

Replicating Robitzsch (2022): A simulation protocol

Leonard Kosanke

The git hash is: 990b4

Credit

This simulation protocol is based on the ADEMP-PreReg Template for Simulation Studies (Siepe et al. 2023).

General Information

What is the title of the project?

Replication of Robitzsch (2022): Comparing the Robustness of the Structural after Measurement (SAM) Approach to Structural Equation Modeling (SEM) against Local Model Misspecifications with Maximum Likelihood estimation (ML) and Unweighted least square estimation (ULS)

Who are the current and future project contributors?

Leonard Kosanke, Valentin Kriegmair and Aaron Peikert.

Provide a description of the project.

This preregistration is for the individual replication study of the study by Robitzsch (2022), performed by Leonard Kosanke: We will compare SAM and traditional SEM methods in terms of bias, Mean Squared Error (MSE), and convergence rates in small to moderate samples. Additionally, We will investigate the impact of model misspecifications on the performance of these methods by varying residual correlations and cross-loadings from low sample sizes up to the population level. In order to do so in a comprehensive manner, We will closely align the structure of my replication to the one in the original study. In consequence, this project consists of multiple studies that will have the same naming as the original study. For sections that differ across studies, We will provide subsections for each individual study. For the other sections, it can be assumed that they are equal across all studies.

Did any of the contributors already conduct related simulation studies on this specific question?

No, we did not conduct previous simulation studies but base our studies on the previous literature on the topic (Robitzsch 2022).

Aims

What is the aim of the simulation studies?

The aim of this simulation study is to compare standard SEM ML and ULS estimation with LSAM, GSAM-ML and GSAM-ULS estimation of SEM in small to moderate sample sizes, as well as in presence of model misspecifications up to the population level.

Data-Generating Mechanism

How will the parameters for the data-generating mechanism (DGM) be specified?

Overview Studies 1-3

In the first set of simulation studies, we specify a two factor Confirmatory Factor Analysis (CFA) model, with each factor being measured by three items. The parameter of interest is the correlation between the two factors. We generate data parametrically, with all observed variables being multivariate normally distributed. The samples of size N is drawn independently of samples units. We choose the same data-generating parameters as in the original paper (see Equation 1).

Study 1

$$\Lambda = \begin{pmatrix} 0.55 & 0 \\ 0.55 & 0 \\ 0.55 & 0 \\ 0 & 0.45 \\ 0 & 0.45 \\ 0 & 0.45 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1.0 & \\ & 1 \end{pmatrix}, \quad \Psi = \begin{pmatrix} 0.6975 & & & & & \\ 0 & 0.6975 & & & & \\ 0 & 0 & 0.6975 & & & \\ \psi_{X1Y1} & 0 & 0 & 0.7975 & & \\ 0 & \psi_{X2Y2} & 0 & 0 & 0.7975 & \\ 0 & 0 & 0 & 0 & 0 & 0.7975 \end{pmatrix} \quad (1)$$

We estimate the CFA model by allocating the items X_i ($i = 1, 2, 3$) to Factor F_X , and Y_i ($i = 1, 2, 3$) to Factor F_Y . The model has a diagonal error covariance matrix and thus, is misspecified. Figure 1 contains the data-generating model.

Multiple data-generating scenarios will be simulated. First, with no correlated residuals ($\psi_{X1Y1} = \psi_{X2Y2} = 0$). Secondly, the only non-zero residual correlation will be ψ_{X1Y1} , which will be assigned a value of either 0.12 or -0.12. Lastly, both ψ_{X1Y1} and ψ_{X2Y2} are assigned values of either 0.12 or -0.12.

Study 1b

In this focused simulation study, the small-sample bias of the LSAM estimation method is investigated. We rely on the same model as Study one as portrayed in Equation 2, with 3 differences. Firstly, all factor loadings have the same value and are manipulated as detailed in section the next section. The factor correlation is also going to be varied as detailed in the same section. Lastly, we assume a correctly specified model without residual correlations present. Here, the paper is contradictory as in the text it says the model has

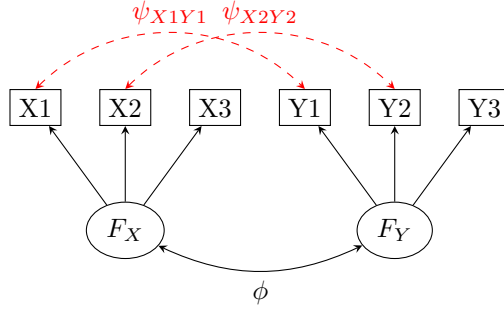


Figure 1: Simulation Study 1: Data-generating model. Residual correlations are shown with curved dashed red double arrows.

two residual correlations, but results are only reported for a model without residual correlations (Robitzsch 2022).

