



## Sample Size Requirements for Bifactor Models

Martina Bader, Lisa J. Jobst & Morten Moshagen

**To cite this article:** Martina Bader, Lisa J. Jobst & Morten Moshagen (2022) Sample Size Requirements for Bifactor Models, Structural Equation Modeling: A Multidisciplinary Journal, 29:5, 772-783, DOI: [10.1080/10705511.2021.2019587](https://doi.org/10.1080/10705511.2021.2019587)

**To link to this article:** <https://doi.org/10.1080/10705511.2021.2019587>



Published online: 05 Apr 2022.



Submit your article to this journal [↗](#)



Article views: 554



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

## Sample Size Requirements for Bifactor Models

Martina Bader , Lisa J. Jobst  and Morten Moshagen 

Ulm University

### ABSTRACT

Despite the widespread application of bifactor models, little research has considered required sample sizes for this type of model. As universal sample size recommendations are often misleading, we illustrate how to determine sample size requirements of bifactor models using Monte Carlo simulations in R. Furthermore, we present results of an extensive simulation study investigating the effects of the number of specific factors and indicators, loading magnitude, the relative general factor strength, and the validity of the proportionality condition on sample size requirements. Although a sample size of 500 was often sufficient to obtain acceptable convergence rates and parameter estimates, the exact sample size requirements depended on various model characteristics.

### KEYWORDS

Bifactor model; Monte Carlo simulation; R; sample size; simsem

In recent years, there has been a major upsurge of interest in the bifactor model as a specific type of confirmatory factor analysis (CFA) model (Reise, 2012; Zhang et al., 2021).<sup>1</sup> Across a broad range of disciplines, bifactor models are now routinely applied to study the latent structure of multidimensional data (e.g., Arens et al., 2021; Jorge-Botana et al., 2019; Moshagen et al., 2018; Shevlin et al., 2017; Waldman et al., 2020). In the standard bifactor model, as introduced by Holzinger and Swineford (1937), a general factor ( $G$ ) directly influencing all manifest variables  $i$  ( $i = 1, \dots, I$ ) is assumed along with a number of  $K$  orthogonal specific factors ( $S_k$ ;  $k = 1, \dots, K$ ) additionally influencing distinct subsets of variables. The score ( $Y_{ik}$ ) on a standardized manifest variable  $i$  of subset  $k$  can be expressed as

$$Y_{ik} = \lambda_{G_i}G + \lambda_{S_{ik}}S_k + \varepsilon_{ik}, \quad (1)$$



where  $\lambda_{G_i}$  and  $\lambda_{S_{ik}}$  denote the factor loadings on the general and specific factor, respectively, and  $\varepsilon_{ik}$  is a residual variable that is assumed to be uncorrelated with  $G$ , all  $S_1, \dots, S_K$ , and all  $\varepsilon_{i'k}$  for which  $i \neq i'$ .

Bifactor models thus partition the total variance of the manifest variables into three independent components: variance due to the general factor, variance due to the specific factor, and residual variance of each manifest variable. This variance decomposition makes the bifactor model attractive for many applications. For instance, it is instrumental for studying the psychometric properties of a measurement instrument, such as determining the relative strength of the general factor versus the specific factors, quantifying the degree of unidimensionality, and examining the reliability of

total and subscale scores (for details, see Rodriguez et al., 2016a). Furthermore, bifactor modeling lends itself well to testing the predictive validity of hierarchical constructs through scrutinizing the unique predictive contribution of the general and specific factors (e.g., Chen et al., 2006; Moshagen, *in press*; Zhang et al., 2021).

Statistically nested within the bifactor model is the higher-order model which assumes a set of first-order factors directly influencing distinct subsets of manifest variables, a second-order general factor directly influencing the first-order factors, and a residual factor for each first-order factor representing variance unaccounted for by the second-order factor (Rindskopf & Rose, 1988; Yung et al., 1999). However, higher-order models differ from bifactor models in the so-called proportionality condition, meaning that they imply an equal ratio of variance explained by the general factor and the first-order factor for each manifest variable in a subset (for details, see Mansolf & Reise, 2017; Yung et al., 1999). Bifactor models, on the other hand, are less restrictive and may accommodate data both conforming to and deviating from the proportionality condition (Bader & Moshagen, *in press*; Reise, 2012).

In general, the relative strength of the general factor to the specific factors is of central interest in many applications of the bifactor model (e.g., Bader et al., 2021; Carnovale et al., 2021; Kofler et al., 2020), and can be quantified using the *explained common variance* (ECV; Ten Berge & Sočan, 2004). The ECV indexes the proportion of common variance in the manifest variables accounted for by the general factor compared to the specific factors, and is

**CONTACT** Martina Bader  [martina.bader@uni-ulm.de](mailto:martina.bader@uni-ulm.de)  Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 47, 89081, Ulm, Germany.

This article was originally published with errors, which have now been corrected in the online version. Please see Correction (<http://dx.doi.org/10.1080/10705511.2022.2123641>)

<sup>1</sup>Also, several exploratory bifactor analysis methods have been developed (for a review and comparison, see Giordano & Waller, 2020). However, this study focuses exclusively on the confirmatory variant.

computed from the standardized factor loadings as follows (Rodriguez et al., 2016a):

$ECV =$

$$\frac{\left(\sum_{i=1}^I \lambda_{G_i}^2\right)}{\left(\sum_{i=1}^I \lambda_{G_i}^2\right) + \left(\sum_{i=1}^I \lambda_{S_{i1}}^2\right) + \left(\sum_{i=1}^I \lambda_{S_{i2}}^2\right) + \cdots + \left(\sum_{i=1}^I \lambda_{S_{iK}}^2\right)} \quad (2)$$

A higher  $ECV$  indicates a higher relative general factor strength, with possible values ranging from zero (no common variance is explained by the general factor) to one (all of the common variance is explained by the general factor; Reise, 2012).

Notwithstanding the popularity and advantages of the bifactor model, sample size requirements for this type of CFA model remained largely unaddressed. This is unfortunate given that numerous studies demonstrated the limitations of absolute or rule-of-thumb-based sample size recommendations, but highlighted the importance of individual model characteristics for determining an adequate sample size (e.g., Gagné & Hancock, 2006; MacCallum et al., 1999; Marsh et al., 1998; Moshagen & Musch, 2014; Wolf et al., 2013). In light of the complexity of bifactor models—typically estimating many free parameters—a dedicated examination of the sample size requirements for this type of CFA model is warranted.

### Sample Size Requirements in Factor Analysis

Determining an adequate sample size is vital for ensuring appropriate rates of proper model convergence, accurate estimates of parameters and standard errors, and sufficient statistical power to test substantive hypotheses regarding the overall model and individual model parameters. However, the sample size requirements of a factor analytic model depend on many factors (Muthén & Muthén, 2002). Key determinants include the degree of factor determination and the communality of the manifest variables, with a larger number of manifest variables per factor and stronger loadings improving model convergence and parameter recovery (e.g., Gagné & Hancock, 2006; MacCallum et al., 1999; MacCallum et al., 2001; Marsh et al., 1998; Velicer & Fava, 1998). These two factors were also shown to interact, such that stronger loadings may partially compensate for a smaller number of indicators, and vice versa, more indicators may partially compensate for weaker loadings (Gagné & Hancock, 2006; Velicer & Fava, 1998), though the exact interactive effects are complex (Gagné & Hancock, 2006; Moshagen & Musch, 2014). A number of studies further investigated the effect of the number of latent factors on sample size requirements, overall suggesting that models including more latent factors require larger sample sizes (e.g., Jackson et al., 2013; Wolf et al., 2013). However, there were also exceptions to this finding, so that the number of latent factors appears to have less strong and straightforward effects on sample size requirements (Jackson et al., 2013; Wolf et al., 2013).

