Methods and Approaches to Assessing Distal Mediation

Thomas D. Fletcher

University of Missouri – St. Louis

FletcherT@umsl.edu

www.umsl.edu/~fletchert

Fletcher, T. D. (2006, August). *Methods and approaches to assessing distal mediation*. Paper presented at the 66th annual meeting of the Academy of Management, Atlanta, GA.

Abstract

This paper briefly reviews the concept of mediation. The basic mediation model, as tested with the indirect effect or product of paths approach is extended to include multiple mediators in a causal chain ($X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$). The concept of distal mediation is explained and examined with regression, structural equation modeling and bootstrapping. It is shown through the use of the bootstrap technique that the sampling distribution of the indirect effect is non-normal creating asymmetrical confidence intervals about the estimate. A hypothetical example of distal mediation is used to illustrate the concepts.

KEYWORDS:

Mediation, bootstrapping, indirect effect

Methods and Approaches to Assessing Distal Mediation

Tests of mediation are common in the organizational and behavioral sciences. Methods for assessing mediation have been discussed extensively over several decades (Baron & Kenny, 1986; James & Brett, 1984; Rozeboom, 1956). For an extensive review of methods across several disciplines see the work by MacKinnon and colleagues (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). James and Brett (1984) describe mediation as a process in which "the influence of an antecedent is transmitted to a consequence through an intervening mediator" (p. 307). The organizational literature is replete with examples of mediational processes. Some theoretical models involving mediation include *theory of reasoned action* (Fishbein & Ajzen, 1975), *goal-setting theory* (Locke & Latham, 1990), and *attribution theories* (Mitchell, Green, & Wood, 1981; Weiner, 1985).

Researchers may want to extend the antecedent – mediator – consequence models to include more than one mediating variable. Table 1 shows several such examples. A researcher investigating these models may be interested in the mediation or indirect effect of the antecedent (X) on the consequence (Y) as it is mediated first by one mediator (M_1) and then another mediator (M_2). This process has been labeled distal mediation (Kenny, Kashy, & Bolger, 1998). However, to date, most of the research and examples on tests of mediation have involved only one mediator in the X,Y relationship. The purpose of this paper is to extend the mediation model to include more than one mediator in a causal sequence (i.e., $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$) and describe several plausible methods for assessing the mediation relationship.

INSERT TABLE 1 ABOUT HERE

Basic Mediation Models

Perhaps the most ubiquitous method to assess mediation in psychology is that proposed by Baron and Kenny (1986). Baron and Kenny (1986) suggest a causal steps approach where a consequence (*Y*) is first regressed on an antecedent (*X*), and then *Y* is regressed on a mediator (*M*) and *M* is regressed on *X*. However, the method used in most path analytic frameworks (e.g., structural equation modeling) involves the estimation of the indirect effect (MacKinnon et al., 2002). MacKinnon et al. further demonstrated that the Baron and Kenny (1986) causal steps method had the least power to detect mediation. One reason for this lack of power is explained below under distal mediation. Figure 1 depicts the simple path model involving an antecedent (*X*), a mediator (*M*) and a consequence (*Y*). The estimation of the indirect effect is described below.

INSERT FIGURE 1 ABOUT HERE

Indirect Effect Estimation

The model in Figure 1 is tested with the following equations:

$$Y = \tau X + \varepsilon_1 \tag{1}$$
$$Y = bM + \tau' X + \varepsilon_2 \tag{2}$$
$$M = aX + \varepsilon_3 \tag{3}$$

Here, *Y* is the consequence, *X* is the antecedent, *M* is the mediator, ε is the disturbance or error terms from each equation and a, b, τ, τ' are the corresponding regression coefficients for each equation. From Equations 1-3, it can be demonstrated that $\tau - \tau' = ab$ (MacKinnon et al., 2002). The mediated effect ($\tau - \tau'$) is the product of the indirect paths ($a \times b$).

James and Brett (1984) argued that the test of mediation is the joint significance of path a and path b in Figure 1. However, in their articulation, there was no requirement for controlling for the antecedent (*X*) when regressing the consequence (*Y*) on the mediator (*M*). In other words, James and Brett (1984) did not argue for the inclusion of path τ'. Without controlling for *X*, ($a \times b$) is not equivalent with ($\tau - 0$), unless *X* is completely unrelated to *Y*. One might have reasons to accurately assess the effect size of the indirect effect. For instance, in a utility sense, one might need to know the mediated effect of *X* on *Y*. Dropping *X* from Equation 2 would likely result in a biased estimate of b.

Standard Error of Indirect Effect

The calculation of the standard error is useful in tests of significance or to construct confidence intervals about the point estimate. However, many researchers have demonstrated problems with using the standard error of the indirect effect for hypothesis testing (Bollen & Stine, 1990; MacKinnon et al., 2002). Sobel (1982) approximated the standard error as Equation 4:

$$s_{ab} = \sqrt{a^2 s_b^2 + b^2 s_a^2}\ , \qquad (4)$$

where a and b are defined as above, and s is the corresponding standard errors for $a, b,$ and ab. The standard error is then used to construct a Z test (ab / s_{ab}) or construct confidence intervals ($ab \pm z_{(1-\alpha/2)} \cdot s_{ab}$).

Distal Mediation Models

Most of the examples in the literature involving causal chains with two or more mediators are estimated within a structural equation modeling framework (e.g., Ahearne, Mathieu, & Rapp, 2005; Thompson, 2005). Other sources show the estimation of the indirect effect without a corresponding test of its significance (cf. Cohen, Cohen, West, & Aiken, 2003). However, the steps involved can be explicated with ordinary regression techniques. The next section describes a simple distal mediation process followed by a section on the estimation of the standard error of the indirect effect.

