Organizational Research Methods
Volume 11 Number 2
April 2008 296-325

10.1177/1094428107300343
http://orm.sagepub.com
hosted at
http://online.sagepub.com

Testing Mediation and Suppression Effects of Latent Variables

Bootstrapping With Structural Equation Models

Gordon W. Cheung
The Chinese University of Hong Kong
Rebecca S. Lau
Virginia Polytechnic Institute and State University

Because of the importance of mediation studies, researchers have been continuously searching for the best statistical test for mediation effect. The approaches that have been most commonly employed include those that use zero-order and partial correlation, hierarchical regression models, and structural equation modeling (SEM). This study extends MacKinnon and colleagues (MacKinnon, Lockwood, Hoffmann, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004, MacKinnon, Warsi, & Dwyer, 1995) works by conducting a simulation that examines the distribution of mediation and suppression effects of latent variables with SEM, and the properties of confidence intervals developed from eight different methods. Results show that SEM provides unbiased estimates of mediation and suppression effects, and that the bias-corrected bootstrap confidence intervals perform best in testing for mediation and suppression effects. Steps to implement the recommended procedures with Amos are presented.

Keywords: *mediating effects; suppression effects; structural equation modeling*

One of the major objectives of social science research is to make *predictions*. However, prediction merely allows us to know the relationship between independent variables and dependent variables. To gain knowledge, we also need to *explain* the relationships, which is another major objective of social science research. One way to understand how or why the variables are associated in a certain manner is to investigate the mechanisms that underlie the relationships. To understand relationships better, social science researchers have been examining the presence of mediators, also known as intervening variables, in relationships among variables. For instance, trust-in-management has been found to mediate the relationship between perceptions of organizational support and employee commitment (Whitener, 2001), whereas the effect of supervision on organizational citizenship behavior is mediated by procedural justice (Zellars, Tepper, & Duffy, 2002). At group level, team member demographic heterogeneity mediates the effects of perceptions of cooperative norms on team effectiveness (Chatman & Flynn, 2001). At organization level, knowledge

Authors' Note: Please address all correspondence to Gordon W. Cheung, Department of Management, The Chinese University of Hong Kong, 425C K. K. L. Building, Shatin, Hong Kong; phone: (852) 2609–7778.

acquisition by a firm from its foreign parent acts as a mediator in the relationship between the firm's absorptive capacity and its performance (Lane, Salk, & Lyles, 2001).

Because of the importance of mediation studies, researchers have been continuously searching for the best statistical test for mediation effect. The approaches that have been most commonly employed include those that use zero-order and partial correlation, hierarchical regression models and structural equation modeling (SEM). In the 1970s and early 1980s, researchers relied mainly on zero-order and partial correlation coefficients to examine mediation effects (e.g. Cheloha & Farr, 1980). However, the correlation approach is subject to the influence of measurement errors and is restricted to the use of measured variables. Correlation coefficients are also nondirectional and are therefore unable to distinguish between the independent variable and the dependent variable. Furthermore, the correlation approach is difficult to apply in the analysis of complex models with multiple mediators.

Another frequently employed approach to examine mediation is hierarchical regression (e.g. Brown, Ganesan, & Challagalla, 2001; Lester, Meglino, & Korsgaard, 2002; Schaubroeck & Lam, 2002). Many studies using this approach have relied on the Sobel test (1982) to examine the significance of mediation effect. However, there is evidence that the distribution of mediation effect is not normal (Bollen & Stine, 1990; MacKinnon & Dwyer, 1993; Stone & Sobel, 1990), and the utilization of a significance test, such as the Sobel test, which assumes a normal distribution when examining the mediation effect, is not appropriate. MacKinnon and his colleagues (MacKinnon, Lockwood, Hoffmann, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004, MacKinnon, Warsi, & Dwyer, 1995) have conducted several simulation studies to examine the accuracy of various tests on mediation effects estimated with the hierarchical regression approach. Most recently, MacKinnon, Lockwood, and Williams (2004) examined the accuracy of confidence intervals for the indirect effect and demonstrated that the bias-corrected (BC) bootstrap method produces the most accurate confidence intervals. However, a problem with the hierarchical regression approach is that it assumes that the variables are measured without errors whereas variables are usually measured with errors in practice. The existence of measurement errors might result in biased estimation of the mediation effects and confidence intervals.

The recently developed SEM is another important statistical tool to investigate mediation, such as in the mediation effects between participation in decision making and satisfaction (Roberson, Moye, & Locke, 1999), between network structure and career success (Seibert, Kraimer, & Liden, 2001), and between proactive personality and career success (Seibert, Kraimer, & Crant, 2001). SEM has several advantages over the hierarchical regression approach to mediational analyses. First, SEM provides a better statistical tool to investigate latent variables with multiple indicators (Holmbeck, 1997). Second, measurement errors in the model can be controlled for when relationships among variables are examined,[1] thus avoiding complications from measurement errors and the underestimation of mediation effects (Baron & Kenny, 1986; Hoyle & Smith, 1994). Third, the SEM approach allows for the analysis of a more complicated model, for example, when a model with more than one mediator and dependent variable can be considered simultaneously (Hoyle & Smith, 1994). Fourth, SEM depicts a clear model that helps ensure that all relevant paths can be included and tested, without omitting any (Baron & Kenny, 1986).

A simulation is therefore conducted to examine if the findings for the hierarchical regression approach by MacKinnon and collegues (MacKinnon et al., 1995; 2002; 2004) can be generalized to the SEM approach that controls for uniqueness variance in estimation of mediation effect of latent variables. This simulation will first demonstrate the biasing effect of measurement errors on the estimation of mediation effects by comparing the results from the hierarchical regression approach with those from the SEM approach. Then, the accuracy of confidence intervals generated by eight different methods for the indirect effects estimated with SEM is examined. This simulation also extends prior work by including an examination of suppression effects, which have been generally ignored in previous simulations and empirical studies. Practical recommendation for the examination of the mediation and suppression effects of latent variables will be given.

Testing Mediation and Suppression Effects

Mediation effect is frequently referred to as indirect effect, where the effect of the independent variable X_1 on the dependent variable Y goes through a mediator X_2. The mediation effect is commonly defined as the reduction in the regression coefficient of X_1 on Y, when the effects of X_2 are controlled for (Baron & Kenny, 1986; Judd & Kenny, 1981). The relationships among X_1, X_2, and Y are shown in Figure 1. Change in the regression coefficient of X_1 on Y when the effects of X_2 are controlled for is operationalized as $\beta_{Y1} - \beta_{Y1.2}$. Suppose X_1 and X_2 are scaled such that r_{1y} and $r_{2y} > 0$, mediation is concluded when $\beta_{Y1} - \beta_{Y1.2} > 0$. In empirical studies, mediation effect is more frequently operationalized as the product of β_{21} and $\beta_{Y2.1}$, which has been shown to be equal to $\beta_{Y1} - \beta_{Y1.2}$ (MacKinnon et al., 1995). The most commonly employed method for examining the statistical significance of mediation effect is the Sobel test (Sobel, 1982), in which the null hypothesis H_0: $\beta_{21}\beta_{Y2.1} = 0$ is tested. The test statistic S, which is approximately distributed as Z, is obtained by dividing the estimated mediation effect $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$ by the standard error in Equation 1:

$$\hat{\sigma}_{\beta_{21}\beta_{Y2.1}} = \sqrt{\hat{\beta}_{21}^2 \hat{\sigma}_{\beta_{Y2.1}}^2 + \hat{\beta}_{Y2.1}^2 \hat{\sigma}_{\beta_{21}}^2}. \quad (1)$$

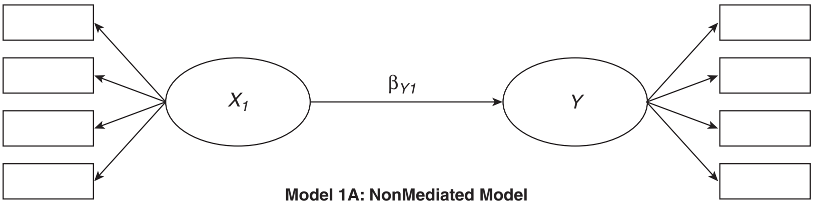
There are two other variations of Equation 1 that provide a standard error for the mediation effect. Baron and Kenny (1986) used a population formula for the standard error for testing the mediation effect, which is based on the first- and second-order Taylor series approximation (Aroian, 1944), where the product of the two variances is added to the variance of mediation effect in Equation 1:

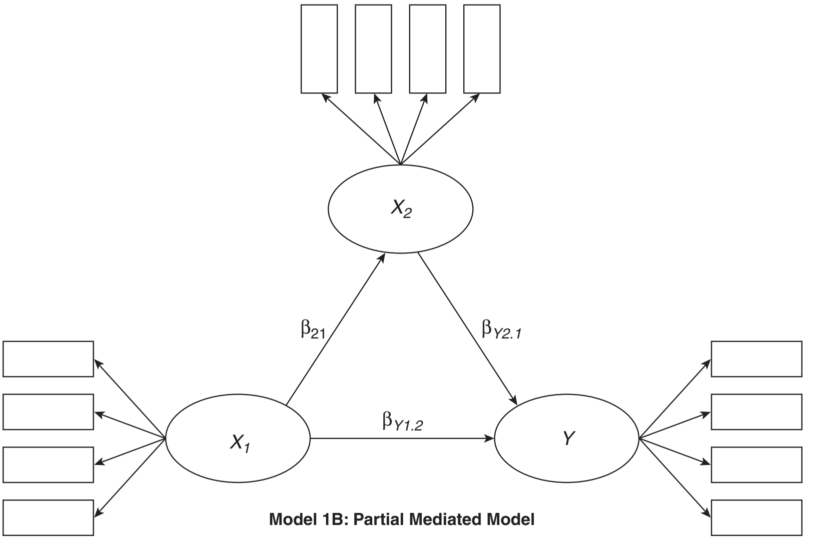
$$\hat{\sigma}_{\beta_{21}\beta_{Y2.1}} = \sqrt{\hat{\beta}_{21}^2 \hat{\sigma}_{\beta_{Y2.1}}^2 + \hat{\beta}_{Y2.1}^2 \hat{\sigma}_{\beta_{21}}^2 + \hat{\sigma}_{\beta_{Y2.1}}^2 \hat{\sigma}_{\beta_{21}}^2}. \quad (2)$$

The second variation is the sample-based estimated standard error of the product of two normal variables derived by Goodman (1960), in which the product of the two variances is subtracted from the variance of the mediation effect in Equation 1:

$$\hat{\sigma}_{\beta_{21}\beta_{Y2.1}} = \sqrt{\hat{\beta}_{21}^2 \hat{\sigma}_{\beta_{Y2.1}}^2 + \hat{\beta}_{Y2.1}^2 \hat{\sigma}_{\beta_{21}}^2 - \hat{\sigma}_{\beta_{Y2.1}}^2 \hat{\sigma}_{\beta_{21}}^2}. \quad (3)$$

Figure 1
Model for Testing the Significance of Mediation Effects With Structural Equation Modeling





Suppression Effects

A suppressor is defined as a third variable that increases the regression coefficient between the independent variable and dependent variable by its inclusion in a regression equation (Conger, 1974).[2] In other words, the relationship between X_1 and Y is hiding or suppressed by the suppressor X_2. When the suppression effect is not controlled for, the relationship between X_1 and Y would appear to be smaller or even of opposite sign (Cohen & Cohen, 1983). Similar to the examination of mediation effect, a suppression effect is operationalized as $\beta_{Y1} - \beta_{Y1.2}$, and suppression is concluded when $\beta_{Y1} - \beta_{Y1.2} < 0$. Alternatively, a suppression effect can be operationalized as the product of β_{21} and $\beta_{Y2.1}$, and suppression is concluded when $\beta_{21}\beta_{Y2.1} < 0$.

