Effect sizes of model misfit in structural equation models:
Standardized residual covariances and residual correlations

Alberto Maydeu-Olivares

Dexin Shi

University of South Carolina

Maydeu-Olivares, A., & Shi, D. (2017). Effect Sizes of Model Misfit in Structural Equation Models: Standardized Residual Covariances and Residual Correlations. *Methodology*, *13*(Supplement 1), 23–30. http://doi.org/10.1027/1614-2241/a000129

Authors note

Correspondence concerning this article should be addressed to Alberto Maydeu-Olivares, Dept. of Psychology, University of South Carolina. Barnwell College. 1512 Pendleton St. Columbia, SC 29208. This work was supported by the National Science Foundation under Grant SES-1659936, E-mail: amaydeu@sc.edu. We are indebted to Yves Rosseel for implementing these procedures in Lavaan.

Abstract

Residual correlations and covariances provide effect sizes of the misfit of covariance structure models. In a simulation study, we found that accurate CIs for standardized residual covariances are obtained even in small samples ($N = 100$), regardless of model size, degree of model misspecification, and data distribution. Standardized residual covariances also provide information about the source of misfit in poorly fitting models. From this viewpoint, they may be considered an alternative to modification indices. We compared the empirical Type I errors and power rates of standardized residual covariances and modification indices and found that both procedures provide nearly identical rates across the simulation conditions investigated.

Effect sizes of model misfit in structural equation models:
Standardized residual covariances and residual correlations

Structural equation modeling (SEM) is a general set of methods that can be applied to fit multivariate data arising from discrete or continuous outcomes. In this paper, we focus on models for continuous outcomes. Furthermore, for ease of exposition we focus on models with unconstrained means, that is, on covariance structure models. Given the multivariate nature of covariance structure models, assessing the fit of these models (i.e., the discrepancy between the data generating mechanism and the fitted model) is also a multivariate problem.

Residual covariances (i.e., the difference between the sample covariances and the covariances expected under the fitted model) provide a natural estimate of the fit of covariance structure models: the larger (in absolute value) the residual covariance, the worse the fit. However, the magnitude of a covariance is difficult to interpret and, as a result, residual covariances are ill suited as effect sizes of the misfit of a model. To overcome this problem, we can divide the residual covariances by their sample standard deviations leading to the standardized residual covariance. However, standardized residual covariances need not be in an interval from (-1, 1). From this point of view, residual correlations may be preferable to standardized residual covariances.

Population standardized residual covariances (or alternatively, residual correlations) provide standardized effect sizes of the (multivariate) misfit of a structural equations model. Confidence intervals for these population effect sizes can be used to gauge the precision with which they are estimated. Alternatively, z statistics may be performed to test the hypothesis that these population parameters are zero (or equal to some alternative, arbitrary, small value). These z statistics are simply the ratio of the estimated statistic, say a residual correlation, to its standard error. Given that so many z tests are examined simultaneously, one should control the error rate for multiple testing. The most convenient way to do so is by using a Bonferroni adjustment of the significance level using the total number of nonredundant covariances or correlations. Of course, a more precise estimate of the overall magnitude of model misfit can be obtained by combining the standardized residual covariances (or residual correlations) into an overall statistic, such as the standardized root mean squared residual (SRMR) or the correlation root mean square residual (CRMR). These overall effect sizes of the misfit can be interpreted (roughly) as the model's average standardized residual covariance and residual correlation, respectively. Of course, in applications we wish to estimate the population SRMR or CRMR and construct a confidence interval for it (Maydeu-Olivares, 2017).

Yet, it is only meaningful to examine the overall size of model misfit when there is little variability of the standardized residual covariances (or residual correlations) around the sample SRMR (or CRMR). Thus, it is important to inspect the largest values (in absolute value) of the statistically significant standardized residual covariances (or residual correlations). Although it has been repeatedly advised to examine the full matrix of residual covariances or correlations (see for instance, McDonald and Ho, (2002) –or at least the most extreme values within the matrix- this examination is seldom performed in applications. Also, SEM programs typically provide information on the statistical significance (i.e., *z* statistics) of (unstandardized) residual covariances or they provide standardized residual covariances, but not both. Yet, both should to be examined. If a residual covariance is statistically significant, we wish to examine its standardized counterpart, so that we can judge qualitatively the size of the misfit. Conversely, if a residual correlation suggests that the misfit is unacceptably large, we wish to examine its confidence interval (particularly if sample size is small) to ensure that the magnitude of the observed residual cannot be attributed to chance.

In addition to provide an effect size of the model misfit, standardized residual covariances and residual correlations may be used to assess the source of misfit in poorly fitting models. To be useful to this aim, it is important to organize the observed variables in consonance with the fitted model. For instance, if the substantive model under consideration is a two factor model with indicators x_1 to x_4 and y_1 to y_4, respectively, the covariance matrix should match this model-induced ordering of the indicators. No pattern between these residuals should be apparent; rather, the residuals should be well scattered. However, unlike score tests (known as modification indices in the SEM literature, Sörbom, 1989), standardized residual covariances do not immediately suggest how to modify the model to obtain a better fit. As a result, standardized residual covariances have not received as much attention as modification indices and little is known about their small sample behavior, as well as the behavior of their standard errors. In this article, we aim at filling this gap by performing a small simulation study to investigate the small sample behavior of standardized residual covariances (and of residual correlations) and their standard errors. In particular, we examine coverage rates for standardized residual covariances and residual correlations for different degrees of model misspecification and non-normality. In addition, we compare the empirical Type I errors and power rates of standardized residual covariances and modification indices.

