STRUCTURAL EQUATION MODELING, *11*(4), 621–637


TEACHER'S CORNER

# Using the Delta Method for Approximate Interval Estimation of Parameter Functions in SEM

Tenko Raykov
*Fordham University*

George A. Marcoulides
*California State University, Fullerton*

In applications of structural equation modeling, it is often desirable to obtain measures of uncertainty for special functions of model parameters. This article provides a didactic discussion of how a method widely used in applied statistics can be employed for approximate standard error and confidence interval evaluation of such functions. The described approach is illustrated with data from a cognitive intervention study, in which it is used to estimate time-invariant reliability in multiwave, multiple indicator models.

Structural equation modeling (SEM) enjoys widespread popularity in the behavioral, social, and educational sciences. A major reason for its frequent use is the widely appreciated fact that SEM allows one to readily account for measurement error and to model complex multivariable relations among observed and latent variables whereby direct and indirect (mediated) effects are straightforwardly evaluated along with indexes of their estimation precision. To use the SEM methodology to its fullest potential, it is essential that a proposed model be developed in a manner consistent with all available theoretical and research-accumulated knowledge in a given substantive domain. When conceptualizing the model, a typical requirement is that it

Requests for reprints should be sent to Tenko Raykov, Department of Psychology, Fordham University, Bronx, NY 10458. E-mail: raykov@fordham.edu

includes parameters pertaining to (a) studied research question(s). Frequently aspects of such questions are reflected directly in one or more model parameters, such as structural regression coefficients, latent variances and covariances, or factor loadings. Use of the model then provides the researcher with estimates of these parameters that possess certain optimality properties (e.g., Bollen, 1989), along with indexes of their stability across repeated sampling from the studied population. These indexes—the parameter standard errors—play, as is well known, an instrumental role in constructing confidence intervals for particular population parameters of interest (e.g., Hays, 1994; Raykov & Marcoulides, 2000).

Oftentimes, however, a research query may be concerned not with a single or a group of separately considered parameters in a model but with a function of several of them. Simple examples of such instances arise when a researcher is interested in (a) a ratio of two parameters (e.g., factor loading to error variance or squared factor loading to error variance—e.g., in scale development contexts); (b) a sum or difference between parameters (e.g., in multigroup models); (c) a correlation coefficient between two latent variables (e.g., in models with special constraints on pertinent factor loadings); or (d) more involved functions of model parameters, such as sum score reliability or maximal reliability coefficients (e.g., McDonald, 1999; Raykov, 2001; Raykov & Penev, in press). Point estimation of such parametric functions usually proceeds without difficulty, by simply applying the function on the parameter estimates obtained once the model is fitted to data. For example, when for this purpose the popular maximum likelihood (ML) method is used, the resulting parameter function estimates will themselves be ML estimates, due to the invariance property of ML estimators (e.g., Pawitan, 2001). However, an informed substantive analysis requires also measures of stability of these estimates (e.g., Hays, 1994). In some cases, a reparameterization of the model that maps some of these parametric functions into single new parameters (e.g., a latent correlation when the scales of involved factors can be set by fixing their variances to unity for identifiability purposes or through judicious restriction of pertinent factor loadings) may be possible. These new parameters are readily estimated using conventional SEM, which also renders standard errors for them. In other cases, however, such a parameterization may be difficult to find or potentially nonexistent or may lead to the loss of substantive interpretability of other relevant aspects of a given model (e.g., Raykov & Penev, 2002). Although computer-intensive resampling methods can be utilized in such settings to obtain an approximate idea of sampling distributions of parametric functions of concern and in particular to furnish confidence intervals for them, (a) they are not readily and widely applicable with many currently available software, (b) they can be relatively expensive in terms of time and related activities on the part of the researcher, and (c) there are no generally applicable results as yet of how good at a given finite sample size in a practical application the underlying approximation of sampling by pertinent resampling distributions is within the framework of SEM (e.g., Bollen & Stine, 1993).

This article discusses an analytic alternative that is readily and widely employable in SEM applications. The alternative procedure is represented by the so-called delta method that is frequently utilized in applied statistics for purposes of interval estimation and allows one to obtain an approximate standard error of any smooth function (see next section) of parameters that are estimated in a given modeling session. A prerequisite for using the method is that parameter estimates and their variances and covariances are available. The latter requirement can be routinely fulfilled in SEM applications, which renders this interval estimation approach readily applicable in empirical research. The underlying approximation procedure is discussed next and illustrated in a subsequent section on data from a cognitive intervention study in which it is used to obtain an approximate standard error and confidence interval for time-invariant reliability in multiwave, multiple indicator models while accounting for observed variable specificity.

## AN APPLICATION-ORIENTED DESCRIPTION OF THE DELTA METHOD

### Smooth Parametric Functions

The practice of SEM in applied research is often concerned with special types of functions of parameters in a considered model that are referred to later as "smooth functions." In simple terms, a graph of such a function with regard to any parameter has the properties of (a) being continuous (i.e., not having breaks) and (b) not abruptly changing direction when moving from smaller to larger values on that parameter (while fixing the other parameters). For example, sums, differences, products, and ratios of model parameters in SEM applications are all functions that fulfill this requirement. However, not only simple functions such as these are smooth, in fact many functions that are more involved are also smooth. For example, consider the function of the scale reliability coefficient $\rho_X$ for a set $X_1, X_2, \ldots, X_k$ of congeneric measures (Jöreskog, 1971, $k > 2$), that is, measures assessing a common underlying dimension (denoted $\eta$) with possibly different units of measurement and error variances:

