Adversarial Simulation

A case study

Leonard Kosanke

The git hash is: a9e23

Deviations from the simulation protocol

Across studies

For pragmatic purposes of the simulation, we were unable to track and include non-convergence explicitly and exclude cases where it occured.

Studies 1,2,3

For all these studies, 3 deviations are present. Firstly, the total number of conditions turned out to be higher, as the calculation in the protocol forgot the addition of N=50 to be added to the factor sample size. Secondly, It was not detailed what type of LSAM estimation will be implemented. For completeness sake, we included ML- as well as ULS-LSAM estimation. Lastly, instead of just calculating the relative bias as originally planned, we also calculated the absolute bias to mitigate biases of opposite directions canceling each other out, as rightly pointed out in study 4 of Robitzsch (2022).

Study 1b

In Study 1b, contrary to the simulation protocol, we added another sample size with N=50 to get a more nuanced understanding of the small-sample bias of LSAM. For the same reasons as in studies 1,2 and 3, we used ML- as well as ULS-LSAM estimation. Calculation of the relative bias was not feasible in this study, as for the conditions where phi is 0, dividing by 0 would lead to infinite values.

Study 4

Again, I forgot the addition of N=50 in the original condition calculation. Importantly, the model parameters and resulting population-level covariance matrices for the different DGM's provided by Robitzsch (2022) in their github repository were not the same as in the first paper by Rosseel and Loh (2022): They omitted, for example, the cross-loading of the last factor in DGM2, without an apparent reason. Additionally, they had different values for the residual variances without giving insight into why they chose to do so and how (there was no calculation in the simulation script provided). For these reasons, and because of the fact that the calculation of the values in Rosseel and Loh (2022) was replicable, we chose to use the values of Rosseel and Loh (2022) by generating them with the same functions they used. In consequence, a fourth cross-loading was present in DGM 2. The only difference in our data-simulation is that we only used the 3 out of 4 misspecification conditions that were present in Rosseel and Loh (2022). Importantly, due to the implementation of the functions in the paper by Rosseel and Loh (2022), the naming of the DGM's changed. In their paper, what Robitzsch (2022) called DGM1 is implemented with a value of 0. DGM 2 and 3 are thus also implemented with values 1 and 2. To avoid confusion, I will use the terminology of the code from Rosseel and Loh (2022). Thus, the naming is different when interpreting the results later on.

Study 4a

I decided to include a variation of study 4a after all, as the relation between beta/phi values and N is interesting: The original paper showed that results differed between estimators, depending on the size of the factor correlation in Studies 5, 6 and 1b, as well as differential effects due to N in Studies 1,2 and 3. We deemed it interesting to investigate these results in a model with a different number of factors. These findings, combined with our goal to investigate realistic data conditions, rather than just the population level, left us to include Study 4a, with the addition of an additional factor sample size. As DGM1 neither contained misspecification (which is central to our research question), nor did it lead to interesting results in the original study 4, it was omitted. Similarly, as no substantial conclusions were drawn with regards to RMSE in the original papers study 1 and 2, we omitted its calculation as well. Lastly, for the same reasons as argued for Study 4, we again created the DGM's based on the functions used in Rosseel and Loh (2022). As in Study 4, from now on we will also adhere to the naming that follows from the code taken from Rosseel and Loh (2022), so that DGM's 1-3 are now DGM's 0-2.

Studies 5,6

As we are not interested in the comparative performance of SAM vs. SEM at the population level for our research question, but in the performance comparison under realistic conditions, especially in small to moderate sample sizes, we decided to not replicate Studies 5 and 6 after all. We deemed this a fair deviation from our simulation protocol, especially because no substantial performance differences arose in the results of these studies, and the most relevant aspect for us, the differential effect due to the size of factor correlation, is already covered in two different models in studies 1b and 4a.

Result analysis

Original paper

Bias

In the original paper, Robitzsch (2022) had some inconsistencies when it came to the interpretation of the results of their simulation studies. They did not thoroughly define an a prior criterium for when an estimator outperformed another. Only in their Studies 5 and 6, they defined an absolute bias equal to or larger than 0.05 as relevant. Additionally, when an estimator had a minimum percentage error of 20% higher than another, they defined it as being better than the other, in these studies. Across the other studies, interpretations sometimes varied. In some studies (e.g. Study 1), they seemed to orient on the same value of 0.05, but this time as an absolute difference value, without the addition of a certain percentage cut-off. In other studies (e.g. Study 2), they seemed to deviate from this value to define relevant difference in absolute bias. Importantly, this deviation was present in both directions, and therefore did not appear to be consciously biased towards one direction of possible outcomes. In consequence, it appears that interpretations aren't coherent across the studies.

Another important aspect to highlight in the original papers results refers to the general presence of a negative small sample bias in LSAM estimation. This bias makes LSAM appear to perform superior in conditions with positive values in unmodelled cross-loadings and residual correlation, but worse when the values are negative. Even though this bias is correctly identified and explained already in Study 1, Robitzsch (2022) conduct multiple follow-up studies only investigating positive values in unmodelled cross-loadings and residual correlation, namely in study 3. This is done without a general note of some sorts, that this bias should apply in these data-generating scenarios the same as in Studies 1 and 2. In consequence, the interpretations Robitzsch (2022) make, are somewhat misleading on first sight, if one does not account for this bias. This does not, however, explain why they did not include more studies or conditions with negatively valenced residual correlations or cross-loadings. We hypothesize, that the goal of the paper was to show that even in these favorable constellations, SAM does not perform better than SEM.

To be fair, it should be stated that most of the final conclusions drawn in the paper are correct, anyway. By saying that SAM- should not "generally" be preferred over SEM-estimation, they are right in that there are many data constellations, where SAM does not outperform traditional SEM. They also acknowledge, that based on their results for non-saturated structural models, SAM seems to be the better choice. Unfortunately, they use the term "generally" in a second context, when they argue in the discussion that still, SAM should still generally be used. Here, however, they say it should be used to avoid confounding the definition of measurement models from the estimation of structural models. This, according to Robitzsch (2022), often happens in applied research by including cross-loadings and residual correlations at will, without substantive reasons. In this sense, SAM should generally be used.

For our studies, in order to have a more coherent interpretation of results, we will define an estimator to outperform another, based on the cut-off of 0.05 for bias. Additionally, we will only use the term "generally" to refer to the first meaning, that it outperforms across multiple data generating scenarios or conditions.

SD and RMSE

Even though Robitzsch (2022) report tables with results for SD and RMSE, they rarely explicitly mentioned them in the results sections or give cut-offs for substantial relevance.

SD is only reported and mentioned in the conditions without misspecification in Study 1, even though also claimed to be calculated in studies 2 and 3. In this study, they mentioned a 6.6% reduction of SD for LSAM in comparison to ML. This calculation, however, is wrong as the reduction amounts to 8%. For our interpretation, we will report percentage reductions larger than 5% as well and compare them with the original study.

RMSE (and average RMSE) is reported in Studies 1-4, but only in studies 2 and 4 explicitly interpreted for our estimators of interest. In study 2, differences of up to 0.03 are present and interpreted as not being substantial differences. In study 4, differences of 0.01 are present and interpreted as slight efficiency gains. We infer that this lack of mentioning and consistency can be interpreted as there being no differences of interest. Thus, we expect our studies to yield similar results. If, however, there are differences large than 0.03, we will report them.

Preliminary information for results

Recalculation of CI's

We did multiple mistakes with the calculation of the confidence intervals, as we did not include the absolute, but relative mean in the two limits for studies 1b, 4 and 4a. Also, the standard error was wrongly calculated as relative. For Studies 1, 2 and 3, we misscalculated the standard error, as well as the confidence interval by including absolute values in their calculation, even though we were interested in relative values in these studies. Thus, we have to reextract the results for Bias correctly, which was done in postprocess_results.R. All we did there, was reapply the corrected extract_results() function and the same report_bias() function from the individual simulations, one after another. For the results of bias, we load these processed files, instead of the original ones, from the newly created folder: SimulationResultsProcessed.

Results Simulation 1: 2-factor-CFA with 0,1 or 2 correlated residuals

Error, warnings and messages

First, lets check for any messages, warnings or errors:

```
#Save all errors, warnings and messages
all_messages_s1 <- readRDS("SimulationResults/sim1_results_error.rds")


#errors
```

```r
errors_s1 <- all_messages_s1$errors
errors_sum_s1 <- unlist(errors_s1)
length(errors_sum_s1)
```

```
[1] 0
```

```r
#warnings
warnings_s1 <- all_messages_s1$warnings
warnings_sum_s1 <- unlist(warnings_s1)
length(warnings_sum_s1)
```

```
[1] 245
```

```r
#messages
messages_s1 <- all_messages_s1$messages
messages_sum_s1 <- unlist(messages_s1)
length(messages_sum_s1)
```

```
[1] 0
```

No errors and messages were present. There were, however, 245 warnings, we will investigate in detail.

Warnings investigation

Investigating the 'warnings' object list, it became apparent that all the warnings are the same: "no non-missing arguments to min; returning Inf".

```r
unique(warnings_s1)
```

```
[[1]]
character(0)

[[2]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"

[[3]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
```

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	-0.04 [-0.19-0.10]	-0.01 [-0.08-0.06]	0.00 [-0.02-0.03]	-0.00 [-0.02-0.01]	0.00 [-0.00-0.01]	-0.00 [-0.00-0.00]	-0.00 [-0.00–0.00]
SEM_ULS	0.02 [-0.11-0.15]	0.02 [-0.04-0.09]	0.01 [-0.01-0.04]	0.00 [-0.01-0.01]	0.01 [-0.00-0.01]	-0.00 [-0.00-0.00]	-0.00 [-0.00-0.00]
LSAM_ML	-0.39 [-0.50–0.28]	-0.27 [-0.33–0.21]	-0.11 [-0.14–0.08]	-0.06 [-0.07–0.04]	-0.02 [-0.03–0.02]	-0.01 [-0.01–0.01]	-0.00 [-0.00–0.00]
LSAM_ULS	-0.39 [-0.50–0.28]	-0.27 [-0.33–0.21]	-0.11 [-0.14–0.08]	-0.06 [-0.07–0.04]	-0.02 [-0.03–0.02]	-0.01 [-0.01–0.01]	-0.00 [-0.00–0.00]
GSAM_ML	-0.39 [-0.51–0.28]	-0.27 [-0.33–0.21]	-0.11 [-0.14–0.08]	-0.06 [-0.07–0.04]	-0.02 [-0.03–0.02]	-0.01 [-0.01–0.01]	-0.00 [-0.00–0.00]
GSAM_ULS	-0.39 [-0.51–0.28]	-0.27 [-0.33–0.21]	-0.11 [-0.14–0.08]	-0.06 [-0.07–0.04]	-0.02 [-0.03–0.02]	-0.01 [-0.01–0.01]	-0.00 [-0.00–0.00]

Note:

Condition: 0_0.12

```
[3] "no non-missing arguments to min; returning Inf"
[4] "no non-missing arguments to min; returning Inf"

[[4]]
[1] "no non-missing arguments to min; returning Inf"

[[5]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"
[4] "no non-missing arguments to min; returning Inf"
[5] "no non-missing arguments to min; returning Inf"
[6] "no non-missing arguments to min; returning Inf"
```

This warning could indicate either convergence issues or improper solutions due to small sample size in one or multiple estimators. Unfortunately, due to the implementation of the study, we are unable to check for more details, as whether there is any kind of pattern with regards to the warning occurrence. There is, however, evidence that this occurs frequently in standard SEM estimation and was one claimed advantage of SAM- over SEM-estimation (Rosseel and Loh 2022). Even though in Robitzsch (2022) the authors claim that constraining estimation to realistic values (e.g. loadings between 0 and 1) should solve this, applying these same constraints still lead to these issues in our study. The combined study should check this more properly to evaluate these claims thoroughly. As the 245 are less than 0.5% of the 525000 estimations performed, we will continue with the interpretation as planned, as their impact is negligible.

Bias

Condition without correlated residuals (0_0.12)

Both standard SEM estimators are unbiased (i.e. relative Bias < 0.05) across all N conditions. The 4 SAM estimators are negatively biased for sample sizes up to 500 and have nearly identical estimation results. Across all estimators, the relative Bias decreases with increasing N.

