Supplemental material to accompany Preacher and Hayes (2008)

Kristopher J. Preacher — *University of Kansas*
Andrew F. Hayes — *The Ohio State University*

The multivariate delta method for deriving the asymptotic variance of a total indirect effect with three mediators

Bollen (1987, 1989; also see Brown, 1997), extending work by Alwin and Hauser (1975), Fox (1980, 1985), Greene (1977), and others, provides matrix formulae for obtaining point estimates and first-order *SE*s for any indirect effect in a simultaneous equation model. Bollen's method can be understood most easily by considering all variables in the system as endogenous latent variables or indicators of endogenous latent variables (any path analytic model, including multiple linear regression, can be reduced to the latter), even if they are theoretically unaffected by variables outside the system (the so-called "all-*y*" model). The matrix expression for the all-*y* structural model is:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \tag{1}$$

where $\boldsymbol{\eta}$ is an $m \times 1$ vector of m endogenous latent variables, $\mathbf{B}$ is an $m \times m$ matrix of path coefficients (direct effects) linking these variables, and $\boldsymbol{\zeta}$ is an $m \times 1$ vector of disturbance terms. The rows and columns of $\mathbf{B}$ correspond to the m endogenous latent variables, such that each element represents the effect of the column variable on the row variable. Thus, $\mathbf{B}$ contains point estimates of direct effects of every variable on all other variables. We restrict discussion to recursive models (i.e., no feedback loops), in which $\mathbf{B}$ can be arranged as lower triangular.

The data model linking the p dependent variables in the vector $\mathbf{y}$ ($p \times 1$) to the m latent variables in Equation 1 is:

$$\mathbf{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{2}$$

For simplicity, we let $p = m$, $\mathbf{\Lambda}_y = \mathbf{1}$, and $\boldsymbol{\varepsilon} = \mathbf{0}$ in Equation 2 to yield the path analysis model as a special case of SEM. Using the **B** matrix of path coefficients, point estimates of total effects can be obtained by the infinite sum (Folmer, 1981; Fox, 1980; Greene, 1977):

$$\mathbf{T} = \sum_{k=1}^{\infty} \mathbf{B}^k \tag{3}$$

where, as with **B**, rows and columns of **T** correspond to latent variables. A simpler formula (Bollen, 1987, 1989; Folmer, 1981; Sobel, 1988) is:

$$\mathbf{T} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} \tag{4}$$

where **I** is an $m \times m$ identity matrix. Because total effects are the sum of direct and total indirect effects, and the elements of **B** represent direct effects, point estimates of total indirect effects are given by Sobel (1986) as:

$$\mathbf{F} = \mathbf{T} - \mathbf{B} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} - \mathbf{B}. \tag{5}$$

In a model as depicted in Figure S1, **F** will contain all zeros except for a single element equal to $\Sigma_i(a_i b_i)$, $i = 1$ to j where j is the number of proposed mediator variables.

