# Evaluating the Sensitivity of Goodness-of-Fit Indices to Data Perturbation: An Integrated MC-SGR Approach

Massimiliano Pastore[1] and Luigi Lombardi[2]

[1] Department of Psychology
University of Cagliari
Via Is Mirrionis, 1 I-09123 Cagliari, Italy
(e-mail: `mpastore@unica.it`)
[2] Department of Cognitive Sciences and Education
University of Trento
Via Matteo del Ben, 5 I-38068 Rovereto (TN), Italy
(e-mail: `lombardi@form.unitn.it`)

**Abstract.** In this paper, we address the problem of evaluating goodness-of-fit indices in structural equation modeling when corrupted data are considered. Starting from the introduction of a new method, called MC-SGR, we evaluate the sensitivity of four different fit indices (two absolute fit-indices: GFI and AGFI, and two incremental fit-indices: CFI and NNFI) to structured perturbations.

## 1 Introduction

The issue of perturbations in real or simulated data has been substantially neglected in evaluating the adequacy of fit indices used to test covariance structure modeling. Nevertheless, it is certainly legitimate to wonder whether fit indices are reliably sensitive to data corruption. In particular, we would expect that a good index should approach its maximum under correct model specification and uncorrupted data, but also degrade substantially under massive data perturbation. In this paper we provide a possible methodological solution to the problem of evaluating the sensitivity of fit indices in structural equation modeling when perturbed data are considered. In particular, in our study the sensitivity of four different fit indices (two absolute fit-indices: GFI and AGFI, and two incremental fit-indices: CFI and NNFI) to perturbed data is examined in three different factorial models. The sensitivity evaluation is carried out by means of a new integrated approach which combines standard Monte Carlo (MC) simulations and a recent data generating procedure called Sample Generation by Replacements (SGR, [Lombardi *et al.*, 2004]).

The paper is organized as follows. Section 2 outlines the integrated MC-SGR approach. Section 3 describes the simulation study for evaluating the

goodness-of-fit indices under perturbed data scenarios. In Section 4 we discuss results of the simulation study. Finally, Section 5 reports some concluding remarks.

## 2   Integrated approach: MC + SGR

In this section we describe how to integrate the SGR procedure with MC simulations in order to evaluate the sensitivity of fit-indices in structured scenarios of data perturbation.

### 2.1   Generating data replacements: the SGR method

We think of the full dataset as being represented by an $n \times m$ matrix $\mathbf{D}$ (that is, $n$ observations, each containing $m$ elements), of which a certain portion $\mathbf{D}^c$ is actually represented by corrupted-data (corruption due to possibly fake data points in $\mathbf{D}$). The corrupted-portion $\mathbf{D}^c$ of $\mathbf{D}$ together with the uncorrupted portion $\mathbf{D}^u$ of $\mathbf{D}$, constitutes the full data set, that is to say $\mathbf{D} = \mathbf{D}^c \cup \mathbf{D}^u$. The general idea is the following: under the assumption of $\varrho \leq n \times m$ corrupted data points in $\mathbf{D}$, we replace some portions $\mathbf{D}_1, \ldots, \mathbf{D}_s$ of $\mathbf{D}$, each of which contains exactly $\varrho$ elements, with new components $\mathbf{X}_1^r, \ldots, \mathbf{X}_s^r$ in such a way that for all $h = 1, \ldots, s$, all the corresponding elements in $\mathbf{X}_h^r$ and $\mathbf{D}_h$ are different. The exact uncorrupted portion $\mathbf{D}^u$ is assumed to be unknown and only the value of $\varrho$ is supposed to be known. Moreover, all entries in $\mathbf{D}$ are also assumed to be equally likely in the process of replacements. In the SGR approach the final step consists in analyzing the complete new datasets $\mathbf{X}_1, \ldots, \mathbf{X}_s$ (with $\mathbf{X}_h = \mathbf{X}_h^r \cup \mathbf{D}_h^u$; $h = 1, \ldots, s$).

