

Replication Report Rhemtulla et al 2012

Anna Lohmann¹, Arjan Huizing²

¹ Leiden University Medical Center

² TNO (Netherlands Organization for Applied Scientific Research)

May 24, 2022

Abstract

This documents the replication attempt of the simulation study reported in Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>. The study compared two different estimation methods (robust Maximum Likelihood (ML) and categorical least squares (cat-LS/ULSMV)) for fitting confirmatory factor analysis models in the context of categorical variables. Our replication involved writing simulation code based on the information provided in the manuscript and the corresponding supplemental material. Information provided in the original study was detailed and well structured, thus allowing us to reimplement the study to the best of our knowledge. Detailed result tables provided in the supplemental material allowed us to compare our replicated results to the original results.

Correspondence concerning this replication report should be addressed to:
a.l.lohmann@lumc.nl

1 Introduction

This replication report documents the replication attempt of the simulation study

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>

Following the definition of Rougier et al. (2017) we understand the replication of a published study as writing and running new code based on the description provided in the original publication with the aim of obtaining the same results.

2 Method

2.1 Information basis

The replication attempt was based on the information provided in the original manuscript as well as the supplemental material accompanying the publication. The main text provided a link to the supplements (<http://dx.doi.org/10.1037/a0029315.supp>) which referred to the website of the publisher where an additional pdf document with extensive result tables was freely available.

2.2 Data Generating Mechanism

The information provided indicated that the following simulation factors were systematically varied in a full-factorial design for generating the artificial data.

Simulation factor	No. levels	Levels
<i>Varied</i>		
CFA model size	2	10 indicators, 20 indicators
Underlying distribution	2	normal, non-normal
Number of categories	6	2,3,4,5,6,7
Threshold symmetry	5	symmetry, moderate asymmetry, moderate asymmetry alternative, extreme asymmetry
Sample Size	4	100, 150, 350, 600
<i>Fixed</i>		
factor loadings		0.3, 0.4, 0.5, 0.6, 0.7
factor correlation		0.3

This results in a total of 480 scenarios under which data is generated. Each of these conditions was simulated with 1000 repetitions.

Generating data consisted of two steps. (1) Data was generated based on the underlying distribution, CFA model and sample size. (2) The generated data was categorized based on the given category thresholds corresponding to a given number of categories and threshold symmetry.

2.2.1 CFA model

The CFA models underlying data generation were described as “Model 1 was a two-factor CFA model with five indicators per factor, for a total of 10 indicators. Factor loadings for the five indicators were .3, .4, .5, .6, .7. [...] The model was identified by fixing the variances of each latent variable to 1. Generated continuous variables had unit variance (prior to categorization). Model 2 was identical to Model 1, but with 10 indicators per factor.” (p.359) We translated this information into the following matrices:

$$\Lambda = \begin{bmatrix} 0.3 & 0 \\ 0.4 & 0 \\ 0.5 & 0 \\ 0.6 & 0 \\ 0.7 & 0 \\ 0 & 0.3 \\ 0 & 0.4 \\ 0 & 0.5 \\ 0 & 0.6 \\ 0 & 0.7 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We used these matrices as input for the `model()` function of the `simsem` package.

2.2.2 Underlying distribution, CFA model size and Sample Size

The original study indicated that data were generated using the Fleishman (1978) and Vale Maurelli (1983) method. We emulated this approach using the `generate()` function from the `simsem` package with the parameter `inDist` set to `NULL` in the normal case and to `simsem::bindDist(skewness = 2, kurtosis = 7)` in the non-normal case. The `model` parameter from the `generate()` was specified as detailed above. This constituted the first step of the data generation.

2.2.3 Number of categories and Threshold symmetry

After data was generated based on the given CFA model and the underlying distribution the resulting data was categorized into the number of categories for the scenario at hand. For each number of categories and each threshold symmetry, Z-scores for category thresholds could be obtained from the first table of the supplemental material. The sample covariance matrix of the resulting categorized data was tested for positive definiteness. In case it was found to be non-positive definite data was resampled with a different seed until it was. Additionally, it was ensured that none of the generated variables had zero variance. These measures are not documented in the original study. Hence, we do not know whether or at which point in the simulation pipeline these issues were dealt with.

2.3 Investigated Methods

The study compares the performance of robust normal theory maximum likelihood (ML) and robust categorical least squares (ULS) methodology for estimating confirmatory factor analysis (CFA) with ordinary variables. The underlying CFA model was fit using each of the two methods under investigation.

2.3.1 Robust normal theory maximum likelihood (ML)

CFA's were carried out using the `cfa()` function of the `lavaan` package. For the *Robust normal theory maximum likelihood* approach we set the `estimator` argument to `MLVM`.

2.3.2 Robust categorical least squares (ULS)

The *Robust categorical least squares (ULS)* approach was also implemented using the `cfa()` function from the `lavaan` package. In this case the `estimator` argument was set to `ULSMV`. Additionally, the `ordered` argument was set to `TRUE`.

<Describe how the second method is defined and implemented. You can include equations and or R code. If applicable, mention specialized R packages, their version as well as, parameters of specific functions.>

2.4 Performance measures

The models estimated using the two methods described above were compared on various performance measures.

2.4.1 Convergence Failures

The original article assessed the number of convergence failures. We implemented convergence failure via the `lavInspect()` function with the `what` argument set to "converged".

2.4.2 Improper solutions

The original study reports assessing the number of improper solutions. The paper defines improper solution as *"when cat-LS estimation produced a factor loading greater than 1 or continuous ML estimation produced a standardized factor loading greater than 1"* (p.361) We implemented convergence failure via the `lavInspect()` function with the `what` argument set to "post.check".

2.4.3 Parameter Estimates

We extracted parameter estimates from the fitted lavaan object using the `lavInspect()` function.

2.4.4 Parameter Bias

The parameter bias was calculated as the difference of the mean estimate per scenario and the true value.

2.4.5 Coverage

For each iteration of each scenario it was assessed whether the estimated parameter fell within 1.96 standard errors of the true value. We used robust standard errors from the estimated model for this assessment.

2.5 Power

In addition to the above mentioned analyses the study included a brief evaluation of the relative power of the ML-based and the ULS-based robust test statistics to detect a least major model misspecification. For this purpose the authors fit a *"one-factor model to the data generated by Model 1 (the 10-indicator, two factor model) for the subset of conditions in which the underlying distribution was normal and thresholds were symmetrically distributed."* (p.369). This subset corresponds to 60 of the 480 scenarios. We interpreted the above to indicate that the same generated data as for the rest of the simulation study was used. We hence filtered the generated data sets to only retain the scenarios including model 1, normally distributed variables and symmetrically distributed thresholds for categorization and fit a one-factor model to each of the data sets that fit these criteria.

A p-value < 0.05 of the robust χ^2 statistic was used to indicate a model misspecification.

2.6 Technical implementation

The original simulation study was carried out in EQS (Version 6.1) as well as Mplus (Version 6.11). The authors of the original study report that data generation was carried out in EQS and data analysis was conducted using both EQS as well as MPlus. However, only results from the Mplus analysis are reported. Our replication was implemented using the R programming environment (details regarding software versions can be obtained from the section Reproducibility Information). The corresponding R code can be obtained from <https://github.com/replisims/rhemtulla-2012>.

2.7 Replicator degrees of freedom

The following table provides an overview of replicator degrees of freedom, i.e. decisions that had to be made by the replicators because of insufficient or contradicting information. Issues were resolved by discussion among the replicators.

Issue	Replicator decision	Justification
Data basis fig 1&2, tab 1	Simulate just one variable	it seemed unlikely that dozens of variables from the models were collapsed
Factor loadings of Model 2	each factor loading is assumed to occur twice	most likely
Error handling	Case-wise deletion	Text indicated that “cases” were removed

2.7.1 Data basis for Figures 1 and 2

<More details on how the information provided was insufficient, unclear or vague> “Some weird quote from the original article that you could not make any sense of” (p.XY) The text indicated that the data underlying figures 1 and 2 as well as table 1 were generated for each “scenario” and a sample size of 1000000. We interpreted this to mean that one variable of length 1000000 was generated according to the specifications of each scenario although each scenario technically generated data according to an entire CFA model.