The remainder of this article is organized as follows: First, we review statistical theory regarding confidence intervals for standardized residual covariances and residual correlations both under normality assumptions and under the asymptotically distribution free assumptions set forth by Browne (1982, 1984). Next, we report the results of our small simulation study. Then, we present an applied example to illustrate the use of standardized residual covariances vs. modification indices. We conclude with a general discussion and some recommendations for applied users.

Statistical theory for standardized residual covariances

Let p denote the number of observed variables being modeled, N denote sample size, and let $\boldsymbol{\sigma}$ denote a $t = p(p+1)/2$ vector of population covariances. We consider a covariance structure $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $q \leq t$ vector of parameters to be estimated from the data. The two best known procedures for estimating the model parameters involve minimizing the discrepancy functions

$$F = \left(\mathbf{s} - \boldsymbol{\sigma}_0\right)' \hat{\mathbf{W}} \left(\mathbf{s} - \boldsymbol{\sigma}_0\right) \ , \tag{1}$$

and

$$F_{ML} = \ln\left|\boldsymbol{\Sigma}_0\right| - \ln\left|\mathbf{S}\right| + \operatorname{tr}\left(\mathbf{S}\boldsymbol{\Sigma}_0^{-1}\right) - p \ , \tag{2}$$

with respect to $\boldsymbol{\theta}$. In (2), $\mathbf{S}$ and $\boldsymbol{\Sigma}_0$ denote the sample and model-implied population covariance matrices, respectively, whereas in (1), $\mathbf{s}$ is a t-vector of sample covariances and $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$, a fixed matrix. In covariance structure analysis, maximum likelihood (ML) parameter estimates under normality assumptions are frequently obtained by minimizing (2). In contrast, (1) defines a class of functions. Different choices of the weight matrix $\hat{\mathbf{W}}$ lead to different estimators such as unweighted least squares (ULS), diagonally weighted least squares (DWLS), or the asymptotically distribution free (ADF) weighted least squares (WLS) proposed by Browne (1982, 1984).

Under the null hypothesis, for any member of the class of estimators (1), as well as for the ML estimator (for details see Browne & Arminger, 1995), the covariance matrix of the estimated parameters is

$$\text{acov}(\hat{\boldsymbol{\theta}}) \overset{a}{=} N^{-1}\left(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\boldsymbol{\Delta}\left(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta}\right)^{-1} \ . \qquad (3)$$

In (3) $\boldsymbol{\Delta} = \dfrac{\partial \boldsymbol{\sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$ and $N\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of the sample covariances, **s**, i.e.,

$$\text{acov}(\mathbf{s}) = N^{-1}\boldsymbol{\Gamma} \ . \qquad (4)$$

$\boldsymbol{\Gamma}$ can be computed under normality assumptions ($\boldsymbol{\Gamma}_{NT}$) or under ADF assumptions ($\boldsymbol{\Gamma}_{ADF}$) (e.g., non-normality). For asymptotically optimal (AO) estimators, that is, for estimators with minimum variance within this class of estimators, (3) simplifies to

$$\text{acov}(\hat{\boldsymbol{\theta}}) \overset{a}{=} N^{-1}\left(\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}\right)^{-1} \ . \qquad (5)$$

Under normality assumptions, the AO estimators are the (NT, normal theory) generalized least squares (GLS) estimator and the ML estimator. Under ADF assumptions, the AO estimator is the WLS estimator.

We now turn to the residual covariances (and variances) $\mathbf{e}_u = \mathbf{s} - \hat{\boldsymbol{\sigma}}$, where $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})$, with typical element $s_{ij} - \hat{\sigma}_{ij}$. For any for the estimators considered in this paper, their asymptotic covariance matrix is

$$\text{acov}(\mathbf{e}_u) \overset{a}{=} N^{-1}\left(\mathbf{I} - \boldsymbol{\Delta}\left(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}'\mathbf{W}\right)\boldsymbol{\Gamma}\left(\mathbf{I} - \boldsymbol{\Delta}\left(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}'\mathbf{W}\right)' \ . \qquad (6)$$

Using $\boldsymbol{\Gamma}_{NT}$ or $\boldsymbol{\Gamma}_{ADF}$ in (6), one obtains standard errors for the residual covariances under normality or robust to non-normality using the square root of the diagonal elements of (6). For AO estimators, (6) simplifies to

$$\mathrm{acov}(\mathbf{e}_u) \overset{a}{=} N^{-1}\left(\mathbf{\Gamma} - \Delta\left(\Delta\mathbf{\Gamma}^{-1}\Delta\right)^{-1}\Delta'\right) = \mathrm{acov}(\mathbf{s}) - \Delta\,\mathrm{acov}(\hat{\theta})\Delta' \;. \tag{7}$$

Note that if ML estimation is used with robust standard errors, equation (6) should be used, not (7), as the ML estimator is not asymptotically optimal under ADF assumptions.

