Schmitt, M. J., & Steyer, R. (1993b). Mother–daughter attachment and family cohesion: Latent trait–state models of current and retrospective perceptions. In R. Steyer, H. Graser, & K. Widaman (Eds.), *Consistency and specificity: Latent state–trait models in differential psychology* (pp. 21–42). New York: Springer.

Speilberger, C. D. (1985). Anxiety, cognition, and affect: A state–trait perspective. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 171–182). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Steyer, R., Majcen, A. M., Schwenkmezger, P., & Buchner, A. (1989). A latent trait–state anxiety model and its application to determine consistency and specificity coefficients. *Anxiety Research, 1,* 281–299.

Steyer, R., & Schmitt, M. J. (1994). The theory of confounding and its application in causal modeling with latent variables. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 36–67). Thousand Oaks, CA: Sage.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait–multimethod data. *Applied Psychological Measurement, 9,* 1–16.

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of Research in Education: Vol. 15* (pp. 345–422). Washington, DC: American Educational Research Associations.

Willett, J. B., & Sayer, G. S. (1996). Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 159–194). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Windle, M. (1996). An alcohol involvement topology for adolescents: Convergent validity and longitudinal stability. *Journal of Studies on Alcohol, 57,* 627–637.

Windle, M., & Miller-Tutzauer, C. (1992). Confirmatory factor analyses and concurrent validity of the perceived social support–family measure among adolescents. *Journal of Marriage and the Family, 54,* 777–787.

Wolins, L. (1982). *Research mistakes in the social and behavioral sciences.* Ames: Iowa State University Press.

Wood, P., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin, 116,* 166–186.

Wothke, W. (1996). Models for multitrait–multimethod analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 7–56). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zucker, R. A., & Barron, F. H. (1973). Parental behaviors associated with problem drinking and antisocial behavior among adolescent males. In M. E. Chafetz (Ed.), *Proceedings of the First Annual Alcoholism Conference of the National Institute on Alcohol Abuse and Alcoholism: Research on alcoholism: I. Clinical problems and special populations* (pp. 276–296). Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

STRUCTURAL EQUATION MODELING, 5(4), 411–419

TEACHER'S CORNER

A Note on Multiple Sample Extensions of the RMSEA Fit Index

James H. Steiger
Department of Psychology
University of British Columbia

Generalization of the Steiger–Lind root mean square error of approximation fit indexes and interval estimation procedure to models based on multiple independent samples is discussed. In this article, we suggest an approach that seems both reasonable and workable, and caution against one that definitely seems inappropriate.

Steiger and Lind (1980) introduced the root mean square error of approximation (RMSEA) fit index for evaluating covariance structure models. Steiger (1989, 1990, 1994, 1995) presented a detailed theoretical rationale for this index. For additional discussion and applications, see Browne and Cudeck (1993).

Here we briefly sketch the background for the RMSEA. Suppose that the sample covariance matrix S based on $N = n + 1$ observations has been fit to a model $M(\theta)$ that is a function of a vector of t parameters θ, and that the fitting has been accomplished by minimizing a discrepancy function $F(S,M(\theta))$ under choice of θ.

Define $\hat{\theta}$ as the minimizer of the discrepancy function, that is,

$$\hat{\theta} = \text{Arg Min}_{\theta} F(S, M(\theta)) \tag{1}$$

Under the standard regularity conditions, if $F(S,M(\theta))$ satisfies certain restrictions, the quantity $nF(S, M(\hat{\theta}))$ will have a large sample distribution that is closely approximated by $\chi^2_{\nu,\lambda}$ (i.e., a noncentral chi-square variate with ν degrees of freedom,

Requests for reprints should be sent to James H. Steiger, Department of Psychology, 2136 West Mall, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. E-mail: steiger@unixg.ubc.ca

and noncentrality parameter λ). If the parameters are unrestricted (i.e., mathematically independent and variable), then $\nu = p(p+1)/2 - t$. The noncentrality parameter is

$$\lambda = nF^* = nF(\Sigma, M(\hat{\theta})) \tag{2}$$

The noncentrality parameter is thus n times the *population discrepancy function*, $F^* = F(\Sigma, M(\hat{\theta}))$, which is the value of the discrepancy function one would obtain if one actually knew the population covariance matrix Σ, and fit the model to it.