Despite the similarity in the procedures for testing mediation and suppression, suppression is rarely examined in organizational and psychological research. However, studying suppression may in fact contribute to theoretical development. Variance of the independent variable can be partitioned into criterion-relevant and criterion-irrelevant components, and inclusion of the suppressor in the analysis helps to partial out the criterion-irrelevant variance (Tzelgov & Henik, 1991). For example, Burton, Lee, and Holtom (2002) found that an employee's age suppressed the relationship between motivation to attend and overall absenteeism. They suggested that the weak relationship between motivation to attend and overall absenteeism found in past research might be because of the failure to include suppressors in the analysis. Another commonly encountered potential suppressor is the halo effect, which usually conceals the relationship between psychological scales and their criterion. Henik and Tzelgov (1985) have demonstrated that the predictive power of psychological scales can be improved by including the halo effect as a suppressor in the analysis.

SEM Approach

Although hierarchical regression models have been commonly used for mediational analysis, they are subject to measurement errors. If the variables are measured with errors, then the significance of the mediation effect is likely to be underestimated because the effect of the independent variable on the dependent variable without the mediator is likely to be underestimated, and the direct effect of the independent variable on the dependent variable is likely to be overestimated (Kenny, Kashy, & Bolger, 1998).

Researchers have turned to latent variables with multiple indicators and SEM to deal with this measurement error problem. The structural equation model for examining mediation effects is shown as Model 1B in Figure 1. Most SEM software packages (such as EQS and LISREL) currently appear to use the Sobel σ in Equation 1 (Sobel, 1982) for examining the significance of indirect effect.[3] However, distribution of the mediation effect is normal only when β_{21} and $\beta_{Y2.1}$ are equal to 0 (MacKinnon et al., 2004). Hence, utilizing the Sobel test and similar approaches that assume normal distribution may not be appropriate for examining the significance of the mediation effect. MacKinnon and colleagues (MacKinnon et al., 2002; 2004) suggest using the bootstrap method to define the confidence intervals for mediation effects estimated with the hierarchical regression approach. Similarly, confidence intervals can be created for parameters estimated with SEM by the bootstrap method. Although confidence intervals can also be used for null hypothesis testing, they are superior to null hypothesis testing because they provide a range of plausible population values for the mediation effect.

The standard errors calculated from Equations 1 to 3 can also be used to establish confidence intervals by assuming a normal distribution and by substituting the respective standard errors into Equation 4:

$$(\hat{\beta}_{21}\hat{\beta}_{Y2.1} - z_{\alpha/2}\hat{\sigma}_{\beta_{21}\beta_{Y2.1}}, \hat{\beta}_{21}\hat{\beta}_{Y2.1} + z_{\alpha/2}\hat{\sigma}_{\beta_{21}\beta_{Y2.1}}). \tag{4}$$

When sample size is small, $z_{\alpha/2}$ in Equation 4 is substituted by $t_{\alpha/2}$ with appropriate degrees of freedom. Bootstrapping confidence intervals are usually more accurate than the confidence intervals based on Equation 4 for significance testing because they do not depend on an assumption of normality.

Bootstrapping involves "resampling" the data many times with replacement to generate an empirical estimation of the entire sampling distribution of a statistic (Mooney & Duval, 1993). It is particularly useful when the statistic does not have known distribution (such as sample median) or when distribution assumptions have been violated. The bootstrap procedure first involves defining a resampling space R, which is usually the observed sample with size n. Then B number (usually 500 or 1,000) of bootstrap samples of n observations is randomly drawn from R with replacement. The desired statistics or parameters are obtained for each bootstrap sample. In this study, the hypothesized structural equation model is fitted to each bootstrap sample. Finally, confidence intervals for the estimated parameters are constructed. Four methods are commonly used to define confidence intervals based on bootstrapping: the percentile method; the bootstrap-t method; the BC method; and the bias-corrected and accelerated (BC_a) method.

Percentile Method

The simplest method for constructing confidence intervals from bootstrapping is the percentile method. The (1 - α) confidence intervals are defined by

$$(\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}). \tag{5}$$

where $\hat{\theta}^*_{\alpha/2}$ is the $\alpha/2^{th}$ percentile of the bootstrap sampling distribution of $\hat{\beta}_{Y1}\hat{\beta}_{Y2.1}$.

Bootstrap-*t* Method

Unlike statistical tests such as the Sobel test that assume normal distribution, the bootstrap-t method generates the distribution of Z directly from the data (Efron & Tibshirani, 1993, p. 160) using the following:

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{\hat{\sigma}^*(b)}, \tag{6}$$

where $Z^*(b)$ is the Z value for the bth bootstrap sample, $\hat{\theta}$ is the original sample estimate of the mediation effect $\hat{\beta}_{Y1}\hat{\beta}_{Y2.1}$, and $\hat{\theta}^*(b)$ and $\hat{\sigma}^*(b)$ are the estimated value and standard error of $\hat{\beta}_{Y1}\hat{\beta}_{Y2.1}$ for the bth bootstrap sample. The $\alpha/2^{th}$ and $(1 - \alpha/2)^{th}$ percentiles of $Z^*(b)$ are used to form the confidence intervals:

$$(\hat{\beta}_{21}\hat{\beta}_{Y2.1} - Z^*_{\alpha/2}\hat{\sigma}_{\beta_{21}\beta_{Y2.1}}, \hat{\beta}_{21}\hat{\beta}_{Y2.1} - Z^*_{1-\alpha/2}\hat{\sigma}_{\beta_{21}\beta_{Y2.1}}). \tag{7}$$

BC Method

The percentile method assumes unbiasedness for the distribution of $\hat{\theta}^*$. That is, $\hat{\theta}^*$ is an unbiased estimator of $\hat{\theta}$ and $\hat{\theta}$ is an unbiased estimator of θ (Mooney & Duval, 1993). The BC method adjusts the bootstrapped distribution of $\hat{\theta}$ with:

$$\hat{z}_0 = \Phi^{-1}\left\{\text{Prob}(\hat{\theta}^*(b) < \hat{\theta})\right\}, \tag{8}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal cumulative distribution function (Efron & Tibshirani, 1993, p. 186). The confidence intervals are defined as follows:

$$(\hat{\theta}^*_{\alpha_1}, \hat{\theta}^*_{\alpha_2}). \tag{9}$$

where

$$\alpha_1 = \Phi\left(2\hat{z}_0 + z_{\alpha/2}\right), \text{ and} \tag{10}$$
$$\alpha_2 = \Phi\left(2\hat{z}_0 + z_{1-\alpha/2}\right). \tag{11}$$

BC$_a$ Method

The BC$_a$ method further adjusts the distribution of $\hat{\theta}^*$ by an acceleration (Efron & Tibshirani, 1993, p. 186):

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}}, \text{ where} \tag{12}$$

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}\left(\hat{z}_0 + z_{\alpha/2}\right)}\right), \text{ and} \tag{13}$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}\left(\hat{z}_0 + z_{1-\alpha/2}\right)}\right). \tag{14}$$

The terms $\hat{\theta}_{(\cdot)}$ and $\hat{\theta}_{(i)}$ are obtained from the jackknife method described below. The BC$_a$ confidence intervals are obtained by substituting α_1 and α_2 from Equations 13 and 14 into Equation 9.

Jackknife Method

The jackknife method is closely associated with the bootstrap method. Instead of resampling n observations with replacement in each bootstrap sample, the jackknife method samples by leaving out one observation at a time:

$$J_{(i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_n). \tag{15}$$

Although the jackknife method can also be used to estimate the standard error and confidence intervals of an estimate, it is more commonly used to estimate the bias of the estimate. In this study, $\hat{\theta}_{(i)}$ is the value of $\hat{\beta}_{Y1}\hat{\beta}_{Y2.1}$ for the i^{th} jackknife sample and $\hat{\theta}_{(\cdot)}$ is the mean of $\hat{\theta}_{(i)}$ across n jackknife samples. The jackknife confidence intervals are given by:

$$(\hat{\theta}_{(\cdot)} - z_{\alpha/2}\sigma_J, \hat{\theta}_{(\cdot)} + z_{\alpha/2}\sigma_J). \tag{16}$$

where σ_J is the standard deviation of $\hat{\theta}_{(i)}$ across n jackknife samples.

Simulation

Method

LISREL 8.54 (Jöreskog & Sörbom, 1996) was used to conduct the simulation and bootstrapping procedures. In this study, three levels of sample size (100, 200, and 500), two levels of item reliability ($\rho = 0.75$ and 0.90), three levels of factor loadings (all 0.6, all 0.8, and combination), and two levels of direct effect from the independent variable to the dependent variable ($\beta_{Y1.2} = 0$ for full mediation, $\beta_{Y1.2} = 0.2$ for partial mediation and suppression) were studied. Following MacKinnon et al. (2004), the zero mediation effect was simulated by four combinations of β_{21} and $\beta_{Y2.1}$ ($\beta_{21} = 0$; $\beta_{Y2.1} = 0$, 0.14, 0.39, and 0.59). The small to large mediation effects were simulated by six combinations of β_{21} and $\beta_{Y2.1}$ (β_{21}, $\beta_{Y2.1}$ = 0.14, 0.14; 0.14, 0.39; 0.14, 0.59; 0.39, 0.39; 0.39, 0.59; and 0.59, 0.59). This study also examined the performance of confidence intervals for the suppression effect. Six combinations of suppression effects were simulated by reversing the sign of $\beta_{Y2.1}$ so that the products of β_{21} and $\beta_{Y2.1}$ became negative. Table 1 shows the mediation and suppression effects represented by various combinations of β_{21}, $\beta_{Y2.1}$ and $\beta_{Y1.2}$. These combinations of specifications resulted in the production of 576 models with different population parameters. Two hundred samples (data sets) were simulated for each model.