$$\rho_X = \rho_X(b_1, b_2, \ldots, b_k, \theta_1, \theta_2, \ldots, \theta_k) = \frac{(b_1 + b_2 + \ldots + b_k)^2}{(b_1 + b_2 + \ldots + b_k)^2 + \theta_1 + \theta_2 + \ldots + \theta_k}, \tag{1}$$

where $b_1, b_2, \ldots, b_k$ are the loadings of the measures on the common latent dimension $\eta$ (with variance assumed 1, for identifiability reasons), $\theta_1, \theta_2, \ldots, \theta_k$ are the variances of the associated error terms, and $X = X_1 + X_2 + \ldots + X_k$ is the scale score (unit-weighted composite; e.g., Bollen, 1989; we use in Equation 1 the notation $\rho_X(b_1, b_2, \ldots, b_k, \theta_1, \theta_2, \ldots, \theta_k)$ to emphasize that the reliability coefficient $\rho_x$ is a function of the altogether $2k$ loading and error variance parameters applied in its arguments). Under the assumption of nonvanishing loadings and error variances, the function $\rho_X(b_1, b_2, \ldots, b_k, \theta_1, \theta_2, \ldots, \theta_k)$, with respect to any of its $2k$ parameters, is smooth across its domain. Similarly, the maximal reliability coefficient for a given set of congeneric measures is then a smooth function of factor loadings and error variances (e.g., Raykov & Penev, 2004). In the rest of this article, we are only concerned with smooth functions (see the next section in which this notion is made more precise).

## Linear Approximation: Basis

A smooth function can be shown to be locally linear (i.e., it can be approximated by a linear function with regard to its arguments), as can be readily seen using its so-called Taylor expansion (Stewart, 1991). Specifically, if a given function of say a single variable or argument, $f(x)$, is $n + 1$ times differentiable in an interval containing a point of interest $x_0$ from its domain (in which sense we refer to it as smooth), then

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \ldots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0) + R_n, \tag{2}$$

where $f^{(k)}(x_0)$ denotes the $k$th derivative of $f$, evaluated at $x_0$, and $R_n$ is a remainder term ($0 < k < n + 1, n > 0$). Equation 2 is frequently referred to as the $n$th-order Taylor series expansion and is the basis of the following approximation of $f(x)$, resulting from dropping the remainder $R_n$ in Equation 2:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \ldots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0), \tag{3}$$

in terms of the $n$th degree polynomial in its right-hand side (with '$\approx$' denoting approximately equal). The quality of this approximation improves as $x$ gets closer to the point $x_0$.

Typically, of particular interest in empirical applications is the first-order approximation:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0). \tag{4}$$

It can also be seen as following from geometry-related considerations when approximating the functional value at $x$ by a pertinent linear increase along the derivative $f'(x_0)$ when moving from $x_0$ to $x$ (e.g., Tan, 1990).

In the context of SEM applications, it is common to consider as a variable (argument) a parameter estimator, denoted $\hat{\gamma}$, of a parameter $\gamma$ under consideration. In that case, Equation 4 takes the form

$$f(\hat{\gamma}) \approx f(\gamma_0) + (\hat{\gamma} - \gamma_0) f'(\gamma_0), \tag{5}$$

where $f(\gamma)$ denotes a parameter function of interest and $\gamma_0$ can be taken to be the population value of $\gamma$. With parameter functions of more than a single argument (parameter), say $\gamma_1, \gamma_2, \ldots, \gamma_p$ ($p > 1$), the corresponding first-order approximation is (e.g., Tan, 1990)

$$\begin{aligned} f(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p) \approx\ & f(\gamma_{10}, \gamma_{20}, \ldots, \gamma_{p0}) + D_1(\hat{\gamma}_1 - \gamma_{10}) + \\ & D_2(\hat{\gamma}_2 - \gamma_{20}) + \ldots + D_p(\hat{\gamma}_p - \gamma_{p0}), \end{aligned} \tag{6}$$

where for simplicity of notation and ease of following developments $D_j = \dfrac{\partial f(\gamma_{10}, \gamma_{20}, \ldots, \gamma_{p0})}{\partial \gamma_j}$ symbolizes the partial derivative of $f$ with regard to $\gamma_j$, when evaluated at the population parameter ($j = 1, \ldots, p$). We note in passing that Equation 5 is a special case of Equation 6, which results when $p = 1$ (i.e., $f$ is a function of a single argument or parameter). Higher order approximations (expansions) are also readily obtained in the case of functions of multiple arguments (e.g., Stewart, 1991) but will not be of interest to us because they often provide limited improvement, if any, of considerable practical value, over first-order approximations.

As in the single variable case, (a) the quality of the approximation in Equation 6 depends on how close to linear the function $f$ is at the point $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$, and (b) the underlying assumption is that $f$ is $n + 1$ times continuously partially differentiable in an open interval in its domain, which contains $\gamma$ and the true parameter vector $\gamma_0 = (\gamma_{10}, \gamma_{20}, \ldots, \gamma_{p0})$ and all points between them (Stewart, 1991). Due to the smoothness of $f$, as this point $\gamma$ moves closer to $\gamma_0$, the linear approximation improves (e.g., Tan, 1990). This approximation is the essence of what can be considered the first step of an application of the delta method.