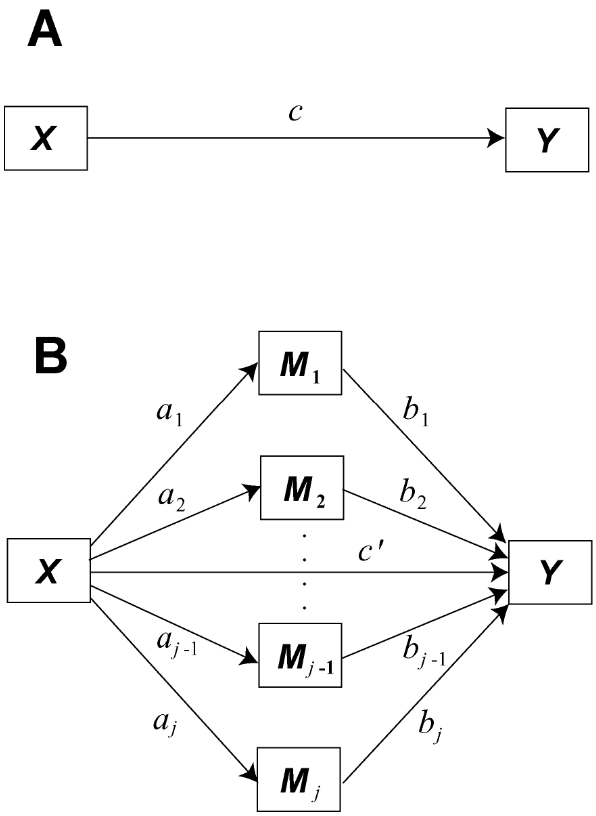

Figure S1. Illustration of a multiple mediation design with j mediators. X is hypothesized to exert indirect effects on Y through $M_1, M_2, \ldots, M_j$.

Equation 5 provides only point estimates for total indirect effects. Bollen (1987, 1989) and Sobel (1982, 1988) explain how the multivariate delta method can be used to determine first-order *SE*s of these indirect effects (assuming that maximum likelihood or generalized least squares minimization has been employed); these *SE*s, in turn, permit significance testing and CI construction. *SE*s for indirect effects are obtained as the square roots of diagonal elements of the asymptotic covariance matrix of **F**, given by:

$$\mathbf{\Sigma}(\mathbf{F}) = N^{-1}\left[\left(\frac{\partial \mathbf{f}}{\partial \hat{\boldsymbol{\theta}}_N}\right)' \mathbf{V}(\hat{\boldsymbol{\theta}}_N)\left(\frac{\partial \mathbf{f}}{\partial \hat{\boldsymbol{\theta}}_N}\right)\right], \tag{6}$$

where $\hat{\boldsymbol{\theta}}$ is a vector of sample estimates of free model parameters, **f** is a vector containing differentiable elements of **F**, and $N^{-1}\mathbf{V}(\hat{\boldsymbol{\theta}}_N)$ is the sample estimate of the asymptotic covariance matrix of $\boldsymbol{\theta}_N$. Consider the special case of multiple mediation with $j = 3$ mediators. In this case,

$$\mathbf{B} = \begin{array}{c} \\ X \\ M_1 \\ M_2 \\ M_3 \\ Y \end{array} \begin{array}{c} \begin{array}{ccccc} X & M_1 & M_2 & M_3 & Y \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ a_1 & 0 & 0 & 0 & 0 \\ a_2 & 0 & 0 & 0 & 0 \\ a_3 & 0 & 0 & 0 & 0 \\ c' & b_1 & b_2 & b_3 & 0 \end{bmatrix} \end{array}. \quad (7)$$

By Equation 5,

$$\mathbf{F} = \begin{array}{c} \\ X \\ M_1 \\ M_2 \\ M_3 \\ Y \end{array} \begin{array}{c} \begin{array}{ccccc} X & M_1 & M_2 & M_3 & Y \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_1b_1 + a_2b_2 + a_3b_3 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}. \quad (8)$$

There is only one total indirect effect, $\mathbf{f} = a_1b_1 + a_2b_2 + a_3b_3$, and the parameter vector is:

$$\boldsymbol{\theta}_N = [a_1 \quad a_2 \quad a_3 \quad b_1 \quad b_2 \quad b_3]'. \quad (9)$$

The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_N$ is:

$$N^{-1}\mathbf{V}(\hat{\boldsymbol{\theta}}_N) = \left[\begin{array}{c|c} \boldsymbol{\Sigma}(\hat{a}) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}(\hat{b}) \end{array}\right]. \quad (10)$$

where $\boldsymbol{\Sigma}(\hat{a})$ and $\boldsymbol{\Sigma}(\hat{b})$ are the full, symmetric covariance matrices of the a coefficients and b coefficients, respectively. The derivative of **f** with respect to $\boldsymbol{\theta}_N$ is:

$$\frac{\partial \mathbf{f}}{\partial \hat{\boldsymbol{\theta}}_N} = \begin{bmatrix} b_1 & b_2 & b_3 & a_1 & a_2 & a_3 \end{bmatrix}'. \quad (11)$$