Model 1B in Figure 1 was first fitted for each of the 115,200 simulated data sets using LISREL 8.54. This procedure generated the indirect effect ($\hat{\beta}_{21}\hat{\beta}_{Y2.1}$), direct effects ($\hat{\beta}_{21}$ and $\hat{\beta}_{Y2.1}$), and the corresponding standard errors for the calculation of the standard error of indirect effect using Equations 1 to 3. Then jackknife samples were created for each of the 115,200 simulated data sets. Model 1B was fitted to each jackknife sample, and the jackknife indirect effects ($\hat{\beta}_{21}\hat{\beta}_{Y2.1}$) and the corresponding standard errors were computed. Finally, 500 bootstrap samples were generated for each of the 115,200 simulated data sets. The same SEM model was fitted to each bootstrap sample and the estimated parameters were used for defining the various confidence intervals.

Bias Estimation

The parameters $\hat{\beta}_{21}$, $\hat{\beta}_{Y2.1}$, and $\hat{\beta}_{Y1.2}$ of Model 1B estimated with SEM, as well as the product of $\hat{\beta}_{21}$ and $\hat{\beta}_{Y2.1}$ for each of the 115,200 simulated data sets were compared with the true values. Biases were estimated by subtracting the true values from the estimated parameters. In addition, three observed variables X_1, X_2 and Y were created by averaging the values of the corresponding items. The parameters $\hat{\beta}_{21}$, $\hat{\beta}_{Y2.1}$, and $\hat{\beta}_{Y1.2}$ of Model 1B for each of the 115,200 simulated data sets were estimated using the hierarchical regression approach. The estimated parameters, together with $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$, were compared with the true values.

Confidence Limits

Both the 90% and 95% confidence intervals for each of the eight methods and 115,200 simulated data sets were calculated. Estimates of $\hat{\beta}_{21}$ and $\hat{\beta}_{Y2.1}$ and the corresponding standard errors were inserted into Equations 1 to 3 to obtain standard errors for $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$. The

Table 1
Mediation and Suppression Effects Represented by Various Combinations of Population Parameters

	$\beta_{Y1.2} = 0$	$\beta_{Y1.2} = 0.2$
$\beta_{21} = 0, \beta_{Y2.1} \geq 0$ (indirect effect $= 0$)	No mediation and suppression effects	No mediation and suppression effects
$\beta_{21} > 0, \beta_{Y2.1} > 0$ (positive indirect effect)	Full mediation effects	Partial mediation effects
$\beta_{21} > 0, \beta_{Y2.1} < 0$ (negative indirect effect)	Full mediation effects	Partial suppression effects

standard errors were then substituted into Equation 4 to obtain the confidence intervals. Jackknife confidence limits were obtained by substituting the jackknife estimate of $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$ and its standard error into Equation 16. Confidence intervals based on the bootstrap percentile method, bootstrap-*t* method, BC bootstrap method, and BC_a bootstrap method were calculated according to the methods described above.

Accuracy (Robustness)

The accuracy of confidence limits was examined by comparing the proportion of times that the true value of the mediation effect fell outside the confidence limits versus the expected values (Efron & Tibshirani, 1993, p. 321). For example, in assessing the accuracy of the 95% confidence intervals, it is expected that among the 200 simulated samples for each combination of population parameters, the true value of the mediation effect will fall to the left of the lower limit in five samples (2.5% of 200 samples), and will fall to the right of the upper limit in five samples. In addition to testing the statistical significance of the deviations, we use the liberal robustness criterion suggested by Bradley (1978), that is, $0.5\alpha \leq \rho \leq 1.5\alpha$, to examine if the deviation of the confidence limits from the expected values was substantial. For the 95% confidence limits, proportions were considered to be robust if they fell within the range of 1.25% to 3.75%. The confidence intervals were also used to test against the null hypothesis which assumes that the mediation and suppression effect is zero. The effect was considered to be significant if zero was not within the confidence intervals. The observed Type I error rate and statistical power for testing both mediation and suppression effects were examined.

Results

Bias Estimation

Table 2 reports the bias estimation for the parameters for both the hierarchical regression approach and the SEM approach. All the parameters estimated with the hierarchical regression approach deviate significantly from the true values when the true value is nonzero ($p < .0001$). Results of regression analysis show that the magnitude of bias is significantly affected by the size of the measurement errors ($p < .0001$) and the true values of the parameters ($p < .0001$). When the size of the measurement errors is moderate ($\rho = 0.75$), the true mediation effect is on average underestimated by 12.4% and the true suppression effect is underestimated by 14.3%. Maximum deviation occurs when there is a partial suppression effect at 0.3481 ($\beta_{21} = 0.59$, $\beta_{Y2.1} = -0.59$, and $\beta_{Y1.2} = 0.2$), where the true suppression effect was underestimated by 16.2%.[4] When the measurement errors are small ($\rho = 0.9$), the true mediation and suppression effects are, on average, underestimated by 5.5%. On the other hand, none of the parameters estimated from the SEM approach deviates significantly from the true value ($p > .05$). The average deviation of estimated mediation effect from the true value is less than 0.5%, and the deviation for the suppression effect is less than 0.1%. The maximum deviation occurs when the true partial mediation effect is at 0.2301 ($\beta_{21} = 0.14$, $\beta_{Y2.1} = 0.59$ and $\beta_{Y1.2} = 0.2$), where the estimated mediation effect is smaller than the true value by 1.39%.

Distribution of Mediation and Suppression Effects

A major concern of using the Sobel test in examining the significance of mediation and suppression effect is that the distribution of mediation and suppression effect is not normal. The skewness and kurtosis of the mediation and suppression effects for each of the 576 population models are examined. It was found that mediation effects skew to the right when sample size is 100 or 200, whereas suppression effects skew to the left. The effects of β_{21}, $\beta_{Y2.1}$, $\beta_{Y1.2}$, and sample size on the skewness and kurtosis were examined with a regression analysis. It was found that both sample size and the magnitude of $\beta_{Y2.1}$ have significant negative relationships with the skewness ($p < .01$). Moreover, the skewness was largest when β_{21} and $\beta_{Y2.1}$ are at similar magnitude. Similar results were found for kurtosis. Because the magnitude of $\beta_{Y1.2}$ has no significant effect on the distribution of the mediation and suppression effects, Table 3 reports the skewness, kurtosis and Kolmogorov-Smirnov Z(test for normality) for combinations of β_{21}, $\beta_{Y2.1}$, and sample size only.

Confidence Intervals

Problems occurred when calculating the Goodman σ of indirect effects (Equation 3). When the true values of β_{21} and $\beta_{Y2.1}$ are small, the estimated parameters $\hat{\beta}_{21}$ and $\hat{\beta}_{Y2.1}$ are usually smaller than the corresponding standard errors. As a result, the last term of the standard error of the Goodman formula is larger than the sum of the first two terms, and the σ in Equation 3 is undefined (taking the square root of a negative number). Results in subsequent sections for the Goodman σ method are based only on those cases with feasible solutions, which may make the results look better than the other standard error based methods. Hence, results for the Goodman σ method should be interpreted cautiously.

Both 90% and 95% confidence limits for the mediation and suppression effects were calculated for the eight methods. The effects of simulated population parameters including factor loadings, reliability, sample size, and β_{21}, $\beta_{Y2.1}$ and $\beta_{Y1.2}$ on the number of times the true mediation effect was greater than or less than the confidence intervals generated by each method were examined using MANOVA. Only sample size, β_{21} and $\beta_{Y2.1}$ had

Table 2
Biasing Effect of Measurement Error on Parameter Estimates

	Average bias ($\hat{\theta} - \theta$)							
	Hierarchical Regression				Structural Equation Modeling			
	$\rho = 0.75$		$\rho = 0.90$		$\rho = 0.75$		$\rho = 0.90$	
Average True Parameter Values (θ)	$\beta_{Y1.2} = 0$	$\beta_{Y1.2} = 0.2$	$\beta_{Y1.2} = 0$	$\beta_{Y1.2} = 0.2$	$\beta_{Y1.2} = 0$	$\beta_{Y1.2} = 0.2$	$\beta_{Y1.2} = 0$	$\beta_{Y1.2} = 0.2$
Null models								
$\beta_{21} = 0.0000$	0.0014	0.0005	0.0002	0.0009	0.0017	0.0005	0.0003	0.0010
$\beta_{Y2.1} = 0.2800$	−0.0161*	−0.0155*	−0.0078*	−0.0068*	0.0007	0.0013	−0.0010	0.0001
$\beta_{21}\beta_{Y2.1} = 0.0000$	0.0000	−0.0001	0.0005	0.0002	0.0000	−0.0002	0.0006	0.0002
Full or partial mediation models (true indirect effects > 0)								
$\beta_{21} = 0.2983$	−0.0180*	−0.0189*	−0.0080*	−0.0077*	−0.0003	−0.0013	−0.0008	−0.0005
$\beta_{Y2.1} = 0.4483$	−0.0308*	−0.0282*	−0.0135*	−0.0124*	0.0005	−0.0012	−0.0004	−0.0008
$\beta_{21}\beta_{Y2.1} = 0.1479$	−0.0187*	−0.0183*	−0.0085*	−0.0080*	0.0006	−0.0009	−0.0004	−0.0007
Full mediation and partial suppression models (true indirect effects < 0)								
$\beta_{21} = 0.2983$	−0.0176*	−0.0193*	−0.0078*	−0.0077*	0.0001	−0.0018	−0.0005	−0.0004
$\beta_{Y2.1} = -0.4483$	0.0314*	0.0356*	0.0141*	0.0155*	0.0002	0.0008	0.0011	0.0007
$\beta_{21}\beta_{Y2.1} = -0.1479$	0.0189*	0.0212*	0.0084*	0.0090*	−0.0005	0.0005	0.0003	0.0000

Note: Full mediation ($\beta_{Y1.2} = 0$; $\beta_{21}\beta_{Y2.1} \neq 0$); Partial mediation ($\beta_{Y1.2} = 0.2$; $\beta_{21}\beta_{Y2.1} > 0$); partial suppression ($\beta_{Y1.2} = 0.2$; $\beta_{21}\beta_{Y2.1} < 0$).
*$p < .0001$.