## Linear Approximation: Example

According to the preceding subsection, the value of a smooth parameter function at a point close to the population parameter equals approximately the sum of (a) the value of the function at the latter and (b) a linear combination of the differences for each of its arguments from their own population value weighted by the derivatives of the function with respect to these arguments evaluated at the true parameter point. In particular, if in the previous scale reliability example $(b_{10}, b_{20}, \ldots, b_{k0}, \theta_{10}, \theta_{20}, \ldots, \theta_{k0})$ denotes the unknown population parameter with coordinates the population loadings and error variances, then the value of the scale reliability coefficient at a point close to that true parameter can be approximated as

$$\begin{aligned}
&\rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0}) + \\
&+ \frac{\partial \rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0})}{\partial b_1}(b_1 - b_{10}) + \\
&+ \frac{\partial \rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0})}{\partial b_2}(b_2 - b_{20}) + ... \\
&+ \frac{\partial \rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0})}{\partial b_k}(b_k - b_{k0}) + \\
&+ \frac{\partial \rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0})}{\partial \theta_1}(\theta_1 - \theta_{10}) + \\
&+ \frac{\partial \rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0})}{\partial \theta_2}(\theta_2 - \theta_{20}) + ... \\
&+ \frac{\partial \rho_X(b_{10}, b_{20}, ..., b_{k0}, \theta_{10}, \theta_{20}, ..., \theta_{k0})}{\partial \theta_k}(\theta_k - \theta_{k0})
\end{aligned} \tag{7}$$

where

$\frac{\partial \rho_X(b_{10}, b_{20}, \ldots, b_{k0}, \theta_{10}, \theta_{20}, \ldots, \theta_{k0})}{\partial b_i}$ and $\frac{\partial \rho_X(b_{10}, b_{20}, \ldots, b_{k0}, \theta_{10}, \theta_{20}, \ldots, \theta_{k0})}{\partial \theta_i}$ $(i = 1, 2, \ldots, k)$ denote the derivative of the reliability function in Equation 1 with respect to each loading and error variance, respectively, evaluated at the population parameter (e.g., Raykov, 2002). We note that Equation 7 is not identical to $\rho_X(b_1, b_2, \ldots, b_k, \theta_1, \theta_2, \ldots, \theta_k)$, the value of the function of interest at a point close to the population parameter, but only approximates that value. If we now denote for notational simplicity $\gamma_1 = b_1, \gamma_2, = b_2, \ldots, \gamma_k = b_k$, $\gamma_{k+1} = \theta_1$, $\gamma_{k+2} = \theta_2, \ldots$, $\gamma_p = \theta_k$, and by $D_j = \frac{\partial \rho_X(b_{10}, b_{20}, \ldots, b_{k0}, \theta_{10}, \theta_{20}, \ldots, \theta_{k0})}{\partial \gamma_j}$ the derivative of the reliability coefficient with respect to the $j$th parameter (i.e., considering all other parameters as constants), evaluated at the true parameter point $(j = 1, 2, \ldots, p)$, then approximation Equation 7 of the parameter function (1) can be written as

$$\rho_X(\gamma_1, \gamma_2, ..., \gamma_p) \approx \rho_X(\gamma_{10}, \gamma_{20}, ..., \gamma_{p0}) + D_1(\gamma_1 - \gamma_{10}) + D_2(\gamma_2 - \gamma_{20}) + ... + D_p(\gamma_p - \gamma_{p0}), \tag{8}$$

that up to notation of the function under consideration and its arguments is identical to Equation 6 (cf. Raykov, 2002).

Hence, approximation Equation 6 gives a rather general tool for representing in a linear form a smooth function of any number of variables. In particular, for a given structural equation model this first step of a utilization of the delta method, embodied in Equation 6, provides a fairly widely applicable means of approximating smooth functions of any number of its parameters. To enhance the practical relevance of this discussion, and specifically of Equation 6 underlying the rest of the article, a list of rules for frequently used derivatives is provided in the Appendix. These rules can be employed essentially automatically to initiate an application of the delta method in empirical research.

## Approximate Variance

Once the approximation of a smooth parametric function $f$ under consideration is obtained, approximate variance evaluation of $f$ is facilitated. To this end, let us return to the previous approximate Equation 6, which for convenience in the following developments is restated next:

$$f(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_p) \approx f(\gamma_{10}, \gamma_{20}, ..., \gamma_{p0}) + D_1(\hat{\gamma}_1 - \gamma_{10}) + D_2(\hat{\gamma}_2 - \gamma_{20}) + ... + D_p(\hat{\gamma}_p - \gamma_{p_0}) \tag{9}$$

Because the parameter estimates appearing in the right-hand side of Equation 9 can be considered realizations of random variables, sampling variability is inherent in its left-hand side as well. In fact, just like these parameter estimates, $f(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p)$ can also be considered a random variable itself. To obtain its approximate variance, we can apply the well-known formula for variance of a sum of random variables (e.g., Hays, 1994) and obtain the following:

$$\hat{\sigma}^2(f) \approx D_1{}^2\sigma^2(\hat{\gamma}_1) + D_2{}^2\sigma^2(\hat{\gamma}_2) + ... + D_p{}^2\sigma^2(\hat{\gamma}_p) + 2D_1D_2\sigma(\hat{\gamma}_1, \hat{\gamma}_2) + 2D_2D_3\sigma(\hat{\gamma}_2, \hat{\gamma}_3) + ... + 2D_{p-1}D_p\sigma(\hat{\gamma}_{p-1}, \hat{\gamma}_p) \tag{10}$$