2.7.2 Factor loadings of model 2

The original article indicated that “Model 2 was identical to Model 1, but with 10 indicators per factor.”(p.359) No additional information regarding the factor loadings for these additional factor loadings was provided. We hence assumed that additional indicators reused the same set of factor loadings such that each loading occurred twice.

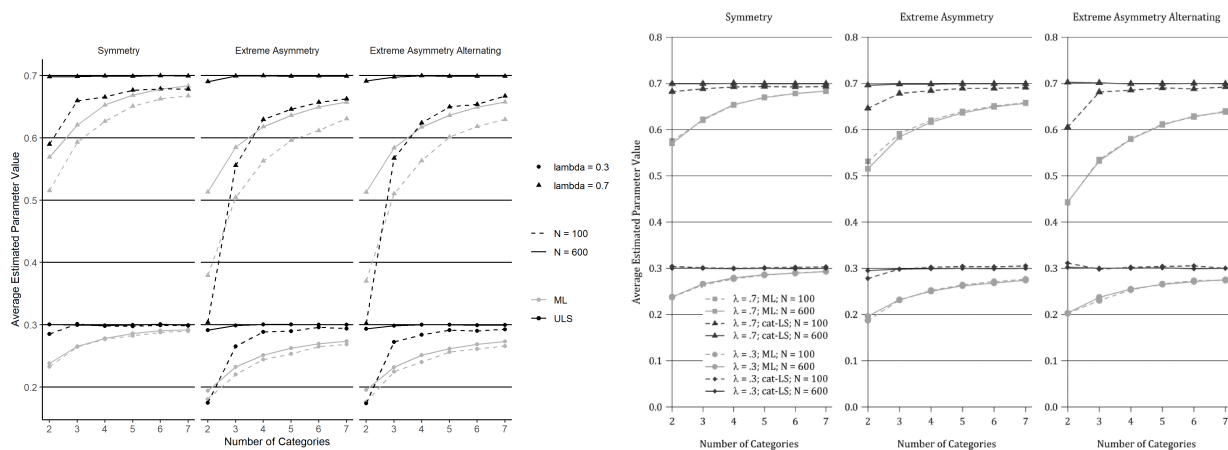


Figure 1: Parameter estimates (factor loadings, underlying distribution is normal). Values are averaged across model size and across all loadings for which the true parameter value was the same. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper set of lines represents results for a true parameter value of .7. The lower set of lines represents results for a true parameter value of .3. Vertical panels represent different levels of threshold symmetry. Left figure: replication; right figure original study.

2.7.3 Error handling

<More details on how the information provided was insufficient, unclear or vague> “Some weird quote from the original article that you could not make any sense of” (p.XY)

3 Results

3.1 Replication of result figures

The original study provides descriptives for the simulated data in two figures. Figure 1 and Figure 2 of the original manuscript

3.1.1 Figure 3 and 4 Parameter estimates (factor loadings)

The results pertaining to the robust ML estimator are largely comparable to the original results both in magnitude as well as regarding trend. Contrary to the original results our replication exhibited a larger downwards bias for $N = 100$.

For $N = 600$ the results pertaining to the ULS estimator closely align with the original results. For $N = 100$ we obtained parameter estimates that exhibited noticeably more downwards bias especially for lower numbers of categories.

These patterns also hold for the non-normal scenarios. The only exception being the 2-category scenario

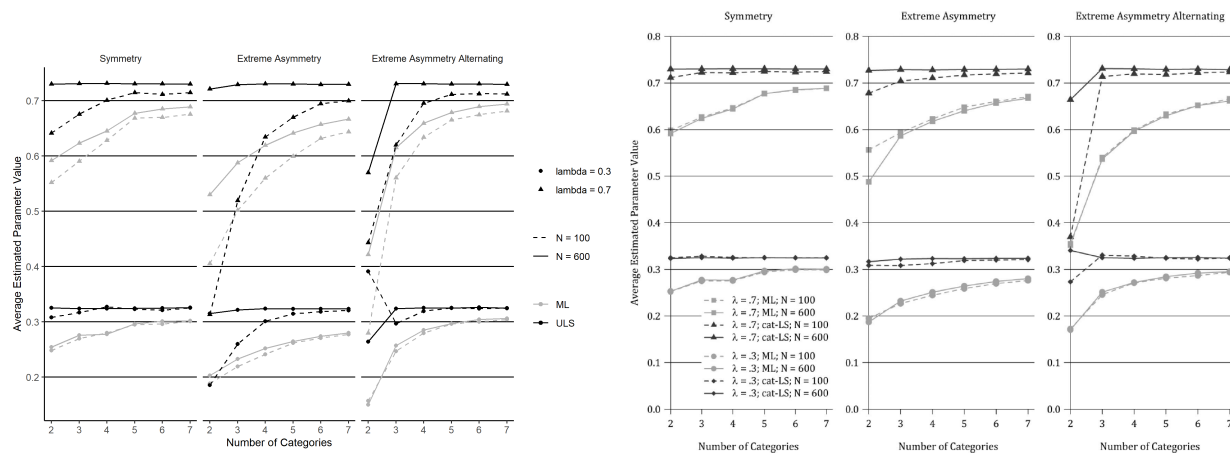


Figure 2: Parameter estimates (factor loadings, underlying distribution is nonnormal; skew 2, kurtosis 7). Values are averaged across model size and across all loadings for which the true parameter value was the same. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper set of lines represents results for a true parameter value of .7. The lower set of lines represents results for a true parameter value of .3. Vertical panels represent different levels of threshold symmetry. Left figure: replication; right figure original study.

where large discrepancies can be observed for the ULS estimator and $N = 600$.

3.1.2 Figure 5 Parameter estimates (factor correlation)

Parameter estimates for the factor correlations largely align with the original results. For scenarios where $N = 100$, we observed a larger downwards bias, especially for scenarios with a low number of categories.

3.1.3 Figure 6 and 7 Coverage (factor loadings)

Regarding coverage the trends in our results correspond to the original findings. Regarding magnitude, our results show consistently lower coverage especially with ML estimator and lower number of categories.

3.1.4 Figure 8 Coverage (factor correlations)

Type I error of mean-and variance adjusted test statistic roughly aligns for symmetry and extreme asymmetry scenarios. In the Extreme Asymmetry Alternating scenarios the original study finds considerably higher type I error rates for scenarios pertaining to the ML estimator and $N = 600$.

Regarding coverage of the factor correlation our results closely align with the original findings considering trends. Considering magnitude, coverage in the $N=100$ scenarios is consistently lower.

3.1.5 Type I error rate

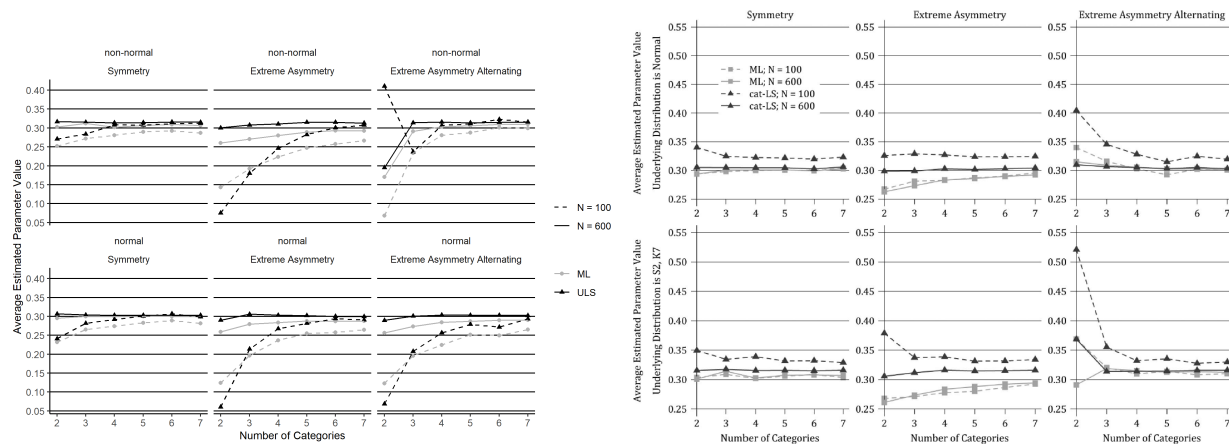


Figure 3: Parameter estimates (factor correlation, true value is .3). Values are averaged across model size. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper panel corresponds to conditions in which the underlying distribution is normal; the lower panel corresponds to conditions in which the underlying distribution is nonnormal (skew 2, kurtosis 7). Vertical panels represent different levels of threshold symmetry. Left figure: replication; right figure original study.

