Measures of Fit in Structural Equation Modeling: An Introduction

James H. Steiger

Besides the chi square value and its probability level, modern structural equation modeling programs print a number of indices of fit, which can be used to interpret how well a model fits the data. The indices discussed here are all single model indices, i.e., values which can be computed from a single model tested on one data set.

General Theoretical Orientation

When attempting to assess how well a model fits a particular data set, one must realize at the outset that the classic hypothesis-testing approach is inappropriate. Consider common factor analysis. When maximum likelihood estimation became a practical reality, the chi square "goodness-of-fit" statistic was originally employed in a sequential testing strategy. According to this strategy, one first picked a small number of factors, and tested the null hypothesis that this factor model fit the population $\mathbf{\Sigma}$ perfectly. If this hypothesis was rejected, the model was assumed to be too simple (i.e., to have too few common factors) to fit the data. The number of common factors was increased by one, and the preceding procedure repeated. The sequence continued until the hypothesis test failed to reject the hypothesis of perfect fit. Steiger and Lind (1980) pointed out that this logic was essentially flawed, because, for any population $\mathbf{\Sigma}$ (other than one constructed as a numerical example directly from the common factor model) the *a priori* probability is essentially 1 that the common factor model will not fit perfectly so long as degrees of freedom for the chi square statistic were positive.

In essence, then, population fit for a covariance structure model with positive degrees of freedom is almost never really perfect. Testing whether it is perfect makes little sense. It is what statisticians sometimes call an "accept-support" hypothesis test, because accepting the null hypothesis supports what is generally the experimenter's point of view, i.e., that the model does fit.

Accept-support hypothesis tests are subject to a host of problems. In particular, of course, the traditional priorities between Type I and Type II error are reversed. If the proponent of a model simply performs the chi square test with low enough power, the model can be supported. As a natural consequence of this, hypothesis testing approaches to the assessment of model fit *should* make some attempt at power evaluation. Steiger and Lind (1980) demonstrated that performance of statistical tests in common factor analysis could be predicted from a noncentral chi square approximation. A number of papers dealing with the theory and practice of power evaluation in covariance structure analysis have been published (Matsueda & Bielby, 1986; Satorra and Saris, 1985; Steiger, Shapiro, & Browne, 1985). Unfortunately, power estimation in the analysis of a multivariate model is a difficult, somewhat arbitrary procedure, and such power estimates have not, in general, been reported in published studies.

The main reason for evaluating power is to gain some understanding of precision of estimation in a particular situation, to guard against the possibility that a model is "accepted" simply because of insufficient power. An alternative (and actually more direct) approach to the evaluation of precision is to *construct a confidence interval on the population noncentrality parameter* (or some particularly useful function of it). This approach, first suggested in the context of covariance structure analysis by Steiger and Lind (1980) offers two worthwhile pieces of information at the same time. It allows one, for a particular model and data set to express (1) how bad fit is in the population, and (2) how precisely the *population* badness-of-fit has been determined from the *sample* data.

Noncentrality-Based Parameter Estimates and Confidence Intervals

Let $\mathbf{S}$ be the sample covariance matrix based on N observations, and for notational convenience, define $n = N - 1$. $\mathbf{M}(\boldsymbol{\theta})$ is the attempt to reproduce $\mathbf{S}$ with a particular model and a particular parameter vector $\boldsymbol{\theta}$. $\mathbf{M}(\boldsymbol{\theta}_{ML})$ is the corresponding matrix constructed from the vector of maximum likelihood estimates $\boldsymbol{\theta}_{ML}$ obtained by minimizing the maximum likelihood discrepancy function,

$$F_{ML}(\mathbf{S}, \mathbf{M}(\boldsymbol{\theta})) = \ln\left|\mathbf{M}(\boldsymbol{\theta})\right| - \ln\left|\mathbf{S}\right| + \mathrm{Tr}(\mathbf{M}(\boldsymbol{\theta})^{-1}) - p \qquad (1)$$