Standardized residual covariances and residual correlations

Residual covariances cannot be substantively interpreted as their magnitude depends on the units of the variables. To overcome this problem, we can divide the residual covariances by their sample standard deviations leading to the standardized residual $\frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$, or in matrix form $\mathbf{e}_s = \hat{\mathbf{G}}^{-1/2}\mathbf{e}_u = \hat{\mathbf{G}}^{-1/2}(\mathbf{s} - \hat{\sigma})$, where $\hat{\mathbf{G}} = \mathrm{diag}\left(\mathrm{vecs}(\mathbf{s}_{ii}\mathbf{s}_{ii}')\right)$ and $\mathbf{s}_{ii}$ denotes a p vector of sample variances. Instead of using standardized residual covariances, we could use the $t - p$ residual correlations $\frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} - \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} = r_{ij} - \hat{\rho}_{ij}$. In matrix form they are written as

$\mathbf{e}_r = \hat{\mathbf{G}}^{-1/2}\mathbf{s} - \hat{\mathbf{G}}_0^{-1/2}\hat{\sigma} = \mathbf{r} - \hat{\rho}$, where $\hat{\mathbf{G}}_0 = \mathrm{diag}\left(\mathrm{vecs}(\hat{\sigma}_{ii}\hat{\sigma}_{ii}')\right)$.

Asymptotic standard errors for the standardized residuals can be obtained (Ogasawara, 2001) as the square root of the diagonal elements of

$$\mathrm{acov}(\mathbf{e}_s) \overset{a}{=} \mathbf{G}^{-1/2}\,\mathrm{acov}(\mathbf{e}_u)\mathbf{G}^{-1/2} \;, \tag{8}$$

where $\mathbf{G} = \mathrm{diag}\left(\mathrm{vecs}(\sigma_{ii}\sigma_{ii}')\right)$, and $\mathrm{acov}(\mathbf{e}_u)$ is given by (6). For AO estimators only, we may use (7) instead of (6). Similarly, asymptotic standard errors for the residual correlations can be obtained as the square root of the diagonal elements of

$$\mathrm{acov}(\mathbf{e}_s) \overset{a}{=} \mathbf{F}\,\mathrm{acov}(\mathbf{e}_u)\mathbf{F} \;, \tag{9}$$

where $\mathbf{F} = \left.\dfrac{\partial(\mathbf{r} - \hat{\boldsymbol{\rho}})}{\partial \mathbf{s}'}\right|_{\mathbf{s}=\boldsymbol{\sigma}}$ (Maydeu-Olivares, 2017).

Confidence intervals and Wald (z) statistics

Letting $\boldsymbol{\sigma}$ denote the population covariances, under parameter drift assumptions, the standardized residual covariances and residual correlations are asymptotically normally distributed with mean equal to their population counterparts

$$\boldsymbol{\varepsilon}_s = \mathbf{G}^{-1/2}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_0), \qquad \text{and } \boldsymbol{\varepsilon}_r = \mathbf{G}^{-1/2}\boldsymbol{\sigma} - \mathbf{G}_0^{-1/2}\boldsymbol{\sigma}_0, \tag{10}$$

and covariance matrices given by (8) and (9). As a result, in large samples, the (100 – α)% confidence interval for ε (that is, ε_s or ε_r), can be obtained using

$$\Pr\left(\hat{\varepsilon} - z_{\alpha/2}\widehat{\mathrm{ASE}}(\hat{\varepsilon}) \le \varepsilon \le \hat{\varepsilon} + z_{\alpha/2}\widehat{\mathrm{ASE}}(\hat{\varepsilon})\right) = 1 - \alpha, \tag{11}$$

where $\hat{\varepsilon} = e$ denotes the estimate of the effect size (e.g., the sample standardized residual covariance) and ASE denotes its asymptotic standard error. Similarly, we can test whether a population standardized residual covariance or residual correlation equals some cut-off criteria *c* (usual zero) using

$$z = \frac{\hat{\varepsilon} - c}{\widehat{\mathrm{ASE}}(\hat{\varepsilon})}. \tag{12}$$

Note that the *z* statistics for standardized residual covariances must equal the *z* statistics for unstandardized residual covariances. This is because $\hat{\varepsilon}_s = \dfrac{\hat{\varepsilon}_u}{\sqrt{s_{ii}s_{jj}}}$ and $\widehat{\mathrm{ASE}}(\hat{\varepsilon}_s) = \dfrac{\sqrt{\mathrm{var}(\hat{\varepsilon}_u)}}{\sqrt{s_{ii}s_{jj}}}$.