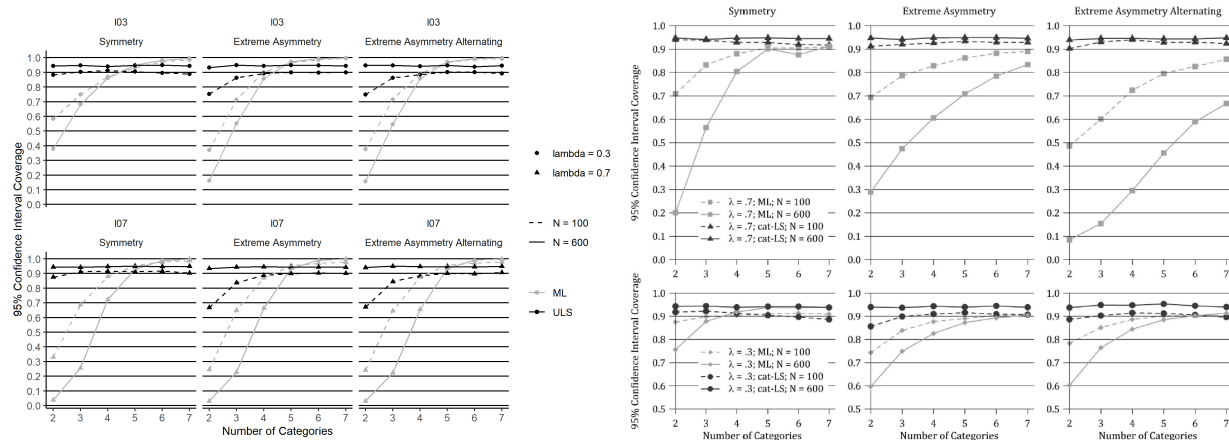


Figure 4: Coverage by number of categories (.7 and .3 factor loadings); underlying distribution is normal. Values are averaged across model size and across all loadings for which the true parameter value was the same. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper panel represents results for a true parameter value of .7. The lower panel represents results for a true parameter value of .3. Vertical panels represent different levels of threshold symmetry. Left figure: replication; right figure original study.

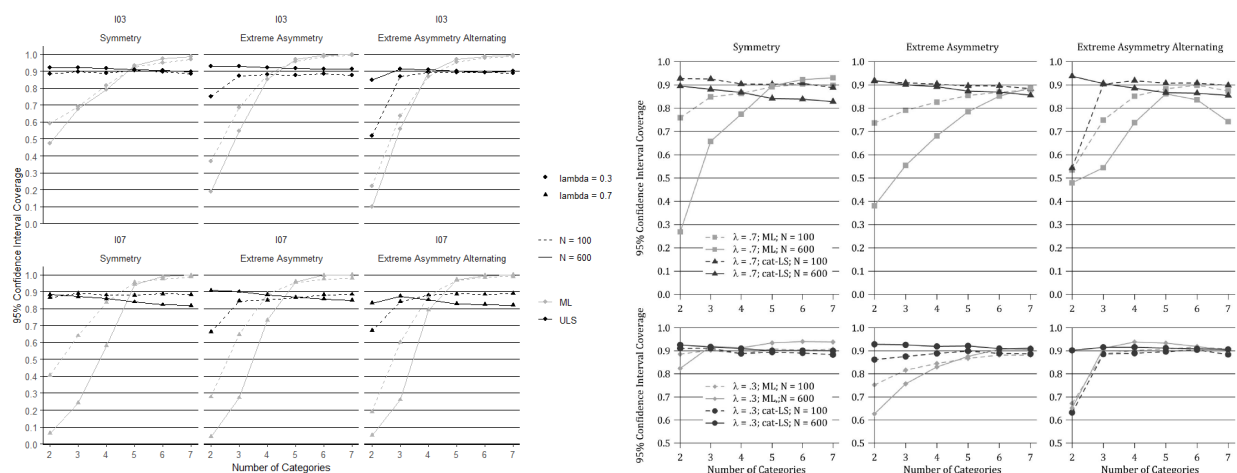


Figure 5: Coverage by number of categories (.7 and .3 factor loadings); underlying distribution is nonnormal (skew 2, kurtosis 7). Values are averaged across model size, and across all loadings for which the true parameter value was the same. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper panel represents results for a true parameter value of .7. The lower panel represents results for a true parameter value of .3. Vertical panels represent different levels of threshold symmetry. Left figure: replication; right figure original study.

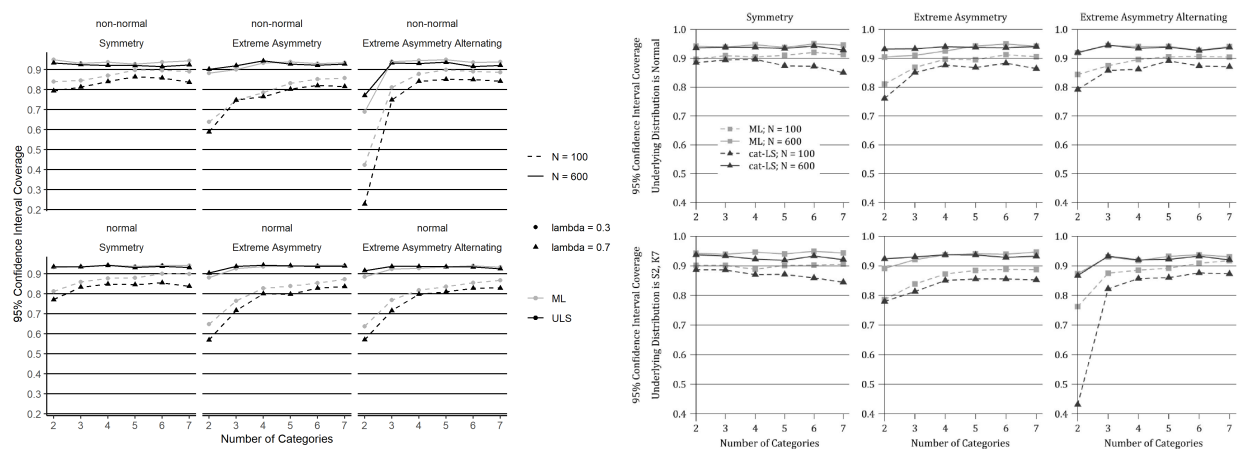


Figure 6: Coverage by number of categories (factor correlation). Values are averaged across model size. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper panel corresponds to conditions in which the underlying distribution is normal; the lower panel corresponds to conditions in which the underlying distribution is nonnormal (skew 2, kurtosis 7). Vertical panels represent different levels of threshold symmetry. Left figure: replication; right figure original study.