In the right-hand side of Equation 10, the variances of the parameter estimates (i.e., $\sigma^2(\hat{\gamma}_1)$, $\sigma^2(\hat{\gamma}_2), \ldots, \sigma^2(\hat{\gamma}_p)$) are the squared parameter standard errors provided by an SEM application. These variances, as is well-known, also make up the main diagonal elements of the inverse of the so-called observed information matrix associated with the fitted model, which is routinely made available by popular SEM software such as LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 2004), and M*plus* (Muthén & Muthén, 2004). The covariances appearing in the right-hand side of (10) (i.e., $\sigma_2(\hat{\gamma}_1, \hat{\gamma}_2)$, $\sigma(\hat{\gamma}_2, \hat{\gamma}_3), \ldots, \sigma(\hat{\gamma}_{p-1}, \hat{\gamma}_p)$) are indexes of interrelation between these parameter estimates and make up the off-diagonal elements of that matrix. Finally, the partial derivatives in the equation are specific for each function *f* under consideration, and for the purpose of approximate variance estimation can be evaluated at the parameter estimates (which represent the best numerical guesses of the unknown parameters involved in *f*). Thus, an SEM application effectively furnishes all quantities appearing in the right-hand side of Equation 10 and hence permits approximate variance evaluation of the parametric function *f* of interest, which can be considered the second step of an application of the delta method. With a large number of parameters appearing in *f,* the following compact matrix formula proves handy in empirical research (e.g., Rao, 1973, chap. 6):

$$\hat{\sigma}^2(f) \approx \frac{\partial f}{\partial \hat{\gamma}'} Cov(\hat{\gamma}) \frac{\partial f}{\partial \hat{\gamma}}$$

where $\frac{\partial f}{\partial \hat{\gamma}'}$ symbolizes the $p \times 1$ row-vector of partial derivatives of *f* with regard to each consecutive parameter (element of the vector $\gamma$ containing all parameters involved in *f*) that is evaluated at the model solution, and $Cov(\hat{\gamma})$ is the part of the inverse observed information function associated with the fitted model, which pertains to these parameters (in the order that they appear in $\gamma$). This matrix formula is readily used with widely available software capable of matrix computation (e.g., SPSS, SAS, and MATLAB) and can save tedious calculation work and corresponding programming for a potentially long right-hand side of Equation 10. A special case of this formula results when $\gamma$ contains only one parameter, in which case the approximate variance of *f* obtains the following convenient form: $\sigma^2(f) \approx [f'(\hat{\gamma})]^2 \sigma^2(\hat{\gamma})$, where $f'(\hat{\gamma})$ denotes the derivative of *f* with regard to this parameter, evaluated at its estimate, and $\sigma^2(\hat{\gamma})$ is the squared standard error associated with this estimate.

## Approximate Standard Error and Confidence Interval

Once the variance approximation is obtained, an approximate standard error of the function *f* under consideration obviously results from Equation 10 as (e.g., Hays, 1994)

$$\hat{\sigma}(f) = [\hat{D}_1{}^2\sigma^2(\hat{\gamma}_1) + \hat{D}_2{}^2\sigma^2(\hat{\gamma}_2) + \ldots + \hat{D}_p{}^2\sigma^2(\hat{\gamma}_p) \\ + 2\hat{D}_1\hat{D}_2\sigma(\hat{\gamma}_1, \hat{\gamma}_2) + 2\hat{D}_2\hat{D}_3\sigma(\hat{\gamma}_2, \hat{\gamma}_3) + \ldots + 2\hat{D}_{p-1}\hat{D}_p\sigma(\hat{\gamma}_{p-1}, \hat{\gamma}_p)]^{1/2} \quad (11)$$

where $\hat{D}_j$ stands for derivative of $f$ with respect to its $j$th argument, $\gamma_j$, evaluated at the model solution (i.e., when estimates of the model parameters participating in this derivative are substituted into its expression; $j = 1, \ldots, p$).

Assuming availability of a large sample, the SEM parameter estimator follows approximately a normal distribution (as long as the fit function appropriate for the observed variable distribution is used; e.g., Bollen, 1989). Because the parametric function $f$ is assumed to be smooth, it follows that its estimator $\hat{f}$ also could be considered approximately normally distributed (e.g., Rao, 1973, chap. 6a; as long as none of the partial derivatives of $f$ vanish). Thus, an approximate $100(1 - \delta)\%$ – confidence interval for $f$ $(0 < \delta < 1)$ results then with Equation 11 as (e.g., Hays, 1994)

$$(\hat{f} - z_{1-\delta/2}\hat{\sigma}(f), \hat{f} + z_{1-\delta/2}\hat{\sigma}(f)), \tag{12}$$

where $z_{1-\delta/2}$ is the $\delta/2$th quantile of the standard normal distribution. Obtaining the expressions in Equation 11 and in particular Equation 12 can be considered the final step of an application of the delta method. Their computation in empirical research can be carried out using the SPSS command file also provided in the Appendix.

## APPLICATION OF THE DELTA METHOD FOR INTERVAL ESTIMATION OF TIME-INVARIANT RELIABILITY

In this section, to demonstrate the use of the delta method for obtaining approximate standard error and confidence interval of parametric functions in SEM, we employ data from a cognitive intervention study by Baltes, Dittmann-Kohli, and Kliegl (1986). The goal of their investigation was to examine reserve capacity of elderly adults in aging-sensitive tasks of fluid abilities. For our illustrative purposes, we utilize data from three repeated assessments with three fluid intelligence measures, the so-called ADEPT Induction, ADEPT Figural Relations, and Thurstone's Standard Induction tests, which can be considered markers of the construct fluid intelligence (e.g., Horn, 1982). Further details on the original study can be found in Baltes et al. (1986).

