The Relationship between the Standardized Root Mean Square Residual and Model Misspecification in Factor Analysis Models

Dexin Shi
University of South Carolina
Alberto Maydeu-Olivares
University of South Carolina
University of Barcelona
Christine DiStefano
University of South Carolina

Multivariate Behavioral Research, *53*(5), 676–694.
https://doi.org/10.1080/00273171.2018.1476221

Correspondence concerning this article should be addressed to Dexin Shi, Dept. of Psychology, University of South Carolina. Barnwell College. 1512 Pendleton St. Columbia, SC 29208. E-mail: shid@mailbox.sc.edu. We are indebted to Yves Rosseel for implementing the unbiased SRMR in *Lavaan*. We thank the anonymous reviewers for their valuable suggestions. We also acknowledge the Research Computing Center at the University of South Carolina for providing the computing resources that contributed to the results of this paper. This research was supported by the National Science Foundation under Grant No. SES-1659936.

Abstract

We argue that the definition of close fitting models should embody the notion of substantially ignorable misspecifications (SIM). A SIM model is a misspecified model that might be selected, based on parsimony, over the true model should knowledge of the true model be available. Because in applications the true model (i.e., the data generating mechanism) is unknown, we investigate the relationship between the population Standardized Root Mean Square Residual (SRMR) values and various model misspecifications in factor analysis models to better understand the magnitudes of the SRMR. Summary effect sizes of misfit such as the SRMR are necessarily insensitive to some non-ignorable localized misspecifications (i.e., the presence of a few large residual correlations in large models). Localized misspecifications may be identified by examining the largest standardized residual covariance. Based on the findings, our population reference values for close fit are based on a two-index strategy: a) largest absolute value of standardized residual covariance $\leq .10$, and b) SRMR $\leq .05 \times \bar{R}^2$ the average R^2 of the manifest variables; for acceptable fit our values are .15 and $.10 \times \bar{R}^2$, respectively.

Keywords: Structural Equation Modeling (SEM); Standardized Root Mean Square Residual (SRMR); Close Fit

The Relationship between the Standardized Root Mean Square Residual and Model Misspecification in Factor Analysis Models

In structural equation modeling (SEM), the assessment of model-data fit has long been an important, but difficult, issue. The most common test of fit, the Likelihood Ratio (LR) based chi-square test, is typically used to evaluate the discrepancy between a proposed model and the data. The results of the test suggest if the model is an adequate representation of the data. However, the LR chi-square is a test of exact fit, meaning it is testing that there is *no* discrepancy between the hypothesized model and the data. In most empirical situations, the model under consideration is to some degree incorrect i.e., misspecified (Box, 1979; MacCallum, 2003). Thus, in large samples, the use of the LR chi-square test will often suggest an unacceptable fit, even when the model misspecification is relatively minor.

In applications, tests of exact fit often reject the fitted model and researchers are keenly interested in determining whether their misfitting model is actionable (i.e., it could be retained and substantive inferences could be drawn from it), or should be rejected and a better model should be sought. Current practices involving the decision of whether a mispecified model is actionable (whether it fits ‘closely’) or not are often based on goodness of fit indices (e.g. the Comparative Fit Index, CFI; Bentler, 1990). Sample values of these goodness of fit indices are compared to a fixed cutoff value that have been proposed in the literature (e.g., Hu & Bentler, 1999). If a sample estimate meets the recommended cutoff value (e.g. CFI ≥ 0.95), the model is retained as a “close fitting” model. Otherwise, the model is rejected.

Researchers have pointed out several problems with the practice of using goodness of fit indices with this “hypothesis testing” approach. One source of problems involves the use of goodness of fit indices (Barrett, 2007; Maydeu-Olivares, 2017; Yuan, 2005). Another source of problems involves the definition (or lack thereof) of 'close fitting' model. We describe each of these topics in turn.

Goodness of fit indices vs. effect sizes of model misfit

One main concern involving goodness of fit indices is that the procedures are largely heuristic, and are not grounded in statistical theory. A decision is made solely by evaluating the estimated value (i.e., sample statistic), which may be a biased estimator of the population parameter of interest. In addition, the sampling variability of the statistic is blatantly ignored. Thus, researchers may not know how widely estimates vary across samples.

A better alternative for assessing close fit is to use effect size measures of misfit. Effect sizes of model misfit are population parameters that capture the discrepancy between the fitted model and the data generating process. For effect sizes of model misfit, statistical theory is available, which enables the construction of confidence intervals, and if of interest, statistical tests (Maydeu-Olivares, 2017)[1]. Various forms of effect sizes exist in the SEM literature, including measures which are unstandardized (e.g. the Root Mean Squared Error of Approximation, or RMSEA; Browne & Cudeck, 1993; Steiger, 1989, 1990), standardized (e.g. the Standardized Root Mean Squared Error, or SRMR; Bentler, 1995; Joreskog & Sorbom, 1988), or relative (e.g. the Goodness of Fit Index, or GFI; Jöreskog & Sörbom, 1988; Maiti & Mukherjee, 1990; Steiger, 1989).

Currently, the most widely used effect size of model misfit is the RMSEA (Browne & Cudeck, 1993; Steiger, 1990). The RMSEA measures the unstandardized discrepancy between the population and the fitted model, adjusted by the degrees of freedom (*df*) of the model. Formal statistical inferences can be made by testing the hypothesis that $RMSEA \leq c$, where c is the reference cut-off in the population suggesting close fit. The most commonly used cut-off value is based on the recommendation from Browne and Cudeck (1993, p. 144), where the authors stated that "practical experience has made us feel that a value of the RMSEA of about 0.05 or less would indicate a close fit of the model in relation to the degrees of freedom". However, the population RMSEA is impossible to interpret because it is in an unstandardized metric. As Edwards (2013, p. 213) puts it "We do not know what a 0.01 difference in RMSEA values means. We do not know that a model with an RMSEA of 0.12 is incapable of telling us something useful about the world. We do not know that a model with an RMSEA of 0.01 is telling us anything useful about the world." Besides the level of model misspecification, the population RMSEA is dependent on other characteristics of the population model (i.e., "incidental parameters", Saris, Satorra, & van der Veld, 2009). For example, the same population RMSEA (say 0.05) may hold a different meaning in terms of the model misspecification when models differ in terms of the magnitude of factor loadings and model size (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Savalei, 2012; Shi, Lee & Maydeu-Olivares, 2018).

Standardized effect sizes are preferable to unstandardized measures as they facilitate the interpretation of the magnitude of misfit. The most popular standardized effect size of misfit is the Standardized Root Mean Squared Residual (SRMR), which can be crudely interpreted as the average standardized residual covariance. Recently, Maydeu-Olivares (2017) derived an unbiased estimator of the population SRMR, its asymptotic standard error, and suggested using a standard normal reference distribution to approximate its asymptotic distribution. As a result, the SRMR can be used to provide a statistical test of close fit (i.e. $SRMR \leq c$) and it is an attractive alternative to the use of the RMSEA. A major advantage of using SRMR over RMSEA is that its value can be substantively interpreted. Also, in finite samples, Maydeu-Olivares, Shi & Rosseel (2017) showed that compared to RMSEA, SRMR yielded more accurate empirical rejection rates and better coverage to its population value, especially when the number of observed variables was larger than 30.

A few studies have examined the behavior of the sample SRMR under various types of model misspecifications (Beauducel & Wittmann, 2005; Fan & Sivo, 2005, 2007; Garrido, Abad, & Ponsoda, 2016; Hu & Bentler, 1998, 1999), including the most influential simulation study conducted by Hu and Bentler (1999). In trying to find balance between minimizing Type I errors (i.e., rejecting a correctly specified model) and maximizing power (i.e., rejecting a misspecified model), Hu and Bentler (1999) suggested that models with sample SRMR values less than .08 generally indicated adequate fit. These previous studies have focused on simulating data through use of finite samples, which inherently include sampling error. In addition, the formula used is a biased estimate of the population SRMR, and the resultant estimate may be noticeably different from its population counterpart in small samples (Maydeu-Olivares, 2017). Given that previous studies have focused on the *sample* SRMR, researchers would benefit from greater understanding of how *population* SRMR is affected by model misspecifications.

SEM misfit may best be characterized as a multivariate problem, and it requires examining all statistically significant standardized residual covariances, especially the ones with largest absolute value (Maydeu-Olivares & Shi, 2017; McDonald & Ho, 2002b; Raykov, 2000) [2]. An examination of the full matrix of residual covariances, or at least the most extreme value within that matrix, is seldom performed in practice. However, being (crudely) the average standardized residual covariance, researchers have suggested that it is only meaningful to interpret the SRMR when there is little variability among the standardized residual covariances and there are no clear outliers (i.e., some standardized residual covariances much larger than the rest: McDonald & Ho, 2002; Raykov, 2000). Also, the SRMR may not be sensitive to model misspecifications which cause only a small proportion of residuals to be large in a residual covariance matrix that may include many zeros or small values. However, these aspects have not been thoroughly explored, and as of yet, the relationship between the largest standardized residual covariance and model misspecification has not been fully explored, and therefore no clear criteria are available to assess close fit.

A definition of close fit: Substantively ignorable misfit (SIM)

Although the population SRMR is easy to interpret in a standardized metric, it is not intuitive to researchers what a specific value of SRMR (for example, .05) implies in terms of misspecification(s) in the fitted model. Therefore, it is necessary to investigate the relationship between the magnitudes of effect size (i.e., the population SRMR) and some common types of model misspecification to gain greater understanding of the meaning of the population parameter. Thus, we aim at addressing Edwards's concern (using the SRMR) and help applied researchers to make a more informed decision on whether to retain or reject a misspecified model.

We recognize that whether to retain or reject a misspecified model depends on the purpose of the application, and therefore it is necessarily subjective. Yet, in many instances an application may serve several purposes, or the study be purely exploratory. In these instances, researchers may find helpful that some reference cut-off values be provided. Choosing a reference cut-off value (c) is a difficult but unavoidable issue, because a decision needs to be made regarding the model's use. A direct analogy involves *p*-values and significance levels. Statistics has often been described as the quantification of uncertainty. From this view point, statistics finishes once a *p*-value (or a confidence interval) is obtained. However, if a decision must be made, the use of an agreed upon significance level (a cut-off value) greatly facilitates scientific communication.

Putting forth suggested cut-off values to distinguish between close and non-close fitting models is made more difficult by the absence of any definition of close fitting model in the literature. For instance, based on their practical experience, Browne and Cudeck (1993) simply defined a model as fitting closely when RMSEA is less than or equal to 0.05. To overcome this shortcoming we define a model to provide a close fit to the data generating mechanism if its misspecification is substantively ignorable. More specifically, we define a model with substantively ignorable misspecifications (SIM) as a mispecified model that might be selected, based on parsimony, over the true model should knowledge of the true model be available. To the best of our knowledge, the first research that used the notion of SIM (without explicitly defining it) to specify criteria for close fit is Maydeu-Olivares and Joe (2014) who used SIM to establish cut-off values for the use of the RMSEA in IRT models. Also, our use of SIM is consistent with some of the most influential writings in goodness-of-fit assessment in SEM (e.g., Saris et al., 2009).