Thus, SEM estimation outperforms SAM in small to moderate sample sizes, when there is no misspecification.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	-0.12 [-0.27-0.03]	-0.09 [-0.16–0.02]	-0.07 [-0.10–0.05]	-0.08 [-0.10–0.07]	-0.07 [-0.08–0.07]	-0.08 [-0.08–0.08]	-0.08 [-0.08–0.08]
SEM_ULS	-0.06 [-0.20-0.09]	-0.06 [-0.12-0.01]	-0.07 [-0.09–0.04]	-0.08 [-0.09–0.07]	-0.07 [-0.08–0.07]	-0.08 [-0.08–0.08]	-0.08 [-0.08–0.08]
LSAM_ML	-0.45 [-0.56–0.34]	-0.33 [-0.39–0.27]	-0.19 [-0.21–0.16]	-0.14 [-0.16–0.13]	-0.11 [-0.12–0.10]	-0.10 [-0.10–0.10]	-0.09 [-0.09–0.09]
LSAM_ULS	-0.45 [-0.56–0.34]	-0.33 [-0.39–0.27]	-0.19 [-0.21–0.16]	-0.14 [-0.16–0.13]	-0.11 [-0.12–0.10]	-0.10 [-0.10–0.10]	-0.09 [-0.09–0.09]
GSAM_ML	-0.45 [-0.56–0.34]	-0.33 [-0.39–0.27]	-0.19 [-0.21–0.16]	-0.14 [-0.16–0.13]	-0.11 [-0.12–0.10]	-0.10 [-0.10–0.10]	-0.09 [-0.09–0.09]
GSAM_ULS	-0.45 [-0.56–0.34]	-0.33 [-0.39–0.26]	-0.19 [-0.21–0.16]	-0.14 [-0.16–0.13]	-0.11 [-0.12–0.10]	-0.10 [-0.10–0.10]	-0.09 [-0.09–0.09]

Note:

Condition: 1_-0.12

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.04 [-0.10-0.18]	0.09 [0.02-0.16]	0.11 [0.08-0.13]	0.10 [0.09-0.12]	0.11 [0.11-0.12]	0.11 [0.10-0.11]	0.11 [0.11-0.11]
SEM_ULS	0.10 [-0.02-0.23]	0.12 [0.06-0.18]	0.11 [0.09-0.14]	0.10 [0.09-0.11]	0.10 [0.10-0.11]	0.10 [0.10-0.10]	0.10 [0.10-0.10]
LSAM_ML	-0.34 [-0.46–0.23]	-0.21 [-0.28–0.15]	-0.03 [-0.05–0.00]	0.03 [0.02-0.04]	0.07 [0.06-0.07]	0.08 [0.07-0.08]	0.09 [0.09-0.09]
LSAM_ULS	-0.34 [-0.46–0.23]	-0.21 [-0.28–0.15]	-0.03 [-0.05–0.00]	0.03 [0.02-0.04]	0.07 [0.06-0.07]	0.08 [0.07-0.08]	0.09 [0.09-0.09]
GSAM_ML	-0.34 [-0.46–0.23]	-0.21 [-0.28–0.15]	-0.03 [-0.05–0.00]	0.03 [0.02-0.04]	0.07 [0.06-0.07]	0.08 [0.07-0.08]	0.09 [0.09-0.09]
GSAM_ULS	-0.34 [-0.46–0.23]	-0.21 [-0.28–0.15]	-0.03 [-0.05–0.00]	0.03 [0.02-0.04]	0.07 [0.06-0.07]	0.08 [0.07-0.08]	0.09 [0.09-0.09]

Note:

Condition: 1_0.12

Condition with one negative residual correlation (1_-0.12)

All estimators are negatively biased (i.e. relative Bias >= -0.05) across all N conditions. Across all estimators except SEM-ULS, the Bias decreases substantially with increasing N. For SEM-ULS, it increases slightly but unsubstantially up to N=500, but is lowest overall. The 4 SAM estimators have nearly identical estimation results. In relative comparison, the two SEM estimators outperform all SAM estimators in every condition of N, even though this difference is not substantive for N=1000 and higher.

Thus, SEM estimation outperforms SAM in small to moderate sample sizes, when there is one negative residual.

Condition with one positive residual correlation (1_0.12)

Besides SEM-ML in N=50, both SEM estimators are positively biased across all N, with no visible trend for increasing N. All SAM estimators have nearly identical values and increasing relative bias for increasing N. They are negatively biased up to N=250 and positively biased for the other N conditions. For N=250 and N=500, their relative bias is unsubstantial. In relative comparison, the two SEM estimators outperform all SAM estimators for sample sizes up to 100. For all higher N, the SAM estimators outperform the SEM estimators, even though the difference is unsubstantial starting from N=1000.

Thus, SEM estimation outperforms SAM in smaller samples (N=50-100), whereas SAM outperforms SEM in moderate samples (N250-500), and slightly in higher sample sizes.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	-0.20 [-0.36–0.05]	-0.18 [-0.25–0.10]	-0.17 [-0.19–0.14]	-0.17 [-0.18–0.16]	-0.16 [-0.17–0.16]	-0.17 [-0.17–0.16]	-0.16 [-0.16–0.16]
SEM_ULS	-0.14 [-0.29-0.01]	-0.14 [-0.21–0.08]	-0.16 [-0.18–0.13]	-0.17 [-0.18–0.15]	-0.16 [-0.17–0.16]	-0.17 [-0.17–0.17]	-0.17 [-0.17–0.17]
LSAM_ML	-0.50 [-0.61–0.39]	-0.39 [-0.45–0.32]	-0.27 [-0.30–0.25]	-0.23 [-0.24–0.21]	-0.20 [-0.20–0.19]	-0.19 [-0.19–0.19]	-0.18 [-0.18–0.18]
LSAM_ULS	-0.50 [-0.61–0.39]	-0.39 [-0.45–0.32]	-0.27 [-0.30–0.25]	-0.23 [-0.24–0.21]	-0.20 [-0.20–0.19]	-0.19 [-0.19–0.19]	-0.18 [-0.18–0.18]
GSAM_ML	-0.50 [-0.61–0.39]	-0.39 [-0.45–0.32]	-0.27 [-0.30–0.25]	-0.23 [-0.24–0.21]	-0.20 [-0.20–0.19]	-0.19 [-0.19–0.19]	-0.18 [-0.18–0.18]
GSAM_ULS	-0.50 [-0.61–0.38]	-0.39 [-0.45–0.32]	-0.27 [-0.30–0.25]	-0.23 [-0.24–0.21]	-0.20 [-0.20–0.19]	-0.19 [-0.19–0.19]	-0.18 [-0.18–0.18]

Note:
Condition: 2_-0.12

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.11 [-0.03-0.24]	0.17 [0.11-0.23]	0.20 [0.17-0.22]	0.19 [0.18-0.20]	0.20 [0.19-0.20]	0.19 [0.19-0.20]	0.20 [0.20-0.20]
SEM_ULS	0.18 [0.06-0.30]	0.19 [0.14-0.25]	0.20 [0.18-0.22]	0.19 [0.18-0.20]	0.19 [0.19-0.20]	0.19 [0.19-0.19]	0.19 [0.19-0.19]
LSAM_ML	-0.30 [-0.41–0.19]	-0.13 [-0.20–0.07]	0.05 [0.03-0.08]	0.11 [0.10-0.12]	0.15 [0.15-0.16]	0.17 [0.16-0.17]	0.18 [0.18-0.18]
LSAM_ULS	-0.30 [-0.41–0.18]	-0.13 [-0.20–0.07]	0.05 [0.03-0.08]	0.11 [0.10-0.12]	0.15 [0.15-0.16]	0.17 [0.16-0.17]	0.18 [0.18-0.18]
GSAM_ML	-0.30 [-0.41–0.18]	-0.13 [-0.20–0.07]	0.05 [0.03-0.08]	0.11 [0.10-0.12]	0.15 [0.15-0.16]	0.17 [0.16-0.17]	0.18 [0.18-0.18]
GSAM_ULS	-0.30 [-0.41–0.18]	-0.13 [-0.20–0.07]	0.05 [0.03-0.08]	0.11 [0.10-0.12]	0.15 [0.15-0.16]	0.17 [0.16-0.17]	0.18 [0.18-0.18]

Note:
Condition: 2_0.12

Condition with two negative residual correlations (2_-0.12)

All estimators are negatively biased across all N conditions. Across all estimators except SEM-ULS, the Bias decreases substantially with increasing N. For SEM-ULS, it slightly, but unsubstantially increases up to N=500, but is lowest overall in these conditions. The 4 SAM estimators have nearly identical estimation results. In relative comparison, the two SEM estimators outperform all SAM estimators in every condition of N, even though this difference is not substantive for N=1000 and higher.

Thus, SEM estimation outperforms SAM in small to moderate sample sizes, when there are two unmodelled negative residual correlations.

Condition with two positive residual correlation (2_0.12)

Both SEM estimators are positively biased across all N, with no visible trend for increasing N for ULS, but a substantial increase for ML. All SAM estimators have nearly identical values and increasing relative bias for increasing N. They are negatively biased up to N=100 and positively biased for the other N conditions. In relative comparison, the two SEM estimators outperform all SAM estimators for N=50. For all higher N, the SAM estimators outperform the SEM estimators (substantially, besides ML in N=100), even though the difference is unsubstantial starting from N=1000.

Thus, SEM estimation outperforms SAM in small samples (N=50), whereas SAM outperforms SEM in small to moderate samples (N100-500), and slightly in higher sample sizes.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.2994117	0.2080360	0.1218107	0.0831350	0.0596527	0.0373128	0.0056354
SEM_ULS_sd	0.2762314	0.1938677	0.1179754	0.0819869	0.0591705	0.0371478	0.0056359
LSAM_ML_sd	0.2324899	0.1900756	0.1232512	0.0840204	0.0593394	0.0373247	0.0056366
LSAM_ULS_sd	0.2332835	0.1900221	0.1232517	0.0840205	0.0593394	0.0373247	0.0056366
GSAM_ML_sd	0.2347972	0.1901166	0.1232512	0.0840204	0.0593394	0.0373247	0.0056366
GSAM_ULS_sd	0.2373658	0.1900492	0.1232502	0.0840203	0.0593394	0.0373247	0.0056366

Note:
Condition: 0_0.12

Summary

in 0.5% of estimations, a warning referring to potential problems with convergence or improper solutions due to small sample sizes in one or multiple estimators occurred. This should be investigated more explicitly in the joint study.

In conditions with no, one or two negative residual correlations, SEM outperforms SAM in small to moderate sample sizes. This seems to be independent of the amount of misspecification (1 vs. 2).

Comparing the conditions with one and two positive residual correlations, the results suggest that SAM outperforms SEM in smaller to moderate samples, with increasing amount of misspecifiation (even better for 2 than for 1 residual correlation). In small samples, SEM tends to perform better.

Comparison with original paper

Comparing these results with the original paper, the results are similar and can be seen as mostly replicated. The only difference lies in the potential convergence issues in a small number of estimations. Besides that, we draw the same conclusions as Robitzsch (2022): All SAM approaches appear to be generally negatively biased for small to moderate samples in the 2-factor-CFA model. This leads to the comparatively worse performance in the conditions with no, one or two negative residual correlations. Additionally, this negative bias corrects for unmodelled positive residual correlations and makes SAM appear superior in these conditions.

SD

No residual correlations (0_0.12)

Across all estimators, the standard deviation decreases with increasing N.

There are only two substantial differences: Substantial percentage reduction of SD up to 22.4% when comparing SEM-ML and LSAM-ML in N=50: 0.299x = *0.232 0.232/0.299 = 0.775 1-(0.232/0.299) =* 0.224*

Substantial percentage reduction of SD up to 8.6% when comparing SEM-ML and LSAM-ML in N=100:

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.3210532	0.2127884	0.1245407	0.0848174	0.0609095	0.0384101	0.0058608
SEM_ULS_sd	0.3028143	0.1979163	0.1204037	0.0832098	0.0600621	0.0379880	0.0058037
LSAM_ML_sd	0.2327865	0.1871164	0.1229734	0.0862784	0.0611392	0.0388350	0.0059686
LSAM_ULS_sd	0.2336178	0.1871355	0.1229738	0.0862784	0.0611392	0.0388350	0.0059686
GSAM_ML_sd	0.2347754	0.1871332	0.1229734	0.0862784	0.0611392	0.0388350	0.0059686
GSAM_ULS_sd	0.2356162	0.1871443	0.1229724	0.0862784	0.0611392	0.0388350	0.0059686

Note:

Condition: 1_-0.12

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.2894206	0.1969607	0.1202117	0.0826682	0.0583952	0.0365446	0.0055962
SEM_ULS_sd	0.2682900	0.1805294	0.1161134	0.0807541	0.0575612	0.0362207	0.0055498
LSAM_ML_sd	0.2400034	0.1960049	0.1258344	0.0830220	0.0603064	0.0374891	0.0056808
LSAM_ULS_sd	0.2398938	0.1961046	0.1258349	0.0830220	0.0603064	0.0374891	0.0056808
GSAM_ML_sd	0.2425296	0.1960305	0.1258344	0.0830220	0.0603064	0.0374891	0.0056808
GSAM_ULS_sd	0.2424237	0.1961125	0.1258333	0.0830220	0.0603064	0.0374891	0.0056808

Note:

Condition: 1_0.12

Thus, SAM outperforms SEM in terms of SD reduction in smaller samples

One negative residual correlation (1_-0.12)

Across all estimators, the standard deviation decreases with increasing N.