INSERT FIGURE 2 ABOUT HERE

Indirect Effect Estimation

The models involved in a simple distal mediation example are depicted in Figure 2. The models in Figure 2 can be represented by Equations 5-8.

$$Y = \tau X + \varepsilon_1 \quad (5)$$
$$Y = \tau' X + fM_1 + cM_2 + \varepsilon_2 \quad (6)$$
$$M_2 = eX + bM_1 + \varepsilon_3 \quad (7)$$
$$M_1 = aX + \varepsilon_4 \quad (8)$$

Equation 5 represents the total effect of *X* on *Y*. Equations 6-8 represent the mediated effect of *X* on *Y*. In structural equation modeling programs (e.g., LISREL; Jöreskog & Sörbom, 1996), equations 6-8 are estimated simultaneously. There are several indirect paths from *X* to *Y*. Therefore, the indirect effect in this set of equations is not simply the product of paths abc, but is equal to the sum of all the indirect paths: $abc + af + ec$. The mediated effect is equal to the total indirect effect: $\tau - \tau' = abc + af + ec$. It should be readily seen that the total effect of *X* on *Y* (τ) will be small when there are many variables mediating the *X,Y* relationship. The total effect cannot exceed the sum of the indirect effects. This relationship leads to power issues. That is, the more distal the *X,Y* relationship, the lower the power to detect the total effect and the greater the need for increased sample size.

Standard Error of Indirect Effect

When the indirect effect involves a distal mediation process, the computation of the standard error can be generalized from the standard error of the product of two paths using the multivariate delta method (Sobel, 1982). The standard error of the indirect effect ($abc + af + ec$) is computed with Equation 9:

$$s_{(abc+af+ec)} = \sqrt{s_{abc}^2 + s_{af}^2 + s_{ec}^2 + 2*\left(s_{abc,af} + s_{abc,ec} + s_{af,ec}\right)}, \qquad (9)$$

where s_{ec} and s_{af} are computed similarly to the product of ab above (Equation 4), and s_{abc} is computed as (Eq. 14 in Bollen & Stine, 1990):

$$s_{abc} = \sqrt{a^2b^2s_c^2 + a^2c^2s_b^2 + b^2c^2s_a^2} \qquad (10)$$

Finally, $s_{abc,af}, s_{abc,ec}, s_{af,ec}$ are the covariances among the individual indirect effects. Their estimation using the multivariate delta method is presented in the Appendix.

The standard error in Equation 9 is then used for hypothesis testing or construction of confidence intervals in the usual manner. There is sufficient reason to suspect that the indirect effect of a distal mediation process is not normally distributed diminishing the utility of this approximated standard error in both hypothesis testing and the construction of symmetrical confidence intervals (Bollen & Stine, 1990; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002).

Partial vs. Full Mediation

Before describing different methods for assessing the indirect effect as described above, I must address the issue of full versus partial mediation. Given the model in Figure 2, partial versus full mediation entails the estimation of paths e, f, and τ' or not. That is, if e, f, and τ' are each non-significant, one might conclude full-mediation and that these paths should not be estimated in the model. If the paths are significant, then one has partial mediation, provided the indirect effect is itself significant.

However, the issue is not so much in the significance of each individual path (e, f, and τ'), but in the potential biasing effect of not estimating these paths. Shrout and Bolger (2002) demonstrated several cases in which a researcher might conclude full-mediation in a simple *X* → *M* → *Y* relationship, but that the direct path *X* → *Y* should also be included in estimation. That is, the effect of *X* on *Y* should be controlled for in Equation 2 (see above). Non-significance is not necessarily equivalent with nil (Cohen, 1994). Suppose Equation 2 is modified by dropping *X* and assuming full-mediation. The resulting model in this simple mediation example is:

$$Y = b'M + \varepsilon \qquad (11)$$

where b' is the regression coefficient when *not* controlling for *X*. Using equations 1, 11 and 3, $\tau - 0 \neq ab'$ as one might expect. This is because τ' rarely exactly equals zero (0), and b in Equation 2 is the partial regression coefficient after controlling for *X*. With distal mediation, the complication is exacerbated because there are many individual indirect effects and the potential for many statistically non-significant paths influencing the estimation of the significant paths. If the model is exploratory, then the direct paths should be estimated to reduce concerns of producing biased estimates of the indirect effect. However, if the model is a confirmatory test of well known relationships, one might exclude the estimation of the direct paths. In either case, a researcher should make explicit the intentions.

Methods for Assessing Mediation

Ordinary Least Squares

The preponderance of research on mediation uses ordinary least squares (OLS) or simply linear regression (see MacKinnon et al., 2002). With OLS, the three equations (Equations 6-8) are each separately estimated. When the model is recursive (i.e., no feedback loops), separate estimation poses no problem. When the assumptions are met (e.g., the residuals are all normally distributed), OLS produces unbiased estimates of the regression coefficients. These coefficients are then used to compute the indirect effect as described above (e.g., $abc + af + ec$). However, using OLS, the standard errors of each coefficient may not be asymptotically correct (Bollen, 1989). Further, it has been described elsewhere that in simple mediation models, the product of the paths ($a \times b$) divided by its standard error is not distributed as standard normal (Bollen & Stine, 1990; MacKinnon et al., 2002; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002). It has been demonstrated repeatedly that the sampling distribution of the product of the paths ($a \times b$) in Figure 1 is asymmetrical. Therefore, the development of symmetrical confidence intervals is problematic. One can only think these problems to be greater in models involving distal mediation.

