cited White (1982). Later, just after formula (11), we added some comments in order to motivate the use of the bootstrap.

4."Several data sets have been randomly generated in order to assess the quality of the different strategies for computing bootstrap...". Please rewrite as it is obvious when doing a simulation study to generate several data sets.

Done

5.The Monte Carlo simulation setup needs to be better motivated. The setup is well explained but a reader needs to know in what sense the setup is empirically relevant, i.e. when doing empirical research what lessons can be learned from this simulation? Further, they need to motivate the number of replicates and similar choices.

We did this. In particular, in Section 4, in the set-up of the two simulation studies, we now motivated all the choices we made for the levels of the design variables. In the Discussion (Section 5), we discussed the relevance of the studies by explaining why they are helpful when doing empirical research. Essentially, we think that “The simulation experiment offered recommendations on the computation of CIs for the PCOVR parameters when doing empirical research. In fact, some differences in the statistical behavior of the CIs emerged with respect to the adopted variant, the parameter matrix, and the characteristics of the data. In some cases, the obtained CIs appeared to be reasonably good estimates, in some others, the quality degraded” (see page 29)

6.Dufour and Kiviet (1998) shows that the number of replicates should be selected such that $(B+1)\alpha$ should be an integer. As B=1000 this makes minor difference.

The use of B = 1000 bootstrap sample was chosen according to Timmerman et al. (2007). We specified it in the revised version of the manuscript and we cited the result by Dufour and Kiviet (1998).

Dufour, J-M. and Kiviet, J.F. (1998) Exact inference methods for first-order autoregressive distributed lag models, Econometrica 66 (1), 79-104.

White, H. (1982) Maximum likelihood estimation of misspecified models, Econometrica 50 (1), 1-25.

Dear Reviewer 2,

we thank you for all the suggestions, comments and remarks. The points you raised and the corrections we made are discussed below. (your points are in a different font). Note that, for easy comparison of versions, in the revised manuscript we made all changes using the Word option "track changes" (including slight modifications that we found desirable ourselves, without being requested by the Reviewers).

Reviewer #2: BJMSP.19.0085_R1 Bootstrap confidence intervals for principal covariates regression

The paper has been improved considerably compared to the previous version. The theoretical background and the bootstrap strategies are much more clearly explained. The simulation studies seem proper and offer useful insight into the behavior of the four strategies examined. I like the addition of Simulation study number 2, assessing the behavior under the extraction of fewer and equal numbers of components than present at the population level.

Thanks

To facilitate the application of the strategies discussed, it is necessary that code is made available to estimate the CIs for PCOVR, e.g., in R and/or Matlab. Further, in view of the transparency it would be good to make publicly available (or alternatively upon request) the code that was used for the simulation study and analysis.

All the analyses have been carried out by using R. In the revised version of the manuscript, we explicitly stated it in several occasions (pages 6, 15, 17). We also added that the code and the functions to obtain CIs are available upon request.

Specific comments

p. 1, Abstract: Please be explicit in what the four strategies for estimating bootstrap confidence intervals (CIs) entail. Make also explicit what the 'few exceptions' are to the appropriate statistical behavior.

Done

p. 3, l. 9: The phrasing '…several alternative methods are available…' suggest to me alternatives to the bootstrap. Please rephrase such that it is clear that you mean are variants of the bootstrap (or I am mistaken: make explicit what alternatives you mean).

We meant variants of the bootstrap. To this purpose, we rewrote the sentence.

p. 6, Section 2.1: I missed what software was used to perform the PCOVR analysis. I have the same remark for the simulation study.

As clarified above, we used R, in particular the package PCovR.

p. 8, l. 14 (and at other places): Replace 'multi-normality' by 'multivariate normality' (i.e., I do not know the term, and even Google does not help me further).

Done

p. 21, l. -3 to -8: The explanation of the unsatisfactory behavior of the QSPB does not belong in the Results section, but to the Discussion section. Please remove here, and devote attention to this in the Discussion section.

We moved this explanation to Section 5 (Discussion section).

p. 23, l. 7: Replace 'forth' by 'fourth'.

Thanks

p. 23, l. 14-15: Please clarify what the figures exactly mean, for example by explicating what '4% vs 46%' implies precisely, and to what condition this complies. To me it is not clear exactly what this entails - maybe I would after consulting Vervloet et al. (2016), but the text should be understandable without that.

We added an example in order to better explain their meaning (see page 24)

p. 24, l. 17-19: I do not understand the sentence 'Namely … level' - please rephrase. I guess it should be something like '… parameter matrices, except for p_y for which…'?

Done: the term “except” was missing.

p. 26-28: The Discussion section is a bit meagre. The summary of the aims, the study and the results is rather long. In my view, this should be shortened. Further, I miss critical reflections on 1) the smaller accuracy of the CIs for W and WP_Y, compared to P_X and P_Y; 2) the smaller accuracy of QSPB, especially in the complex condition.

We reconsidered the Discussion section by shortening the summary and focusing in more detail on critical reflections. To this purpose, we moved here the explanation about QSPB (point 2). Concerning point 1, we believe that the smaller accuracy of $\mathbf{WP}_Y$ can be explained by the fact that it indeed refers to two sources of uncertainty, i.e., those for $\mathbf{W}$ and $\mathbf{P}_Y$. The smaller accuracy of $\mathbf{W}$ is more obscure to us: it deserves future research.

`p. 52, Fig. S7: The figure is incorrectly displayed. That is, only pink filled circles are visible in my figure - which is unlike the description.`

Corrected

BOOTSTRAP CONFIDENCE INTERVALS FOR PRINCIPAL COVARIATES REGRESSION

BOOTSTRAP CI'S FOR PCOVR

Paolo Giordani*[1], and Henk A.L. Kiers[2]

[1] Department of Statistical Sciences, Sapienza University of Rome

[2] Department of Psychology, University of Groningen

*Corresponding author information: Paolo Giordani, P.le Aldo Moro, 5, 00185 Rome, Italy, (e-mail: paolo.giordani@uniroma1.it).

Abstract:

Principal Covariate Regression (PCOVR) is a method for regressing a set of criterion variables with respect to a set of predictor variables when the latter ones are many and/or collinear. This is done by extracting a limited number of components that simultaneously synthetize the predictor variables and predict the criterion ones. So far, no procedure has been offered for estimating statistical uncertainties of the obtained PCOVR parameter estimates. The present article shows how this goal can be achieved, conditionally on the model specification, by means of the bootstrap approach. Four strategies for estimating bootstrap confidence intervals (CIs) are derived and their statistical behavior in terms of coverage is assessed by means of a simulation experiment. Such strategies are distinguished by the use of the varimax and quartimin procedures and by the use of Procrustes rotations of bootstrap solutions towards the sample solution. In general, the four strategies showed appropriate statistical behavior with coverage tending to the desired level for increasing sample sizes. The main exception involved strategies based on the quartimin procedure in cases characterized by complex underlying structures of the components. The appropriateness of the statistical behavior was higher when the proper number of components was extracted.

Keywords:

Principal Covariate Regression, Confidence intervals, Bootstrap.

Data availability statement:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements:

We thank the three anonymous reviewers whose comments and suggestions greatly improved the quality of the paper.

Bootstrap confidence intervals for principal covariates regression

Abstract: Principal Covariate Regression (PCOVR) is a method for regressing a set of criterion variables with respect to a set of predictor variables when the latter ones are many and/or collinear. This is done by extracting a limited number of components that simultaneously synthetize the predictor variables and predict the criterion ones. So far, no procedure has been offered for estimating statistical uncertainties of the obtained PCOVR parameter estimates. The present article shows how this goal can be achieved, conditionally on the model specification, by means of the bootstrap approach. Four strategies for estimating bootstrap confidence intervals (CIs) are derived and their statistical behavior in terms of coverage is assessed by means of a simulation experiment. ~~In general, apart from a few exceptions~~Such strategies are distinguished by the use of the varimax and quartimin procedures and by the use of Procrustes rotations of bootstrap solutions towards the sample solution. In general, the four strategies showed appropriate statistical behavior with coverage tending to the desired level for increasing sample sizes. The main exception involved strategies based on the quartimin procedure in cases characterized by complex underlying structures of the components. The appropriateness of the statistical behavior was higher when the proper number of components was extracted.

Keywords: Principal Covariate Regression, Confidence intervals, Bootstrap.

1. Introduction

There exist several methods for analyzing the dependence between two sets of variables, say **Y** and **X**, observed on the same group of units. The simplest strategy is to perform multivariate linear regression. However, such an approach may lead to poor results, especially when the number of predictor variables in **X** is large, due to a high risk of multicollinearity of the predictor variables. If so, **X** is (close to) rank deficient implying that the estimates of the regression coefficients ~~grow~~tend to be large as well as the variances of the corresponding estimators. For this reason, alternative methods have been developed. At least two main classes of methods can be distinguished (see, e.g., Hastie, Tibshirani, & Friedman, 2001, pp. 55-75). The first one involves regularization methods such as the well-known Ridge regression, proposed by Hoerl, & Kennard (1970) following Tikhonov (1943). These techniques consider different regularizations of the regression coefficients in order to prevent the estimates and the variances from growing. The second class involves the use of dimension reduction techniques in order to reduce the space of the predictor variables by means of a limited number of underlying components. For an overview, the interested reader may refer to De Jong, & Kiers (1992), Kiers, & Smilde (2007) and Vervloet (2017). A popular technique belonging to this class is Principal Component Regression (Jolliffe, 1982), briefly PCR, where components are firstly extracted from **X** and then the variables in **Y** are regressed on the scores of these components. On the one hand, this solves the multicollinearity problem because the components are orthogonal. On the other hand, it is not guaranteed that the components explain the criterion variables in **Y** reasonably well, hence the prediction error of **Y** may be large. An alternative strategy is represented by Reduced-Rank Regression (Anderson, 1951; Anderson, 1958; Izenman, 1975), briefly RRR. The primary goal of RRR is the prediction of **Y** by looking for the components of **X** such that the prediction error of **Y** is minimized, even if they may synthesize **X** in a poor way. ~~We must note that such predictor components may not explain the criterion variables reasonably well. This occurs when the variables in **X** have very limited predictive power.~~

~~Both~~Therefore, both PCR and RRR extract components, but differ in the way the components are found. The former approach gives components accounting for the variance of **X** at best (regardless of the correlation with **Y**), the latter one gives components correlated with **Y** as much as possible (regardless the explained variance of **X**). In order to take into account both the objectives, Principal COVariates Regression (De Jong, & Kiers, 1992), briefly PCOVR, has been proposed. In PCOVR, the components are found by minimizing a linear combination of the two previously-described criteria. In this way, the PCOVR components are able to summarize **X** and predict **Y**.

The effectiveness of PCOVR, also in comparison with its potential competitors, has been shown in several papers. See, for instance, Kiers, & Smilde (2007) and Vervloet, Van Deun, Van den Noortgate, & Ceulemans (2016). To judge to what extent the PCOVR estimates can be generalized to the population, inferential statistics are needed. In this paper, we examine a bootstrap procedure to estimate confidence intervals for PCOVR parameters. The use of bootstrapping is motivated by the fact that it works well in Principal Component Analysis and, therefore, we expect the same for PCOVR. In principal component methods, several ~~alternative methods~~variants of the bootstrap are available; see, e.g., Timmerman, Kiers, & Smilde (2007). The most relevant ones are recalled and applied to the PCOVR domain; furthermore, a simulation experiment is carried out in order to assess their quality. The paper is structured as follows. In the next section, PCOVR is presented. Section 3 focuses on some strategies for estimating confidence intervals of the PCOVR parameter estimates. The results of the simulation experiment are reported in Section 4. A final discussion in Section 5 ends the paper.

2. Principal COVariates Regression

Let **Y** and **X** be two matrices of order ($N \times K$) and ($N \times J$), respectively, where N denotes the number of units and K and J (usually, but not necessarily, $K < J$) are the number of criterion variables and that of predictor variables, respectively. In PCOVR, R ($< J$) components, expressed as a linear combination of **X**, are sought:

$$\mathbf{T} = \mathbf{XW}, \tag{1}$$

where $\mathbf{T}$ is the component score matrix of order ($N \times R$) and $\mathbf{W}$ is the component weight matrix of order ($J \times R$). The component scores in $\mathbf{T}$ play the role of explaining both $\mathbf{Y}$ and $\mathbf{X}$, namely,

$$\mathbf{Y} = \mathbf{TP}_{\mathbf{Y}} + \mathbf{E}_{\mathbf{Y}}, \tag{2}$$

$$\mathbf{X} = \mathbf{TP}_{\mathbf{X}} + \mathbf{E}_{\mathbf{X}}, \tag{3}$$

where $\mathbf{P}_{\mathbf{Y}}$ of order ($R \times K$) is the matrix of the regression weights for the K criterion variables on the R components and $\mathbf{P}_{\mathbf{X}}$ of order ($R \times J$) is the matrix of the component loadings for the J predictor variables on the components. Finally, $\mathbf{E}_{\mathbf{Y}}$ and $\mathbf{E}_{\mathbf{X}}$ are the error matrices for $\mathbf{Y}$ and $\mathbf{X}$, respectively.

The parameter matrices $\mathbf{P}_{\mathbf{Y}}$, $\mathbf{P}_{\mathbf{X}}$ and $\mathbf{T}$ (and implicitly $\mathbf{W}$ according to (1)) are estimated by minimizing a linear combination of the sum of squares of $\mathbf{E}_{\mathbf{Y}}$ and $\mathbf{E}_{\mathbf{X}}$, expressed as:

$$f(\mathbf{P}_{\mathbf{Y}}, \mathbf{P}_{\mathbf{X}}, \mathbf{T}) = (1-\alpha)\frac{\|\mathbf{Y}-\mathbf{TP}_{\mathbf{Y}}\|^2}{\|\mathbf{Y}\|^2} + \alpha\frac{\|\mathbf{X}-\mathbf{TP}_{\mathbf{X}}\|^2}{\|\mathbf{X}\|^2}, \tag{4}$$

where the symbol $\|\cdot\|$ denotes the Frobenius norm of matrices and $\alpha \in [0, 1]$ is a tuning parameter for the variances of $\mathbf{X}$ and $\mathbf{Y}$ explained by $\mathbf{T}$. When $\alpha = 0$, PCOVR reduces to RRR, whilst setting $\alpha = 1$ leads to PCR. The automatic selection of α has been widely investigated in several papers. See, De Jong, & Kiers (1992), Kiers, & Smilde (2007), Vervloet, Van Deun, Van den Noortgate, & Ceulemans (2013), and Vervloet, Van Deun, Van den Noortgate, & Ceulemans (2016).

Given α and R, the estimation of the parameter matrices can be found in closed form. To do it, we observe that (4) can be rewritten as

$$f(\mathbf{P_Y}, \mathbf{P_X}, \mathbf{T}) = \left\| \begin{bmatrix} \frac{(1-\alpha)^{1/2}}{\|\mathbf{Y}\|}\mathbf{Y} \\ \frac{\alpha^{1/2}}{\|\mathbf{X}\|}\mathbf{X} \end{bmatrix} - \mathbf{T} \begin{bmatrix} \frac{(1-\alpha)^{1/2}}{\|\mathbf{Y}\|}\mathbf{P_Y} \\ \frac{\alpha^{1/2}}{\|\mathbf{X}\|}\mathbf{P_X} \end{bmatrix} \right\|^2. \tag{5}$$

De Jong & Kiers (1992) showed that **T** contains in its columns the first *R* eigenvectors of

$$(1-\alpha)\frac{\mathbf{H}_X\mathbf{YY}'\mathbf{H}_X}{\|\mathbf{Y}\|^2} + \alpha\frac{\mathbf{XX}'}{\|\mathbf{X}\|^2}, \tag{6}$$

where $\mathbf{H_X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, the role of which is to project **Y** on the space spanned by the columns of **X**. Once the estimate of **T** is found, $\mathbf{P_Y}$ and $\mathbf{P_X}$ are estimated by

$$\mathbf{P_Y} = \mathbf{T}'\mathbf{Y}, \tag{7}$$

$$\mathbf{P_X} = \mathbf{T}'\mathbf{X}. \tag{8}$$

Finally, the component weights are estimated by

$$\mathbf{W} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{T}. \tag{8}$$

See, for further details, De Jong, & Kiers (1992).