There are only two substantial differences: Substantial percentage reduction of SD up to 27.4% when comparing SEM-ML and LSAM-ML in N=50:

Substantial percentage reduction of SD up to 12.2% when comparing SEM-ML and LSAM-ML in N=100:

Thus, SAM outperforms SEM in terms of SD reduction in smaller samples.

One positive residual correlation (1_0.12)

Across all estimators, the standard deviation decreases with increasing N.

The SAM estimators outperform both SEM estimators in N=50 with a percentage reduction of SD up to 16.9%. ULS-SEM outperforms all SAM estimators with a percentage reduction of SD up to 8.16% in N=100-250. No substantial differences between the two SEM and all SAM estimators starting from moderate sample sizes (N=500-2500).

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.3273680	0.2186402	0.1248222	0.0873407	0.0619126	0.0383017	0.0059566
SEM_ULS_sd	0.3158052	0.2041129	0.1198360	0.0849046	0.0606796	0.0378193	0.0058653
LSAM_ML_sd	0.2318215	0.1905338	0.1230386	0.0887802	0.0622210	0.0387604	0.0060471
LSAM_ULS_sd	0.2327260	0.1906172	0.1230391	0.0887802	0.0622210	0.0387604	0.0060471
GSAM_ML_sd	0.2336877	0.1906955	0.1230386	0.0887802	0.0622210	0.0387604	0.0060471
GSAM_ULS_sd	0.2346232	0.1907682	0.1230381	0.0887802	0.0622210	0.0387604	0.0060471

Note:

Condition: 2_-0.12

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.2820486	0.1871067	0.1180181	0.0808494	0.0571806	0.0358058	0.0054829
SEM_ULS_sd	0.2540941	0.1757545	0.1138675	0.0793859	0.0565420	0.0354570	0.0054503
LSAM_ML_sd	0.2388948	0.1979236	0.1264714	0.0824969	0.0589894	0.0370433	0.0055570
LSAM_ULS_sd	0.2391389	0.1979099	0.1264721	0.0824969	0.0589894	0.0370433	0.0055570
GSAM_ML_sd	0.2402856	0.1979794	0.1264714	0.0824969	0.0589894	0.0370433	0.0055570
GSAM_ULS_sd	0.2405700	0.1979344	0.1264698	0.0824967	0.0589894	0.0370433	0.0055570

Note:

Condition: 2_0.12

Thus, SAM outperforms SEM in terms of SD reduction in small samples, whereas SEM outperforms SAM in moderate samples.

Two negative residual correlations (2_-0.12)

Across all estimators, the standard deviation decreases with increasing N.

The SAM estimators outperform both SEM estimators in N=50 with a percentage reduction of SD up to 29% and in N=100 with up to 12.8%. For higher N, there are no substantial differences.

Thus, SAM outperforms SEM in terms of SD reduction in smaller samples.

Two positive residual correlations (2_0.12)

Across all estimators, the standard deviation decreases with increasing N.

The SAM estimators outperform both SEM estimators in N=50 with a percentage reduction of SD up to 15.2%. Both SEM estimators ourperform the SAM estimators in N=100 with a percentage reduction of SD up to 11.1% and in N=250 with up to 10.3%. For higher N, there are no substantial differences.

Thus, SAM outperforms SEM in terms of SD reduction in small samples, but SEM outperforms SAM in smaller to moderate samples.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.3005163	0.2080780	0.1217712	0.0831305	0.0596670	0.0373185	0.0056336
SEM_ULS_rmse	0.2765139	0.1942412	0.1181585	0.0819662	0.0592561	0.0371396	0.0056341
LSAM_ML_rmse	0.3314332	0.2497986	0.1399163	0.0904993	0.0607282	0.0379424	0.0056375
LSAM_ULS_rmse	0.3314782	0.2496858	0.1399168	0.0904994	0.0607282	0.0379424	0.0056375
GSAM_ML_rmse	0.3332450	0.2498205	0.1399162	0.0904993	0.0607281	0.0379423	0.0056375
GSAM_ULS_rmse	0.3344564	0.2496787	0.1399151	0.0904992	0.0607281	0.0379423	0.0056375

Note:
Condition: 0_0.12

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.3289817	0.2195613	0.1322372	0.0983096	0.0753561	0.0610843	0.0465557
SEM_ULS_rmse	0.3045969	0.2009511	0.1269533	0.0959100	0.0747192	0.0617589	0.0484429
LSAM_ML_rmse	0.3570748	0.2711957	0.1664049	0.1220156	0.0897678	0.0716146	0.0542805
LSAM_ULS_rmse	0.3569695	0.2710438	0.1664053	0.1220156	0.0897678	0.0716146	0.0542805
GSAM_ML_rmse	0.3582141	0.2712039	0.1664049	0.1220156	0.0897678	0.0716146	0.0542805
GSAM_ULS_rmse	0.3580409	0.2710251	0.1664038	0.1220156	0.0897678	0.0716146	0.0542805

Note:
Condition: 1_-0.12

Summary and comparison with original paper

Overall, SAM outperforms SEM in terms of SD reduction in small to smaller samples, especially for negatively correlated residuals. In smaller to moderate samples, SEM outperforms SAM only for positively correlated residuals. For conditions with no misspecification, which were the only conditions reported in Robitzsch (2022), the percentage reduction favoring SAM is nearly identical for N=100, and the reduction seems to be even more significantly in favor of SEM in the smaller sample size.

RMSE

No residual correlations (0_0.12)

Across all estimators, the RMSE decreases with increasing N. In N=100, both SEM estimators were more efficient than all SAM estimators. Besides, there are no substantial differences present.

One negative residual correlation (1_-0.12)

Across all estimators, the RMSE decreases with increasing N. For small to moderate samples (N=50-250), both SEM estimators were more efficient than all SAM estimators. For higher N, there are no substantial differences present.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.2903666	0.2041826	0.1367211	0.1037104	0.0886955	0.0735649	0.065652
SEM_ULS_rmse	0.2752874	0.1937697	0.1338199	0.1006104	0.0851725	0.0691692	0.060096
LSAM_ML_rmse	0.3161203	0.2333349	0.1268449	0.0851536	0.0722432	0.0594033	0.053989
LSAM_ULS_rmse	0.3155952	0.2334030	0.1268454	0.0851536	0.0722432	0.0594033	0.053989
GSAM_ML_rmse	0.3180434	0.2333511	0.1268449	0.0851536	0.0722432	0.0594033	0.053989
GSAM_ULS_rmse	0.3174417	0.2333899	0.1268438	0.0851536	0.0722432	0.0594033	0.053989

Note:
Condition: 1_0.12

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.3495919	0.2425600	0.1594906	0.1335079	0.1149357	0.1068168	0.0987182
SEM_ULS_rmse	0.3264740	0.2217173	0.1529855	0.1312674	0.1149851	0.1087381	0.1016315
LSAM_ML_rmse	0.3781482	0.2995712	0.2042734	0.1616319	0.1332475	0.1200147	0.1079898
LSAM_ULS_rmse	0.3780533	0.2995780	0.2042739	0.1616319	0.1332475	0.1200147	0.1079898
GSAM_ML_rmse	0.3789630	0.2996346	0.2042734	0.1616319	0.1332475	0.1200147	0.1079898
GSAM_ULS_rmse	0.3788108	0.2996180	0.2042728	0.1616319	0.1332475	0.1200147	0.1079898

Note:
Condition: 2_-0.12

One positive residual correlation (1_0.12)

Across all estimators, the RMSE decreases with increasing N. For smaller samples (N=50-100), SEM-ULS was more efficient than all SAM estimators. For higher N, there are no substantial differences present.

Two negative residual correlations (2_-0.12)

Across all estimators, the RMSE decreases with increasing N. For small to moderate samples (N=50-500), The two SEM estimators were more efficient than all SAM estimators. For higher N, there are no substantial differences present.

Two positive residual correlations (2_0.12)

Across all estimators, the RMSE decreases with increasing N. For moderate samples (N=250-500), all SAM estimators were more efficient than the two SEM estimators. For higher and lower N, there are no substantial differences present.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.2889074	0.2136481	0.1677122	0.1402285	0.1316671	0.1210553	0.1172049
SEM_ULS_rmse	0.2763714	0.2106218	0.1659391	0.1386034	0.1293397	0.1180913	0.1137134
LSAM_ML_rmse	0.2986753	0.2137068	0.1306533	0.1066594	0.1094951	0.1062431	0.1077363
LSAM_ULS_rmse	0.2984241	0.2136190	0.1306539	0.1066594	0.1094951	0.1062430	0.1077363
GSAM_ML_rmse	0.2997323	0.2137495	0.1306532	0.1066594	0.1094951	0.1062431	0.1077363
GSAM_ULS_rmse	0.2994252	0.2136240	0.1306519	0.1066593	0.1094951	0.1062430	0.1077363

Note:
Condition: 2_0.12

Summary

Overall, classic SEM estimation was more efficient than SAM estimation in small to moderate samples. Only in the case of two positive residual correlation, SAM was more efficient than SEM in moderate samples.

Summary Study 1

We gained the following insights with regards to the 2-factor-CFA model: Unlike the original paper by Robitzsch (2022), we experienced some issues that are related to convergence or improper solutions in small samples.

In terms of bias in small to moderate samples, SEM outperforms SAM in conditions with no, one and two negative residual correlations. SAM appears to outperform SEM for positive residual correlations. This, however, is an artifact of its negative small sample bias and not its performance. The results align with the ones obtained by the original paper.

With regards to SD and RMSE, results are incoherent with no clear pattern emerging.

Thus, it cannot be stated that one estimation method is generally performing better than the other, based on the results of Study 1.

Results Simulation 1b: 2-factor-CFA with 2 residual correlations and varying lambda and phi

Error, warnings and messages

First, lets check for any messages, warnings or errors:

```
[1] 0
```

```
[1] 414
```

```
[1] 0
```

No errors and messages were present. There were, however, 414 warnings, we will investigate in detail.

Warnings investigation

Investigating the 'warnings' object list, it became apparent that all the warnings are the same: "no non-missing arguments to min; returning Inf".

```
unique(warnings_s1b)
```

```
[[1]]
character(0)

[[2]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"
[4] "no non-missing arguments to min; returning Inf"

[[3]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
```

As before, this warning could indicate either convergence issues or improper solutions due to small sample size in one or multiple estimators. Again, we are unable to check for more details, but hypothesize that this occurs mainly in the standard SEM estimators. There is, however, evidence that this occurs frequently standard SEM estimation and was one claimed advantage of SAM- over SEM-estimation (Rosseel and Loh 2022). The combined study should check this more properly to evaluate these claims thoroughly. As the 414 estimation are around 0.5% of the 75000 estimations performed, we will continue with the interpretation as planned, as their impact is negligible.

Bias

```
#Load processed results
bias_ci_s1b <- readRDS("SimulationResultsProcessed/sim1b_abs_bias_ci.rds")
```

Study 1b: Absolute Bias of LSAM-ML for N=50					
	Phi values				
Lambda	**phi=0**	**phi=0.2**	**phi=0.4**	**phi=0.6**	**phi=0.8**
0.4	0.202 [0.193, 0.210]	0.202 [0.193, 0.210]	0.258 [0.249, 0.267]	0.346 [0.335, 0.356]	0.444 [0.432, 0.456]
0.5	0.187 [0.179, 0.195]	0.179 [0.172, 0.187]	0.203 [0.196, 0.211]	0.245 [0.236, 0.253]	0.305 [0.294, 0.315]
0.6	0.176 [0.169, 0.182]	0.165 [0.158, 0.171]	0.166 [0.160, 0.172]	0.170 [0.164, 0.177]	0.190 [0.183, 0.197]
0.7	0.164 [0.158, 0.170]	0.150 [0.144, 0.156]	0.139 [0.133, 0.144]	0.122 [0.117, 0.127]	0.116 [0.111, 0.121]
0.8	0.150 [0.144, 0.155]	0.135 [0.130, 0.140]	0.119 [0.115, 0.124]	0.098 [0.094, 0.102]	0.074 [0.071, 0.077]

Study 1b: Absolute Bias of LSAM-ULS for N=50					
	Phi values				
Lambda	**phi=0**	**phi=0.2**	**phi=0.4**	**phi=0.6**	**phi=0.8**
0.4	0.204 [0.195, 0.213]	0.203 [0.194, 0.211]	0.258 [0.249, 0.267]	0.345 [0.335, 0.355]	0.440 [0.428, 0.452]
0.5	0.188 [0.180, 0.196]	0.180 [0.172, 0.187]	0.203 [0.196, 0.211]	0.244 [0.235, 0.252]	0.304 [0.294, 0.315]
0.6	0.176 [0.169, 0.182]	0.165 [0.158, 0.171]	0.166 [0.159, 0.172]	0.170 [0.163, 0.177]	0.190 [0.183, 0.197]
0.7	0.164 [0.158, 0.170]	0.150 [0.144, 0.156]	0.139 [0.133, 0.144]	0.122 [0.117, 0.127]	0.116 [0.111, 0.121]
0.8	0.150 [0.144, 0.155]	0.135 [0.130, 0.140]	0.119 [0.115, 0.124]	0.098 [0.094, 0.102]	0.074 [0.071, 0.077]

Conditions with N=50

Comparing LSAM-ML and LSAM-ULS, the values are nearly identical. Therefore, the interpretation is the same for both estimators.