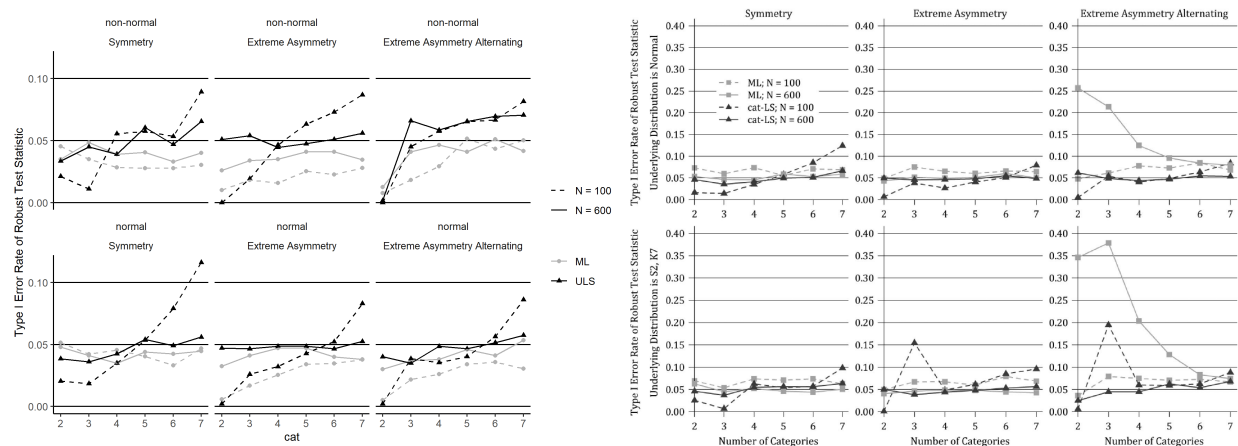


Figure 7: Type I error of mean-and-variance adjusted test statistic by number of categories. Values are averaged across model size. Lines represent different estimators and different sample sizes (see legend). ML = robust continuous maximum likelihood estimation; cat-LS = robust categorical least squares estimation. The upper panel corresponds to conditions in which the underlying distribution is normal; the lower panel corresponds to conditions in which the underlying distribution is nonnormal (skew 2, kurtosis 7). Vertical panels represent different levels of threshold symmetry.

3.2 Replication of result tables

<Compare any tabulated data to the original>

3.2.1 Table 1

Table 1 presents the “*Skew and Kurtosis of Observed Categorical Variables by Threshold Distribution, Underlying Distribution, and Number of Categories*” (p.363). The “[v]alues in this table were obtained by generating samples of size $N = 1,000,000$ for each condition and recording the skew and kurtosis of the observed distributions.” (p.363) As discussed above we understood “each condition” to only include underlying distribution, number of categories and threshold symmetry. We hence only simulated one variable of sample size 1,000,000 per condition in order to replicate figure 1, figure 2 as well as table 1.

Underlying distribution	Categories	Symmetry		Mod. Asym		Mod. Asym-Alt		Ext. Asym-Alt		Ext. Asym-Alt	
		S	K	S	K	S	K	S	K	S	K
non-normal	2	0.49	-1.76	1.11	-0.78	-0.22	-1.95	2.27	3.15	14.74	-4.09
non-normal	3	0.00	0.29	0.29	-0.96	-0.03	-0.59	1.84	1.75	0.56	-1.25
non-normal	4	0.92	-0.05	1.08	0.44	-0.13	-0.66	1.57	0.94	-0.82	-0.69
non-normal	5	0.73	-0.16	1.11	1.07	0.20	-0.80	1.38	0.47	-1.11	-0.42
non-normal	6	0.80	0.19	1.52	1.93	0.17	-0.60	1.28	0.30	-1.19	-0.26
non-normal	7	0.93	0.30	1.33	1.16	0.32	-0.39	1.26	0.37	-1.19	-0.18
normal	2	0.00	-2.00	0.59	-1.65	-0.59	-1.66	1.97	1.87	1.90	-1.98
normal	3	0.00	-0.53	0.13	-1.09	-0.13	-1.09	1.41	0.44	0.45	-1.41
normal	4	0.00	-0.53	0.69	-0.23	-0.69	-0.22	1.10	-0.25	-0.26	-1.10
normal	5	0.00	-0.47	0.59	-0.21	-0.59	-0.20	0.90	-0.59	-0.58	-0.90
normal	6	0.00	-0.43	0.62	-0.10	-0.62	-0.10	0.80	-0.69	-0.68	-0.80
normal	7	0.00	-0.41	0.52	-0.29	-0.52	-0.29	0.78	-0.62	-0.62	-0.78

Note:

Values in this table were obtained by generating samples of size N = 1,000,000 for and recording the skew and kurtosis of the observed distributions. Mod. Asym= Moderate Asymmetry; Mod.Asym-Alt = Moderate Asymmetry-Alternating; Ext.Asym = Extreme Asymmetry; Ext. Asym-Alt = Extreme Asymmetry-Alternating; S = skew; K = kurtosis

3.2.2 Observed Power (Table 2)

N	2 categories		3 categories		4 categories		5 categories		6 categories		7 categories	
	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS
100	0.398	0.408	0.602	0.667	0.713	0.806	0.769	0.890	0.809	0.927	0.849	0.955
150	0.654	0.702	0.840	0.889	0.936	0.960	0.971	0.988	0.976	0.990	0.979	0.993
350	0.994	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note:

Type I error was assessed by fitting a one-factor model to two-factor simulated data. ML = robust continuous maximum likelihood estimation; ULS = robust categorical least squares estimation.

Results regarding observed

power closely aligned with the original findings. The scenarios exhibiting a power below .8 matched the ones identified in the original study.

3.3 REplication of supplemental results

The following tables correspond to tables presented in the supplemental material of the original study which can be accessed at <http://dx.doi.org/10.1037/a0029315.supp>

3.3.1 Number of nonconverged cases per 1000 replications

Distribution	Model	cats	Ext. Asym.						Ext. Asym.-Alt						Mod. Asym.						Mod. Asym.-Alt						Symmetric											
			N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600					
			ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS						
non-normal	1	2	159	51	70	32	3	1	0	0	405	21	10	3	1	0	0	0	4	0	0	0	0	2	0	0	0	0	0	0	0	0	0					
non-normal		3	40	30	9	3	0	0	0	0	21	10	3	1	0	0	0	0	0	4	0	0	0	0	2	0	0	0	0	0	0	0	0					
non-normal		4	13	8	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	1	0	0	0	3	2	0	0	0	0	0	0	0					
non-normal		5	5	2	1	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal		6	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0					
non-normal		7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal	2	2	8	0	0	0	0	0	0	149	0	79	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal		3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal		4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal		5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal		6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
non-normal		7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
normal	1	2	193	88	72	39	3	1	0	165	74	76	49	1	1	0	0	0	33	10	5	2	0	0	0	24	8	8	1	0	0	0	23	6	5	1	0	0
normal		3	32	15	3	1	0	0	0	40	24	3	3	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	4	2	0	0	0		
normal		4	11	1	1	1	0	0	0	12	4	2	0	0	0	0	0	0	4	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0			
normal		5	6	4	1	0	0	0	0	2	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	2	1	0	0	0	0	1	0	0	0			
normal		6	1	1	0	1	0	0	0	2	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0			
normal		7	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0			
normal	2	2	16	2	0	0	0	0	0	11	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0			
normal		3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
normal		4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
normal		5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
normal		6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
normal		7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				

Note:
 Mod Asym = Moderate Asymmetry;
 Mod Asym-Alt = Moderate Asymmetry-Alternating;
 Ext Asym = Extreme Asymmetry;
 Ext Asym-Alt = Extreme Asymmetry-Alternating;
 ML = robust normal-theory maximum likelihood;
 ULS = robust categorical least squares.