The columns of **T** are usually normalized so that, in each column, the elements have variance fixed at 1. In this case, when **X** contains standardized data and the extracted components are orthogonal, the component loadings in $\mathbf{P_X}$ are equal to the correlations between components and variables. Such a normalization of **T** is not sufficient to fully identify the solution. In fact, equally fitting solutions can be found by premultiplying $\mathbf{P_Y}$ and $\mathbf{P_X}$ by a rotation matrix **B**, provided that this rotation is compensated by postmultiplying **T** by $\mathbf{B}^{-1}$. Letting $\mathbf{P_Y}^R = \mathbf{BP_Y}$, $\mathbf{P_X}^R = \mathbf{BP_X}$ and $\mathbf{T}^R = \mathbf{TB}^{-1}$, we have

$$\mathbf{Y} = \mathbf{T}^R\mathbf{P}_\mathbf{Y}^R + \mathbf{E}_\mathbf{Y} = \mathbf{T}\mathbf{B}^{-1}\mathbf{B}\mathbf{P}_\mathbf{Y} + \mathbf{E}_\mathbf{Y} = \mathbf{T}\mathbf{P}_\mathbf{Y} + \mathbf{E}_\mathbf{Y}, \tag{9}$$

$$\mathbf{X} = \mathbf{T}^R\mathbf{P}_\mathbf{X}^R + \mathbf{E}_\mathbf{X} = \mathbf{T}\mathbf{B}^{-1}\mathbf{B}\mathbf{P}_\mathbf{X} + \mathbf{E}_\mathbf{X} = \mathbf{T}\mathbf{P}_\mathbf{X} + \mathbf{E}_\mathbf{X}. \tag{10}$$

The rotational freedom can be exploited in order to simplify the interpretation of the components by means of, e.g., the (normalized or non-normalized) Varimax procedure (Kaiser, 1958). Throughout the present paper, we opted for the non-normalized version.

For practical purposes, the applicability of PCOVR is enhanced by the availability of the `R` (R Core Team, 2020) package `PCovR` (Vervloet, Kiers, Van den Noortgate, & Ceulemans, 2015). It was used to perform the PCOVR analysis in the following example as well as in the simulation experiment of Section 4.

2.1. Example

A PCOVR analysis is applied to the Rohwer data (Timm, 1975) for illustrative purposes. The data refer to an experiment on $N = 69$ children. The research interest lies in assessing whether and how a set of $J = 5$ paired-associate (PA) tasks predicts the performance on $K = 3$ measures of aptitude and achievement. The PA tasks, labelled named (N), still (S), named still (NS), named action (NA), and sentence still (SS), vary in how the stimuli are presented. The three measures are a student achievement test (SAT), the Peabody Picture Vocabulary test (PPVT) and the Raven Progressive matrices test (RPMT). The need for PCOVR arises because several predictor variables are highly correlated (see Table 1).

TABLE 1 HERE

PCOVR is run on such data varying R in the set {2, 3, 4} and setting α following the sequential procedure proposed by Vervloet, Van Deun, Van den Noortgate, & Ceulemans (2013). First, the value of α is determined on the basis of maximum likelihood principles. Then, by using the selected weighting value, a generalization of the well-known scree test is used to determine R. We get the optimal values $R = 3$ and $\alpha = 0.91$. In order to interpret the solution, we inspect the varimax rotated component loading matrix reported in Table 2.

TABLE 2 HERE

We can see that Component 1 has high positive loadings for all the tasks with particular reference to SS, NS and NA. Components 2 and 3 are mainly related to the remaining two tasks. Specifically, Component 2 is positively related to S and Component 3 to N. In order to assess how the components predict the criterion variables, we observe the regression weight matrix given in Table 3.

TABLE 3 HERE

We can discover that positive relations emerge. PPVT are well predicted by Component 1. Therefore, good performances of SS, NS and NA are predictors of good performances in PPVT. The predictions of SAT and RPMT are more complex. SAT appears to be related most to Components 2 and 3. Hence, good performances of SAT primarily reflects good performances for S and N. Concerning RPMT, to a certain extent a connection with Components 1 and 2 is visible. To explicitly inspect the relationship between the criterion and predictor variables, the matrix $\mathbf{WP_y}$ can be analyzed. This matrix gives the regression weight for estimating **Y** from **X** bypassing the interpretation of the components. We have the weights reported in Table 4.

TABLE 4 HERE

Obviously, by inspecting Table 4 we can draw conclusions consistent with those based on Tables 2 and 3. Furthermore, we can highlight the role of S in predicting RPMT. The matrix **T** (not reported here) may also be considered to study how the units take different scores with respect to the extracted components.

3. Bootstrap methodology in Principal COVariates Regression

Once the PCOVR solution has been obtained, it would be interesting to assess the inferential properties of the obtained estimates. In particular, a relevant point to address is whether the obtained solution represents a good estimate of the population one. This can be done by computing either Standard Errors (SEs) or Confidence Intervals (CIs) for the parameter matrices. In the particular domain of PCOVR, this problem has not been covered yet. Therefore, we will use the results in the PCA framework as a starting point taking into account the existing connection between PCOVR and PCA reported in (5). For this purpose, we rely strongly on the paper by Timmerman et al. (2007). We must note that our results are developed conditionally on the PCOVR model specification. Specifically, on the basis of the available data, we first choose the value of α and the number of components R. Then, given this specification, we study uncertainty of the parameter estimates.

The computation of SEs or CIs in PCA can be done following at least two strategies depending on the ~~multi-~~ multivariate normality assumption. If it holds, assuming that the scores are independently randomly sampled from identically distributions (iid), analytic results can be derived (see, for an overview, Anderson, 1984). Refinements for particular cases are also available. For instance, Ogasawara (2000) studies asymptotic SEs for component loadings when data are standardized. The same author investigates SEs for rotated component loadings (Ogasawara, 1999, 2002).

3.1. Bootstrap

When the ~~multi-~~multivariate normality assumption is violated, as is often the case, analytic results are no longer available. ~~A more flexible~~It must be noted that several studies on the problem of misspecification and how it affects parameter estimates and standard errors (see, e.g., White, 1982). An alternative approach, adopted in this paper, is represented by the use of resampling techniques such as the bootstrap, originally introduced by Efron (1979). Let θ be a population parameter to be estimated. It can be estimated by the sample statistic t that represents the realization of the estimator T. The bootstrap leads to an assessment on the uncertainty associated with T by computing SEs or CIs. This is done by mimicing the sampling process by using the observed sample.

Let t_b be the estimate of θ obtained by using the b-th bootstrap sample ($b = 1, \ldots, B$), where B is the number of bootstrap samples. The bootstrap SE of T is

$$SE(T) = \sqrt{\sum_{b=1}^{B} \left(t_b - \frac{\sum_{b=1}^{B} t_b}{B} \right)^2 / B}. \tag{11}$$

The use of the bootstrap is motivated by at least two main reasons. First of all, there is no need to specify the unknown distribution for correct inference. Moreover, the bootstrap distribution can be expected to be a better approximation to the small sample distribution than the standard asymptotic approximation.

Given the confidence level $1 - \gamma$ (usually $\gamma = 0.05$), the bootstrap CI can be computed according to at least three approaches. The bootstrap Wald CI can be built by using the bootstrap SE defined in (11) as, for instance, $t \pm z_{1-\gamma/2} SE(T)$, where $z_{1-\gamma/2}$ denotes the $100(1 - \gamma/2)$th percentile of the standard normal distribution. The Wald CI is not invariant under monotonic transformations of the sample statistic t and its values are not confined to the range of what is to be estimated. The bootstrap percentile CI is based on the percentiles of the bootstrap distribution of t_b, $b = 1, \ldots, B$.

The CI is then defined as ($t_{\gamma/2}$, $t_{1-\gamma/2}$), where $t_{\gamma/2}$ and $t_{1-\gamma/2}$ are the 100(γ/2)th and 100(1 – γ/2)th percentiles of the bootstrap distribution, respectively. Differently from the Wald CI, the percentile CI interval is transformation respecting and range preserving. The bias-corrected and accelerated percentile bootstrap CI, proposed by Efron (1987), represents an enhancement of the percentile one. In detail, Efron (1987) improves the accuracy of the percentile CI by using adaptive percentiles for bias corrections and acceleration adjustments in order to correct the standard percentile CI for bias and skewness in the bootstrap distribution. A description of the procedure is provided in the appendix. The bias-corrected and accelerated percentile bootstrap CI, henceforth denoted as BC_a percentile CI, is transformation respecting, range preserving and has a lower coverage error in comparison with percentile CI (see, e.g., Efron, & Tibshirani, 1993).

3.2. Bootstrap confidence intervals in Principal COVariates Regression

Timmerman et al. (2007) illustrate different strategies for estimating confidence intervals in PCA. They show that the bootstrap approach should be preferred in comparison with the asymptotic approach and, among the various alternatives to estimate CIs by bootstrapping, the BC_a percentile method performed best by far. Since the PCOVR solution coincides with the PCA one of a particular matrix, we reasonably expect that BC_a percentile CI works well also in the PCOVR domain and hence we have used it in the present paper. As described by Timmerman et al. (2007), bootstrapping can be fruitfully applied in order to estimate CIs of the component loadings. By means of these estimated CIs, we hope to find intervals of values that contain the true population parameter with probability 1 – γ.

The main difficulty relies in the non-identifiability of the PCA solution in terms of sign and axis position because the bootstrap component loadings must be made consistent with the loadings from the observed sample. Timmerman et al. (2007) discuss the non-identifiability of the PCA solution by presenting three different cases related to different interpretations of the component loadings.

The first one consists of limiting the attention to the principal axes, i.e., the component loadings corresponding to the eigenvectors associated with the largest eigenvalues of the covariance/correlation data matrix. In such a case, the non-uniqueness of the PCA solution is related to the sign indeterminacy. In practice, the bootstrap component loadings must be multiplied by ±1 such that they optimally resemble the ones obtained from the observed sample and the same must be done for the population component loadings. In this case, the interpretation involves the unrotated component loadings. This is rarely done in psychology, where one usually applies rotations of the component loadings in order to simplify their interpretation. For this reason, we are not going to consider such a strategy in the present paper.

In order to achieve simplicity, the second strategy naturally arises. It consists of applying the same rotation method to the sample component loadings as to the bootstrap ones. For instance, if the sample component loadings are varimax rotated, the varimax rotation is also applied to the bootstrap component loadings. For comparative purposes, the sign indeterminacy and the ordering of the components should be checked. The previously described approach is usually referred to as the bootstrap fixed rotation method in order to emphasize that a fixed rotation method (e.g., the varimax) is employed for all the (bootstrap) samples. In this case, particular emphasis in the interpretation is put on the selected rotation method. This strategy will be referred to as 'Fixed criterion'.

Finally, the third approach consists of first rotating the original sample component loading matrix to simplicity, and then rotating the bootstrap component loading matrices in such a way that they resemble the (simplicity rotated) original one as much as possible. The idea is that each bootstrap solution just spans the space of a class of infinitely many rotated loadings matrices, and from these we wish to identify one that is interpreted (as much as possible) similar to the (simplicity rotated) sample loading matrix. This can be done by means of Procrustes rotation of each bootstrap loading matrix towards the (simplicity rotated) sample loading matrix as a target. Hence, the strategy can be called bootstrap target rotation method. Such a case implies that the rotation method as such is no longer relevant for interpretative purposes of the bootstrap solutions. In fact, all attention is paid to the *outcome* of the simple structure rotation procedure for the sample loading matrix only, and this sample solution is taken very seriously, and serves as the target for the other ones. The rationale is that this sample loading matrix supposedly is well interpretable, and therefore serves as a good reference basis. This strategy will be referred to as 'Procrustes rotation'.

The performances of alternative strategies for computing CIs in PCA have been deeply investigated by Timmerman et al. (2007), who focused on CIs for varimax rotated component loadings. In the study, the authors considered both bootstrap and asymptotic CIs and analyzed their quality in terms of coverage. Specifically, the coverage of a CI is appropriate when the probability of a $100(1 – \gamma)\%$ CI not covering the true population parameter θ from above and below is equal to $\gamma/2$. Setting $\gamma = 0.05$, this means that the true population parameter θ belongs to the CI with a probability equal to 0.95 (the confidence level). With probability equal to 0.025, θ is lower (higher) than the lower (higher) bound of the interval.

It is important to note that the non-identifiability of the PCA solution implies that it is not clear what the population parameters are. In fact, the above-recalled different interpretations of the component loadings and the related rotation variants also imply a different stance on the population component loadings, because also the population loading matrix actually spans the space of a class of infinitely many rotated loadings matrices. Because in the current approach, we have decided that the (simple structure rotated) sample loading matrix serves as a reference matrix, the population loading matrix must be transformed optimally towards the sample loading matrix in order to identify it, and in order to see to what extent the CIs we set up as estimates of confidence intervals cover comparable population values. This must be done in the same way as for the bootstrap component loadings because the rotation applied to the population component loadings must comply with the one applied to the bootstrap component loadings. See, for a deeper discussion, Kiers (2004).

Timmerman et al. (2007) found that the bootstrap CIs performed better than asymptotic CIs. Furthermore, among the different methods to estimate CIs by means of bootstrapping, the BC_a percentile CIs generally performed best.

The results of Timmerman et al. (2007) represent the starting point for estimating CIs in PCOVR. However, such results cannot be straightforwardly applied to PCOVR due to its higher level of complexity. With respect to PCA, the scope of the bootstrap is broadened. In fact, it is reasonable to estimate CIs for several parameter matrices, namely:

- the component loading matrix for the predictor variables on the components ($\mathbf{P_X}$),
- the regression weight matrix for the criterion variables on the components ($\mathbf{P_Y}$),
- the component weight matrix ($\mathbf{W}$),
- the regression weight matrix for estimating the criterion variables from the prediction variables ($\mathbf{WP_Y}$).

We decided to limit our attention to the bootstrap approach and, in particular, to the BC_a percentile method due to its valuable performances. Moreover, we extended the analysis by considering not only the (orthogonal) varimax rotation, but also the (oblique) quartimin one (Carroll, 1953). Such criteria will be considered for both the 'Fixed criterion' and the 'Procrustes rotation' strategies.

In order to compute bootstrap CIs, the following ordered steps for managing the indeterminacy of the PCOVR solution should be made. Let us consider the varimax case.

1. The sample component loading matrix, say $\mathbf{P}_{\mathbf{X}\mathrm{S}}$, is rotated to simple structure according to the varimax criterion and the rotation is compensated in the remaining parameter matrices.
2. The bootstrap samples solution, setting R and α as for the sample case, is rotated to be consistent with the rotated sample solution. This can be done by means of two alternative strategies.
   a. In the 'Fixed criterion' strategy, the component loading matrix for the b-th bootstrap sample ($b = 1, \ldots, B$), say $\mathbf{P}_{\mathbf{X}b}$, is varimax rotated. It is not yet guaranteed that $\mathbf{P}_{\mathbf{X}\mathrm{S}}$ and $\mathbf{P}_{\mathbf{X}b}$ are comparable because of the signs and the ordering of the components (in the

rows), but this can be fixed easily by adjusting signs and order of the rows of $\mathbf{P}_{\mathbf{X}b}$ such that they optimally resemble those of $\mathbf{P}_{\mathbf{X}\mathrm{S}}$ in terms of the Tucker congruence coefficient (Tucker, 1951). All these transformations are finally compensated in the remaining parameter matrices.

b. In the 'Procrustes rotation' strategy, $\mathbf{P}_{\mathbf{X}b}$ is rotated by means of an orthogonal rotation matrix $\mathbf{B}$ so as to resemble the sample component loading matrix $\mathbf{P}_{\mathbf{X}}$ as much as possible. Since the rows refer to the different components, the Procrustes rotation is applied to $\mathbf{P}_{\mathbf{X}b}'$ with respect to $\mathbf{P}_{\mathbf{X}}'$. Hence, the function

$$||\mathbf{P}_{\mathbf{X}b}'\mathbf{B} - \mathbf{P}_{\mathbf{X}}'||^2 \tag{16}$$

is minimized over $\mathbf{B}$. The minimization problem in (16) is usually referred to as orthogonal Procrustes rotation (Cliff, 1966). Once $\mathbf{B}$ is found, the rotation is compensated in the remaining parameter matrices, i.e., by postmultiplying $\mathbf{T}$ by $(\mathbf{B}')^{-1}$ provided that $\mathbf{P}_{\mathbf{Y}b}$ is premultiplied by $\mathbf{B}'$.

In the quartimin case, the previously described steps still hold provided that the varimax rotation is replaced by the quartimin one in the 'Fixed criterion' strategy and the minimum of (16) is achieved with respect to an oblique rotation matrix $\mathbf{B}$ (oblique Procrustes rotation problem, Jennrich, 2002) in the 'Procrustes rotation' strategy.