Across all conditions, both estimators were biased.

For higher values of lambda, the estimators had less absolute Bias. This difference increased for higher phi values.

For lower values of lambda, the absolute bias increased with increasing phi. This trend reversed for higher values of lambda, where the absolute bias decreased with higher phi values.

Conditions with N=100

Again, the values and therefore the interpretation of LSAM-ML and LSAM-ULS are identical.

Across all conditions, both estimators were biased. For higher values of lambda, the estimators had less absolute Bias. This difference increased for higher phi values. For lower values of lambda, the absolute bias increased with increasing phi. This trend reversed for higher values of lambda, where the absolute bias decreased with higher phi values.

Study 1b: Absolute Bias of LSAM-ML for N=100					
	Phi values				
Lambda	**phi=0**	**phi=0.2**	**phi=0.4**	**phi=0.6**	**phi=0.8**
0.4	0.168 [0.161, 0.175]	0.164 [0.158, 0.171]	0.209 [0.202, 0.216]	0.277 [0.268, 0.286]	0.349 [0.338, 0.360]
0.5	0.157 [0.150, 0.163]	0.155 [0.149, 0.161]	0.153 [0.148, 0.159]	0.171 [0.165, 0.177]	0.200 [0.192, 0.208]
0.6	0.140 [0.135, 0.146]	0.135 [0.130, 0.139]	0.124 [0.119, 0.129]	0.116 [0.112, 0.121]	0.110 [0.106, 0.114]
0.7	0.119 [0.115, 0.124]	0.112 [0.108, 0.116]	0.104 [0.100, 0.107]	0.088 [0.085, 0.091]	0.071 [0.068, 0.074]
0.8	0.106 [0.102, 0.110]	0.097 [0.094, 0.101]	0.088 [0.084, 0.091]	0.073 [0.070, 0.076]	0.055 [0.053, 0.057]

Study 1b: Absolute Bias of LSAM-ULS for N=100					
	Phi values				
Lambda	**phi=0**	**phi=0.2**	**phi=0.4**	**phi=0.6**	**phi=0.8**
0.4	0.168 [0.161, 0.175]	0.164 [0.157, 0.171]	0.209 [0.202, 0.216]	0.276 [0.268, 0.285]	0.348 [0.337, 0.359]
0.5	0.157 [0.151, 0.163]	0.155 [0.149, 0.161]	0.153 [0.148, 0.159]	0.171 [0.165, 0.177]	0.200 [0.192, 0.208]
0.6	0.140 [0.135, 0.146]	0.135 [0.130, 0.139]	0.124 [0.119, 0.129]	0.116 [0.112, 0.121]	0.110 [0.106, 0.114]
0.7	0.119 [0.115, 0.124]	0.112 [0.108, 0.116]	0.104 [0.100, 0.107]	0.088 [0.085, 0.091]	0.071 [0.068, 0.074]
0.8	0.106 [0.102, 0.110]	0.097 [0.094, 0.101]	0.088 [0.084, 0.091]	0.073 [0.070, 0.076]	0.055 [0.053, 0.057]

Additionally, there seems to be no differential effect due to N, as the interpretations between the different N-conditions are identical. Naturally, the bias values are larger for N=50, still.

Comparison to original paper

In contrast to the original paper by Robitzsch (2022), we computed absolute bias to avoid infinite values in some conditions. Additionally, computing absolute instead of relative Bias solves the problem of positive and negative biases canceling each other out. Results partly align in terms of absolute bias: For higher values of lambda, LSAM has less absolute Bias. This difference increases for higher phi values. Only for low values of lambda, the bias increases for higher values of phi. Unlike in this study, the original paper did not find a reversal effect for higher lambda values. Another difference is that we, unlike Robitzsch (2022), found estimates to be biased for low phi-values. This might be due to the difference of using absolute bias, thereby avoiding cancellation of positive and negative values. Still, our general conclusion corresponds to the original paper: LSAM has a general negative small sample bias. The bias is larger in magnitude for lower loadings and higher true factor correlations. Importantly, the bias seems to not disappear for low values of phi and lambda, unlike in the original paper.

Results Simulation 2: 2-factor-CFA with 1 or 2 cross-loadings

Error, warnings and messages

First, lets check for any messages, warnings or errors:

```
[1] 0
```

```
[1] 612
```

```
[1] 0
```

No errors and messages were present. There were, however, 612 warnings we will investigate in detail.

Warnings investigation

Investigating the ‘warnings’ object list, it became apparent that all the warnings are the same: “no non-missing arguments to min; returning Inf”.

Sometimes 4-6 warnings per repetition, more than before (up to 2 per repetition)

```
unique(warnings_s2)
```

```
[[1]]
character(0)

[[2]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"

[[3]]
[1] "no non-missing arguments to min; returning Inf"

[[4]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"
[4] "no non-missing arguments to min; returning Inf"

[[5]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"
[4] "no non-missing arguments to min; returning Inf"
[5] "no non-missing arguments to min; returning Inf"
[6] "no non-missing arguments to min; returning Inf"

[[6]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"

[[7]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"
```

Condition:

	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	-0.209 [-0.372–0.046]	-0.157 [-0.234–0.080]	-0.134 [-0.163–0.106]	-0.137 [-0.151–0.124]	-0.130 [-0.137–0.123]	-0.133 [-0.136–0.131]	-0.130 [-0.130–0.130]
SEM_ULS	-0.144 [-0.303-0.015]	-0.123 [-0.195–0.050]	-0.128 [-0.155–0.101]	-0.140 [-0.154–0.127]	-0.136 [-0.143–0.130]	-0.142 [-0.145–0.139]	-0.140 [-0.140–0.140]
LSAM_ML	-0.535 [-0.648–0.422]	-0.406 [-0.471–0.342]	-0.264 [-0.292–0.237]	-0.206 [-0.219–0.193]	-0.173 [-0.180–0.166]	-0.162 [-0.165–0.159]	-0.150 [-0.150–0.150]
LSAM_ULS	-0.534 [-0.646–0.421]	-0.406 [-0.470–0.342]	-0.264 [-0.292–0.237]	-0.206 [-0.219–0.193]	-0.173 [-0.180–0.166]	-0.162 [-0.165–0.159]	-0.150 [-0.150–0.150]
GSAM_ML	-0.535 [-0.649–0.421]	-0.406 [-0.471–0.342]	-0.264 [-0.292–0.237]	-0.206 [-0.219–0.193]	-0.173 [-0.180–0.166]	-0.162 [-0.165–0.159]	-0.150 [-0.150–0.150]
GSAM_ULS	-0.533 [-0.647–0.419]	-0.406 [-0.470–0.342]	-0.264 [-0.292–0.237]	-0.206 [-0.219–0.193]	-0.173 [-0.180–0.166]	-0.162 [-0.165–0.159]	-0.150 [-0.150–0.150]

Note:

Condition: 1_-0.3

Condition:

	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.160 [0.042-0.279]	0.190 [0.131-0.249]	0.205 [0.183-0.228]	0.205 [0.194-0.216]	0.212 [0.207-0.218]	0.207 [0.205-0.209]	0.208 [0.208-0.208]
SEM_ULS	0.202 [0.094-0.311]	0.204 [0.149-0.258]	0.200 [0.178-0.223]	0.195 [0.184-0.205]	0.199 [0.194-0.205]	0.193 [0.190-0.195]	0.193 [0.193-0.193]
LSAM_ML	-0.253 [-0.364–0.142]	-0.099 [-0.163–0.036]	0.071 [0.046-0.097]	0.137 [0.126-0.149]	0.173 [0.167-0.178]	0.184 [0.182-0.186]	0.196 [0.196-0.196]
LSAM_ULS	-0.251 [-0.363–0.140]	-0.099 [-0.162–0.036]	0.071 [0.046-0.097]	0.137 [0.126-0.149]	0.173 [0.167-0.178]	0.184 [0.182-0.186]	0.196 [0.196-0.196]
GSAM_ML	-0.252 [-0.364–0.140]	-0.099 [-0.163–0.036]	0.071 [0.046-0.097]	0.137 [0.126-0.149]	0.173 [0.167-0.178]	0.184 [0.182-0.186]	0.196 [0.196-0.196]
GSAM_ULS	-0.250 [-0.362–0.138]	-0.099 [-0.162–0.036]	0.071 [0.046-0.097]	0.137 [0.126-0.149]	0.173 [0.167-0.178]	0.184 [0.182-0.186]	0.196 [0.196-0.196]

Note:

Condition: 1_0.3

```
[4] "no non-missing arguments to min; returning Inf"
[5] "no non-missing arguments to min; returning Inf"
```

As before, this warning could indicate either convergence issues or improper solutions due to small sample size in one or multiple estimators. Again, we are unable to check for more details, but hypothesize that this occurs mainly in the standard SEM estimators. There is, however, evidence that this occurs frequently standard SEM estimation and was one claimed advantage of SAM- over SEM-estimation (Rosseel and Loh 2022). The combined study should check this more properly to evaluate these claims thoroughly. As the 612 estimation are around 1.5% of the 42000 estimations performed, we will continue with the interpretation as planned, as their impact is negligable.

Bias

One negative cross-loading (1_-0.3)

All estimators are negatively biased (i.e. relative Bias >= -0.05) across all N conditions. Across all estimators except SEM-ULS, the Bias decreases substantially with increasing N. For SEM-ULS, the bias stays constant across N, and is lowest by a substantial marging compared to all SAM estimators up until N=500. For higher N, SEM-ML is the best, but the differences between all estimators are negligible The 4 SAM estimators have nearly identical estimation results. In relative comparison, the two SEM estimators outperform all SAM estimators in every condition of N, even though this difference is not substantive for N=1000 and higher.