### 2.2   Extended MC simulations

Usually, in a Monte Carlo experiment, a hypothesized model is used to generate new data under various conditions. Therefore, the simulated data are used to evaluate some characteristics of the model. This, of course, implies that the distribution of the random component in the assumed model must be known, and it must be possible to generate pseudorandom samples from that distribution under the desired conditions planned by the researcher. In order to evaluate the impact of perturbed data on fit-indices we ought to generate for each MC simulated data $\mathbf{D}_k$ $(k = 1, \ldots, t)$ a family $\mathcal{R}(\mathbf{D}_k, \varrho)$ of SGR perturbed data matrices with exactly $\varrho$ replacements. Therefore, we may think of each new perturbed data $\mathbf{X} \in \mathcal{R}(\mathbf{D}_k, \varrho)$ as an alternative "informative scenario" which is directly derived from the original simulated MC sample $\mathbf{D}_k$. Next, the behavior of a target fit-index can be evaluated with respect to the perturbed samples. In this case, of course, the distributional properties of the fit-index are not those that simply hold under a particular model hypothesis (like for standard Monte Carlo simulation studies); rather

they are the properties under a model whose parameters corresponds to values fitted from both the MC generating process and the structured collection of perturbed samples that are generated from the given MC data sets.

## 3    Simulation study

In this simulation study, four fit-indices were examined with respect to structured perturbation of data. Of the four indices, two were absolute fit-indices (Goodness of Fit Index, GFI, and Adjusted Goodness of Fit Index, AGFI [Jöreskog and Sörbom, 1994]), and two incremental fit-indices (Comparative Fit Index, CFI [Bentler, 1990], and Nonnormed Fit Index, NNFI [Bentler and Bonnett, 1980] or TLI [Tucker and Lewis, 1973]). In this evaluation, three different types of target models were involved.

### 3.1    Target Models

We selected three target models that [Paxton *et al.*, 2001] considered were commonly encountered in applied research (see Figures 1, and 2). The first model, Model 1, contained nine measured variables and three latent factors. Each variable loaded on a single factor. Further, Factor 2 was regressed on Factor 1, and Factor 3 was regressed on Factor 2. The second model, Model 2, had the same basic structure as Model 1 but contained 15 measured variables, with five indicators per factor. Finally, Model 3 contained 13 measured variables with the same measurement structure as Model 1 (three indicators per factor) but added four observed exogenous variables. Factor 1 depended on all four correlated exogenous variables.

Parameter values were chosen on the basis of effect size ($R^2$ values) and statistical significance. For Model 1, the primary factor loadings were set to a standardized value of .70 (with $R^2 = .49$). The regression parameters among the latent factors were set to a standardized value of .60 ($R^2 = .36$). For Model 2, all the values were exactly the same as those of Model 1 except for the addition of two measured variables per factor. Finally, for Model 3, we included four exogenous variables. The primary factor loadings were set to .87, .82 and .72 for the first, the second and the third latent factor, respectively.

### 3.2    Simulation design

The following procedural steps were repeated for each target model $M_j$ ($j = 1, 2, 3$).