Structural Equation Modeling

In structural equation modeling (SEM), a model is formulated and tested such that Equations 6-8 are simultaneously estimated. Further, most packages (e.g., LISREL; Jöreskog & Sörbom, 1996) will decompose the effects and their corresponding standard errors. That is, the output will produce the total effect (τ), the direct effect (τ') and the indirect effect ($abc + af + ec$) along with each of the standard errors. This is done with the assistance of the computer rather than the tedious computation of the standard errors. The added benefit is that the maximum likelihood estimates of the standard errors for the individual paths are asymptotically correct (Bollen, 1989). There are however drawbacks to using SEM to test distal mediation. Only the total indirect effect is provided by the software; the specific indirect effects are not provided (see Bollen, 1989; Bollen & Stine, 1990). For instance, if we wanted to know, given the model in

Figure 2, the indirect effect *X* on *Y* specifically through M_1 (i.e., af), we would have to re-specify the model potentially biasing the parameters. Finally, SEM does not resolve the issue of the sampling distribution of the indirect effect. Because the sampling distribution is believed to be skewed, the standard errors are not useful in constructing confidence intervals.

Bootstrapping

There are a number of introductions to the basic principles of the bootstrap (Efron & Tibshirani, 1993; Davison & Hinkley, 1997; Mooney & Duval, 1993; Wood, 2005). Likewise, a number of researchers have heralded the bootstrap procedure as the new method to deal with the non-normality inherent in the estimation of the indirect effect (Bollen & Stine, 1990; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002). Indeed, resampling methods such as the bootstrap procedure do not require the same assumptions as parametrical inference tests and are now available in a wide variety of software platforms. For example, R and S-Plus have packages built into the program corresponding to several texts (Davison & Hinkley, 1997; Efron & Tibshirani, 1993; Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2005). Shrout and Bolger (2002) have created macros for use in SPSS (SPSS, 1999) and demonstrate the use of bootstrapping in EQS (Bentler, 1997) for computation of the indirect effect. Preacher and Hayes (2004) have created macros for SAS and SPSS to compute the indirect effect and bootstrap confidence intervals. LISREL too has bootstrapping capabilities built into the program (Jöreskog & Sörbom, 1996). The ease with which these are useful in the present context varies greatly across programs.

I will illustrate the basic principles of the bootstrap procedure with a simple example. First, to utilize bootstrapping, one needs to work with the raw data, not a covariance matrix. Suppose the data consist of the following five scores on some variable: [2, 3, 3, 4, 5]. Now suppose a single bootstrap sample is taken with replacement. This means that at random, five scores will be randomly culled from the original sample, each original value having an equal probability of being chosen. The bootstrap sample could be: [2, 2, 3, 5, 5]. Notice that in the bootstrap sample, 2 and 5 are each taken twice, 3 is only taken once and 4 is not taken at all. A second bootstrap sample may contain [2, 2, 3, 3, 4]. From the samples, the desired statistic is computed. This procedure is repeated *R* times. There is considerable debate as to what constitutes an adequate number of bootstrap replications, but Hinkley and Davison (1997) have suggested 999 to be sufficient for most purposes. For the distal mediation depicted in Figure 2, the indirect effect ($abc + af + ec$) would be estimated from each sample and stored for analysis. The standard deviation of these estimates is considered the bootstrapped standard error. Finally, the 2.5^{th} and 97.5^{th} percentile represents the 95% confidence interval. That is, to obtain the $(1 - \alpha)*100$ % confidence interval, take the middle $(1 - \alpha)*100$ percentage of sorted observations. There are a number of other more sophisticated methods (e.g., the bias-corrected bootstrap), but the percentile method has been demonstrated to be reasonably adequate for estimating indirect effects (MacKinnon et al., 2004; Shrout & Bolger, 2002). With 999 replicates, one simply sorts the replicates and takes the 25^{th} and 975^{th} sorted observation to be the 95% confidence interval.

It is easy enough given the computing power at one's fingertips to use the bootstrap methodology for estimating the indirect effect. While the standard error is simply the standard deviation of the bootstrapped sampling distribution, the construction of confidence intervals does not rely on the estimation of the standard error. When skew is present in the sampling distribution, the confidence intervals need not be symmetrical. A drawback of the bootstrapping procedure is that it may require some programming depending on the complexity of the estimate to be bootstrapped and the software package of choice. For instance, I have created functions for use in R (and S-Plus) to estimate both proximal and distal mediation and to bootstrap the estimates for the illustrations below. The bootstrapping procedures are based on the functions in R (and S-Plus) that correspond to Davison and Hinkley (1997).

Illustrative (Hypothetical) Example

To illustrate the concepts, similarities and differences in the methods for assessing distal mediation I have constructed a hypothetical correlation matrix in Table 2. The pattern of correlations is consistent with what might be obtained in a prototypical distal mediation situation. Variables closer to one another in the causal chain are more closely correlated. From this hypothetical correlation matrix, separate, independent datasets were generated corresponding to sample sizes of 100, 200, 400 and 10,000. The first three are typical sample sizes found in multiple regression and SEM research in the social sciences. The latter sample size was estimated to approach population estimates. Each sample was made independent by changing the random seed generator. This correlation matrix corresponds to true population values for the regression coefficients $a, b, c, e, f, \tau,$ and τ' as .300, .280, .280, .066, .062, .075, and .015 respectively. These correspond to a total indirect effect ($abc + af + ec$) of .060. Each independent sample is expected to vary from the population values due to sampling variability.

INSERT TABLE 2 ABOUT HERE

For each sample, the model in Figure 2 and Equations 6-8 was estimated using OLS, SEM and bootstrapping. Only the total indirect effect was examined, although specific indirect effects could be estimated using OLS and bootstrapping with little additional effort. The estimated indirect effect, the standard error of the indirect effect and the 95% confidence intervals are presented in Table 3. The most obvious characteristic in Table 3 is that all three methods produced exactly the same estimate of the indirect effect for each sample. While the standard errors are quite similar for each approach there is some variability. Finally, the bootstrap confidence intervals based on the percentile method are far from symmetrical. To further explore this asymmetry, the density of the sampling distribution (i.e., smoothed histogram) is plotted for each sample with a corresponding normal curve with means and standard deviations equal to the bootstrap sampling distribution (see Figure 3). While the bootstrap sampling distributions approach normality, they are not distributed exactly as normal. As sample size increases, the bootstrap distribution appears to approximate normality.