Thus, SEM estimation outperforms SAM in small to moderate sample sizes, when there is one negative cross-loading.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	-0.420 [-0.614–0.226]	-0.376 [-0.478–0.274]	-0.305 [-0.343–0.268]	-0.290 [-0.308–0.273]	-0.270 [-0.278–0.261]	-0.274 [-0.277–0.271]	-0.268 [-0.268–0.268]
SEM_ULS	-0.355 [-0.558–0.153]	-0.320 [-0.424–0.217]	-0.282 [-0.317–0.246]	-0.283 [-0.300–0.266]	-0.271 [-0.279–0.262]	-0.279 [-0.282–0.276]	-0.276 [-0.276–0.276]
LSAM_ML	-0.697 [-0.815–0.579]	-0.614 [-0.682–0.546]	-0.482 [-0.512–0.452]	-0.412 [-0.427–0.396]	-0.366 [-0.374–0.358]	-0.351 [-0.354–0.348]	-0.335 [-0.335–0.335]
LSAM_ULS	-0.695 [-0.814–0.576]	-0.613 [-0.681–0.545]	-0.482 [-0.512–0.452]	-0.412 [-0.427–0.396]	-0.366 [-0.374–0.358]	-0.351 [-0.354–0.348]	-0.335 [-0.335–0.335]
GSAM_ML	-0.695 [-0.817–0.574]	-0.614 [-0.682–0.546]	-0.482 [-0.512–0.452]	-0.412 [-0.427–0.396]	-0.366 [-0.374–0.358]	-0.351 [-0.354–0.348]	-0.335 [-0.335–0.335]
GSAM_ULS	-0.693 [-0.815–0.570]	-0.613 [-0.681–0.545]	-0.482 [-0.512–0.452]	-0.412 [-0.427–0.396]	-0.366 [-0.374–0.358]	-0.351 [-0.354–0.348]	-0.335 [-0.335–0.335]

Note:
Condition: 2_-0.3

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.315 [0.216-0.414]	0.366 [0.321-0.411]	0.393 [0.376-0.410]	0.391 [0.382-0.400]	0.398 [0.394-0.403]	0.393 [0.391-0.394]	0.394 [0.394-0.394]
SEM_ULS	0.355 [0.273-0.438]	0.368 [0.324-0.411]	0.382 [0.365-0.400]	0.376 [0.367-0.385]	0.383 [0.378-0.387]	0.376 [0.374-0.378]	0.377 [0.377-0.377]
LSAM_ML	-0.061 [-0.169-0.047]	0.115 [0.053-0.176]	0.295 [0.273-0.316]	0.350 [0.340-0.360]	0.388 [0.383-0.393]	0.400 [0.398-0.402]	0.412 [0.412-0.412]
LSAM_ULS	-0.060 [-0.169-0.048]	0.115 [0.053-0.176]	0.295 [0.273-0.316]	0.350 [0.340-0.360]	0.388 [0.383-0.393]	0.400 [0.398-0.402]	0.412 [0.412-0.412]
GSAM_ML	-0.060 [-0.168-0.049]	0.115 [0.053-0.176]	0.295 [0.273-0.316]	0.350 [0.340-0.360]	0.388 [0.383-0.393]	0.400 [0.398-0.402]	0.412 [0.412-0.412]
GSAM_ULS	-0.059 [-0.168-0.050]	0.115 [0.053-0.176]	0.295 [0.273-0.316]	0.350 [0.340-0.360]	0.388 [0.383-0.393]	0.400 [0.398-0.402]	0.412 [0.412-0.412]

Note:
Condition: 2_0.3

One positive cross-loadings (1_0.3)

All estimators are biased across all N conditions. For the two SEM-estimators, the relative Bias is always positive and shows no trend across N. All SAM estimators have nearly identical values and increasing relative bias for increasing N. Their bias turn from being negative for N=50-100 to positive from N=250 on. In relative comparison, the two SEM estimators outperform all SAM estimators for N=50. For N=100-2500, the SAM estimators outperform the SEM estimators, even though the difference is unsubstantial starting from N=1000.

Thus, SEM estimation outperforms SAM in small samples (N=50), whereas SAM outperforms SEM in smaller to moderate samples (N100-500), and slightly in higher sample sizes.

Two negative cross-loadings (2_-0.3)

All estimators are negatively biased across all N conditions. Across all estimators, the Bias decreases substantially with increasing N. The 4 SAM estimators have nearly identical estimation results. In relative comparison, the two SEM estimators substantially outperform all SAM estimators in every condition of N.

Thus, SEM estimation outperforms SAM, when there are two negative residuals.

Two positive cross-loadings (2_0.3)

All estimators are biased across N conditions. Both SEM estimators are positively biased across all N, with no visible trend for increasing N for ULS, but a substantial increase for ML. All SAM estimators have nearly identical values and increasing relative bias for increasing N. They are negatively biased in N=50 and positively biased for the other N conditions.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.3443256	0.2317790	0.1361158	0.0918224	0.0656054	0.0407611	0.0062861
SEM_ULS_sd	0.3356916	0.2195551	0.1305268	0.0898937	0.0644590	0.0402754	0.0062310
LSAM_ML_sd	0.2381340	0.1939230	0.1311746	0.0901785	0.0649837	0.0409721	0.0063155
LSAM_ULS_sd	0.2381944	0.1939300	0.1311751	0.0901785	0.0649837	0.0409721	0.0063155
GSAM_ML_sd	0.2413660	0.1939225	0.1311746	0.0901785	0.0649837	0.0409721	0.0063155
GSAM_ULS_sd	0.2416841	0.1939194	0.1311729	0.0901785	0.0649837	0.0409721	0.0063155

Note:
Condition: 1_-0.3

In relative comparison, the four SAM estimators outperform the two SEM estimators for N=50-500, even though unsubstantial for N=500. For all higher N, no estimator substantially differs from another, but the two SEM slightly outperform.

Thus, SAM outperforms SEM in small to moderate samples (N50-250).

Summary

Again, in 1.5% of estimations, a warning referring to potential problems with convergence or improper solutions due to small sample sizes in one or multiple estimators occured. This should be investigated more explicitly in the joint study.

For small to moderate samples, SEM outperforms SAM in conditions with negative cross-loadings. SAM outperforms SEM in conditions with positive cross-loadings.

Comparison with original paper

Comparing these results with the original paper, the results are similar and can be seen as replicated. The only difference lies in the potential convergence issues in a small number of estimations.

Besides that, we draw the same conclusions as Robitzsch (2022): All SAM approaches appear to be generally negatively biased for small to moderate samples in the 2-factor-CFA model. This leads to the comparatively worse performance in the conditions with no, one or two negative residual correlations. Additionally, this negative bias corrects for unmodelled positive cross-loadings and makes SAM appear superior in these conditions.

SD

One negative cross-loading (1_-0.3)

Across all estimators, the standard deviation decreases with increasing N. There are only two substantial differences: Substantial percentage reduction of SD up to 30.8% when comparing SEM-ML and LSAM-ML

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.2502131	0.1779720	0.1090078	0.0743139	0.0537844	0.0336956	0.0050816
SEM_ULS_sd	0.2287762	0.1657764	0.1067319	0.0740809	0.0538710	0.0337516	0.0051060
LSAM_ML_sd	0.2343461	0.1915603	0.1222473	0.0808618	0.0566806	0.0353171	0.0053249
LSAM_ULS_sd	0.2347949	0.1914066	0.1222482	0.0808619	0.0566806	0.0353171	0.0053249
GSAM_ML_sd	0.2356775	0.1916719	0.1222473	0.0808618	0.0566806	0.0353171	0.0053249
GSAM_ULS_sd	0.2361282	0.1914923	0.1222452	0.0808617	0.0566806	0.0353172	0.0053249

Note:
Condition: 1_0.3

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.4088216	0.3080475	0.1803231	0.1182111	0.0831909	0.0507446	0.0078367
SEM_ULS_sd	0.4273783	0.3122573	0.1721577	0.1153720	0.0828871	0.0505562	0.0077926
LSAM_ML_sd	0.2489034	0.2049341	0.1442852	0.1062293	0.0767270	0.0494213	0.0078145
LSAM_ULS_sd	0.2511176	0.2052430	0.1444518	0.1062293	0.0767270	0.0494213	0.0078146
GSAM_ML_sd	0.2564566	0.2053801	0.1442832	0.1062293	0.0767270	0.0494213	0.0078145
GSAM_ULS_sd	0.2590250	0.2057016	0.1444452	0.1062288	0.0767270	0.0494213	0.0078145

Note:
Condition: 2_-0.3

in N=50: Substantial percentage reduction of SD up to 16% when comparing SEM-ML and LSAM-ML in N=100:

Thus, SAM outperforms SEM in terms of SD reduction in smaller samples.

One positive cross-loading (1_0.3)

Across all estimators, the standard deviation decreases with increasing N.

The SAM estimators outperform both SEM estimators in N=50 with a percentage reduction of SD up to 6.4%. ULS-SEM outperforms all SAM estimators with a percentage reduction of SD up to 14.1% in N=100, and of 12.3% in N=250 and of 8.6% in N=500. Above that point, overall SD is very small and negligible in all estimators No substantial differences between the two SEM and all SAM estimators starting from moderate sample sizes (N=500-2500).

Thus, SAM outperforms SEM in terms of SD reduction in small samples, whereas SEM outperforms SAM in smaller to moderate samples.

Two negative cross-loadings (2_-0.3)

Across all estimators, the standard deviation decreases with increasing N.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.2082148	0.1349762	0.0841629	0.0605771	0.0439456	0.0277925	0.0042264
SEM_ULS_sd	0.1751410	0.1308633	0.0853242	0.0621478	0.0454727	0.0287615	0.0043774
LSAM_ML_sd	0.2286619	0.1865175	0.1041352	0.0679428	0.0476475	0.0300411	0.0045870
LSAM_ULS_sd	0.2287458	0.1865255	0.1041356	0.0679428	0.0476475	0.0300411	0.0045870
GSAM_ML_sd	0.2293452	0.1865702	0.1041352	0.0679428	0.0476475	0.0300411	0.0045870
GSAM_ULS_sd	0.2293400	0.1865498	0.1041342	0.0679428	0.0476475	0.0300411	0.0045870

Note:
Condition: 2_0.3

The SAM estimators substantially, but decreasingly outperform both SEM estimators in samples N=50-500 with a SD percentage reduction of 10.1% remaining in N=500. For higher N, there are no substantial differences and overall SD is very small across estimators.

Thus, SAM outperforms SEM in terms of SD reduction in small to moderate samples.

Two positive cross-loadings (2_0.3)

Across all estimators, the standard deviation decreases with increasing N.

Both SEM estimators outperform the SAM estimators in N=50-500 with a SD percentage reduction up to 11.7% remaining in N=500. For higher N, there are no substantial differences.

Thus, SEM outperforms SAM in small to moderate samples.

Summary

For negative cross-loadings, SAM tends to outperform SEM in terms of SD reduction small to moderate samples. For positive cross-loadings, SEM tends to outperform SAM in terms of SD reduction smaller to moderate samples. There seems to be an effect due to the amount of misspecification, such that the conditions with 2 cross-loadings had larger differences in performance (in both directions).

RMSE

One negative cross-loading (1_-0.3)

Across all estimators, the RMSE decreases with increasing N. For small to moderate samples (N=50-500), both SEM estimators were more efficient than all SAM estimators. For higher N, there are no substantial differences present.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.3663843	0.2501193	0.1581473	0.1233194	0.1018884	0.0897700	0.0782766
SEM_ULS_rmse	0.3465368	0.2315299	0.1514690	0.1232106	0.1040559	0.0942171	0.0842276
LSAM_ML_rmse	0.3997701	0.3115102	0.2057931	0.1530333	0.1225556	0.1053797	0.0901037
LSAM_ULS_rmse	0.3989967	0.3113319	0.2057937	0.1530333	0.1225556	0.1053797	0.0901037
GSAM_ML_rmse	0.4015660	0.3115101	0.2057931	0.1530332	0.1225556	0.1053797	0.0901037
GSAM_ULS_rmse	0.4008041	0.3112959	0.2057911	0.1530333	0.1225556	0.1053797	0.0901037

Note:

Condition: 1_-0.3

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.2680044	0.2113311	0.1644895	0.1437515	0.1382670	0.1285665	0.1249105
SEM_ULS_rmse	0.2588713	0.2058474	0.1607456	0.1382830	0.1310986	0.1204395	0.1159952
LSAM_ML_rmse	0.2792001	0.2005604	0.1294806	0.1154805	0.1180471	0.1159376	0.1179137
LSAM_ULS_rmse	0.2790329	0.2003719	0.1294813	0.1154805	0.1180470	0.1159376	0.1179137
GSAM_ML_rmse	0.2799369	0.2006526	0.1294806	0.1154805	0.1180470	0.1159376	0.1179137
GSAM_ULS_rmse	0.2797014	0.2004302	0.1294789	0.1154805	0.1180470	0.1159376	0.1179137

Note:

Condition: 1_0.3

One positive cross-loading (1_0.3)

Across all estimators, the RMSE decreases with increasing N. For smaller and large samples (N=50-100 and N=1000-100000), none of the estimators differed from another. For N=250, all 4 SAM estimators substantially outperformed the two SEM estimators. For N=500, all 4 SAM estimators substantially outperformed SEM-ML estimation.

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.4801738	0.3817372	0.2569580	0.2105236	0.1819276	0.1721264	0.1611536
SEM_ULS_rmse	0.4774695	0.3665606	0.2411851	0.2054523	0.1823036	0.1747702	0.1657878
LSAM_ML_rmse	0.4864230	0.4214094	0.3231554	0.2687705	0.2323712	0.2161971	0.2010006
LSAM_ULS_rmse	0.4867413	0.4210569	0.3231112	0.2687705	0.2323712	0.2161971	0.2010006
GSAM_ML_rmse	0.4895815	0.4217053	0.3231551	0.2687705	0.2323712	0.2161971	0.2010006
GSAM_ULS_rmse	0.4897305	0.4213287	0.3231026	0.2687698	0.2323712	0.2161971	0.2010006

Note:

Condition: 2_-0.3

Condition:							
	Sample Size						
Method	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.2811007	0.2577686	0.2503512	0.2421059	0.2429587	0.2372349	0.2365280
SEM_ULS_rmse	0.2759391	0.2564477	0.2446896	0.2340556	0.2339779	0.2274803	0.2263631
LSAM_ML_rmse	0.2314802	0.1987154	0.2051523	0.2206486	0.2373739	0.2419782	0.2472435
LSAM_ULS_rmse	0.2315006	0.1987204	0.2051523	0.2206486	0.2373739	0.2419782	0.2472435
GSAM_ML_rmse	0.2320582	0.1987755	0.2051523	0.2206486	0.2373739	0.2419782	0.2472435
GSAM_ULS_rmse	0.2319736	0.1987644	0.2051522	0.2206486	0.2373739	0.2419782	0.2472435

Note:
Condition: 2_0.3

Two negative cross-loadings (2_-0.3)

Across all estimators, the RMSE decreases with increasing N. For small samples with N=50, none of the estimators differed in terms of efficiency. For all other conditions, The two SEM estimators were more efficient than all SAM estimators.

