Monte Carlo Simulation to Study Fit Index Performances under the Varying Modeling Conditions: Comparing R and SAS Results

Dongho Choi

University of Nebraska-Lincoln

Abstract/Rationale

In this study, Monte Carlo methods were used to investigate the impact on fit indexes of varying sample size, model size, model type, and degree of model misspecification.

Monte Carlo simulation is one of statistical techniques that use pseudo-random sampling (e.g., pulling random numbers from the given distribution) to derive approximate solutions to a variety of mathematical problems.

Monte Carlo simulation offers researchers an alternative to the theoretical approach by creating empirical sampling distributions under certain assumptions. This is important because many situations exist in which implementing a theoretical approach is difficult or not possible at all (e.g., RMR, CFI, TLI have no known statistical distributions).

Furthermore, when some theoretical assumptions are violated in the actual data, the validity of the estimates is often compromised and uncertain. Because the statistical properties of most SEM estimators do not necessarily hold under such conditions, researchers must turn to simulation research to support the validity of their inferences.

This study specifically addresses the sensitivity of various fit indices to various model modeling conditions such as sample size, model size, model type, and degree of model misspecification. Previous literatures have provided the guidelines for which fit indices must be used and proposed cutoff criteria for those fit indexes in model fit assessment. According to Hu and Bentler (1998), a good index should approach its maximum under correct specification but also degrade substantially under misspecification. It also should be sensitive to model misspecification, but not to types of models.

Recently, however, many fit indices are found not only to be sensitive to model misspecification (desirable), but also be sensitive to other modeling conditions such as types of model (e.g., CFA, SR, or LGM; undesirable), model complexity, and sample sizes, thus invalidating the previously proposed cutoff criteria. Therefore, this study revisits the notion of so-called global cutoff for model fit indices and examine the performance of those indices while considering relevant factors.

The results were consistent with the findings from Fan & Sivo (2007) to the some degrees, but did not perfectly replicate them. Implications and Limitations are discussed at the end.

Method

Monte Carlo simulation was run to investigate the impact on fit indexes of varying sample size, model size, model type, and degree of model misspecification.

Dependent variable: stand-alone model fit indices (χ^2, RMSEA, RMR, SRMR, GFI, AGFI, CFI, NFI, TLI, Gamma hat).

Independent variable: model type (CFA & SR), model size (big & small), sample size(100, 200, 500, & 1000), and degree of model misspecification (True, LV1, & LV2).

The total number of cells was 48 (3 degrees of model misspecification × 2 model types × 2 levels of model complexity x 4 sample sizes). Within each cell condition, 500 replications were implemented. The total number of cases is 48 x 500 = 24000 cases.

To quantify model misspecification, a single replication was run for each of the 12 models (3 degrees of model misspecification × 2 model types × 2 levels of model complexity), using 1,000,000 cases to obtain the "population" gamma hat.

For True model, gamma = 1.

For LV1 model, gamma = 0.9686 to 0.9762 (M = 0.9724, SD = 0.0003).

For LV2 model, gamma = 0.9237 to 0.9442 (M = 0.9374, SD = 0.0077).

4-way between-factor ANOVA was conducted to determine the extent to which variation in the fit indexes across cells was attributable to the modeling conditions that were manipulated by the study design, as quantified by η^2.

the η^2 value of a fit index represents the sensitivity of the fit index to different independent variables (types of models, model misspecification, sample size, and interactions) and their interactions.

Method (cont'd)

Two program languages – SAS and R – were used for implementing data simulation, model fitting and estimation, to generate but supposedly equivalent data. Then, SAS PROC GLM was solely used for secondary data analysis. The results of analysis on generated data using SAS and R were later compared to each other.

Table 1. Package or procedure, estimation method, and optimization method used

	Package or procedure	Estimation Methods	Optimization Method
R	Lavaan MASS::mvrnorm	Maximum Likelihood (ML)	Levenberg-Marquardt Optimization
SAS	PROC FACTOR PROC CALIS PROC IML	Maximum Likelihood (ML)	non-linear minimization with box constraints (nlminb)