Study 2

$$\Lambda = \begin{pmatrix} 0.55 & \delta_{X1} \\ 0.55 & 0 \\ 0.55 & 0 \\ \delta_{X2} & 0.45 \\ 0 & 0.45 \\ 0 & 0.45 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1.0 & \\ 0.6 & 1 \end{pmatrix}, \quad \Psi = \begin{pmatrix} 0.6975 & & & & \\ 0 & 0.6975 & & & \\ 0 & 0 & 0.6975 & & \\ 0 & 0 & 0 & 0.7975 & \\ 0 & 0 & 0 & 0 & 0.7975 \\ 0 & 0 & 0 & 0 & 0 & 0.7975 \end{pmatrix} \quad (2)$$

We estimate the CFA model by allocating the items X_i ($i = 1, 2, 3$) to Factor F_Y , and Y_i ($i = 1, 2, 3$) to Factor F_X . The model is misspecified as it does not model the cross-loadings. Figure 2 contains the data-generating model.

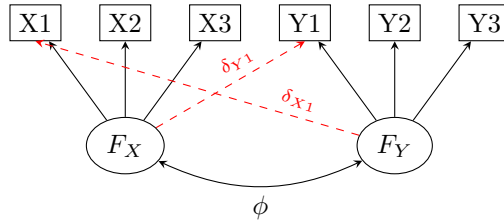


Figure 2: Simulation Study 2: Data-generating model. Cross-loadings are shown with dashed red arrows.

Multiple data-generating scenarios will be simulated. First, for one cross-loading, δ_{X1} will be either set to 0.3 or -0.3 and δ_{Y1} to 0. For the condition with 2 cross-loadings, both will be set to either 0.3 or -0.3. Again, the condition of having both positive or and negative cross-loadings present, is omitted, similar to Study 1.

Study 3

$$\Lambda = \begin{pmatrix} 0.55 & \delta_{X1} \\ 0.55 & 0 \\ 0.55 & 0 \\ 0 & 0.45 \\ 0 & 0.45 \\ 0 & 0.45 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1.0 & \\ 0.6 & 1 \end{pmatrix}, \quad \Psi = \begin{pmatrix} 0.6975 & & & & \\ 0 & 0.6975 & & & \\ 0 & 0 & 0.6975 & & \\ 0 & 0 & 0 & 0.7975 & \\ 0 & \psi_{X2Y2} & 0 & 0 & 0.7975 \\ 0 & 0 & 0 & 0 & 0 & 0.7975 \end{pmatrix} \quad (3)$$

We estimate the CFA model by allocating the items X_i ($i = 1, 2, 3$) to Factor F_X , and Y_i ($i = 1, 2, 3$) to Factor F_Y . The model is misspecified as it does not model the misspecification. Figure 3 contains the data-generating model.

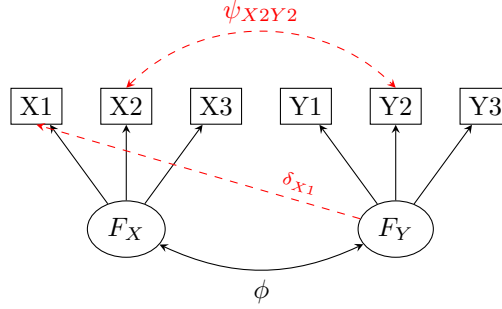


Figure 3: Simulation Study 3: Data-generating model. Cross-loading is shown with dashed red arrows, residual correlation is shown with curved dashed red double arrow.

One data-generating scenarios will be simulated, where the cross-loading δ_{X1} is set to 0.3 and the residual correlation ϕ_{X2Y2} to 0.12.

