

The following description of the simulation studies is based on the established structure for simulation studies coaligned with the other conducted studies to facilitate a potential collaboration:

Aims, objectives and research questions

All studies aimed to evaluate the performance of vanilla SEM (with maximum likelihood) compared to global SAM (gSAM), local SAM with maximum likelihood (lSAM-ML), and local SAM with unweighted least squares (lSAM-ULS) under various conditions. The two research questions we established prior to conducting the studies served as general basis for both studies:

1. How do SAM and traditional SEM methods (including ML and ULS) compare in terms of bias, Mean Squared Error (MSE), and convergence rates in small to moderate samples?
2. What is the impact of model misspecifications, such as residual correlations and cross-loadings, on the performance of SAM compared to traditional SEM methods?

The first two studies represent a conceptual replication of the studies conducted by Rosseel and Loh (2022) and Dhaene and Rosseel (2023). However, these studies just served as a basis for the design and several adjustments were made as described below. The third study was conducted as a result of the collaboration with Kosanke and the respective studies conducted. After examining the studies conducted by Kosanke and study 1 and 2 we did not jointly arrive at the conclusion that a collaborative simulation study is warranted (see Discussion for more detail). However I identified several arguments for setting up another simulation.

Population Models and Data Generation Mechanisms

Study 1:

Data were generated based on a 5-factor population structural model with 3 indicators for each factor. Four different models were simulated (see figure 1-4). In line with Rosseel and Loh (2022) this model design was chosen to represent a realistic model with sufficient

complexity to pose a challenge for the estimation methods, especially in the presence of misspecifications:

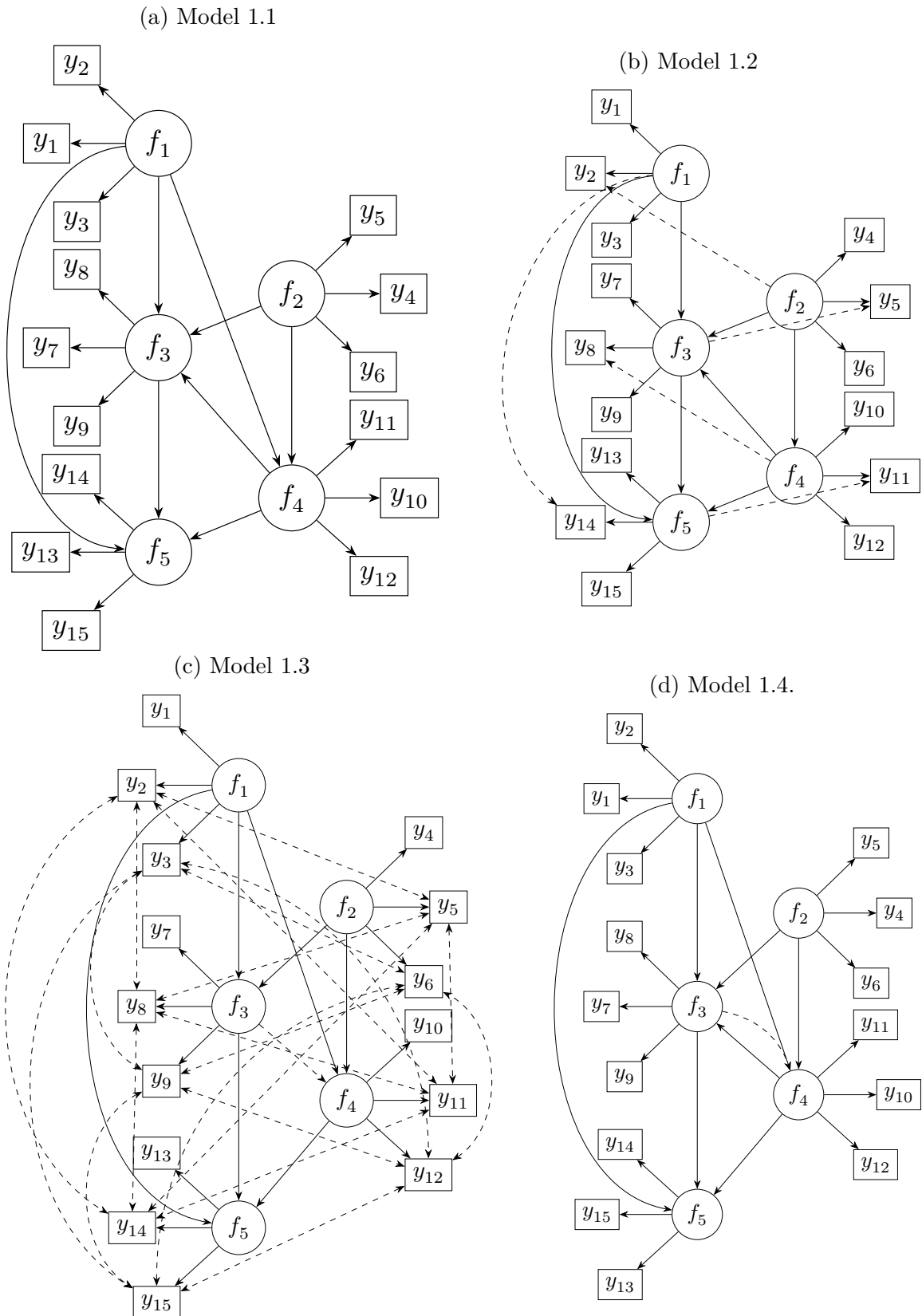
- Model 1.1: Correctly specified model.
- Model 1.2: Misspecified with cross-loadings in the population model that are ignored in the estimation model (model 1.1)
- Model 1.3: Misspecified with correlated residuals and a reversed structural path between the third and fourth latent factors in the population model that are ignored in the estimation model (model 1.1)
- Model 1.4: Misspecified with a bidirectional structural relation between factors 3 and 4 specified as only one directional