3.3.2 Number of improper solutions per 1000 replications

Distribution	Model	Ext. Asym.												Ext. Asym.-Alt												Mod. Asym.												Mod. Asym.-Alt												Symmetric											
		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600																					
		ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S	ML	U.L.S																				
non-normal	1	2	377	447	175	331	8	64	0	17	704	581	606	537	248	242	70	176	122	231	38	116	0	7	0	1	62	135	8	47	0	0	1	61	144	20	61	0	1	0	0	0																			
non-normal		3	190	297	59	124	0	14	0	0	91	167	18	59	0	1	0	0	21	35	5	13	0	2	0	0	21	44	0	4	0	1	0	28	84	4	27	0	1	0	0	0																			
non-normal		4	89	119	14	46	0	2	0	0	25	51	2	8	0	0	0	0	38	49	4	8	0	0	0	0	17	44	1	11	0	0	0	26	54	2	10	0	0	0	0																				
non-normal		5	67	68	21	21	0	0	0	0	23	24	4	8	0	0	0	0	31	31	2	4	0	0	0	0	14	32	2	5	0	0	0	16	35	1	1	0	0	0	0																				
non-normal		6	51	52	6	12	0	0	0	0	15	21	0	2	0	0	0	0	40	30	6	6	0	0	0	0	10	19	2	5	0	0	0	16	17	1	2	0	0	0	0																				
non-normal		7	35	34	6	7	0	0	0	0	10	13	2	4	0	0	0	0	32	12	5	6	0	0	0	0	6	7	3	5	0	0	0	25	19	1	5	0	0	0	0																				
non-normal	2	2	27	140	3	49	0	1	0	0	365	645	217	580	17	114	0	10	2	31	0	5	0	0	0	0	7	0	1	0	0	0	9	0	0	0	0	0	0	0																					
non-normal		3	2	37	0	7	0	0	0	0	13	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0																					
non-normal		4	1	8	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0																						
non-normal		5	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																					
non-normal		6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																					
non-normal		7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																					
normal	1	2	416	468	225	330	11	63	0	6	386	474	225	354	9	61	1	8	120	170	41	79	0	6	0	0	99	170	32	74	0	4	0	90	145	25	59	0	2	0	0																				
normal		3	148	178	41	72	0	3	0	0	154	175	44	73	1	2	0	0	36	39	4	9	0	0	0	0	37	38	4	13	0	0	0	36	44	3	14	0	1	0	0																				
normal		4	75	71	7	24	0	1	0	0	77	79	15	25	0	0	0	0	30	29	6	8	0	0	0	0	31	30	6	4	0	0	0	23	26	3	7	0	0	0																					
normal		5	64	58	10	8	0	0	0	0	57	40	7	12	0	0	0	0	32	23	4	3	0	0	0	0	27	23	2	5	0	0	0	16	19	1	1	0	0	0																					
normal		6	34	29	8	9	0	0	0	0	41	29	10	7	0	0	0	0	28	18	2	2	0	0	0	0	18	17	4	4	0	0	0	9	13	2	3	0	0	0																					
normal		7	36	24	4	3	0	0	0	0	32	25	5	3	0	0	0	0	26	14	0	2	0	0	0	0	18	16	1	2	0	0	0	10	3	4	3	0	0	0																					
normal	2	2	41	113	3	35	0	1	0	0	28	116	0	31	0	0	0	0	11	0	1	0	0	0	0	0	12	0	0	0	0	0	1	5	0	0	0	0	0																						
normal		3	1	5	0	1	0	0	0	0	1	10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																					
normal		4	1	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																						
normal		5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																						
normal		6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																						
normal		7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																						

Note:

Mod Asym = Moderate Asymmetry;

Mod Asym-Alt = Moderate Asymmetry-Alternating;

Ext. Asym = Extreme Asymmetry;

Ext. Asym-Alt = Extreme Asymmetry-Alternating;

ML = robust normal-theory maximum likelihood;

U.L.S. = robust categorical least squares.

	N = 100	
	param.	cells
lambda = 3	ML	0.02
	ML	-0.02
	ML	-0.01
	ML	-0.01
	ML	-0.01
lambda = 4	ML	-0.03
	ML	-0.03
	ML	-0.03
	ML	-0.03
	ML	-0.03
lambda = 5	ML	-0.04
	ML	-0.04
	ML	-0.04
	ML	-0.04
	ML	-0.04
lambda = 6	ML	-0.05
	ML	-0.05
	ML	-0.05
	ML	-0.05
	ML	-0.05
lambda = 7	ML	-0.06
	ML	-0.06
	ML	-0.06
	ML	-0.06
	ML	-0.06
lambda = 8	ML	-0.07
	ML	-0.07
	ML	-0.07
	ML	-0.07
	ML	-0.07
lambda = 9	ML	-0.08
	ML	-0.08
	ML	-0.08
	ML	-0.08
	ML	-0.08
lambda = 10	ML	-0.09
	ML	-0.09
	ML	-0.09
	ML	-0.09
	ML	-0.09
lambda = 11	ML	-0.10
	ML	-0.10
	ML	-0.10
	ML	-0.10
	ML	-0.10
lambda = 12	ML	-0.11
	ML	-0.11
	ML	-0.11
	ML	-0.11
	ML	-0.11
lambda = 13	ML	-0.12
	ML	-0.12
	ML	-0.12
	ML	-0.12
	ML	-0.12
lambda = 14	ML	-0.13
	ML	-0.13
	ML	-0.13
	ML	-0.13
	ML	-0.13
lambda = 15	ML	-0.14
	ML	-0.14
	ML	-0.14
	ML	-0.14
	ML	-0.14
lambda = 16	ML	-0.15
	ML	-0.15
	ML	-0.15
	ML	-0.15
	ML	-0.15
lambda = 17	ML	-0.16
	ML	-0.16
	ML	-0.16
	ML	-0.16
	ML	-0.16
lambda = 18	ML	-0.17
	ML	-0.17
	ML	-0.17
	ML	-0.17
	ML	-0.17
lambda = 19	ML	-0.18
	ML	-0.18
	ML	-0.18
	ML	-0.18
	ML	-0.18
lambda = 20	ML	-0.19
	ML	-0.19
	ML	-0.19
	ML	-0.19
	ML	-0.19
lambda = 21	ML	-0.20
	ML	-0.20
	ML	-0.20
	ML	-0.20
	ML	-0.20
lambda = 22	ML	-0.21
	ML	-0.21
	ML	-0.21
	ML	-0.21
	ML	-0.21
lambda = 23	ML	-0.22
	ML	-0.22
	ML	-0.22
	ML	-0.22
	ML	-0.22
lambda = 24	ML	-0.23
	ML	-0.23
	ML	-0.23
	ML	-0.23
	ML	-0.23
lambda = 25	ML	-0.24
	ML	-0.24
	ML	-0.24
	ML	-0.24
	ML	-0.24
lambda = 26	ML	-0.25
	ML	-0.25
	ML	-0.25
	ML	-0.25
	ML	-0.25
lambda = 27	ML	-0.26
	ML	-0.26
	ML	-0.26
	ML	-0.26
	ML	-0.26
lambda = 28	ML	-0.27
	ML	-0.27
	ML	-0.27