Therefore, $z_s = \dfrac{\hat{\varepsilon}_u}{\widehat{\mathrm{ASE}}(\hat{\varepsilon}_u)} = z_u$. Similarly, in well-fitting models, the variances estimated under the model will be close to the sample variances, or $\hat{\boldsymbol{\sigma}}_{ii} \approx \mathbf{s}_{ii}$. In such cases $\hat{\mathbf{G}}^{-1/2} \approx \hat{\mathbf{G}}_0^{-1/2}$, and the residual correlations $\mathbf{e}_r = \hat{\mathbf{G}}^{-1/2}\mathbf{s} - \hat{\mathbf{G}}_0^{-1/2}\hat{\boldsymbol{\sigma}}$ will be very similar to the standardized residual covariances $\mathbf{e}_s = \hat{\mathbf{G}}^{-1/2}\mathbf{s} - \hat{\mathbf{G}}^{-1/2}\hat{\boldsymbol{\sigma}} = \hat{\mathbf{G}}^{-1/2}(\mathbf{s} - \hat{\boldsymbol{\sigma}})$ and their z statistics will be very similar as well.

A Monte-Carlo comparison of standardized residual covariances, correlation residuals, and modification indices

To investigate the coverage rates of standardized residual covariances and residual correlations, we performed a small simulation study with 16 conditions. The conditions were obtained by crossing

a) sample size (small, *N* = 100; large, *N* = 1000),

b) model size (small, *p* = 10 observed variables; large, *p* = 30),

c) model misspecification (correctly specified, and misspecified)

d) distribution of the data (normal, non-normal).

For each condition 1000 replications were obtained.

More specifically, we generated multivariate normal data with mean zero and a covariance structure conforming to an independent clusters two factor model with factor loadings equal to 0.7 and uniqueness equal to 0.51 for all variables. The data was then discretized into 7 categories coded 0 to 6 where the thresholds were selected so that the observed data had the desired skewness and kurtosis. A one factor model was fitted in all cases using ML estimation using the Lavaan package in R (Rosseel, 2012). Standard errors were computed under normality assumptions, and also under ADF assumptions (see Satorra & Bentler, 1994).

The misspecified model was obtained by setting the correlation between the factors to ρ = 0.8 when generating the data. The correctly specified model was obtained by setting the correlation to ρ = 1. The threshold values used for all variables in the normal and non-normal condition are provided in Table 1. These thresholds yield a population skewness of 0 and (excess) kurtosis of 0 for the normal condition, and of -2 and 3, respectively, for the non-normal conditions. Table 1 also contains the population Root Mean Square Error of Approximation (RMSEA: Browne & Cudeck, 1993), population Standardized Root Mean squared Residual (SRMR) and Correlation Root Mean squared Residual (CRMR). The computation of these population quantities is described in Maydeu-Olivares (2017). The computation of the population skewness and kurtosis for discretized normal variables is described in Maydeu-Olivares, Coffman and Hartmann (2007).

Insert Table 1 about here

We see in Table 1 that the population RMSEA values for misspecified models range from 0.046 to 0.075; population SRMR values range from 0.039 to 0.042, and population CRMR values range from 0.043 to 0.046. Note that these population values are smaller for non-normal data than for normal data.

We provide in Table 2 the empirical Type I error rates (for correctly specified models) and empirical power rates (for misspecified models) of z statistics (12) for testing whether the population standardized residual covariances and residual correlations equal zero (both under normality and ADF assumptions). We also provide in this table the empirical Type errors and power rates for modification indices for comparison.

Now, the fitted model is a one factor model, $\mathbf{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{\Theta}$, where $\boldsymbol{\lambda}$ denotes the p vector of factor loadings and $\mathbf{\Theta}$ denotes the $p \times p$ covariance matrix among the unique errors in the factor model. For this choice of fitted model, only modification indices for the off-diagonal elements of $\mathbf{\Theta}$ are computed. For each model, there are $t - q$ modification indices, $t - q$ z

statistics for residual correlations; we used only the $t - q$ z statistics for residual covariances (we excluded the tests for residual variances). For each replication and choice of test statistic, we computed the minimum, maximum and median *p*-value across the $t - q$ tests of the population parameter being zero. Table 2 provides the average of these minimum, maximum and medians across the 1000 replications of each condition. For correctly specified models ($\rho = 1$), rejection rates should be as close as possible to the nominal rate (0.05); for incorrectly specified models ($\rho = 0.8$), rejection rates should be as large as possible.

We see in Table 2 that rejection rates for modification indices, standardized residual covariances and residual correlations are very similar. When the statistics are computed assuming normality and the data is actually non-normal, all three statistics lack power to detect that the model is misspecified (*p* values are too large). In all other conditions, rejection rates are right on target across all conditions, even at the smallest sample size ($N = 100$) and largest model ($p = 30$) considered.

In examining rejection rates for the misspecified model, results obtained under normality assumptions when the data is non-normal should be ignored, as empirical Type I errors are incorrect. Inspecting the other conditions, we see that a) power is very similar across the three statistics considered (modification indices, *z* statistics for residual covariances, and *z* statistics for residual correlations), b) power logically increases for increasing sample size, c) power does not increase for increasing model size, and d) power does not increase when data is normally distributed and normality is assumed. Since there appears to be no gain in assuming normality, combining the Type I and power results we conclude that it is preferable to use ADF assumptions.