To investigate additional possible and realistic scenarios beyond the ones studied by Rosseel and Loh (2022) model 1.3 here included a combination of measurement and structural misspecifications as opposed to only measurement misspecifications to introduce an even more severely misspecified model under which SAM methods might perform even better than traditional SEM. Further, model 1.4 included a (not modelled) bidirectional structural relation between factors 3 and 4 as opposed to the unidirectional reversed one. For all models, the population-level values of the structural parameters were set to 0.1. Factor loadings were fixed across all reliability conditions, with the first indicator of each factor serving as the scaling indicator ($\lambda = 1.0$), and the other two indicators having loadings of 0.7. Indicator reliability levels were manipulated by adjusting the measurement error variances in the Θ matrix. Specifically, the a reliability value was set at different levels (low = 0.3, moderate = 0.5 or high = 0.7) to compute the respective error variances on the diagonal of Θ : $\Theta^* = \text{Var}(\eta)\Lambda^T \times \frac{1}{r-1}$.

Study 2:

Data were generated based on a 5-factor population structural model with 3 indicators for each factor with loadings set to 1, 0.9 and 0.8 for each factor and reliability modulated like in study 1. Regression weights were set to either 0.183 and 0.224 (low) or 0.365 and 0.447 (medium). This should represent varying variance explained (R^2) by the endogenous factors set at low ($R^2 = 0.1$) or medium ($R^2 = 0.4$). Note however that the computation of this was a simplification and does not accurately result in said R^2 values. The aim here was only to generally modulate between lower and higher regression weights. The