Because in applications the data generating mechanism (or in short, the true model, although the expression is an oxymoron) is unknown, it is necessary to probe different combinations of true and fitted models and consider for each combination which misspecifications are substantively ignorable. In this article we focus on factor analysis models and we consider three classes of misspecifications: 1) misspecified dimensionality (e.g., fitting a one-factor model when the true model is a two-factor model); 2) fitting an factor model with independent clusters when the true model contains small cross-loadings; 3) fitting a model with uncorrelated errors when the true model contains correlated errors.

Consider the following specific example of the first scenario: choosing between a one factor model and a two factor model with independent clusters (i.e, every item loads on a single factor) whose factors correlate ρ. What magnitude of ρ is substantively ignorable? Certainly, not 0, but neither it is .3. We believe that most researchers confronted with a choice between a one factor model and a two factor model with $\rho = .99$ would choose a one factor model. Consequently, the misspecification obtained when fitting a one factor model in this case is substantively ignorable. We also believe that the misfit corresponding to $\rho = .90$ is substantively ignorable, but we do not wish to go further. We prefer to err on the safe side (the same spirit was used to establish cut-off values – significance levels– for *p*-values). Therefore we define a one factor model fitted to a two factor model with $\rho \geq .90$ as fitting closely, and not fitting closely otherwise. We use similar criteria for other combinations of true and fitted models. For instance, we consider (standardized) cross-loadings less than or equal to .10 as substantively ignorable (provided they do not follow a substantively meaningful pattern). We believe most researchers would prefer to fix cross-loadings to zero when their standardized magnitude is ≤ .10, but again we prefer to err on the safe side, and we do not wish to go further. Similarly, we consider correlations among the residuals less than or equal to .10 as substantively ignorable (provided the correlations are not patterned).

To summarize, to help researchers to make a more informed decision on whether to retain or reject a misspecified model, we investigate the relationship between the population SRMR and model misspecifications in the context of the factor analysis models. In addition to the SRMR, the behavior of the largest standardized residual covariance (i.e., defined in terms of absolute values) is inspected. In particular, we are interested in examining whether models with substantively ignorable misfit (SIM, our definition of close fit) can be distinguished from non-SIM models using these two parameters.

The remainder of this paper is organized as follows. We first review statistical theory and clarify the formula for unbiased estimation of the population SRMR. Next, using population covariance matrices, we explore the behaviors of the population SRMR and the largest standardized residual covariance under different types and degrees of model misspecification. That is, for an array of true and fitted models we specify which ones we consider close fitting or acceptable (i.e., actionable) and we report the corresponding values of the SRMR (and the largest standardized residual covariance). We conclude this discussion by offering practical guidance for empirical research when using the SRMR to assess goodness of fit.

Statistical Theory for the SRMR

Let σ_{ij} denote the unknown population covariance between variables *i* and *j* (or the variance if $i = j$) and σ_{ij}^0 denote the population covariance (or variance) under the fitted model. Then, the population SRMR (P.SRMR) is defined as (Maydeu-Olivares, 2017):

$$P.SRMR = \sqrt{\frac{\boldsymbol{\varepsilon}_s'\boldsymbol{\varepsilon}_s}{t}} = \sqrt{\frac{1}{t}\sum_{i \le j} \frac{\left(\sigma_{ij} - \sigma_{ij}^0\right)^2}{\sqrt{\sigma_{ii}\sigma_{jj}}}} \;. \tag{1}$$

Here, $\boldsymbol{\varepsilon}_s$ is the vector of the population standardized residual covariances, $t = p(p+1)/2$ signifies the number of unique elements in the (residual) covariance matrix, where *p* denotes the number of observed variables being modeled. Thus, Eq. 1 approximates the average population standardized residual covariance.

In finite samples, let s_{ij} be the sample covariance, $\hat{\sigma}_{ij}$ denote the model implied covariance, $\mathbf{e}_s$ be the *t* vector of the standardized residual covariances with elements

$$e_{ij} = \frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{s_{ii} s_{jj}}}, \tag{2}$$

and Ξ_s represent the asymptotic covariance matrix of $\mathbf{e}_s$. Maydeu-Olivares (2017) showed that regardless of the discrepancy function and distributional assumptions used, an asymptotically unbiased estimate of the population SRMR of Eq. 1 can be expressed as:

$$SRMR_u = \hat{k}_s^{-1} \sqrt{\frac{\max\left(\mathbf{e}_s{}'\mathbf{e}_s - \mathrm{tr}(\hat{\Xi}_s), 0\right)}{t}}, \quad \text{where } \hat{k}_s = 1 - \frac{\mathrm{tr}(\hat{\Xi}_s^2) + 2\mathbf{e}_s'\hat{\Xi}_s\mathbf{e}_s}{4\left(\mathbf{e}_s'\mathbf{e}_s\right)^2}. \tag{3}$$

However, the asymptotic covariance matrix, Ξ_s, depends on the discrepancy function used to estimate the model, and on whether normality or asymptotically distribution free (ADF: Browne, 1982) assumptions are used.

In typical applications, SEM software programs compute a sample counterpart of the population SRMR in Eq. 1 as:

$$SRMR_b = \sqrt{\frac{\mathbf{e}_s{}'\mathbf{e}_s}{t}} = \sqrt{\frac{1}{t} \sum_{i \le j} \frac{\left(s_{ij} - \hat{\sigma}_{ij}\right)^2}{\sqrt{s_{ii} s_{jj}}}}, \tag{4}$$

where the elements in the equation are defined earlier. It is noted that the sample SRMR shown in Equation 4 is a biased estimate of the population SRMR. Following Maydeu-Olivares (2017), we derived the expected value of the biased estimate $SRMR_b$. This can be approximated in large samples using

$$E[SRMR_b] = \sqrt{\frac{\mathrm{tr}(\Xi_s) + \boldsymbol{\varepsilon}_s'\boldsymbol{\varepsilon}_s}{t}} \frac{8[\mathrm{tr}(\Xi_s) + \boldsymbol{\varepsilon}_s'\boldsymbol{\varepsilon}_s]^2 - 2\,\mathrm{tr}(\Xi_s) - 4\boldsymbol{\varepsilon}_s'\Xi_s\boldsymbol{\varepsilon}_s}{8[\mathrm{tr}(\Xi_s) + \boldsymbol{\varepsilon}_s'\boldsymbol{\varepsilon}_s]^2}. \tag{5}$$

In appendix 1, we support the accuracy of this approximation by demonstrating the biases of the *SRMRb* and *SRMRu* in estimating the population SRMR using a small simulation example. It is noted that the *SRMRb* (Eq. 4) generally reported in software packages is upwardly biased, meaning the index typically suggests worse model fit than is actually present. The amount of bias is compounded when sample size is small and low standardized factor loadings are present.

Confidence intervals for the SRMR and tests of close fit can be obtained using its unbiased estimate and a reference normal distribution. Specifically, with large samples, a (100 – α)% confidence interval for the SRMR can be obtained using

$$\Pr\left(SRMR_u - z_{\alpha/2}SE(SRMR_u) \le SRMR \le SRMR_u + z_{\alpha/2}SE(SRMR_u)\right) = 1-\alpha\,, \tag{6}$$

where SE () denotes asymptotic standard error, which is given as (Maydeu-Olivares, 2017):

$$\mathrm{SE}(SRMR_u) = \sqrt{k_s^{-2}\,\frac{\mathrm{tr}(\boldsymbol{\Xi}_s^2) + 2\mathbf{e}_s'\boldsymbol{\Xi}_s\mathbf{e}_s}{2t\;\mathbf{e}_s'\mathbf{e}_s}}\,. \tag{7}$$

In addition, statistical test for model close fit can be conducted with hypotheses such that

$$H_0 : SRMR \le c \text{ vs. } H_1 : SRMR > c\,,$$

where $c > 0$ is a reference cut-off value for close fit, p-values are obtained using $p = 1 - \Phi(z)$, where $\Phi()$ denotes a standard normal distribution function and $z = \dfrac{SRMR_u - c}{SE(SRMR_u)}$.

Population SRMR and Model Misspecification

The relationship between the population SRMR and misspecification in factor analysis models is explored in this section. In addition to the population SRMR, the behavior of the largest standardized residual covariance (i.e., defined in terms of absolute value) was inspected across different model misspecifications. The performance of these two parameters was evaluated through a simulation study in which we generated population covariance matrices as the purpose of this study is to better understand the behavior of the population SRMR.

The population model used was a confirmatory factor analysis (CFA) model with two correlated factors. The following three types of misspecification, that are often observed in practice when fitting CFA models, were considered.

a) *Misspecified dimensionality*. The population model was as two factor, independent clusters model with correlated factors. A one factor model was fit to the two-dimensional structure.

b) *Omitting cross-loadings*. Items related to each factor via an independent clusters structure; however, one or multiple indicators cross-loaded on both factors. The fitted model assumed an independent clusters structure for both factors, where the cross-loading value(s) was incorrectly fixed to zero.

c) *Omitting residual correlations*. In the population model, one or multiple residual correlations (covariances) were present. A simple structure model with no correlated error was fit.

To produce these population covariance matrices, factor variances were fixed to one. The error variances were set such that all factor loadings (including cross-loadings) are on a standardized scale. Other characteristics that were manipulated are as follows:

Magnitude of Factor Loadings. The population factor loadings (λ) included low (.40), medium (.60), and high (.80). The primary factor loadings are the same for all items across factors.

Model size. The model size is indicated by the number of observed variables (*p*; Moshagen, 2012; Shi, Lee, & Terry, 2015, 2018)[3]. Model size ranged from small to very large, including p = 12, 36, 72, or 144. Equal number of items loaded on each factor.

Magnitude of Model Misspecification. In the scenario in which multidimensionality was ignored, the level of misspecification was manipulated by changing the degree of correlation between the factors in the true model. The true correlation coefficients ranged from .60 to .90, in increments of .10. Given that the estimated model collapsed across two factors in the population model, a smaller inter-factor correlation indicated a greater level of misspecification. When the misspecification included omitted cross-loadings or residual correlations, the inter-factor correlation was fixed to .30 and the level of misspecification was determined by the population value of the omitted parameters. The omitted cross-loadings or residual correlations ranged from .10 to .40, in increments of .10. Here, larger values indicated a higher level of model misspecification.

Number of Omitted Parameters. When the misspecifications were introduced by omitting cross-loadings or residual correlations, we manipulated the number of omitted cross-loadings/ residual correlations. The number of omitted parameters ranged from one to four, in increments of one. In summary, the number of conditions examined were 48 = 3 (factor loading levels) × 4 (model size levels) × 4 (factor inter-correlation levels) for misspecified dimensionality. When investigating situations with cross-loadings or residual correlations, the number of conditions examined were 192 = 3 (factor loading levels) × 4 (model size levels) × 4 (magnitudes of omitted parameters) × 4 (number of omitted parameters).