INSERT TABLE 3 ABOUT HERE

INSERT FIGURE 3 ABOUT HERE

The four examples were generated with normal data for all variables. However, it is not uncommon in organizational research to encounter variables with slight to moderate skew (e.g., attitudes such as job satisfaction tend to be negatively skewed and studies involving compensation frequently encounter skewed data; Russell & Dean, 2000). To see what affect this might have on the three methods (i.e., OLS, SEM and bootstrap), a new data set was generated with the hypothetical correlation matrix in Table 2 for N = 200. The consequence (*Y*) was created with moderate negative skew. The results of the three methods with the skewed outcome variable are as follows. For all three methods, the indirect effect is .032. The standard errors are .0284, .0283, and .0269 for OLS, SEM and bootstrapping respectively. Finally, the 95% confidence intervals are (-.0240, .0874), (-.0238, .0872), (-.0215, .0871). OLS and SEM are very similar with respect to the standard errors. For all three methods, the conclusion is that the indirect effect is not significantly different from zero, but for the bootstrapping method, the intervals are both more narrow and asymmetrical. Figure 4 shows the density curve for the bootstrap sampling distribution with a normal curve with equal mean and standard deviation superimposed. Here, the impact of non-normality is more evident. Here, I have only demonstrated the effect of skewness on one outcome variable, but in distal mediation, there are three outcome variables that could potentially be in violation of the assumption of normality.

INSERT FIGURE 4 ABOUT HERE

Distribution of the Indirect Effect

The distribution of a product is a somewhat complicated affair. Goodman (1960) worked out the variance of product terms some four decades ago. However, the indirect effect of a distal mediation process is not simply the product of two variables. It involves the sum of three products. The distribution of such a variable is unknown. To explore how $abc + af + ec$ might be distributed, five variables were generated with means and standard deviations corresponding to the population values and their standard errors (for N = 400) based on the correlation matrix shown in Table 2. That is for $a, b, c, e,$ and f respectively, the estimates and standard deviations were simulated as: .30 (.05), .28 (.05), .28 (.05), .07 (.05), .06 (.05). Data were generated for five variables corresponding to $a, b, c, e,$ and f with N = 10,000 observations. The sum of the products, $abc + af + ec$ was then computed. The density of this distribution is shown in Figure 5 with a corresponding normal curve superimposed. The density of this distribution is more leptokurtic (i.e., narrower and peaked) and slightly skewed relative to a normal distribution with equal mean and standard deviation. This is more evidence that bootstrapping may be a better approach to estimating complex variables such as the indirect effect when distal mediation is present.

INSERT FIGURE 5 ABOUT HERE

Summary and Conclusion

There are indeed a number of theoretical models in the organizational and behavioral sciences that rely on tests of mediation. MacKinnon and colleagues (MacKinnon et al., 2002; MacKinnon et al., 2004) have advanced the understanding of mediation in a number of ways. In their comprehensive review, MacKinnon et al. (2002) compared several respected methods for assessing mediation from a number of separate disciplines (e.g., psychology, sociology, epidemiology). In that review, they recommend the product of coefficients approach, but note that ab/s_{ab} is not distributed as standard normal as previously thought. MacKinnon et al. (2004) further explored the distribution of ab by using various resampling techniques. This paper built on those works by extending the product of coefficients approach and the use of the bootstrap resampling technique to models involving distal mediation.

In this paper I demonstrated with a running hypothetical example that SEM and OLS tend to return similar results under conditions of normality. The major pitfall of both is that the indirect effect is not normally distributed as is assumed when constructing confidence intervals. Consistent with what might be expected based on several studies involving mediation and bootstrapping, I found the bootstrapping technique to produce similar results to SEM and OLS, but the results were marginally more accurate because bootstrapping confidence intervals reflect asymmetry found in the distribution of the indirect effect. I only demonstrated these principles using five independent samples of varying sizes. To fully understand the variability (if any) or to generalize beyond this paper, a full Monte Carlo simulation is warranted. Simple examples were used because the purpose of this paper was to extend known concepts related to mediation to distal mediation not to comprehensively investigate differences in estimation methods.

Another benefit of the bootstrap articulated here is that bootstrapping estimates of the indirect effect does not require the complexity involved with the multivariate delta method (Sobel, 1982). SEM programs such as LISREL are useful for decomposing the effects (i.e., computing the direct, indirect and total effects), but at present only produce the total indirect effect in complex models. For simple models such as those discussed in this paper (see Figure 1 and Figure 2), I have put together a set of functions for use in R (R Development Core Team, 2005) that can be used to test mediation and bootstrap the estimates. R is a freely available statistical platform that is both flexible and powerful. These functions will produce each effect in the models, their standard error based on the multivariate delta method, *t*-test, *p*-value, and the ratio of the indirect effect to the total effect. Examples are included with the functions to show how to bootstrap the estimates to create 95% confidence intervals.

References

Ahearne, M., Mathieu, J., & Rapp, A. 2005. To empower or not to empower your sales force? An empirical examination of the influence of leadership empowerment behavior on customer satisfaction and performance. ***Journal of Applied Psychology***, 90: 945-955.

Baron, R. M. & Kenny, D. A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. ***Journal of Personality and Social Psychology***, 51: 1171-1182.

Bentler, P. 1997**. *EQS for Windows*** (Version 5.6) [Computer software]. Los Angeles: Multivariate Software, Inc.