Suppose one has obtained maximum likelihood estimates. Then under conditions (i.e., the "population drift" conditions in Steiger, Shapiro, and Browne, 1985) designed to simulate the situation where the model fits well but not perfectly, nF_{ML} has an asymptotic noncentral chi square distribution with $p(p+1)/2-t$ degrees of freedom, where t is the number of free parameters in the model, and p is the order of $\mathbf{S}$. The noncentrality parameter of the noncentral chi square distribution is nF^*, where F^* is the value of the discrepancy function in Equation 1 obtained if $\mathbf{S}$ is replaced by the population covariance matrix $\mathbf{\Sigma}$, and maximum likelihood estimation is performed on $\mathbf{\Sigma}$ instead of $\mathbf{S}$. Hence, the noncentrality parameter is in effect the "population badness-of-fit statistic."

Interestingly, if one divides the noncentrality parameter by n, one obtains a measure of *population* badness-of-fit which depends only on the model, $\mathbf{\Sigma}$, and the method of estimation.

If one has a single observation from a noncentral chi square distribution, it is very easy to obtain an unbiased estimate of the noncentrality parameter of that distribution. By well known theory, if noncentral chi square variate X has noncentrality parameter λ and degrees of freedom ν, the expected value of X is given by

$$E(X) = \nu + \lambda \tag{2}$$

whence it immediately follows that an unbiased estimate of λ is simply $X-\nu$. Consequently a large sample "biased corrected" estimate of $F*$ is $(X-\nu)/\lambda$. Since $F*$ can never be negative, the simple unbiased estimator is generally modified in practice by converting negative values to zero. The estimate

$$F^+ = \max\{X-\nu)/n, 0\} \tag{3}$$

is the result.

It is also possible, by a variety of methods, to obtain a maximum likelihood estimate of the noncentrality parameter λ from a single observation from a non-central chi square distribution with ν degrees of freedom, and a confidence interval for λ as well. (See, e.g., Saxena and Alam, 1982; Spruill, 1986.)

Before continuing, recall some very basic statistical principles.

- Under very general conditions, if $\hat{\theta}$ is a maximum likelihood estimator for a parameter θ, then for any monotonic strictly increasing function $f(\ )$, $f(\hat{\theta})$ is a maximum likelihood estimator of $f(\theta)$.
- Moreover, if x_{low} and x_{high} are valid limits of a $100(1-\alpha)\%$ confidence interval for θ, $f(x_{low})$ and $f(x_{high})$ are valid limits of a $100(1-\alpha)\%$ confidence interval for $f(\theta)$.

These principles immediately imply that, since one can obtain a maximum likelihood estimate and confidence interval for nF^*, one can obtain a confidence interval and maximum likelihood estimate for F^* by dividing by n.

Structural equation modeling programs obtain a point estimate and confidence interval for nF^* by iterative methods. 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 of λ which place the observed value of the chi square statistic at the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentile points of a $\chi^2_{\nu,\lambda}$ distribution.

Steiger-Lind RMSEA Index

The Population Noncentrality Index F^* (PNI) offers some significant virtues as a measure of badness-of-fit (see, e.g., Steiger & Lind, 1980; McDonald, 1989). First, it is a weighted sum of discrepancies. Second, unlike the Akaike information criterion, for example, it is relatively unaffected by sample size.

However, there are two obvious problems with using the population noncentrality index as an index of population badness-of-fit.

The PNI is not in the metric of the original standardized parameters.

The PNI fails to compensate for model complexity. In general, for a given $\mathbf{\Sigma}$, the more complex the model the better it fits. A method for assessing population fit which fails to compensate for this will inevitably lead to choosing the most complex models, even when much simpler models fit the data nearly as well. The PNI fails to compensate for the size or complexity of a model. Hence it has limited utility as a device for comparing models.

The RMSEA index, first proposed by Steiger and Lind (1980), takes a relatively simplistic (but not altogether unreasonable) approach to solving these problems. Since model complexity is reflected directly in the number of free parameters, and inversely in the number of degrees of freedom, the PNI is divided by degrees of freedom, then the square root is taken to return the index to the same metric as the original standardized parameters.