For each condition, a population covariance matrix was computed. Then, the population values of SRMR were calculated by fitting the misspecified model to the population covariance matrices with maximum likelihood (ML) estimation, using the *lavaan* package in R (Rosseel, 2012; R Core Team, 2015). In addition to the SRMR values, every standardized residual covariance was obtained, and the largest absolute value for the standardized residual covariance was used as an alternative index for evaluating model fit. Analyses of variance (ANOVAs) were conducted to identify conditions which affected the outcome of interest. An eta squared (η^2) value above 10% was used to identify conditions that contributed to sizeable amounts of variability in the outcome. To better demonstrate the behaviors of the population SRMR and the largest standardized residual covariance, all results were presented in the form of figures; tables that include complete simulation results are provided as supplementary materials.

Behavior of the population SRMR and largest standardized residual covariance in models with mispecified dimensionality

The behavior of SRMR and the behavior of the largest absolute value of the standardized residual covariance under the misspecification of disregarding multidimensionality are shown in Figure 1 (panel A and B, respectively). In these figures, different markers were used to indicate different values of inter-factor correlations (i.e., degree of model misspecification) and for each level of inter-factor correlation, the population SRMR and the largest (absolute value of) standardized residual covariance were plotted against the magnitudes of the factor loadings. Given a fixed level of factor loading and inter-factor correlation, cases with different model sizes were labeled separately. By comparing Figures 1A and 1B, we can see that when dimensionality is mispecified, the SRMR value and largest standardized residual covariance provide almost the same information. A linear regression model predicting the SRMR from the largest standardized covariance showed that the two parameters are almost perfectly correlated (R^2 =.988). Therefore, only results for the SRMR are discussed to avoid redundancy.

The results from the ANOVA showed that the most important sources of population SRMR variance were the magnitude of factor loadings ($\eta^2 = .50$) and inter-factor correlations ($\eta^2 = .39$). Specifically, Figure 1A demonstrates SRMR's sensitivity to factor's structural misspecification; as the inter-factor correlation decreased (indicating a more severe misspecification), the population SRMR increased accordingly. In addition, at any fixed level of model misspecification (e.g., $\rho = .8$), the population SRMR depends on the values of the standardized factor loadings according to a curvilinear relationship. The higher the factor loadings, the higher the population SRMR. Moreover, the effects of model misspecification and factor loading are multiplicative, meaning that the higher the level of model misspecification, the higher the effect of the factor loading. Finally, model size did not have a great impact on the results. For a fixed level of factor loading and inter-factor correlation, variability displayed in Figure 1A due to differences in model size was ignorable ($\eta^2 < .01$).

As seen in Figure 1A, the commonly used cutoff of SRMR = .08 (the solid horizontal line) appears to be too liberal a criterion for identifying misspecification in the factor structure. For example, when factor loadings are .6 or less, fitting a one-factor model to a two-factor structure with a low inter-factor correlation (e.g. ρ =.6), the reference of SRMR $\leq$ .08 would always suggest acceptable model fit. This finding holds regardless of model size. In fact, because the population SRMR heavily depends on the size of factor loadings, it is difficult to find a single reference to separate practically well-fitting models (e.g. $\rho > .9$) from non-closely fitting models that should be rejected (e.g. $\rho = .6$).

A curvilinear relationship was observed between SRMR and the magnitude of the standardized factor loading (see Figure 1A). Therefore, we considered to use and evaluate the SRMR in light of the communality (i.e., the squared standardized loading, or λ^2)[4]. Using the ratio of SRMR to the communality (SRMR / λ^2) as the outcome variable, the results from the ANOVA indicated that the inter-factor correlation explained a very large amount of the variance ($\eta^2 = .99$). Figure 1C depicts the relationship between SRMR / λ^2 and the standardized factor loadings across levels of inter-factor correlation (ρ) and model size. As shown, population SRMR / λ^2 only depended on the level of model misspecification (ρ), regardless of the magnitude of the factor loadings and model size. Thus, if close fit is defined as fitting a one-factor model to two-factor data when inter-factor correlations are greater than or equal to .90, such models imply that SRMR / $\lambda^2 \leq .05$ (the solid line in Figure 1C).

Insert Figure 1 around here

Behavior of the population SRMR and largest standardized residual covariance in models with omitted cross-loadings

When the models were misspecified by omitting cross-loadings, ANOVA results showed that the most important sources of population SRMR variance were the size of cross-loading ($\eta^2 = .28$), magnitude of the factor loadings ($\eta^2 = .18$), and model size ($\eta^2 = .16$). In Figure 2, we plotted the behavior of SRMR in the presence of one to four omitted cross-loadings in four separate panels (i.e. panel A-D). These figures show that population SRMR increased as the magnitude of the omitted cross-loading(s) increased. However, the commonly used cutoff is not sensitive to detect misspecification caused by omitting cross-loading(s). Almost all misspecified models would be retained using the guideline of .08 for close fit (the solid horizontal line), even when the size of the omitted cross-loading(s) reached .40. In addition, by comparing the four panels in Figure 2, the population SRMR slightly increased as the number of omitted cross-loadings increased. However, the effect of number of omitted cross-loadings on SRMR was quite small ($\eta^2 = .06$).

The population SRMR was also influenced by the magnitude of the factor loadings; higher factor loadings were associated with a larger population SRMR, especially when the omitted cross-loading value was large. For example, when the number of observed variables was

12 and one cross-loading of .40 magnitude was omitted, the population SRMR increased from .029 to .085 as the magnitude of the (primary) loading values increased from .40 to .80.

For a fixed level of primary factor loading and cross-loading(s), the variability of the population SRMR was still noticeable, indicating a substantial effect of model size. Specifically, holding other conditions constant, larger model size was associated with smaller SRMR values. Moreover, the effect of model size was more noticeable as the level of model misspecification and the magnitude of the factor loading increased. For example, when the factor loadings were set to .80 and one cross-loading of .40 was omitted, the population SRMR decreased from .085 (p = 12) to .024 (p = 144). With a smaller factor loading (i.e., .40) and lower level of model misspecification (i.e., cross-loading = .20), model size had a smaller effect on the SRMR, yielding values which ranged from .016 (p = 12) to .006 (p = 144).

Insert Figure 2 around here

The behavior of SRMR / λ^2 in the presence of omitted cross-loadings is shown in Figure 3. ANOVA results indicated that the most important sources of variance in SRMR / λ^2 were the size of cross-loading (s) (η^2 = .46) and model size (η^2 = .24). It is noted that the effect of model size on SRMR / λ^2 was less noticeable as the level of model misspecification decreased. For example, when the factor loading was .40 and one cross-loading with size of .40 was omitted, the SRMR to communality ratio (SRMR / λ^2) decreased from .183 (p = 12) to .074 (p = 144). When the model misspecification was less severe (i.e. cross-loading = .10), the range of SRMR / λ^2 was smaller across model sizes, with values from .053 (p = 12) to .019 (p = 144). Generally speaking, if researchers would accept models that omit minor cross-loadings (i.e. cross-loading $\leq .10$), SRMR / $\lambda^2 \leq .05$ (the solid line in Figure 3) could be used as a reference.

Insert Figure 3 around here

Using the largest absolute value of the standardized residual covariance as the outcome variable, ANOVA results indicated that the level of cross-loadings ($\eta^2 = .37$), the magnitude of factor loadings ($\eta^2 = .17$), model size ($\eta^2 = .15$), and number of omitted cross-loadings ($\eta^2 = .14$) were the major sources that explained the majority of the variance. Each panel in Figure 4 illustrates the behavior of the largest absolute value of the standardized residual covariance in terms of the magnitudes of factor loadings, levels of cross-loadings and model size across various number of cross-loadings. We can see that the largest absolute value of the standardized residual covariance increased when the magnitude of factor loadings and level of cross-loadings increased.

In addition, the largest standardized residual covariance generally increased as the number of omitted cross-loading increased. However, as more cross-loadings were omitted, the amount of increase in the largest standardized residual covariance "leveled off" (i.e., smaller increases were observed). For example, when $p= 144$, and the magnitude of factor loading was 0.6, if omitting one cross-loading of .20, the largest standardized residual covariance was .107; the largest standardized residual covariance increased to .224 as two cross-loadings were omitted. Nevertheless, as the number of omitted cross-loadings kept increasing, the largest standardized residual covariance tended to remain stable, yielding SRMR= .222 (omitting three cross-loadings) and .219 (omitting four cross-loadings).

Given a fixed level of factor loading and cross-loading, it can be that the variability of the points displayed in Figure 4 is much smaller than those in Figure 2. This implies that the effect of model size on the largest absolute value of the standardized residual covariance was much smaller than the effect on the SRMR. Moreover, the effects of model size on SRMR and the largest standardized residual covariance were in opposite directions. That is, as *p* increased, the largest standardized residual covariance tended to slightly decrease. As shown in Figure 4, if close fit is defined as omitting cross-loadings less than or equal to .10, such models can usually be identified by applying the cut-off with the largest absolute value of the standardized residual covariance $\leq .10$ (the solid horizontal line).

Insert Figure 4 around here

Behavior of the population SRMR and largest standardized residual covariance in models with omitted correlations among the residuals

In the presence of omitted residual correlations, ANOVA results showed that the most important sources of population SRMR variance were model size ($\eta^2 = .53$) and the size of residual correlations ($\eta^2 = .13$). In Figure 5, we present the behavior of population SRMR across the study conditions. Panels A-D represent conditions where one to four residual correlations were omitted. For each panel from Figure 5 (i.e., same number of omitted residual correlations), given a fixed level of factor loadings and residual correlations, the variability of the population SRMR was very large. This indicates that model size was the dominant factor for the values of population SRMR. Specifically, models with more observed variables yielded noticeably smaller population SRMR values. For example, when the factor loadings were .40, and four residual correlations of .40 were ignored, the population SRMR decreased from .070 (p = 12) to .007 (p = 144). In addition, for a fixed number and level of residual correlations, the population SRMR slightly decreased as the magnitude of factor loadings increased; yet, the effect size of the magnitude of factor loadings was quite small ($\eta^2 = .07$). For example, when p = 12, and four residuals correlations of .10 were ignored, the population SRMR dropped marginally, from .018 (λ=.40) to .008 (λ=.80).

The figures also show that population SRMR increased as the number and magnitude of the omitted residual correlations increased. However, the value of population SRMR was still rather small, even when four residual correlations with size of .40 were ignored. The commonly used cutoff for SRMR failed to detect the misspecification caused by omitting residual correlations. That is, all misspecified models considered would be retained using the guideline of .08 for close fit (i.e., denoted by solid horizontal line). Based on the results, when the magnitude of model misspecifications increased, the related change in the population SRMR could be small. Moreover, the population SRMR was greatly impacted by model size. Therefore, it is problematic to use SRMR (or SRMR / λ^2) as the criteria to detect model misspecifications from omitting residual correlations.