Page 17 / 26

3.3.4 Parameter Bias, Model1, Underlying Distribution = Skew2, Kurtosis 7

param.	Ext. Asym.												Ext. Asym.-Alt												Mod. Asym.												Mod. Asym.-Alt												Symmetric											
	N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600		N = 100		N = 150		N = 350		N = 600																					
	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS	ML	ULS																				
lambda = 3	2	-0.086	-0.014	-0.098	-0.012	-0.104	0.001	-0.097	0.016	-0.078	-0.054	0.146	-0.138	0.021	-0.153	-0.019	-0.057	0.015	-0.062	0.017	-0.061	0.023	-0.060	0.025	-0.042	0.021	-0.043	0.025	-0.044	0.024	-0.041	0.027	-0.044	0.026	-0.050	0.020	-0.046	0.026	-0.046	0.025																				
	3	-0.069	-0.002	-0.069	0.010	-0.068	0.021	-0.067	0.022	-0.050	0.011	-0.045	0.022	-0.046	0.020	-0.043	0.025	-0.013	0.024	-0.009	0.027	-0.011	0.026	-0.016	0.024	-0.012	-0.012	-0.012	-0.033	0.020	-0.027	0.024	-0.026	0.024	-0.026	0.021	-0.026	0.021																						
	4	-0.065	0.012	-0.061	0.018	-0.061	0.021	-0.047	0.024	-0.018	0.023	-0.017	0.024	-0.017	0.023	-0.017	0.024	-0.017	0.023	-0.014	0.026	-0.015	0.024	-0.009	0.030	-0.010	0.026	-0.010	0.025	-0.017	0.031	-0.021	0.027	-0.022	0.025	-0.022	0.025																							
	5	-0.037	0.017	-0.038	0.019	-0.037	0.022	-0.036	0.022	-0.003	0.027	-0.005	0.024	-0.006	0.022	-0.003	0.026	-0.014	0.022	-0.012	0.022	-0.007	0.024	-0.007	0.025	-0.002	0.028	-0.003	0.026	-0.004	0.024	-0.005	0.023	-0.003	0.027	-0.006	0.024	-0.003	0.026																					
	6	-0.029	0.021	-0.028	0.022	-0.025	0.024	-0.028	0.021	0.001	0.027	0.001	0.026	0.005	0.028	0.005	0.027	0.024	0.021	-0.025	0.020	-0.019	0.024	-0.019	0.025	0.003	0.026	0.006	0.028	0.004	0.024	0.005	0.025	-0.003	0.023	0.000	0.027	0.002	0.027	0.001	0.025																			
	7	-0.020	0.024	-0.022	0.021	-0.022	0.021	-0.021	0.023	0.008	0.029	0.006	0.027	0.008	0.027	0.006	0.026	-0.019	0.019	-0.010	0.027	-0.012	0.024	-0.010	0.025	0.013	0.026	0.008	0.028	0.010	0.025	0.006	0.029	0.001	0.026	0.003	0.027	0.002	0.026																					
	8	-0.098	-0.052	-0.076	-0.009	-0.038	0.088	-0.022	0.118	-0.180	0.199	-0.164	0.134	-0.138	0.047	-0.113	0.038	-0.026	0.041	-0.007	0.065	0.022	0.128	0.021	0.127	0.010	0.064	0.025	0.110	0.042	0.131	0.040	0.129	0.028	0.087	0.050	0.116	0.063	0.127	0.064	0.128																			
lambda = 4	2	-0.068	-0.025	-0.076	-0.008	-0.038	0.088	-0.022	0.118	-0.180	0.199	-0.164	0.134	-0.138	0.047	-0.113	0.038	-0.026	0.041	-0.007	0.065	0.022	0.128	0.021	0.127	0.010	0.064	0.025	0.110	0.042	0.131	0.040	0.129	0.028	0.087	0.050	0.116	0.063	0.127	0.064	0.128																			
	3	-0.041	0.025	-0.021	0.063	0.010	0.118	0.014	0.122	-0.009	0.046	0.020	0.092	0.045	0.129	0.046	0.130	0.062	0.108	0.076	0.124	0.081	0.127	0.084	0.130	0.060	0.113	0.072	0.123	0.079	0.128	0.080	0.129	0.048	0.103	0.064	0.123	0.069	0.128	0.072	0.131																			
	4	-0.006	0.059	0.017	0.099	0.040	0.127	0.039	0.127	0.058	0.108	0.071	0.122	0.079	0.130	0.079	0.129	0.059	0.105	0.074	0.122	0.081	0.130	0.081	0.129	0.063	0.110	0.077	0.128	0.082	0.129	0.083	0.129	0.048	0.103	0.064	0.123	0.069	0.128	0.072	0.131																			
	5	0.032	0.088	0.039	0.106	0.067	0.128	0.056	0.127	0.080	0.116	0.083	0.122	0.091	0.127	0.093	0.130	0.073	0.119	0.082	0.125	0.088	0.129	0.088	0.128	0.081	0.118	0.088	0.125	0.083	0.131	0.091	0.127	0.079	0.116	0.093	0.129	0.064	0.129	0.092	0.128																			
	6	0.038	0.086	0.080	0.119	0.068	0.127	0.056	0.125	0.089	0.119	0.088	0.123	0.101	0.129	0.100	0.128	0.049	0.114	0.061	0.124	0.078	0.132	0.074	0.128	0.092	0.122	0.096	0.128	0.101	0.128	0.102	0.128	0.094	0.114	0.091	0.127	0.099	0.129	0.099	0.130																			
	7	0.061	0.107	0.065	0.120	0.074	0.127	0.078	0.130	0.097	0.123	0.096	0.123	0.103	0.127	0.103	0.128	0.065	0.118	0.079	0.123	0.086	0.128	0.088	0.130	0.098	0.118	0.107	0.131	0.107	0.128	0.109	0.129	0.079	0.113	0.093	0.125	0.099	0.130	0.097	0.126																			
	8	-0.239	-0.197	-0.214	-0.146	-0.165	-0.022	-0.144	0.016	-0.360	0.036	-0.298	-0.011	-0.277	-0.078	-0.255	-0.088	-0.159	-0.083	-0.132	-0.026	-0.100	0.024	-0.094	0.031	-0.114	-0.033	-0.095	0.005	-0.075	0.030	-0.075	0.030	-0.120	-0.026	-0.101	-0.003	-0.080	0.027	-0.077	0.032																			
lambda = 5	3	-0.173	-0.109	-0.137	-0.044	-0.102	0.019	-0.096	0.029	-0.132	-0.069	-0.097	-0.015	-0.068	0.029	-0.064	0.032	-0.043	0.031	-0.033	0.024	-0.024	0.033	-0.023	0.033	-0.049	0.010	-0.040	0.022	-0.030	0.031	-0.031	0.029	-0.093	-0.018	-0.060	0.020	-0.048	0.033	-0.049	0.030																			
	4	-0.128	-0.063	-0.098	-0.003	-0.072	0.027	-0.068	0.029	-0.053	0.006	-0.034	0.027	-0.030	0.028	-0.028	0.031	-0.060	-0.004	-0.037	0.021	-0.027	0.029	-0.025	0.030	-0.045	0.012	-0.035	0.025	-0.024	0.034	-0.025	0.032	-0.065	-0.002	-0.046	0.025	-0.039	0.030	-0.039	0.030																			
	5	-0.085	-0.014	-0.071	0.001	-0.049	0.032	-0.048	0.032	-0.032	0.009	-0.020	0.023	-0.012	0.029	-0.012	0.038	-0.034	0.019	-0.024	0.028	-0.014	0.033	-0.016	0.030	-0.031	0.015	-0.014	0.033	-0.013	0.032	-0.014	0.030	-0.033	0.012	-0.015	0.028	-0.014	0.029	-0.012	0.031																			
	6	-0.074	-0.011	-0.051	0.019	-0.038	0.027	-0.036	0.029	-0.015	0.023	-0.008	0.028	-0.004	0.031	-0.002	0.031	-0.068	0.002	-0.045	0.025	-0.031	0.031	-0.029	0.032	-0.016	0.017	-0.004	0.031	-0.002	0.031	-0.001	0.032	-0.022	0.015	-0.008	0.029	-0.005	0.032	-0.003	0.032																			
	7	-0.068	-0.008	-0.035	0.023	-0.028	0.031	-0.027	0.031	-0.018	0.012	-0.003	0.026	0.001	0.029	0.002	0.031	-0.057	0.012	-0.031	0.018	-0.018	0.030	-0.017	0.029	-0.008	0.017	0.001	0.029	0.007	0.032	0.006	0.031	-0.019	0.018	-0.004	0.033	-0.004	0.030	-0.005	0.030																			
	8	-0.285	-0.254	-0.245	-0.183	-0.186	-0.033	-0.160	0.018	-0.452	-0.044	-0.379	-0.080	-0.313	-0.122	-0.277	-0.104	-0.184	-0.107	-0.144	-0.030	-0.112	0.025	-0.113	0.026	-0.139	-0.054	-0.109	0.004	-0.089	0.031	-0.091	0.029	-0.148	-0.047	-0.122	-0.014	-0.094	0.028	-0.094	0.029																			
	lambda = 6	3	-0.166	-0.138	-0.152	-0.056	-0.112	0.018	-0.104	0.029	-0.161	-0.094	-0.115	-0.027	-0.081	0.027	-0.079	0.027	-0.057	-0.002	-0.042	0.022	-0.034	0.031	-0.034	0.030	-0.066	-0.001	-0.047	0.025	-0.044	0.027	-0.040	0.030	-0.113	-0.029	-0.075	-0.010	-0.054	0.023	-0.049	0.029																		
4		-0.148	-0.073	-0.112	-0.018	-0.078	0.026	-0.077	0.029	-0.069	-0.006	-0.046	0.018	-0.036	0.029	-0.034	0.031	-0.067	-0.011	-0.040	0.019	-0.028	0.031	-0.022	0.030	-0.041	0.012	-0.025	0.028	-0.023	0.028	-0.020	0.033	-0.040	0.004	-0.019	0.029	-0.019	0.031	-0.019	0.029																			
5		-0.096	-0.029	-0.083	0.000	-0.066	0.029	-0.057	0.029	-0.037	0.012	-0.020	0.028	-0.018	0.030	-0.017	0.038	-0.046	0.010	-0.030	0.024	-0.020	0.031	-0.022	0.030	-0.041	0.012	-0.025	0.028	-0.023	0.028	-0.020	0.033	-0.040	0.004	-0.019	0.029	-0.019	0.031	-0.019	0.029																			
6		-0.074	-0.010	-0.050	0.020	-0.040	0.027	-0.040	0.029	-0.022	0.018	-0.012	0.026	-0.008	0.030	-0.007	0.030	-0.074	0.002	-0.038	0.024	-0.039	0.027	-0.037	0.030	-0.028	0.012	-0.012	0.027	-0.008	0.029	-0.010	0.028	-0.020	0.015	-0.012	0.027	-0.009	0.032	-0.009	0.032																			
7		-0.068	-0.006	-0.043	0.019	-0.033	0.027	-0.030	0.030	-0.017	0.014	-0.005	0.027	-0.004	0.029	-0.002	0.030	-0.058	0.008	-0.028	0.024	-0.018	0.031	-0.020	0.030	-0.009	0.017	-0.005	0.025	0.004	0.033	0.003	0.031	-0.023	0.014	-0.013	0.024	-0.008	0.031	-0.007	0.031																			
8		-0.333	-0.316	-0.283	-0.234	-0.200	-0.045	-0.173	0.016	-0.483	-0.122	-0.416	-0.146	-0.359	-0.174	-0.320	-0.147	-0.215	-0.151	-0.177	-0.089	-0.127	0.022	-0.128	0.026	-0.170	-0.089	-0.133	-0.015	-0.104	0.029	-0.06	0.029	-0.171	-0.077	-0.144	-0.035	-0.110	0.025	-0.108	0.031																			
lambda = 7		3	-0.229	-0.185	-0.175	-0.088	-0.119	0.019	-0.112	0.030	-0.173	-0.116	-0.133	-0.049	-0.060																																													