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

Bollen, K. A. & Stine, R. 1990. Direct and indirect effects: Classical and bootstrap estimates of variability. ***Sociological Methodology,*** 20: 115-140.

Cohen, J. 1994. The earth is round ($p < .05$). ***American Psychologist,*** 49: 997-1003.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. 2003. ***Applied multiple regression/correlation analysis for the behavioral sciences*** (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Davison, A. C. & Hinkley, D. V. 1997. ***Bootstrap methods and their application.*** Cambridge, UK: Cambridge University Press.

Efron, B., & Tibshirani, R. 1993. ***An introduction to the bootstrap***. New York: Chapman & Hall.

Fishbein, M, & Ajzen, I. 1975. ***Belief, attitude, intention and behavior: An introduction to theory and research***. Reading, MA: Addison-Wesley.

Goodman, L. A. 1960. On the exact variance of products. ***Journal of the American Statistical Association***, 55: 708-713.

Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. 2005. Bootstrap methods and permutation tests. In D. S. Moore & G. P. McCabe (Eds.), ***Introduction to the practice of statistics*** (5th ed.). New York: W. H. Freeman and Company.

James, L. R. & Brett, J. M. 1984. Mediators, moderators, and tests for mediation. ***Journal of Applied Psychology***, 69: 307-321.

Jöreskog, K. & Sörbom, D. 1996. ***LISREL 8: User's reference guide***. Lincolnwood, IL: Scientific Software International.

Locke, E. A. & Latham, G. P. 1990. ***A theory of goal setting and task performance***. Englewood Cliffs, NJ: Prentice Hall.

Kenny, D. A., Kashy, D. A., & Bolger, N. 1998. Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), ***Handbook of social psychology (4th ed.):*** Vol. 1, pp. 233-265: New York: McGraw-Hill.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. 2000. Equivalence of the mediation, confounding and suppression effect. ***Prevention Science***, 1: 173-181.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. 2004. Confidence limit for indirect effect: distribution of the product and resampling methods. ***Multivariate Behavioral Research***, 39: 99-128.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. 2002. A comparison of methods to test mediation and other intervening variable effects. ***Psychological Methods***, 7: 83-104.

Mitchell, T. R., Green, S. G., & Wood, R. E. 1981. An attributional theory of leadership and the poor performing subordinate: Development and validation. ***Research in Organizational Behavior***, 3: 197-234.

Mooney, C. Z., & Duval, R. D. 1993. ***Bootstrapping: A nonparametric approach to statistical inference***. Newbury Park, CA: Sage.

Preacher, K. J. & Hayes, A. F. 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. ***Behavior Research Methods, Instruments, and Computers***, 36: 717-731.

R Development Core Team. 2005. ***R: A language and environment for statistical computing***. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org

Rozeboom, W. W. 1956. Mediation variables in scientific theory. ***Psychological Review,*** 63: 249-264.

Russell, C. J. & Dean, M. A. 2000. To log or not to log: bootstrap as an alternative to the parametric estimation of moderation effects in the presence of skewed dependent variables. ***Organizational Research Methods***, 3: 166-185.

Shrout, P. E. & Bolger, N. 2002. Mediation in experimental and nonexperimental studies: New procedures and recommendations. ***Psychological Methods***, 7: 422-445.

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. ***Sociological Methodology***, 16: 159-186.

SPSS Inc. (1999). ***SPSS*** (Version 9). [Computer software]. Chicago: Author.

Thompson, J. A. 2005. Proactive personality and job performance: A social capital perspective. ***Journal of Applied Psychology***, 90: 1011-1017.

Weiner, B. 1985. An attributional theory of achievement motivation and emotion. ***Psychological Review***, 92: 548-573.

Wood, M. 2005. Bootstrapped confidence intervals as an approach to statistical inference. ***Organizational Research Methods***, 8: 454-470.

Appendix

For the model in Figure 2 and Equations 6-8 we have the regression coefficients a, b, c, e, f and the individual indirect effects abc, af, and ec. Using the multivariate delta method, a matrix is constructed based on the indirect effects and the corresponding regression coefficients (or parameter estimates) – see Table A1. If this matrix is called *F*, and the variance-covariance matrix of the regression parameters is called *S*, then the variance-covariance matrix of the indirect effects is computed as the matrix multiplication *FSF'* (see Equation A1).

$$FSF' = \begin{bmatrix} bc & ac & ab & 0 & 0 \\ f & 0 & 0 & 0 & a \\ 0 & 0 & e & c & 0 \end{bmatrix} \begin{bmatrix} s_a^2 & 0 & 0 & 0 & 0 \\ 0 & s_b^2 & 0 & s_{b,e} & 0 \\ 0 & 0 & s_c^2 & 0 & s_{c,f} \\ 0 & s_{b,e} & 0 & s_e^2 & 0 \\ 0 & 0 & s_{c,f} & 0 & s_f^2 \end{bmatrix} \begin{bmatrix} bc & f & 0 \\ ac & 0 & 0 \\ ab & 0 & e \\ 0 & 0 & c \\ 0 & a & 0 \end{bmatrix} \qquad \text{(A1)}$$

The result is a 3 x 3 symmetric matrix. The square-root of the diagonal gives the standard error for each indirect effect (abc, af, and ec). The square-root of the sum of the matrix elements gives the standard error of the total indirect effect ($\tau - \tau' = (abc) + (af) + (ec)$). Equation A2 depicts the symmetric matrix without the elements above the diagonal.

$$\begin{bmatrix} s_{abc}^2 & - & - \\ s_{abc,af} & s_{af}^2 & - \\ s_{abc,ec} & s_{af,ec} & s_{ec}^2 \end{bmatrix} = \begin{bmatrix} b^2c^2s_a^2 + a^2c^2s_b^2 + a^2b^2s_c^2 & - & - \\ bcfs_a^2 + a^2bs_{c,f} & f^2s_a^2 + a^2s_f^2 & - \\ ac^2s_{b,e} + abes_c^2 & aes_{c,f} & e^2s_c^2 + c^2s_e^2 \end{bmatrix} \qquad \text{(A2)}$$