Insert Figure 5 around here

On the other hand, for the largest absolute value of the standardized residual covariance, the ANOVA results indicated that the level of model misspecification (residual correlations) could explain the majority of the variance ($\eta^2 = .62$). The magnitude of the factor loading was an important source of variance ($\eta^2 = .31$); however, the main effect of the model size was negligible ($\eta^2 < .01$). Figure 6 illustrates the behavior of the largest absolute value of the standardized residual covariance when omitted residual correlations were present. We can see that the largest absolute value of the standardized residual covariance was more sensitive than the SRMR to identify misspecified models with omitted residual correlations. The largest absolute value of the standardized residual covariance increased as the size of the residual correlations increased, indicating a more severe misspecification. In addition, the value of the largest standardized residual covariance was much larger than the SRMR obtained from the same misspecified model. The largest absolute value of the standardized residual covariance was negatively associated with the magnitude of the factor loading. In addition, as the magnitude of factor loading increased, the decrease in the largest standardized residual covariance tended to be more gradual when the level of model misspecification was smaller. For example, for a fixed model size (i.e., $p = 12$), when the model was misspecified by omitting one residual correlation = .40, the largest absolute value of the standardized residual covariance decreased from .313 ($\lambda = .40$) to .139 ($\lambda = .80$). If the omitted residual correlation = .10, the largest absolute value of the standardized residual covariance dropped more gradually from .080 ($\lambda = .40$) to .035 ($\lambda = .80$). Therefore, when the misspecifications occurred by omitting residual correlations, the largest absolute value of the standardized residual covariance was a more suitable index than the SRMR for assessing model fit. As shown in Figure 6, close fitting models can be identified by the largest absolute value of the standardized residual covariance ≤ .10 (the solid horizontal line), which approximately corresponds to omitting correlated residuals with correlations of .10.

Insert Figure 6 around here

A numerical example: Fitting a five factor model to the SPRI-R

We provide a numerical example to illustrate our discussion. The Social Problem Solving Inventory-Revised (SPSI-R: D'Zurilla, Nezu, & Maydeu-Olivares, 2002) is a 52 item questionnaire that according to its authors measures five attributes. Each item is scored using five ordered categories. Maydeu-Olivares and D'Zurilla (1996) report confirmatory (CFA) and exploratory (EFA) factor analyses fitted to a sample of $N = 601$ individuals. These data will be reanalyzed here. We do not report parameter estimates as these are available in the original sources. Rather, we wish to focus on answering the question 'do these models fit closely enough?'. If the answer is negative, a better model should be sought for these data.

The data is quite normal, the excess kurtosis for all items is well below one; only one item has skewness larger than $|1|$. Consequently, we simply provide fit results under normality assumptions to simplify our presentation. First we fitted the independent clusters CFA model of Maydeu-Olivares and D'Zurilla (1996): every item is loaded only by its factor. The CFA yields χ^2 (i.e., the likelihood ratio statistic) = 3,209.87 on 1264 df, $p < .001$. A 90% confidence interval (CI) for the RMSEA is (.048; .053), and the unbiased SRMR is .053. The model does not fit the data exactly, but by current standards of 'close' fit, we would conclude based on this information that the model provides a close fit to the data. In so doing, all that we have done is to apply some cutoff values recommended in the literature; the definition of 'close' is based on those cutoffs. A model fits closely if the estimated values are below the cutoffs, and it does not fit closely otherwise (Barrett, 2007).

In this paper, we have put forth a definition of close fit. We deem a model to provide a close fit to the data generating process if we would choose the fitted model over the data generating process based on parsimony, i.e., if the fitted model showed substantively irrelevant misfit. We estimate a 90% CI for the SRMR as (.048; .057), we obtained the average R^2 of the items, $\bar{R}^2 = .42$, and we inspected the largest (in absolute value) standardized residual covariances. There are two residuals > |.2| and over 30 residuals > |.1|, all statistically significant at the 5% level after applying a Bonferroni correction for the (53×52)/2 = 1,378 residuals inspected. We cannot ignore such large residuals and we conclude that the model does not fit closely the unknown data generating process. In fact, we conclude that the fit is not even acceptable, in spite of the estimated RMSEA. Therefore, a better model must be sought. Our conclusion is supported by an inspection of the modification indices. Five of the modification indices involving the factor loadings involved expected standardized cross-loadings larger than |.3| and as high as |.45|; the modification indices for these five loadings ranged from 44 to 58 (on 1 df).

To obtain a better fitting model we estimated an EFA model using a target rotation (Browne, 1972) instead of adding cross-loadings (or correlated residuals) based on the modification indices results. An examination of the estimated loadings reveals that the structure conforms to that put forth by Maydeu-Olivares and D'Zurilla (1996): all secondary loadings are smaller than the primary loadings. Regarding the fit of this model, we obtained $X^2 = 2{,}233.67$ on 1076 df, $p < .001$. A 90% CI for the RMSEA is (.040; .045), and the unbiased SRMR is .021. The average communality is $\bar{R}^2 = .45$, a 90% CI for the SRMR is (.020, .023). There are only four standardized residual covariances larger than |.1|, a 90% confidence interval for the largest is (-.18, -.14). Now, $\text{SRMR}/\bar{R}^2 = .051$, an inspection of Figure 1C reveals that for this $\bar{R}^2$, the estimated SRMR is only slightly above what we would have obtained if we had fitted a one factor model to a two-factor independent clusters model with correlation .9. As we would rather use a one factor model than a two factor model with correlation .9, this lead us to believe that this EFA model fits rather closely the unknown data generating process. Also, we inspected Figures 4 and 6 involving the relationship between the largest standardized residual and the size of localized misfit. To properly interpret these figures it is convenient to estimate the average factor loading. To do so, we simply took $\lambda \approx \sqrt{\bar{R}^2} = .67$. Inspection of Figures 4 to 6 gives us pause to conclude that the EFA model fits closely. Rather, the estimated value of the largest residual suggests that there is some localized misfit in the model that exceeds our standards. Taking together the estimated values of the SRMR and largest standardized residual, we conclude that the model provides an acceptable fit to the data generating model: on average, the model provides a good overall fit to the data, but the model does not capture well enough every association among these 52 items. The conclusion is not surprising: with 52 items, finding a model with substantively ignorable misspecifications requires a lot of work.

Discussion

This research investigated the relationship between the population SRMR (and largest standardized residual covariance) and different types and degrees of misspecification in factor analysis models. Population covariance matrices were used to determine the impact of study conditions on the population SRMR without the impact of sampling error. Conditions of factor loading size and model size were manipulated. Findings showed that the population SRMR was sensitive to model misspecification due to fitting a one-factor model to two-factor data or ignoring cross-loadings, but less sensitive to misspecifications introduced by omitting residual correlations. These findings are consistent with previous studies (Fan & Sivo, 2005; Hu & Bentler, 1998;1999) using sample estimates of the SRMR.

The SRMR can be approximately interpreted as the average standardized residual covariance. As such, it is sensitive to model misspecifications that cause a substantial proportion of large elements (in absolute value) in the standardized residual covariance matrix. When the nature of the model misspecification results in only a small proportion of large residuals, and/or many structural zeros appear in the residual covariance matrix, the average residual covariance cannot provide an accurate representation of the misspecification. Therefore, the SRMR was not sensitive to model misspecifications caused by omitting a few residual correlations, especially when the number of observed variables (and thus, the number of residual covariances) is large. In such situations, examining the largest (in absolute value) standardized residual covariance is more useful. In fact, our findings revealed that both global fit and local fit are important concepts to consider when evaluating models (DiStefano, 2016; West, Taylor, & Wu, 2012; McDonald & Ho, 2002b; Raykov, 2000). That is, good global fit implies that even in the presence of some local misspecifications, the overall model holds up well, and should not be completely discarded. On the other hand, the information about local misfit can help researchers identify the sources of poor global fit and thereby improve the model.

We also explored the effects of two possible “incidental parameters” (i.e., model size and the magnitude of the factor loading). In general, larger standardized factor loadings were associated with higher population SRMR values when the model misspecifications occurred by ignoring the multidimensionality or omitting cross-loadings. However, when the model was misspecified by ignoring correlated residuals, for a fixed level of residual correlations, the population SRMR tended to slightly decrease as factor loading increased. The findings of the current study are consistent with previous methodological research regarding the influence of measurement quality (i.e., standardized factor loadings) on fit indices. Methodologists have shown that for a given level of model misspecification, poor measurement quality is associated with better model fit (i.e., the reliability paradox; see Hancock & Mueller, 2011; Heene, Hilbert, Draxler, & Ziegler, 2011; McNeish, An, & Hancock, 2018). This result can be somewhat counter-intuitive. In particular, researchers would judge the value of an ignorable secondary factor loadings against the values of the primary loadings. For a given size of the secondary loading, it should be more ignorable when the primary loadings are larger; however, the population SRMR tends to increase as the values of the primary loadings increase, suggesting worse fit.

In addition, when omitting cross-loadings or residual correlations, the population SRMR decreased as the number of observed variables increased. However, when misspecifying dimensionality (i.e., fitting a one-factor model to two-factor data), the effect of the model size on the population SRMR is quite small. The patterns described above are slightly different to the behavior of the population RMSEA (Savalei, 2012). In Savalei's (2012) study, she noted that the population RMSEA generally decreased as the magnitude of the factor loading decreased and the number of observed variables increased, despite the type of model misspecification. According to its definition, the RMSEA penalizes model complexity by incorporating a degree of freedom in the formulation, and it measures the discrepancy due to approximation per *df*. Therefore, for models with a fixed level of misspecification, the population RMSEA generally decreases as p increases because a higher p is typically associated with larger degrees of freedom (*df*). The SRMR, on the other hand, is (approximately) the average standardized residual covariance. When misspecifying dimensionality (e.g., fitting a one-factor model to two-factor data), the majority of the residual covariances are impacted by the misspecifications. Therefore, their average values are less sensitive to the size of the residual covariance matrix (as a function of p).

We showed that the population SRMR is not only determined by the size of the model misspecification, but may be influenced by other factors, including the type of model misspecification, the magnitude of the factor loading, and model size. As a result, the practice of using the SRMR with a single cut-off for testing model close fit may be misleading. For example, using the existing criterion of sample SRMR $\leq .08$, would lead us to retain as a close fitting model a one factor model fitted to data generated according to two-factor model with an inter-factor correlation of $\rho = .6$ if the standardized factor loadings were lower than or equal than .60 (e.g., $\lambda \leq .60$). Moreover, in spite of the magnitude of the factor loading or the model size, the currently accepted close fit criteria cannot detect misspecifications caused by omitting correlated residuals, even if the number and size of the omitted residual correlations are fairly large (e.g., four residual correlations of size .40).