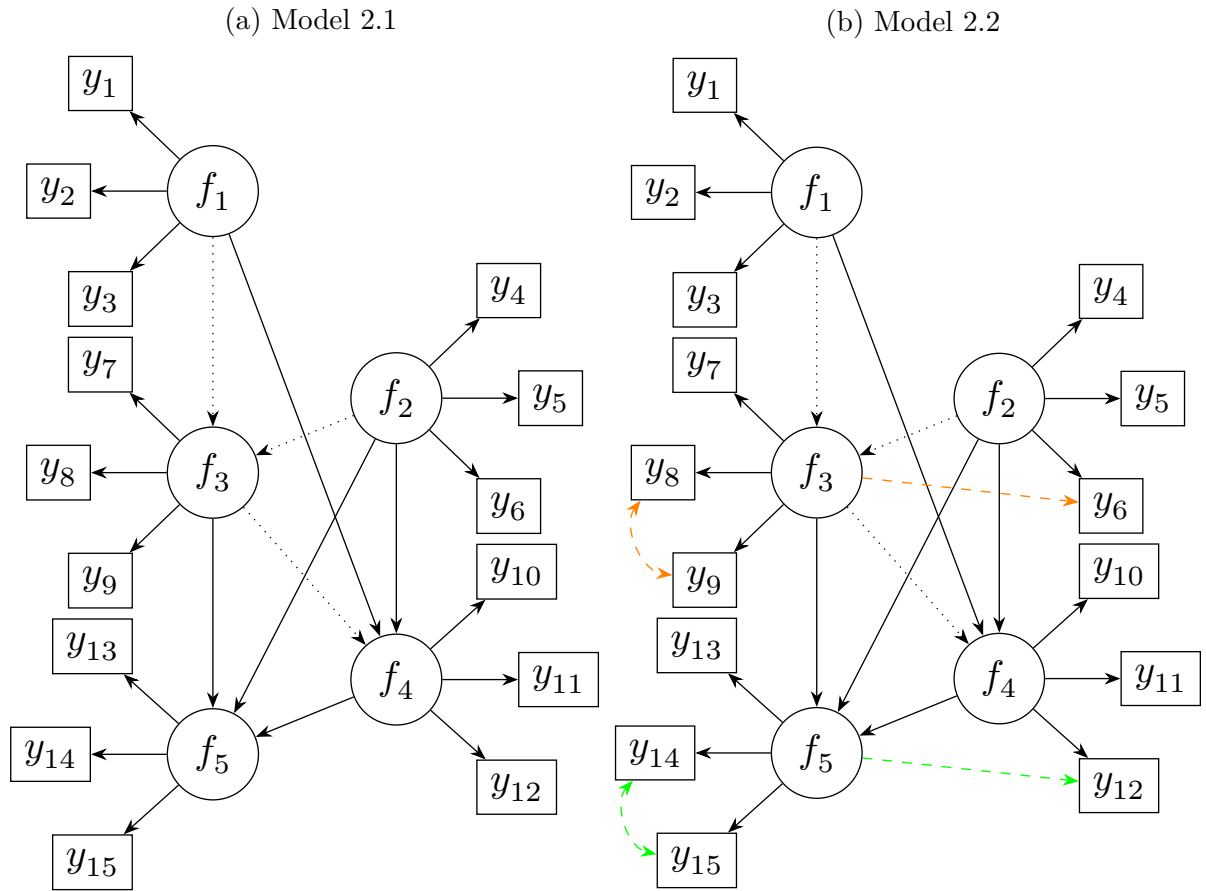
Figure 1: Four Population Model Variations of Study 1. *Note.* Error terms are not explicitly shown in the figure. Dashed lines represent relations omitted in the estimation model present in the population model.



population models resulted in the following model types with varying misspecification in the estimation model:

- Model 2.1: structural misspecification with falsely specified paths in the estimation model not present in the population model (Figure 5)
- Model 2.2: correlated residuals and a factor cross-loading in either the exogenous (Model 2.2-exo), endogenous (Model 2.2-endo) part of the model or both (Model 2.2-both) with falsely specified paths in the estimation model not present in the population model (see figure 5-6).