### The Present Work

Although previous simulation studies provided valuable insights into the sample size requirements of CFA models in general, they offer only limited guidance to substantive researchers deciding on an appropriate sample size for a bifactor model at hand. Specifically, previous findings on the sample size determinants of standard CFA models may be not fully generalizable to bifactor models as they are associated with some particularities (see e.g., Khojasteh & Lo, 2015). For instance, bifactor models are more complex than standard CFA models, estimating multiple loadings per manifest variable to allow for a decomposition of variance into general, specific, and residual components. As such, sample size requirements might be distinctly influenced by the magnitude of the general factor and specific factor loadings as well as the relative strength of general to specific factors. Furthermore, as the specific factors are residualized for the variance due to the general factor, they often exhibit rather small variance, which, in turn, may increase the sample size requirements of bifactor models. Also, bifactor modeling allows for determining specific outcomes such as the  $ECV$  which are frequently relied on in substantive research and should thus be accurately estimated. Finally, given the various determinants of sample size requirements and their interactive effects, forming recommendations that apply to all model and data characteristics is highly challenging (Muthén & Muthén, 2002). Often, a superior approach is to conduct a Monte Carlo simulation tailored to the model and outcomes of interest. In a Monte Carlo simulation, random data of size  $N$  are repeatedly generated from a hypothesized population model with known parameter values and then analyzed to evaluate whether the realized sample size is sufficient to ensure stable model convergence, an accurate recovery of population parameter values, or adequate statistical power (Paxton et al., 2001).

The purpose of this article is twofold. First, we aim to illustrate how to conduct a Monte Carlo simulation to determine the sample size requirements of bifactor models using the *simsem* package (Pornprasertmanit, Miller, Schoemann, Jorgensen, & Quick, 2021) for R (R Core Team, 2021). Although there are existing articles on performing Monte Carlo simulations using the software *Mplus* (e.g., Muthén & Muthén, 2002), the *simsem* package has the advantage that it is open source and free of charge. Furthermore, *simsem* offers various options for model specification and data generation, and it is highly flexible in extracting specific simulation outcomes given that all results are stored in an accessible object in the R environment (for more information, see <https://simsem.org>). The second purpose is to perform an extensive Monte Carlo simulation study to examine the impact of various model characteristics (i.e., number of specific factors and indicators per specific factors, loading magnitude, relative factor strength, and validity of the proportionality condition) on the sample size requirements of bifactor models. In doing so, several outcomes (i.e., convergence rates, parameter estimates, standard errors, and  $ECV$  estimates) are evaluated. Using an illustrative model, we will also study the impact of missing data,

non-normally distributed indicators, and ordinal manifest variables on sample size requirements.

Taken together, the present work provides a comprehensive guide for researchers and users engaging in sample size planning for bifactor models as it provides a step-by-step illustration on performing a Monte Carlo simulation study, gives insights into the key determinants of sample size requirements of bifactor models, and additionally demonstrates how practically relevant conditions affect the required sample size.

## Determining Sample Size Requirements via Monte Carlo Simulation

In this section, we will illustrate how to perform a Monte Carlo simulation to decide on the sample size requirements of bifactor models using the R package *simsem* (Pornprasertmanit et al., 2021). As *simsem* relies on the *lavaan* package (Rosseel, 2012), both packages should be installed either in their official version from CRAN or in their latest development version from GitHub.<sup>2</sup> Following installation, the packages can be activated using the command `library(simsem)`. The R scripts of this tutorial are provided at the Open Science Framework (OSF; <https://osf.io/uq5yb/>).

### Model Specification

The first step in performing a Monte Carlo simulation is to specify the population model for which we aim to determine the required sample size. In our example, the population model is a bifactor model comprising one general factor and three specific factors with three indicators each. In addition, the values of all parameters of the population model need to be defined, for instance, by relying on theoretical assumptions or prior evidence. For simplicity, let us assume general factor loadings of .60 and specific factor loadings of .40 for our bifactor model. We can then specify the population model in line with syntax style of *lavaan*<sup>3</sup> and assign it (`<-`) to a variable denoted `populationModel` as follows:

```
populationModel <- '
# define loadings
G =~ .6*y1 + .6*y2 + .6*y3 + .6*y4 + .6*y5 +
.6*y6 + .6*y7 + .6*y8 + .6*y9
S1 =~ .4*y1 + .4*y2 + .4*y3
S2 =~ .4*y4 + .4*y5 + .4*y6
S3 =~ .4*y7 + .4*y8 + .4*y9
# define residual variances of indicators (= 1 -
.6^2 - .4^2)
y1 =~ .48*y1
```

```
y2 =~ .48*y2
y3 =~ .48*y3
y4 =~ .48*y4
y5 =~ .48*y5
y6 =~ .48*y6
y7 =~ .48*y7
y8 =~ .48*y8
y9 =~ .48*y9
# set factor variances to 1
G =~ 1*G
S1 =~ 1*S1
S2 =~ 1*S2
S3 =~ 1*S3
# set factor covariances to 0
S1 + S2 + S3 =~ 0*G
S1 + S2 =~ 0*S3
S1 =~ 0*S2
,
```

First, we define the respective indicators ( $y_1$  to  $y_9$ ) by which the general factor ( $G$ ) and the specific factors ( $S_1$  to  $S_3$ ) are measured ( $\sim$ ), while also specifying ( $*$ ) the magnitude of each individual factor loading. Next, we define the variances ( $\sim\sim$ ) of the residuals of each indicator by subtracting the squared factor loadings from 1 and set the variances of the latent factors to 1, so that all factor loadings are in a standardized metric. Finally, we set the factor covariances to zero to obtain orthogonal factors, which matches the standard bifactor specification.

The thereby defined population model will be used by *simsem* for data generation. Additionally, we need to specify a model that *simsem* will use for data analysis. Here, we specify an analysis model that correctly represents the latent structure in the population, and assign it to an object labeled `analysisModel`:

```
analysisModel <- '
G =~ NA*y1 + y2 + y3 + y4 + y5 + y6 + y7 + y8
+ y9
S1 =~ NA*y1 + y2 + y3
S2 =~ NA*y4 + y5 + y6
S3 =~ NA*y7 + y8 + y9
G =~ 1*G
S1 =~ 1*S1
S2 =~ 1*S2
S3 =~ 1*S3
S1 + S2 + S3 =~ 0*G
S1 + S2 =~ 0*S3
S1 =~ 0*S2
,
```

<sup>2</sup>The analyses in this article were performed using *simsem* development version 0.5-16.908.

<sup>3</sup>Alternatively, *simsem* allows to specify the population model using a matrix style approach or the *OpenMx* software. For details, see <https://github.com/simsem/simsem/wiki/Vignette>.

Note that the analysis model does not constrain individual parameter values, but only specifies the general model structure. The `NA*` argument preceding the first indicator of each factor is used to freely estimate the respective loading because the latent factors are identified by setting their variance to 1.

Having defined the models, we can now perform a Monte Carlo simulation using the `sim` command and store the results in an object labeled `results`:

```
results <- sim(n = 150,
generate=populationModel,
model=analysisModel, lavaanfun = "sem", nRep
= 500, completeRep = 5000, seed = 2021)
```

The first argument, `n`, defines the to-be-realized sample size. Using the `generate` and `model` arguments, we define the population and analysis model, respectively. The `lavaanfun` argument specifies the name of the function that is used for data analysis, which, in our case, is the `sem` function of *lavaan*. Next, we define the replications of our simulation (i.e., the number of random samples to be drawn), which should be sufficiently high to ensure stable results.<sup>4</sup> Here, we set `nRep` to 500 and `completeRep` to 5000, meaning that we aim to obtain a minimum of 500 properly converged solutions while defining a maximum number of 5,000 replications that stops the simulation even if our target number of 500 proper solutions has not been reached. Finally, we set a random seed to render the results reproducible.