In order to better distinguish close fitting models from models with substantially non-ignorable misspecification, we recommend examining both the largest standardized residual covariance (in absolute value) and the SRMR in light of the average estimated communality, that is R^2. Using this two-index strategy will capture different types of model misspecifications. In addition, for each index, to identify closely fitting models, a reference that is relatively robust to the magnitude of the factor loading and model size can be proposed. In this study, we considered three instances of misspecified models with substantively ignorable misspecification (i.e., providing a close fit to the true model): 1) fitting a one-factor model to two-factor data when inter-factor correlations larger than or equal to .90; 2) omitting cross-loadings with standardized values less than or equal to .10; or 3) ignoring correlated residuals with correlations less than or equal to .10. At the population level, such closely fitting models can be identified when: a) the largest absolute value of the standardized residual covariance $\leq .10$, and b) an SRMR to average communality (i.e., R^2) ratio, $\text{SRMR} / \overline{R}^2 \leq .05$ (or $\text{SRMR} \leq .05 \times \overline{R}^2$).

In Appendix 2, we further evaluate the generalizability of the proposed reference values in more complex situations (i.e. the population model has three correlated factors with different magnitudes of factor loadings). The two-index strategy also performed reasonably well under the complex study conditions.

Researchers may differ in what they consider a SIM, and we acknowledge that our own personal criteria for SIM, and hence, close fit are quite stringent. We believe many researchers would also consider a) fitting a one-factor model to two-factor data with an inter-factor correlation of $\rho \geq .80$, b) setting cross-loadings $\leq .20$, or c) setting residual correlations $\leq .20$ to zero to be substantively ignorable misspecifications. As a result, we believe models with these levels of misspecification are acceptable. Our results indicate that models with acceptable misfit can be distinguished using our two-index strategy using the following criteria: a) largest absolute value of the standardized residual covariance $\leq .15$, and b) $\text{SRMR} \leq .10 \times \overline{R}^2$. For convenience, we provide in Table 1 the values of SRMR corresponding to our definitions of close and acceptable fit for various levels of average communality (R^2 of the observed indicators). We are unable to provide values of the biased SRMR corresponding to our definitions of close and acceptable fit because the expected value of the biased SRMR is a function of $1/\sqrt{N}$ (see Appendix 1).

Insert Table 1 around here

It is worth noting that these reference values for close and adequate fit do not have to serve as fixed criteria. The recommendations are based on our personal cutoff for deciding whether to retain or reject a misspecified model. We acknowledge that the definition of closely fitting model is subjective; in practice, researchers may differ in what they consider a closely or acceptably fitting model, and they need not agree with our classification for “close” and “acceptable” fit models. However, we provide enough information for them to make an informed decision on what cutoff to use given their qualitative decision on which models are close enough to be actionable given the true models. That is, different reference values can be obtained from the figures[5], or easily generated by following the paradigm from the current study, which allow researchers to employ a criterion that meets the unique needs of their proposed research.

In the current study, we focus on population parameters. Of course, in applications only sample estimates are available. It is noted that with sampling errors, the reference values we provided are not sample cutoffs. That is, it may not be proper to compare the sample SRMR (or the largest standardized residual covariance) with the population cutoffs, unless the sample size is very large. To account for the sampling variability, we recommend applying the population criteria with the confidence intervals (or statistical tests for model close fit; e.g., SRMR $\leq c$), which allows researchers to make population inference using sample data. Specifically, for both SRMR and an individual standardized residual covariance term, statistical theory is available to obtain their asymptotic standard errors, and, in turn, to compute the confidence interval (Maydeu-Olivares & Shi, & Rosseel, 2017; Maydeu-Olivares & Shi, 2017; Maydeu-Olivares, 2017; Ogasawara, 2001). Simulation studies have also shown that in finite samples, the point estimations and the confidence intervals are quite accurate for both SRMR and the individual standardized residual covariance even in samples of size 100 and very large models (Maydeu-Olivares & Shi & Rosseel, 2018; Maydeu-Olivares & Shi, 2017). Of course, when considering the statistical significance (or confidence intervals) for individual standardized residual it is important to adjust for multiple testing. In our experience (Maydeu-Olivares & Shi, 2017), the

Bonferroni method suffices for this purpose although more complex procedures (e.g., Benjamini & Hochberg, 1995) may be employed. Additional works are need to further explore the usage of individual standardized residual to assess the overall fit of the model. We hope that this paper enlightens SEM researchers and provides additional information to assist them when conducting the difficult task of assessing model-data fit.

Appendix 1:

Computation of the Population SRMR and Bias of Its Unbiased and Biased Estimates

To illustrate the biases of the biased and unbiased estimates of the population SRMR we have provided a simulation example. The population model was an independent cluster confirmatory factor model with two correlated factors (correlation coefficient $\rho = .80$). Each factor had five normally distributed indicators. The population factor loadings were set to be either .80 (error variance = .36) or .40 (error variance = .84). Misspecification was introduced by ignoring the multidimensional structure and fitting a common factor model to the two-factor data. Table 2 provides the mean across 1,000 replications of the biased and unbiased SRMRs obtained across a number of sample sizes (ranging from 50 to 10,000 observations). Maximum likelihood estimation was used. All computations were performed using the *lavaan* package in R (R Development Core Team, 2015; Rosseel, 2012).

Following Maydeu-Olivares (2017), the population SRMR was computed as follows: The population covariance matrix was input into *lavaan* as if it were the sample covariance matrix, along with the target sample size. The estimated standardized residuals were then used to compute the population SRMR. As a side product, we also computed the expected value of the biased sample SRMR given in Eq. 4. The results are summarized in Table 2. In this Table, we see how the behavior of the sample SRMRs depended on the factor loading. More accurate SRMR values were obtained with higher factor loadings, undoubtedly because they were more accurately estimated (Gerbing & Anderson, 1985; Hoogland & Boomsma, 1998). When the factor loading was .4, the unbiased SRMR provided estimates with a relative bias of less than 10% in samples of 100 observations and higher, but a sample size of 1,000 was needed to obtain a relative bias less than 1%. In contrast, the relative bias of the unbiased SRMR estimate did not exceed 1%, even when sample size was 50.

The results of Table 2 also reveal that the biased SRMR overestimated the population SRMR. This suggests that under such conditions, model fits may appear poorer than they actually are. The bias was not exceptionally large when the factor loading was high (.8) and sample size was at least 200. For a population value of .058, the average of the biased SRMR was .063 at a sample size of 200 (the unbiased average was .058). However, when the factor loading was low (.4), the bias of the biased SRMR was unacceptable, even with a sample size of 10,000. Thus, for a population value of .014, the average of the biased SRMR was .069 at a sample size of 200 (the unbiased average was .016).

We also see in Table 2 that the asymptotic approximation of the average behavior of the biased SRMR was rather accurate. At low factor loadings, our approximation was accurate to three digits, even with a sample size of 50. At high factor loadings, three-digit accuracy was obtained as soon as the sample size reached 300. Even in the worst case (50 observations), the approximation was fairly accurate (an expected mean of .074 vs. a .076 actual mean).

Insert Table 2 and Figure 7 around here

Although it is not apparent in Equation (5) –as N is embedded in Ξ_s – the expected value of the biased sample SRMR can be well approximated by a function of $1/\sqrt{N}$ for each value of the population SRMR. This is illustrated graphically in Figure 7.

Appendix 2:

Using the Standardized Root Mean Square Residual and the Largest Standardized Residual Covariance to Identify Close Fitting Models

This appendix evaluates the performance of the two-index strategy and examines the generalizability of the proposed reference values under more complex modeling situations. The population model used was a confirmatory factor analysis (CFA) model with three correlated factors. The total number of observed variables (*p*) included 36, 72, or 144, resulting in 12, 24, or 48 variables per factor. The population variances for all three factors were set to one. The error variances were set such that all factor loadings were on a standardized scale. To reflect a more realistic situation, items loading on the same factor were of different magnitudes. . The size of the factor loadings included low (.40), medium (.60), or high (.80) on the same factor. For a given factor, the number of items with each of the three levels of factor loadings was the same (e.g., $p = 12$, three items of each loading size). Three types of model misspecifications were manipulated:

Misspecified dimensionality. The population model consists of three correlated factors. The fitted models were either a one factor model or a two factor model, collapsing the correlation between factors one and two. The level of misspecification was manipulated by changing the degree of correlation among the factors in the population. When the unidimensional structure was fitted, smaller inter-factor correlation(s) in the population model were indicative of a greater level of misspecification. Seven patterns of inter-factor correlations were included and summarized in Table 3. In total, 42 cases (3 model sizes × 7 inter-factor correlation patterns × 2 fitted models) were considered. Following the definition in this study, closely fitting models imply that a unidimensional structure was fit when the inter-factor correlations were larger than or equal to .90.

Omitting cross-loadings. The population model consists of three correlated factors; however, three indicators loaded on two different factors[6]. The fitted model ignored cross loadings, with values incorrectly fixed to zero. The inter-factor correlations were set to .30, and the level of misspecification was determined by the population value of the omitted cross-loadings. The magnitudes of (standardized) cross-loadings included .10, .20, or .30. We also manipulated the locations of cross-loadings, where the cross-loadings occurred on items with low (.40), medium (.60), or high (.80) primary factor loadings. The number of conditions considered was 27 = 3 (model size levels) × 3 (magnitudes of omitted cross-loadings) × 3 (locations of cross-loadings). Closely fitting models were defined as omitting cross-loadings with standardized values less than or equal to .10.

Omitting residual correlations. In the population model, three correlated residuals were present[7]. A simple structure three-factor CFA model with no correlated residuals was fit. The correlated residuals were introduced between items with either low (.40), medium (.60), or high (.80) factor loadings. The level of model misspecification was indicated by the values of residual correlations (i.e. .10, .20, or .30); larger values implied a higher level of model misspecification. Under omitting residual correlations, the total number of conditions was 27 = 3 (model size levels) × 3 (magnitudes of omitted residual correlations) × 3 (locations of correlated residuals). If correlated residuals with correlations less than or equal to .10 were ignored, the (misspecified) model is considered closely fitting.

Following the same procedure discussed in the earlier section, for each simulated condition, we computed the population SRMR and the single standardized residual covariance.

The ratio of SRMR to average communality (i.e., the squared standardized factor loading) and the largest absolute value of standardized residual covariance were calculated and used as the two indices for examining model close fit. In Figure 7, for each type of model misspecification, we presented the relationship between the SRMR/λ^2 and the largest standardized residual covariance, along with the recommended reference values for model close fit (i.e. indicated by the solid horizontal line and the dashed vertical line, respectively). In addition, markers were used to differentiate cases that would be denoted as a close fitting model, and those that are not (based on the definitions discussed above). As shown in the figures, the proposed reference values performed well in terms of identifying close fitting models. All close fitting models yielded the largest absolute value of standardized residual covariance less than or equal to .10, and SRMR/λ^2 less than or equal to .05 (i.e. fall in the left lower quadrant created by the two reference lines). On the other hand, by using the two-index cutoffs, almost all models with more severe misspecifications could be successfully identified.