Define

$$s = \text{vecs}(S) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{22} \\ s_{31} \\ \vdots \\ s_{pp} \end{bmatrix} \tag{3}$$

to be a vector composed of the $p(p+1)/2$ nonredundant (lower triangular) elements of S. Likewise, define $\hat{\sigma} = \text{vecs}(M(\hat{\theta}))$ and $e = s - \hat{\sigma}$. e is the vector of discrepancies between the observed covariance matrix and the covariance matrix reproduced by the model and the parameter estimates. If the discrepancy function belongs to the class of generalized least squares discrepancy functions, it may be written in the form

$$F(S, M(\theta)) = e'W^{-1}e \tag{4}$$

where W is an appropriately chosen weight matrix. This form shows that the discrepancy function is a weighted sum of squared discrepancies. Because the chi-square statistic may be written

$$X = ne'W^{-1}e = e'G^{-1}e \tag{5}$$

it also is a weighted sum of squared discrepancies, where

$$G = \frac{W}{n} \tag{6}$$

We can obtain generalized least squares estimates by minimizing either X or $F(S,M(\theta))$. The basic form for the RMSEA in the population is

$$R^* = \sqrt{\frac{F^*}{\nu}} \tag{7}$$

where F^* is the population discrepancy function, calculated by fitting the covariance structure model to the population covariance matrix Σ. The degree of freedom parameter for the model is ν.

Because, with generalized least squares discrepancy functions, F^* is a weighted sum of squared discrepancies, R^* can be interpreted as a root mean square standardized measure of badness of fit of a particular model to Σ.

The RMSEA index has a number of advantages (Steiger, 1989, 1995). Particularly important is the fact that a coherent estimation strategy exists—both a point estimate and a confidence interval are available. As a result, the problems and paradoxes inherent in testing models with large sample sizes are reduced. However, the RMSEA is not a panacea, and should be considered a helpful tool for guiding complicated judgments about model utility, rather than a substitute for such judgments. In particular, rationales that use specific numerical RMSEA "cutoff" values (such as .05) for determining adequacy of model fit may not be able to stand up to careful logical scrutiny.

At this writing, a number of structural modeling software packages calculate and report the RMSEA index. The Steiger and Lind (1980) presentation dealt only with the case where the model concerned only one population, and was tested only on a single sample of size N. Here, we discuss the extension of the RMSEA to more than one sample. We discover that this extension is not entirely straightforward. Indeed, simply applying a well-known computational formula for the one-sample RMSEA to the chi-square statistic from a K-sample analysis can yield a measure of fit that is clearly incorrect.

A MULTIPLE SAMPLE RMSEA

For simplicity, we refer here to the two-sample case, but the generalization from two to K groups is immediate and obvious. Suppose there are two distinct populations, P_1 and P_2, and we take samples of size N_1 and N_2 on random vectors y_1 and y_2 of order $p_1 \times 1$ and $p_2 \times 1$, respectively. The random vectors y_1 and y_2 have corresponding population covariance matrices and Σ_1 and Σ_2. Then the combined random vector

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{8}$$

has covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{9}$$

The sample covariance matrices, S_1 and S_2 and from the two samples, may be combined into one matrix

$$S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \tag{10}$$

The overall model $M(\theta)$ may be partitioned into submodels, that is,

$$M(\theta) = \begin{bmatrix} M_1(\theta) & 0 \\ 0 & M_2(\theta) \end{bmatrix} \tag{11}$$

In this notational system, generalization of the single sample equations to the two sample case is straightforward. Specifically, we have

$$\begin{aligned} X &= e'G^{-1}e = [e_1' \; e_2'] \begin{bmatrix} G_1^{-1} & 0 \\ 0 & G_2^{-1} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \\ &= [e_1' \; e_2'] \begin{bmatrix} n_1 W_1^{-1} & 0 \\ 0 & n_2 W_2^{-1} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \\ &= n_1 e_1' W_1^{-1} e_1 + n_2 e_2' W_2^{-1} e_2 \end{aligned} \tag{12}$$

Alternatively, letting

$$n_\bullet = n_1 + n_2 \tag{13}$$

and

$$c_i = n_i / n_\bullet \tag{14}$$

we may write

$$\begin{aligned} X &= n_\bullet F(S, M(\hat{\theta})) \\ &= n_\bullet (c_1 e_1' W_1^{-1} e_1 + c_2 e_2' W_2^{-1} e_2) \\ &= n_\bullet (c_1 F(S_1, M_1(\hat{\theta})) + c_2 F(S_2, M_2(\hat{\theta}))) \end{aligned} \tag{15}$$