Table 3
Tests for Normality Assumption for the Distribution of Mediation and Suppression Effects

		$n = 100$			$n = 200$			$n = 500$		
β_{21}	$\beta_{Y2.1}$	Skewness	Kurtosis	K-S Z	Skewness	Kurtosis	K-S Z	Skewness	Kurtosis	K-S Z
0	0	−0.093	7.900**	5.957**	.428**	5.382**	6.343**	−.293**	6.358**	5.534**
0	.14	−0.063	2.517**	4.073**	.056	2.041**	2.645**	.053	1.260**	1.640
0	.39	−0.167*	1.013**	1.222	−.062	0.270	1.075	.044	0.095	0.427
0	.59	−0.004	0.078	0.489	−.020	−0.046	0.422	−.114	−0.020	0.767
.14	.14	1.103**	2.067**	4.142**	.834**	1.274**	3.365**	.569**	0.349*	2.017*
.14	.39	0.283**	0.279	1.370	.163	0.245	1.334	.002	−0.023	0.458
.14	.59	−0.011	0.186	0.539	−.010	−0.035	0.612	.045	−0.013	0.543
.39	.39	0.396**	0.186	1.996*	.220**	0.090	1.186	.190*	0.225	1.022
.39	.59	0.229**	0.140	0.970	.169*	−0.022	1.068	.079	0.062	0.742
.59	.59	0.273**	0.292	1.253	.165*	−0.012	1.237	.155	0.020	0.696
.14	−.14	−1.019**	2.071**	3.741**	−.965**	1.666**	3.153**	−.616**	0.483**	2.240**
.14	−.39	−0.318**	0.403**	1.525	−.220**	0.209	1.156	−.116	0.120	0.664
.14	−.59	−0.092	0.215	0.747	−.095	0.089	1.053	.048	0.063	0.505
.39	−.39	−0.374**	−0.009	1.819	−.303**	0.141	1.368	−.150	−0.043	0.899
.39	−.59	−0.217**	0.329*	0.932	−.192*	−0.119	1.137	.090	2.311**	0.853
.59	−.59	−0.264**	0.131	1.131	−.250**	0.050	1.179	−.159	0.010	0.963

Note: K-S Z = Kolmogorov-Smirnov Z.
$^{*}p < .001$. $^{**}p < .0001$.

significant effects on the accuracy of confidence limits. In the interest of conserving space, only the results for 95% confidence intervals are presented. Results for the 90% confidence intervals are comparable and are available from the first author.

Table 4 shows the percentage of times that the confidence intervals do not include zero for models with no mediation effect (true value of $\beta_{21}\beta_{Y2.1} = 0$). When both β_{21} and $\beta_{Y2.1}$ are zero, all methods reject the null hypothesis less frequently than the expected rate of 5%. When the true value of $\beta_{Y2.1}$ increases, all confidence intervals reject the null hypothesis more frequently. The percentages obtained for $\beta_{Y1.2} = 0$ (left hand-side of the table) are compared with those obtained for $\beta_{Y1.2} = 0.2$ (right-hand side of the table). None of the comparisons show significant differences at an alpha level of 0.001. Based on the percentages reported in Table 4, the eight methods can be clustered into three groups. The first group is the four methods that assume normal distribution in estimating confidence intervals, including the Sobel σ, Aroian σ, Goodman σ, and jackknife σ methods. These methods produce percentages that are not significantly different from each other ($p > .05$), but are significantly smaller than the other four bootstrap methods ($p < .0001$). The second group is the percentile method. Although the bootstrap percentile method improves accuracy of the confidence limits, the percentages reported are still significantly lower than the expected values, and lower than the percentages reported by the other three bootstrap confidence intervals ($p < .001$). The third group is the bootstrap-t, BC bootstrap and BC_a bootstrap methods. Overall, confidence intervals produced by these three methods are closest to the expected values.

Table 4
Percentage of Cases With True Values to Left and Right of 95% Confidence Intervals—Models With No Mediation Effect ($\beta_{12}\beta_{Y2.1} = 0$)

			$\beta_{Y1.2} = 0$						$\beta_{Y1.2} = 0.2$					
	Effect Size		$n = 100$		$n = 200$		$n = 500$		$n = 100$		$n = 200$		$n = 500$	
Method	β_{21}	$\beta_{Y2.1}$	L	R	L	R	L	R	L	R	L	R	L	R
	0	0												
Sobel σ			0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Aroian σ			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Goodman σ			0.56	1.27	0.13	0.67	1.71	0.57	0.41	1.50	0.94	0.94	0.69	0.41
Jackknife σ			0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Percentile			0.25	0.17	0.17	0.00	0.00	0.33	0.00	0.08	0.17	0.00	0.08	0.08
Bootstrap-*t*			0.58	0.42	0.75	0.08	0.08	0.33	0.25	0.50	0.33	0.17	0.08	0.08
BC			0.58	0.50	1.17	0.17	0.33	0.58	0.08	0.58	0.67	0.08	0.33	0.17
BC_a			0.58	0.42	1.00	0.17	0.33	0.58	0.08	0.50	0.67	0.17	0.25	0.17
	0	.14												
Sobel σ			0.08	0.08	0.17	0.00	0.58	0.25	0.00	0.08	0.08	0.25	0.25	0.75
Aroian σ			0.08	0.00	0.08	0.00	0.42	0.25	0.00	0.00	0.00	0.17	0.25	0.75
Goodman σ			0.21	0.41	0.55	0.18	0.67	0.59	0.71	0.10	0.27	0.54	0.59	0.84
Jackknife σ			0.17	0.00	0.25	0.08	0.50	0.33	0.17	0.00	0.08	0.33	0.42	0.75
Percentile			0.75	1.00	0.67	0.83	1.58	1.33	0.50	0.33	1.17	0.92	1.58	1.58
Bootstrap-*t*			1.33	1.42	1.92	2.08	3.00	2.42	1.00	1.17	2.75	1.75	3.17	2.67
BC			1.50	1.67	1.75	2.17	3.42	3.00	1.67	1.92	2.50	1.92	2.58	3.42
BC_a			1.42	1.75	1.83	2.25	3.25	3.00	1.67	2.00	2.58	1.92	2.50	3.33
	0	.39												
Sobel σ			1.33	1.08	2.33	1.92	2.08	1.83	1.25	0.92	1.58	2.00	3.42	2.17
Aroian σ			1.00	0.75	2.17	1.67	1.92	1.83	0.83	0.92	1.42	1.75	3.25	2.17
Goodman σ			1.42	1.17	2.50	1.92	2.25	1.83	1.59	1.00	1.67	2.17	3.42	2.17
Jackknife σ			1.42	1.50	2.08	2.25	2.67	2.25	1.92	1.08	1.50	2.42	3.42	2.00
Percentile			2.83	2.75	2.75	2.75	2.75	2.67	3.00	2.58	1.83	3.33	4.08	2.50
Bootstrap-*t*			4.83	4.17	4.58	4.33	3.42	3.08	5.00	4.67	3.50	4.75	4.75	3.08
BC			3.83	3.75	4.08	3.25	3.00	2.75	4.67	4.08	2.67	4.17	3.92	2.50
BC_a			4.00	3.58	3.83	3.25	3.00	2.75	4.67	4.42	2.67	4.33	3.92	2.50
	0	.59												
Sobel σ			2.17	2.25	1.67	1.58	1.92	2.67	2.58	1.33	2.50	2.00	2.92	2.33
Aroian σ			2.08	2.08	1.67	1.42	1.83	2.67	2.25	1.33	2.42	2.00	2.83	2.33
Goodman σ			2.25	2.58	1.75	1.58	1.92	2.75	2.58	1.75	2.50	2.00	2.92	2.33
Jackknife σ			2.58	2.92	2.00	1.75	2.00	2.75	2.83	1.58	3.08	2.25	2.92	2.25
Percentile			3.17	3.25	2.42	1.50	2.08	2.58	2.92	2.08	3.00	2.75	3.25	2.75
Bootstrap-*t*			5.67	4.92	2.92	2.33	2.33	3.33	4.42	3.58	4.83	3.25	3.58	2.50
BC			4.00	3.67	2.50	1.83	2.25	2.92	3.33	3.00	3.67	3.17	2.92	2.83
BC_a			4.08	3.75	2.67	2.75	2.25	2.92	3.50	3.17	3.67	3.25	3.08	2.92
	Overall													
Sobel σ			0.90	0.88	1.04	0.88	1.15	1.19	0.96	0.58	1.04	1.06	1.65	1.31
Aroian σ			0.79	0.71	0.98	0.77	1.04	1.19	0.77	0.56	0.96	.98	1.58	1.31
Goodman σ			1.23	1.42	1.37	1.16	1.63	1.54	1.46	1.09	1.41	1.48	2.04	1.55
Jackknife σ			1.04	1.13	1.08	1.02	1.29	1.33	1.23	0.67	1.17	1.25	1.69	1.25
Percentile			1.75	1.79	1.50	1.27	1.60	1.73	1.60	1.27	1.54	1.75	2.25	1.73

(continued)

Table 4 (continued)

	Effect Size		$\beta_{Y1.2} = 0$						$\beta_{Y1.2} = 0.2$					
			$n = 100$		$n = 200$		$n = 500$		$n = 100$		$n = 200$		$n = 500$	
Method	β_{21}	$\beta_{Y2.1}$	L	R	L	R	L	R	L	R	L	R	L	R
Bootstrap-*t*			3.10	2.73	2.54	2.21	2.21	2.29	2.67	2.48	2.85	2.48	2.90	2.08
BC			2.48	2.40	2.38	1.85	2.25	2.31	2.44	2.40	2.38	2.33	2.44	2.23
BC_a			2.52	2.38	2.33	2.10	2.21	2.31	2.48	2.52	2.40	2.42	2.44	2.23

Note: Percentages in italics are significantly different from the expected value (2.5%) at alpha = .001. Percentages underlined are outside Bradley's (1978) liberal robustness criteria (< 1.25% or > 3.75%). Percentages in italics and underlined are both statistically significant and outside the liberal robustness criteria.