Insert Table 3 and Figure 8 around here

Footnotes

1. In this paper, we distinguish between effect sizes of model misfit and goodness of fit indices. The term goodness of fit indices is reserved for sample statistics used to adjudge model fit disregarding their sampling variability and without referencing the population parameter.
2. When interpreting the standardized residual covariances, we ignore the signs and refer the "largest standardized residual covariances" as the one with the largest *absolute value*.
3. The size of an SEM model has been indicated by different indices, including the number of observed variables (p), the number of parameters to be estimated (q), the degrees of freedom ($df = p\ (p + 1)/2 - q$), and the ratio of the observed variables to latent factors (p/f). Recent studies have suggested that the number of observed variables (p) is the most important determinant of model size effects (Moshagen 2012; Shi, Lee & Terry, 2015, 2017). Therefore, in the current study, we define large models as SEM models with many observed variables.
4. To compute the ratio of SRMR to communality, we used the average (standardized) factor loadings estimated by fitting the misspecified models to the population covariance matrices.
5. Also see the supplemental tables.
6. The three indicators loaded on factors 1 & 2, factors 1 & 3, and factors 2 & 3, respectively.
7. The three correlated residuals were introduced between items from factors 1 & 2, factors 1 & 3, and factors 2 & 3, respectively.

References

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Beauducel, A., & Wittmann, W. W. (2005). Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(1), 41–75. https://doi.org/10.1207/s15328007sem1201_3

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, *57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). EQS 5 [Computer Program]. Encino, CA: Multivariate Software Inc.

Box, G. E. P. (1979). Some Problems of Statistics and Everyday Life. *Journal of the American Statistical Association*, *74*(365), 1–4. https://doi.org/10.1080/01621459.1979.10481600

Browne, M. W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*(2), 207–212. https://doi.org/10.1111/j.2044-8317.1972.tb00492.x

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., & 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.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research*, *36*(4), 462–494. https://doi.org/10.1177/0049124108314720

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). Manual of the Social Problem-Solving Inventory-Revised. North Tonawanda, NY: Multi-Health Systems, Inc.

DiStefano, C. (2016). Using Fit Indices in Structural Equation Modeling. In *Principles and Methods of Test Construction: Standards and Recent Advancements.* (pp.166-196). K. Schweizer & C. DiStefano (Eds.) Göttingen, Germany: Hogrefe Publishers.Edwards, M. C. (2013). Purple Unicorns, True Models, and Other Things I've Never Seen. *Measurement: Interdisciplinary Research & Perspective*, *11*(3), 107–111. https://doi.org/10.1080/15366367.2013.835178

Fan, X., & Sivo, S. A. (2005). Sensitivity of Fit Indexes to Misspecified Structural or Measurement Model Components: Rationale of Two-Index Strategy Revisited. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 343–367. https://doi.org/10.1207/s15328007sem1203_1

Fan, X., & Sivo, S. A. (2007). Sensitivity of Fit Indices to Model Misspecification and Model Types. *Multivariate Behavioral Research*, *42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are Fit Indices Really Fit to Estimate the Number of Factors With Categorical Variables ? Some Cautionary Findings via Monte Carlo Simulation. *Psychological Methods*, *21*(1), 93–111. https://doi.org/10.1037/met0000064

Gerbing, D. W., & Anderson, J. C. (1985). The Effects of Sampling Error and Model

Characteristics on Parameter Estimation for Maximum Likelihood Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *20*(3), 255–271. https://doi.org/10.1207/s15327906mbr2003_2

Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement*, *71*(2), 306-324.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological methods*, *16*(3), 319-336.

Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling. *Sociological Methods & Research*, *26*(3), 329–367. https://doi.org/10.1177/0049124198026003003

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. https://doi.org/10.1037//1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jöreskog, K. G., & Sörbom, D. (1988). LISREL 7. A guide to the program and applications (2nd ed.). Chicago, IL: International Education Services.

MacCallum, R. C. (2003). Working with Imperfect Models. *Multivariate Behavioral Research*, *38*(1), 113–139. https://doi.org/10.1207/S15327906MBR3801_5

Maiti, S. S., & Mukherjee, B. N. (1990). A note on distributional properties of the Jöreskog-

Sörbom fit indices. *Psychometrika*, *55*(4), 721–726. https://doi.org/10.1007/BF02294619

Maydeu-Olivares, A. (2017). Assessing the Size of Model Misfit in Structural Equation Models. *Psychometrika*, *82*(3), 533–558. https://doi.org/10.1007/s11336-016-9552-7

Maydeu-Olivares, A., & D'Zurilla, T. J. (1996). A factor-analytic study of the Social Problem-Solving Inventory: An integration of theory and data. *Cognitive Therapy and Research*, *20*(2), 115–133. https://doi.org/10.1007/BF02228030

Maydeu-Olivares, A., & Joe, H. (2014). Assessing Approximate Fit in Categorical Data Analysis. *Multivariate Behavioral Research*, *49*(4), 305–328. https://doi.org/10.1080/00273171.2014.911075

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. https://doi.org/10.1027/1614-2241/a000129

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(3), 389-402.

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of personality assessment*, *100*(1), 43-52.

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

Moshagen, M. (2012). The Model Size Effect in SEM: Inflated Goodness-of-Fit Statistics Are Due to the Size of the Covariance Matrix. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(1), 86–98. https://doi.org/10.1080/10705511.2012.634724

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

R Development Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Raykov, T. (2000). On Sensitivity of Structural Equation Modeling to Latent Relation Misspecifications. *Structural Equation Modeling*, *7*(4), 596–607. https://doi.org/10.1207/S15328007SEM0704_4

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://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. https://doi.org/10.1080/10705510903203433

Savalei, V. (2012). The Relationship Between Root Mean Square Error of Approximation and Model Misspecification in Confirmatory Factor Analysis Models. *Educational and Psychological Measurement*, *72*(6), 910–932. https://doi.org/10.1177/0013164412452564

Shi, D., Lee, T., & Terry, R. A. (2015). Abstract: Revisiting the Model Size Effect in Structural Equation Modeling (SEM). Multivariate Behavioral Research, 50(1), 142–142. http://doi.org/10.1080/00273171.2014.989012

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 25(1), 21-40. https://doi.org/10.1080/10705511.2017.1369088

Shi, D., Lee, T., & Maydeu-Olivares, A. (2018). Understanding the Model Size Effect on SEM Fit Indices. *Manuscript submitted for publicaiton*.

Steiger, J. H. (1989). EzPATH: A supplementary module for SYSTAT and SYGRAPH. Evanston, IL: Systat, Inc.

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), Handbook of Structural Equation Modeling. New York: Guilford Press.

Yuan, K.-H. (2005). Fit Indices Versus Test Statistics. *Multivariate Behavioral Research*, *40*(1), 115–148. https://doi.org/10.1207/s15327906mbr4001_5

Table 1: SRMR reference values that meet the criteria for close ($SRMR \leq .05 \times \bar{R}^2$) and adequate fit ($.05 \times \bar{R}^2 < SRMR \leq .10 \times \bar{R}^2$) for selected values of average communality ($\bar{R}^2$)

$\bar{R}^2$	SRMR reference value for close fit	SRMR reference value for adequate fit
0.010	0.001	0.001
0.040	0.002	0.004
0.090	0.005	0.009
0.160	0.008	0.016
0.250	0.013	0.025
0.360	0.018	0.036
0.490	0.025	0.049
0.640	0.032	0.064

Table 2: Population SRMR and average sample Estimates

N	**Factor Loading = 0.8**				**Factor Loading = 0.4**			
	P.SRMSR	E[SRMSRb]	SRMSR$_b$	SRMSR$_u$	P.SRMSR	E[SRMSRb]	SRMSR$_b$	SRMSR$_u$
50	.058	.074	.076	.057	.014	.097	.097	.021
100	.058	.066	.067	.058	.014	.069	.069	.016
200	.058	.062	.063	.058	.014	.050	.050	.014
300	.058	.061	.061	.058	.014	.042	.042	.013
400	.058	.060	.060	.058	.014	.037	.037	.013
500	.058	.059	.060	.058	.014	.034	.034	.013
600	.058	.059	.059	.058	.014	.031	.031	.013
700	.058	.059	.059	.058	.014	.029	.030	.013
800	.058	.059	.059	.058	.014	.028	.028	.013
900	.058	.059	.059	.058	.014	.027	.027	.014
1000	.058	.058	.059	.058	.014	.026	.026	.014
2000	.058	.058	.058	.058	.014	.021	.021	.014
5000	.058	.058	.058	.058	.014	.017	.017	.014
10000	.058	.058	.058	.058	.014	.016	.016	.014

Table 3: Patterns of inter-factor correlations considered in appendix 2.

Inter-factor Correlations			**Close Fit?**	
$\boldsymbol{\rho_{12}}$	ρ_{23}	ρ_{13}	1F CFA	2F CFA
0.9	0.9	0.9	Yes	Yes
0.9	0.9	0.8	No	Yes
0.9	0.9	0.7	No	Yes
0.9	0.8	0.8	No	Yes
0.9	0.8	0.7	No	Yes
0.8	0.8	0.8	No	No
0.7	0.7	0.7	No	No

Figure 1: Behavior of (A) the population SRMR, (B) largest standardized residual covariance, and (C) SRMR/communality in models with misspecified dimensionality

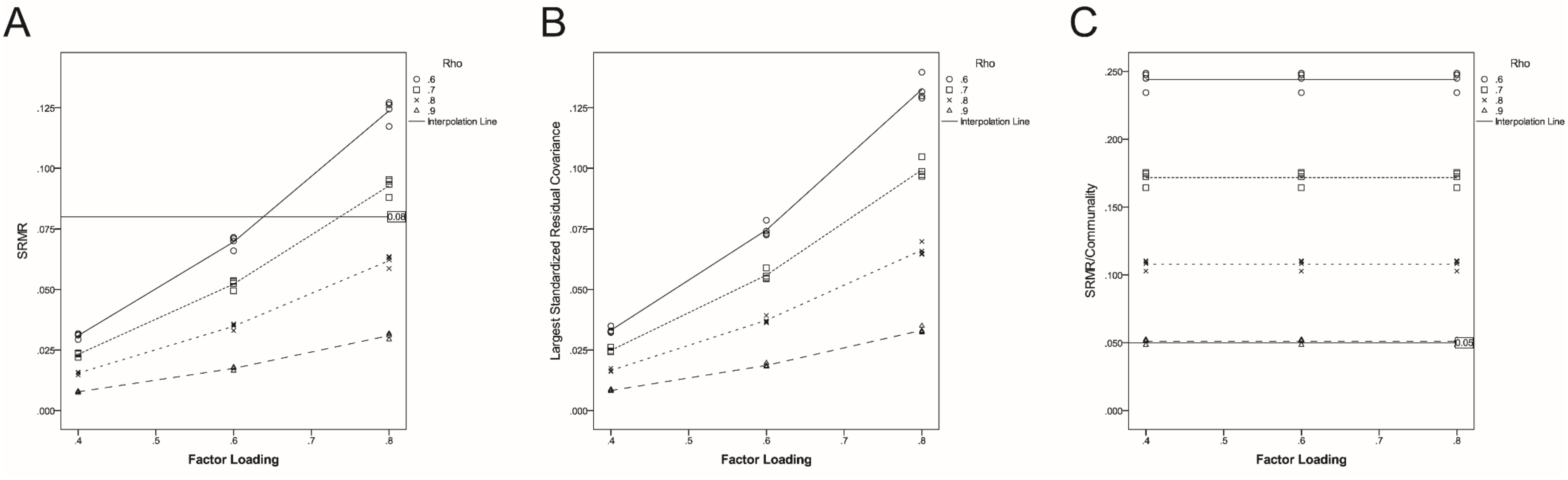

Figure 2: Behavior of the population SRMR in models with one (A) to four (D) omitted cross-loadings