Insert Tables 2 and 3 about here

Table 3 provides 95% coverage rates for the standardized residual covariances and residual correlations. Population values ranged from 0.041 to 0.051 (in absolute value) in both cases. Consistent with the results presented for the test statistics, we see in this Table that the coverage rates are rather accurate, across all conditions of model size, normality, sample size, and model misspecification. The exception is confidence intervals computed under normality assumptions when the data is non-normal. In this case, confidence intervals are too narrow: median coverage rate is 0.86 for 95% intervals.

An example: Modeling the LOT

The Life Orientation Test (LOT: Scheier & Carver, 1985) consists of eight items designed to measure generalized outcome expectancies plus four filler items. The response scale for the items is graded, consisting of five points (0,…, 4). Four of the items are positively worded, while the remaining items are negatively worded.

The LOT was designed to measure a single construct. However, several factor analytic studies (e.g., Chang, D'Zurilla, & Maydeu-Olivares, 1994; Chang & McBride-Chang, 1996; Marshall, Wortman, Kusulas, Hervig, & Vickers, 1992; Robinson-Whelen, Kim, MacCallum, & Kiecolt-Glaser, 1997; Scheier & Carver, 1985) have revealed that a one-factor model does not fit well. Instead, a correlated two-factor model in which all positively worded items load on one factor and all negatively worded items load on another factor provides a substantially better fit to the data (but see Maydeu-Olivares & Coffman, 2006).

Here, we fitted a one factor model by ML to data gathered from 389 respondents by Chang et al. (1994). The model does not fit well. Under normality assumptions the 90%

confidence interval (CI) for the RMSEA is (.15; .19); for the SRMR it is (.08;.12), and for the CRMR it is (.09, .12). These data is quite normal: Excess kurtosis ranges from -0.72 to 0.01; skewness ranges from -0.57 to 0.71. As a result CIs for the RMSEA, SRMR and CRMR computed under ADF assumptions equal those computed under normality to three decimals. The CI for the RMSEA robust to non-normality was computed using the mean adjusted RMSEA as described in Brosseau-Liard, Savalei and Li (2012).

Next, we can examine the standardized residual covariances or residual correlations to gauge the extent of the misfit. In this example residual covariances and residual correlations (and their z statistics) are equal up to two decimals. As a result, only standardized residual covariances are reported and they are shown in Table 4. Residual variances in this example are zero and they have not been printed in this table. In addition, we have ordered the LOT items as {1, 4, 5, 11, 3, 8, 9, 12}. The first four items are positively worded (they reflect optimism), and the last four are negatively worded (they reflect pessimism). We see in this Table that residuals among the negatively worded items are small. Residuals are also small between positively and negatively worded items. Yet, they are large among the positively worded items. Thus, after suitably reordering the items according to a substantive theory with two factors (optimism and pessimism), the pattern we observe in this table is the typical pattern we would observe if a two factor model is the data generating model but a one factor model is fitted.

In inspecting standardized residual covariances it is important to take into account their sampling variability. One way to do so is to indicate, as we have done in Table 4, which of these effect sizes are statistically significant. More specifically, we have boldfaced in this table the standardized residual covariances that are statistically significant at the $\alpha = 0.05$ level using a Bonferroni correction. That is, since there are $t = 28$ residual covariances, we used $\alpha = 0.05/28 =$

0.0018. Of course, a more precise procedure, such as the Benjamini-Hochberg procedure (see Thissen, Steinberg, & Kuang, 2002) could have been used.

Alternatively, confidence intervals for the standardized residual covariances can be constructed. We display in Table 5 90% confidence intervals for these population parameters. Do modification indices provide a similar picture in this application? To address this question, we provide in Table 6 the modification indices for this model both under normality and non-normality assumptions. Comparing Tables 4 and 6 we see that modification indices and standardized residual covariances provide very much the same picture for these data, both under normality and under ADF assumptions. Under normality, both sets of statistics suggests that there is there is some misfit associated to item 9, in addition to the block misfit due to the misspecification of the latent trait dimensionality. Standardized residual covariances immediately convey the magnitude of the misfit: it is not large. Furthermore, these statistics fail to be significant when the precision with which they are estimated is computed under ADF assumptions.

Insert Tables 4 to 6 about here

Discussion and conclusions

Bivariate standardized effect sizes (standardized residual covariances, and residual correlations) may be a helpful tool to detect the source of mispecified covariance structure models. In factor analysis applications, after suitably ordering the observed variables, we look for blocks of residuals as an indication of mispecified latent trait dimensionality, as well as for rows of residuals as an indication of mispecified secondary loadings. Factor analysis is often used in test construction. Within this context it is often of interest to remove poor fitting indicators. To this aim, we find helpful printing the average of the absolute value of the residual correlations (or standardized residual covariances) for each item (Maydeu-Olivares, 2015). It is also possible to use residuals to detect the source of misfit in path analysis models, but see Costner and Schoenberg (1978). Admittedly, when the assumptions underlying the use of modification indices are met (misfit caused by a few omitted paths), one should expect modification indices to be more helpful than residuals in suggesting ways to improve the fit of the model. Otherwise, as the example presented here illustrates, modification indices and residuals provide similar information.