Hence

$$R^* = \sqrt{\frac{F^*}{\nu}} \qquad 4$$

The RMSEA index R^* can be thought of roughly as a root mean square standardized residual. Values above .10 indicate an inadequate fit, values below .05 a very good fit. Point estimates below .01 indicate an outstanding fit, and are seldom obtained.

In practice, point and interval estimates of the population RMSEA index are calculated as follows. First, we obtain point and interval estimates of the PNI. (Negative point estimates are replaced by zero.) Since all these are non-negative, and R^* is a monotonic transform of the PNI, point estimates and a confidence interval for R^* are obtained by inserting the corresponding values for F^* in Equation 4. It may be shown easily that a bound on the point estimate of R^* implies a corresponding bound on the ratio of the chi square statistic to its degrees of freedom. Specifically, suppose, for example, you have decided that, for your purposes, the point estimate of the RMS index should be less than some value c. Manipulating the interval, we start with

$$R^* < c$$

Letting $\chi^2 = nF$, the expression becomes

$$\sqrt{\frac{\frac{\chi^2 - \nu}{n}}{\nu}} < c$$

This in turn implies that

$$\frac{\chi^2}{\nu} < 1 + nc^2 \qquad (5)$$

So, for example, the rule of thumb that, for "close fit," RMS should be less than .05 translates into a rule that

$$\frac{\chi^2}{\nu} < 1 + \frac{n}{400} \tag{6}$$

With this criterion, if $n = 400$, the ratio of the *chi square* to its degrees of freedom should be less than 2. Note that this rule implies a *less stringent* criterion for the ratio χ^2 / ν as sample size increases.

Rules of thumb that cite a single value for a critical ratio of χ^2 / ν ignore the point that the chi square statistic has an expected value that is a function of degrees of freedom, population badness of fit, *and N*. Hence, for a fixed level of population badness of fit, the expected value of the chi square statistic will increase as sample size increases. The rule of Equation 5 compensates for this, and hence it may be useful as a quick and easy criterion for assessing fit.

To avoid misinterpretation, I should emphasize at this point that my primary emphasis is on a confidence interval based approach, rather than one based on point estimates. The confidence interval approach incorporates information about precision of estimate into the assessment of population badness of fit. Simple rules of thumb (such as that of Equation 5) based on point estimates ignore these finer statistical considerations.

Population *Gamma* Index

Tanaka and Huba (1985, 1989) have provided a general framework for conceptualizing certain fit indices in covariance structure analysis. In their first paper, Tanaka and Huba (1985, their Equation 19) gave a general form for the *sample* fit index for covariance structure models under arbitrary generalized least squares estimation.

In the Tanaka-Huba treatment, it is assumed that a covariance structure model has been fit by minimizing an arbitrary generalized least squares (GLS) discrepancy function of the form

$$F(\mathbf{S}, \mathbf{M}(\boldsymbol{\theta}) \mid \mathbf{V}) = \tfrac{1}{2} \mathrm{Tr} \left\{ \left[\mathbf{S} - \mathbf{M}(\boldsymbol{\theta}) \right] \mathbf{V} \right\}^2 \tag{7}$$

or, equivalently (see Browne, 1974)

$$F(\mathbf{s}, \hat{\boldsymbol{\sigma}} \mid \mathbf{W}) = \left(\mathbf{s} - \hat{\boldsymbol{\sigma}}\right)' \mathbf{W} \left(\mathbf{s} - \hat{\boldsymbol{\sigma}}\right) \tag{8}$$