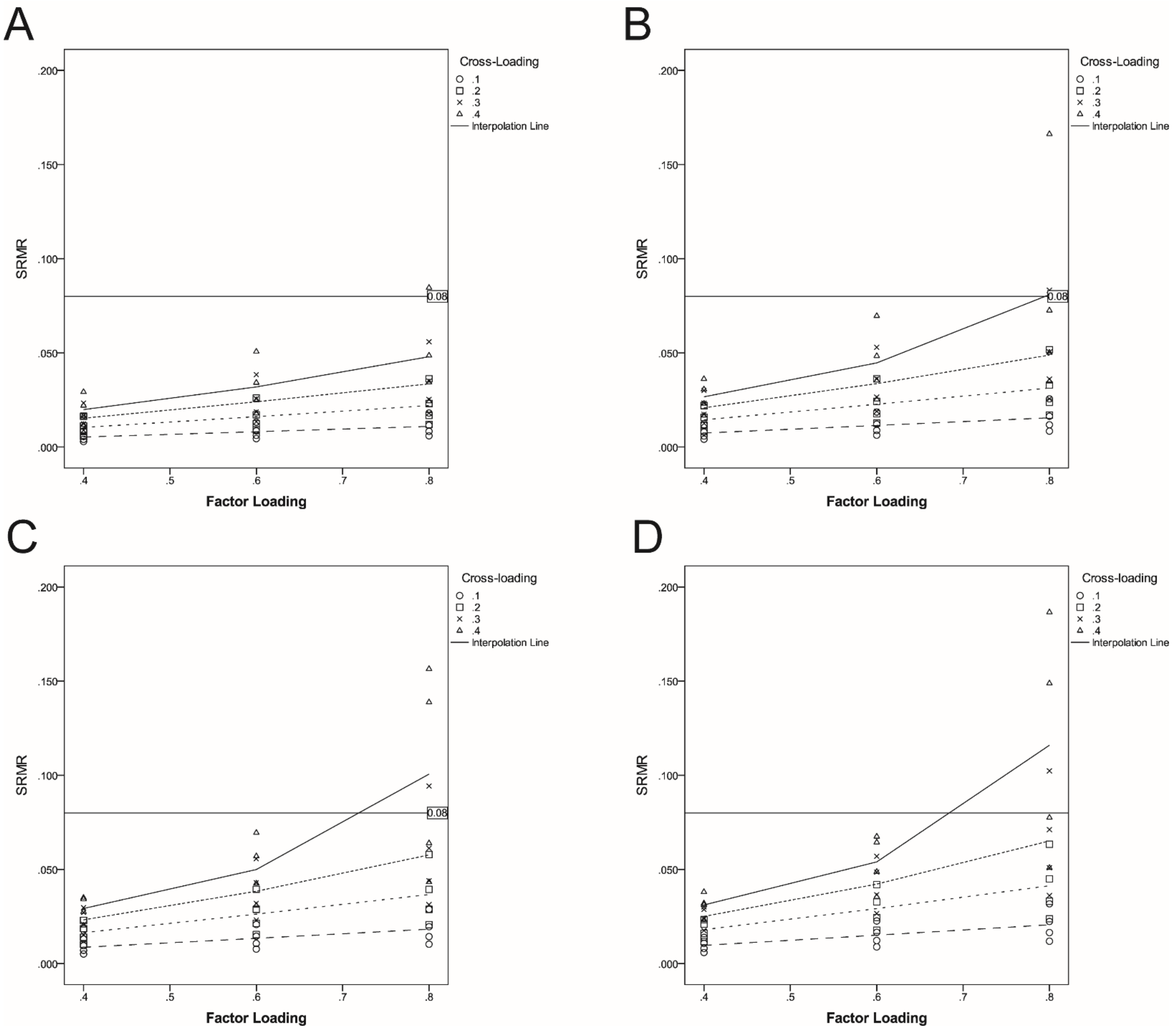

Figure 3: Behavior of SRMR/communality in models with one (A) to four (D) omitted cross-loadings

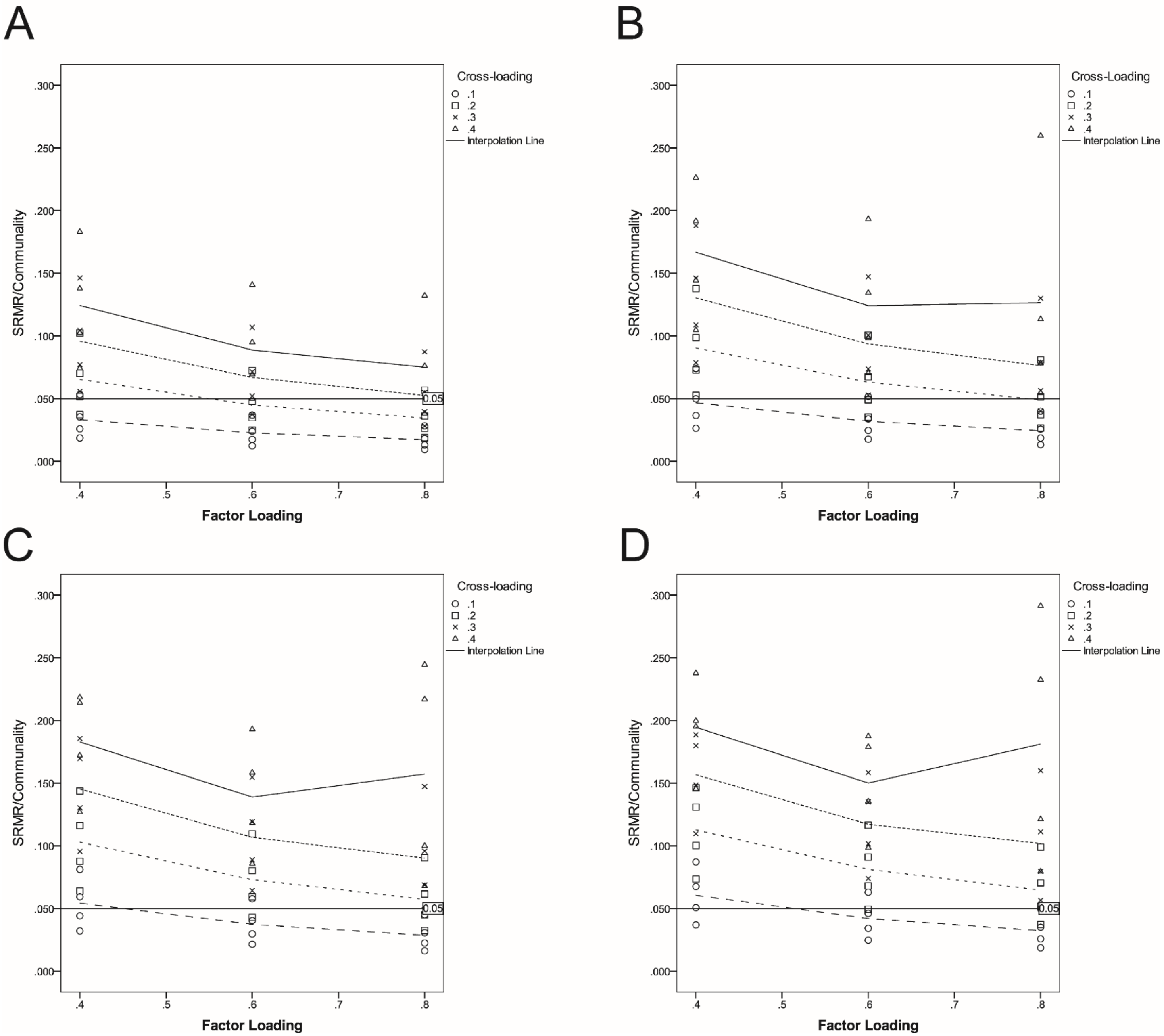

Figure 4: Behavior of the largest standardized residual covariance in models with one (A) to four (D) omitted cross-loadings

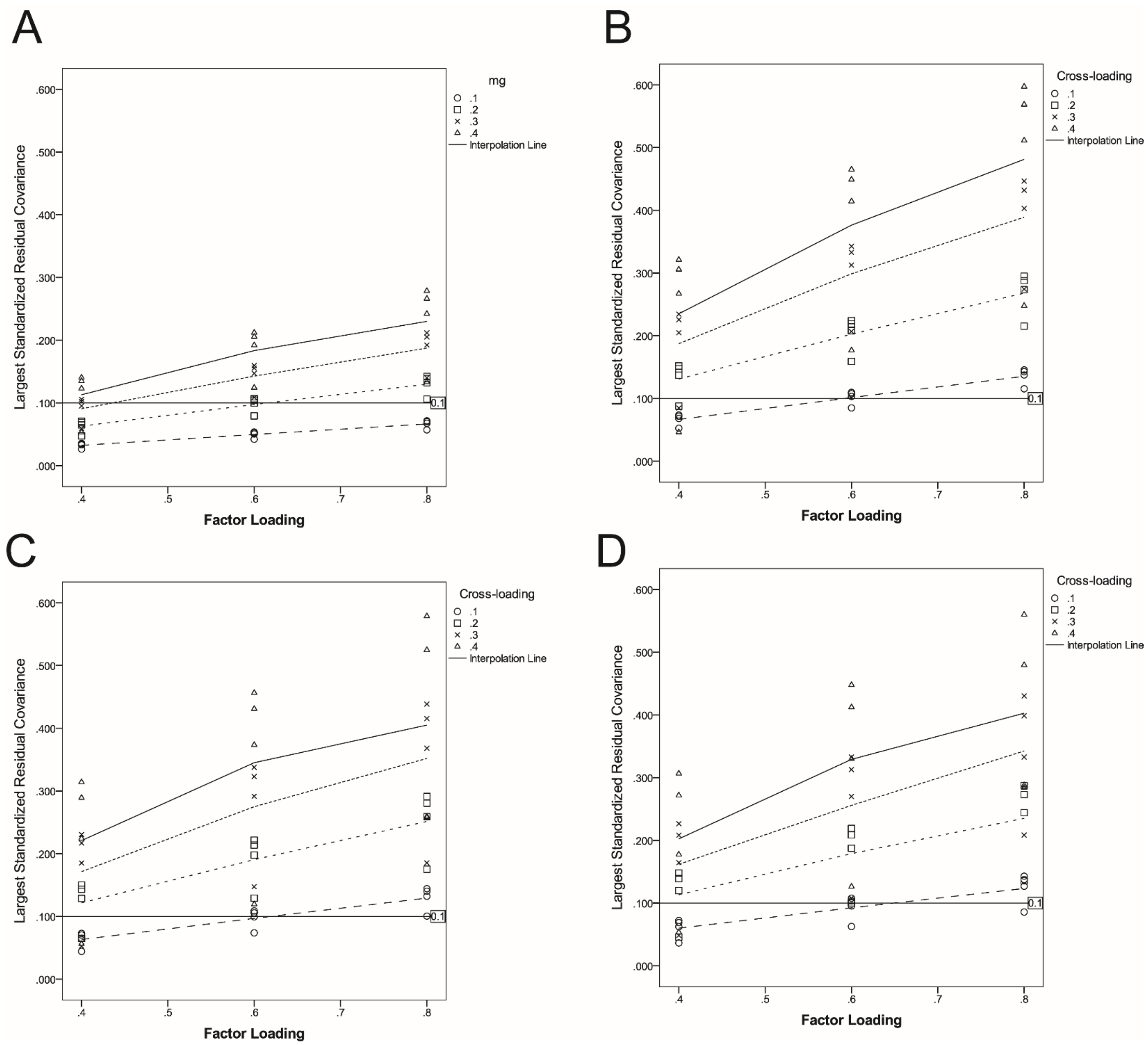

Figure 5: Behavior of the population SRMR in models with one (A) to four (D) omitted residual correlations