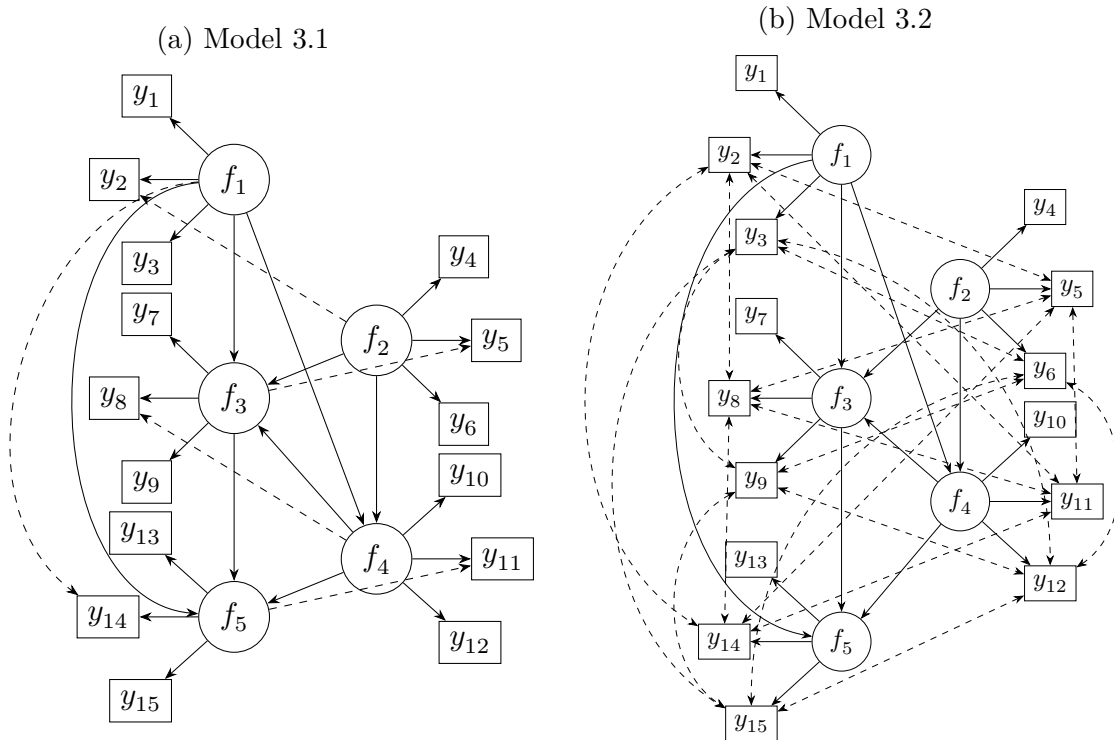
Figure 2: Population Model Variations of Study 2. *Note.* Error terms are not explicitly shown in the figure. Dotted paths represent relations specified in the estimation model not present in the population model. For Model 2.2, orange lines represent misspecifications in the exogenous part of the model, and green lines represent misspecifications in the endogenous part of the model. These types of misspecifications result in different realisations of model 2.2 when they are modulated as factors in study 2 but are subsumed under one model here.



Study 3

As in study 1 and 2 data for study 3 was generated based on a 5-factor population structural model with 3 indicators for each factor. Factor loadings and indicator reliability was computed in the same way as in study 1. Two different population models were simulated that resulted in misspecifications of either omitted crossloadings (model 3.1) or omitted correlated residuals (model 3.2). The population-level values of the structural parameters were set to 0.1. Reliability levels were manipulated as in Study 1. The omitted crossloadings (see figure 7) could either be all positive or negative and were set to be 10% lower in absolute values than the factor loadings. Correlated residuals were also either all positive or all negative and were set to not exceed a factor of 0.6 of the residual variances of the indicators.

Figure 3: Model 3.1 and Model 3.2. *Note.* Error terms are not explicitly shown in the figure. Dashed lines represent relations omitted in the estimation model present in the population model. Unspecified crossloadings and correlated residuals could be either positive or negative resulting in 2 modulations of model 3.1 and 3.2 in the study.



Experimental Design of simulation procedures

Study 1

Study 1 varied three main conditions: (1) sample sizes of small ($N = 100$), moderate ($N = 400$), and large ($N = 6400$); (2) Indicator reliability of low ($= 0.3$), moderate (0.5),

high ($= 0.7$); (3) Model specifications: correctly specified model and misspecified with not specified cross loadings in the population model (see figure 2), misspecified with not-specified correlated residuals and a reversed structural path between the the third and the fourth latent factor in the population model (see figure 3) and a recursive structural relation between factor 3 and 4 in the population specified as only one directional (see figure 4).

Study 2

Study 2 varied five conditions: (1) sample sizes: small ($N = 100$), medium ($N = 400$), and large ($N = 6400$). (2) Variance explained by endogenous factors: low ($R^2 = 0.1$) and medium ($R^2 = 0.4$). (3) Indicator reliability: low (0.3), moderate (0.5), and high (0.7). (4) Model misspecifications: varying the population model by omitting a residual covariance and a factor cross-loading in different parts of the model. (5) Number of measurement blocks: separate measurement model per latent variable ($b = 5$) and joint measurement model for all exogenous variables ($b = 3$) for the local SAM condition (lSAM-ML).