Two positive cross-loadings (2_0.3)

Only for the two SEM estimators, the RMSE decreases with increasing N, but remains at higher values between 0.22-0.25 for all estimators.

For small to moderate samples (N=50-250), all SAM estimators were more efficient than the two SEM estimators. For higher N, there are no substantial differences present.

Summary

In moderate samples, SAM generally outperformed SEM in terms of RMSE. The only exception is the condition with 1 negative cross-loading, where SEM outperformed SAM in small to moderate samples. In the condition with 2 positive cross-loadings, SAM was more efficient in small samples as well. Unlike in Study 1, there seems to be no effect due to amount of misspecification.

Summary Study 2

We gained the following insights with regards to the 2-factor-CFA model: Unlike the original paper by Robitzsch (2022), we experienced some issues that are related to convergence or improper solutions in small samples.

In terms of bias in small to moderate samples, SEM outperformed SAM in conditions with negative cross-loadings. SAM appeared to outperform SEM in conditions with positive cross-loadings. This, however, is again an artifact of its negative small sample bias and not its performance. The results align with the ones obtained by the original paper.

With regards to SD reduction, findings were different. In conditions with negative cross-loadings, SAM tended to outperform SEM, whereas the opposite was true for conditions with positive cross-loadings. Here, an effect due to the amount of misspecification seemed to be present, like for RMSE in Study 1.

In terms of RMSE, results differed again. Generally, SAM was more efficient than SEM in moderate samples. For small samples, the results were not as clear. Unlike in Study 1, there seems to be no effect due to amount of misspecification.

Overall, we can infer that findings are somewhat incoherent and it can not be stated that one estimation method is generally performing better than the other, based on the results of Study 2.

Results simulation 3:2-factor-CFA with one cross-loading and one residual correlation

Error, warnings and messages

First, lets check for any messages, warnings or errors:

```
[1] 0
```

```
[1] 193
```

```
[1] 0
```

No errors and messages were present. There were, however, 193 warnings we will investigate in detail.

Warnings investigation

Investigating the 'warnings' object list, it became apparent that all the warnings are the same: "no non-missing arguments to min; returning Inf".

```
unique(warnings_s3)
```

```
[[1]]
character(0)

[[2]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"
[3] "no non-missing arguments to min; returning Inf"
[4] "no non-missing arguments to min; returning Inf"
```

Study 3:

	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rel_bias	0.209 [0.094-0.324]	0.270 [0.213-0.328]	0.283 [0.260-0.306]	0.289 [0.277-0.300]	0.289 [0.284-0.295]	0.282 [0.280-0.285]	0.284 [0.284-0.284]
SEM_ULS_rel_bias	0.250 [0.145-0.356]	0.284 [0.231-0.336]	0.277 [0.255-0.300]	0.280 [0.269-0.291]	0.279 [0.273-0.285]	0.271 [0.269-0.273]	0.272 [0.272-0.272]
LSAM_ML_rel_bias	-0.232 [-0.343–0.121]	-0.061 [-0.128-0.007]	0.127 [0.100-0.153]	0.211 [0.199-0.223]	0.246 [0.240-0.252]	0.261 [0.259-0.263]	0.276 [0.276-0.276]
LSAM_ULS_rel_bias	-0.229 [-0.341–0.118]	-0.060 [-0.127-0.007]	0.127 [0.100-0.153]	0.211 [0.199-0.223]	0.246 [0.240-0.252]	0.261 [0.259-0.263]	0.276 [0.276-0.276]
GSAM_ML_rel_bias	-0.230 [-0.342–0.119]	-0.060 [-0.128-0.007]	0.127 [0.100-0.153]	0.211 [0.199-0.223]	0.246 [0.240-0.252]	0.261 [0.259-0.263]	0.276 [0.276-0.276]
GSAM_ULS_rel_bias	-0.228 [-0.340–0.116]	-0.060 [-0.127-0.007]	0.127 [0.100-0.153]	0.211 [0.199-0.223]	0.246 [0.240-0.252]	0.261 [0.259-0.263]	0.276 [0.276-0.276]

```
[[3]]
[1] "no non-missing arguments to min; returning Inf"
[2] "no non-missing arguments to min; returning Inf"

[[4]]
[1] "no non-missing arguments to min; returning Inf"
```

As before, this warning could indicate either convergence issues or improper solutions due to small sample size in one or multiple estimators. Again, we are unable to check for more details, but hypothesize that this occurs mainly in the standard SEM estimators. There is, however, evidence that this occurs frequently standard SEM estimation and was one claimed advantage of SAM- over SEM-estimation (Rosseel and Loh 2022). The combined study should check this more properly to evaluate these claims thoroughly. As the 193 estimation are around 1.8% of the 10500 estimations performed, we will continue with the interpretation as planned, as their impact is negligable.

Bias

All estimators were biased for every condition of N. The SAM estimators had nearly identical values. Besides SEM-ULS, all estimators' Bias increased substantially, even though the increase was stronger for the SAM estimators. Only the SAM estimators had negative Bias for N=50-100. For N=50, none of the estimators outperformed another. For N=100-500, all SAM estimators outperformed the two SEM estimators. For higher N, none of the estimators substantially differed from another.

Thus, all SAM estimators outperformed the two SEM estimators in smaller to moderate samples.

Comparison with original paper

Comparing these results with the original paper, our results were somewhat different. Even though the results were similar for most conditions (Robitzsch (2022) had the same results in conditions N=100-250 and N=1000-100000), the results were different for N=500. This, combined with the findings from the additional condition of N=50 revealed that SAM estimation outperformed SEM estimation in this data generating scenario in small to moderate samples. Another difference lies, again, in the potential convergence issues in a small number of estimations.

Study 3:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_sd	0.2425692	0.1734508	0.1102019	0.0768395	0.0550863	0.0337544	0.0054908
SEM_ULS_sd	0.2229101	0.1594261	0.1085021	0.0765384	0.0550831	0.0339035	0.0055079
LSAM_ML_sd	0.2344909	0.2033577	0.1280126	0.0843593	0.0576502	0.0352977	0.0056783
LSAM_ULS_sd	0.2347393	0.2034004	0.1280133	0.0843593	0.0576502	0.0352977	0.0056783
GSAM_ML_sd	0.2362469	0.2035047	0.1280126	0.0843593	0.0576502	0.0352977	0.0056783
GSAM_ULS_sd	0.2365296	0.2035117	0.1280109	0.0843592	0.0576502	0.0352977	0.0056783

Study 3:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML_rmse	0.2729248	0.2373975	0.2023008	0.1893898	0.1821868	0.1728205	0.1705186
SEM_ULS_rmse	0.2686908	0.2331863	0.1984849	0.1846220	0.1761678	0.1662260	0.1632532
LSAM_ML_rmse	0.2725827	0.2065099	0.1488332	0.1520811	0.1585422	0.1605814	0.1655320
LSAM_ULS_rmse	0.2720746	0.2065173	0.1488336	0.1520811	0.1585422	0.1605814	0.1655320
GSAM_ML_rmse	0.2736547	0.2066425	0.1488332	0.1520811	0.1585422	0.1605814	0.1655320
GSAM_ULS_rmse	0.2730890	0.2066077	0.1488320	0.1520811	0.1585422	0.1605814	0.1655320

SD

Across all estimators, the standard deviation decreases with increasing N. There are thre substantial differences: Substantial percentage reduction of SD up to 5.5% in favor of SEM-ULS when comparing with GSAM-ULS in N=50. Substantial percentage reduction of SD up to 21.7% when comparing SEM-ULS when comparing with GSAM-ULS in N=100. Substantial percentage reduction of SD up to 15.6% when comparing SEM-ULS when comparing with GSAM-ULS in N=250. For higher N, the SD still partly differs but becomes negligibly small.

Thus, SEM outperforms SAM in terms of SD reduction in small to moderate samples.

RMSE

Across all estimators, the RMSE decreased with increasing N. For N=50 and N=1000-100000, none of the estimators differed from another. For N=100-500, the two SEM-estimators outperform all SAM-estimators in terms of efficiency.

Thus, SEM outperforms SAM in smaller to moderate samples.

Summary Study 3

As before, we experienced some issues that are related to convergence or improper solutions in small samples.

In terms of Bias, all SAM estimators outperformed the two SEM estimators in smaller to moderate samples.

These results differed from the ones obtained with regards to SD and RMSE, where SEM tended to outperform SAM in smaller to moderate samples.

Importantly, one has to bear in mind that same as before, SAM only appears to be more robust against unmodelled positive residual correlations and cross-loadings due to its general negative bias in small samples. As in this study, same as in Robitzsch (2022), the misspecification values were positive, we again cannot conclude that SAM performs better than SEM.

Summary 2-factor-CFA models (Studies 1-3)

SAM does not generally outperform SEM in small to moderate samples. SAM exhibits a negative small sample bias that makes SAM appear superior in conditions with unmodelled positive cross-loadings and residual correlations. This bias is especially strong for lower lambda and higher phi values. If there is no misspecification or unmodelled negative cross-loadings and residual correlations, SAM tends to perform worse than traditional SEM.

Results Simulation 4: 5-factor-model under 3 different data-generating mechanisms

Importantly, for this study, we do not investigate CFA's anymore, but models with regressions included. Also, keep in mind that the naming of the DGM changed to correspond with the code implementation. What Robitzsch (2022) called DGM 1-3, we now will call DGM 0-2.

Error, warnings and messages

First, lets check for any messages, warnings or errors:

```
[1] 0
```

```
[1] 9007
```

```
[1] 0
```

No errors and messages were present. There were, however, 9007 warnings, we will investigate in detail.

Condition:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.147 [0.145, 0.149]	0.098 [0.096, 0.099]	0.060 [0.059, 0.060]	0.041 [0.041, 0.042]	0.029 [0.029, 0.030]	0.018 [0.018, 0.019]	0.003 [0.003, 0.003]
SEM_ULS	0.191 [0.173, 0.210]	0.101 [0.100, 0.103]	0.061 [0.060, 0.062]	0.042 [0.041, 0.042]	0.030 [0.029, 0.030]	0.019 [0.018, 0.019]	0.003 [0.003, 0.003]
LSAM_ML	0.132 [0.130, 0.134]	0.092 [0.091, 0.093]	0.058 [0.057, 0.058]	0.040 [0.039, 0.041]	0.029 [0.028, 0.029]	0.018 [0.018, 0.018]	0.003 [0.003, 0.003]

Note:
Condition: DGM_0

Condition:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.253 [0.248, 0.258]	0.171 [0.168, 0.175]	0.115 [0.113, 0.117]	0.095 [0.093, 0.096]	0.085 [0.084, 0.086]	0.080 [0.079, 0.081]	0.075 [0.074, 0.076]
SEM_ULS	0.445 [0.408, 0.481]	0.291 [0.286, 0.295]	0.263 [0.259, 0.266]	0.254 [0.250, 0.257]	0.249 [0.246, 0.252]	0.248 [0.245, 0.251]	0.247 [0.244, 0.250]
LSAM_ML	0.150 [0.148, 0.152]	0.112 [0.111, 0.114]	0.083 [0.082, 0.084]	0.072 [0.071, 0.073]	0.065 [0.065, 0.066]	0.061 [0.060, 0.062]	0.057 [0.056, 0.058]

Note:
Condition: DGM_1

Warnings investigation

Warning Message
lavaan WARNING: covariance matrix of latent variables is not positive definite; use lavInspect(fit, "cov.lv") to inv
lavaan WARNING: number of observations (100) too small to compute Gamma
lavaan WARNING: number of observations (50) too small to compute Gamma
lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

The three warnings that amount to 7 overall, are referring to problems with regards to positive definite matrices or model identification. This suggests potential issues with multicollinearity or redundancy among latent variables, but only in a negligible number of estimations. The 9000 warnings with regards to Gamma computation in smaller samples is negligible as well, as we are not interested in the computation of robust fit indices.