In this paper we are going to compute CIs based on either the varimax or quartimin rotation and either the 'Fixed criterion' or 'Procrustes rotation' strategy. This leads to the following four variants:

- Varimax rotated Sample and Bootstrap solutions (VSB);
- Quartimin rotated Sample and Bootstrap solutions (QSB);
- Varimax rotated Sample and Procrustes rotated Bootstrap solutions (VSPB);

- Quartimin rotated Sample and Procrustes rotated Bootstrap solutions (QSPB).

These four methods have been implemented in R (R Core Team, 2020) and are available upon request from the corresponding author. It must be underlined that the above list is not exhaustive. Many other variants could be used, adopting different rotational criteria.

3.3. Example

In order to assess the statistical uncertainty of the PCOVR solution for the Rohwer data reported in Section 2.1, we estimate the BC_a percentile CIs for the parameter matrices given in Tables 2 to 4. Since the sample solution is varimax rotated, and we now apply the variants VSB and VSPB. Setting $\gamma = 0.05$ and $B = 1000$, we get the CIs given in Tables 5 to 7.

TABLE 5 HERE

TABLE 6 HERE

TABLE 7 HERE

From Table 5, the inspection of the CIs for $\mathbf{P_X}$ highlights that the variability of the estimates is quite large. This especially occurs for Component 1 and for the low loadings in absolute sense of Components 2 and 3. The variables playing a relevant role in interpreting the components are well captured by the observed sample. In fact, all these estimates are clearly above zero, but for Component 1, these values are highly uncertain. The most stable component loadings appear to be the one of S on Component 2 and the one of N on Component 3. Such findings hold for both the fixed varimax and the Procrustes varimax strategies, although the bounds of the CIs differ considerably. The analysis of the CIs for $\mathbf{P_Y}$ and $\mathbf{WP_Y}$ reported in Tables 6 and 7 can be carried out in a similar way as for $\mathbf{P_X}$.

Apart from the CIs for $\mathbf{WP_Y}$ (any rotation does not alter $\mathbf{WP_Y}$: $\mathbf{WP_Y} = \mathbf{WB}^{-1}\mathbf{BP_Y}$), the use of VSB and VSPB leads to different CIs. This occurs because the former CIs consider sampling fluctuations due to the rotation criterion (varimax) and the latter CIs express sampling fluctuations related to the chosen target loading matrix. As such, CIs estimated by VSB are expected to be wider than those estimated by VSPB. This comment also holds for QSB and QSPB, with the difference involving the use of the (oblique) quartimin criterion.

Summing up, since different CIs are obtained when considering the four strategies summarized in Table 5, it appears to be important to assess their statistical behavior. For this purpose, we carried out a simulation experiment, the results of which are reported in the next section.

4. Simulation experiment

The aim of the simulation experiment was to assess how the statistical behavior of the bootstrap CIs was affected by the data structure. To this purpose, we explored different situations to see whether good results were obtained and also to discover whether under such conditions differences in the quality of the CIs emerged. These situations seemed reasonably representative of what one might encounter in practice and, therefore, the simulation experiment offers practical recommendations and suggestions on the computation of bootstrap CIs for the PCOVR parameter matrices. These recommendations and suggestions concern, among others, the required sample size and, in general, highlight particular data structures leading to a poorer quality of the CIs.

The simulation experiment was split into two studies. In the first study, we considered the case with more than one criterion variable. Data were generated according to different scenarios with a given number of components, which were expected to have similar relevance and strength. PCOVR solutions and bootstrap CIs were estimated by using the true number of components. In the second study, taking inspiration from Vervloet et al. (2016), a different set-up was considered where the relevance and strength of the components varied and PCOVR solutions and bootstrap CIs were estimated by using fewer components than those used for generating the data. The R (R Core Team, 2020) code that was used for the simulation experiment is available upon request from the corresponding author.

4.1. Simulation study n. 1

4.1.1. Set-up

~~Several data sets have been randomly generated in~~In order to assess the quality of the different strategies for computing bootstrap BC_a percentile CIs for the PCOVR solutions, ~~W~~we assumed to deal with a population of 10,000 units on which K criterion and J predictor variables were observed. We simulated population data according to the PCOVR model in (2) and (3). Denoting by R the number of components, we had:

$$\mathbf{Y}^{POP} = \mathbf{T}^{POP}\mathbf{P_Y}^{POP} + \varepsilon_Y \mathbf{E_Y}^{POP}, \tag{17}$$

$$\mathbf{X}^{POP} = \mathbf{T}^{POP}\mathbf{P_X}^{POP} + \varepsilon_X \mathbf{E_X}^{POP}, \tag{18}$$

where $\mathbf{Y}^{POP}$, of order (10,000 × K), and $\mathbf{X}^{POP}$, of order (10,000 × J), were the population data matrices for the criterion and predictor variables, respectively, $\mathbf{T}^{POP}$ was the population component score matrix of order (10,000 × R), $\mathbf{P_Y}^{POP}$ of order (R × K) was the population matrix of the regression weights for the criterion variables on the components and $\mathbf{P_X}^{POP}$ of order (R × J) was the population matrix of the component loading for the predictor variables on the components. Furthermore, $\mathbf{E_Y}^{POP}$ and $\mathbf{E_X}^{POP}$ were the population error matrices for $\mathbf{Y}^{POP}$ and $\mathbf{X}^{POP}$, of orders (10,000 × K) and (10,000 × J), respectively. Thus, the population data matrices can be split into a structural part (either $\mathbf{T}^{POP}\mathbf{P_Y}^{POP}$ or $\mathbf{T}^{POP}\mathbf{P_X}^{POP}$) and an error part (either $\mathbf{E_Y}^{POP}$ or $\mathbf{E_X}^{POP}$). The elements of $\mathbf{E_Y}^{POP}$ and $\mathbf{E_X}^{POP}$ were randomly generated from the standard normal distribution and multiplied by a scalar (one for each matrix) such that $\| \mathbf{T}^{POP}\mathbf{P_Y}^{POP} \| = \| \mathbf{E_Y}^{POP} \|$ and $\| \mathbf{T}^{POP}\mathbf{P_X}^{POP} \| = \| \mathbf{E_X}^{POP} \|$. The scalars ε_Y and ε_X tuned the amount of noise in the population data matrices. We considered two levels of noise for $\mathbf{Y}^{POP}$ ($\varepsilon_Y = 0.1$, *low noise Y*, and $\varepsilon_Y = 0.3$, *high noise Y*) and for $\mathbf{X}^{POP}$ ($\varepsilon_X = 0.1$, *low noise X*, and $\varepsilon_X = 0.3$, *high noise X*~~).~~) in order to assess whether a larger amount of noise deteriorated the statistical behavior of the CIs. Concerning the structural part, we operated as follows. We set $K = 6$ and we considered two levels for the ratio between criterion and prediction variables, i.e., in the *small ratio* case, $J/K = 2$, hence $J = 12$, in the *large ratio* case, $J/K = 3$, hence $J = 18$. This choice was made to assess whether differences in the numbers of prediction and criterion variables affected the quality of CIs. The number of components was set to $R = 2$ (*two-component* case) or $R = 3$ (*three-component* case~~).~~) in order to inspect whether the selection of the number of components had impact on the obtained CIs. The elements of $\mathbf{T}^{POP}$ were randomly drawn from the standard normal distribution. Two levels of simplicity for $\mathbf{P_Y}^{POP}$ and $\mathbf{P_X}^{POP}$ were assumed. In the simple case the matrices were constructed in such a way that every (criterion or prediction) variable was related to only one component and every row of $\mathbf{P_Y}^{POP}$ and $\mathbf{P_X}^{POP}$ contained 50% (if $R = 2$) or 66% (if $R = 3$) of zero elements. For instance, when $K = 6$, the structure of $\mathbf{P_Y}^{POP}$ with, respectively, $R = 2$ and $R = 3$, was

$$\mathbf{P_Y}^{POP} = \begin{bmatrix} x & x & x & 0 & 0 & 0 \\ 0 & 0 & 0 & x & x & x \end{bmatrix}, \qquad (19)$$

$$\mathbf{P_Y}^{POP} = \begin{bmatrix} x & x & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & 0 & 0 \\ 0 & 0 & 0 & 0 & x & x \end{bmatrix}, \qquad (20)$$

where the symbol 'x' denotes randomly generated values from the uniform distribution in [0.5, 1]. Such a case for $\mathbf{P_Y}^{POP}$ and $\mathbf{P_X}^{POP}$ was labelled *simple structure*. In the complex case, variables were related to at least one component (exactly two when R = 3) and every row had 33% of zero elements. Thus, (19) and (20) were replaced by

$$\mathbf{P_Y}^{\text{POP}} = \begin{bmatrix} x & x & x & x & 0 & 0 \\ 0 & 0 & x & x & x & x \end{bmatrix}, \tag{21}$$

$$\mathbf{P_Y}^{\text{POP}} = \begin{bmatrix} x & x & x & x & 0 & 0 \\ 0 & 0 & x & x & x & x \\ x & x & 0 & 0 & x & x \end{bmatrix}. \tag{22}$$

The complex structure case for $\mathbf{P_Y}^{\text{POP}}$ and $\mathbf{P_X}^{\text{POP}}$ was referred to as *complex structure*. Hence, by considering simple and complex structures for the component loadings and the regression weights we were interested in assessing whether and how the quality of the estimated CIs was affected by situations where some criterion and prediction variables are influenced by more than one component. In fact, we expected this leads to less accurate estimated CIs.

Once the population data were available, we randomly generated the sample data by varying the sample size that was equal to N = 50 (*small size*), N = 100 (*medium size*) and N = 500 (*large size*). By this, we aimed at studying whether a limited number of units would prevent us from obtaining good results in the CIs. For each sample size, 1,000 sample data sets were randomly generated from the population data. The design was fully crossed leading to 2 (*low noise Y* and *high noise Y*) × 2 (*low noise X* and *high noise X*) × 2 (*small ratio* and *large ratio*) × 2 (*two-component* and *three-component*) × 2 (*simple structure* and *complex structure*) × 3 (*small size*, *medium size* and *large size*) × 1,000 (sample replications) = 96,000 sample data sets.

4.1.2. Quality criteria

The simulation study aimed at assessing the statistical behavior of the estimated CIs in terms of coverage of the proportion of observations outside them.

For coverage, we checked whether the probabilities of the 100(1 – γ)% CIs covering the true population parameter θ were equal to (1 – γ) or not. This was assessed for the estimated CIs of $\mathbf{P}_X$, $\mathbf{P}_Y$, $\mathbf{W}$ and $\mathbf{WP}_Y$. The true population parameter matrices were the ones estimated by applying PCOVR to the population data by using the proper number of components and selecting α according to Vervloet, Van Deun, Van den Noortgate, & Ceulemans (2013). Note that the overall average value of α was 0.70. Following Kiers (2004), the population solutions were transformed optimally towards the sample ones in order to interpret the CIs we set up as estimates of confidence intervals. This was done by either the 'Fixed criterion' (varimax or quartimin) or the 'Procrustes rotation' strategy applied to $\mathbf{P}_X$ consistently with the four examined strategies for computing CIs (VSB, QSB, VSPB, QSPB) so that the same type of rotations was applied to the population and bootstrap solutions. Note that, as pointed out by Timmerman et al. (2007), the coverage was assessed with respect to the estimated parameters using the population data and not those used to generate the population data, because the bootstrap is not adopted for compensating a model mis-specification and the related systematic bias.

For each sample data set, we performed the PCOVR analysis by using the same values of R and α found for the corresponding population data and the solution was rotated to simple structure by applying either the varimax or the quartimin procedure to $\mathbf{P}_X$ consistently with the choice made for the population solution. Then, we estimated the bootstrap BC_a percentile CIs by means of the above-mentioned four strategies. This was done by considering B = 1,000 bootstrap samples. Such a value should be chosen in order to reduce the simulation error to an acceptable level. In doing so, we followed the choice made by Timmerman et al. (2007). However, as a reviewer observed, Dufour & Kiviet (1998) showed that the number of bootstrap samples should be such that (B + 1) γ is an integer. However, since B = 1,000, differences are negligible. For instance, in the QSPB case, $\mathbf{P}_X$ of the sample solution was quartimin rotated and the component loadings for the predictor variables of the population and bootstrap solutions were obliquely Procrustes rotated towards the sample ones. Obviously, in each solution, the rotation was compensated in the remaining parameter matrices for getting an equally fitting solution.

Once the solutions were rotated, we could properly evaluate the coverage because the rotated population solution was fully comparable with the bootstrap CIs. Exactly like Timmerman et al. (2007), the coverage was calculated as the percentage of times for which the true rotated parameter was inside the $100(1 - \gamma)\%$ estimated CIs. In practice, setting $\gamma = 0.05$, we calculated the percentages of times for which the true parameter was lower than the 2.5% bound or higher than the 97.5% bound of the CI. Denoting these exceeding percentages by 2.5% EP and 97.5% EP, respectively, the coverage was determined as

$$\text{Coverage} = 100\% - 2.5\%\ \text{EP} - 97.5\%\ \text{EP}. \tag{23}$$

Note that this percentage was computed for every condition, i.e., for every combination of all the levels of all the design variables, by using the 1,000 replications. Coverages close to 95% indicate good statistical behavior.

4.1.3. Results

The results on the coverages of the BC_a percentile CIs for $\mathbf{P_X}$, $\mathbf{P_Y}$, $\mathbf{W}$ and $\mathbf{WP_Y}$ are reported by distinguishing with respect to the levels of the design variables and to the four strategies. Moreover, concerning $\mathbf{P_X}$ and $\mathbf{P_Y}$ we also distinguished the coverage with respect to the 'high' and 'low' values, i.e., those denoted by, respectively, '*x*' and '0' in (20)-(23). The mean 95% coverage levels computed with respect to all the 96,000 datasets and distinguished by parameter matrix are displayed in Figure 1.

FIGURE 1 HERE

First of all, we can see that a more or less pronounced undercoverage was registered on average. Such a comment especially holds for the CIs estimated by QSPB. For the remaining three strategies, mean coverages were very similar. The estimated CIs for $\mathbf{P_Y}$ were, on average, the best ones in terms of coverage. For VSB and QSB, these were almost equal to the desired level (94.83% and 94.81%, respectively). With respect to $\mathbf{P_X}$, VSB (93.10% on average), QSB (92.78%) and VSPB (92.77%) had essentially the same mean coverages. Virtually no differences emerged for the mean coverages of the CIs estimated for **W** ranging from 92.73% for VSB to 92.94% for VSPB. Finally, the estimated CIs for $\mathbf{WP_Y}$ (the same for all strategies) were characterized by a quite large undercoverage (91.65% on average).

A deeper insight into the assessment of the statistical behavior of the estimated CIs was attained by analyzing the mean coverages distinguished by the levels of the design variables. In particular, we checked whether, for each scenario, the mean coverages became closer to 95% for increasing values of N. There were in total $2^5 = 32$ scenarios corresponding to the combinations of the levels of *noise Y*, *noise X*, *ratio*, *component* and *structure*. Figure 2 contains the results for $\mathbf{P_X}$, while outcomes for the remaining parameter matrices are reported in Figures S1-S7 as Supplementary Material. Note that the mean coverages for $\mathbf{P_X}$ and $\mathbf{P_Y}$ are also distinguished by high and low loadings. By inspecting Figure 2, we can see an appropriate statistical behavior, i.e., the mean coverage tends to 95% when the sample size N increases. Except for QSPB, for all the remaining three variants, this is often visible when N = 500 and, sometimes, even when N = 100. Specifically, in scenarios characterized by *large ratio*, *three-component* and *complex structure*, the mean coverages for VSB, QSB and VSPB were slightly lower than 95% for samples of size N = 100. Conversely, in such scenarios, CIs computed by QSPB presented undercoverage, even more severe when N = 500. In general, the statistical behavior of CIs built according to QSPB was not satisfactory in *complex structure* scenarios. Possibly, this is because the simple structure rotated sample loadings in the complex condition were rather unstable, because in these cases different rotations may lead to fairly similar quartimin values. Since these matrices were taken as references to which population and bootstrap loading matrices were rotated, it is conceivable that this instability led to weaker quality confidence intervals for QSPB.