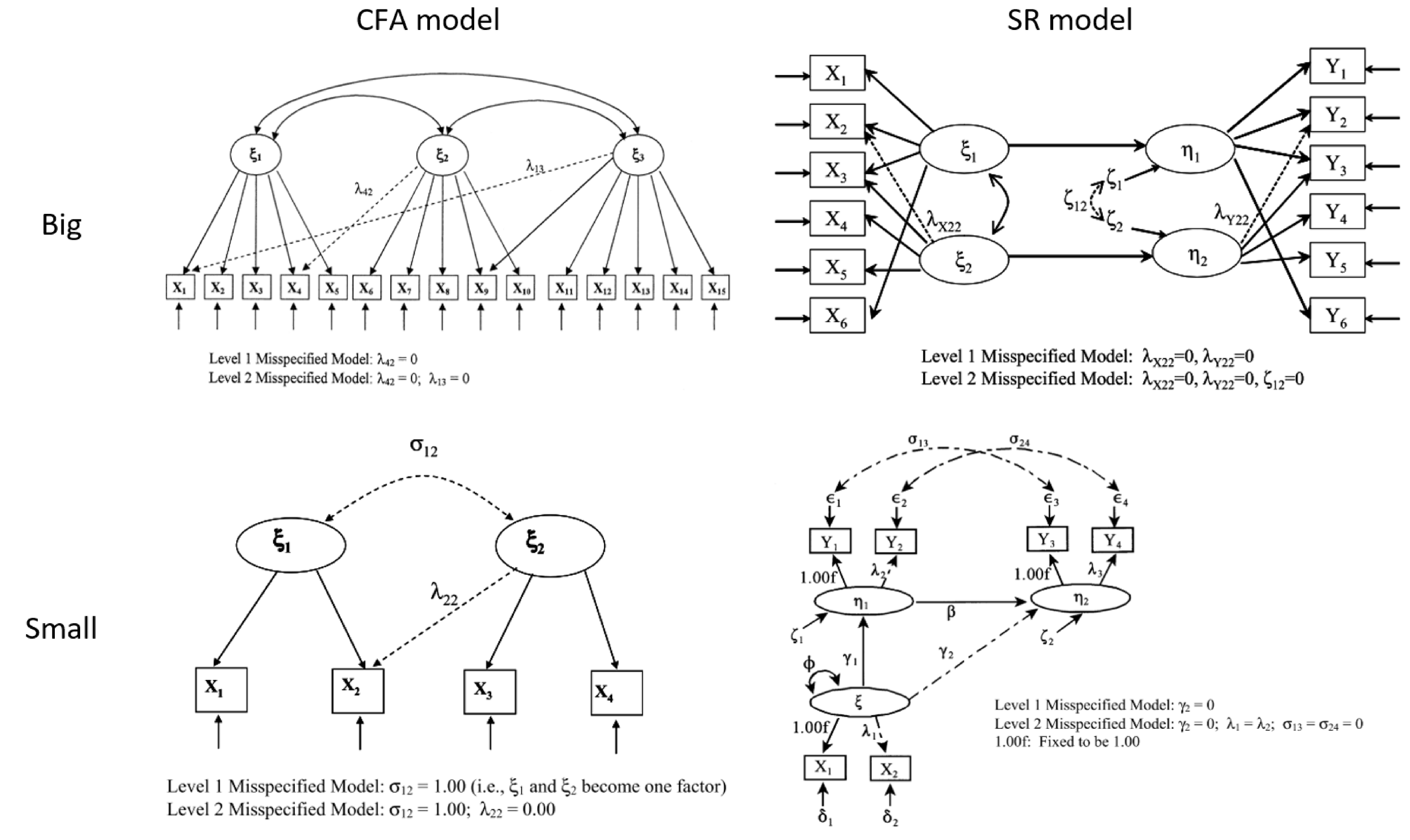

Figure 1. Models being tested

Outcomes

Results are presented in tabular and graphical form in Tables 2 & Figures 2, respectively.

Table 2. Values of semi-partial η² for the Effects of the independent Variables on the fit indices