Bias

Conditions without misspecification (DGM 0)

All estimators are biased for small to moderate samples (N=50-250). For higher N, no estimator is biased. In N=50, LSAM outperforms SEM-ULS. For all other conditions, there is no substantial difference between any of the estimators biases.

Thus, for the most part, no estimator generally outperforms another in the correctly specified model conditions.

Conditions with four unmodelled cross-loadings (DGM 1)

All estimators were biased across all conditions of N. Still, their bias decreases for higher N. LSAM-ML substantially outperforms the two SEM-estimators in smaller samples (N=50-100). It outperforms SEM-

Condition:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.159 [0.157, 0.162]	0.107 [0.106, 0.109]	0.071 [0.070, 0.072]	0.056 [0.055, 0.057]	0.048 [0.048, 0.049]	0.044 [0.043, 0.044]	0.043 [0.043, 0.043]
SEM_ULS	0.295 [0.205, 0.384]	0.108 [0.106, 0.109]	0.067 [0.066, 0.068]	0.049 [0.049, 0.050]	0.039 [0.039, 0.040]	0.031 [0.031, 0.032]	0.029 [0.028, 0.029]
LSAM_ML	0.136 [0.135, 0.138]	0.097 [0.096, 0.098]	0.065 [0.064, 0.066]	0.051 [0.050, 0.052]	0.043 [0.043, 0.044]	0.039 [0.038, 0.039]	0.038 [0.037, 0.038]

Note:
Condition: DGM_2

ULS in every other condition of N, as well. With regards to SEM-ML, LSAM-ML only slightly outperforms, for higher N.

Thus, SAM tends to outperforms SEM in presence of unmodelled cross-loadings.

Conditions with 20 unmodelled residual correlation (DGM 2)

All estimators are biased for small to moderate samples (N=50-500). Only SEM-ULS is unbiased for N=500. For higher N (N=1000-100000), no estimator is biased. In N=50, LSAM outperforms SEM-ULS estimation. For all other conditions, there is no substantial difference between any of the estimators bias.

Summary

LSAM-estimation appears to generally outperform SEM-ULS estimation in smaller samples across all DGM's. Moreover, LSAM- appeared to outperform both SEM-estimators in smaller samples, when unmodelled cross-loadings are present. Importantly, as we found LSAM to have a negative small sample bias in conditions with low values of phi as well in our simulation study 1b, we have to take this into account: Even though LSAM appears to outperform SEM-estimation in DGM 1, this might be attributable to its general negative small sample bias. Thus, it cannot simply be stated, that SAM estimation should be preferred in these scenarios.

In conditions without misspecification (DGM 0) and in the presence of unmodelled residual correlations (DGM 2), there is no general advantage of one estimation method over another, for smaller to moderate samples.

Comparison with original paper

Interestingly, unlike Robitzsch (2022), we have found some bias in DGM 0 even though the model was correctly specified in that scenario.

For DMG 1, results are very similar, with LSAM-ML seemingly outperforming SEM-ULS estimation. A new insight is that LSAM-ML seems to outperform SEM-ML in smaller samples as well.

In DGM 2, we did not find the disadvantage of SEM-ML in comparison to the other estimators that Robitzsch (2022) found.

An additional finding is, that LSAM-ML tends to generally perform better than SEM-ULS in smaller samples.

Condition:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.188	0.123	0.075	0.051	0.037	0.023	0.004
SEM_ULS	1.062	0.128	0.077	0.053	0.037	0.023	0.004
LSAM_ML	0.165	0.115	0.072	0.050	0.036	0.023	0.004

Note:
Condition: DGM_0

Condition:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.373	0.257	0.166	0.124	0.107	0.100	0.095
SEM_ULS	2.070	0.373	0.320	0.306	0.300	0.298	0.296
LSAM_ML	0.188	0.141	0.103	0.089	0.080	0.075	0.071

Note:
Condition: DGM_1

Importantly, our conclusion with regards to LSAM's performance is different, due to the different findings in our Study 1b, that found LSAM to be biased even for low values of phi and lambda.

RMSE

Conditions without misspecification (DGM 0)

Across increasing N, the average RMSE decreases substantially for all estimators. For N=50, SEM-ULS performs substantially worse than the other two estimators. For all other conditions, there is no difference between any of the estimators.

Thus, no estimator generally outperforms another in conditions with a correctly specified model.

Conditions with four unmodelled cross-loadings (DGM 1)

Across increasing N, the average RMSE decreases substantially for all estimators. Across all N, SEM-ULS performs substantially worse than the other two estimators. For N=50-500, SEM-ML performs substantially worse than LSAM-ML. For higher N, SEM-ML performs worse than LSAM-ML as well, but only slightly.

Thus, LSAM-ML appears to outperform the two SEM-estimators for small to moderate samples sizes, in the conditions with four unmodelled cross-loadings.

Condition:							
	Sample Size						
Method/Metric	**50**	**100**	**250**	**500**	**1000**	**2500**	**1e+05**
SEM_ML	0.203	0.136	0.089	0.069	0.058	0.050	0.044
SEM_ULS	5.011	0.136	0.084	0.062	0.048	0.038	0.029
LSAM_ML	0.170	0.121	0.081	0.063	0.052	0.044	0.038

Note:
Condition: DGM_2

Conditions with 20 unmodelled residual correlation (DGM 2)

Across increasing N, the average RMSE decreases substantially for all estimators. For N=50, SEM-ULS performs substantially worse than the other two estimators. For all other conditions, there is no difference between any of the estimators.

Thus, no estimator generally outperforms another in conditions with 20 unmodelled residual correlation.

Summary

LSAM-ML appears to outperfoms the two SEM-estimators in conditions with unmodelled cross-loadings. For conditions without misspecifications and unmodelled residual correlations, no difference arises besides a worse performance of SEM-ULS.

Comparison with original paper

The results align closely to the findings in Robitzsch (2022). A new insight is the comparatively worse performance of SEM-ULS in small samples (N=50).

Summary Study 4

Overall, the main performance difference is that LSAM appears to be outperforming the two SEM-estimators in smaller to moderate samples sizes, in conditions with unmodelled cross-loadings (DGM1). Even though LSAM appears to outperform SEM-estimation in DGM 1, this might be attributed to its general negative small sample bias, that persisted even for low phi values in our study1b. Thus, it cannot simply be stated, that SAM estimation should be preferred in these scenarios.

In conditions without misspecification (DGM 0) and in the presence of unmodelled residual correlations (DGM 2), there is no general advantage of one estimation method over another. A new insight is the comparatively worse performance of SEM-ULS in small samples (N=50). Due to our different findings with regards to the small sample bias of LSAM in Study 1b, we arrive at different conclusions than Robitzsch (2022).

Results Simulation 4a: 5-factor-model under 2 different data-generating mechanisms and varying beta size

Again, we investigated a five factor model with regressions included, and this time under vary size of the regression coefficient beta. Unlike the original paper, were the investigation was only conducted on the population level, we also varied the sample size as a factor in our simulation. Also, keep in mind that the naming of the DGM changed to correspond with the code implementation. What Robitzsch (2022) called DGM 1-3, we now will call DGM 0-2.

Error, warnings and messages

First, lets check for any messages, warnings or errors:

```
[1] 0
```

```
[1] 24089
```

```
[1] 0
```

No errors and messages were present. There were, however, 9007 warnings, we will investigate in detail.

Warnings investigation

Warning Message
lavaan WARNING: covariance matrix of latent variables is not positive definite; use lavInspect(fit, ”cov.lv”) to inv
lavaan WARNING: number of observations (100) too small to compute Gamma
lavaan WARNING: number of observations (50) too small to compute Gamma
lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

Study 4a: Absolute Bias of SEM-ML for DGM 1							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.253 [0.248, 0.258]	0.172 [0.168, 0.175]	0.115 [0.113, 0.118]	0.095 [0.093, 0.096]	0.085 [0.084, 0.086]	0.080 [0.079, 0.081]	0.075 [0.074, 0.076]
0.2	0.327 [0.311, 0.342]	0.220 [0.215, 0.224]	0.162 [0.159, 0.165]	0.139 [0.136, 0.141]	0.122 [0.121, 0.124]	0.114 [0.112, 0.115]	0.109 [0.108, 0.111]
0.3	0.545 [0.466, 0.624]	0.270 [0.266, 0.275]	0.218 [0.215, 0.221]	0.206 [0.203, 0.209]	0.195 [0.192, 0.197]	0.190 [0.188, 0.192]	0.180 [0.179, 0.182]
0.4	0.981 [0.863, 1.099]	0.336 [0.328, 0.344]	0.270 [0.266, 0.274]	0.257 [0.254, 0.260]	0.252 [0.249, 0.255]	0.251 [0.249, 0.254]	0.254 [0.251, 0.256]

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi
lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be posi

Similarly as in Study 4, next to the warnings regarding the Gamma calculation, all 89 other warnings referred to problems with positive definite matrices or model identification. This suggests potential issues with multicollinearity or redundancy among latent variables, but only in a negligible number of estimations. The 9000 warnings with regards to Gamma computation in smaller samples is negligible as well, as we are not interested in the computation of robust fit indices.

Bias

```r
# Load the absolute bias results
bias_ci_s4a <- readRDS("SimulationResultsProcessed/sim4a_abs_bias_ci.rds")
```

Conditions with four unmodelled cross-loadings (DGM 1)

All three estimators were biased across all conditions. The only exception is LSAM for very large samples (N=2500-100.000) and beta=0.3. For all three estimators, across increasing N, the bias decreased substan-

Study 4a: Absolute Bias of SEM-ULS for DGM 1							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.445 [0.408, 0.481]	0.291 [0.286, 0.295]	0.263 [0.259, 0.266]	0.254 [0.250, 0.257]	0.249 [0.246, 0.252]	0.248 [0.245, 0.251]	0.247 [0.244, 0.250]
0.2	0.697 [0.593, 0.800]	0.324 [0.319, 0.329]	0.285 [0.282, 0.289]	0.276 [0.273, 0.279]	0.270 [0.267, 0.273]	0.269 [0.266, 0.272]	0.267 [0.264, 0.270]
0.3	1.079 [0.953, 1.204]	0.429 [0.405, 0.453]	0.322 [0.318, 0.326]	0.308 [0.305, 0.312]	0.299 [0.296, 0.303]	0.298 [0.294, 0.301]	0.294 [0.291, 0.298]
0.4	1.991 [1.796, 2.186]	0.654 [0.607, 0.700]	0.385 [0.380, 0.390]	0.357 [0.353, 0.362]	0.344 [0.340, 0.348]	0.339 [0.335, 0.343]	0.333 [0.329, 0.336]

Study 4a: Absolute Bias of LSAM-ML for DGM 1							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.150 [0.148, 0.152]	0.112 [0.111, 0.114]	0.083 [0.082, 0.084]	0.072 [0.071, 0.073]	0.065 [0.065, 0.066]	0.061 [0.060, 0.062]	0.057 [0.056, 0.058]
0.2	0.145 [0.143, 0.147]	0.108 [0.107, 0.110]	0.077 [0.076, 0.078]	0.066 [0.065, 0.067]	0.059 [0.058, 0.060]	0.056 [0.056, 0.057]	0.053 [0.053, 0.054]
0.3	0.141 [0.139, 0.143]	0.104 [0.102, 0.105]	0.073 [0.072, 0.074]	0.060 [0.059, 0.061]	0.052 [0.051, 0.053]	0.047 [0.047, 0.048]	0.043 [0.042, 0.043]
0.4	0.151 [0.149, 0.154]	0.115 [0.113, 0.116]	0.088 [0.086, 0.089]	0.076 [0.075, 0.077]	0.070 [0.069, 0.071]	0.067 [0.066, 0.068]	0.065 [0.064, 0.066]

tially, in all conditions of beta. For higher beta, the bias increased substantially, in all conditions of N for SEM-ML and SEM-ULS. This was not the case for LSAM-ML.

Across all conditions, LSAM-ML substantially outperformed the two SEM-estimators.

The bias difference between LSAM-ML and the two other estimators is higher for lower N.

The only exception was for beta =0.1 in N=250 and all higher N. Here, there was only a slightly better performance of LSAM- over SEM-ML estimation, though not substantial.

This effect was stronger for higher values of beta. Additionally, SEM-ML outperformed SEM-ULS in all conditions as well.

Thus, LSAM-ML generally appeared to outperform SEM-estimation and SEM-ML outperformed SEM-ULS estimation, in the presence of four unmodelled cross-loadings. This was especially evident for higher beta values and lower N.