where $\mathbf{s} = \text{vecs}(\mathbf{S})$, and $\hat{\boldsymbol{\sigma}} = \text{vecs}(\mathbf{M}(\boldsymbol{\theta}))$. $\mathbf{V}$ in Equation 7 and $\mathbf{W}$ in Equation 8 are arbitrary matrices. Appropriate choice of $\mathbf{V}$ or $\mathbf{W}$ can yield GLS or IRGLS estimators. For example, minimization of Equation 7 with $\mathbf{V} = \mathbf{S}^{-1}$ if $\mathbf{S}$ has a Wishart distribution yields the well-known GLS estimators (Browne, 1974). Setting $\mathbf{V} = \left[\mathbf{M}(\boldsymbol{\theta})\right]^{-1}$ yields "Iteratively Reweighted Generalized Least Squares (IRGLS) estimators. Bentler (1989, page 216), citing Lee & Jennrich, 1979, states that (IRGLS) estimators are equivalent to ML estimators. Setting $\mathbf{V} = \left[\mathbf{M}(\hat{\boldsymbol{\theta}}_{ML})\right]^{-1}$ yields a discrepancy function which, according to Browne (1974), is usually minimized by the same which minimizes the maximum likelihood discrepancy function.

The Tanaka-Huba fit index can be written as

$$\gamma = 1 - \left[\mathbf{e}'\mathbf{W}\mathbf{e} \,/\, \mathbf{s}'\mathbf{W}\mathbf{s}\right] \tag{9}$$

where $\mathbf{e} = \mathbf{s} - \hat{\boldsymbol{\sigma}}$ is the vectorized model residuals. Tanaka and Huba (1989) demonstrate a deceptively simple, but important result which holds for models which are invariant under a constant scaling function (ICSF). A covariance structure model is ICSF if multiplication of any covariance matrix which fits the model by a positive scalar yields another covariance matrix which also satisfies the model exactly (though possibly with different free parameter values).

If a model which is ICSF has been estimated by minimizing a discrepancy function of the form given in Equations 7 and 8 then

$$\mathbf{e}'\mathbf{W}\hat{\boldsymbol{\sigma}} = \mathbf{0} \tag{10}$$

$\mathbf{e}$ and $\mathbf{s}$ are orthogonal "in the metric of $\mathbf{W}$," and, consequently,

$$\mathbf{s}'\mathbf{W}\mathbf{s} = \hat{\boldsymbol{\sigma}}'\mathbf{W}\hat{\boldsymbol{\sigma}} + \mathbf{e}'\mathbf{W}\mathbf{e} \tag{11}$$

If Equation 11 holds, then γ may be written

$$\gamma = \hat{\boldsymbol{\sigma}}'\mathbf{W}\hat{\boldsymbol{\sigma}} \,/\, \mathbf{s}'\mathbf{W}\mathbf{s} = 1 - \mathbf{e}'\mathbf{W}\mathbf{e} \,/\, \mathbf{s}'\mathbf{W}\mathbf{s} \tag{12}$$

In this form, γ defines a weighted coefficient of determination.

In what follows, we shall adopt the simplifying notational convention $\hat{\boldsymbol{\Sigma}}_{ML} = \mathbf{M}(\hat{\boldsymbol{\theta}}_{ML})$ Under the conditions of Equation 12, with maximum likelihood estimation, one immediately obtains

$$F\left(\mathbf{S}, \hat{\boldsymbol{\Sigma}}_{ML}\right) = \mathbf{e}'\mathbf{W}\mathbf{e} = \tfrac{1}{2}\mathrm{Tr}\left(\left[\mathbf{S} - \hat{\boldsymbol{\Sigma}}_{ML}\right]\hat{\boldsymbol{\Sigma}}_{ML}^{-1}\right)^2 \tag{13}$$

$$\mathbf{s}'\mathbf{W}\mathbf{s} = \tfrac{1}{2}\mathrm{Tr}\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}_{ML}^{-1}\right)^2 \tag{14}$$

whence

$$\gamma_{ML} = 1 - \frac{\mathrm{Tr}\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}_{ML}^{-1} - \mathbf{I}\right)^2}{\mathrm{Tr}\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}_{ML}^{-1}\right)^2} \tag{15}$$

which is equivalent to the Jöreskog and Sörbom (1984) GFI index included in the output of many structural equation modeling programs. Moreover, if the model is ICSF, then, under maximum Wishart likelihood estimation, there is the simplifying result (Browne, 1974, Proposition 8)

$$\mathrm{Tr}\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}_{ML}^{-1}\right) = p \tag{16}$$