Equation A3 is equivalent to Equation 9 in the text.

$$s_{abc+af+ec} =$$

$$\sqrt{\left(b^2c^2s_a^2 + a^2c^2s_b^2 + a^2b^2s_c^2\right) + \left(f^2s_a^2 + a^2s_f^2\right) + \left(e^2s_c^2 + c^2s_e^2\right) + 2*\left([bcfs_a^2 + a^2bs_{c,f}] + [ac^2s_{b,c} + abes_c^2] + [aes_{c,f}]\right)} \qquad \text{(A3)}$$

TABLE 1
Examples of Distal Mediation in the Organizational and Behavioral Sciences

X →	M_1 →	M_2 →	Y
Leader-Member Exchange	Organizational Commitment	Turnover Intentions	Actual turnover
Encouragement and training in goal setting	Actual Goal-setting	Behavioral Intentions	Performance
Extraversion	Job Satisfaction	Job Dedication	Task Performance
Team Cohesion	Team Monitoring	Backup Behaviors	Team Performance

TABLE 2
Hypothetical Correlation Matrix for Variables Involving Distal Mediation

	X	M_1	M_2	Y
X	1.000			
M_1	0.300	1.000		
M_2	0.150	0.300	1.000	
Y	0.075	0.150	0.300	1.000

TABLE 3
Results of Three Methods for Four Samples and Sample Sizes

				95% CI	
N	**Method**	**Effect Est.**	**SE**	**LCL**	**UCL**
	OLS	0.0948	0.0512	-0.0054	0.1951
100	SEM	0.0948	0.0507	-0.0045	0.1941
	Bootstrap	0.0948	0.0576	-0.0077	0.2193
	OLS	0.0664	0.0254	0.0166	0.1162
200	SEM	0.0664	0.0254	0.0167	0.1161
	Bootstrap	0.0664	0.0245	0.0259	0.1225
	OLS	0.0536	0.0230	0.0084	0.0987
400	SEM	0.0536	0.0230	0.0085	0.0986
	Bootstrap	0.0536	0.0232	0.0091	0.0989
	OLS	0.0525	0.0040	0.0447	0.0602
10,000	SEM	0.0525	0.0040	0.0447	0.0602
	Bootstrap	0.0525	0.0041	0.0441	0.0606

Note. Effect est. is $abc + af + ec$. OLS is ordinary least squares. SEM is structural equation modeling. For OLS and SEM, the 95% confidence interval is computed as estimate +/- 1.96 * standard error. For Bootstrap, the 95% CI is computed as the 25th and 975th sorted observation from 999 replicates.

TABLE A.1
Heuristic F Matrix Depicting the Indirect Effects and the Regression Coefficients

Indirect Effects	Regression coefficients				
	a	*b*	*c*	*e*	*f*
abc	*bc*	*ac*	*ab*	0	0
af	*f*	0	0	0	*a*
ec	0	0	*e*	*c*	0

Note. F is the 3 x 5 matrix elements of this table.

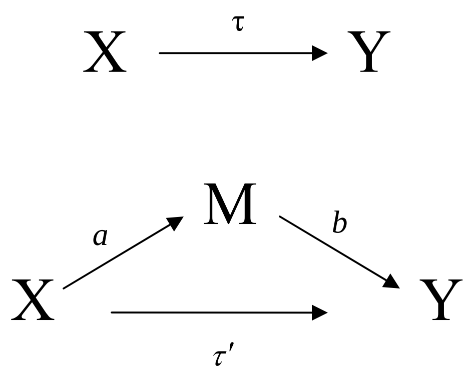

FIGURE 1
Upper model depicts total effect of X on Y and lower model depicts effect of X on Y mediated by M.

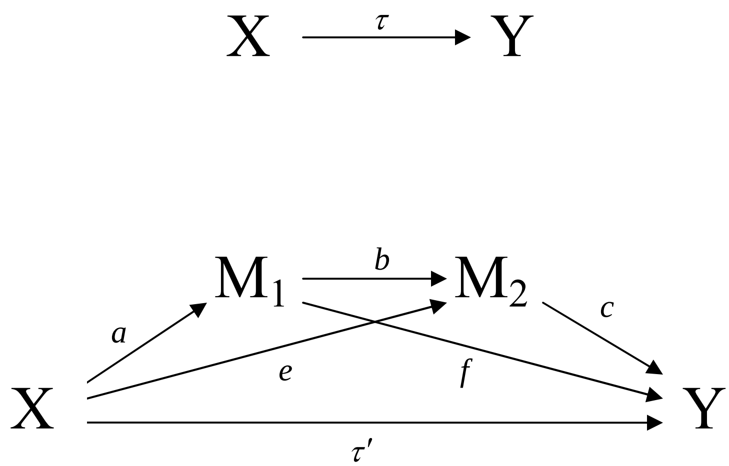

FIGURE 2
Distal mediation models. Upper model depicts total effect of X on Y and lower model depicts distal effect of X on Y mediated by M_1 and M_2.