## Results

To decide on the required sample size, various criteria may be assessed. First, we examine the rate of proper convergence, which is typically considered acceptable when being

.90 or higher (Gagné & Hancock, 2006; Moshagen & Musch, 2014). Using the function `summaryConverge(results)`, the following output is obtained:

```
$Converged
      num.converged num.nonconverged
           511           147

$Nonconvergent Reasons
Nonconvergent                count
Improper SE                  22
Nonpositive definite matrix   0
Optimal estimates were not guaranteed 125
Optimal estimates were not guaranteed 0
```

The first part of the output indicates that 511 out of 658 replications converged properly<sup>5</sup>, corresponding to an unacceptable convergence rate of .78. Furthermore, the reasons for nonconvergence are summarized, showing that solutions were discarded due to non-convergence of the estimation function and due to non-positive definiteness of latent or manifest (residual) covariance matrices. In the case of bifactor models with orthogonal factors, only the manifest residual covariances can be non-positive definite.

Another criterion to consider is the accuracy of parameter and standard error estimates. To this end, we use the function `summaryParam(results, detail=TRUE, improper=FALSE)`, where we specify our results object and indicate that we wish to obtain detailed summary statistics while excluding improper solutions. An excerpt of the obtained output is shown below:

The rows refer to the various parameters of the model, beginning with the general factor loadings followed by the specific factor loadings. The columns provide several summary statistics of the parameter estimates and standard

|        | Estimate<br>Average | Estimate SD | Average SE | Power (Not<br>equal 0) | Average<br>Bias | Coverage | Rel Bias | Rel SE Bias |
|--------|---------------------|-------------|------------|------------------------|-----------------|----------|----------|-------------|
| G=~y1  | 0.59                | 0.09        | 0.09       | 1.00                   | -0.01           | 0.93     | -0.01    | -0.05       |
| G=~y2  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.95     | -0.01    | 0.00        |
| G=~y3  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.95     | 0.00     | 0.01        |
| G=~y4  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.93     | -0.01    | -0.06       |
| G=~y5  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.93     | 0.00     | -0.03       |
| G=~y6  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.95     | 0.00     | -0.02       |
| G=~y7  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.96     | 0.01     | 0.00        |
| G=~y8  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.96     | 0.00     | 0.00        |
| G=~y9  | 0.60                | 0.09        | 0.09       | 1.00                   | 0.00            | 0.94     | 0.00     | 0.01        |
| S1=~y1 | 0.40                | 0.13        | 0.16       | 0.79                   | 0.00            | 0.98     | 0.01     | 0.21        |
| S1=~y2 | 0.40                | 0.13        | 0.16       | 0.78                   | 0.00            | 0.97     | 0.00     | 0.22        |
| S1=~y3 | 0.40                | 0.14        | 0.16       | 0.79                   | 0.00            | 0.97     | 0.01     | 0.16        |
| S2=~y4 | 0.39                | 0.13        | 0.16       | 0.75                   | -0.01           | 0.98     | -0.02    | 0.21        |
| S2=~y5 | 0.41                | 0.15        | 0.16       | 0.76                   | 0.01            | 0.96     | 0.01     | 0.12        |
| S2=~y6 | 0.40                | 0.14        | 0.17       | 0.76                   | 0.00            | 0.97     | 0.00     | 0.19        |
| S3=~y7 | 0.40                | 0.13        | 0.17       | 0.77                   | 0.00            | 0.98     | 0.00     | 0.29        |
| S3=~y8 | 0.40                | 0.14        | 0.16       | 0.78                   | 0.00            | 0.98     | 0.01     | 0.21        |
| S3=~y9 | 0.40                | 0.13        | 0.17       | 0.78                   | 0.00            | 0.98     | -0.01    | 0.30        |

<sup>4</sup>Users may test the stability of their results by running the simulation multiple times specifying different random numbers in the `seed` argument of the `sim` function.

<sup>5</sup>As the number of proper solutions specified in the `nRep` argument indicates only the minimum target number, the final number of proper solutions may exceed this number. This is because if there are non-converged solutions, *simsem* increases the total number of replications based on the estimated convergence rate such that it is expected to yield the target number of proper solutions. However, it is possible that within those replications more proper solutions are obtained.



**Table 1.** Convergence rates and relative biases in estimated parameters, standard errors (SE) and explained common variance (ECV) as a function of sample size.

| N   | Convergence rate | General factor loadings<br>Mean (min, max) | Specific factor loadings<br>Mean (min, max) | SE general factor loadings<br>Mean (min, max) | SE specific factor loadings<br>Mean (min, max) | ECV   |
|-----|------------------|--|---|---|--|-------|
| 200 | 0.89             | 0.00 (−0.01, 0.01)                         | 0.00 (−0.01, 0.02)                          | 0.00 (−0.04, 0.04)                            | 0.14 (0.01, 0.32)                              | −0.03 |
| 250 | 0.94             | 0.00 (−0.01, 0.00)                         | 0.00 (−0.01, 0.01)                          | 0.00 (−0.06, 0.08)                            | 0.05 (−0.03, 0.13)                             | −0.02 |
| 300 | 0.97             | 0.00 (0.00, 0.01)                          | 0.00 (−0.01, 0.00)                          | −0.01 (−0.05, 0.01)                           | 0.02 (−0.02, 0.08)                             | −0.01 |

errors. The first two columns summarize the average and standard deviation of the estimated parameters across all simulation solutions. The standard deviation of parameter estimates over a large number of replications may be regarded as the population value of the corresponding standard errors, which can be compared against the empirically estimated standard errors given in the column Average SE. The next column, Power (Not Equal 0), provides the proportion of solutions in which the parameters differed significantly from zero (given an  $\alpha$ -level of 0.05 per default). Here, the power was very high for the general factor loadings, but somewhat lower for the specific factor loadings. In the column Average Bias, the difference between the parameter estimates and the population parameter values are summarized. The Coverage column gives the proportion of solutions in which the  $(1 - \alpha)\%$  confidence interval covered the true population value, which should ideally be close to .95 given an  $\alpha$ -level of 0.05. Overall, the coverage in our example appears acceptable, ranging from 0.93 to 0.98 (Muthén & Muthén, 2002). The final two columns presented provide the average relative biases of parameter estimates and standard errors, calculated as  $RB_0 = (\hat{\theta} - \theta)/\theta$ , where  $\hat{\theta}$  is the estimated parameter (or standard error) and  $\theta$  is the corresponding population value. Absolute  $RB$  values less than .05 are typically considered negligible, values between 0.05 and 0.10 are considered moderate, and values larger than .10 are considered strongly (and thus unacceptably) biased (e.g., Bonifay et al., 2015; Flora & Curran, 2004; Muthén & Muthén, 2002). As can be seen, the relative bias in all factor loadings was negligible. The estimated standard errors of the general factor loadings, however, exhibited small to moderate biases, and the standard errors of the specific factor loadings were strongly biased.

Finally, the accuracy in estimating specific bifactor indices may be evaluated. To this end, we wrote an additional function (available at the OSF) that provides summary statistics and the relative bias for the estimated  $ECV$ . When running this function using `summaryECV(results, improper = FALSE)`, the following output is obtained:

|     | Population Value | Estimate Average | Estimate SD | Average Bias | Relative Bias |
|-----|------------------|------------------|-------------|--------------|---------------|
| ECV | 0.69             | 0.67             | 0.05        | −0.02        | −0.03         |

The results indicate that the  $ECV$  on average underestimates the population  $ECV$  by 0.02, corresponding to a relative bias of −0.03, which is acceptable.