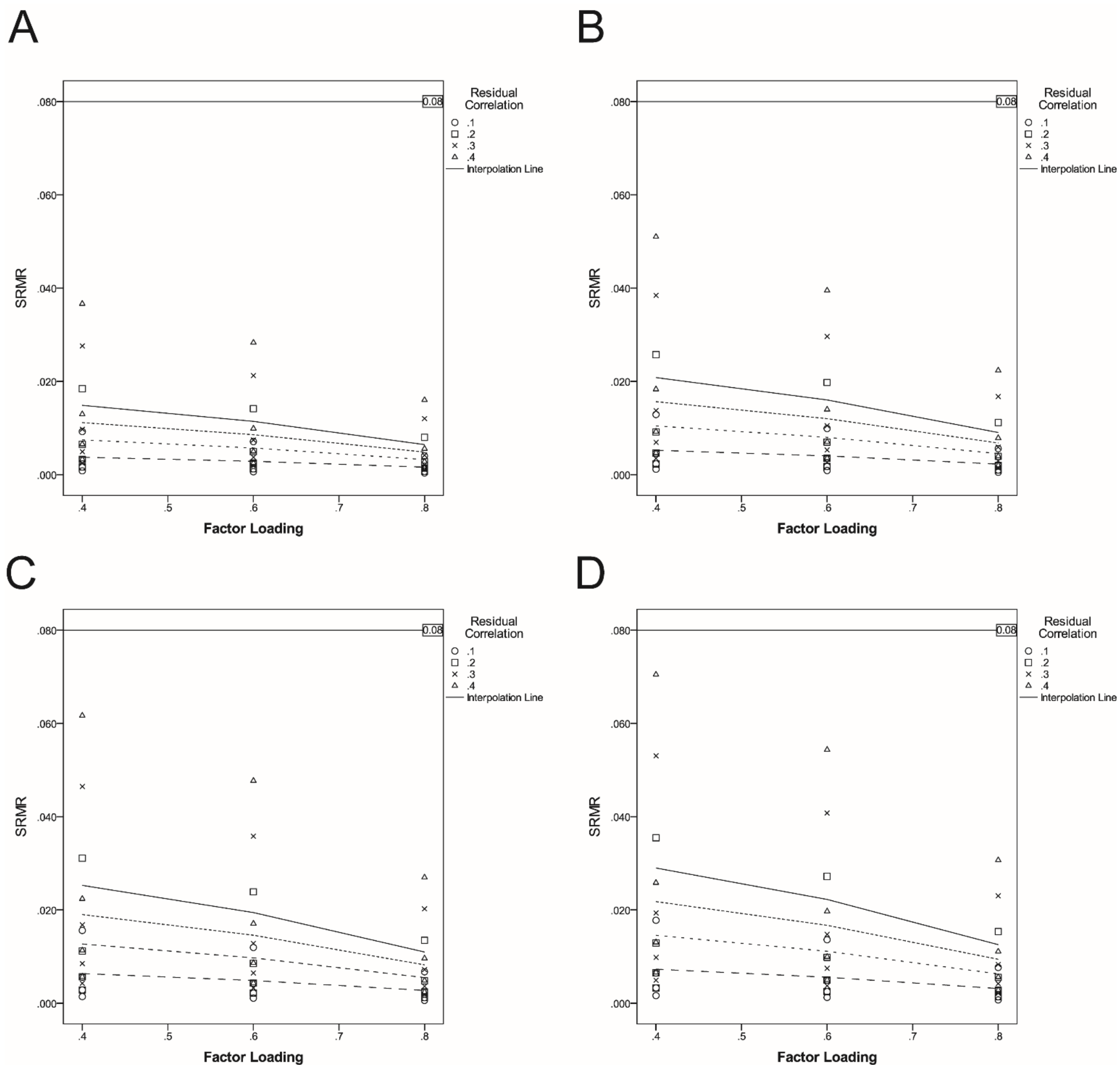

Figure 6: Behavior of the largest standardized residual covariance in models with one (A) to four (D) omitted residual correlations

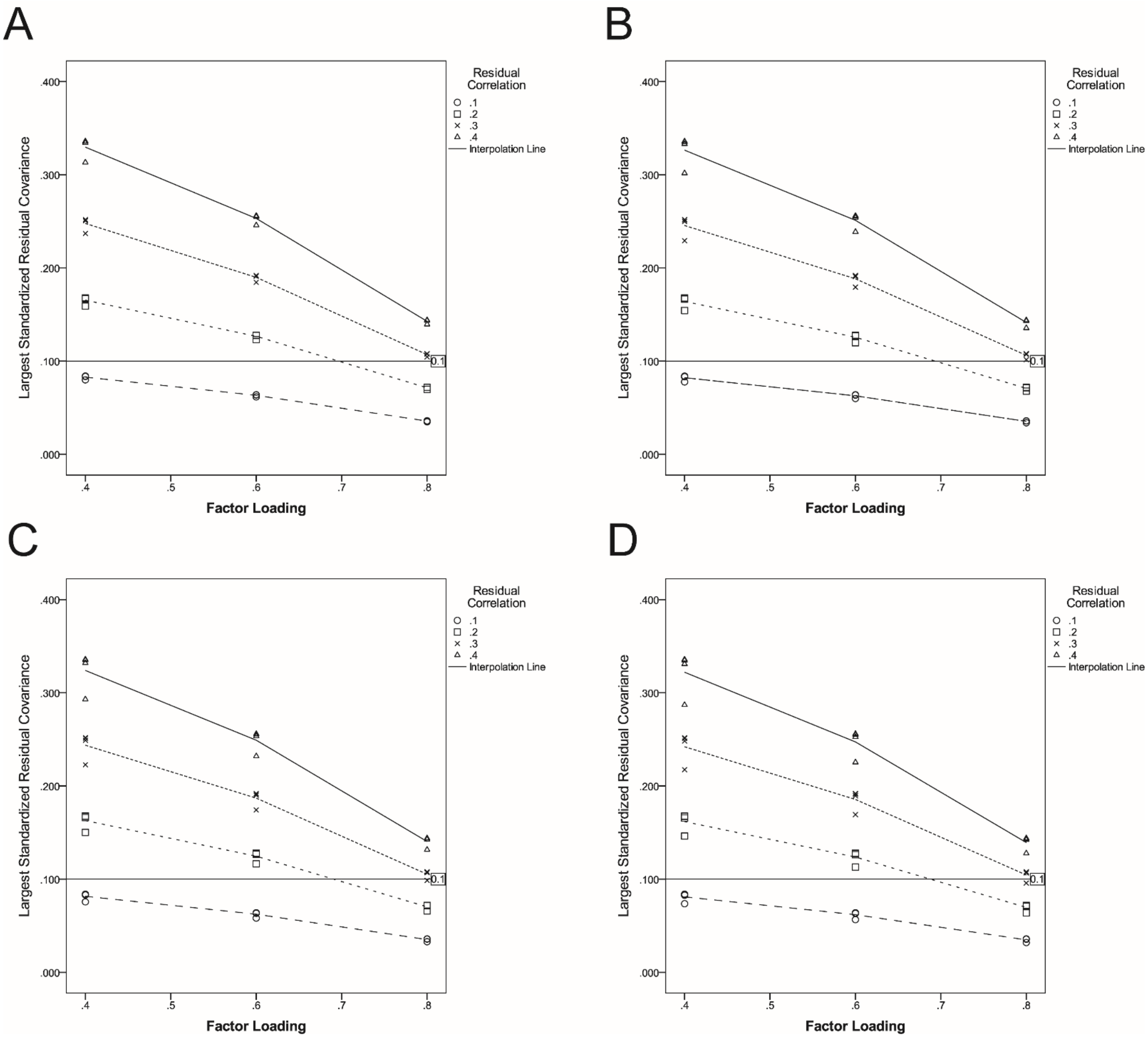

Figure 7: Relationship between the expected value of the biased sample SRMR, population SRMR, and sample size (*N*)

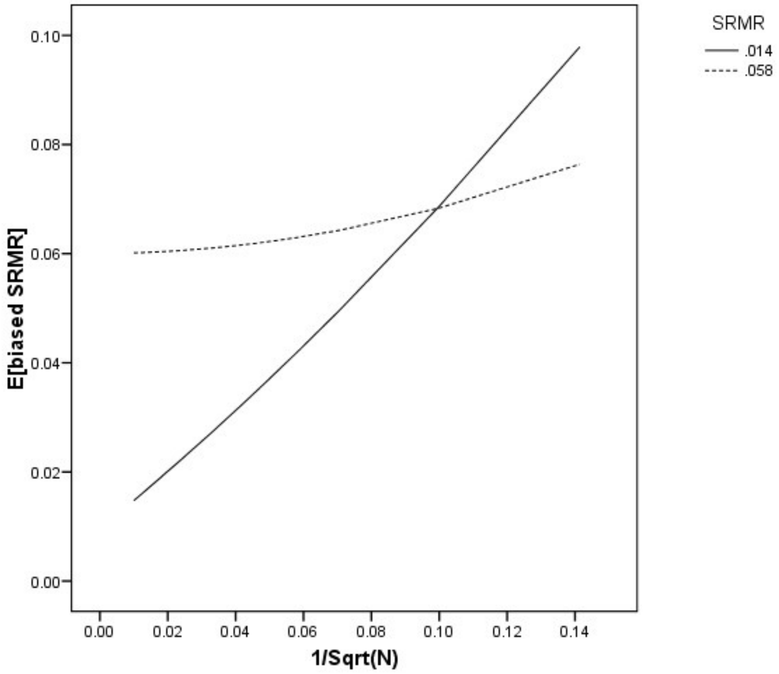

Figure 8: Using the two-index strategy for identifying close fitting models