The previous comments also hold for the mean coverages of CIs referring to the high and low loadings in $\mathbf{P}_\mathbf{X}$ (Figures S1 and S2, respectively). Some minor differences were visible in the mean coverages, but, as far as we saw, no general trends emerged. The mean coverage levels for $\mathbf{P}_\mathbf{Y}$ are displayed in Figures S3-S5, where the latter two figures concern the high and low regression weights. All in all, as we already observed, for each variant, the mean coverages of the CIs for $\mathbf{P}_\mathrm{Y}$ were by far the best ones in comparison with the other parameter matrices. Consistently with the results for $\mathbf{P}_\mathrm{X}$, appropriate statistical behaviors were observed for all the variants except for QSPB. Specifically, as for $\mathbf{P}_\mathbf{x}$, a pattern of undercoverage occurred. The mean coverage levels for $\mathbf{W}$ are reported in Figure S6. The differences in the performances of the four criteria were less noticeable with respect to $\mathbf{P}_\mathrm{X}$. However, similar statistical behaviors as with $\mathbf{P}_\mathbf{X}$ emerged. Finally, the mean coverage levels for $\mathbf{WP}_\mathbf{Y}$ are plotted in Figure S7. As already observed, the same CIs were found for all the criteria since $\mathbf{WP}_\mathbf{Y}$ is not affected by rotations. In general, we observed undercoverage for the CIs independently of the design variables.

4.2. Simulation study n. 2

A smaller simulation study has also been implemented aiming at assessing the quality of the estimated CIs when the components underlying the data had different relevance and strength. In doing so, we were interested in studying what happened when fewer components than those used to generate the data were extracted. In fact, in the previous simulation study, we used the true number of components for fitting PCOVR to the data. In the current simulation study, we checked whether and possibly how a wrong choice of R affected the quality of the estimated CIs.

4.2.1. Set-up

The set-up of this simulation study took inspiration from the one of “Simulation study 1” of Vervloet et al. (2016). We assumed to deal with a population of 10,000 units. on which K = 1 criterion and J = 24 or J = 48 predictor variables were observed. Following the usual notation, we simulated population data as:

$$\mathbf{y}^{POP} = \mathbf{T}^{POP}\mathbf{p}_{\mathbf{Y}}^{POP} + \varepsilon_{Y}\mathbf{e}_{\mathbf{Y}}^{POP}, \tag{24}$$

$$\mathbf{X}^{POP} = \mathbf{T}^{POP}\mathbf{P}_{\mathbf{X}}^{POP} + \varepsilon_{X}\mathbf{E}_{\mathbf{X}}^{POP}, \tag{25}$$

where $\mathbf{y}^{POP}$ and $\mathbf{e}^{POP}$ are now vectors of length 10,000 and $\mathbf{p}_{\mathbf{Y}}^{POP}$ is a vector of length 4, i.e., R = 4 components were used to generate data. The vector $\mathbf{p}_{\mathbf{Y}}^{POP}$ was constructed according to three cases. Namely, R/2 (= 2) elements (the second and the fourth) were always set to 0 and the remaining R/2 values (respectively, the first and the third) were equal to 0.71 and 0.71 (case labelled *no difference*), 0.60 and 0.80 (case *small difference*) and 0.44 and 0.90 (case *large difference*). As in Vervloet et al. (2016), this structure of $\mathbf{p}_{\mathbf{Y}}^{POP}$ allowed for varying the relevance of the components. In the current study, we were interested in assessing whether and how this affected the quality of the obtained CIs. The matrix $\mathbf{P}_{\mathbf{X}}^{POP}$ was such that, for every variable, one loading was equal to 1 and the other ones were 0. For every component, the numbers of loadings equal to one differed in order to consider various levels of strength of the components. Such numbers were chosen in the same way as in Vervloet et al. (2016) and reported in Table 8 leading to six cases labelled ‘*4% vs 46%*’, ‘*8% vs 42%*’, ‘*13% vs 38%*’, ‘*17% vs 33%*’, ‘*21% vs 29%*’ and ‘*25% vs 25%*’. For instance, ‘*4% vs 46%*’ means that, when J = 24, two components had (1/24)×100% ≈ 4% of loadings equal to 1 and the other two components (11/24)×100% ≈ 46% loadings equal to 1. When J = 48, such percentages corresponded to about (2/48)×100% and (22/48)×100%, respectively. Such a design variable allowed us to see whether under such conditions differences between the quality of the CIs could be discerned. The elements of $\mathbf{T}^{POP}$ were randomly drawn from the standard normal distribution and the same was done for the elements of $\mathbf{e_Y}^{POP}$ and $\mathbf{E_X}^{POP}$. Next, these were normalized such that $\| \mathbf{e_Y}^{POP} \| = \| \mathbf{T}^{POP}\mathbf{p_Y}^{POP} \|$, and $\| \mathbf{E_X}^{POP} \| = \|\mathbf{T}^{POP}\mathbf{P_X}^{POP} \|$, respectively. In addition, to tune the amount of noise in the population data matrices, $\mathbf{e_Y}^{POP}$ and $\mathbf{E_X}^{POP}$ were also multiplied by the scalars ε_Y and ε_X. These two scalars took three levels (0.05, 0.25 and 0.45).

TABLE 8 HERE

Once the population data were available, we randomly generated sample data with $N = 200$. For each condition, 25 sample data sets were randomly generated from the population data. The design was fully crossed leading to 3 (levels of noise for **y**, ε_Y) × 3 (levels of noise for **X**, ε_X) × 2 (numbers of predictor variables, J) × 3 (levels of relevance: *no difference, small difference*, *large difference*) × 6 (levels of strength: '*4% vs 46%*', '*8% vs 42%*', '*13% vs 38%*', '*17% vs 33%*', '*21% vs 29%*', '*25% vs 25%*') × 25 (sample replications) = 8,100 sample data sets.

4.2.2. Results

The PCOVR analysis (including the selection of α) and the assessment of the coverage of the CIs of $\mathbf{P_X}$, $\mathbf{p_Y}$, **W** and $\mathbf{WP_Y}$ was done as described in Section 4.1.2. The most relevant difference is that the true population parameter matrices and the sample ones were the ones estimated by applying PCOVR to the population data by extracting not only the proper number of components ($R = 4$), but also a lower number ($R = 3$) to assess whether this affected the coverage.

The mean 95% coverage levels computed with respect to all the 8,100 datasets and distinguished by parameter matrix are displayed in Figure 3.

FIGURE 3 HERE

The results were consistent with those registered for the previous simulation study. Namely, once again, we observed undercoverage on average with respect to the CIs for all the parameter matrices except for $\mathbf{p_Y}$ for which the mean coverage was around the desired level. Although some differences occurred, such a comment largely holds for all the estimation strategies. An interesting finding was that, in case of misspecified model, i.e., fewer components extracted than the true number, undercoverage was more severe on average. This was mainly observed for $\mathbf{W}$ and $\mathbf{Wp_Y}$ and, to a limited extent, for $\mathbf{P_X}$ and $\mathbf{p_Y}$. Given the impact of a wrong number of components on the coverage, we further inspected this situation. For this purpose, we considered a subset of 3 × 3 × 2 × 3 × 6 = 324 randomly generated data sets corresponding to the first replication of every scenario and studied the width of the CIs for $\mathbf{p_Y}$. Note that the components have been ordered with respect to the CI widths for the elements of $\mathbf{p_Y}$, in increasing order. Figure 4 contains the boxplots of the CI widths for $\mathbf{p_Y}$ distinguished by the number of extracted components R and the criterion for estimating CIs.

FIGURE 4 HERE

We discovered that extracting a smaller number of components led to rather unstable solutions and, therefore, the CIs were relatively wide when R = 3 compared to when R = 4. In particular, one component (C3) appeared to be extremely unstable. Such a phenomenon occurred for all the criteria. Now one might think that such wider intervals would not lead to undercoverage, but rather to overcoverage (as wider intervals cover more values). However, the instability of the parameters can be expected to also make the estimates of the intervals less stable, and could therefore lead to more generally less predictable and desired behavior. Why then this led to undercoverage and not to overcoverage is not fully clear to us, but obviously, given that the nominal coverage should be 95% it can not easily get higher, whereas it can easily get lower. Furthermore, as expected, the use of

'Fixed criterion' strategies led to CIs wider than the corresponding 'Procrustes rotation' strategies, i.e., VSB vs VSPB and QSB vs QSPB. This was observed for both the cases with R = 3 and R = 4. To gather more information, we investigated whether and how the coverage differed with respect to the levels of the design variables. From the previous simulation study, we observed a slight impact on the estimated CIs by the number of (predictor) variables and the levels of noise. For this reason, we checked how such design variables affected the coverage. Such mean coverages for each parameter matrix are reported in Figures S8-S11 (Supplementary Material). All in all, the mean coverages remained stable even if a few minor differences were found especially for $\mathbf{W}$ and $\mathbf{W}\mathbf{p}_{\mathbf{Y}}$. Based on this result, we then studied whether coverage depended on the different relevance and strength of the components.

FIGURE 5 HERE

Figure 5 contains the mean coverages for all the parameter matrices distinguished by all combinations of levels of the design variables relevance and strength. First, we noticed a more pronounced undercoverage on average when R = 3 components were extracted in comparison with the R = 4 case. This occurred for all the parameter matrices except for $\mathbf{p}_{\mathbf{Y}}$. As far as we saw, the relevance of the components had a negligible impact on the mean coverage levels. On the contrary, they appeared to be related to the strength. However, this especially holds for $\mathbf{W}$ and $\mathbf{W}\mathbf{p}_{\mathbf{Y}}$. In particular, the mean differences in coverage between the R = 3 and R = 4 cases were larger when the components had almost the same strength ('*17% vs 33%*', '*21% vs 29%*', '*25% vs 25%*'). If so, the model misspecification remarkably increased the risk of estimating CIs with undercoverage. Differences were smaller for the remaining levels of strength, in particular for '*4% vs 46%*'. In such a case, the mean coverages were essentially the same, apart from $\mathbf{W}\mathbf{p}_{\mathbf{Y}}$ in the *no difference* case. We also observed that, for the '*4% vs 46%*' scenario, the worst mean coverage (severe undercoverage) of the CIs for $\mathbf{W}$ and $\mathbf{W}\mathbf{p}_{\mathbf{Y}}$ was registered. Hence, such CIs should be taken with a pinch of salt if underlying components are expected to have rather different strength. Note that the previous comments hold for all the four criteria for estimating CIs.

5. Discussion

The paper discussed some variants for computing CIs for the parameter matrices of the PCOVR solution. In the PCA framework, this point has been investigated and there is a certain consensus towards the use of bootstrap CIs for making inference on the PCA solution. Among the various alternatives for computing CIs, those based on the bootstrap BCa percentile method, appear to be the most valuable choice (see Timmerman et al., 2007). In the PCOVR framework, as far as we know, no studies have been carried out yet. This stimulated us to study the performance of the bootstrap BCa percentile method for computing CIs for the parameter matrices of a PCOVR solution~~. By comparing PCA and PCOVR, some common and distinctive features are visible. The most important distinctive feature concerns the higher level of complexity for PCOVR because inference can be made for more than one parameter matrix, while for PCA attention is paid only to the component loading matrix. In fact, CIs can be computed for the matrices~~, i.e, $\mathbf{P}_{\mathbf{X}}$~~ and~~, $\mathbf{P}_{\mathbf{Y}}$, ~~but also for the component weight matrix~~ $\mathbf{W}$ and the combined matrix $\mathbf{W}\mathbf{P}_{\mathbf{Y}}$~~ expressing the regression weights. The most relevant common aspect involves~~. To handle the non-uniqueness of the PCOVR solutions~~. Thus, suitable strategies are needed. In the paper~~, we considered four ~~variants~~strategies involving the varimax and quartimin rotations of $\mathbf{P}_{\mathbf{X}}$ and distinguishing the use of a fixed criterion and the Procrustes rotation.

~~The results of~~By a simulation experiment~~ aiming at assessing the quality of the four above-mentioned variants for handling the rotational freedom in PCOVR solutions and analyzing their impact on the CIs for the four above-mentioned parameter matrices have been reported. In particular, by the simulation experiment~~, we studied whether the coverage of the CIs obtained by using the four above-mentioned strategies tended to expected levels as the sample size increased. In doing so, we did not make a comparative assessment among the variants because the variants determine CIs expressing different types of uncertainty and, hence, are not comparable.

~~The~~The simulation experiment offered recommendations on the computation of CIs for the PCOVR parameters when doing empirical research. In fact, some differences in the statistical behavior of the CIs emerged with respect to the adopted variant, the parameter matrix, and the characteristics of the data. In some cases, the obtained CIs appeared to be reasonably good estimates, in some others, the quality degraded. Going into details, the simulation experiment was split into two parts. In the first one, we considered the case with more than one criterion variable and studied the statistical behavior of the estimated CIs in different scenarios. We found that, in general, a relatively small number of units (about 100) allows for building CIs that are reasonably good in terms of coverage. When the underlying structure of the components is complex, i.e., with predictor and criterion variables related to more than one component, a larger number of units should be used for computing CIs. This seems to hold especially for those variants based on the quartimin rotation. In particular, in such complex scenarios, the statistical behavior of CIs built according to QSPB was quite poor. Possibly, this is because the simple structure rotated sample loadings in the complex condition were rather unstable, because in these cases different rotations may lead to fairly similar quartimin values. Since these matrices were taken as references to which population and bootstrap loading matrices were rotated, it is conceivable that this instability led to weaker quality CIs for QSPB.

Apart from the adopted variant, some differences in the quality of the CIs for the parameter matrices emerged. In particular, we can state that the CIs for $\mathbf{P}_\mathbf{X}$ and ~~$\mathbf{P}_\mathbf{Y}$ are more accurate than those for $\mathbf{W}$ and $\mathbf{WP}_\mathbf{Y}$~~, especially, $\mathbf{P}_\mathbf{Y}$ are more accurate than those for $\mathbf{W}$ and $\mathbf{WP}_\mathbf{Y}$. The lowest quality of the CIs for $\mathbf{WP}_\mathbf{Y}$ can possibly be explained by the fact that such CIs are affected by a double source of uncertainty, i.e., those concerning the CIs for $\mathbf{W}$ and $\mathbf{P}_\mathbf{Y}$. The poorer quality of the

CIs for **W** in comparison with those for $\mathbf{P_X}$ and $\mathbf{P_Y}$ is less easy to understand. Future research will be needed to understand this result in detail.

The previous results were observed when PCOVR was applied by setting the true number of components and avoiding considering components having different relevance and strength. This point was addressed in the latter part where datasets involving one criterion variable were generated. We discovered that, to some extent, the relevance and strength of the components affected the coverage of the CIs, in particular, for **W** and $\mathbf{WP_Y}$. A much more relevant role is played by the choice of the number of components. In fact, the quality of the CIs deteriorated when fewer components are used than those used to generate the data. This suggests that, if the number of components used is too low, one should be careful in interpreting CIs, because they may actually have coverages considerably lower than their nominal value.

Acknowledgments

We thank the three anonymous reviewers whose comments and suggestions greatly improved the quality of the paper.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, *22*, 327-351.

Anderson, T. W. (1958). Introduction to Multivariate Statistical Analysis. New York: Wiley.

Anderson, T. W. (1984). An introduction to multivariate statistical analysis (2nd ed.). New York: Wiley.

Carroll, J. B. (1953). An analytic solution for approximating simple structure in factor analysis. *Psychometrika, 18,* 23-38. doi:10.1007/BF02289025

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33–42. doi:10.1007/BF02289455

De Jong, S., Kiers, H. A. L. (1992). Principal covariates regression: part I. Theory~~,~~. *Chemometrics and Intelligent Laboratory Systems*, *14*, 155-164. doi:10.1016/0169-7439(92)80100-I

Dufour, J. - M., Kiviet, J. F. (1998). Exact inference methods for first-order autoregressive distributed lag models. *Econometrica*, *66*, 79-104. doi:10.2307/2998541

Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *The Annals of Statistics, 7,* 1-26. doi:10.1214/aos/1176344552

Efron, E. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*, 171-185. doi:10.2307/2289144

Efron, B., Tibshirani, R. (1993). An Introduction to the Bootstrap. Boca Raton: Chapman & Hall/CRC.