Study 4

Here, we will specify a five factor Confirmatory Factor Analysis (CFA) model, with each factor being measured by three items. The parameters of interest are the correlation between the factors. We generate data parametrically, with all observed variables being multivariate normally distributed. The samples of size N is drawn independently of samples units. We choose the same data-generating parameters and mechanisms as in the original paper. The paper proposes 4 data-generating models (DGM):

DGM1

In this model, no misspecifications are present. Item loadings were selected to achieve a reliability coefficient of 0.70, as measured by McDonald's Omega. The primary factor loading of each item was established to be greater than the secondary and tertiary loadings. Consequently, unlike in Simulation Studies 1 through 3, the data generation process did not assume equal factor loadings. For the replication, we chose to use the exact model parameters and resulting population-level covariance matrices as the original study by

Robitzsch (2022), which can be found at https://github.com/alexanderrobitzsch/supplement_sam/tree/main/Study4.

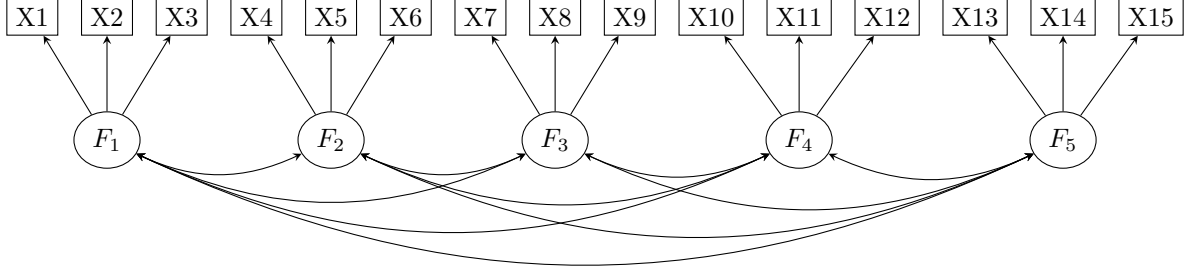


Figure 4: Simulation Study 4: Data-generating model 1 (DGM1).

We estimate the CFA model by allocating the items to the Factors as shown in Figure 4. The model is correctly specified and has a non-saturated covariance structure for the factors. This structure contains a common regression coefficient β as shown in Equation 4, which was set to 0.1.

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta & \beta & 0 & \beta & 0 \\ \beta & \beta & 0 & 0 & 0 \\ \beta & 0 & \beta & \beta & 0 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} \quad (4)$$

The vector comprising the residual factor variables, denoted as ξ_d ($d = 1, \dots, 5$), follows an independent normal distribution. Variances have been specified to ensure that $\text{Var}(F_d) = 1$ for all factor variables F_d ($d = 1, \dots, 5$). Thus, the model-implied covariance matrix $\text{Var}(F) = \Phi = \Phi(\beta)$ is a function of the regression parameter β .

DGM2

In this model, three cross-loadings are present, each of which are defined as $0.9 \cdot \lambda_0$, with λ_0 as the primary loading value. Besides, everything is the same as detailed in DGM1. DGM2 is displayed in Figure 5.

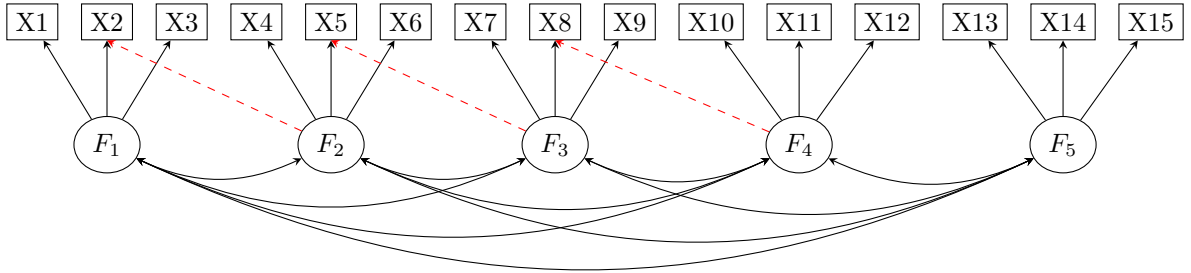


Figure 5: Simulation Study 4: Data-generating model 2 (DGM2).

The structural model is identical to DGM1 as detailed in Equation 4.

DGM3

In the third DGM, 20 residual correlations are present, whose values are detailed in the github repository of the original paper provided in DGM1. Besides, everything is the same as detailed in DGM1. DGM3 is displayed in Figure 6.