Taken together, although the results demonstrate that a sample size of 150 was sufficient to obtain accurate parameter and  $ECV$  estimates, the unacceptable convergence rate and the moderate to strong relative biases in the standard error

estimates indicate that a larger sample size is required. Thus, a reasonable next step is to repeat the Monte Carlo study—successively increasing the sample size—until satisfactory results for all outcomes are obtained. Table 1 summarizes the convergence rate and relative biases in the estimated parameters, standard errors, and  $ECV$  when increasing the sample size in steps of 50. As can be seen, with a sample size of 300, a high convergence rate and acceptable relative biases in all parameters and standard errors are achieved.

## Simulation Study

The goal of the simulation study was to systematically investigate the impact of several model characteristics on the sample size requirements of bifactor models. To this end, a broad set of population models differing in the number of specific factors ( $K = 3, 5, 7$ , or  $10$ ), the number of indicators per specific factor ( $I_k = 3, 5, 7$ , or  $10$ ), the average magnitude of standardized general factor loadings ( $\lambda_G = 0.3, 0.5$ , or  $0.7$ ), and the relative strength of the general factor compared to the specific factors ( $ECV = 0.25, 0.50$ , or  $0.75$ ) was realized. Additionally, we manipulated whether the model structure conformed to the proportionality condition to investigate whether the greater complexity of data deviating from proportionality results in larger sample size requirements. To determine the necessary sample size, several outcomes (i.e., convergence rates, parameter and standard error estimates as well as the estimated  $ECV$ ) were evaluated. Based on previous theoretical and empirical work (e.g., Gagné & Hancock, 2006; Jackson et al., 2013; MacCallum et al., 1999, 2001), we predicted an improvement in outcomes and thus smaller required sample sizes for models with more indicators per specific factor and higher loadings on general and specific factors. However, we expected no genuine advantages of a higher relative factor strength because, in isolation, a high  $ECV$  does not guarantee high indicator communalities (MacCallum et al., 1999). Overall, we hypothesized larger sample size requirements for models with more specific factors, but expected rather small effects given somewhat mixed previous findings (e.g., Jackson et al., 2013; Wolf et al., 2013).

## Methods

### Model Specification

Population models were specified as standard bifactor models, comprising one general factor along with three to ten orthogonal specific factors. Each indicator loaded onto the general factor and on one specific factor, whereby the number of indicators was held equal across specific factors. All item residuals were constrained to mutual orthogonality.

Thus, a wide range of model sizes was realized, ranging from models with a total of  $I=9$  indicators (3 indicators  $\times$  3 specific factors) to models with  $I=100$  indicators (10 indicators  $\times$  10 specific factors). General factor loadings were sampled from a uniform distribution bounded by  $\lambda_G \pm 0.10$ , with  $\lambda_G$  conforming to the respective condition.

To construct population models perfectly conforming to the proportionality condition, the *ECV* across all indicators was set to 0.25, 0.50, or 0.75. However, to increase external validity, the *ECV* across the indicators of a particular specific factor varied slightly within population models (the standard deviation of  $ECV_k$  was 0.03 on average). Given a particular general factor loading and  $ECV_k$ , the corresponding standardized specific factor loading was then calculated by solving Equation (2) for  $\lambda_{Sik}$ :

$$\lambda_{Sik} = \sqrt{\frac{\lambda_{Gi}^2(1-ECV_k)}{ECV_k}} \quad (3)$$

This procedure ensured an equal proportion of variance explained by the general factor to the respective specific factor in all indicators within each subset  $k$ .

To construct population models deviating from the proportionality condition, we first calculated the average specific factor loading corresponding to the average general factor loading and the *ECV* of the respective condition, and then sampled the individual specific factor loadings from a uniform distribution bounded by  $\lambda_S \pm 0.10$ . In so doing, the ratio of variance explained by the general factor to the respective specific factor varied between the indicators within each subset as well as between the specific factors within population models (the standard deviation of  $ECV_k$  was again 0.03 on average).

To avoid inadmissible parameter values and computational problems due to high indicator communalities,<sup>6</sup> only particular *ECV* conditions were realized depending on the magnitude of general factor loadings (Table 2), resulting in a non-fully crossed simulation design. In total, 2 (proportionality condition)  $\times$  4 (number of specific factors)  $\times$  4 (number of indicators per specific factor)  $\times$  6 (combinations of  $\lambda_G$  and *ECVs*) = 192 different population models were defined.

### Data Generation and Analysis

Data generation and analysis proceeded through the *simsem* package. Multivariate normally distributed data with sample sizes  $N=150, 300, 500, 1,000$ , and  $2,000$  were generated and analyzed using maximum likelihood estimation (as implemented in *lavaan*). Analysis models were correctly specified according to the respective population bifactor structure, and were identified by setting the variance of all factors to 1. For each of the 960 conditions (192 population models  $\times$  5 sample sizes), we intended to obtain 500 properly

**Table 2.** Realized conditions of average general factor loadings ( $\lambda_G$ ), average specific factor loadings ( $\lambda_S$ ) and corresponding explained common variance (*ECV*).

| $\lambda_G$ | $\lambda_S$ | <i>ECV</i> |
|-------------|-------------|------------|
| 0.30        | 0.52        | 0.25       |
| 0.30        | 0.30        | 0.50       |
| 0.30        | 0.17        | 0.75       |
| 0.50        | 0.50        | 0.50       |
| 0.50        | 0.29        | 0.75       |
| 0.70        | 0.40        | 0.75       |

Note. Individual factor loadings were sampled from a uniform distribution.

converged solutions, but set a maximum number of 5,000 replications in case some conditions led to convergence rates  $< 0.10$ . In cases where *simsem* produced more than 500 proper solutions, we included only the first 500 replications to ensure comparability across all conditions. Non-converged or improper solutions (e.g., negative variance estimates) were excluded from further analyses (Chen et al., 2001).

### Outcomes

As outcomes, we considered the rate of proper convergence as well as the accuracy of parameter estimates and standard errors. Convergence rates were calculated as the number of proper solutions divided by the total number of replications for a condition. The accuracy of parameter estimates was evaluated with respect to the relative bias in the loadings on the general and specific factors and the estimated *ECV*, given that these outcomes are arguably of most interest to substantive researchers. For the factor loading estimates, we first calculated the *RB* for each individual parameter, and then calculated the mean *RB* across all respective parameters (e.g., all general factor loadings) within a model. Analogously, the accuracy of the estimated standard errors of general and specific factor loadings was evaluated based on *RBs* comparing individual standard error estimates with their population counterparts, and then calculating the mean *RB* across all respective parameters within a model.