Substituting in Equation 13, one finds

$$F\left(\mathbf{S}, \hat{\boldsymbol{\Sigma}}_{ML}\right) = \tfrac{1}{2}\left(\mathrm{Tr}\left[\mathbf{S}\hat{\boldsymbol{\Sigma}}_{ML}^{-1}\right]^2 - p\right) \tag{17}$$

and so

$$\gamma_{ML} = \frac{p}{2\mathbf{s}'\mathbf{W}\mathbf{s}} = \frac{p}{\mathrm{Tr}\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}_{ML}^{-1}\right)^2} \tag{18}$$

Tanaka and Huba (1985, 1989) based their derivation of γ on sample quantities. However, in principle one is interested in a sample index primarily as a vehicle for estimating the corresponding *population* index. Define $\boldsymbol{\Sigma}_{ML}$ as the reproduced covariance matrix resulting from fitting the model to the population covariance matrix $\boldsymbol{\Sigma}$ with maximum likelihood estimation. The corresponding population quantities are obtained by substituting $\boldsymbol{\Sigma}$ for $\mathbf{S}$, and $\boldsymbol{\Sigma}_{ML}$ for $\hat{\boldsymbol{\Sigma}}_{ML}$ in Equations 17 and 18. One obtains

$$\Gamma_1 = \frac{p}{\mathrm{Tr}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{ML}^{-1}\right)^2} \tag{19}$$

Γ_1 can be thought of as a *weighted population coefficient of determination for the multivariate (ICSF) model*. (It may also be thought of as the population equivalent of the Jöreskog-Sörbom GFI index.)

An accurate point estimate for Γ_1 will provide useful information about the extent to which a model reproduces the information in $\boldsymbol{\Sigma}$. A confidence interval, however, provides even more useful information, because it conveys not only the size of Γ_1, but also the precision of our estimate.

Let F^* be the Population Noncentrality Index $F(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{ML}^{-1})$. From Equations 17 and 18, it is easy to see that

$$\Gamma_1 = \frac{p}{2F^* + p} \tag{20}$$

Equation 20 demonstrates that, under maximum likelihood estimation with ICSF models, Γ_1 can be expressed solely as a function of the Population Noncentrality Index and p, the number of manifest variables. Any consistent estimate of F^* will yield a consistent estimate of Γ_1 when substituted in Equation 20. Equation Similarly, substitution of the endpoints of the confidence interval for F^* in Equation 20 will generate a confidence interval for Γ_1.

Equation 20 and the accompanying derivation were first presented in Steiger (1989). Maiti and Mukherjee (1990), working completely independently of Steiger (1989), produced the identical result (their Equation 17). Steiger (1989) had suggested that the sample GFI was a biased estimator of the population value. Maiti and Mukerjee quantified the bias with the following (their Equation 16) approximate expression (for a chi square statistic with ν degrees of freedom).

$$E\left(\gamma_{ML}\right) \approx \frac{p}{p + 2F^* + \frac{2}{N}\nu} \tag{21}$$

This can be rewritten in perhaps a more revealing form as

$$E\left(\gamma_{ML}\right) = \Gamma_1\left(\frac{p}{p + (2\nu / N)\Gamma_1}\right) \qquad 22$$

Adjusted Population Gamma Index

Γ_1, like F^*, fails to compensate for the effect of model complexity. Consider a sequence of *nested* models, where the models with more degrees of freedom are special cases of those with fewer degrees of freedom. (See Steiger, Shapiro, and Browne, 1985, for a discussion of the statistical properties of chi square tests with nested models.) For a nested sequence of models, the more complex models (i.e., those with more free parameters and fewer degrees of freedom) will always have Γ_1 coefficients as low or lower than those which are less complex.

Goodness of fit, as measured by Γ_1, improves more or less inevitably as more parameters are added. The adjusted population *gamma* index Γ_2 attempts to compensate for this tendency.