EFFECT INDEX	χ^2		RMSEA		RMR		SRMR		GFI		AGFI		CFI		NFI		TLI		GAMMA HAT	
	SAS	R	SAS	R	SAS	R	SAS	R	SAS	R	SAS	R	SAS	R	SAS	R	SAS	R	SAS	R
MODEL TYPES (mtype)	0.0123	0.0111	0.0127	0.0120	0.3201	0.3209	0.0754	0.0647	0.0078	0.0080	0.0111	0.0113	0.0245	0.0220	0.0660	0.0642	0.0107	0.0100	0.0038	0.0028
MODEL COMPLEXITY (msize)	0.2941	0.2915	0.1974	0.1958	0.2398	0.2400	0.0142	0.0197	0.2005	0.2043	0.0178	0.0160	0.0383	0.0404	0.0178	0.0179	0.1541	0.1541	0.0016	0.0011
mtype*msize	0.0250	0.0242	0.0255	0.0269	0.2395	0.2406	0.0146	0.0087	0.0048	0.0053	0.0067	0.0064	0.0156	0.0137	0.0850	0.0830	0.0022	0.0016	0.0009	0.0014
MISSPECIFICATION (mlevel)	0.2220	0.2263	0.4708	0.4783	0.0455	0.0455	0.5053	0.5213	0.3748	0.3742	0.4292	0.4380	0.6146	0.6222	0.4113	0.4184	0.4442	0.4538	0.7433	0.7521
mtype*mlevel	0.0006	0.0006	0.0072	0.0070	0.0390	0.0389	0.0637	0.0740	0.0008	0.0007	0.0078	0.0073	0.0143	0.0147	0.0068	0.0070	0.0087	0.0087	0.0029	0.0033
msize*mlevel	0.0379	0.0376	0.0856	0.0878	0.0233	0.0231	0.1413	0.1366	0.0063	0.0065	0.0688	0.0705	0.0627	0.0609	0.0489	0.0481	0.1003	0.1027	0.0105	0.0089
mtype*msize*mlevel	0.0025	0.0029	0.0148	0.0134	0.0259	0.0257	0.0524	0.0414	0.0015	0.0010	0.0141	0.0126	0.0046	0.0049	0.0028	0.0032	0.0010	0.0008	0.0030	0.0020
SAMPLE SIZE (nsize)	0.1843	0.1846	0.0005	0.0003	0.0097	0.0090	0.0247	0.0249	0.1877	0.1850	0.1525	0.1513	0.0024	0.0016	0.1348	0.1319	0.0005	0.0001	0.0035	0.0027
mtype*nsize	0.0006	0.0003	0.0000	0.0000	0.0092	0.0085	0.0004	0.0009	0.0017	0.0017	0.0000	0.0001	0.0003	0.0003	0.0118	0.0124	0.0001	0.0001	0.0001	0.0001
msize*nsize	0.0381	0.0362	0.0007	0.0008	0.0083	0.0079	0.0052	0.0067	0.1074	0.1088	0.0284	0.0304	0.0011	0.0012	0.0558	0.0587	0.0005	0.0008	0.0026	0.0028
mtype*msize*nsize	0.0023	0.0022	0.0001	0.0000	0.0080	0.0076	0.0039	0.0039	0.0070	0.0063	0.0026	0.0021	0.0007	0.0004	0.0264	0.0248	0.0003	0.0001	0.0004	0.0002
mlevel*nsize	0.1293	0.1322	0.0049	0.0045	0.0011	0.0013	0.0025	0.0024	0.0001	0.0001	0.0012	0.0012	0.0003	0.0004	0.0002	0.0002	0.0000	0.0000	0.0007	0.0008
mtype*mlevel*nsize	0.0005	0.0004	0.0001	0.0000	0.0011	0.0012	0.0001	0.0003	0.0001	0.0001	0.0000	0.0001	0.0000	0.0001	0.0002	0.0002	0.0000	0.0001	0.0000	0.0000
msize*mlevel*nsize	0.0233	0.0228	0.0008	0.0006	0.0010	0.0011	0.0005	0.0005	0.0008	0.0007	0.0000	0.0000	0.0000	0.0000	0.0005	0.0004	0.0000	0.0000	0.0000	0.0000
mtyp*msiz*mlev*nsize	0.0015	0.0017	0.0000	0.0001	0.0010	0.0011	0.0003	0.0002	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001

(a*b = interaction between a & b)
(Underscored values are > 0.1)
(Bolded indices are also shown in graphs)

Outcomes (cont'd)