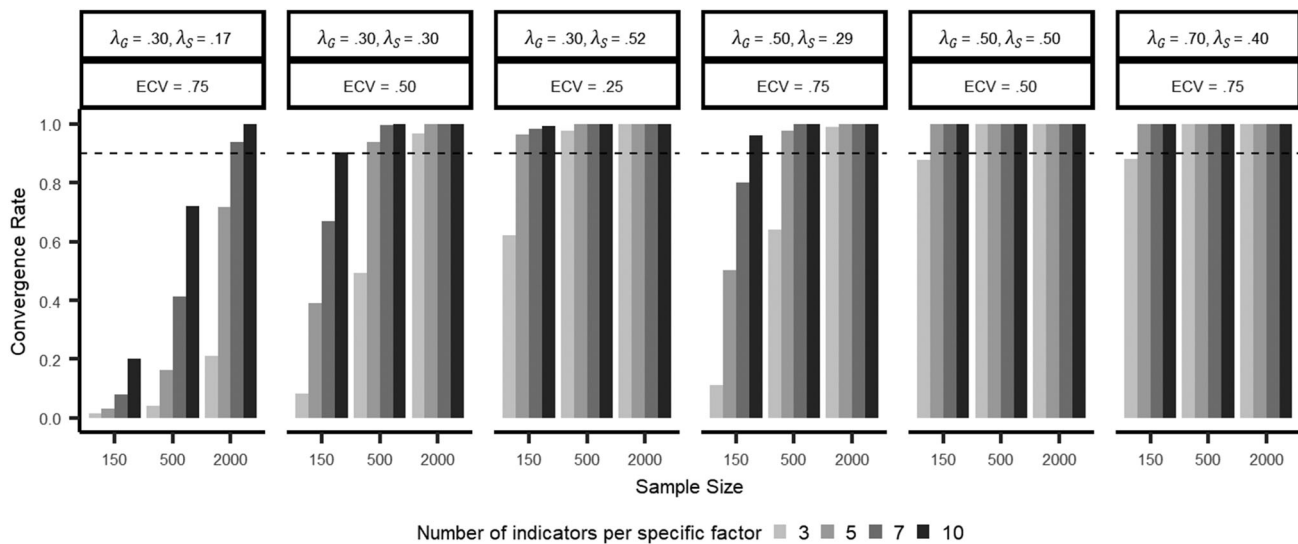
### Results

In what follows, a selected subset of results for the sample size conditions  $N=150, 500$ , and  $2,000$  is presented to illustrate the main results of the study. The complete results and supplemental materials are provided at the OSF. In general, increasing the sample size expectedly improved all outcomes. The validity of the proportionality condition had no meaningful effect on any of the outcomes and is thus not considered further.

### Convergence Rates

As can be seen in Figure 1, rates of proper convergence increased with sample size, the number of indicators per specific factor, and the magnitude of general factor loadings. In general, a sample size of 500 was sufficient to guarantee proper convergence in most conditions, whereas a sample size of 150 only led to satisfactory convergence rates when both general and specific loadings were strong. There was

<sup>6</sup>Specifically,  $\lambda_G = 0.70$  combined with  $ECV=0.50$  would translate into  $\lambda_S = 0.70$ , which yields a very high indicator communality ( $0.70^2 + 0.70^2 = 0.98$ ). Also,  $\lambda_G = 0.70$  combined with  $ECV=0.25$  translating into an inadmissibly high  $\lambda_S = 1.21$ , and  $\lambda_G = 0.50$  combined with  $ECV=0.25$  translating into  $\lambda_S = 0.87$  would yield indicator communalities  $> 1$ .



**Figure 1.** Median rates of proper convergence depending on the magnitude of factor loadings, explained common variance (ECV), number of indicators per specific factor, and sample size. *Note.* The dashed line indicates the minimum acceptable convergence rate of 0.90.

no main effect for the *ECV*, but conditional effects based on the absolute magnitude of general factor loadings, such that, given a constant magnitude of general factor loadings, convergence rates increased with higher specific factor loadings (and thus decreased with the *ECV*). The number of specific factors overall exhibited only small and varying effects on convergence rates (for details, see Table S1 at the OSF).<sup>7</sup>

### Parameter Estimates

The median relative bias in the estimated general factor loadings was negligible in virtually all conditions with acceptable convergence rates, and was at most moderate in conditions with unacceptable convergence rates (Figure S1 at the OSF). The estimated specific factor loadings were only negligibly biased given an acceptable convergence rate, but were moderately to strongly overestimated in conditions with very low convergence rates (Figure S2 at the OSF). Sampling variability of the estimated factor loadings crucially depended on the loading magnitude, whereby higher loadings on the general and specific factors, respectively, were associated with lower sampling variability. However, unlike the effect on convergence rates, given a constant loading magnitude, sampling variability decreased with higher *ECV*.

The median relative bias in the estimated *ECV* shown in Figure 2 was negligible in most conditions with acceptable convergence rates. However, in many conditions with unacceptable convergence rates, the *ECV* tended to be moderately to strongly underestimated (see Figure S3 at the OSF for details). A sample size between 150 and 300 was associated with acceptable median estimates of the *ECV* for the majority of population models, but the sampling variability

was still substantial in some conditions even at a sample size of 500.

### Standard Errors

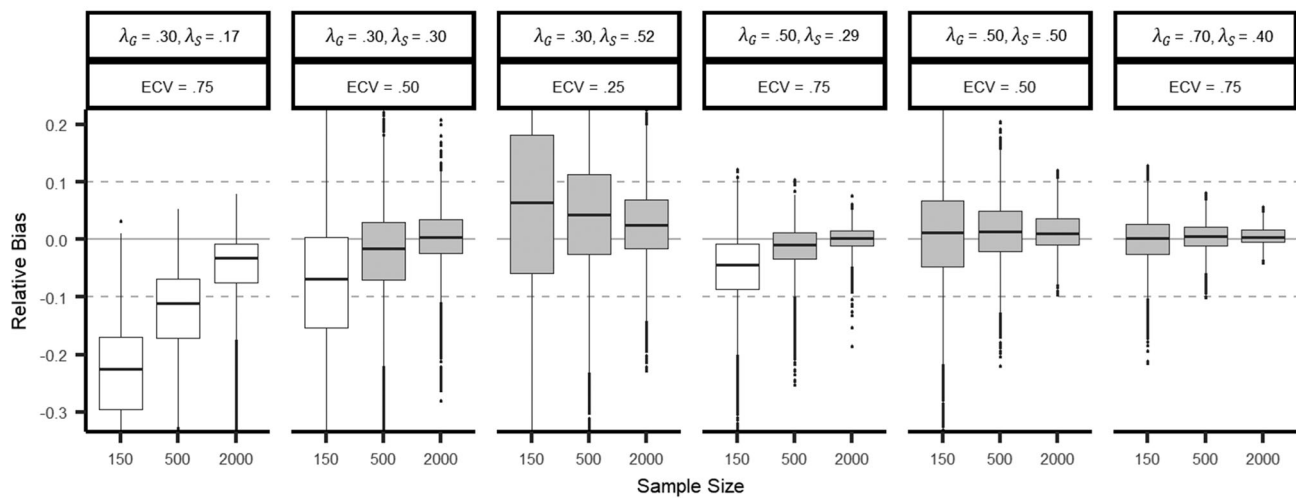
Figure 3 illustrates the relative bias in the estimated standard errors of the loadings on the general factor and the specific factors. In the majority of conditions with acceptable convergence rates, the standard errors of the general factor loadings were only negligibly to moderately underestimated. A stronger bias occurred for the estimated standard errors of the specific loadings even when the magnitude of general factor and specific factor loadings was comparable. In many conditions, the standard errors of the specific factor loadings were moderately to strongly underestimated. The relative bias in the standard error estimates decreased with an increasing magnitude of loadings on the general factor and the specific factors, respectively, and, given a constant magnitude of factor loadings, with higher *ECV* (for details, see Figures S4 and S5 at the OSF). Generally speaking, whereas sample sizes of around 300 were sufficient to obtain acceptable estimates for the standard errors of general factor loadings for the majority of population models, sample size requirements for acceptable estimates of the standard errors of specific factor loadings for the majority of population models were around 500.