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

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, *5*, 248-264. doi:10.1016/0047-259X(75)90042-1

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, *67,* 7-20. doi:10.1007/BF02294706

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Journal of Royal Statistical Society Series C*, *31*, 300-303. doi:10.2307/2348005

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187-200. doi:10.1007/BF02289233

Kiers, H. A. L. (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics*, *18*, 22-36. doi: 10.1002/cem.841

Kiers, H. A. L., Smilde, A. K. (2007). A comparison of various methods for multivariate regression with highly collinear variables. *Statistical Methods and Applications, 16*, 193-228. doi:10.1007/s10260-006-0025-5

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1991). Approximating confidence intervals for factor loadings. *Multivariate Behavioral Research, 26,* 421-434. doi:10.1207/s15327906mbr2603_3

Ogasawara, H. (1999). Standard errors for Procrustes solutions. *Japanese Psychological Research, 41*, 121-130. doi:10.1111/1468-5884.00111

Ogasawara, H. (2000). Standard errors of the principal component loadings for unstandardized and standardized variables. *British Journal of Mathematical and Statistical Psychology*, *53*, 155-174. doi:10.1348/000711000159277

Ogasawara, H. (2002). Concise formulas for the standard errors of component loading estimates. *Psychometrika*, *67*, 289-297. doi:10.1007/BF02294847

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Tikhonov, A. N. (1947). On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, *39*, 195-198.

Timm, N. H. (1975). Multivariate Analysis with Applications in Education and Psychology, Belmont, CA: Wadsworth (Brooks/Cole).

Timmerman, M. E., Kiers, H. A. L., Smilde, A. K. (2007). Estimating confidence intervals for principal component loadings: a comparison between the bootstrap and asymptotic results. *British Journal of Mathematical and Statistical Psychology*, *60*, 295-314. doi:10.1348/000711006X109636

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No. 984*. Washington, DC: Department of the Army.

Vervloet, M., (2017). Unraveling and unlocking the assets of principal covariates regression, Ph.D. Thesis, Leuven: KU Leuven.

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. doi:10.1016/j.chemolab.2013.02.005

Vervloet, M., Kiers, H. A. L., Van den Noortgate, W., Ceulemans, E. (2015). PCovR: an R package for principal covariates regression. *Journal of Statistical Software, 65*, 1-14. doi:10.18637/jss.v065.i08

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. doi:10.1016/j.chemolab.2015.12.004

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1-25. doi:10.2307/1912526

Appendix

The lower and upper bounds of the BC_a percentile CIs, depend on the bootstrap distribution and on two parameters, z_0 and a. The bias-correction parameter, z_0, is determined by considering the proportion of bootstrap estimates that are less than the observed sample statistic t. It can be estimated as

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\sum_{b=1}^{B} I(t_b < t)}{B}\right), \quad \text{(A1)}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function and $I(t_b < t)$ is the indicator function equal to 1, if $t_b < t$, and to 0, otherwise. The acceleration parameter, a, is proportional to the skewness of the bootstrap distribution. A common way to estimate a is based on the jackknife method. The jackknife is another resampling method that estimates a parameter of interest by using the observed sample and adding or removing single units (positive and negative jackknife, respectively). According to Lambert et al. (1991), we can estimate a as

$$\hat{a} = \frac{\sum_{i=1}^{N}\left[\left(t^{+i}-t\right)/(N+1)\right]^{3}}{6\left\{\sum_{i=1}^{N}\left[\left(t^{+i}-t\right)/(N+1)\right]^{2}\right\}^{3/2}} = \frac{\sum_{i=1}^{N}\left(t^{+i}-t\right)^{3}}{6\left\{\sum_{i=1}^{N}\left(t^{+i}-t\right)^{2}\right\}^{3/2}}, \tag{A2}$$

where t^{+i} is the positive jackknife estimate of θ computed by considering the original sample adding unit i, $i = 1, \ldots, N$. By using (A1) and (A2), the lower and upper bounds of the BC$_a$ percentile CIs at the confidence level $1 - \gamma$ are given by

$$\Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\gamma/2}}{1-\hat{a}\left(\hat{z}_0 + z_{\gamma/2}\right)}\right) \tag{A3}$$

and

$$\Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\gamma/2}}{1-\hat{a}\left(\hat{z}_0 + z_{1-\gamma/2}\right)}\right), \tag{A4}$$

respectively. For further details on the bootstrap refer to, e.g., Efron, & Tibshirani (1993).

Table 1. Correlation matrix for the PA tasks

	N	S	NS	NA	SS
N	1.00	0.25	0.51	0.49	0.46
S		1.00	0.34	0.55	0.43
NS			1.00	0.68	0.66
NA				1.00	0.72
SS					1.00

Table 2. Varimax rotated component loading matrix for the predictor variables ($\mathbf{P_X}'$). Component loadings higher than 0.30 in absolute sense are in bold.

	Component 1	Component 2	Component 3
N	0.25	0.11	**0.96**
S	0.20	**0.96**	0.10
NS	**0.83**	0.09	**0.32**
NA	**0.74**	**0.45**	0.27
SS	**0.84**	0.26	0.21

Table 3. Regression weight matrix for the criterion variables on the components ($\mathbf{P}_{\mathbf{Y}}'$).

	Component 1	Component 2	Component 3
SAT	0.15	0.23	0.24
PPVT	0.47	0.17	0.25
RPMT	0.23	0.26	0.16

Table 4. Regression weight matrix for the criterion variables on the predictor variables (**$\mathbf{WP_y}$**).

	SAT	PPVT	RPMT
N	0.20	0.11	0.09
S	0.18	0.01	0.18
NS	0.00	0.19	0.04
NA	0.08	0.17	0.11
SS	0.02	0.19	0.07

Table 5. BC_a percentile CIs at the confidence level 0.95 for $\mathbf{P}_X'$.

VSB				VSBP		
Component 1	Component 2	Component 3		Component 1	Component 2	Component 3
(0.16, 0.39)	(–0.02, 0.22)	(0.94, 0.98)	N	(0.15, 0.36)	(–0.04, 0.22)	(0.94, 0.99)
(0.11, 0.30)	(0.93, 0.99)	(–0.01, 0.20)	S	(0.02, 0.34)	(0.91, 1.00)	(–0.02, 0.21)
(0.35, 0.94)	(–0.16, 0.27)	(0.14, 0.66)	NS	(0.46, 0.96)	(–0.10, 0.36)	(0.02, 0.57)
(0.47, 0.87)	(0.24, 0.68)	(0.11, 0.45)	NA	(0.54, 0.84)	(0.30, 0.60)	(0.12, 0.46)
(0.48, 0.94)	(0.10, 0.56)	(–0.02, 0.59)	SS	(0.61, 0.95)	(0.11, 0.46)	(–0.08, 0.54)

Table 6. BC_a percentile CIs at the confidence level 0.95 for $\mathbf{P_Y}'$.

VSB				VSBP		
Component 1	Component 2	Component 3		Component 1	Component 2	Component 3
(–0.15, 0.44)	(0.01, 0.45)	(–0.18, 0.47)	SAT	(–0.09, 0.41)	(0.00, 0.44)	(–0.19, 0.48)
(0.17, 0.65)	(–0.04, 0.41)	(0.04, 0.45)	PPVT	(0.23, 0.64)	(–0.04, 0.38)	(0.04, 0.50)
(0.05, 0.44)	(0.00, 0.44)	(–0.04, 0.35)	RPMT	(0.05, 0.43)	(0.00, 0.45)	(–0.03, 0.35)

Table 7. BC_a percentile CIs at the confidence level 0.95 for $\mathbf{WP_y}$.

VSB & VSBP			
	SAT	PPVT	RPMT
N	(–0.23, 0.44)	(–0.15, 0.32)	(–0.12, 0.27)
S	(–0.08, 0.38)	(–0.22, 0.19)	(–0.10, 0.37)
NS	(–0.36, 0.13)	(0.06, 0.30)	(–0.18, 0.17)
NA	(–0.07, 0.13)	(0.06, 0.25)	(0.02, 0.18)
SS	(–0.18, 0.10)	(0.08, 0.33)	(–0.13, 0.15)

Table 8. Number of component loadings equal to 1 per component (Simulation study n. 2).

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

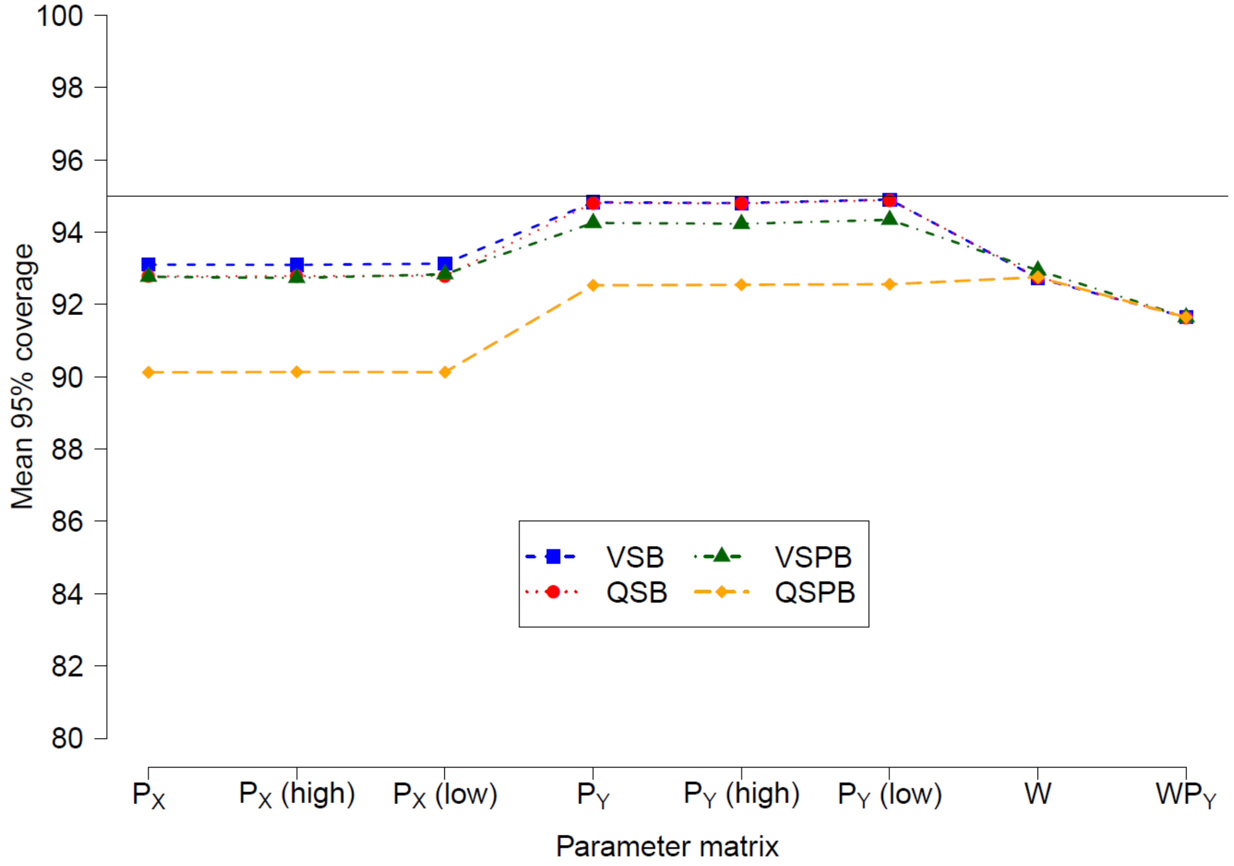

Figure 1. Mean 95% coverage levels of the CIs for the parameter matrices (Simulation study n. 1). The horizontal line indicates the desired coverage level.

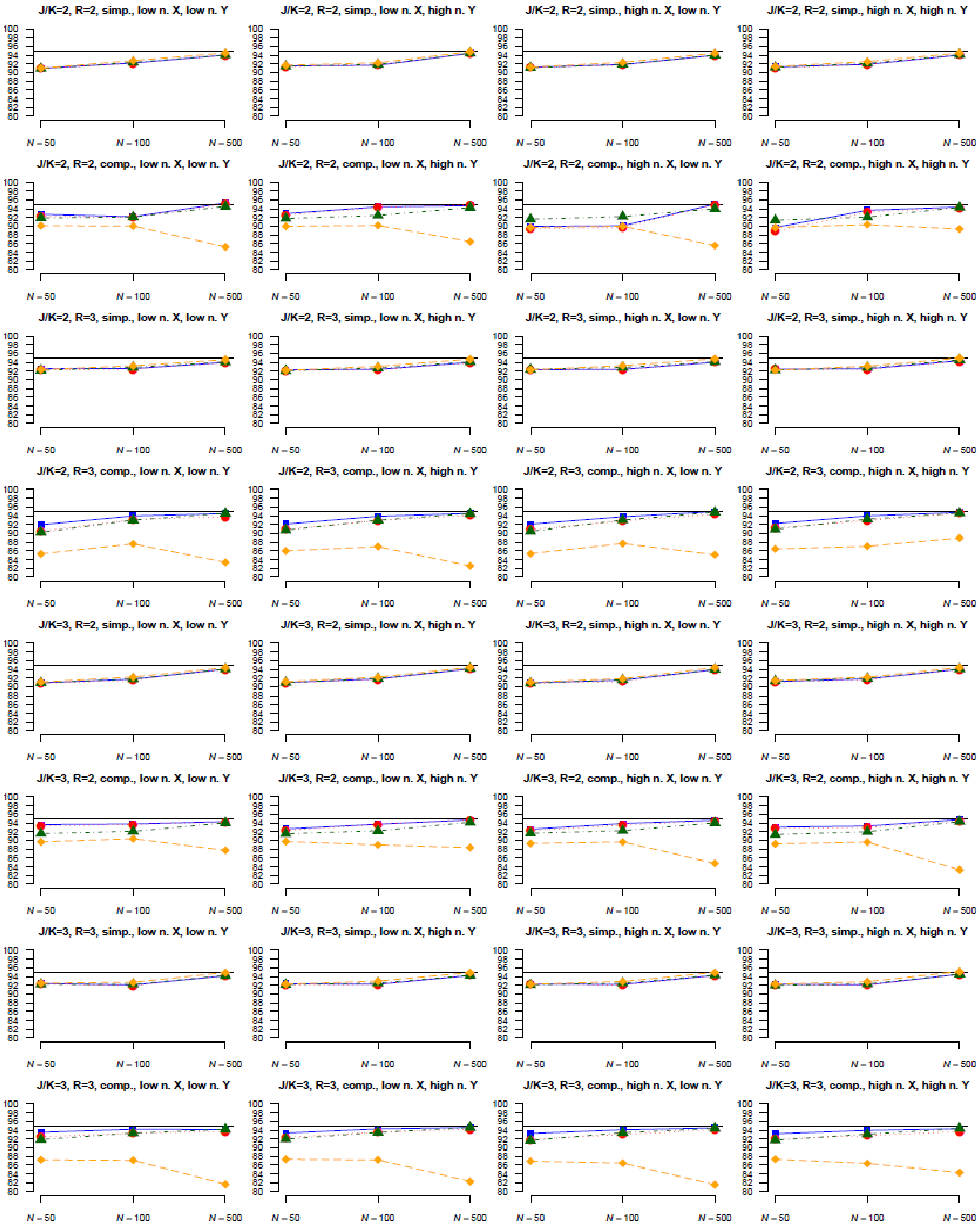

Figure 2. Mean 95% coverage levels of the CIs for $\mathbf{P}_X$ (Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

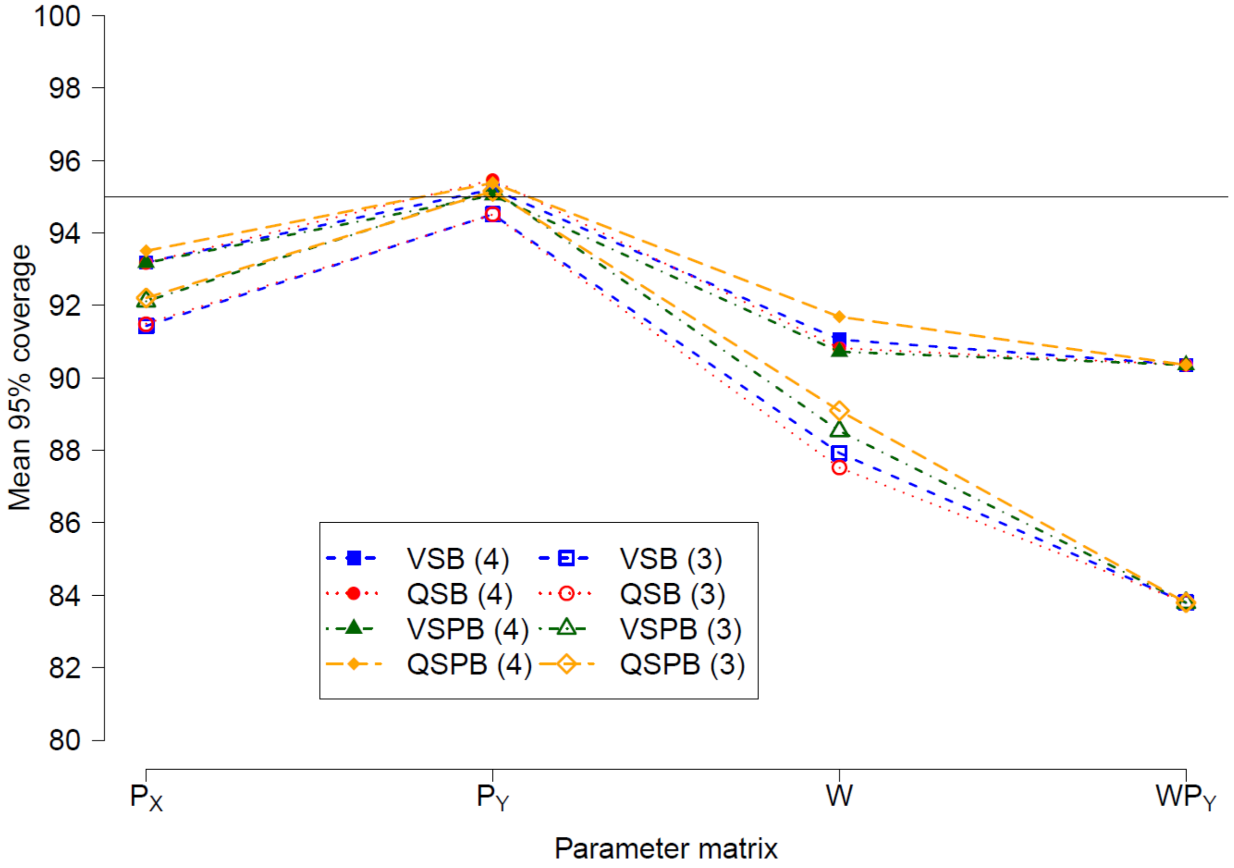

Figure 3. Mean 95% coverage levels of the CIs for the parameter matrices (Simulation study n. 2). The horizontal line indicates the desired coverage level.