Suppose we are interested in the reliability coefficients of these measures. Capitalizing on the approach outlined in Raykov and Tisak (2004) for testing their stability across time and their point estimation, we apply here the delta method to obtain an approximate standard error and confidence interval for them. The essence of the Raykov and Tisak approach consisted in the following: (a) fitting to the covariance matrix of the nine observed variables (three markers of fluid intelligence evaluated at three occasions) a three-wave, three-indicator confirmatory factor analysis model that for completeness of this discussion is

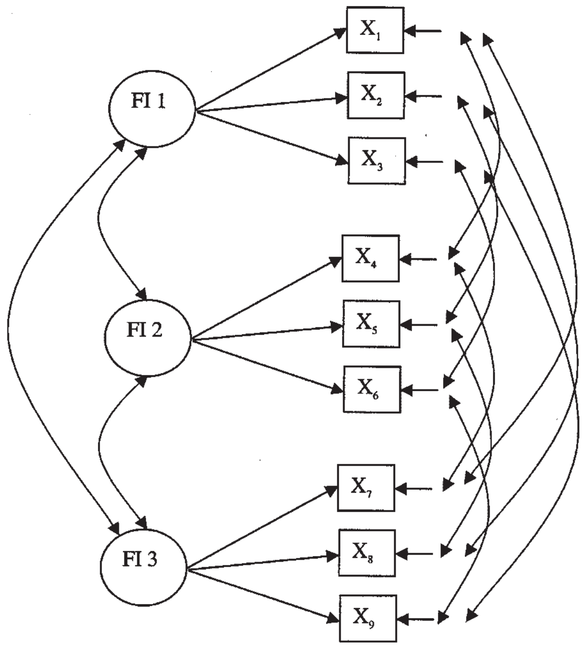


FIGURE 1 Three-wave, three-indicator confirmatory factor analysis model in Raykov and Tisak (2004) for examining stability in reliability accounting for variable specificity.

*Note.* FI 1, FI 2, FI 3 = fluid intelligence construct at first, second, and third measurement, respectively; remaining notation as in Table 1. Small horizontal one-way arrows denote measurement errors; for each of the three repeatedly administered tests, measurement error covariances for any two assessment occasions are introduced and set identical; factor loadings and error variances of same test set equal across time.

presented in Figure 1 as is their covariance matrix in Table 1 (Raykov & Tisak, 2004); (b) introducing in the model a single additional parameter per repeatedly presented test for its three cross-time measurement error covariances, denoted $\theta_1$*, $\theta_2$* and $\theta_3$* for the ADEPT Induction, ADEPT Figural Relations, and Thurstone's Induction tests, respectively; (c) imposing the requirement of identical factor loadings across time for each test, to ensure measurement invariance over time (e.g., Tisak & Tisak, 2002); and (d) showing that in their context under the assumption of time-invariance in reliability (which Raykov & Tisak, 2004, demonstrated in the underlying generic model to be tantamount to stability of pertinent error variance over time), these test reliability coefficients equal the following expression that accounts for indicator specificity variance (a main, confounding part of error variance; e.g., Mulaik, 1972):

TABLE 1
Fluid intelligence test covariance matrix ($N$ = 161; Raykov & Tisak, 2004)

| *Variable* | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ |
|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 328.112 | | | | | | | | |
| $X_2$ | 200.050 | 199.235 | | | | | | | |
| $X_3$ | 324.969 | 220.124 | 377.212 | | | | | | |
| $X_4$ | 310.488 | 200.811 | 330.119 | 338.014 | | | | | |
| $X_5$ | 202.093 | 173.222 | 224.539 | 202.286 | 213.317 | | | | |
| $X_6$ | 322.896 | 225.760 | 365.100 | 333.189 | 228.492 | 392.469 | | | |
| $X_7$ | 305.800 | 197.065 | 321.715 | 312.179 | 202.484 | 324.459 | 345.960 | | |
| $X_8$ | 219.358 | 187.923 | 244.277 | 220.427 | 197.847 | 242.081 | 224.301 | 251.710 | |
| $X_9$ | 315.551 | 220.475 | 355.877 | 325.068 | 223.498 | 358.247 | 326.814 | 243.735 | 376.839 |

*Note*. Matrix identical to that used in Raykov & Tisak (2004); $N$ = sample size; $X_1$, $X_4$, $X_7$ = ADEPT Induction test at 1st, 2nd and 3rd assessments, respectively; $X_2$, $X_5$, $X_8$ = ADEPT Figural Relations test at 1st, 2nd and 3rd assessments; $X_3$, $X_6$, $X_9$ = Thurstone's Standard Induction test at 1st, 2nd and 3rd assessments.

$$\rho_i = \frac{\lambda_i^2 + \theta_i^*}{\lambda_i^2 + \theta_i} \tag{13}$$

where $\lambda_i$ is the time-invariant loading of the $i$th repeated test, $\rho_i$ its constant reliability, and $\theta_i$ its stable error variance ($i = 1, 2, 3$ for the ADEPT Induction, ADEPT Figural Relations, and Thurstone's tests, respectively).

The approach in Raykov and Tisak (2004) described by (a) through (d) in the preceding paragraph allows one, for each longitudinally presented fluid measure, to accomplish two aims: (a) test for time-invariance in reliability and, in the affirmative case, (b) point estimate this coefficient accounting for measure specificity. For the data under consideration in this section (Table 1), Raykov and Tisak (2004) found with this approach that the underlying model in Figure 1 fitted well and yielded time-invariant reliability estimates as follows (estimates of constant loading, error covariances and variances correspondingly provided in right-hand sides; cf. Equation 13):

$$\hat{\rho}_1 = \frac{\hat{\lambda}_1^2 + \hat{\theta}_1^*}{\hat{\lambda}_1^2 + \hat{\theta}_{11}} = \frac{17.230^2 + 17.791}{17.230^2 + 41.662} = .929 \text{ for the ADEPT Induction test,} \tag{14}$$

$$\hat{\rho}_2 = \frac{\hat{\lambda}_2^2 + \hat{\theta}_2^*}{\hat{\lambda}_2^2 + \hat{\theta}_{22}} = \frac{12.124^2 + 39.580}{12.124^2 + 73.449} = .846 \quad \text{for the ADEPT Figural Relations test,} \tag{15}$$

and

$$\hat{\rho}_3 = \frac{\hat{\lambda}_3^{\,2} + \theta_3^{\,*}}{\hat{\lambda}_3^{\,2} + \theta_{33}} = \frac{19.070^2 + .700}{19.070^2 + 18.999} = .952 \quad \text{for Thurstone's Standard Induction test.} \tag{16}$$