Just as Γ_1 is computed by subtracting a ratio of sums of squares from 1, Γ_2 is obtained by subtracting a corresponding ratio of mean squares from 1. Let $p^* = p(p+1)/2$. Let $\hat{\boldsymbol{\sigma}}$ be a $p^* \times 1$ vector of non-duplicated elements of the population reproduced covariance matrix $\mathbf{M}(\boldsymbol{\theta})$, as in Equation 11 for a model with ν degrees of freedom, and $\mathbf{e}$ a corresponding vector of residuals. Then

$$\Gamma_2 = 1 - \frac{\mathbf{e}'\mathbf{W}\mathbf{e} / \nu}{\hat{\boldsymbol{\sigma}}'\mathbf{W}\hat{\boldsymbol{\sigma}} / p^*} = 1 - \left(p^* / \nu\right)(1 - \Gamma_1) \qquad (23)$$

Consistent estimates and confidence intervals for Γ_1 may thus be converted into corresponding quantities for Γ_2 by applying Equation 23.

McDonald's Index of Noncentrality

McDonald proposed this index of noncentrality in a 1989 article in the *Journal of Classification*. The index represents one approach to transforming the population noncentrality index F^* into the range from 0 to 1. The index does not compensate for model parsimony, and the rationale for the exponential transformation it uses is primarily pragmatic.

The index may be expressed as

$$\exp(-F^* / 2) \tag{24}$$

Good fit is indicated by values above .95.

Extensions to Multiple Group Analysis

When more than one group is analyzed, the chi square statistic is a weighted sum of the discrepancy functions obtained from the individual groups . If the sample sizes are equal, the noncentrality-based indices discussed above generalize in a way that is completely straightforward. When sample sizes are unequal, this is not so, although *SEPATH* will still compute modified versions of the indices as described below, and these will still be of considerable value in assessing model fit.

With K independent samples, the overall chi square statistic is of the form

$$F = \sum_{k=1}^{K} c_k F_k \tag{25}$$

where

$$c_k = \frac{N_k - 1}{N_{total} - K} \tag{26}$$

and

$$N_{total} = \sum_{k=1}^{K} N_k \tag{27}$$

The chi square statistic is then computed as

$$\chi^2 = (N_{total} - K)F \tag{28}$$

This statistic has, under the assumptions of Steiger, Shapiro, and Browne (1985) a large sample distribution that is approximated by a noncentral chi square distribution, with ν degrees of freedom, and a noncentrality parameter equal to

$$\lambda = \sum_{k=1}^{K} (N_k - 1)F_k^* \tag{29}$$

where F_k^* is the population discrepancy function for the kth group.

One can estimate this noncentrality parameter and set confidence intervals on it. However, inference to relevant population quantitities is less straightforward. Consider, for example, the point estimate analogous to the single sample case. The statistic

$$\frac{\chi^2 - \nu}{N_{total} - K} \tag{30}$$

has an expected value of approximately

$$\frac{\lambda}{N_{total} - K} = \sum_{k=1}^{K} \frac{N_k - 1}{N_{total} - K} F_k^* = \sum_{k=1}^{K} c_k F_k^* \tag{31}$$

where c_k is as defined above. This demonstrates that we can estimate a *weighted average* of the discrepancies for each sample, where the weights sum to 1, and are a function of sample size. If the sample sizes are equal, the weighted average becomes the simple arithmetic average, or mean, and so we can also estimate the unweighted sum of discrepancies.

How one should this information to produce multiple group versions of the RMSEA, and population gamma indices is open to some question when sample sizes are not equal. Perhaps the most natural candidates for the population RMSEA would be an "unweighted" index,

$$RMSEA_{unweighted} = \sqrt{\sum_{k=1}^{K} \frac{F_k^*}{\nu}} \tag{32}$$

and a "weighted" index

$$RMSEA_{weightd} = \sqrt{\frac{\sum_{k=1}^{K} c_k F_k^*}{\nu / k}} \tag{33}$$

When sample sizes are equal, both are the same.

Unfortunately, since we can only estimate the *weighted average* of population discrepancies, we must choose the second option when sample sizes are unequal. The *SEPATH* module in *Statistica* currently reports point and interval estimates for the *weighted* coefficient, which represents the square root of the ratio of a weighted average of discrepancies to an average number of degrees of freedom.