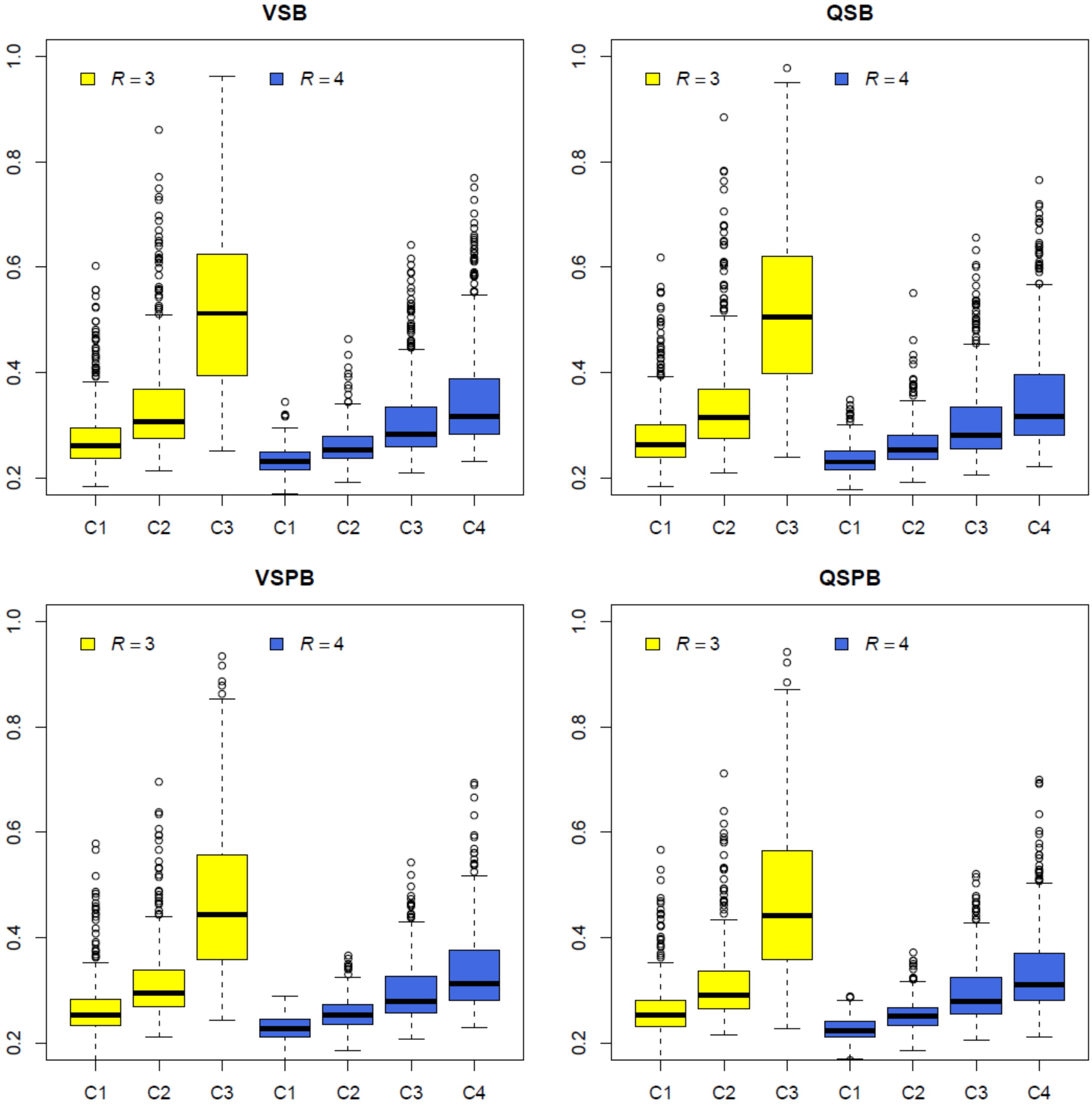

Figure 4. Boxplots of the CI widths for $\mathbf{p_Y}$ (subset of 324 randomly generated data sets in Simulation study n. 2) distinguished by the number of extracted components *R*.

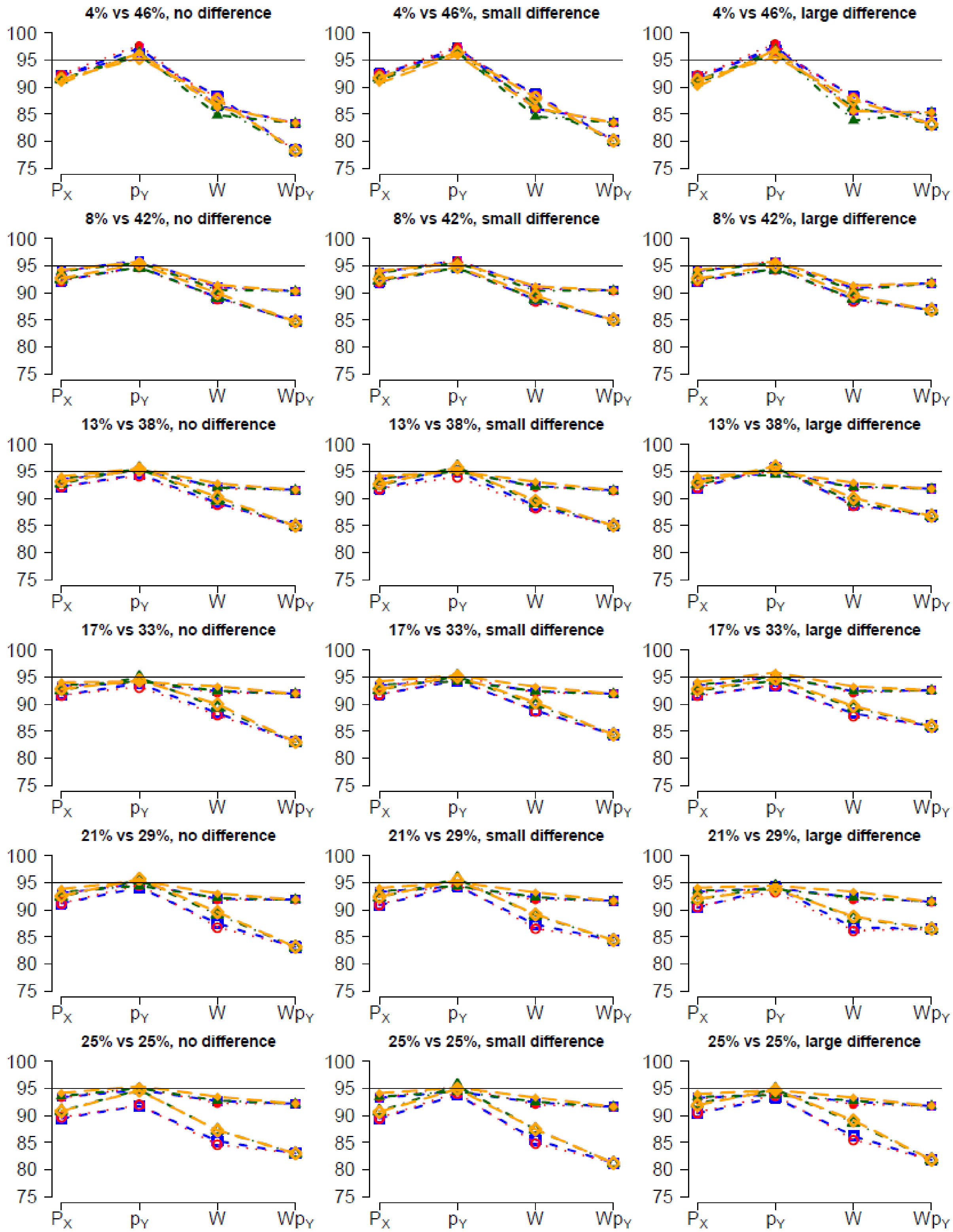

Figure 5. Mean 95% coverage levels of the CIs for all the parameter matrices (Simulation study n. 2) distinguished by levels of relevance and strength of the components. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds when R = 4 and by the corresponding empty symbols when R = 3. The horizontal line indicates the desired coverage level.

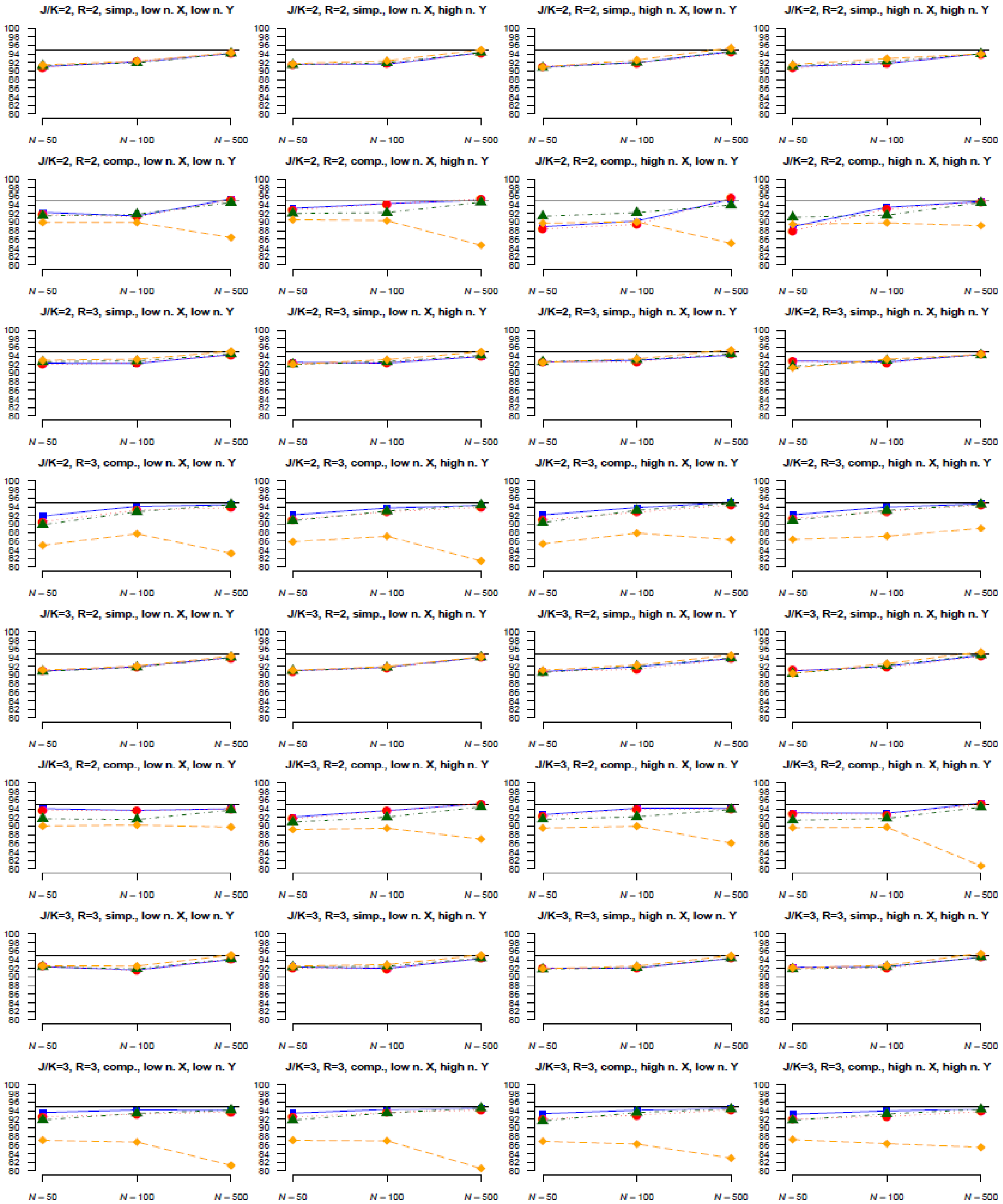

Figure S1. Mean 95% coverage levels of the CIs for $\mathbf{P}_X$ ('high' values, Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

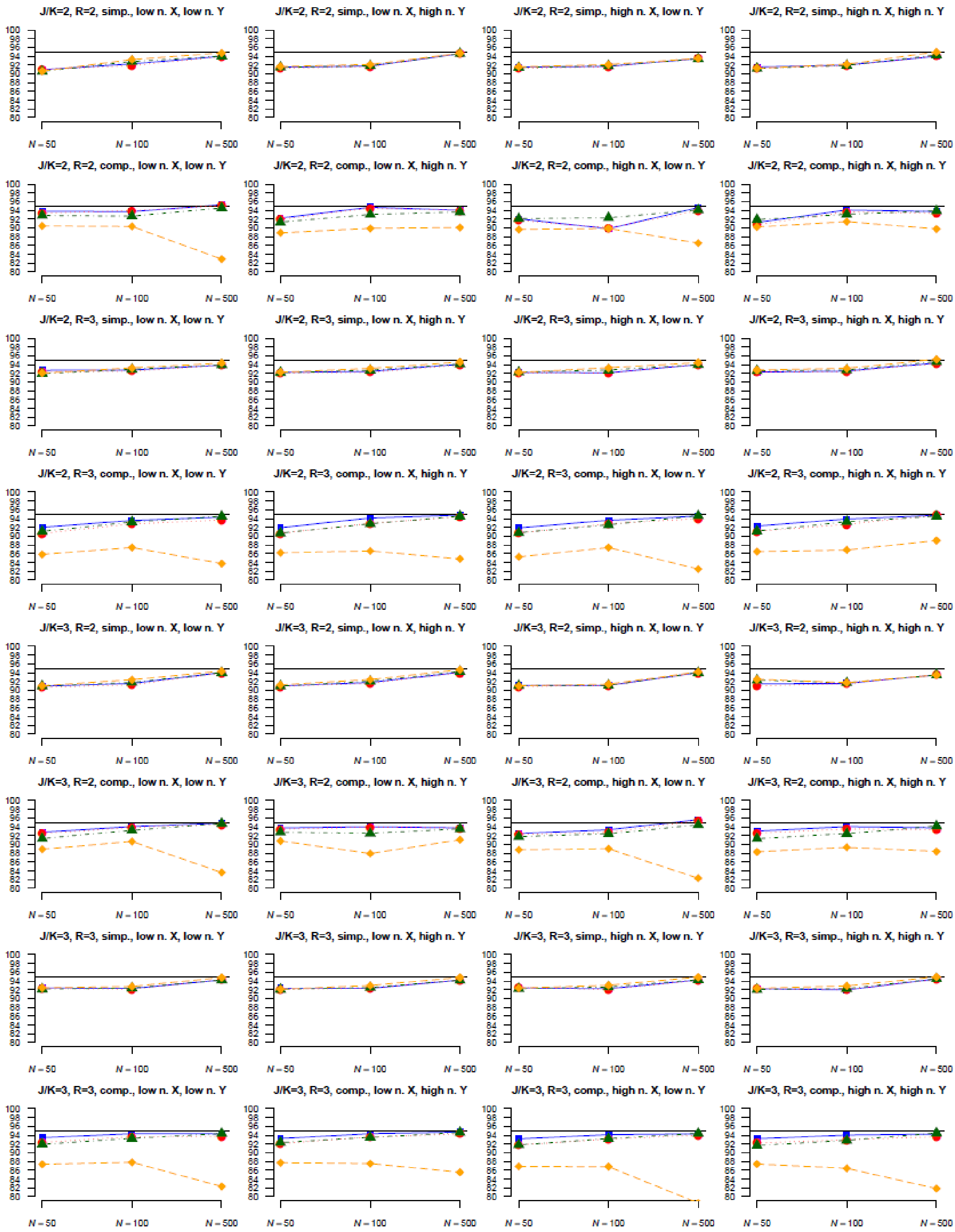

Figure S2. Mean 95% coverage levels of the CIs for $\mathbf{P}_X$ (‘low’ values, Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