The point estimates in Equations 14 through 16, however, do not contain any information as to how far they lie from the population time-invariant reliability coefficients of these fluid measures, which are the quantities of actual interest. To provide such information, we use here the delta method to obtain an approximate standard error and confidence interval for the respective reliability coefficients. To this end, first we recast in our earlier notation the generic formula for reliability, which underlies Equation 13, as follows:

$$\rho = \frac{\lambda^2 + \theta^*}{\lambda^2 + \theta} = \frac{\gamma_1^2 + \gamma_2}{\gamma_1^2 + \gamma_3} = \rho(\gamma_1, \gamma_2, \gamma_3) \tag{17}$$

denoting each involved model parameter by a corresponding γ. Using next differentiation Rules 5, 3, and 2 from the Appendix, we obtain the partial derivatives of the reliability coefficient in Equation 17:

$$D_1 = \frac{\partial \rho}{\partial \gamma_1} = \frac{2\gamma_1(\gamma_3 - \gamma_2)}{(\gamma_1^2 + \gamma_3)^2},\ D_2 = \frac{\partial \rho}{\partial \gamma_2} = \frac{1}{\gamma_1^2 + \gamma_3}, \tag{18}$$
$$D_3 = \frac{\partial \rho}{\partial \gamma_3} = -\frac{\gamma_1^2 + \gamma_2}{(\gamma_1^2 + \gamma_3)^2}.$$

Hence, as indicated previously, an approximate standard error for time-invariant reliability is obtained from Equations 11 and 18 as:

$$\hat{\sigma}(\rho) = [\hat{D}_1^{\,2}\sigma^2(\hat{\gamma}_1) + \hat{D}_2^{\,2}\sigma^2(\hat{\gamma}_2) + \hat{D}_3^{\,2}\sigma^2(\hat{\gamma}_3) + 2\hat{D}_1\hat{D}_2\sigma(\hat{\gamma}_1, \hat{\gamma}_2) + 2\hat{D}_1\hat{D}_3\sigma(\hat{\gamma}_1, \hat{\gamma}_3) + 2\hat{D}_2\hat{D}_3\sigma(\hat{\gamma}_2, \hat{\gamma}_3)]^{1/2} \tag{19}$$

with

$$\hat{D}_1 = \frac{2\hat{\gamma}_1(\hat{\gamma}_3 - \hat{\gamma}_2)}{(\hat{\gamma}_1^{\,2} + \hat{\gamma}_3)^2},\ \hat{D}_2 = \frac{1}{\hat{\gamma}_1^{\,2} + \hat{\gamma}_3},\ \hat{D}_3 = -\frac{\hat{\gamma}_1^{\,2} + \hat{\gamma}_2}{(\hat{\gamma}_1^{\,2} + \hat{\gamma}_3)^2}, \tag{20}$$

where for a given repeatedly presented fluid measure $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{\gamma}_3$ are the estimates of its respective time-invariant loading, error covariance, and variance (obtained when the model in Figure 1 is fitted to the covariance matrix in Table 1).

As mentioned previously, all quantities appearing in the right-hand side of Equation 19 and 20 are provided by the software used for model fitting purposes. In particular, with the inverse information matrix for this model given in

TABLE 2
Inverted Information Matrix Used (for Model in Figure 1)

| *Par* | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_1$* | $\theta_2$* | $\theta_3$* |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda_1$ | 1.092 | | | | | | | | |
| $\lambda_2$ | 0.649 | 0.777 | | | | | | | |
| $\lambda_3$ | 0.992 | 0.702 | 1.183 | | | | | | |
| $\theta_1$ | –.642 | –.239 | 0.618 | 31.589 | | | | | |
| $\theta_2$ | –.173 | –.248 | 0.220 | 6.103 | 40.972 | | | | |
| $\theta_3$ | 0.692 | 0.358 | –.855 | –23.543 | –8.393 | 34.081 | | | |
| $\theta_1$* | –.593 | –.248 | 0.621 | 28.253 | 5.944 | –23.696 | 29.653 | | |
| $\theta_2$* | –.174 | –.193 | 0.222 | 5.948 | 35.783 | –8.484 | 5.956 | 38.322 | |
| $\theta_3$* | 0.678 | 0.338 | –.817 | –23.667 | –8.460 | 31.298 | –23.271 | –8.314 | 32.232 |

*Note*. Matrix pertains to model in Figure 1 (when fitted to covariance matrix in Table 1). Par = parameter; $\lambda_1$ to $\lambda_3$ = time-invariant factor loadings (for the ADEPT Induction, ADEPT Figural Relations, and Thurstone's Standard Induction tests, respectively); $\theta_1$ to $\theta_3$ = time-invariant error variances (in same order); $\theta_1$* to $\theta_3$* = time-invariant error covariances (in same order). Main diagonal elements = squared standard errors of pertinent parameters; off-diagonal elements = estimated covariances for parameter estimator components corresponding to row and column.