Accuracy of confidence intervals are also studied by examining the percentage of times that the confidence intervals do not include the true value of $\beta_{21}\beta_{Y2.1}$ when there are true mediation and suppression effects ($\beta_{21}\beta_{Y2.1} \neq 0$). The percentages obtained for $\beta_{Y1.2} = 0$ (full mediation) are compared with those obtained for $\beta_{Y1.2} = 0.2$ (partial mediation and suppression). Among all 576 comparisons, only 4 show significant differences at an alpha level of 0.001.[5] Hence, Table 5 shows only the combined results for $\beta_{Y1.2} = 0$ and $\beta_{Y1.2} = 0.2$. Although the confidence limits of mediation effect of latent variables based on the Sobel σ, Aroian σ, Goodman σ, and jackknife σ assume normal distributions, Table 5 shows that there is an imbalance in the percentage of times that the true values of mediation effect were greater than or less than the confidence limits. Similar results were found in previous simulated studies on the mediation effect of observed variables (MacKinnon et al., 2004). In the case of mediation (positive indirect effects), the lower confidence limits based on methods with a normality assumption produced much smaller error rates than the expected values, whereas the upper confidence limits produced larger error rates. The imbalance occurs in opposite directions for models with true negative indirect effects (suppression effects). This imbalance is less observable when the true mediation effect is zero (Table 4). In general, the bootstrap methods produce more balanced confidence intervals. The percentages for positive indirect (mediation) effect are compared with the corresponding percentages for negative indirect (suppression) effect. Among all 576 comparisons, all but one comparison have $p > .05$. The only exception is the comparison of the models with sample size = 200, $\beta_{21} = 0.14$ and $\beta_{Y2.1} = 0.14/\text{-}0.14$, where $.01 > p > .001$.

Table 5 also shows whether the percentages are substantially different from the expected value of 2.5%. In general, the confidence limits are more accurate when the sample size increases. Moreover, confidence limits based on the Sobel σ, Aroian σ, and Goodman σ perform worst. Forty-three out of the 72 confidence limits based on the Sobel σ reported in Table 5 have percentages that are significantly different from the expected value. Similarly, 44 confidence limits based on the Aroian σ and 47 confidence limits based on the Goodman σ deviate significantly from the expected values. The jackknife method produces slightly more accurate confidence limits with 32 significant deviations. The four bootstrap methods produce confidence limits that are substantially more accurate than those based on methods with a normality assumption. The bootstrap-*t* method (21 times) is better than the

Table 5
Percentage of Cases With True Values to Left and Right of 95% Confidence Intervals—Models with True Mediation Effect ($\beta_{12}\beta_{Y2.1}\neq 0$)

	Positive Indirect Effect ($\beta_{21}\beta_{Y2.1} > 0$)								Negative Indirect Effect ($\beta_{21}\beta_{Y2.1} < 0$)						
	Effect Size		$n = 100$		$n = 200$		$n = 500$			$n = 100$		$n = 200$		$n = 500$	
Method	β_{21}	$\beta_{Y2.1}$	L	R	L	R	L	R	$b_{Y2.1}$	L	R	L	R	L	R
	.14	.14							−.14						
Sobel σ			*0.25*	*8.54*	1.46	*9.38*	*0.67*	*7.58*		*8.50*	*0.08*	*8.96*	*0.29*	*7.17*	*0.88*
Aroian σ			*0.21*	*0.00*	*0.17*	*5.79*	*0.59*	*6.67*		*0.00*	*0.08*	*5.96*	*0.29*	*6.38*	*0.71*
Goodman σ			*0.37*	*9.82*	*0.38*	*10.52*	*0.67*	*8.25*		*10.81*	*0.08*	*10.30*	*0.38*	*8.17*	*0.88*
Jackknife σ			*0.38*	*8.00*	*0.21*	*9.46*	*0.88*	*7.54*		*8.00*	*0.21*	*9.04*	*0.46*	*6.96*	*0.84*
Percentile			*0.83*	2.42	*0.92*	*4.46*	1.88	*4.67*		2.30	*0.79*	*4.71*	*1.09*	*4.21*	1.50
Bootstrap-*t*			2.05	*17.79*	1.79	*11.50*	2.59	3.55		*18.00*	*1.38*	*10.63*	1.88	*3.67*	2.42
BC			1.54	*7.13*	1.71	*3.83*	2.59	3.42		*8.00*	*1.21*	*4.30*	1.79	2.21	2.25
BC_a			1.50	*6.84*	1.67	*3.92*	2.63	2.63		*8.17*	*1.13*	*4.38*	*1.42*	2.29	2.29
	.14	.39							−.39						
Sobel σ			*0.54*	3.00	*0.71*	*4.00*	*1.00*	3.25		*4.75*	*0.79*	3.38	*0.88*	3.38	1.75
Aroian σ			*0.46*	2.46	*0.63*	3.42	*0.92*	2.96		*4.09*	*0.42*	3.25	*0.71*	3.13	1.59
Goodman σ			*0.54*	*3.75*	*0.71*	*4.42*	*1.00*	3.38		*5.54*	*0.79*	*3.71*	*0.92*	3.46	1.63
Jackknife σ			*0.79*	3.13	*1.13*	*4.25*	1.63	3.42		*4.92*	*1.09*	*3.67*	*1.17*	3.00	1.96
Percentile			1.71	3.00	1.46	*4.08*	2.67	2.67		*4.29*	1.75	3.25	2.13	3.13	2.33
Bootstrap-*t*			3.42	*4.75*	2.46	*5.13*	3.17	3.38		*7.25*	3.04	*4.00*	*3.59*	3.25	2.88
BC			2.34	2.83	2.25	*3.88*	2.42	3.42		*4.25*	2.42	2.88	2.92	2.79	2.84
BC_a			2.38	2.71	2.42	*4.00*	2.42	3.38		*4.34*	2.34	2.88	2.96	2.75	2.92
	.14	.59							−.59						
Sobel σ			1.50	3.04	1.63	3.04	2.09	2.25		*3.67*	2.09	2.25	1.63	2.63	2.00
Aroian σ			*1.25*	2.84	*1.38*	3.04	1.96	2.25		3.50	1.79	2.21	1.54	2.54	1.88
Goodman σ			*1.42*	3.09	1.59	3.17	2.00	2.25		*3.96*	1.96	2.33	1.63	2.54	1.92
Jackknife σ			2.34	2.75	2.50	3.38	2.67	2.21		3.34	2.25	2.00	2.17	2.67	2.21
Percentile			2.67	3.00	2.54	3.42	3.00	2.46		*3.75*	2.84	2.04	2.46	2.75	2.59

(continued)

Table 5 (continued)

	Positive Indirect Effect ($\beta_{21}\beta_{Y2.1} > 0$)									Negative Indirect Effect ($\beta_{21}\beta_{Y2.1} < 0$)						
	Effect Size		$n = 100$		$n = 200$		$n = 500$			$n = 100$		$n = 200$		$n = 500$	
Method	β_{21}	$\beta_{Y2.1}$	L	R	L	R	L	R	$b_{Y2.1}$	L	R	L	R	L	R
Bootstrap-*t*			*4.46*	*4.25*	*4.21*	*3.67*	3.42	2.38		*4.34*	*4.80*	2.34	3.29	2.63	3.17
BC			2.79	3.46	3.08	3.54	3.05	2.29		3.46	*3.75*	2.00	2.75	2.50	2.71
BC_a			2.83	3.63	3.04	3.50	3.09	2.29		3.42	*3.84*	2.04	2.92	2.46	2.75
	.39	.39							−.39						
Sobel σ			*0.30*	*5.33*	*0.75*	*4.30*	*1.38*	3.38		*5.13*	*0.29*	*3.96*	*0.84*	3.38	*1.04*
Aroian σ			*0.30*	*4.80*	*0.75*	*3.92*	*1.30*	3.17		*4.67*	*0.21*	3.67	*0.75*	3.25	*1.00*
Goodman σ			*0.30*	*5.88*	*0.75*	*4.38*	*1.30*	3.25		*5.50*	*0.29*	*4.21*	*0.80*	3.42	*1.04*
Jackknife σ			*0.92*	*5.54*	1.75	*4.71*	2.17	*4.13*		*5.09*	*1.08*	*4.25*	*1.25*	3.71	1.79
Percentile			1.71	*4.00*	2.30	*4.04*	2.71	3.33		*3.67*	1.58	*3.58*	1.71	3.08	2.25
Bootstrap-*t*			2.88	2.63	3.50	2.67	3.38	3.04		2.38	2.75	2.38	2.67	2.54	3.05
BC			2.09	2.59	3.13	2.92	2.92	3.09		2.50	2.50	2.88	2.17	3.09	2.59
BC_a			2.09	2.80	3.04	2.96	2.75	3.09		2.46	2.42	2.92	2.21	3.13	2.63

Table 5 (continued)