With X and F expressed in this form, as in the single sample case, we may obtain our estimates, and the related chi-square statistic, by minimizing either X or F under choice of θ. X has, at large samples, a distribution that is approximately $\chi^2_{\nu,\lambda}$, where

$$\nu = p_1(p_1 + 1)/2 + p_2(p_2 + 1)/2 - t \tag{16}$$

and

$$\lambda = n_\bullet F^* \tag{17}$$

where

$$F^* = c_1 F(\Sigma_1, M_1(\hat{\theta})) + c_2 F(\Sigma_2, M_2(\hat{\theta})) \tag{18}$$

Just as in the single sample case (see Steiger & Lind, 1980; McDonald, 1989), we may estimate F^* by estimating λ with one of several available techniques, then dividing by $n_\bullet$. A point estimate is particularly easy to obtain (McDonald, 1989). Because, for a noncentral chi-square variate,

$$E(X) = \nu + \lambda \tag{19}$$

we have the approximate large sample-result

$$E\left(\frac{X - \nu}{n_\bullet}\right) = \frac{\lambda}{n_\bullet} = F^* \tag{20}$$

The estimate is

$$\hat{F}^* = (X - \nu)/n_\bullet \tag{21}$$

In practice the aforementioned estimate is usually modified to eliminate negative values. Specifically, one uses

$$\hat{F}^+ = \max\{(X - \nu)/n_\bullet, 0\} \tag{22}$$

Interval estimates are obtainable using an iterative technique described by Steiger and Lind (1980), Browne and Cudeck (1993), and Steiger (1995). The $100(1-\alpha)\%$ confidence limits for the noncentrality parameter λ of a $\chi^2_{\nu,\lambda}$ distribution are obtained by finding (via quasi-Newton iteration) the values λ that place the observed value of the chi-square statistic at the $100(\alpha/2)$ and $100(1-\alpha)$ percentile points of a $\chi^2_{\nu,\lambda}$ distribution. Once a confidence interval for $\lambda = n_\bullet F^*$ is found, one simply divides the endpoints by $n_\bullet$ to obtain a confidence interval for F^*.

Note from Equation 18 that F^* is a weighted average of the discrepancies obtained by fitting each population covariance matrix to the model specific to that population, and that the weights sum to 1. Because there are population (and sample) specific quantities corresponding to the covariance matrix and discrepancy

function, it is important to be careful generalizing from the single sample case to that involving two or more samples. Specifically, although we can calculate a discrepancy function for each sample, such a discrepancy function, when multiplied by the n for that sample, need not have a χ^2 distribution if models for the different populations involve the same elements of θ. If each submodel involves completely different subsets of the elements of θ, then of course fitting the multiple group model is precisely the same as fitting each model individually to its covariance matrix. An example of this would be fitting the same confirmatory factor model to two groups, but allowing all parameters to be different across groups. In this case, the overall discrepancy function would be simply the (weighted) sum of the discrepancy functions obtained by fitting each model separately to each covariance matrix. The special case where models involve nonintersecting subsets of θ is not particularly interesting, though it does shed some light on how not to compute a multiple sample RMSEA.

In constructing an appropriate K-sample analog of the RMSEA, it is important to realize that, in general, there is no sample-specific degree of freedom parameter. The degrees of freedom for a multiple sample model are calculated on the basis of the total number of variances and covariances and the total number of free parameters estimated. Hence, in the general case it makes no sense to compute a set of "sample specific" RMSEAs, and then compute some sort of weighted average of them. For the population K-sample RMSEA one must take the overall population discrepancy function (which is a *weighted average* of the sample-based discrepancies), divide by the *average number of degrees of freedom per sample*, then take the square root to obtain a root-mean-square measure. This yields the following definition.

$$\text{RMSEA} = \sqrt{F^*/(\nu / K)} = \sqrt{K}\sqrt{\frac{F^*}{\nu}} \tag{23}$$

One may construct a point estimate for the RMSEA by substituting $\hat{F}^*$ *or* $\hat{F}^+$ in Equation 21 for F^* in Equation 23. With the latter, one obtains

$$\text{RMSEA} = \sqrt{K}\sqrt{\frac{\text{Max}\{X / \nu - 1, 0\}}{n_.}} \tag{24}$$

Interval estimates for the RMSEA are obtained by inserting the endpoints for a confidence interval for F^* in Equation 23.