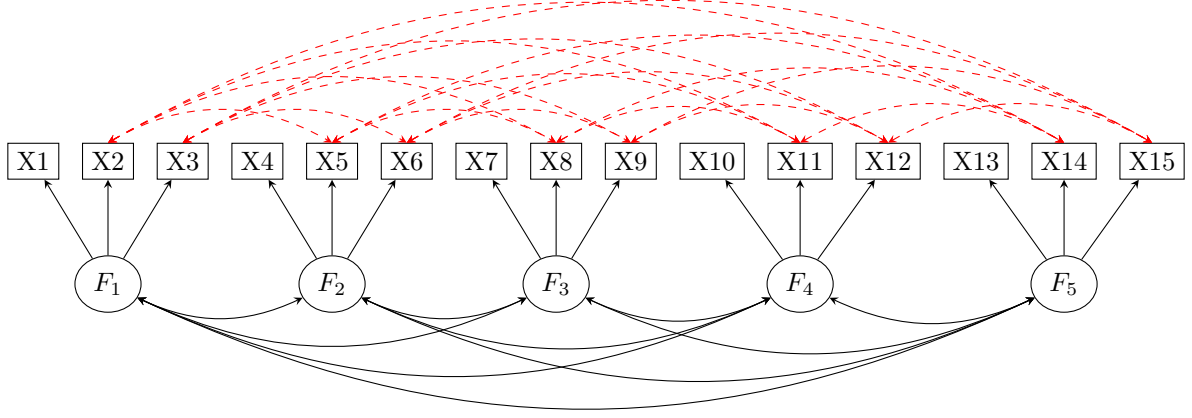


Figure 6: Simulation Study 4: Data-generating model 3 (DGM3).

Study 5

In study 5, we specify a three factor Confirmatory Factor Analysis (CFA) model with cross-loadings, with each factor being measured by three items. The parameters of interest are again the correlations between the factors. We generate data parametrically, with all observed variables being multivariate normally distributed. We will choose the same data-generating parameters as in the original paper: All factor loadings with regards to the same factor will be identical and set to $\lambda_1 = 0.6$, $\lambda_2 = 0.65$, and $\lambda_3 = 0.55$. The cross-loading δ_1 will be fixed to 0.4. The analysis will be conducted on the population level, using the true covariance matrix as input. Figure 7 contains the data-generating model for Study 5.

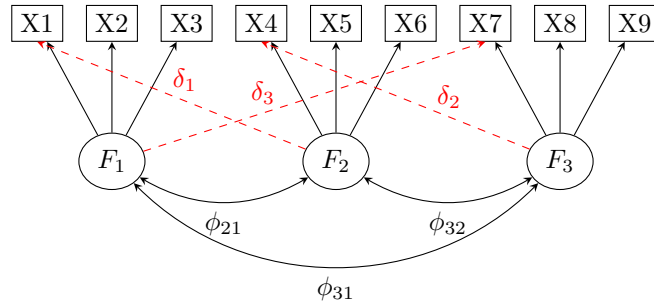


Figure 7: Simulation Study 5: Data-generating model with three factors. Factor correlations and cross-loadings are shown.

The other values are subject to manipulation and will be detailed in the next section.

Study 6

In study 6, we specify a three factor Confirmatory Factor Analysis (CFA) model with residual correlations, with each factor being measured by three items. The parameters of interest are again the correlations between the factors. We generate data parametrically, with all observed variables being multivariate normally distributed. We will choose the same data-generating parameters as in the original paper: The factor loading specification is identical to Study 5. The residual correlations ψ will be fixed to either 0.2 or 0.4 and has the same value for all correlations. The analysis will be conducted on the population level, using the true covariance matrix as input. Figure 8 contains the data-generating model for Study 6.

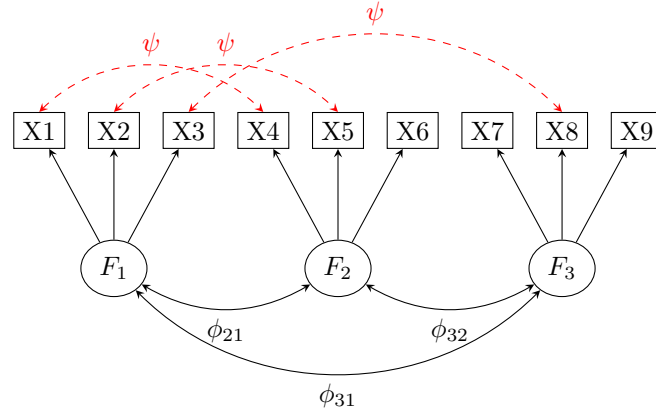


Figure 8: Simulation Study 6: Data-generating model with three factors. Factor correlations and residual correlations are shown.