Standardized residual covariances (or residual correlations) also provide effect sizes of the misfit of the model. This is important, as in models involving a large number of variables one often needs to settle for approximate models due to time constraints. Also, when large sample sizes are used, well-fitting models may be rejected due to high power. In both cases, these standardized effect sizes need to be examined to gauge qualitatively the extent of the misfit. The use of overall effect sizes of the model's misfit of such as the SRMR should not be used in the presence of large standardized residual covariances. Of course, modification indices can also be used to gauge the effect size of the misfit (via the use of expected parameter change, see Saris, Satorra, & van der Veld, 2009). We have shown in this paper that modification indices and residuals show similar Type I and II errors. However, often we find the use of standardized residual covariances more convenient than modification indices and expected parameter changes to convey the magnitude of a model's misfit.

In closing, we have demonstrated that, under the conditions investigated, standardized residual covariances and residual correlations display a similar behavior. However, the conditions investigated here involved scale invariant models. Further research involving models that are not scale invariant is needed in helping choose between both sets of statistics.

References

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An Investigation of the Sample Performance of Two Nonnormality Corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. http://doi.org/10.1080/00273171.2012.715252

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. http://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean-and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York: Plenum.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. s. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chang, E. C., D'Zurilla, T. J., & Maydeu-Olivares, A. (1994). Assessing the dimensionality of optimism and pessimism using a multimeasure approach. *Cognitive Therapy and Research*, *18*(2), 143–160. http://doi.org/10.1007/BF02357221

Chang, L., & McBride-Chang, C. (1996). The Factor Structure of the Life Orientation Test. *Educational and Psychological Measurement*, *56*(2), 325–329. http://doi.org/10.1177/0013164496056002013

Costner, H., & Schoenberg, R. (1978). Diagnosing indicator ills in multiple indicator models. In

A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (Vol. 9, pp. 167–199). New York: Seminar Press.

Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hervig, L. K., & Vickers, R. R. (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of Personality and Social Psychology*, *62*(6), 1067–1074. http://doi.org/10.1037/0022-3514.62.6.1067

Maydeu-Olivares, A. (2015). Evaluating fit in IRT models. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (pp. 111–127). New York: Routledge.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*.

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, *11*(4), 344–62. http://doi.org/10.1037/1082-989X.11.4.344

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, *12*(2), 157–176. http://doi.org/10.1037/1082-989X.12.4.433

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. http://doi.org/10.1037//1082-989X.7.1.64

Ogasawara, H. (2001). Standard errors of fit indices using residuals in structural equation modeling. *Psychometrika*, *66*(3), 421–436. http://doi.org/10.1007/BF02294443

Robinson-Whelen, S., Kim, C., MacCallum, R. C., & Kiecolt-Glaser, J. K. (1997). Distinguishing optimism from pessimism in older adults: Is it more important to be optimistic or not to be pessimistic? *Journal of Personality and Social Psychology*, *73*(6), 1345–1353. http://doi.org/10.1037/0022-3514.73.6.1345

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://doi.org/10.18637/jss.v048.i02

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. http://doi.org/10.1080/10705510903203433

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*(3), 219–247. http://doi.org/10.1037/0278-6133.4.3.219

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*(3), 371–384. http://doi.org/10.1007/BF02294623

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *Journal of Educational and Behavioral Statistics*, *27*(1), 77–83. http://doi.org/10.3102/10769986027001077

Table 1

Population thresholds, skewness, kurtosis, RMSEA, SRMR, and CRMR

p	ρ	*Item thresholds*	*Kur.*	*Skew.*	*RMSEA*	*SRMR*	*CRMR*
10	0.8	-1.64, -1.08, -.52, .52, 1.08, 1.64	0	0	0.075	0.041	0.046
10	0.8	-2.33, -1.88, -1.55, -1.17, -.84, -.55	3.2	-2.0	0.063	0.039	0.043
30	0.8	-1.64, -1.08, -.52, .52, 1.08, 1.64	0	0	0.054	0.044	0.046
30	0.8	-2.33, -1.88, -1.55, -1.17, -.84, -.55	3.2	-2.0	0.046	0.042	0.043

Note: RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean squared Residual, CRMR = Correlation Root Mean squared Residual (CRMR). For correctly specified models (ρ = 1) the same threholds were used, leading to the same values of kurtosis and skewness. In this case, however, the population RMSEA, SRMR, and CRMR are all zero.