*i* ) According to $M_j$, 1000 raw-data sets $\mathbf{D}_k^j$ with $n = 50$ observations were generated. Next, each $\mathbf{D}_k^j$ ($k = 1, \ldots, 1000$) was discretized on a 5-point scale using the method described by [Jöreskog and Sörbom, 1996].
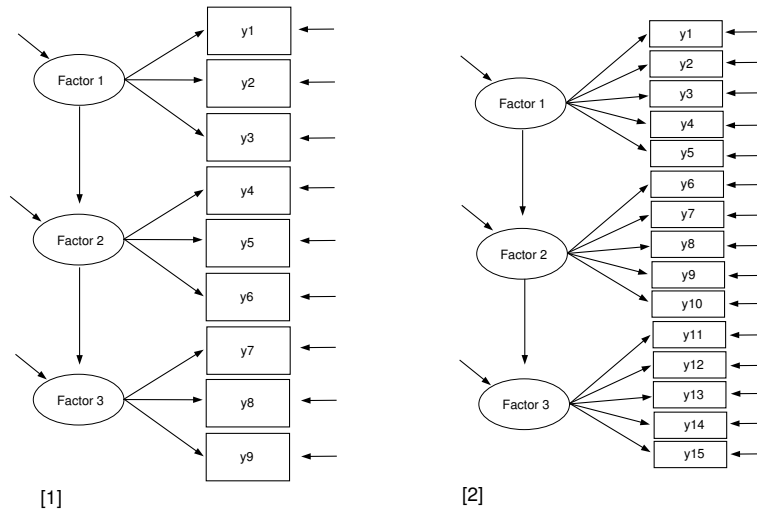
**Fig. 1.** Model [1]: nine observed variables and three factors. Model [2]: 15 observed variables and three factors.
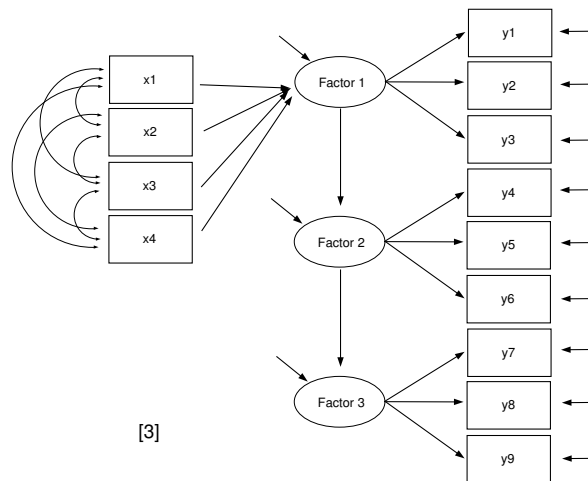


**Fig. 2.** Model [3]: 13 observed variables (four exogenous and nine endogenous) and three factors.

*ii )* For each discretized matrix $\underline{\mathbf{D}}_k^j$ we computed its polychoric correlation matrix and, subsequently, used this correlation matrix as input for $M_j$.

*iii )* Then the one hundred best fitting discretized matrix were selected by applying the following criteria: Chi-square not significant, Standardized Root-Mean-Square Residual (SRMR) $< .09$, Comparative Fit index (CFI) $> .96$, Nonnormed Fit Index (NNFI) $> .95$ [Hu and Bentler, 1999].

*iv )* For each best fitting data $\underline{\mathbf{B}}_h^j$ $(h = 1, \ldots, 100)$ we generated a family $\mathcal{R}(\underline{\mathbf{B}}_h^j, \varrho)$ of 50 SGR perturbed data matrices with exactly $\varrho$ replacements. The exact number $\varrho$ of replacements varied as a factor with 10 different levels $l = 1, 2, \ldots, 10$. Each level $l$ denoted the proportion $(l \times 10)/100$ of replacements with respect to the size of the data set.

*v )* Each perturbed data matrix $\mathbf{X} \in \mathcal{R}(\underline{\mathbf{B}}_h^j, \varrho)$ was subjected to model $M_j$ and the four fit-indices were finally evaluated. The whole procedure generated a total of 50000 new perturbed data matrices $\mathbf{X}$ for each target model.

## 4   Results

Table 1 reports the percentage of Converging Solutions (CS) and Acceptable Solutions (AS) as a function of percentage of replacements for the three considered models[1]. As expected, the percentage of CS decreased with larger percentage of replaced elements. A similar pattern was also observed for AS.