Table 2, using the SPSS input file in the Appendix, one obtains for the constant reliability of the ADEPT Induction test the following derivative estimates: $\hat{D}_1 = .007$, $\hat{D}_2 = .000$, $\hat{D}_3 = -.003$; with them, Equations 11 and 12 yield now $\hat{\sigma}(\rho_1) = .016$ and (.894, .964) as approximate standard error and 95% confidence interval, respectively, for this reliability. These results suggest that the time-invariant reliability of the ADEPT Induction measure may be expected with a high degree of confidence to lie between the high .80s and mid .90s in the studied elderly population (and was estimated in Raykov & Tisak, 2004, at .929; see Equation 14).

Similarly, for the time-invariant reliability of the ADEPT Figural Relations test we obtain (with the same SPSS input file in which now the relevant, for this fluid measure estimates, squared standard errors and parameter covariances are substituted): $\hat{\sigma}(\rho_2)$ and (.789, .903) as approximate standard error and 95% confidence interval, respectively. This indicates that the constant reliability of the ADEPT Figural Relations test can be expected with high confidence to lie between the high .70s and low .90s (and was estimated at .846 in Raykov & Tisak, 2004; see Equation 15). Finally, in the same manner we obtain for Thurstone's Standard Induction measure reliability $\hat{\sigma}(\rho_3) = .016$ and (.920, .983) correspondingly as approximate standard error and 95% confidence interval, indicating that its time-invariant reliability could be expected with high degree of confidence to be located between the low .90s and high .90s (and was estimated in Raykov & Tisak, 2004, at .952; see Equation 16).

By way of summary, the findings in this section provide intervals of plausible values for the time-invariant reliability coefficients of the three repeatedly presented fluid intelligence measures from the Baltes et al. (1986) study, which were estimated at .929, .846, and .952 in Raykov and Tisak (2004), respectively. Accordingly, their constant reliability coefficients can be expected with high degree of confidence to lie in the elderly population of interest between the high .80s and mid .90s (ADEPT Induction test), between the high .70s and low .90s (ADEPT Figural Relations test), and between the low and high .90s (Thurstone's Standard Induction test). Thus, the delta method applied for interval estimation purposes in this section allowed us to supplement results from previous studies with plausible ranges of values for these time-invariant fluid intelligence test reliabilities in the studied older adult population.

## CONCLUSION

This article provided a didactic discussion of an analytic approach for approximate standard error and confidence interval construction, the delta method. The approach is readily and widely applicable in SEM. This procedure permits researchers to attach indexes of stability of estimation to parametric functions of substantive interests in entertained models and to construct ranges of plausible values for them in studied populations. Using rules and code provided in the Appendix, the discussed interval estimation approach can be straightforwardly applied, once a given model is fit to data and the inverse information matrix requested from the SEM software used for these purposes. The method is best employed with large samples due to the following reasons: (a) the SEM methodology underlying its utilization is based on an asymptotic statistical theory (Bollen, 1989) and (b) the linear approximation of a likely nonlinear parametric function of interest improves with larger samples because then due to the consistency property of the SEM parameter estimator (Bollen, 1989) a given model's solution can be expected to be close to the population parameter at which the pertinent first-order approximation of the function is taken.

In conclusion, the discussed application of the delta method for approximate interval estimation enhances researchers' capabilities to ask more complicated questions from their data and used covariance structure models, by providing a method for evaluating indexes of estimation precision and ranges of plausible values in studied populations for potentially complicated model parameter functions that reflect their substantive queries.

## ACKNOWLEDGMENTS

This research was conducted during visits to the Research Institute on Addictions with the State University of New York at Buffalo. Thanks are due to J. Tisak and S. Penev for valuable discussions on time-invariant reliability and approximate interval estimation and to P. B. Baltes, F. Dittmann-Kohli, and R. Kliegl for permission to use data from their project "Aging and Plasticity in Fluid Intelligence" for illustrative purposes.

## REFERENCES

Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve capacity of the elderly in aging-sensitive tasks of fluid intelligence: Replication and extension. *Psychology and Aging, 1,* 172–177.

Bentler, P. M. (2004). *EQS structural equation program manual.* Encino, CA: Multivariate Software.

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

Bollen, K. A., & Stine, R. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Thousand Oaks, CA: Sage.

Hays, W. L. (1994). *Statistics.* Fort Worth, TX: Harcourt Brace Jovanovic.

Horn, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 847–870). New York: McGraw-Hill.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36,* 109-133.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL8 user's reference guide.* Lincolnwood, IL: Scientific Software International.

McDonald, R. P. (1999). *Test theory. A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mulaik, S. A. (1972). *Foundations of factor analysis.* New York: McGraw-Hill.

Muthén, B. O., & Muthén, L. K. (2004). *Mplus user's guide.* Los Angeles: Muthén & Muthén.

Pawitan, Y. (2001). *In all likelihood. Statistical modeling and inference using likelihood.* Oxford, England: Clarendon.

Rao, C. R. (1973). *Linear statistical inference and its applications.* New York: Wiley.

Raykov, T. (2001). Estimation of congeneric scale reliability via covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology, 54,* 315–323.

Raykov, T. (2002). Analytic estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research, 37,* 89–103.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Raykov, T., & Penev, S. (2002). Interval estimation of latent interrelationship indices in a priori constrained covariance structure models. *Understanding Statistics, 1,* 139–156.