	Positive Indirect Effect ($\beta_{21}\beta_{Y2.1} > 0$)								Negative Indirect Effect ($\beta_{21}\beta_{Y2.1} < 0$)						
	Effect Size		$n = 100$		$n = 200$		$n = 500$			$n = 100$		$n = 200$		$n = 500$	
Method	β_{21}	$\beta_{Y2.1}$	L	R	L	R	L	R	$b_{Y2.1}$	L	R	L	R	L	R
	.39	.59							−.59						
Sobel σ			*0.55*	*3.71*	*0.46*	2.21	*1.04*	2.09		*4.88*	*0.79*	2.46	*0.71*	2.50	*1.38*
Aroian σ			*0.42*	3.29	*0.30*	1.88	*0.92*	1.96		*4.63*	*0.71*	2.34	*0.58*	2.34	*1.17*
Goodman σ			*0.42*	*3.75*	*0.30*	2.17	*0.92*	1.96		*4.92*	*0.71*	2.50	*0.63*	2.42	*1.17*
Jackknife σ			*1.13*	*4.08*	1.59	3.08	2.00	2.84		*5.09*	*1.42*	3.04	1.67	3.09	2.38
Percentile			1.67	*3.88*	1.92	2.75	2.42	2.92		*4.29*	2.17	2.63	2.25	2.96	2.54
Bootstrap-*t*			3.04	2.67	3.00	2.21	3.38	2.25		*3.75*	3.05	2.00	3.09	2.21	3.38
BC			2.08	3.29	2.50	2.54	2.42	2.92		*3.71*	2.54	2.46	2.34	2.55	2.96
BC_a			2.08	3.50	2.25	2.46	2.46	2.88		*3.63*	2.50	2.50	2.25	2.50	2.84
	.59	.59							−.59						
Sobel σ			*0.50*	2.67	*0.59*	2.00	*0.67*	*1.34*		2.84	*0.34*	1.96	*0.59*	1.54	1.25
Aroian σ			*0.46*	2.25	*0.50*	1.80	*0.50*	*1.17*		2.75	*0.25*	1.84	*0.50*	*1.34*	1.13
Goodman σ			*0.46*	2.54	*0.50*	1.80	*0.55*	*1.17*		2.88	*0.25*	1.88	*0.50*	1.46	1.13
Jackknife σ			2.00	*3.88*	1.63	2.92	2.63	2.71		3.42	1.46	3.21	1.75	2.88	2.38
Percentile			2.38	3.13	2.00	2.71	3.04	2.67		3.08	2.08	2.92	2.38	2.63	2.75
Bootstrap-*t*			*3.84*	1.67	3.09	2.25	*3.67*	2.21		2.09	3.21	1.88	3.46	2.17	*3.67*
BC			2.88	2.55	2.38	2.38	3.17	2.46		2.34	2.63	2.38	2.83	2.17	3.21
BC_a			2.55	2.79	2.34	2.42	3.05	2.59		2.50	2.50	2.54	2.71	2.34	3.04
	Overall														
Sobel σ			*0.61*	*4.38*	*0.75*	*4.15*	*1.14*	*3.32*	Overall	*4.96*	*0.73*	*3.83*	*0.82*	*3.43*	*1.39*
Aroian σ			*0.52*	2.61	*0.62*	*3.31*	*1.03*	*3.03*		*3.27*	*0.58*	*3.21*	*0.73*	*3.16*	*1.25*
Goodman σ			*0.58*	*4.72*	*0.70*	*4.38*	*1.07*	*3.38*		*5.52*	*0.69*	*4.13*	*0.81*	*3.58*	*1.29*
Jackknife σ			*1.26*	*4.56*	*1.47*	*4.64*	*2.00*	*3.81*		*4.97*	*1.25*	*4.20*	*1.41*	*3.72*	*1.93*
Percentile			*1.83*	*3.24*	*1.86*	*3.58*	2.62	*3.12*		*3.56*	*1.87*	*3.19*	*2.00*	*3.13*	2.33
Bootstrap-*t*			*3.28*	*5.63*	*3.01*	*4.57*	*3.27*	2.80		*6.30*	*3.04*	*3.87*	*3.00*	2.75	*3.10*
BC			2.29	*3.64*	2.51	*3.18*	2.76	*2.94*		*4.05*	2.51	2.82	2.47	2.55	2.76
BC_a			2.24	*3.71*	2.46	*3.21*	2.73	2.81		*4.09*	2.46	2.88	2.41	2.58	2.75

Note: Percentages in italics are significantly different from the expected value (2.5%) at alpha = .001. Percentages underlined are outside Bradley's (1978) liberal robustness criteria (3.75%). Percentages in italics and underlined are both statistically significant and outside the liberal robustness criteria.

methods with normality assumption, but slightly worse than the bootstrap percentile method (17 times). Both the BC bootstrap and the BC_a bootstrap methods produce confidence limits that are the most accurate (9 and 11 times respectively) among the eight methods.

Table 6 shows the observed Type I error rates for each method. When both $\beta_{Y1.2} = 0$ and $\beta_{Y2.1} = 0$, all the methods result in confidence intervals that are significantly lower than the expected values of 5%. When $\beta_{Y2.1}$ becomes larger (0.14), the bootstrap-*t*, BC bootstrap and BC_a bootstrap confidence intervals are closer to the expected levels than other confidence intervals. When $\beta_{Y2.1}$ becomes even larger (0.39 and 0.59), these three methods result in inflated Type I error rates, particularly when $\beta_{Y2.1}$ is large and sample size is small. Under these conditions, the bootstrap percentile method outperforms other methods.

The power for detecting a true mediation or suppression effect for each method is summarized in Table 7. When the effect size is large ($\beta_{21} = .39$, $\beta_{Y2.1} = .59$ or $-.59$; and $\beta_{21} = .59$, $\beta_{Y2.1} = .59$ or $-.59$), all methods have very high power and differences among methods are negligible. When the effect size is medium ($\beta_{21} = .39$, $\beta_{Y2.1} = .39$ or $-.39$), all methods provide acceptable power at all sample sizes. The bootstrap methods have significantly higher power than methods that have the normality assumption only when sample size is at 100 ($p < .0001$). When the effect size is small to medium ($\beta_{21} = .14$, $\beta_{Y2.1} = .39$ or $-.39$; and $\beta_{21} = .39$, $\beta_{Y2.1} = .59$ or $-.59$; and $\beta_{21} = .59$, $\beta_{Y2.1} = .59$ or $-.59$), all methods have adequate power only when the sample size is 500. Although the bootstrap methods provide significantly higher power than the methods assuming normality, the power levels are all inadequate (at 55% or lower) with sample sizes of 100 or 200. When the effect size is very small ($\beta_{21} = .14$, $\beta_{Y2.1} = .14$ or $-.14$) only the bootstrap methods with a sample size of 500 provide adequate power. The power for testing mediation effects was compared with that for testing suppression effects at the same magnitude. All methods have similar power level ($p > .01$) for testing mediation effects and suppression effects when the magnitudes of the true effects are the same.

Discussion

This study shows that measurement errors, which cause the underestimation of both β_{21} and $\beta_{Y2.1,}$ may result in a serious underestimation of mediation and suppression effects when the hierarchical regression approach is employed. When more than one mediator is included in the model, the biasing effect of the measurement error on the estimated mediation effect is less clear. The estimated mediation effect can be either underestimated or overestimated. On the other hand, SEM is very effective in controlling for measurement errors when estimating both the direct and indirect effects.

Many empirical researchers who have examined the mediation effect of latent variables with SEM have tested only for the significance of the direct effect from the independent variable to the mediator ($\hat{\beta}_{21}$) and the direct effect from the mediator to the dependent variable ($\hat{\beta}_{Y2.1}$). However, the significance of these two direct paths does not provide support for a significant mediation effect ($\hat{\beta}_{21}\hat{\beta}_{Y2.1}$) from the independent to the dependent variable through the mediator. Moreover, the mediation effect may be significant even if

Table 6
Type I Error Rates of 95% Confidence Intervals for Each Method

	$\beta_{Y2.1} = 0$			$\beta_{Y2.1} = 0.14$			$\beta_{Y2.1} = 0.39$			$\beta_{Y2.1} = 0.59$			Overall		
	Sample Size			Sample Size			Sample Size			Sample Size			Sample Size		
Method	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500
$\beta_{Y1.2} = 0$															
Sobel σ	*0.08*	*0.00*	*0.00*	*0.17*	*0.17*	*0.83*	*2.42*	4.25	3.92	4.42	3.25	4.58	*1.77*	*1.92*	*2.33*
Aroian σ	*0.00*	*0.00*	*0.00*	*0.08*	*0.08*	*0.67*	*1.75*	3.83	3.75	4.17	3.08	4.50	*1.50*	*1.75*	*2.23*
Goodman σ	*1.83*	*0.80*	*2.29*	*0.62*	*0.73*	*1.26*	*2.58*	4.42	4.08	4.83	3.33	4.67	*2.65*	*2.53*	3.17
Jackknife σ	*0.08*	*0.00*	*0.00*	*0.17*	*0.33*	*0.83*	2.92	4.33	4.92	5.50	3.75	4.75	*2.17*	*2.10*	*2.63*
Percentile	*0.42*	*0.17*	*0.33*	*1.75*	*1.50*	2.92	5.58	5.50	5.42	6.42	3.92	4.67	3.54	*2.77*	3.33
Bootstrap-t	*1.00*	*0.83*	*0.42*	2.75	4.00	5.42	*9.00*	*8.92*	6.50	*10.58*	5.25	5.67	5.83	4.75	4.50
BC	*1.08*	*1.33*	*0.92*	3.17	3.92	6.42	*7.58*	*7.33*	5.75	*7.67*	4.33	5.17	4.88	4.23	4.56
BC_a	*1.00*	*1.17*	*0.92*	3.17	4.08	6.25	*7.58*	7.08	5.75	*7.83*	4.50	5.17	4.90	4.21	4.52
$\beta_{Y1.2} = 0.2$															
Sobel σ	*0.00*	*0.00*	*0.00*	*0.08*	*0.33*	*1.00*	*2.17*	3.58	5.58	3.92	4.50	5.25	*1.54*	*2.10*	2.96
Aroian σ	*0.00*	*0.00*	*0.00*	*0.00*	*0.17*	*1.00*	*1.75*	3.17	5.42	3.58	4.42	5.17	*1.33*	*1.94*	*2.90*
Goodman σ	*1.92*	*1.88*	*1.11*	*0.81*	*0.81*	*1.43*	*2.59*	4.42	4.08	4.33	4.50	5.25	*2.58*	*2.90*	3.61
Jackknife σ	*0.00*	*0.00*	*0.00*	*0.17*	*0.42*	*1.17*	3.00	3.92	5.42	4.42	5.33	5.17	*1.90*	*2.42*	2.94
Percentile	*0.08*	*0.17*	*0.17*	*0.83*	*2.08*	3.17	5.58	5.17	6.58	5.00	5.75	6.00	*2.88*	3.29	3.98
Bootstrap-t	*0.75*	*0.50*	*0.17*	*2.17*	4.50	5.83	*9.67*	*8.25*	*7.83*	*8.00*	*8.08*	6.08	5.15	5.33	4.98
BC	*0.67*	*0.75*	*0.50*	3.58	4.42	6.00	*8.75*	6.83	6.42	6.33	6.83	5.75	4.83	4.71	4.67
BC_a	*0.58*	*0.83*	*0.42*	3.67	4.50	5.83	*9.08*	7.00	6.42	6.67	6.92	6.00	5.00	4.81	4.67

Note: Type I error rates in italics are significantly different from the expected level (5%) at an alpha level of .001. Type I error rates underlined are outside Bradley's (1978) liberal robustness criteria (< 2.5% or > 7.5%). Type I error rates in italics and underlined are both statistically significant and outside the liberal robustness criteria.