### Summary

To summarize, the simulation study demonstrated that higher factor loadings and a larger number of indicators exhibited particularly beneficial effects on convergence rates and the recovery of parameters and standard errors. Nevertheless, although a diverse set of population models was realized, other influencing factors such as violations of distributional assumptions, missing data, or ordinal manifest variables were not considered. Thus, we will next illustrate how researchers and practitioners may also incorporate

<sup>7</sup>Note that in 6% of all conditions (55 out of 960), no 500 properly converged solutions were obtained within the maximum number of 5,000 replications (see Table S1). For these conditions, the remaining outcomes should be interpreted with caution.





**Figure 2.** Relative bias in the estimated explained common variance as a function of the magnitude of factor loadings, explained common variance (ECV), and sample size. *Note.* Conditions with acceptable median convergence rates of at least 90% are in gray. The boxes represent the median relative percentage bias along with the lower and upper quartile. The whiskers include 1.5 times the interquartile range above the upper quartile and below the lower quartile. Values outside the dashed lines indicate strong relative biases.

these factors into their Monte Carlo study when deciding on sample size.

## Additional Examples

### Non-Normal Indicators and Missing Data

Non-normally distributed manifest variables and incomplete data are prevalent in substantive research (e.g., Cain et al., 2017). Importantly, the effect of these factors on sample size requirements may also be considered in a Monte Carlo simulation. The first step is again to define the `populationModel` and `analysisModel` as shown above. Additionally, we specify the univariate distribution of the manifest variables using the `bindDist` function and store these in an object labeled `distribution`:

```
distribution <- bindDist(skewness = 0.5,
  kurtosis = 4, p = 9)
```

In our example, we aim to obtain a skewness of 0.5 and a kurtosis of 4 (values that are typical of non-normal psychological data; Cain et al., 2017) for all nine indicators, which `simsem` will generate using the approach by Vale and Maurelli (1983).<sup>8</sup>

To set the missing data mechanism and proportion of incomplete data, we employ the `miss` function as follows:

```
missing <- miss(pmMCAr = 0.10, m = 0)
```

We use the `pmMCAr` argument to obtain data missing completely at random (Little & Rubin, 2002) and specify that the proportion of missing values in all indicators should be

approximately 10%. By setting the `m` argument to 0, we further indicate that missing data should be addressed using full information maximum likelihood estimation.

We then perform the simulation using the `sim` function, whereby the `distribution` object is supplied to the `indDist` argument and the `missing` object is supplied to the `miss` argument. In addition, given that maximum likelihood estimation with uncorrected test statistics and standard errors yields biased results under the presence of missingness and non-normality, we set the `estimator` to `MLR`, indicating that robust Huber-White standard errors and a scaled test statistic that is asymptotically equal to the test statistic by Yuan and Bentler (2000) should be employed.

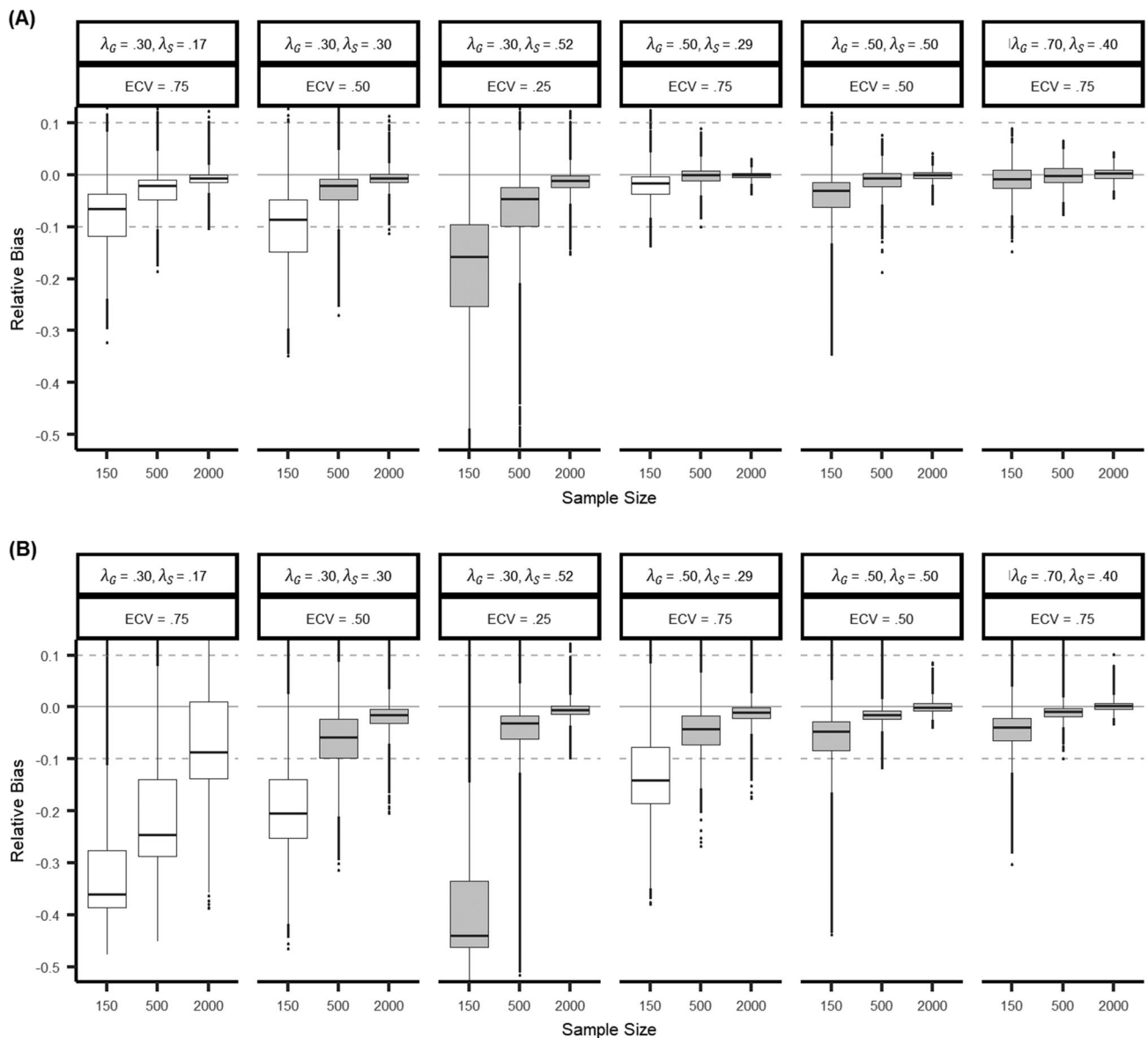
```
results <- sim(n = 300, generate =
  populationModel, model = analysisModel,
  lavaanfun = "sem", nRep = 500, completeRep =
  5000, indDist = distribution, miss = missing,
  estimator = "MLR", seed = 2021)
```

As shown in Table 3, in the presence of non-normal indicators and missing data, a sample size of 300 was not sufficient to obtain acceptable standard error estimates for all indicators in our example model. However, increasing the sample size to 350 led to satisfactory results in all outcomes.

### Ordinal Indicators

Many measures used in the social sciences rely on ordinal manifest variables to assess theoretically continuous constructs. We now want to test whether the sample size requirements of our previously introduced example bifactor model also hold when using five-categorical ordinal instead of continuous indicators. To do so, we first define our `populationModel` and `analysisModel` as shown above. Next, we create a function labeled `discretize` that discretizes continuous indicators into five categories using the thresholds  $-1.50$ ,  $-0.50$ ,  $0.50$ , and  $1.50$  (e.g., Rhemtulla et al., 2012).