Raykov, T., & Penev, S. (in press). A direct method for obtaining an approximate standard error and confidence interval of maximal reliability for composites with congeneric measures. *British Journal of Mathematical and Statistical Psychology.*

Raykov, T., & Tisak, J. (2004). Examining time-invariance in reliability in multi-wave multi-indicator models: A covariance structure analysis approach accounting for indicator specificity. Manuscript submitted for publication.

Stewart, J. (1991). *Calculus.* Monterey, CA: Brooks/Cole.

Tan, S. J. (1990). *Applied calculus.* Boston: Kent.

Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods, 5,* 175–198.

# APPENDIX

## List of Frequently Used Derivative Rules

Rule 1. $\frac{\partial c}{\partial \gamma} = 0$, for any constant $c$ and any parameter $\gamma$; thus, the partial derivative of any given parameter with respect to any other parameter, is zero.

Rule 2. $\frac{\partial \gamma^r}{\partial \gamma} = r\gamma^{r-1}$, for any real number $r$.

Rule 3. $\frac{\partial[c_1 a_1(\gamma) + c_2 a_2(\gamma)]}{\partial \gamma} = \frac{c_1 \partial a_1(\gamma)}{\partial \gamma} + \frac{c_2 \partial a_2(\gamma)}{\partial \gamma}$, for any pair of constants $c_1$ and $c_2$, and any pair of differentiable functions $a_1(\gamma)$ and $a_2(\gamma)$ of $\gamma$; thus, the derivative of a linear combination of several differentiable functions is the same linear combination (i.e., with the same weights) of the derivatives of the functions.

Rule 4. $\frac{\partial[a_1(\gamma) a_2(\gamma)]}{\partial \gamma} = a_2(\gamma)\frac{\partial a_1(\gamma)}{\partial \gamma} + a_1(\gamma)\frac{\partial a_2(\gamma)}{\partial \gamma}$, for any pair of differentiable functions $a_1(\gamma)$ and $a_2(\gamma)$ of $\gamma$.

Rule 5. $\frac{\partial[a_1(\gamma) / a_2(\gamma)]}{\partial \gamma} = \frac{a_2(\gamma)\frac{\partial a_1(\gamma)}{\partial \gamma} - a_1(\gamma)\frac{\partial a_2(\gamma)}{\partial \gamma}}{a_2^{\,2}(g)}$, for any pair of differentiable functions $a_1(\gamma)$ and $a_2(\gamma)$ of $\gamma$ such that $a_2(\gamma) \neq 0$.

*Note*. In all five rules, $\gamma$ denotes a single variable (parameter) rather than a vector of such. Any of the previous rules can be applied on functions of more than a single argument, considering all but one of them as constants, to obtain the partial derivative of the function with respect to that variable argument. To keep the symbolism in these rules as close as possible to the one used in the main text, we chose to utilize partial derivative notation (despite the fact that formally the previous functions depend on a single argument). For variance approximation purposes, in this paper all derivatives are assumed non-vanishing at the model solution point.

## SPSS Input File for Computing Approximate Standard Error and Confidence Interval of Time-Invariant Reliability in Multiwave Multiindicator Models

```
TITLE 'SE AND CI FOR RELIABILITY IN MULTI-WAVE/INDICATOR MODELS'.
COMPUTE RHO = .929.
* ENTER IN PRECEDING LINE THE ESTIMATE OF TIME-INVARIANT
RELIABILITY.
COMPUTE LAMBDA = 17.230.
* ENTER IN PRECEDING LINE THE ESTIMATE OF INVARIANT FACTOR LOADING.
COMPUTE THETASTA = 17.791.
* ENTER IN PRECEDING LINE THE ESTIMATE OF INVARIANT ERROR
COVARIANCE.
COMPUTE THETA = 41.662.
* ENTER IN PRECEDING LINE THE ESTIMATE OF INVARIANT ERROR VARIANCE.
* IN THE NEXT 6 LINES ENTER PERTINENT ENTRIES OF THE MODEL I_I_M.
COMPUTE V1 = 1.092.
COMPUTE V2 = 29.653.
COMPUTE V3 = 31.589.
COMPUTE CG1G2 = -.593.
COMPUTE CG1G3 = -.642.
COMPUTE CG2G3 = 28.253.
* THE FOLLOWING 3 LINES CORRESPOND TO EQUATION 17.
COMPUTE D1 = 2*LAMBDA*(THETA-THETASTA)/(LAMBDA**2+THETA)**2.
COMPUTE D2 = 1/(LAMBDA**2+THETA)**2.
COMPUTE D3 = -(LAMBDA**2+THETASTA)/(LAMBDA**2+THETA)**2.
* THE FOLLOWING LINE CORRESPONDS TO EQUATION 16.
COMPUTE SE.RHO = SQRT(D1**2*V1+D2**2*V2+D3**2*V3+2*D1*D2*CG1G2+
2*D1*D3*CG1G3+2*D2*D3*CG2G3).
* THE FOLLOWING TWO LINES CORRESPOND TO EQUATION 9.
COMPUTE CI95.LO = RHO-1.96*SE.RHO.
COMPUTE CI95.UP = RHO+1.96*SE.RHO.
EXECUTE.
```

*Note*. Numbers entered in compute statements pertain to the ADEPT Induction test in the illustration section. I_I_M = inverse information matrix associated with fitted model (Figure 1; Raykov & Tisak, 2004). V1 to V3 = squared standard errors for time-invariant factor loading, error covariance and error variance (see Equation 16); CG1G2 to CG2G3 = estimated covariances for estimators of time-invariant factor loading, error covariance, and error variance (found as pertinent off-diagonal entries of I_I_M; see Table 2). This input file is unchanged if further waves or indicators are added to the fitted model (with time-invariant reliability; see Figure 1).