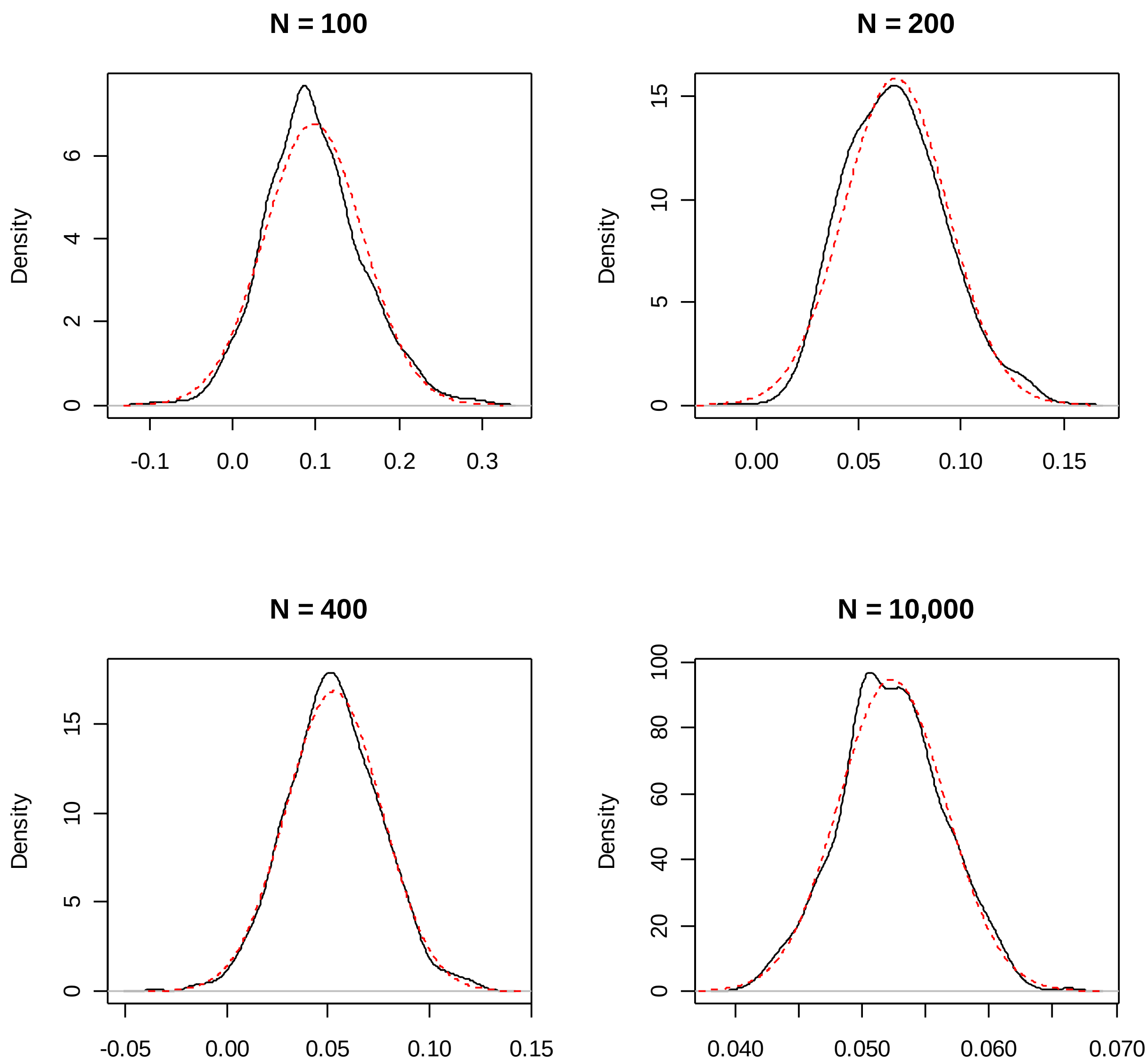

FIGURE 3
Bootstrap sampling distributions for all examples with superimposed Normal distribution with same mean and standard deviation.