Table 2

Empirical rejection rates at $\alpha = 0.05$ for modification indices, standardized residual covariances, and residual correlations

					Modification Indices						Standardized Residual Covariances						Residual Correlations					
p	*N*	ρ	Kur.	Skew.	NT			ADF			NT			ADF			NT			ADF		
					min	max	med	min	max	med	min	max	med	min	max	med	min	max	med	min	max	med
10	100	1	0	0	0.04	0.07	0.05	0.04	0.07	0.05	0.04	0.07	0.05	0.04	0.08	0.06	0.04	0.07	0.05	0.04	0.08	0.06
10	100	1	3.2	-2	0.13	0.18	0.15	0.04	0.08	0.06	0.12	0.17	0.14	0.04	0.08	0.06	0.12	0.17	0.14	0.04	0.08	0.06
10	1000	1	0	0	0.04	0.07	0.05	0.04	0.07	0.05	0.04	0.07	0.05	0.04	0.07	0.05	0.04	0.07	0.05	0.04	0.07	0.05
10	1000	1	3.2	-2	0.13	0.17	0.15	0.03	0.06	0.05	0.13	0.17	0.15	0.04	0.07	0.05	0.13	0.17	0.15	0.04	0.07	0.05
30	100	1	0	0	0.03	0.08	0.05	0.03	0.08	0.05	0.03	0.07	0.05	0.03	0.08	0.06	0.03	0.07	0.05	0.03	0.08	0.06
30	100	1	3.2	-2	0.12	0.18	0.15	0.03	0.08	0.06	0.11	0.17	0.14	0.04	0.09	0.06	0.11	0.17	0.14	0.04	0.09	0.06
30	1000	1	0	0	0.03	0.07	0.05	0.03	0.07	0.05	0.03	0.07	0.05	0.03	0.07	0.05	0.03	0.07	0.05	0.03	0.07	0.05
30	1000	1	3.2	-2	0.12	0.18	0.15	0.03	0.07	0.05	0.12	0.18	0.14	0.03	0.07	0.05	0.12	0.18	0.14	0.03	0.07	0.05
10	100	0.8	0	0	0.1	0.19	0.13	0.1	0.19	0.13	0.11	0.15	0.13	0.12	0.17	0.14	0.11	0.15	0.13	0.12	0.17	0.14
10	100	0.8	3.2	-2	0.14	0.26	0.18	0.04	0.14	0.08	0.14	0.23	0.18	0.06	0.13	0.1	0.14	0.23	0.18	0.06	0.13	0.1
10	1000	0.8	0	0	0.67	0.89	0.72	0.66	0.88	0.71	0.68	0.88	0.73	0.68	0.88	0.72	0.68	0.88	0.73	0.68	0.88	0.72
10	1000	0.8	3.2	-2	0.5	0.7	0.55	0.31	0.52	0.35	0.51	0.69	0.56	0.34	0.47	0.4	0.51	0.69	0.56	0.34	0.47	0.4
30	100	0.8	0	0	0.1	0.17	0.13	0.1	0.17	0.13	0.1	0.15	0.12	0.1	0.17	0.13	0.1	0.15	0.12	0.1	0.17	0.13
30	100	0.8	3.2	-2	0.14	0.23	0.19	0.05	0.12	0.08	0.14	0.22	0.18	0.05	0.14	0.09	0.14	0.22	0.18	0.05	0.14	0.09
30	1000	0.8	0	0	0.66	0.79	0.72	0.66	0.79	0.72	0.66	0.79	0.72	0.66	0.78	0.72	0.66	0.79	0.72	0.66	0.78	0.72
30	1000	0.8	3.2	-2	0.48	0.61	0.55	0.29	0.42	0.36	0.49	0.6	0.54	0.31	0.4	0.36	0.49	0.6	0.54	0.31	0.4	0.36

Notes: *p* = number of observed variables; *N* = Sample Size; ρ = population inter-factor correlation; Kur.= excess kurtosis; Skew.= skewness; NT = normality assumptions; ADF = asymptotically distribution free assumptions; med = median;

Table 3

95% coverage rates for standardized residual covariances and residual correlations

p	N	ρ	Kur.	Skew.	Standardized Residual Covariances						Residual Correlations					
					NT			ADF			NT			ADF		
					min	max	med	min	max	med	min	max	med	min	max	med
10	100	1	0	0	0.93	0.96	0.95	0.92	0.96	0.94	0.93	0.96	0.95	0.92	0.96	0.94
10	100	1	3.2	-2	0.83	0.88	0.86	0.92	0.96	0.94	0.83	0.88	0.86	0.92	0.96	0.94
10	1000	1	0	0	0.93	0.96	0.95	0.93	0.96	0.95	0.93	0.96	0.95	0.93	0.96	0.95
10	1000	1	3.2	-2	0.83	0.87	0.85	0.93	0.96	0.95	0.83	0.87	0.85	0.93	0.96	0.95
30	100	1	0	0	0.93	0.97	0.95	0.92	0.97	0.95	0.93	0.97	0.95	0.92	0.97	0.95
30	100	1	3.2	-2	0.83	0.89	0.86	0.91	0.96	0.94	0.83	0.89	0.86	0.91	0.96	0.94
30	1000	1	0	0	0.93	0.97	0.95	0.93	0.97	0.95	0.93	0.97	0.95	0.93	0.97	0.95
30	1000	1	3.2	-2	0.82	0.88	0.86	0.93	0.97	0.95	0.82	0.88	0.86	0.93	0.97	0.95
10	100	0.8	0	0	0.93	0.96	0.95	0.93	0.96	0.94	0.93	0.96	0.95	0.93	0.96	0.94
10	100	0.8	3.2	-2	0.83	0.9	0.86	0.91	0.96	0.94	0.83	0.9	0.86	0.91	0.96	0.94
10	1000	0.8	0	0	0.92	0.97	0.95	0.92	0.96	0.95	0.92	0.97	0.95	0.92	0.96	0.95
10	1000	0.8	3.2	-2	0.83	0.89	0.86	0.93	0.96	0.95	0.83	0.89	0.86	0.93	0.96	0.95
30	100	0.8	0	0	0.93	0.97	0.95	0.92	0.97	0.94	0.93	0.97	0.95	0.92	0.97	0.94
30	100	0.8	3.2	-2	0.83	0.91	0.87	0.92	0.96	0.94	0.83	0.91	0.87	0.92	0.96	0.94
30	1000	0.8	0	0	0.93	0.97	0.95	0.93	0.97	0.95	0.93	0.97	0.95	0.93	0.97	0.95
30	1000	0.8	3.2	-2	0.82	0.9	0.86	0.93	0.97	0.95	0.82	0.9	0.86	0.93	0.97	0.95