4 Discussion

4.1 Replicability

Due to the high amount of details in the original publication and the corresponding supplemental materials the replication was straight forward. The largest amount of time was spent ensuring that the methods used for data generation and analysis did indeed correspond to what was used in the original study. This is, however, in no way the fault of the authors but rather due to limited documentation of the R packages used for replication. On the contrary the detailed description of the implementation allowed for a close correspondence of methodology which would have otherwise been left to guesswork.

A feature that deserves special praise with regards to facilitating replicability is the high amount of documentation that the authors dedicated to the generation of the simulated data as well as the descriptives of the same. The ability to closely monitor the data generation process and compare features of the simulated data to the original study instilled a great deal of confidence in the replicators and ensured that any potential deviations of results could not be attributed to faulty interpretation and implementation of the data generating mechanism.

Another feature that increased reproducibility was the structure of the manuscript. The very first element of the method section was an overview of the simulation factors. Readability was increased by listing each factor as a separate bullet point. Subsequent sections detailed the implementation of each simulation factor. A separate subheading for each simulation factor made it easy to locate relevant information.

The large number of result tables presented in the supplemental material is another exemplary reporting practice worth highlighting. While the comparison of hundreds of table cells is not an easy endeavor and the general interest in these tables likely limited it protects the authors against any allegations of selective reporting and makes the assessment of replicability possible.

A similar structure could be found for the performance measures which were discussed in separate subsections separated by corresponding heading. While very readable as is, we would have however preferred the performance measures to be elaborated on as part of the method section instead of the result section.

The introduction section included the presentation and discussion of several closely related methods as well as findings from previous studies investigating the same. Due to the large amount of information surrounding highly similar methods and their implementation it took us several readings of the introduction to feel confident about having identified the version actually implemented in the study at hand. A clearer separation of the implemented methods (e.g. in a box) would have facilitated isolating the relevant implementation details.

Finally, a major factor facilitating the reproduction process was the availability of specialized SEM software in the R programming environment. As R is frequently used for simulation studies investigating SEM methodology we were able to build upon a code base that was designed for this very purpose. While such

specialized software has the potential of huge time savings on the coding end and additionally is likely to minimize coding errors on the part of the replicator it consumes a significant amount of time to familiarize oneself with the exact parameters underlying the tools. The inexperienced user is at the mercy of the package documentation and the occasional peek under the hood of a given function. Having a code base from related simulation studies available would increase confidence in using such tools and avoid some trial and error while familiarizing oneself with the functionalities.

<Provide a general statement of how you experienced the replication process. Was it easy? What made it easy or difficult?>

4.2 Replicator degrees of freedom

We judge the replicator degrees of freedom in this replication to be very minimal. The only area for clarification

<Here you can discuss the replicator degrees of freedom. What could the authors have done to make it more clear? Do you think the replicator degrees of freedom are so extensive that they could influence the results?>

4.3 Equivalence of results

<How would you judge the overall equivalence of results? Are the orders of magnitude comparable? Are trends in the same direction? Would you draw the same conclusions as the authors based on your replication? Were some results not comparable because of insufficient figure resolution or labeling? Did the authors omit some results which consequently cannot be compared?>

Although our replicated results do not perfectly align with the original study's findings, the conclusions drawn by the authors largely mirror our own. Due to detailed descriptions of error frequency, we were able to detect that any scenarios with large discrepancies from the original study corresponded to scenarios with high numbers of errors.

Figure 1 and two as well as table 1 suggest that our implementation of the data generating mechanism produced identical results to the original study. Any discrepancies in results are thus likely due to differences in model estimation. Our results indicate poor performance of both estimators at low sample size and low numbers of categories. Given the large number of errors (also encountered in the original study) it would have been advisable to report Monte Carlo errors to allow a more nuanced comparison of the magnitude of discrepancies.

5 Acknowledgments

<Acknowledge the help of anyone who assisted you in the process>

6 Contributions

Authors made the following contributions according to the CRediT framework <https://casrai.org/credit/>

Anna Lohmann:

- Data Curation
- Formal Analysis (lead)
- Investigation
- Software
- Visualization (lead)
- Writing - Original Draft Preparation
- Writing - Review & Editing

Arjan Huizing:

- Formal Analysis (supporting)
- Investigation
- Software (supporting)
- Visualization (supporting)
- Validation
- Writing - Review & Editing

References

- 10 Rougier, Nicolas P., Konrad Hinsén, Frédéric Alexandre, Thomas Arildsen, Lorena A. Barba, Fabien C. Y. Benureau, C. Titus Brown, et al. 2017. "Sustainable Computational Science: The ReScience Initiative." *PeerJ Computer Science* 3 (December): e142. <https://doi.org/10.7717/peerj-cs.142>.