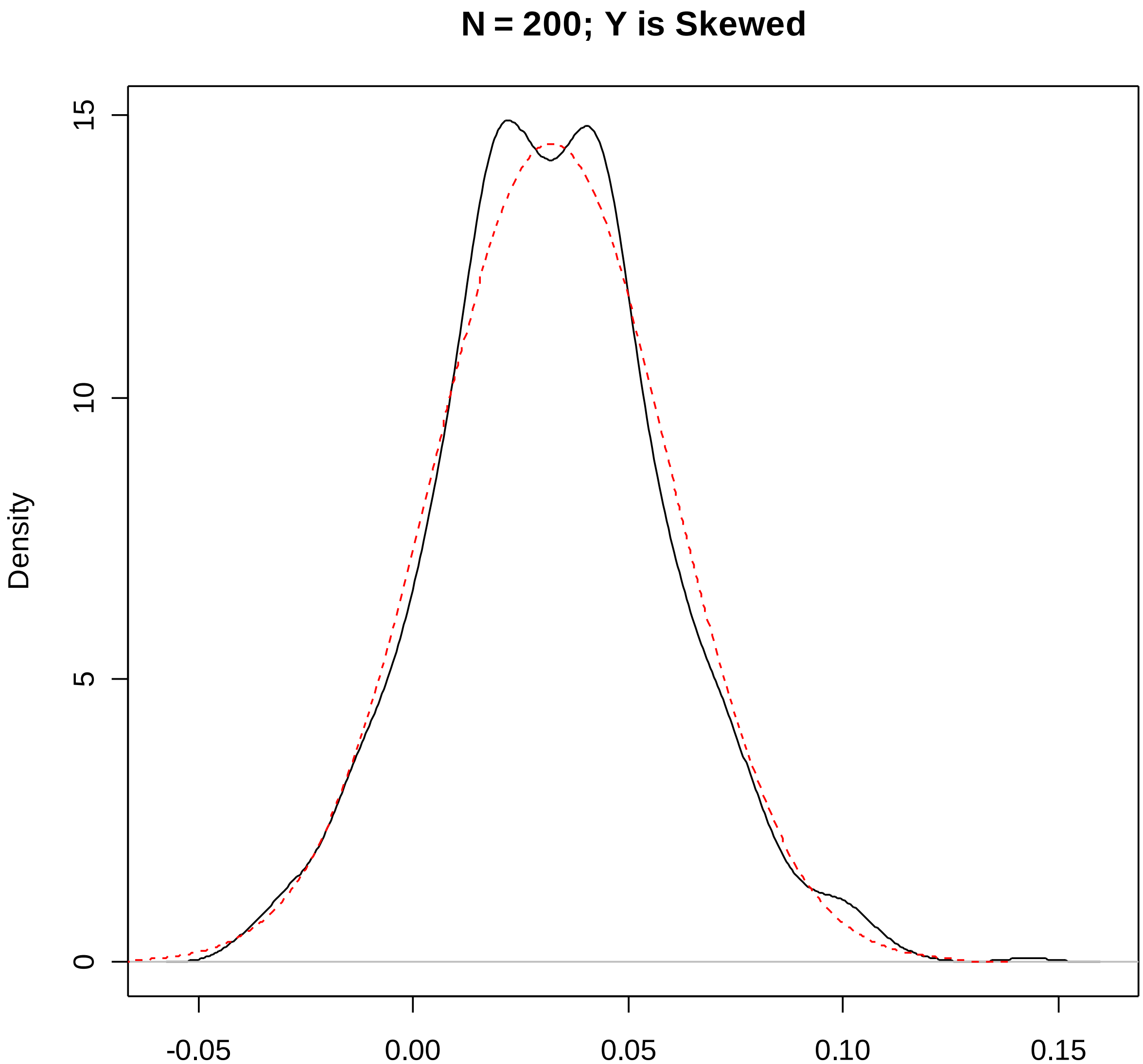

FIGURE 4
Bootstrap sampling distribution for data with a negatively skewed Y variable.

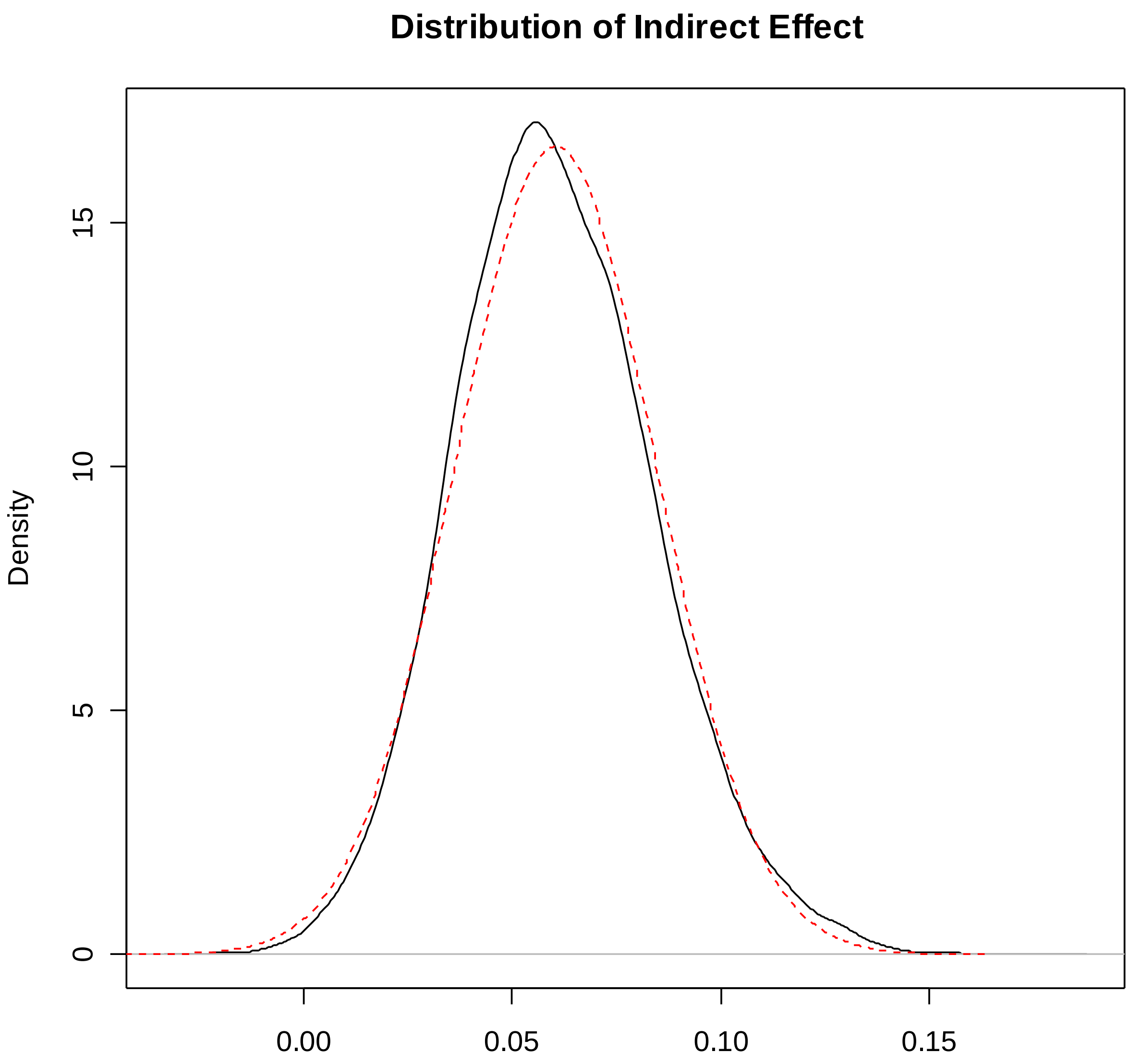

FIGURE 5
Distribution of 10,000 estimates of the indirect effect from a hypothetical correlation matrix and N = 400.