The other values are subject to manipulation and will be detailed in section 4.2.

What will be the different factors of the data-generating mechanism?

Study 1

We will vary the factors:

- extent of misspecification with none, one or two correlated residuals $\psi_{X_1Y_1}$ and $\psi_{X_2Y_2}$, which, if present, take values of either 0.12 or -0.12, but not both negative and positive values in one condition.
- the sample size $N = 50, 100, 250, 500, 1000, 2500, 10^5$

Study 1b

We will vary the factors:

- factor loadings as $\lambda = 0.4, 0.5, 0.6, 0.7, 0.8$
- factor correlation $\phi = 0, 0.2, 0.4, 0.6, 0.8$

Study 2

We will vary the factors:

- extent of misspecification with one or two cross-loadings δ_{X1} and δ_{Y1} , which, if present, take values of either 0.3 or -0.3, but not both negative and positive values in one condition.
- the sample size $N = 50, 100, 250, 500, 1000, 2500, 10^5$

Study 3

We will vary the factor sample size with $N = 50, 100, 250, 500, 1000, 2500, 10^5$

Study 4

We will vary the factor sample size with $N = 50, 100, 250, 500, 1000, 2500, 10^5$.

The extent of misspecification can also be seen as a factor with 3 levels, along the DGMs 1-3.

Study 5

We will vary the factors:

- extent of misspecification with one, two or 3 cross-loadings. Thus, δ_2 and δ_3 , will be set to 0 or 0.4 depending on condition, with δ_1 always being set to 0.4.
- factor correlations that can vary with levels 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6.

Study 6

We will vary the factors:

- size of correlated residuals where ψ will be set to 0.2 or 0.4.
- factor correlations that can vary with levels 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6.

Justification for factor values and Additional simulation settings.

Study 1, 2, 3, 4

Specific values are already stated above and were chosen to be equivalent to the original study (Robitzsch 2022). We deviate from the values in sample size, as our research question explicitly focusses on small to moderate samples. To answer it, we included $N = 50$, to cover this range more specifically.

Studies 1b, 5, 6

The values chosen again are equivalent to the original study. No further justification is given there.

Additional simulation settings across studies

Additionally, to reduce complexity and answer the research question at hand, we refrain from including most of the estimation methods, and focus only on ML, ULS, their GSAM counterparts as well as LSAM estimation. Thus, we omitted all EFA-based methods estimated in the original study, as well as robust moment estimation, the latter of which did not contribute significantly to the results of interest across all studies. Even though In the original papers Study 4, robust moment estimation had very good results in large sample sizes, these are, again, not of interest for us.

If there is more than one factor: How will the factor levels be combined and how many simulation conditions will this create?

Study 1

We will implement a partly factorial design. This will result in 6 (sample size) x 5 (misspecification, with all combinations except one positive and one negative residual correlation) = 30 simulation conditions.

Study 1b

We will implement a fully factorial design. This will result in 5 (factor loadings) x 5 (factor correlations) = 25 simulation conditions.

Study 2

We will implement a partly factorial design. This will result in 6 (sample size) x 4 (misspecification, with all combinations except one positive and one negative residual correlation and no misspecification) = 24 simulation conditions.

Study 3

As there is only one factor sample size, there will be 6 simulation conditions.

Study 4

For the factor sample size, there will be 6 simulation conditions. This will be done across the 3 DGMs, so that one could speak of 18 conditions, if one counts the DGMs as a factor.

Study 5

We will implement a fully factorial design. This will result in $2(\delta_2) \times 2(\delta_3) \times 7(\phi_{21}) \times 7(\phi_{31}) \times 7(\phi_{32}) = 1372$ simulation conditions.

Study 6

We will implement a fully factorial design. This will result in $2(\psi) \times 7(\phi_{21}) \times 7(\phi_{31}) \times 7(\phi_{32}) = 686$ simulation conditions.

Estimands and Targets

What will be the estimands and/or targets of the simulation study?

Across Studies

Across studies, our estimand is the factor correlation estimated by the different estimation methods. We are not interested in the parameter sensitivity due to misspecification at the population level, that Robitzsch (2022) investigated.