Percentage of Replacements

| model | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 92.40 | 62.82 | 28.54 | 12.16 | 7.44 | 6.68 | 7.00 | 6.52 | 6.46 | 6.80 | |
| 2 | 99.68 | 89.76 | 50.38 | 15.24 | 6.14 | 3.20 | 3.52 | 3.86 | 3.08 | 2.72 | CS |
| 3 | 79.44 | 52.42 | 24.94 | 7.58 | 2.50 | 1.92 | 1.68 | 1.62 | 1.56 | 1.76 | |
| 1 | 85.50 | 47.46 | 15.04 | 4.02 | 1.76 | 1.40 | 1.38 | 1.26 | 1.14 | 1.50 | |
| 2 | 99.36 | 83.18 | 39.20 | 9.88 | 3.06 | 1.58 | 1.60 | 1.98 | 1.38 | 1.34 | AS |
| 3 | 79.44 | 52.42 | 24.94 | 7.58 | 2.50 | 1.92 | 1.68 | 1.62 | 1.56 | 1.76 | |

**Table 1.** Percentage of Converging Solutions (CS) (resp. Acceptable Solutions (AS)) as a function of percentage of replacements.

Figure 3 shows the means of GFI and AGFI for the three models. Segments represent standard deviations[2]. Dashed lines represent the cutoff optimal value (.95). Although both indices were constantly less than .95, the GFI (resp. AGFI) mean appeared not to be affected from increasing levels of replacements. Furthermore, very surprisingly, the means of GFI and AGFI increased with larger percentage of replaced elements.

---

[1] All our analysis were based on the Maximum Likelihood estimation algorithm.

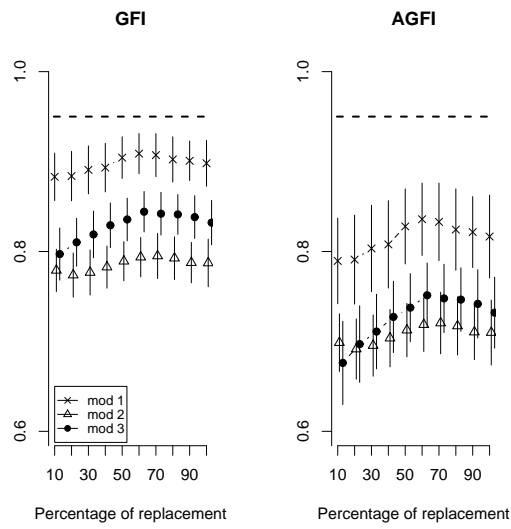[2] For the evaluation of the fit-indices we considered only AS.

**Fig. 3.** Means of GFI and AGFI as a function of percentage of replacements. Segments represent standard deviations.
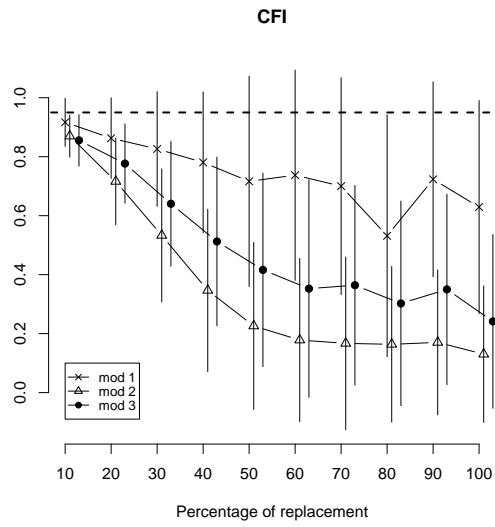
**Fig. 4.** Means of Comparative Fit Index (CFI) as a function of percentage of replacements. Segments represent standard deviations.

Figure 4 shows the means of CFI as a function of percentage of replacements for the three models. The dashed line indicates the cutoff optimal value (.96). By increasing the percentage of replacements, CFI means decreased and, in general, variability increased. The pattern associated to Model 1 showed that this model was less sensitive to replacements than both Model 3 and Model 2, the latter being the most sensitive to percentage of replacements. Notice that the same patterns were shown also by GFI and AGFI.
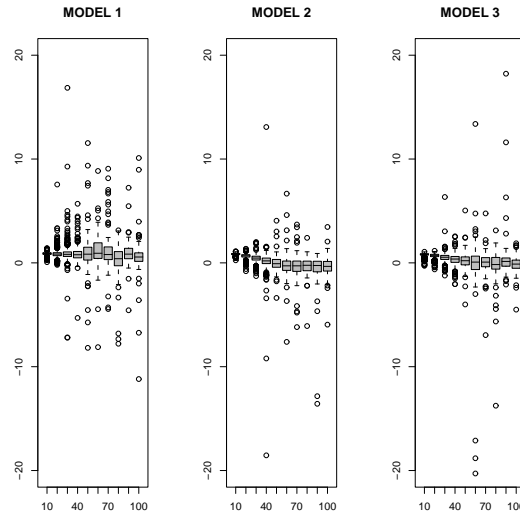