<sup>8</sup>Note that although the approach by Vale and Maurelli (1983) to generate non-normal data is widely used, it is subject to criticism, such as being limited in the range of realizable skewness and kurtosis values or having multivariate properties similar to a multivariate normal distribution (e.g., Foldnes & Grønneberg, 2015; Olvera Astivia & Zumbo, 2015). Thus, users may employ alternative methods for data generation, for instance, using copulas (Mair et al., 2012) or a non-linear structural model (Auerswald & Moshagen, 2015).



**Figure 3.** Relative bias in the estimated standard errors of (A) the general factor loadings and (B) the specific factor loadings as a function of the magnitude of factor loadings, explained common variance (ECV), and sample size. *Note.* Conditions with acceptable median convergence rates of at least 90% are in gray. The boxes represent the median relative percentage bias along with the lower and upper quartile. The whiskers include 1.5 times the interquartile range above the upper quartile and below the lower quartile. Values outside the dashed lines indicate strong relative biases.

**Table 3.** Convergence rates and relative biases in estimated parameters, standard errors (SE) and explained common variance (ECV) in models with nonnormal indicators and missing data as a function of sample size.

| N   | Convergence Rate | General factor loadings<br>Mean (min, max) | Specific factor loadings<br>Mean (min, max) | SE general factor loadings<br>Mean (min, max) | SE specific factor loadings<br>Mean (min, max) | ECV   |
|-----|------------------|--|---|---|--|-------|
| 300 | 0.93             | 0.00 (−0.01, 0.01)                         | 0.00 (−0.02, 0.03)                          | −0.03 (−0.07, 0.00)                           | 0.07 (−0.04, 0.15)                             | −0.02 |
| 350 | 0.94             | 0.00 (−0.01, 0.00)                         | 0.00 (−0.01, 0.01)                          | 0.00 (−0.06, 0.05)                            | 0.03 (−0.03, 0.10)                             | −0.02 |

```

discretize <- function(data) {
  data[data > 1.5] <- 5
  data[data > 0.5 & data <= 1.5] <- 4
  data[data > -0.5 & data <= 0.5] <- 3
  data[data > -1.5 & data <= -0.5] <- 2
  data[data <= -1.5] <- 1
  data
}

```

This function can then be supplied to the `datafun` argument of the `sim` function so that all simulated data sets will be transformed accordingly. As ordinal indicators violate the assumptions of maximum likelihood estimation, we set the estimator to be diagonally weighted least squares estimation based on polychoric correlations combined with robust standard errors and a mean and variance adjusted test statistic ("WLSMV"; Muthén, 1984; Muthén et al., 1997).

**Table 4.** Convergence rates and relative biases in estimated parameters, standard errors (SE) and explained common variance (ECV) in models with ordinal indicators as a function of sample size.

| N   | Convergence rate | General factor loadings<br>Mean (min, max) | Specific factor loadings<br>Mean (min, max) | SE general factor loadings<br>Mean (min, max) | SE specific factor loadings<br>Mean (min, max) | ECV   |
|-----|------------------|--|---|---|--|-------|
| 300 | 0.95             | −0.03 (−0.04, −0.03)                       | −0.03 (−0.05, −0.02)                        | −0.01 (−0.06, 0.04)                           | 0.11 (0.05, 0.32)                              | −0.02 |
| 350 | 0.96             | −0.03 (−0.04, −0.03)                       | −0.03 (−0.05, −0.01)                        | 0.01 (−0.01, 0.02)                            | 0.05 (0.00, 0.12)                              | −0.02 |
| 400 | 0.98             | −0.04 (−0.04, −0.03)                       | −0.03 (−0.07, −0.01)                        | −0.01 (−0.07, 0.04)                           | 0.03 (−0.03, 0.08)                             | −0.02 |

```

results <- sim(n = 300,
generate=populationModel,
model=analysisModel, lavaanfun = "sem", nRep
= 500, completeRep = 5000,
datafun=discretize, estimator = "WLSMV",
seed = 2021)

```

The results indicate whereas a sample size of 300 ensured a high convergence rate and acceptable parameter estimates also for ordinal manifest variables, the relative bias in the standard errors increased to an unacceptable extent for some indicators (Table 4). Successively increasing the sample size up to  $N=400$  mitigated these biases to an acceptable extent.

## Discussion

In light of the widespread applications of bifactor models in many disciplines, the present work sought to provide an in-depth illustration on how researchers and users may decide on the sample size requirements of this type of CFA model by means of a Monte Carlo simulation approach. Employing the user-friendly R package *simsem* (Pornprasertmanit et al., 2021), we guided readers through the process of model specification, data generation and analysis as well as outcome interpretation. Additionally, it was demonstrated how to consider the impact of practically relevant conditions such as non-normally distributed manifest variables, ordinal indicators, and missing data in a Monte Carlo simulation. The results thereby suggested that the required sample size of bifactor models increases under these conditions (see also Muthén & Muthén, 2002; Wolf et al., 2013). Given the versatility of the *simsem* package, several other factors beyond the ones considered herein can be implemented in the simulation design. In particular, users can also easily assess the effect of model misspecification on sample size requirements by specifying an analysis model that differs from the population model used for data generation.

A general challenge in performing a Monte Carlo simulation is that a priori assumptions about the structure and parameter values of the population model are required. In the ideal case, users can rely on prior empirical evidence on the pattern and magnitude of factor loadings and the ECV in a model of interest. Alternatively, one can advance theoretical expectations about the degree of communality and relative general factor strength in a particular construct domain. For instance, an ECV of .70 or higher is typically considered to represent a very strong general factor and essential unidimensionality (Rodríguez et al., 2016b). As a robustness check, it is also possible to run multiple simulations with slightly varying model specifications or to define

random parameter values across the simulation replications in order to ensure that a particular sample size is sufficient under different conditions.

## Determinants of Sample Size Requirements for Bifactor Models

The second purpose of the present work was to perform a comprehensive simulation study to investigate the determinants of sample size requirements for bifactor models. To this end, we considered a broad set of population models differing in the number of specific factors, the number of indicators per specific factor, the magnitude of general factor loadings, the relative strength of the general factor compared to specific factors (i.e., ECV), and the proportionality of general to specific factor loadings, and then evaluated the rates of proper convergence as well as the accuracy of parameter and standard error estimates.

As hypothesized from previous analytical and empirical work (e.g., Gagné & Hancock, 2006; MacCallum et al., 1999, 2001; Wolf et al., 2013), a larger number of indicators per factor and a larger magnitude of general and specific factor loadings exhibited beneficial effects on convergence rates and parameter estimates. In line with our expectations, we found no main effect of the relative general factor strength on the outcomes, but only conditional effects based on the absolute magnitude of general factor loadings. Specifically, given a constant magnitude of general factor loadings, convergence rates typically improved with higher specific factor loadings and thus a lower ECV. By contrast, sampling variability of the estimated factor loadings and the respective standard error estimates benefitted from a higher ECV given a constant loading magnitude. Knowledge of the ECV of a given model is thus only informative of the required sample size in combination with information about the absolute magnitude of factor loadings.

The number of specific factors overall exhibited small effects on most outcomes. However, it should be noted that in the present study the total number of indicators increased with an increasing number of specific factors. In situations with a constant total number of indicators, a larger number of specific factors may have adverse effects on outcomes given the resulting decrease in factor determination (MacCallum et al., 1999). Furthermore, whether the general and specific factor loadings were proportional for all indicators within the specific factors (as implied by a higher-order model) or deviated from the proportionality condition showed no relevant effect on any of the outcomes. Nevertheless, it is important to note that whereas the degree of deviation from the proportionality condition represents a continuum, in the present study, only a small degree of

deviation<sup>9</sup> was realized for each population model, which was necessary to maintain comparability of the average magnitude and variability of factor loadings across conditions. Thus, although our results were largely unaffected by the validity of the proportionality condition, it cannot be ruled out that a strong deviation from proportionality might still have an effect on model estimation.