In calculating analogs of the population gamma indices, *SEPATH* substitutes K times the estimate of the *weighted average* of discrepancies in place of F^* in Equations 20 and 23.

Other Indices of Fit

Jöreskog-Sörbom GFI

This sample based index of fit is computed as γ_{ML} in Equation 18.

Jöreskog-Sörbom Adjusted GFI

This sample based index of fit is computed as

$$1-(p^* / \nu)(1-\gamma_{ML}) \tag{34}$$

where γ_{ML} is the GFI index.

Rescaled Akaike Information Criterion

In a number of situations the user must decide among a number of competing *nested* models of differing dimensionality. (The most typical example is the choice of the number of factors in common factor analysis.) Akaike (1973, 1983) proposed a criterion for selecting the dimension of a model. Steiger and Lind (1980) presented an extensive Monte Carlo study of the performance of the Akaike criterion. Here the criterion is rescaled (without affecting the decisions it indicates) so that it remained more stable across differing sample sizes. The rescaled Akaike criterion is as follows.

Consider a sequence of models M_k fit to a data set. Let $F_{ML,k}$ be the maximum likelihood discrepancy function and f_k be the number of free parameters for the model M_k. Let N be the sample size. Then select the model M_k for which

$$A_k = F_{ML,k} + \frac{2f_k}{N-1} \tag{35}$$

Schwarz's Bayesian Criterion

This criterion (Schwarz, 1978) is similar in use to Akaike's index, selecting, in a sequence of nested models, the model for which

$$S_k = F_{ML,k} + \frac{f_k \ln(N)}{N-1} \tag{36}$$

Browne-Cudeck Single Sample Cross-Validation Index

Browne and Cudeck (1989) proposed a single sample cross-validation index as a follow-up to their earlier (Cudeck & Browne, 1983) paper on cross-validation. Cudeck and Browne had proposed a cross-validation index which, for model M_k in a set of competing models is of the form $F_{ML}\left(\mathbf{S}_\nu, \mathbf{M}_k(\boldsymbol{\theta})\right)$. In this case, F is the maximum likelihood discrepancy function, $\mathbf{S}_\nu$ is the covariance matrix calculated on a cross-validation sample, and $\mathbf{M}_k(\boldsymbol{\theta})$ the reproduced covariance matrix obtained by fitting model M_k to the original calibration sample. In general, better models will have smaller cross-validation indices.

The drawback of the original procedure is that it requires two samples, i.e., the calibration sample for fitting the models, and the cross-validation sample. The new measure estimates the original cross-validation index from a single sample.

The measure is

$$C_k = F_{ML}(\mathbf{S}, \mathbf{M}_k(\boldsymbol{\theta})) + \frac{2f_k}{N-p-2} \tag{37}$$

where f_k is the number of free parameters in model k, p is the number of manifest variables, and N is the sample size.

Independence Model chi square and df

These are the chi square goodness-of-fit statistic, and associated degrees of freedom, for the hypothesis that the population covariances are all zero. Under the assumption of multivariate normality, this hypothesis can only be true if the variables are all independent. The "Independence Model" is used as the "Null Model" in several comparative fit indices.

Bentler-Bonett Normed Fit Index

One of the most historically important and original fit indices, the Bentler-Bonett index measures the *relative decrease* in the discrepancy function caused by switching from a "Null Model" or baseline model, to a more complex model. It is defined as: $B_k = \frac{F_0 - F_k}{F_0}$

F_0 = the discrepancy function for the "Null Model"

F_k = the discrepancy function for the k'th model

This index approaches 1 in value as fit becomes perfect. However, it does not compensate for model parsimony.