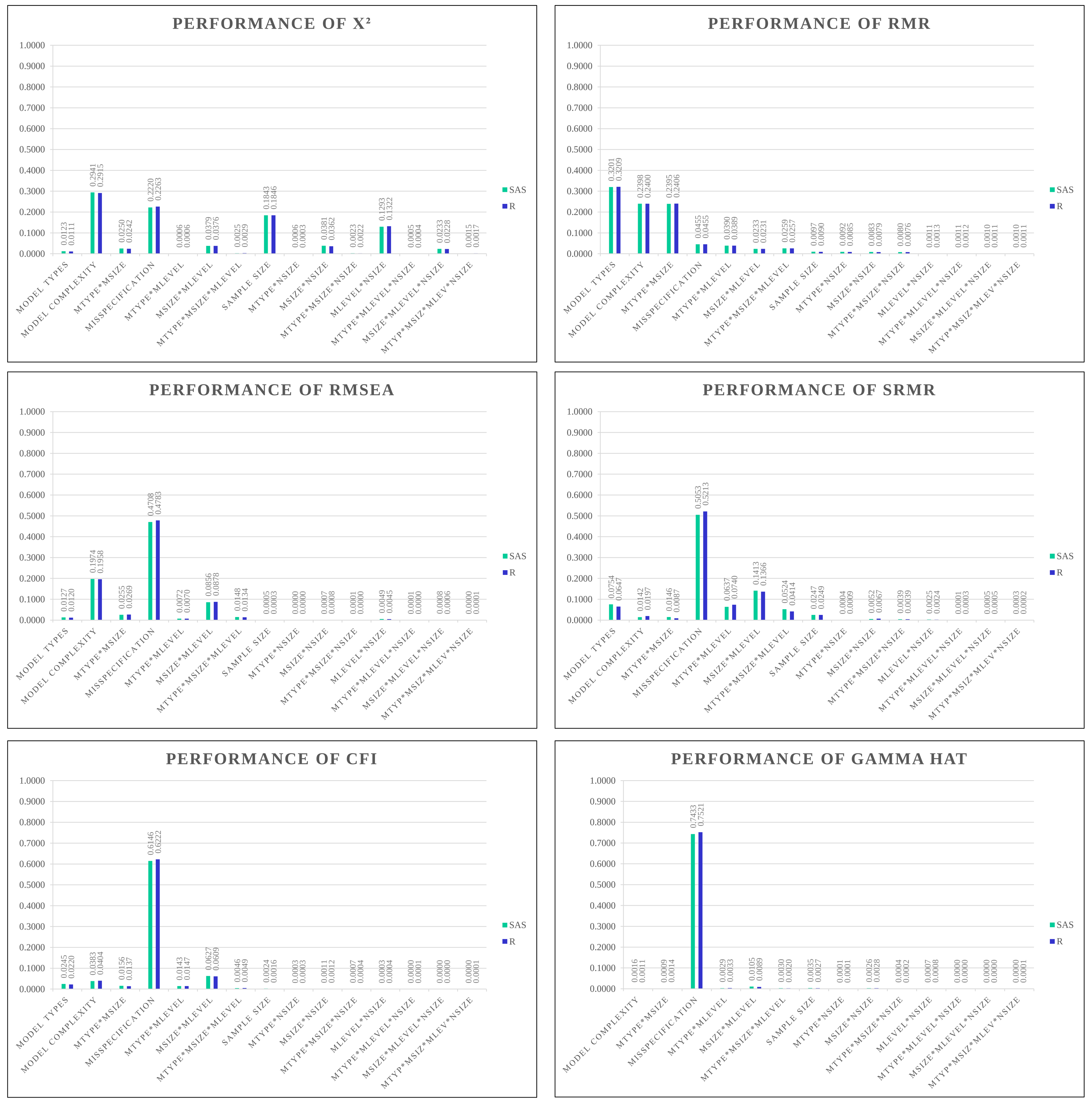

Figure 2. Values of semi-partial η² for the Effects of the independent Variables on the fit indices

Implications

Gamma hat and CFI outperformed other fit indices as a sole function of model misspecification showing robustness to other varying modeling conditions, whereas χ^2 and RMR showed the worst performance.

However, RMSEA is designed to be sensitive to the model complexity on purpose (e.g., RMSEA penalizes smaller models and rewards larger models)

Gamma hat should be used in combination with other indices (e.g., "*what does it mean that gamma hat is good but RMSEA is bad?*").

R is more convenient for writing codes & managing the errors.

Limitations

Levels of model size was not quantified.

Did not incorporate categorical variables in the model.

Did not incorporate missing data structure in the model.

Did not incorporate non-normally distributed data in the model.

Did not incorporate various models in the study (e.g., Latent Growth Model).

Did not distinguish measurement model misspecification and structural model misspecification.