Table 7
Power for Testing Mediation and Suppression Effects ($\beta_{21}\beta_{Y2.1} \neq 0$)

	Positive Indirect Effect ($\beta_{21}\beta_{Y2.1} > 0$)					Negative Indirect Effect ($\beta_{21}\beta_{Y2.1} < 0$)				
	Effect Size		Sample Size			Effect Size		Sample Size		
Method	β_{21}	$\beta_{Y2.1}$	100	200	500	β_{21}	$\beta_{Y2.1}$	100	200	500
	.14	.14				.14	−.14			
Sobel σ			1.33	6.75	55.83			1.17	6.92	53.08
Aroian σ			0.83	5.17	51.17			0.83	5.25	50.50
Goodman σ			2.54	8.78	57.42			2.06	9.37	57.33
Jackknife σ			2.08	7.75	55.00			1.58	7.67	52.83
Percentile			11.58	28.33	79.17			9.92	26.50	78.00
Bootstrap-*t*			6.00	18.92	71.58			5.08	20.08	72.33
BC			11.42	29.25	78.33			11.50	30.08	79.42
BC_a			10.83	28.33	78.42			11.17	29.42	79.42
	.14	.39				.14	−.39			
Sobel σ			17.79	40.25	82.96			14.88	40.58	83.38
Aroian σ			14.92	38.79	82.54			12.58	39.25	82.88
Goodman σ			19.00	41.13	83.17			16.47	41.75	83.73
Jackknife σ			19.42	42.04	83.50			16.71	41.08	83.04
Percentile			37.25	55.42	86.96			34.00	54.25	86.96
Bootstrap-*t*			27.92	46.67	84.38			25.21	46.54	84.00
BC			33.54	50.29	85.08			32.38	50.63	85.29
BC_a			33.71	50.79	85.25			32.29	51.17	85.50
	.14	.59				.14	−.59			
Sobel σ			24.50	47.17	85.50			25.33	46.79	84.33
Aroian σ			23.13	46.54	85.38			24.00	46.04	84.04
Goodman σ			24.71	47.29	85.54			25.83	47.21	84.40
Jackknife σ			26.88	48.17	86.00			25.54	45.83	84.08
Percentile			37.21	54.42	87.33			36.58	54.17	85.58
Bootstrap-*t*			28.71	48.75	85.96			28.54	48.50	84.38
BC			30.00	49.71	86.13			31.21	50.83	85.29
BC_a			29.92	50.08	86.17			31.46	51.38	85.21
	.39	.39				.39	−.39			
Sobel σ			85.17	99.63	100.00			83.46	99.58	100.00
Aroian σ			82.21	99.54	100.00			80.83	99.58	100.00

(continued)

Table 7 (continued)

	Positive Indirect Effect ($\beta_{21}\beta_{Y2.1} > 0$)					Negative Indirect Effect ($b_{21}\beta_{Y2.1} < 0$)				
	Effect Size		Sample Size			Effect Size		Sample Size		
Method	β_{21}	$\beta_{Y2.1}$	100	200	500	β_{21}	$\beta_{Y2.1}$	100	200	500
Goodman σ			86.25	99.63	100.00			85.33	99.63	100.00
Jackknife σ			83.79	99.54	100.00			82.17	99.63	100.00
Percentile			95.50	99.79	100.00			94.38	99.88	100.00
Bootstrap-*t*			91.46	99.71	100.00			90.17	99.75	100.00
BC			93.88	99.79	100.00			93.71	99.83	100.00
BC_a			93.71	99.79	100.00			93.38	99.83	100.00
	.39	.59				.39	− .59			
Sobel σ			96.75	100.00	100.00			95.21	99.96	100.00
Aroian σ			96.42	100.00	100.00			96.25	99.96	99.96
Goodman σ			96.83	100.00	100.00			95.63	99.96	100.00
Jackknife σ			96.67	99.71	100.00			95.21	99.92	100.00
Percentile			98.96	100.00	100.00			97.63	99.96	100.00
Bootstrap-*t*			97.67	99.96	100.00			96.00	99.96	100.00
BC			97.83	100.00	100.00			96.79	99.96	100.00
BC_a			97.83	100.00	100.00			96.67	99.96	100.00
	.59	.59				.59	− .59			
Sobel σ			99.92	100.00	100.00			99.79	100.00	100.00
Aroian σ			99.88	100.00	100.00			99.79	100.00	100.00
Goodman σ			99.92	100.00	100.00			99.83	100.00	100.00
Jackknife σ			99.33	99.25	99.00			98.83	99.63	99.21
Percentile			99.96	100.00	100.00			99.88	100.00	100.00
Bootstrap-*t*			99.92	100.00	100.00			99.92	100.00	100.00
BC			99.92	100.00	100.00			99.79	100.00	100.00
BC_a			99.92	100.00	100.00			99.75	100.00	100.00
	Overall					Overall				
Sobel σ			54.30	65.66	87.11			53.28	65.72	86.76
Aroian σ			52.96	65.04	86.37			52.36	65.13	86.10
Goodman σ			55.73	66.35	87.69			54.99	66.52	87.58
Jackknife σ			54.72	66.09	87.04			53.37	65.67	86.62
Percentile			63.37	72.87	92.10			62.07	72.63	91.79
Bootstrap-*t*			58.68	68.94	90.26			57.49	69.12	90.10
BC			61.14	71.53	91.60			60.78	71.99	91.71
BC_a			61.05	71.62	91.64			60.68	72.10	91.73

only one direct path is significant, but the second direct path is close to significance. Hence, the mediation effect should be examined by testing the significance of the indirect effect, $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$. Currently, most SEM software packages provide the Sobel σ for the indirect effects. One concern of using this standard error in testing the significance of indirect effect is that the indirect effect is not normally distributed. Results in this study show that the distribution of mediation and suppression effects is skewed, particularly when the sample size is small to medium. The normality assumption in general holds when the sample size reaches 500, except when all the components of the mediation or suppression effect (β_{21} and $\beta_{Y2.1}$) are small.

The objective of this study is to examine the accuracy of confidence intervals for the mediation and suppression effect of latent variables produced by various methods. Confidence intervals for models with full mediation ($\beta_{Y1.2} = 0$) were first compared with those for partial mediation ($\beta_{Y1.2} = 0.2$). Results show that neither the distribution of the mediation and suppression effect, nor the accuracy of confidence limits produced by all methods is affected by the magnitude of the direct effect from the independent variable to the dependent variable.

The properties of each of the eight methods for generating confidence intervals for $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$ are summarized in Table 8. Because $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$ is skewed even when both direct paths are normally distributed, application of the Sobel test, which assumes normal distribution to evaluate the significance of mediation effect, may be inappropriate. Results reported in Table 5 show that 60% of the confidence limits based on the Sobel σ deviate significantly from the theoretical values of 2.5% and 97.5%. In this study the modified Sobel test (Baron & Kenny, 1986) is not found to be more accurate than the original Sobel test. Accuracy of the confidence intervals based on Aroian σ (Equation 2) is found to be little different from the confidence intervals based on the Sobel σ. Because the Aroian σ is larger than the Sobel σ, both the Type I error rate and the power of the confidence intervals based on the Aroian σ are slightly lower than those of the confidence intervals based on the Sobel σ. Although the Goodman σ produces slightly more accurate confidence intervals than the Sobel σ and the Aroian σ, it is undefined in many cases when the true values of β_{21} and $\beta_{Y2.1}$ are small. This deficiency precludes the applicability of the Goodman σ to the examination of the significance of the mediation and suppression effects.

Unlike previous simulations on observed variables which show that the jackknife method did not perform better than the Sobel test (MacKinnon et al., 2004), results in this study show that performance of the confidence intervals based on the jackknife method is better than those based on various versions of the Sobel standard errors. This difference in findings is probably because of the inclusion of measurement errors in the current simulation. When measurement errors exist, there is a higher chance for the presence of influential cases. The jackknife method effectively removes some of the effects of the influential cases and hence produces smaller standard errors and narrower confidence intervals.

The bootstrapping approach has received increasing attention in recent years, particularly in the area of establishing confidence intervals for estimated parameters that have no known distributions or have violated the normality assumption. By examining the number of models for which the percentage of true values falls outside the confidence intervals, it was shown that the bootstrap percentile method produces substantially more accurate

Table 8
Properties of Confidence Intervals Created by Various Methods

	Sobel σ	Aroian σ	Goodman σ	Jackknife σ	Percentile	Bootstrap-t	BC	BC_a
Accuracy of confidence intervals								
Nil effect × small N	Fair	Fair	Fair	Fair	Fair	Good	Good	Good
Nil effect × medium N	Fair	Fair	Fair	Fair	Fair	Good	Good	Good
Nil effect × large N	Fair	Fair	Fair	Fair	Fair	Good	Good	Good
Small effect × small N	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair
Small effect × medium N	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair
Small effect × large N	Fair	Fair	Fair	Fair	Fair	Fair	Good	Good
Large effect × small N	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair
Large effect × medium N	Fair	Fair	Fair	Fair	Fair	Good	Good	Good
Large effect × large N	Fair	Fair	Fair	Fair	Good	Fair	Good	Good
Type I error								
Nil effect × small N	Fair	Fair	Fair	Fair	Fair	Good	Good	Good
Nil effect × medium N	Fair	Fair	Fair	Fair	Fair	Good	Good	Good
Nil effect × large N	Fair	Fair	Good	Fair	Good	Good	Good	Good
Power								
Small effect × small N	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair
Small effect × medium N	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair
Small effect × large N	Fair	Fair	Fair	Fair	Good	Good	Good	Good
Large effect × small N	Good	Good	Good	Good	Good	Good	Good	Good
Large effect × medium N	Good	Good	Good	Good	Good	Good	Good	Good
Large effect × large N	Good	Good	Good	Good	Good	Good	Good	Good

Note: Small N: $N = 100$; Medium N: $N = 200$; Large N: $N = 500$. Fair = Substantial deviations from expected (nominal) results; Good = Only minor deviations from expected results. BC bootstrap method slightly outperforms the BC_a bootstrap method in almost all aspects.

confidence intervals than the four methods that assume normal distribution. The percentage of cases with an imbalance of true values to left and right of the confidence intervals is also less. The bootstrap percentile confidence intervals are also associated with more accurate Type I error rate and higher power.

Although the confidence intervals based on the bootstrap-t method perform much better than the confidence intervals with a normality assumption, two problems preclude a recommendation to use this method to examine the mediation and suppression effect. The first problem is that the bootstrap-t method produces confidence intervals that are very inaccurate when the true mediation and suppression effect is small, although the problem is not observable with a sample size of 500. The second problem is that, as shown in Equation 6, the calculation of Z^* requires an estimation of the standard error for the mediation effect for each bootstrap sample. Theoretically, this standard error should be estimated by another level of bootstrapping within each bootstrap sample, which would mean the generation of the confidence intervals for 1,000 bootstrap samples would require the estimation of one million structural equation models. Hence, in this study an approximation of Z^* is estimated by substituting the Sobel σ into Equation 6 for each bootstrap sample.