Notes: p = number of observed variables; N = Sample Size; ρ = population inter-factor correlation; Kur.= excess kurtosis; Skew.= skewness; NT = normality assumptions; ADF = asymptotically distribution free assumptions

Table 4

Results for a one factor model applied to the Life Orientation Test (LOT) data: standardized residual covariances

Item	**1**	**4**	**5**	**11**	**3**	**8**	**9**	**12**
1		**0.31**	**0.25**	0.13	-0.08	-0.05	-0.07	-0.05
4	**0.31**		**0.29**	**0.19**	-0.08	-0.04	**-0.11**	0.01
5	**0.25**	**0.29**		0.08	-0.03	-0.06	-0.07	-0.02
11	0.13	**0.19**	0.08		-0.08	**-0.06**	-0.02	0.05
3	-0.08	-0.08	-0.03	-0.08		0.01	0.03	0.04
8	-0.05	-0.04	**-0.06**	-0.06	0.01		0.05	-0.01
9	**-0.07**	**-0.11**	**-0.07**	-0.02	0.03	**0.05**		0.00
12	-0.05	0.01	-0.02	0.05	0.04	-0.01	0.00	

Note: We have boldfaced the standardized residual covariances that are statistically significant at the 5% level after applying a Bonferroni correction. Above the diagonal we have boldfaced statistics significant under asymptotically distribution free (ADF) assumptions; below the diagonal, statistics significant under normality (NT) assumptions.

Table 5

Results for a one factor model applied to the Life Orientation Test (LOT) data: 90% confidence intervals for standardized residual covariances

Item	**1**	**4**	**5**	**11**	**3**	**8**	**9**	**12**
1		0.19;0.43	0.14;0.37	0.00;0.26	-0.18;0.02	-0.11;0.01	-0.14;0.00	-0.14;0.05
4	0.19;0.44		0.16;0.43	0.06;0.31	-0.17;0.01	-0.11;0.03	-0.17;-0.04	-0.08;0.10
5	0.13;0.38	0.17;0.41		-0.05;0.20	-0.14;0.08	-0.12;0.01	-0.15;0.00	-0.12;0.08
11	0.00;0.26	0.06;0.31	-0.04;0.20		-0.19;0.03	-0.12;-0.01	-0.10;0.06	-0.05;0.14
3	-0.18,0.02	-0.17;0.01	-0.12;0.07	-0.18;0.02		-0.05;0.07	-0.04;0.10	-0.06;0.14
8	-0.10;0.01	-0.09;0.01	-0.11;-0.01	-0.12;0.00	-0.04;0.06		0.00;0.10	-0.06;0.04
9	-0.14,-0.01	-0.16;-0.05	-0.13;-0.01	-0.10;0.05	-0.02;0.09	0.02;0.08		-0.06;0.06
12	-0.14;0.04	-0.07;0.10	-0.10;0.07	-0.05;0.15	-0.04;0.12	-0.05;0.03	-0.05;0.05	

Note: above the diagonal, under ADF assumptions; below the diagonal, under normality assumptions.

Table 6

Results for a one factor model applied to the Life Orientation Test (LOT) data: Modification indices for covariances among error terms

Item	**1**	**4**	**5**	**11**	**3**	**8**	**9**	**12**
1		**52.38**	**32.94**	6.97	4.49	4.40	7.83	1.78
4	**66.06**		**52.10**	**17.17**	5.06	3.82	**20.49**	.11
5	**41.54**	**65.71**		2.76	.63	7.38	8.53	.25
11	8.80	**21.66**	3.49		3.85	7.00	.78	1.64
3	5.67	6.38	.80	4.86		.24	2.74	1.71
8	5.55	4.82	9.31	8.83	.30		**29.75**	.24
9	**9.88**	**25.84**	**10.76**	.98	3.46	**37.52**		.05
12	2.24	.14	.32	2.07	2.16	.31	.06	

Note: Statistics computed under asymptotically distribution free (ADF) assumptions above the diagonal, under normality assumptions (NT) below the diagonal. We have boldfaced the statistics that are significant at the 5% level after applying a Bonferroni correction.