**Fig. 5.** Distributions of Nonnormed Fit Index (NNFI) as a function of percentage of replacement.

Finally, Figure 5 depicts the distributions of Nonnormed Fit Index (NNFI) for the three models. Remember that a model is a good one, when NNFI ranges between .95 and 1. Unlike both GFI and AGFI, NNFI was very sensitive to increasing levels of replacements. This observation is supported by the fact that a very large proportion of values fell outside the acceptable range [.95-1].
Table 2 reports the proportion of NNFI values within the range [.95-1]. We may notice a strong relationship between replacements and NNFI values. For example, in Model 1, we observed less than 10% of acceptable NNFI values, when 20% of replacements were considered.

## 5    Concluding remarks

A dominance relation can be read from Figures 3 and 4 as follows $M_2 \succ M_3 \succ M_1$, where $X \succ Y$ denotes that $X$ is more sensitive to perturbations

| model | Percentage of Replacements | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| 1 | 14.85 | 8.72 | 3.72 | 2.99 | 3.41 | 0.00 | 0.00 | 1.59 | 7.02 | 4.00 |
| 2 | 6.96 | 1.64 | 0.71 | 0.61 | 0.65 | 0.00 | 1.25 | 0.00 | 0.00 | 0.00 |
| 3 | 4.41 | 2.56 | 1.20 | 0.53 | 0.80 | 1.04 | 1.19 | 1.23 | 0.00 | 0.00 |

**Table 2.** Percentage of NNFI in the range [.95-1] as a function of percentage of replacements.

than $Y$. Overall our results suggested that the performance of the models were sensitive to perturbed data sets. This effect was stronger in the second model as it showed a clear replacement effect. In general, we recommend to choose more sensitive criteria (like NNFI) in order to better evaluate the effect in the model of eventual fake data.

Future applications of this methodology may be used in evaluating the robustness of goodness-of-fit criteria in empirical data set. However, more reasonable replacement scenarios based on external knowledge about process corruption should limit the upper bound of replacements. For example, in a personnel selection context the maximal number of fake answers in a personality questionnaire could be limited to 30%.

# References

[Bentler and Bonnett, 1980]P.M. Bentler and D.G. Bonnett. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, pages 588–606, 1980.

[Bentler, 1990]P.M. Bentler. Comparative fit indexes in structural models. *Psychological Bulletin*, pages 238–246, 1990.

[Hu and Bentler, 1999]L. Hu and P.M. Bentler. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, pages 1–55, 1999.

[Jöreskog and Sörbom, 1994]K.G. Jöreskog and D. Sörbom. *LISREL VI user's guide (3rd ed.)*. Scientific Software, Mooresville, IN, 1994.

[Jöreskog and Sörbom, 1996]K.G. Jöreskog and D. Sörbom. *PRELIS 2: User's reference guide*. Scientific Software, Chicago, IL, 1996.

[Lombardi *et al.*, 2004]L. Lombardi, M. Pastore, and M. Nucci. Evaluating uncertainty of model acceptability in empirical applications: A replacement approach. In K. van Montfort, J. Oud, and A. Satorra, editors, *Recent Development on Structural Equation Models*, pages 69–82. Kluwer, Dordrecht, NE, 2004.

[Paxton *et al.*, 2001]P. Paxton, P.J. Curran, K.A. Bollen, J. Kirby, and F. Chen. Monte carlo experiments: Design and implementation. *Structural Equation Modeling*, pages 287–312, 2001.

[Tucker and Lewis, 1973]L.R. Tucker and C. Lewis. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, pages 1–10, 1973.