Equation 23 includes the well-known single sample formula as a special case.

DISCUSSION

Two aspects of Equation 23 are particularly noteworthy. First, F^* is a weighted average of the discrepancies in the individual populations, where the weighting is a function of the sample sizes in the current analysis. When sample sizes are equal, this is equivalent to the unweighted average. When sample sizes differ substantially, the weights applied to the discrepancy in each population will also differ, and the question naturally arises as to whether such weighting is appropriate. The answer to this question is a function of experimenter preference, the type of model employed, and the relative sizes of the experimental populations being examined. The issue essentially vanishes when sample sizes are identical or close to it.

Second, employing the analog of the single-sample formula, say

$$\text{RMSEA} = \sqrt{\frac{\text{Max}\{X / \nu - 1, 0\}}{n_.}} \tag{25}$$

to compute a K-sample RMSEA from the chi-square statistic will yield an incorrect result, that is, all values will be off by $\sqrt{K}$. Such an error may in fact be present in some release versions of distributed software. If so, it may be corrected easily without upgrading the software, simply by multiplying all reported values by $\sqrt{K}$. Testing whether such a correction is necessary is straightforward, as described here.

1. Construct two identical arbitrary data sets (random numbers will suffice).
2. Test one sample with a simple model, for example, a single factor model, and record the RMSEA value.
3. Construct a two-sample model, where each group is tested with the same model used in Step 2, but with no parameters constrained to be equal across populations.
4. The two-sample RMSEA reported by the software should of course be identical to the one-sample value, because two (in theory) completely independent samples have yielded identical fit to two independent versions of the same model. If it is not, check whether multiplying the value by $\sqrt{2}$ yields an identical value. If it does, then it is highly likely that the software is generating incorrect values by using the single-sample formula inappropriately.
5. If the test in Step 4 confirms that an error is present, multiply all point estimates and interval estimates of the RMSEA by $\sqrt{K}$ to obtain correct values.

NUMERICAL EXAMPLE

Suppose the 4 x 4 covariance matrices shown in Tables 1 and 2 were obtained for two independent samples of size 100 each. Suppose further that a one-factor common factor model were fit to each of the two covariance matrices simultaneously by

TABLE 1
Sample Covariance Matrix: Group 1

	VAR1	*VAR2*	*VAR3*	*VAR4*
VAR1	2.000	.800	.600	.800
VAR2	.800	2.000	.600	.800
VAR3	.600	1.200	2.000	.700
VAR4	.800	.800	.700	2.000

TABLE 2
Sample Covariance Matrix: Group 2

	VAR1	*VAR2*	*VAR3*	*VAR4*
VAR1	4.000	1.320	.760	.800
VAR2	1.320	1.000	.600	.400
VAR3	.760	.600	1.000	.170
VAR4	.800	.400	.170	1.000

the method of maximum likelihood, and the loadings were not constrained to be the same across samples. In this case, the overall discrepancy function is the average of the two discrepancy functions that are obtained when the two samples are fit separately, using ordinary single sample methods.

In this case, if a single factor model is fit to the data in Table 1 with the method of maximum likelihood using SEPATH, a maximum likelihood (ML) discrepancy function of .049599 and an iteratively reweighted generalized least squares discrepancy function of 0.050071 are obtained, and the point estimate for the RMSEA (calculated from the IRGLS discrepancy function) is .122206. For the data in Table 2, the corresponding values are .047936, .049066, and .120134. When a simultaneous model is fit to the two samples, one obtains discrepancy functions of .048767 and .049569, and an RMSEA point estimate of .121174. If one calculates the RMSEA from the ML discrepancy function, as is done by some software, the values will be slightly different from those reported here, but one may still employ the check described in the preceding section.

REFERENCES

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

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6,* 97–103.

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

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Steiger, J. H. (1994). Structural equation modeling [Computer program]. In *Statistica/W, Version 4.5.* Tulsa, OK: Statsoft.

Steiger, J. H. (1995). Structural equation modeling [Computer program]. In *Statistica/W, Version 5.0.* Tulsa, OK: Statsoft.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual Spring Meeting of the Psychometric Society, Iowa City, IA.