Overall, most outcomes under scrutiny were influenced by the same model characteristics, and conditions with satisfactory convergence rates often also exhibited an acceptable recovery of population parameters. The estimated standard errors of the general factor loadings and especially the specific factor loadings, on the other hand, proved to be more sensitive and required larger sample sizes to avoid underestimation.

### Sample Size Requirements

In general, sample size requirements for the assessed bifactor models varied strongly depending on the model characteristics and the outcomes of interest, ranging from a required sample size of around 150 to a required sample size larger than 2,000. The findings thus illustrate the limitations of rules-of-thumb based on model complexity, given that a higher number of indicators actually benefitted many outcomes and that an increasing number of specific factors overall exhibited small and unsystematic effects. Generally speaking, in conditions with high general factor loadings  $\lambda_G = 0.70$  and high ECV = 0.75, sample sizes of 150 to 300 were typically sufficient to obtain acceptable convergence rates, parameter and standard error estimates. For models with moderate general factor loadings of  $\lambda_G = 0.50$  and ECV = 0.50, the minimum required sample size was in the range of 150 to 500. However, sample size requirements increased with increasing ECV, leading to a minimum sample size of 300 to 1,000 (particularly for models with few specific factors or few indicators per specific factor) for models with  $\lambda_G = 0.50$  and ECV = 0.75. In conditions with low general factor loadings of  $\lambda_G = 0.30$ , sample size requirements also increased with increasing ECV, with a minimum required sample size of 2,000 and larger for models with ECV = 0.75, a minimum sample size between 500 and 2,000 for most models with ECV = 0.50, and a minimum sample size between 300 and 500 for most models with ECV = 0.25 (except for models with only few specific factors or indicators per specific factor, for which sample sizes of around 1,000 were required). Of course, only a small set of sample size conditions was assessed in the present study, so that the exact sample size requirements are very likely to lie in between the realized conditions. It should also be mentioned that all models under scrutiny

were valid, so that the identified sample size requirements might not be fully generalizable to misspecified models (but see MacCallum et al., 2001). Finally, sample size considerations should also ensure sufficient statistical power for global model fit assessment and the main hypotheses of interest (Jobst et al., *in press*). These limitations notwithstanding, we hope that the herein provided tutorial and insights into the determinants of sample size requirements of bifactor models will aid researchers in future sample size planning.

### Data availability

The Supplementary materials are uploaded at the Open Science Framework (<https://osf.io/uq5yb/>).

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### References

- Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2021). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of Educational Research*, 91, 34–72.
- Auerswald, M., & Moshagen, M. (2015). Generating correlated, non-normally distributed data using a non-linear structural model. *Psychometrika*, 80, 920–937.
- Bader, M., & Moshagen, M. (in press). No probifactor model fit index bias, but a propensity toward selecting the best model. *Journal of Abnormal Psychology*. <https://doi.org/10.1037/abn0000685>
- Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment*, 33, 511–525.
- Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling*, 22, 504–516.
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49, 1716–1735.
- Carnovale, M., Taylor, G. J., Parker, J. D. A., Sanches, M., & Bagby, R. M. (2021). A bifactor analysis of the 20-item Toronto Alexithymia Scale: Further support for a general alexithymia factor. *Psychological Assessment*, 33, 619–628.
- Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41, 189–225.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, 29, 468–508.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466–491.
- Foldnes, N., & Grønneberg, S. (2015). How general is the Vale-Maurelli simulation approach? *Psychometrika*, 80, 1066–1083.
- Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, 41, 65–83.
- Giordano, C., & Waller, N. G. (2020). Recovering bifactor models: A comparison of seven methods. *Psychological Methods*, 25, 143–156.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41–54.

<sup>9</sup>Fitting higher-order models (assuming perfect proportionality) to the population variance-covariance matrices of the bifactor models deviating from the proportionality condition, revealed only a small degree of misspecification: The average population minimum of the maximum likelihood fit function ( $F_0$ ) was 0.10 ( $SD = 0.12$ ), the average root-mean-square error of approximation (RMSEA) was 0.012 ( $SD = 0.01$ ), and the average standardized-root-mean-square residual (SRMR) was 0.013 ( $SD = 0.01$ ).



- Jackson, D. L., Voth, J., & Frey, M. P. (2013). A note on sample size and solution propriety for confirmatory factor analytic models. *Structural Equation Modeling*, 20, 86–97.
- Jobst, L. J., Bader, M., & Moshagen, M. (in press). A tutorial on assessing statistical power and determining sample size for structural equation models. *Psychological Methods*. <https://doi.org/10.1037/met0000423>
- Jorge-Botana, G., Olmos, R., & Luzón, J. M. (2019). Could LSA become a “bifactor” model? Towards a model with general and group factors. *Expert Systems with Applications*, 131, 71–80.
- Khojasteh, J., & Lo, W.-J. (2015). Investigating the sensitivity of goodness-of-fit indices to detect measurement invariance in a bifactor model. *Structural Equation Modeling*, 22, 531–541.
- Kofler, M. J., Singh, L. J., Soto, E. F., Chan, E. S. M., Miller, C. E., Harmon, S. L., & Spiegel, J. A. (2020). Working memory and short-term memory deficits in ADHD: A bifactor modeling approach. *Neuropsychology*, 34, 686–698.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley & Sons, Inc.
- MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36, 611–637.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84–99.
- Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, 47, 547–565.
- Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, 61, 120–129.
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181–220.
- Moshagen, M. (in press). When a truly positive correlation turns negative: How different approaches to model hierarchically structured constructs affect estimated correlations to covariates. *European Journal of Personality*. <https://doi.org/10.1177/08902070211050170>
- Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, 10, 60–70.
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, 125, 656–688.
- Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115–132.
- Muthén, B. O., Du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. [http://www.statmodel.com/bmuthen/articles/Article\\_075.pdf](http://www.statmodel.com/bmuthen/articles/Article_075.pdf).
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620.
- Olvera Astivia, O. L., & Zumbo, B. D. (2015). A cautionary note on the use of the Vale and Maurelli method to generate multivariate, nonnormal data for simulation purposes. *Educational and Psychological Measurement*, 75, 541–567.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8, 287–312.
- Pornprasertmanit, S., Miller, P., Schoemann, A., Jorgensen, T., & Quick, C. (2021). *Package simsem: SIMulated structural equation modeling* (0.5-16.908) [Computer software]. <https://simsem.org>.
- R Core Team. (2021). *R: A language and environment for statistical computing*. <https://www.R-project.org/>.
- Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17, 354–373.
- Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23, 51–67.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21, 137–150.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98, 223–237.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36.
- Shevlin, M., McElroy, E., Bentall, R. P., Reininghaus, U., & Murphy, J. (2017). The psychosis continuum: Testing a bifactor model of psychosis in a general population sample. *Schizophrenia Bulletin*, 43, 133–141.
- Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69, 613–625.
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48, 465–471.
- Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3, 231–251.
- Waldman, I. D., Poore, H. E., Lunningham, J. M., & Yang, J. (2020). Testing structural models of psychopathology at the genomic level. *World Psychiatry*, 19, 350–359.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73, 913–934.
- Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30, 165–200.
- Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64, 113–128.
- Zhang, B., Sun, T., Cao, M., & Drasgow, F. (2021). Using bifactor models to examine the predictive validity of hierarchical constructs: Pros, cons, and solutions. *Organizational Research Methods*, 24, 530–571.