Conditions with 20 unmodelled residual correlation (DGM 2)

All estimators were generally biased across conditions, besides conditions of very high N (N=1000-100.000) and lower to moderate beta (0.1-0.3). For all three estimators, across increasing N, the bias decreased substantially, in all conditions of beta. Only in SEM-ML in N=50 and SEM-ULS in N=100, there was a substantial increase in bias for higher beta values.

For N=50, both SEM- and LSAM-ML substantially outperformed ULS estimation. In these conditions, LSAM-ML also outperformed SEM-ML. for beta 0.2-0.4. Lastly, LSAM-ML outperformed SEM-ULS in N=100 for beta = 0.4.

For all other conditions, none of the estimators differed from another, in terms of bias.

Study 4a: Absolute Bias of SEM-ML for DGM 2							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.159 [0.157, 0.162]	0.107 [0.106, 0.109]	0.071 [0.070, 0.072]	0.056 [0.055, 0.057]	0.048 [0.048, 0.049]	0.044 [0.043, 0.044]	0.043 [0.043, 0.043]
0.2	0.168 [0.165, 0.170]	0.112 [0.110, 0.113]	0.072 [0.071, 0.073]	0.056 [0.056, 0.057]	0.046 [0.046, 0.047]	0.041 [0.040, 0.041]	0.038 [0.038, 0.038]
0.3	0.181 [0.178, 0.184]	0.122 [0.120, 0.124]	0.080 [0.079, 0.082]	0.063 [0.062, 0.064]	0.052 [0.052, 0.053]	0.046 [0.046, 0.047]	0.041 [0.041, 0.041]
0.4	0.218 [0.215, 0.222]	0.145 [0.143, 0.147]	0.098 [0.097, 0.100]	0.080 [0.079, 0.081]	0.069 [0.068, 0.069]	0.063 [0.062, 0.063]	0.055 [0.054, 0.056]

Study 4a: Absolute Bias of SEM-ULS for DGM 2							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.295 [0.205, 0.384]	0.108 [0.106, 0.109]	0.067 [0.066, 0.068]	0.049 [0.049, 0.050]	0.039 [0.038, 0.040]	0.031 [0.031, 0.032]	0.029 [0.028, 0.029]
0.2	0.257 [0.220, 0.294]	0.113 [0.112, 0.115]	0.070 [0.069, 0.071]	0.052 [0.051, 0.053]	0.040 [0.040, 0.041]	0.032 [0.032, 0.032]	0.027 [0.027, 0.028]
0.3	0.353 [0.220, 0.486]	0.125 [0.123, 0.127]	0.078 [0.077, 0.079]	0.059 [0.058, 0.060]	0.047 [0.046, 0.047]	0.038 [0.038, 0.039]	0.032 [0.032, 0.032]
0.4	0.330 [0.283, 0.376]	0.157 [0.148, 0.167]	0.096 [0.094, 0.097]	0.075 [0.074, 0.076]	0.063 [0.062, 0.063]	0.055 [0.055, 0.056]	0.047 [0.047, 0.048]

Study 4a: Absolute Bias of LSAM-ML for DGM 2							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.136 [0.135, 0.138]	0.097 [0.096, 0.098]	0.065 [0.064, 0.066]	0.051 [0.050, 0.052]	0.043 [0.043, 0.044]	0.039 [0.038, 0.039]	0.038 [0.037, 0.038]
0.2	0.130 [0.129, 0.132]	0.091 [0.090, 0.092]	0.058 [0.057, 0.059]	0.043 [0.043, 0.044]	0.033 [0.033, 0.034]	0.025 [0.025, 0.026]	0.019 [0.019, 0.020]
0.3	0.125 [0.123, 0.127]	0.088 [0.086, 0.089]	0.057 [0.056, 0.057]	0.042 [0.042, 0.043]	0.033 [0.033, 0.033]	0.026 [0.025, 0.026]	0.022 [0.022, 0.022]
0.4	0.133 [0.131, 0.134]	0.100 [0.099, 0.102]	0.077 [0.076, 0.078]	0.068 [0.067, 0.069]	0.064 [0.063, 0.064]	0.061 [0.060, 0.061]	0.057 [0.057, 0.058]

Thus, LSAM-ML appeared to perform best in small samples and higher beta. Besides that, none of the estimators generally outperforms another in the presence of 20 unmodelled residual correlations.

RMSE

Conditions with four unmodelled cross-loadings (DGM 1)

Across all conditions of N and beta, LSAM-ML estimation was more efficient than the two SEM-estimators. This difference is stronger in smaller samples. Additionally, SEM-ML outperformed SEM-ULS estimation, in terms of efficiency, in all conditions as well. Both SEM estimators have reduced RMSE values for higher N. All three estimators have higher RMSE values for higher beta. This effect was especially present in the two SEM estimators.

Thus, LSAM-ML appeared to outperform SEM-estimation and SEM-ML appeared to outperform SEM-ULS estimation, in the presence of four unmodelled cross-loadings. This was especially true in lower samples and higher beta values.

Conditions with 20 unmodelled residual correlation (DGM 2)

Across all conditions of N, LSAM-ML generally outperforms the two SEM-estimators for lower to moderate beta values (approx. beta= 0.2-0.4). This effect is stronger for higher values of beta and lower N. SEM-ULS had reduced RMSE values for higher N. All three estimators have higher RMSE values for higher beta. This effect was especially present in the two SEM estimators.

Study 4a: RMSE of SEM-ML for DGM 1							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.451	0.423	0.403	0.394	0.391	0.390	0.390
0.2	0.573	0.532	0.516	0.505	0.498	0.493	0.492
0.3	0.701	0.609	0.595	0.591	0.586	0.584	0.577
0.4	0.949	0.674	0.648	0.644	0.643	0.643	0.645

Study 4a: RMSE of SEM-ULS for DGM 1							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.558	0.515	0.506	0.503	0.502	0.502	0.502
0.2	0.771	0.594	0.577	0.573	0.571	0.570	0.570
0.3	0.924	0.688	0.634	0.629	0.625	0.624	0.623
0.4	1.221	0.816	0.690	0.678	0.673	0.671	0.669

Study 4a: RMSE of LSAM-ML for DGM 1							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.358	0.360	0.365	0.366	0.367	0.367	0.368
0.2	0.425	0.429	0.433	0.435	0.435	0.435	0.436
0.3	0.442	0.449	0.454	0.455	0.455	0.456	0.456
0.4	0.417	0.425	0.429	0.430	0.431	0.431	0.431

Study 4a: RMSE of SEM-ML for DGM 2							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.378	0.370	0.367	0.366	0.365	0.365	0.365
0.2	0.464	0.453	0.448	0.447	0.446	0.445	0.445
0.3	0.524	0.512	0.505	0.503	0.502	0.502	0.501
0.4	0.577	0.557	0.548	0.545	0.543	0.543	0.542

Study 4a: RMSE of SEM-ULS for DGM 2							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.377	0.352	0.347	0.346	0.345	0.344	0.344
0.2	0.452	0.443	0.437	0.435	0.433	0.433	0.433
0.3	0.616	0.506	0.497	0.494	0.493	0.492	0.492
0.4	0.626	0.553	0.542	0.539	0.536	0.536	0.535

Study 4a: RMSE of LSAM-ML for DGM 2							
	Sample size						
beta	**N=50**	**N=100**	**N=250**	**N=500**	**N=1000**	**N=2500**	**N=100.000**
0.1	0.352	0.354	0.356	0.357	0.357	0.357	0.357
0.2	0.417	0.420	0.423	0.423	0.423	0.423	0.424
0.3	0.439	0.443	0.446	0.447	0.447	0.447	0.447
0.4	0.419	0.424	0.427	0.428	0.428	0.428	0.429

Thus, LSAM-ML appeared to generally outperform SEM-estimation, in the presence of 20 unmodelled residual correlation. Even more so, for higher values of beta and lower N.

Summary Study 4a

In terms of Bias, only in the presence of four unmodelled cross-loadings, did LSAM-estimation appear to generally outperform both SEM-estimators. This was especially evident for higher beta values and lower N. In the presence of unmodelled residual correlations, LSAM-ML appeared to perform best in small samples and higher beta as well.

With regards to RMSE, LSAM-ML appeared to outperform SEM-estimation both under unmodelled cross-loadings and residual correlations. Again, this appeared to be especially present in conditions with small samples and high beta values.

Going all the way back to Studies 1-3, we again have to remember that LSAM exhibited a general negative bias in smaller samples for the 2-factor-CFA. This makes it appear to outperform in conditions of lower N and higher phi or beta. As replicated in Study 1b, this bias tends to be even stronger for higher parameter values in the structural model, and thus probably for higher beta values as well.

To paint a complete picture of the comparative performance of SAM vs. SEM estimation, the joint study should include more conditions with negatively valenced unmodelled cross-loadings and residual correlations, as well as include a DGM with correctly specified models to investigate a potential small sample bias in models with regressions.

Comparison to original paper

Importantly, Robitzsch (2022) did not investigate differential effects due to sample size. Nor did they investigate RMSE. Thus, results are only partly comparable.

In DMG1, the original paper indicated the same results, in that LSAM performed best. Interestingly, Robitzsch (2022) did not find the better performance of SEM-ML over SEM-ULS estimation that we found. Also, no differential effect for higher beta values was found in the original paper.

In DGM2, Robitzsch (2022) found SEM-ULS to perform slightly better than the two others. Thus, results are quite similar, as we also did not find substantial differences between the three estimators.

We disagree with the conclusion that Robitzsch (2022) draw with regards to this study, as they deem SEM-ULS and LSAM do not differ in performance, when in fact LSAM outperformed SEM-ULS in DGM1 in their studies.

Additionally, later on in the paper, they posit that SAM can be expected to be more robust than both SEM-estimators in models with non-saturated structural models. With this last conclusion, we disagree for now with regards to the now multiple times mentioned potential negative small sample bias of LSAM, that appears to be present in non-saturated structural models as well.

Summary for 5-factor-model under 3 different data-generating mechanisms (Studies 4 and 4a)

For low beta (0.1), LSAM appeared to outperform the two SEM-estimators in smaller to moderate samples sizes, in conditions with unmodelled cross-loadings (DGM1). In conditions without misspecification (DGM 0) and in the presence of unmodelled residual correlations (DGM 2), there is no general advantage of one estimation method over another.

With increasing values of beta (0.2-0.4), the effect with regards to unmodelled cross-loadings became even stronger (DGM 1). In this context, LSAM appeared to not only outperform SEM-estimation in terms of efficiency in small to moderate samples in DGM1, but also in conditions with unmodelled residual correlations (DGM2).

Very important to keep in mind here is a finding of the earlier study and the original paper, that LSAM exhibited a negative finite sample bias that skewed results in conditions with positive values for unmodelled cross-loadings or residual correlations (Robitzsch 2022). This effect was even stronger for higher values of phi and low N, which aligns with the findings in our studies for models with regressions..

Thus, in the joint study, conditions with negative unmodelled cross-loadings or residual correlations should be included for unsaturated structural models. Additionally, conditions with correctly specified models should be included for higher values of beta. Adding these conditions would shed light onto the question of whether LSAM's small sample bias is present in regression models as well, or whether it actually outperforms SEM in these models.

Overall summary

With regards to all studies conducted, as in the original paper by Robitzsch (2022), SAM does not generally outperform SEM in small to moderate samples. SAM exhibits a negative small sample bias that makes SAM appear superior in conditions with unmodelled positive cross-loadings and residual correlations. This bias is especially strong for lower lambda and higher phi or beta values. Going ahead of what was investigated in Robitzsch (2022), we found that this bias might also present in models with regressions. Thus, it cannot be concluded that SAM is more robust in non-saturated structural model, as of yet. If there is no misspecification or unmodelled negative cross-loadings and residual correlations, SAM tends to perform worse than traditional SEM, as far as can be concluded from the 2-factor-CFA-models.

Starting points for joint simulation

With regards to the joint simulation, three main aspects might be interesting to investigate in further data constellations. Firstly, with regards to the conceptual side of the simulation, we should include the explicit monitoring of non-convergence, as we found potential issues with it in multiple studies. The implementation of the studies should allow to capture and investigate these issues in more detail.

Secondly, conditions with negative unmodelled cross-loadings and residual correlations should be included for unsaturated structural models. This would be interesting to be able to differentiate whether the negative finite sample bias of LSAM, that became apparent in the 2-factor-CFA, is also influencing the results in the 5-factor-model with regressions. For the same reason, conditions resembling the inclusion of DGM 0 (so correctly specified regression models) in Study 4a (under increasing beta and N) should be investigated. Only if the presence of a negativ small sample bias can be ruled out in regression models, one could discuss if SAM generally outperforms SEM in these models.

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.

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.