Study 3

Study 3 varied three conditions: (1) sample sizes of very small ($N = 50$), small ($N = 100$) or moderate ($N = 400$). (2) Indicator reliability of low ($= 0.3$), moderate (0.5) or high ($= 0.7$); (3) Model misspecifications with not-specified cross loadings in the population model that were positive or negative (see figure) or not-specified correlated residuals in the population model that were postive or negative (see figure 8).

Method Selection

All studies compared the performance of four estimation methods: Vanilla SEM with maximum likelihood (ML), Global SAM with maximum likelihood (gSAM), Local SAM with maximum likelihood (lSAM-ML), Local SAM with unweighted least squares (lSAM-ULS).

Performance Measures

For all studies convergence rates and number of improper solutions (converged models that showed negative variances) were tracked. Further estimated strucutral regression

parameters were evaluated regarding, bias, coverage levels of 95% confidence intervals (CIs), and root mean squared errors (RMSE).

Software

All analyses were conducted in R Core Team (2023). Simulation and estimation was done using Rosseel (2012). See <https://github.com/valentinkm/AdversarialSimulation> for more details.

Analysis and Interpretation plan

The results were interpreted by descriptively comparing the performance measures (bias, MSE, convergence rates) of the different estimation methods under varying sample sizes, indicator reliability levels, and model misspecifications. For study 1 and 3 in which all population regressors were set to be equal, we considered the absolute bias. For the varying correctly specified regressors in study 2 we considered the relative bias. Performance of parameters specified in the estimation model not present in the population model was evaluated in terms of absolute bias as well but excluded from aggregation across all parameters in one model.

- **sample mean:**

$$\bar{T} = \frac{1}{K} \sum_{k=1}^K T_k$$

- **Sample variance:**

$$S_T^2 = \frac{1}{K-1} \sum_{k=1}^K (T_k - \bar{T})^2$$

- **Sample skewness (standardized):**

$$g_T = \frac{1}{K S_T^3} \sum_{k=1}^K (T_k - \bar{T})^3$$

- **Sample kurtosis (standardized):**

$$k_T = \frac{1}{K S_T^4} \sum_{k=1}^K (T_k - \bar{T})^4$$

- **Bias $\bar{T} - \theta$**

- **RMSE $\sqrt{\frac{1}{K} \sum_{k=1}^K (T_k - \theta)^2}$**

- **MCSE of Bias: $\sqrt{\frac{S_T^2}{K}}$**

- **MCSE of RMSE: $\sqrt{\frac{K-1}{K} \sum_{j=1}^K (\text{RMSE}_{(j)} - \text{RMSE})^2}$**

- **Relative Bias** $\frac{\bar{T}-\theta}{\theta}$
- **** MCSE for relative Bias**** $\sqrt{\frac{S_T^2}{K\theta^2}}$
- **Relative RMSE** $\sqrt{\frac{(\bar{T}-\theta)^2+S_T^2}{\theta^2}}$
- **MCSE for relative RMSE** $\sqrt{\frac{K-1}{K} \sum_{j=1}^K \left(rRMSE_{(j)} - rRMSE \right)^2}$

Dhaene, Sara, and Yves Rosseel. 2023. “An Evaluation of Non-Iterative Estimators in the Structural After Measurement (SAM) Approach to Structural Equation Modeling (SEM).” *Structural Equation Modeling: A Multidisciplinary Journal* 30 (6): 926–40. <https://doi.org/10.1080/10705511.2023.2220135>.

R Core Team. 2023. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Rosseel, Yves. 2012. “Lavaan: An R Package for Structural Equation Modeling.” *Journal of Statistical Software* 48 (May): 1–36. <https://doi.org/10.18637/jss.v048.i02>.

Rosseel, Yves, and Wen Wei Loh. 2022. “A Structural After Measurement Approach to Structural Equation Modeling.” *Psychological Methods*, No Pagination Specified—. <https://doi.org/10.1037/met0000503>.