Methods

How many and which methods will be included and which quantities will be extracted?

Studies 1,2,3

In the first set of studies, we will investigate ML, ULS, their GSAM counterparts (GSAM-ML and GSAM-ULS) as well as LSAM. The two GSAM estimation methods use ML and ULS in the second estimation step while having the measurement model parameters fixed to those obtained from the single CFA models in the first step. We will implement constrained ML estimation, where the loadings and variance parameters are given the constraint that they have to be positive and larger than 0.01. The reason for choosing these methods is, that they are central to the research question. The exact computation and software used will be based on, and where possible equivalent to the original study and its reported code.

Study 1b

For the focused simulation study 1b, only LSAM performance will be investigated to investigate the small sample bias of the method. The exact implementation will be done as in the code of the original paper.

Study 4

We will apply ML, ULS and LSAM estimation, as in the original paper, while omitting RME for reduced redundancy.

Studies 5 and 6

We will apply ULS and LSAM estimation, as in the original paper.

Performance Measures

Which performance measures will be used?

Studies 1,2,3

We will compute the bias, the standard deviation (SD), as well as the root mean square error (RMSE) for the estimated factor correlation.

Study 1b

We will compute the bias of the estimated factor correlation, as a function of the factor loadings λ and true factor correlation ϕ in the sample size $N = 100$.

Study 4

We will compute the average absolute and the average RMSE of the factor correlations.

Studies 5 and 6

In terms of performance, the same procedure as in the original paper will be utilized: The analysis focused on determining whether SAM was superior to SEM (i.e., ULS) regarding bias. An estimate of factor correlation was considered to have “No bias” (No bias) if both ULS and LSAM estimates presented absolute biases less than 0.05. SAM was deemed superior to SEM (“SAM better”) if the ULS estimate was biased and SAM showed a minimum percentage error reduction of 20% relative to ULS. Conversely, SEM was classified as superior to SAM (“SEM better”) when the LSAM estimate was biased, and ULS showed a minimum percentage error reduction of 20% compared to LSAM. In instances where neither of these conditions was met, both estimates were labeled as biased (“Both biased”). The percentage values of data constellations exhibiting each of these four scenarios was evaluated.

How many simulation repetitions will be used for each condition?

Across studies

We will perform 1500 repetitions for each simulation condition. This is based on the value decided on by Robitsch (2022), who did not provide reasoning for his choice.

How will missing values due to non-convergence or other reasons be handled?

Across studies

We do not expect convergence issues due to choosing constrained ML estimation. If we do observe non-convergence, we exclude these cases and report this.

How do you plan on interpreting the performance measures?

Across studies

We orient on the values and interpretations obtained in the original study for interpretation. This varied to some extent and might be extended to a more objective interpretation of values later on, if deemed necessary.

Other

Which statistical software/packages do you plan to use?

Across studies

All analyses will be conducted in R (R Core Team 2023), as in the original paper. The packages will be identical, where possible. If newer versions of some packages are available, we will use these new versions. For ML, ULS and LSAM estimation, we will use lavaan, using the same syntax provided in the original papers (Robitzsch 2022).

Which computational environment do you plan to use?

Across studies

Likely, we will run the simulation study on a Windows 11 machine. The complete output of `sessionInfo()` will be saved and reported.

Which other steps will you undertake to make simulation results reproducible?

Across studies

For all studies, we will upload the reproducible simulation script and a data set containing all relevant estimates, standard errors etc. for each iteration of the simulation to Github (<https://github.com/valentinkm/AdversarialSimulation>).

Is there anything else you want to preregister?

Across studies

We omitted the subsections referring to Monte Carlo Uncertainty, as they are not of interest for our research.

Bibliography

- R Core Team. 2023. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robitzsch, Alexander. 2022. “Comparing the Robustness of the Structural After Measurement (SAM) Approach to Structural Equation Modeling (SEM) Against Local Model Misspecifications with Alternative Estimation Approaches.” *Stats* 5 (3): 631–72. <https://doi.org/10.3390/stats5030039>.

Siepe, Björn S., František Bartoš, Tim P Morris, Anne-Laure Boulesteix, Daniel W. Heck, and Samuel Pawel. 2023. “Simulation Studies for Methodological Research in Psychology: A Standardized Template for Planning, Preregistration, and Reporting.” Preprint. PsyArXiv. <https://doi.org/10.31234/osf.io/ufgy6>.