Appendix

Additional results

<insert additional results not reported in the original article or results presented in an alternative way>

6.1 Code organization

The code and the files associated are organized in the form of a research compendium which can be found in the following git repository <https://github.com/replisims/rhentulla-2012>

```
## .
## +-- defs.tex
## +-- figures
## |   +-- fig3.png
## |   +-- fig3_original.png
## |   +-- fig4.png
## |   +-- fig4_original.png
## |   +-- fig5.png
## |   +-- fig5_original.png
## |   +-- fig6.png
## |   +-- fig6_original.png
## |   +-- fig7.png
## |   +-- fig7_original.png
## |   +-- fig8
## |   +-- fig8_original.png
## |   +-- fig9.png
## |   +-- fig9_original.png
## |   +-- fig_3.png
## |   +-- fig_4.png
## |   +-- fig_5.png
## |   +-- fig_6.png
## |   +-- fig_8.png
## |   +-- fig_9.png
## |   +-- tabA2_A3.png
## |   +-- tabA4_A5.png
## |   +-- tabA6.png
## |   +-- tabA7.png
## |   +-- table1.png
## |   +-- table2.png
## |   +-- tableA10.html
## |   +-- tableA2_A3.html
## |   +-- tableA4_A5.html
## |   +-- tableA6.html
## |   +-- tableA7.html
## |   +-- tableA8.html
## |   +-- tableA9.html
## +-- flowchart.PNG
## +-- Lato-Black.ttf
## +-- Lato-BlackItalic.ttf
## +-- Lato-Bold.ttf
```



```
## +-- Lato-BoldItalic.ttf
## +-- Lato-Italic.ttf
## +-- Lato-Regular.ttf
## +-- references.bib
## +-- Replication Report Rhemthulla et al 2012.Rmd
## +-- Replication Report Rhemthulla et al 2012.Rmd.bak
## +-- Replication-Report-Rhemthulla-et-al-2012.log
## +-- Replication-Report-Rhemthulla-et-al-2012.pdf
## +-- Replication-Report-Rhemthulla-et-al-2012.Rmd
## +-- Replication-Report-Rhemthulla-et-al-2012.tex
## +-- UbuntuMono-Bold.ttf
## +-- UbuntuMono-BoldItalic.ttf
## +-- UbuntuMono-Italic.ttf
## \-- UbuntuMono-Regular.ttf
```

- foldername: contains <insert description>
- filename: contains <insert description>
- ...

Reproducibility Information

This report was last updated on 2022-05-24 23:34:18. The simulation replication was conducted using the following computational environment and dependencies:

```
## - Session info -----
## setting value
## version R version 4.1.3 (2022-03-10)
## os Windows 10 x64 (build 19043)
## system x86_64, mingw32
## ui RTerm
## language (EN)
## collate English_United States.1252
## ctype English_United States.1252
## tz Europe/Berlin
## date 2022-05-24
## pandoc 2.17.1.1 @ C:/Program Files/RStudio/bin/quarto/bin/ (via rmarkdown)
##
## - Packages -----
## package * version date (UTC) lib source
## assertthat 0.2.1 2019-03-21 [1] CRAN (R 4.1.2)
## cachem 1.0.6 2021-08-19 [1] CRAN (R 4.1.2)
## callr 3.7.0 2021-04-20 [1] CRAN (R 4.1.2)
## cli 3.1.0 2021-10-27 [1] CRAN (R 4.1.2)
## crayon 1.5.1 2022-03-26 [1] CRAN (R 4.1.3)
## DBI 1.1.2 2021-12-20 [1] CRAN (R 4.1.2)
## desc 1.4.1 2022-03-06 [1] CRAN (R 4.1.3)
## devtools 2.4.3 2021-11-30 [1] CRAN (R 4.1.2)
## digest 0.6.29 2021-12-01 [1] CRAN (R 4.1.2)
## dplyr * 1.0.8 2022-02-08 [1] CRAN (R 4.1.2)
## ellipsis 0.3.2 2021-04-29 [1] CRAN (R 4.1.2)
## evaluate 0.15 2022-02-18 [1] CRAN (R 4.1.3)
## fansi 1.0.3 2022-03-24 [1] CRAN (R 4.1.3)
## fastmap 1.1.0 2021-01-25 [1] CRAN (R 4.1.2)
```

```

## fs                1.5.2      2021-12-08 [1] CRAN (R 4.1.2)
## generics          0.1.2      2022-01-31 [1] CRAN (R 4.1.2)
## glue              1.6.2      2022-02-24 [1] CRAN (R 4.1.2)
## htmltools         0.5.2      2021-08-25 [1] CRAN (R 4.1.2)
## knitr              * 1.38     2022-03-25 [1] CRAN (R 4.1.3)
## lifecycle         1.0.1      2021-09-24 [1] CRAN (R 4.1.2)
## magrittr          2.0.2      2022-01-26 [1] CRAN (R 4.1.2)
## memoise           2.0.1      2021-11-26 [1] CRAN (R 4.1.2)
## pillar            1.7.0      2022-02-01 [1] CRAN (R 4.1.2)
## pkgbuild           1.3.1      2021-12-20 [1] CRAN (R 4.1.2)
## pkgconfig          2.0.3      2019-09-22 [1] CRAN (R 4.1.2)
## pkgload            1.2.4      2021-11-30 [1] CRAN (R 4.1.2)
## prettyunits        1.1.1      2020-01-24 [1] CRAN (R 4.1.2)
## processx           3.5.2      2021-04-30 [1] CRAN (R 4.1.2)
## ps                 1.6.0      2021-02-28 [1] CRAN (R 4.1.2)
## purrr              0.3.4      2020-04-17 [1] CRAN (R 4.1.2)
## R6                  2.5.1      2021-08-19 [1] CRAN (R 4.1.2)
## remotes            2.4.2      2021-11-30 [1] CRAN (R 4.1.2)
## ReplisimReport     0.0.0.9000 2022-02-03 [1] Github (replisims/ReplisimReport@5f14003)
## rlang              1.0.1      2022-02-03 [1] CRAN (R 4.1.2)
## rmarkdown          2.13       2022-03-10 [1] CRAN (R 4.1.3)
## rprojroot          2.0.2      2020-11-15 [1] CRAN (R 4.1.2)
## rstudioapi         0.13       2020-11-12 [1] CRAN (R 4.1.2)
## sessioninfo        1.2.2      2021-12-06 [1] CRAN (R 4.1.2)
## stringi            1.7.6      2021-11-29 [1] CRAN (R 4.1.2)
## stringr            1.4.0      2019-02-10 [1] CRAN (R 4.1.2)
## testthat           3.1.1      2021-12-03 [1] CRAN (R 4.1.2)
## tibble             3.1.6      2021-11-07 [1] CRAN (R 4.1.2)
## tidyselect         1.1.2      2022-02-21 [1] CRAN (R 4.1.3)
## usethis             2.1.5      2021-12-09 [1] CRAN (R 4.1.2)
## utf8               1.2.2      2021-07-24 [1] CRAN (R 4.1.2)
## vctrs              0.3.8      2021-04-29 [1] CRAN (R 4.1.3)
## withr              2.5.0      2022-03-03 [1] CRAN (R 4.1.3)
## xfun               0.30       2022-03-02 [1] CRAN (R 4.1.3)
## xtable             * 1.8-4      2019-04-21 [1] CRAN (R 4.1.2)
## yaml               2.3.5      2022-02-21 [1] CRAN (R 4.1.2)
##
## [1] C:/Users/alohmann/Documents/R/win-library/4.1
## [2] C:/Program Files/R/R-4.1.3/library
##
## -----

```

The current Git commit details are:

```

## Local:   test C:/Users/alohmann/Dropbox (Personal)/anna/projects_new/replisims/replications/rhemtulla-2012
## Remote:  test @ origin (https://github.com/replisims/rhemtulla-2012.git)
## Head:    [4d48583] 2022-05-24: Add a lot of figures

```