Figure S3. Mean 95% coverage levels of the CIs for $\mathbf{P_Y}$ (Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

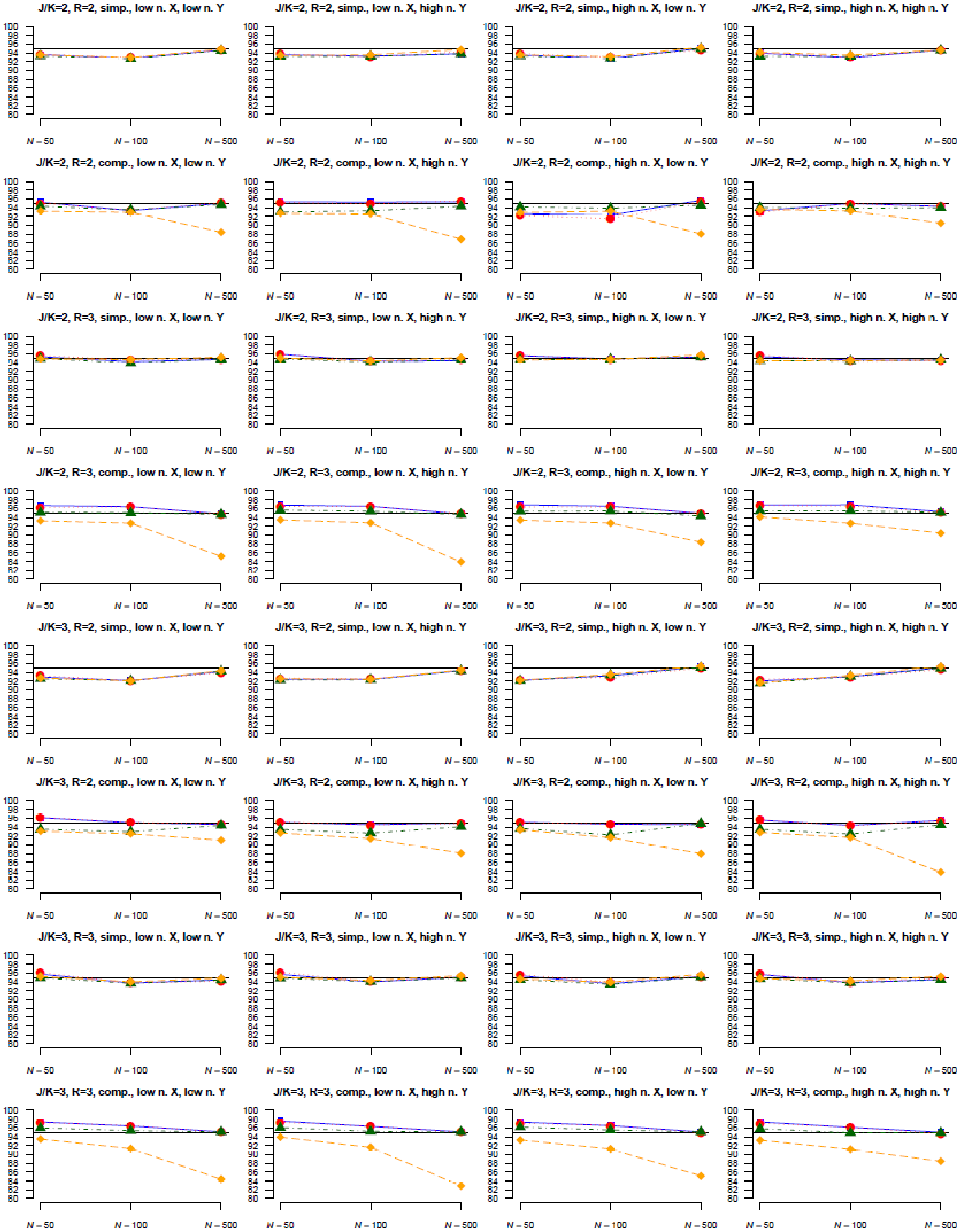

Figure S4. Mean 95% coverage levels of the CIs for $\mathbf{P}_{\mathbf{Y}}$ ('high' values, Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

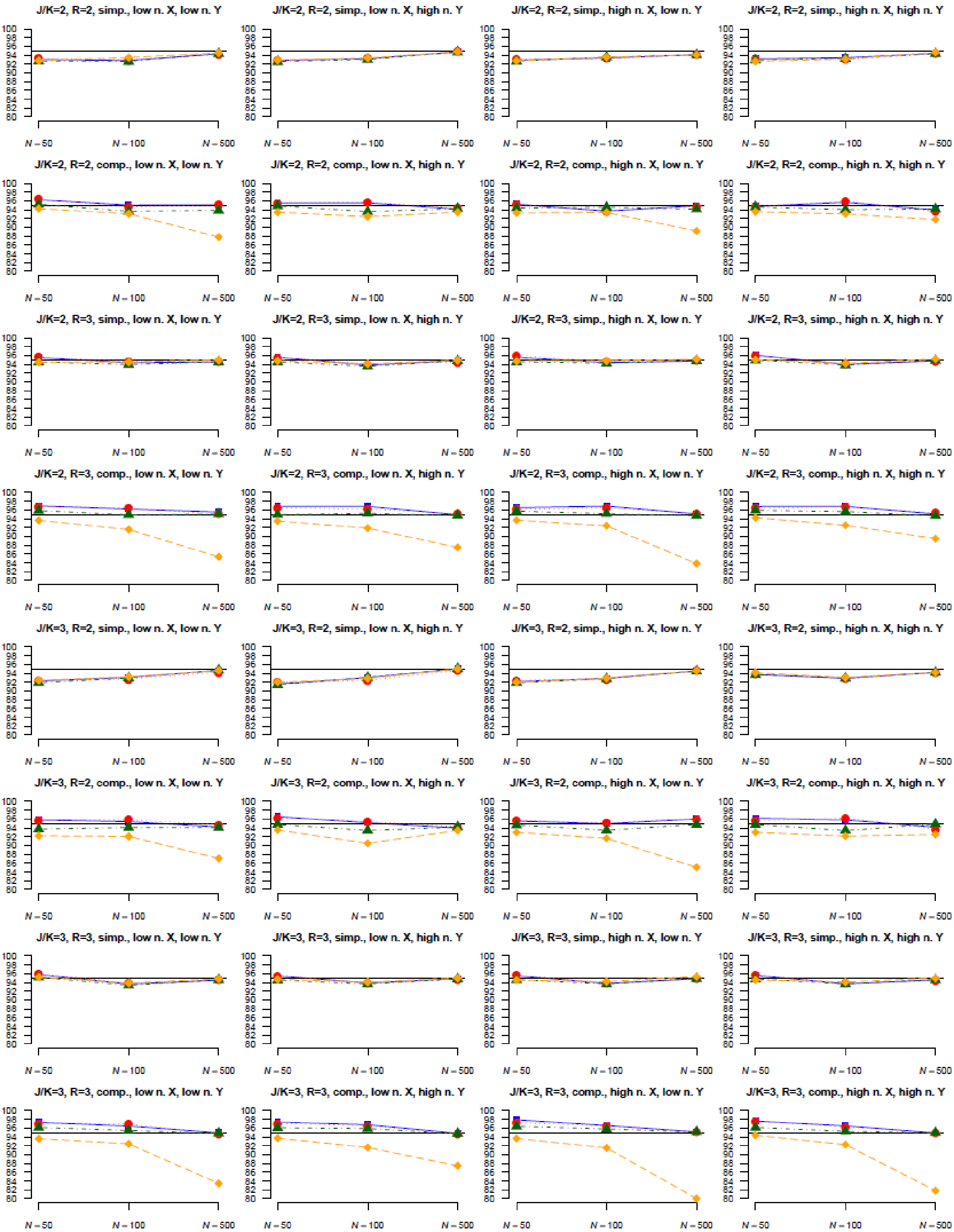

Figure S5. Mean 95% coverage levels of the CIs for $\mathbf{P}_Y$ ('low' values, Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

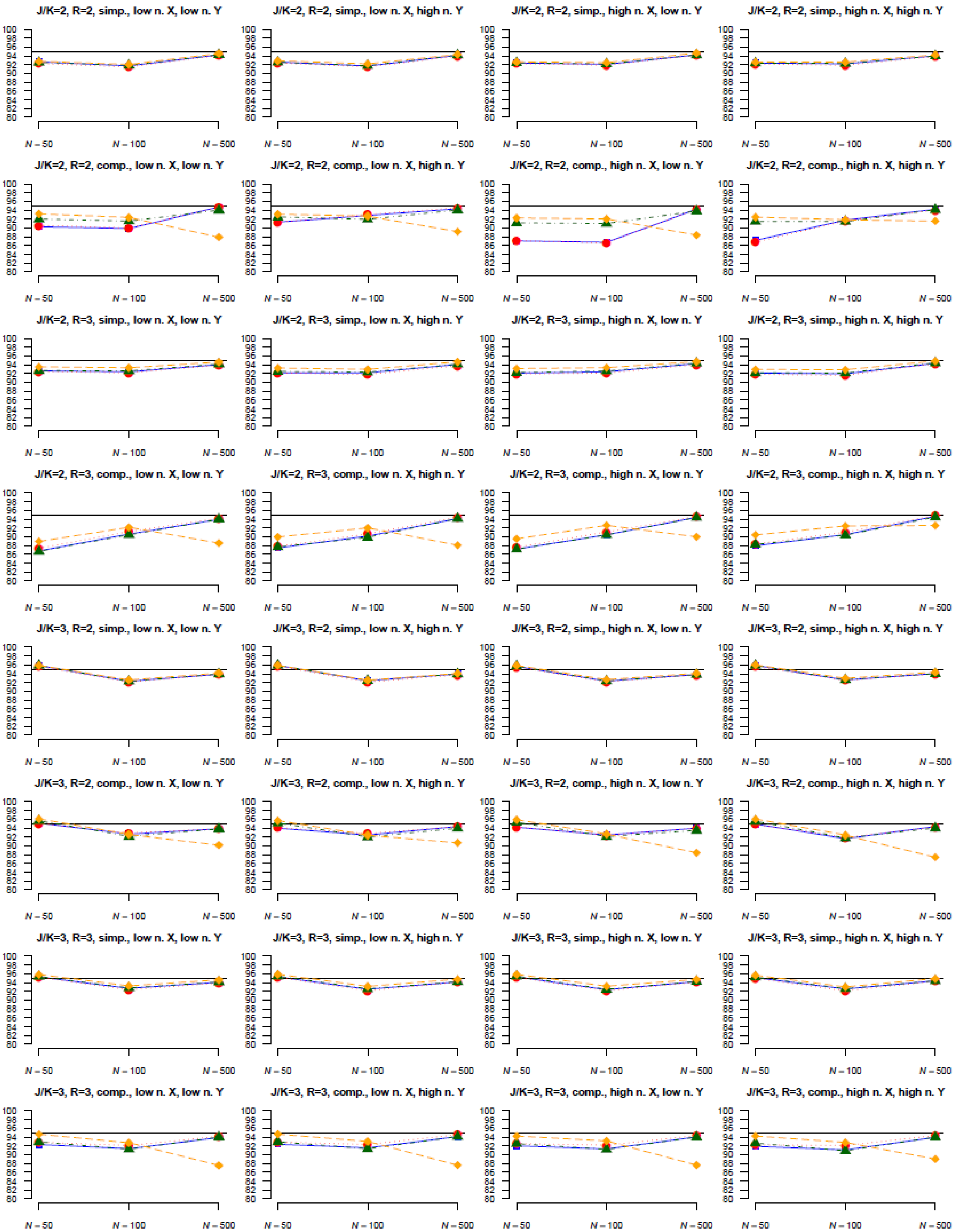

Figure S6. Mean 95% coverage levels of the CIs for **W** (Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by, respectively, blue filled squares, red filled circles, green filled triangles and orange filled diamonds. The horizontal line indicates the desired coverage level.

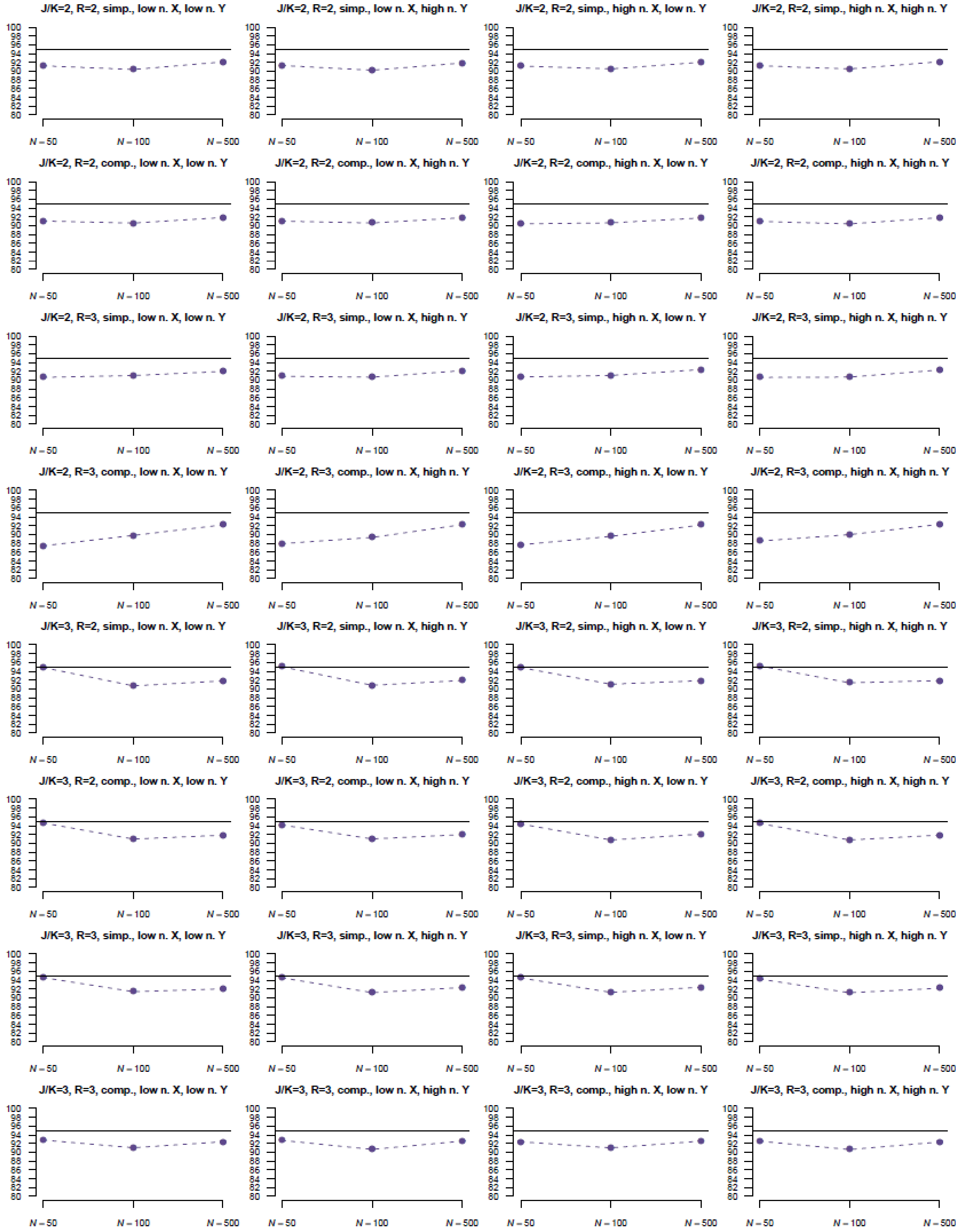

Figure S7. Mean 95% coverage levels of the CIs for **WP**$_{\mathbf{Y}}$ (Simulation study n. 1) for increasing sample sizes distinguished by all combinations of levels of the design variables. Levels for VSB, QSB, VSPB and QSPB are denoted by~~, respectively, blue filled squares, red~~ violet filled circles~~,~~

~~green filled triangles and orange filled diamonds~~. The horizontal line indicates the desired coverage level.