Therefore, the asymptotic covariance matrix of **F** for this special case is:

$$\mathbf{\Sigma}(\mathbf{F}) = \begin{bmatrix} b_1 & b_2 & b_3 & a_1 & a_2 & a_3 \end{bmatrix} \left[\begin{array}{c|c} \mathbf{\Sigma}(\hat{a}) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Sigma}(\hat{b}) \end{array}\right] \begin{bmatrix} b_1 & b_2 & b_3 & a_1 & a_2 & a_3 \end{bmatrix}' \tag{12}$$

$$\begin{aligned} \mathbf{\Sigma}(\mathbf{F}) &= b_1^2\sigma_{a_1}^2 + a_1^2\sigma_{b_1}^2 + b_2^2\sigma_{a_2}^2 + a_2^2\sigma_{b_2}^2 + b_3^2\sigma_{a_3}^2 + a_3^2\sigma_{b_3}^2 \\ &+2(a_1a_2\sigma_{b_1,b_2} + a_1a_3\sigma_{b_1,b_3} + a_2a_3\sigma_{b_2,b_3} + b_1b_2\sigma_{a_1,a_2} + b_1b_3\sigma_{a_1,a_3} + b_2b_3\sigma_{a_2,a_3}), \end{aligned} \tag{13}$$

Bias corrected (BC) and bias corrected and accelerated (BCa) confidence intervals for indirect effects

With a percentile bootstrap 95% CI, the estimates in the 2.5^{th} and 97.5^{th} percentiles in the sorted distribution define the lower and upper bounds of the interval. Define Z_{lower} and Z_{upper} as the corresponding *z*-scores in a standard normal distribution (for a 95% CI, $Z_{\text{lower}} = -1.96$ and $Z_{\text{upper}} = 1.96$). Define Z'_{lower} and Z'_{upper} as the *z*-scores defining the percentiles for the bias corrected and accelerated (BCa) bootstrap CI. Specifically,

$$Z'_{\text{lower}} = Z_0 + \frac{Z_0 + Z_{\text{lower}}}{1 - \hat{a}(Z_0 + Z_{\text{lower}})} \tag{14}$$

where Z_0 is the *z*-score corresponding to the percent of the *k* bootstrap estimates that are less than the estimate in the original sample. Z'_{upper} is defined as in Equation 14, replacing Z_{lower} with Z_{upper}. The *acceleration constant* is defined as

$$\hat{a} = \frac{\sum_{i=1}^{n} (\bar{\theta} - \theta_i)^3}{6\left[\sum_{i=1}^{n} (\bar{\theta} - \theta_i)^2\right]^{3/2}} \tag{15}$$

where θ_i is the i^{th} "jackknife" estimate of *ab*, defined as the indirect effect computed after deleting case *i*, and $\bar{\theta}$ is the mean of the *n* jackknife estimates. Setting *a* to zero rather than estimating it yields a *bias corrected* (BC) confidence interval.

References

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*(February), 37-47.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology*, *17*, 37-69.

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

Brown, R. L. (1997). Assessing specific mediational effects in complex theoretical models. *Structural Equation Modeling*, *4*, 142-156.

Folmer, H. (1981). Measurement of the effects of regional policy instruments by means of linear structural equation models and panel data. *Environment and Planning A*, *13*, 1435-1448.

Fox, J. (1980). Effects analysis in structural equation models. *Sociological Methods and Research*, *9*, 3-28.

Fox, J. (1985). Effects analysis in structural equation models II: Calculation of specific indirect effects. *Sociological Methods and Research*, *14*, 81-95.

Greene, V. L. (1977). An algorithm for total and indirect causal effects. *Political Methodology*, *4*, 369-381.

Preacher, K. J., & Hayes, A. F. (2007). *Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models*. Manuscript submitted for publication.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 290-312). San Francisco: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), *Sociological Methodology 1986* (pp. 159-186). Washington, DC: American Sociological Association.

Sobel, M. E. (1988). Direct and indirect effects in linear structural equation models. In J. S. Long (Ed.), *Common problems / proper solutions*. Newbury Park, CA: Sage.