Bentler-Bonett Non-Normed Fit Index

This comparative index takes into account model parsimony. It may be written as

$$BNN_k = \frac{\frac{\chi_0^2}{\nu_0} - \frac{\chi_k^2}{\nu_k}}{\frac{\chi_0^2}{\nu_0} - 1} = \frac{\frac{F_0}{\nu_0} - \frac{F_k}{\nu_k}}{\frac{F_0}{\nu_0} - \frac{1}{N-1}} \tag{38}$$

χ_0^2 = chi square for the "Null Model"

χ_k^2 = chi square for the k'th model

ν_0 = degrees of freedom for the "Null Model"

ν_k = degrees of freedom for the k'th model

Bentler Comparative Fit Index

This comparative index estimates the relative decrease in population noncentrality obtained by changing from the "Null Model" to the k'th model. The index may be computed as:

$$1 - \frac{\hat{\tau}_k}{\hat{\tau}_0} \tag{39}$$

$\hat{\tau}_k$ = estimated non-centrality parameter for the k'th model

$\hat{\tau}_0$ = estimated non-centrality parameter for the "Null Model"

James-Mulaik-Brett Parsimonious Fit Index

This index was one of the earliest (along with the Steiger-Lind index) to compensate for model parsimony. Basically, it operates by rescaling the Bentler-Bonnet Normed fit index to compensate for model parsimony. The formula for the index is:

$$\pi_k = \frac{\nu_k}{\nu_0} B_k \tag{40}$$

ν_0 = degrees of freedom for the "Null Model"

ν_k = degrees of freedom for the k'th model

B_k = Bentler-Bonnet normed fit index

Bollen's *Rho*

This comparative fit index computes the relative reduction in the discrepancy function per degree of freedom when moving from the "Null Model" to the k'th model. It is computed as:

$$\rho_k = \frac{\frac{F_0}{\nu_0} - \frac{F_k}{\nu_k}}{\frac{F_0}{\nu_0}} \tag{41}$$

ν_0 = degrees of freedom for the "Null Model"

ν_k = degrees of freedom for the *k*th model

F_0 = the discrepancy function for the "Null Model"

F_k = the discrepancy function for the *k*th model

Comparing the two indices, we see that, for even moderate *N,* there is bound to be virtually no difference between Bollen's *Rho* and the Bentler-Bonnet Non-normed fit index.

Bollen's *Delta*

This index is also similar in form to the Bentler-Bonnet index, but rewards simpler models (those with higher degrees of freedom). It is computed as:

$$\Delta_k = \frac{F_0 - F_k}{F_0 - \frac{\nu_k}{N}} \tag{42}$$

ν_k = degrees of freedom for the kth model

F_0 = the discrepancy function for the "Null Model"

F_k = the discrepancy function for the kth model

N = sample size

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute: Proceedings of the 44th Session, Volume 1*. Pages 277–290.

Bentler, P. M, & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles, CA: BMDP Statistical Software.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Browne, M. W., & Cudeck, R. (1990). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*, 445–455.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models*. Beverly Hills, CA: Sage.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317–327.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, *18*, 147–167.

Huba, G. J., & Harlow, L. L. (1987). Robust structural equation models: implications for developmental psychology. *Child Development*, *58*, 147–166.

Jaccard, J., Weber, J., & Lundmark, J. (1975). A multitrait-multimethod factor analysis of four attitude assessment procedures. *Journal of Experimental Social Psychology*, *11*, 149–154.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis. Assumptions, models, and data*. Beverly Hills, CA: Sage Publications.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*, 443–477.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, *19*, 404–416.

Jöreskog, K. G., & Sörbom, D. (1984). *Lisrel VI. Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Mooresville, Indiana: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). *Lisrel 7. A guide to the program and applications*. (2nd Ed.) Chicago, Illinois: SPSS Inc.

Maiti, S. S., & Mukherjee, B. N. (1990). A note on the distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika*, *55*, 721–726.

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

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.

Saxena, K. M. L., & Alam, K. (1982). Estimation of the noncentrality parameter of a chi squared distribution. *Annals of Statistics*, *10*, 1012–1016.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Spruill, M. C. (1986). Computation of the maximum likelihood estimate of a noncentrality parameter. *Journal of Multivariate Analysis*, *18*, 216–224.

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

Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City. May 30, 1980.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, *50*, 253–264.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *38*, 197–201.

Tanaka, J. S., & Huba, G. J. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *42*, 233–239.