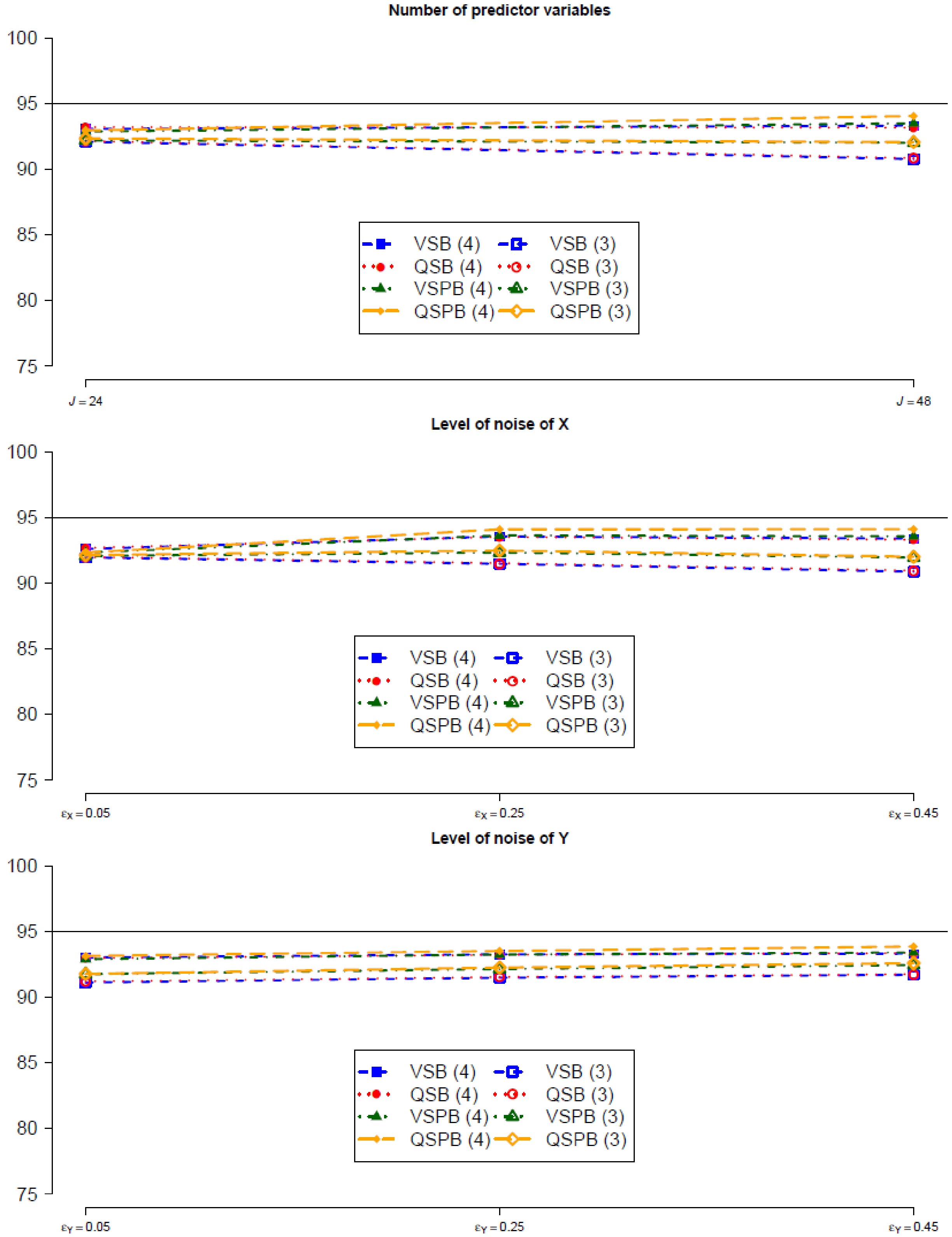

Figure S8. Mean 95% coverage levels of the CIs for $\mathbf{P}_X$ (Simulation study n. 2) distinguished by the design variables. The horizontal line indicates the desired coverage level.

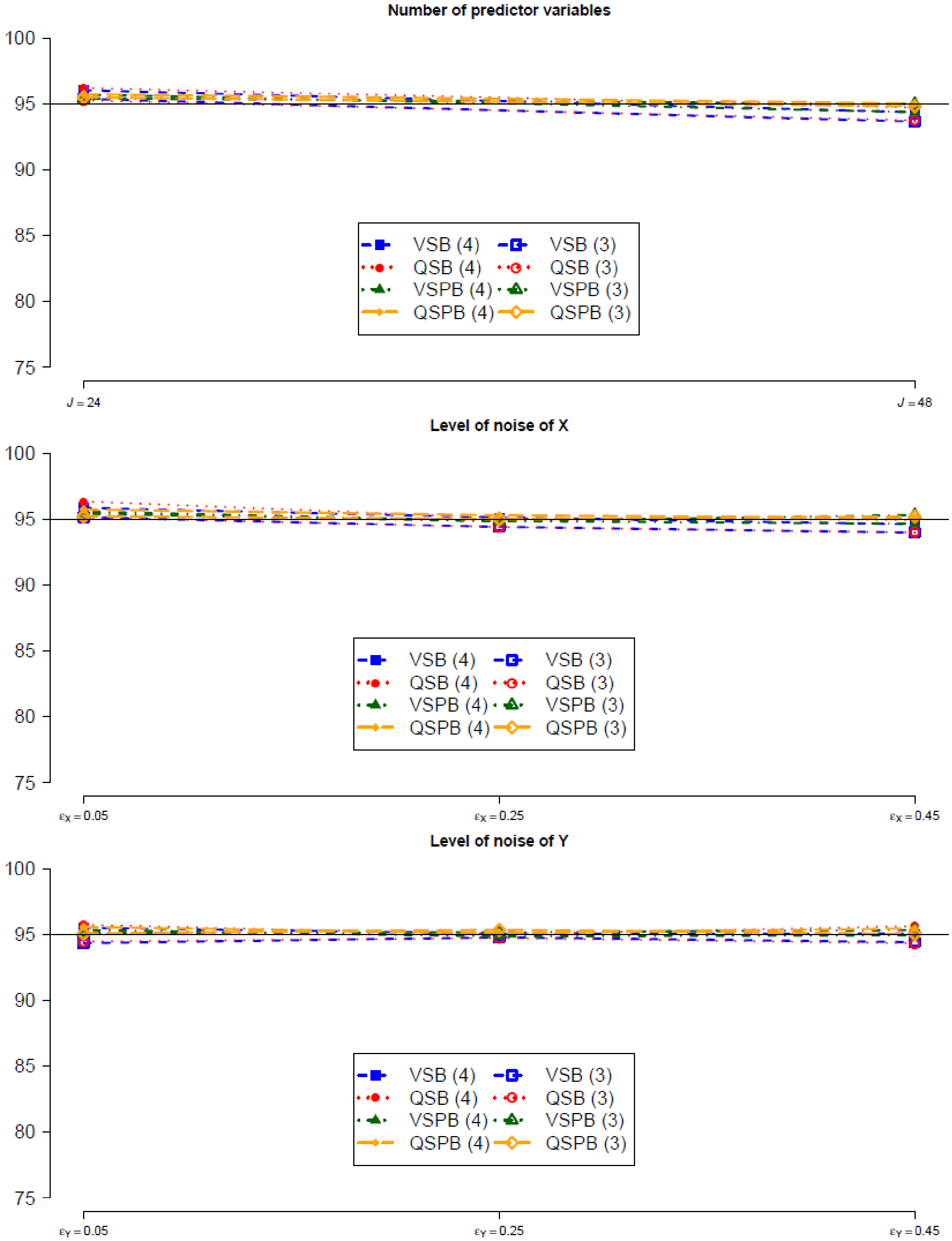

Figure S9. Mean 95% coverage levels of the CIs for $\mathbf{p_Y}$ (Simulation study n. 2) distinguished by the design variables. The horizontal line indicates the desired coverage level.

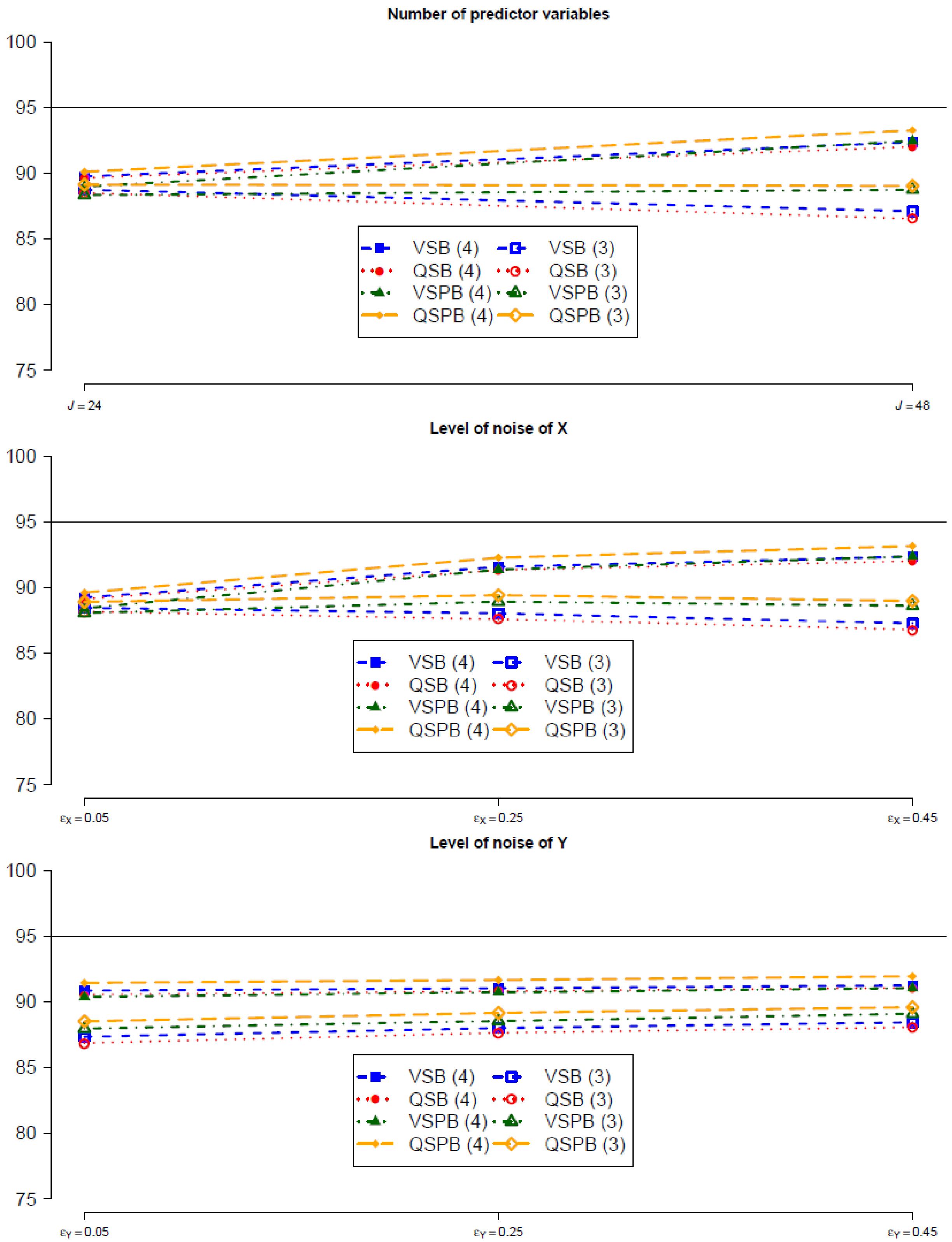

Figure S10. Mean 95% coverage levels of the CIs for **W** (Simulation study n. 2) distinguished by the design variables. The horizontal line indicates the desired coverage level.

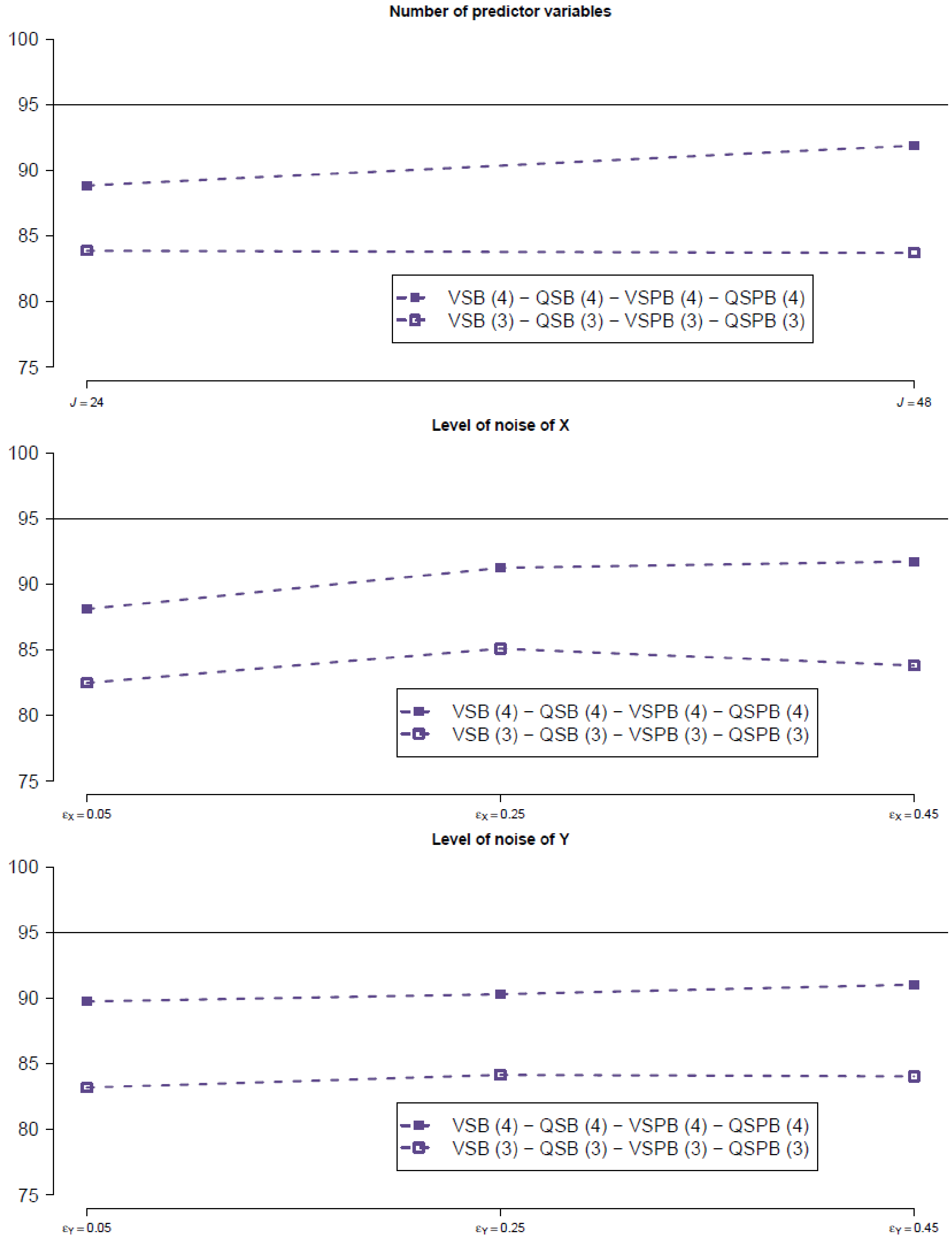

Figure S11. Mean 95% coverage levels of the CIs for **Wp**$_Y$ (Simulation study n. 2) distinguished by the design variables. The horizontal line indicates the desired coverage level.