The BC bootstrap and BC_a bootstrap methods produce similar results. Although the BC_a bootstrap confidence intervals have made more adjustments for biases than the BC bootstrap confidence intervals, the BC bootstrap method slightly outperforms the BC_a bootstrap method in almost all aspects. These two methods produce the most accurate confidence limits, the most accurate Type I error, and have the largest power for detecting mediation and suppression effects. Although the advantages of BC bootstrap and BC_a bootstrap methods over the Sobel test are not obvious when both the effect size and sample size are large, the BC bootstrap and BC_a bootstrap methods detect a significant mediation and suppression effect 10% to 20% more frequently than the Sobel test when the effect size is small. A problem with the BC bootstrap and BC_a bootstrap methods is that these methods may reject the null hypotheses slightly more than the Type I error rate when there is no mediation effect, particularly when the sample size is small and $\beta_{Y2.1}$ is large (Table 6).

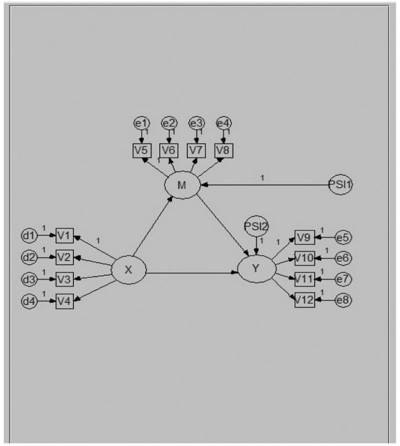
Practical Recommendations

When examining mediation and suppression effects with latent variables, it is recommended that SEM should be used to control for the effects of measurement errors. One should test for mediation and suppression effects by examining the product of the direct path from the independent variable to the mediator and the direct path from the mediator to the dependent variable. Because this product term is not normally distributed, one should employ the BC bootstrap method to establish confidence intervals for the mediation and suppression effects.

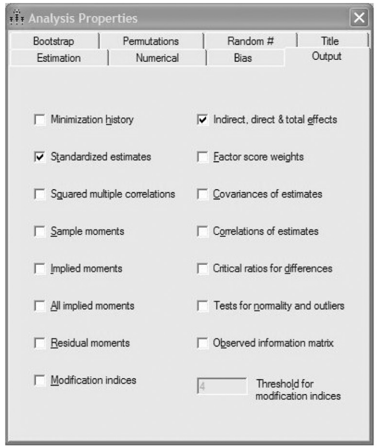
Although the procedure for calculating the BC bootstrap confidence intervals with LISREL is somewhat tedious, it is included as an option in Amos (Arbuckle & Wothke, 1999). Hence, the BC bootstrap confidence intervals of the mediation and suppression effects for 1,000 bootstrap samples can be generated within a few seconds with just a few clicks. The 3-step procedures for examining the mediation effect in Model 1B with Amos 4.0 are shown in Figure 2. First, create an SEM model with Amos as usual. Second, in the Analysis Properties → Output dialogue box, check the "Indirect, direct & total effects" box. For the standardized solutions, check the "Standardized estimates" box. Third, in the Analysis Properties Bootstrap dialogue box, check the "Perform bootstrap" box. Type "1000" in the "Number of bootstrap samples" box. Also check the "Bias-corrected confidence intervals" box and type "95" in the BC confidence level box if 95% confidence intervals are preferred. Keep the "1" in the "Bootfactor" box, which specifies the sample size for the bootstrap samples. Finally, click the "Calculate estimates" icon to run the analysis. Sample outputs are also shown in Figure 2. The results will include a table that contains the lower boundaries of BC bootstrap confidence intervals for indirect effects (Figure 2, Output 1), a table that contains the upper boundaries of BC bootstrap confidence intervals for indirect effects (Figure 2, Output 2), and a table that contains two-tailed significance levels based on the BC bootstrap confidence intervals (Figure 2, Output 3).

The Amos outputs show that the estimated indirect effect from X to Y through M is 0.086. The 95% BC confidence intervals for the indirect effect are between 0.005 and 0.183, with a p-value at 0.039 for two-tailed significance test. Because the Amos does not report standard errors for indirect effect if bootstrapping is not used, the same data set was

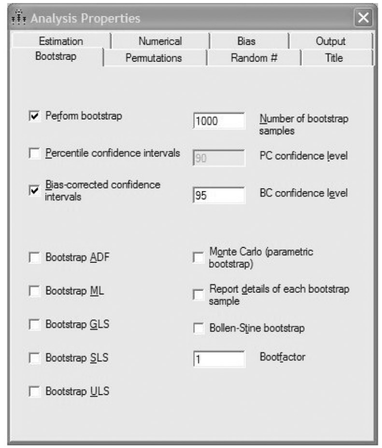
Figure 2
Procedures for Generating Bias-Corrected Confidence Intervals for Indirect Effects With Amos



Step 1: Created an SEM model



Step 2: In the Analysis Properties → Output dialogue box, check the “Indirect, direct & total effects.” If one also wants the standardized solutions, check the “Standardized estimates.”



Step 3: In the Analysis Properties → Bootstrap dialogue box, check “Perform bootstrap” and “Bias-corrected confidence intervals.” Put “1000” for the “Number of bootstrap samples” and “95” for the BC confidence level box if 95% confidence intervals are preferred. Put “1” for “Bootfactor,” which specifies the sample size for the bootstrap samples.

Indirect Effects - Lower Bounds (BC) (Group number 1 - Default model)

	X	M	Y
M	.000	.000	.000
Y	.005	.000	.000

Output 1

Indirect Effects - Upper Bounds (BC) (Group number 1 - Default model)

	X	M	Y
M	.000	.000	.000
Y	.183	.000	.000

Output 2

Indirect Effects - Two Tailed Significance (BC) (Group number 1 - Default model)

	X	M	Y
M	...	...	...
Y	.039	...	...

Output 3

estimated with LISREL 8.71. The estimated indirect effect is 0.0856, with a standard error at 0.0455 and associated p-value at 0.0615. That means if the standard error reported from LISREL is used, one may conclude that the indirect effect is not significant at an alpha level of 0.05, but may conclude that the indirect effect is significant if the BC bootstrap method is employed.

A potential limitation of the BC bootstrap confidence intervals is that two researchers analyzing the same set of data may obtain different confidence intervals because the bootstrap samples generated by each researcher may be different (Gleser, 1996; MacKinnon et al., 2004). However, the differences will be negligible when the number of bootstrap samples is large. Hence, when generating the BC bootstrap confidence intervals, one should generate at least 500 to 1,000 bootstrap samples. Another limitation of the BC bootstrap (as well as BC_a bootstrap) confidence intervals is that the Type I error rate may be higher than the specified level when the sample size is small. Hence, researchers should be very cautious when using this method if the sample size is close to 100. The Type I error rate of BC bootstrap confidence intervals is closer to expected values when the sample size is 200 or more. Researchers may cross-validate their results with the confidence intervals based on the bootstrap percentile method, which have a lower Type I error rate and are also readily available in Amos.

Interpretation of Model Parameters When the Suppression Effect Is Significant

The accuracy of confidence intervals (Table 5) and the power of testing for a true mediation effect (Table 7) were compared with those for a true suppression effect. The results show that all methods perform similarly for both the mediation effect and suppression effect. Although the mediation effect and suppression effect are being examined in similar procedures, the approaches to interpreting results are very different. When examining the mediation effect, one's interest is to decompose the total effect of X_1 on Y into direct effect and indirect (mediation) effect through the mediator X_2. When examining the suppression effect, one's interest is probably to enhance the predictive ability of X_1 on Y by partialing out the criterion-irrelevant variance of X_1, which is achieved by including X_2 in the prediction of Y. When significant suppression effect is identified, one should not conclude that there is a negative direct effect from X_2 on Y. Besides, one should not interpret the total effect of X_1 on Y, but should interpret the direct effect of X_1 on Y in combination with the effect from X_2 on Y. In other words, one should interpret $\beta_{Y1.2}$ as the direct effect of X_1 on Y, after clearing out the criterion-irrelevant variance from X_2 (Maassen & Bakker, 2001; Tzelgov & Henik, 1991).

Limitations

There are several limitations to this study. First, the sample data were generated with normal distributions. Although the BC bootstrap confidence intervals will perform better than the confidence intervals assuming normal distributions, it is not known if the BC bootstrap confidence intervals are accurate when the data are not multivariate normal. Second, to allow for a manageable simulation, only a few levels were examined for each model parameter. Even with this limitation, 576 combinations of population parameters and more than 88 million structural equation models have been estimated. As only models with three factors, each measured with four items are examined in this study, it is not clear if the results of this study can be generalized to models with factors measured by fewer items, or to more complicated models with more factors and items. Finally, this study deals only with the statistical analysis of mediation and suppression effects. Mediational studies require both strong theoretical support and sound research design to rule out alternative explanations for the statistical findings. For example, in a model with only three latent variables, SEM may not be able to distinguish between whether a variable is a mediator or a confounding variable that causes both the independent and dependent variables. In addition to sound theoretical support, a research design that measures the mediator temporarily after the independent variable may weaken the possibility of the confounding effect. Moreover, inclusion of control variables and other mediators in the model may also result in ruling out the confounding effect. Furthermore, the timing of when the mediation effect occurs or is measured may also affect the estimated effect size, as well as the power of the analysis. For the proximal mediator which occurs or is measured close in time to the independent variable, the regression coefficient β_{21} will be larger than the coefficient $\beta_{Y2.1}$. This will result in a smaller estimated mediation effect than the effect of a mediator that has a similar size in β_{21} and $\beta_{Y2.1}$. The same is also true for the distal mediator which is associated with β_{21} which is significantly smaller than the coefficient $\beta_{Y2.1}$. Therefore, researchers should be very cautious in determining when to measure the mediators if they have a choice.

Conclusions

This study contributes to our understanding of the test for mediation effects in several ways. First, it extends MacKinnon and his colleagues' (MacKinnon et al., 1995; 2002; 2004) work by examining mediation effect of latent variables. The results show that measurement errors bias the parameter estimates for mediation effect in the hierarchical regression approach, but can be effectively controlled for in the SEM approach. This study also provides statistical evidence for the non-normal distribution of mediation effect when sample size is small or medium. Therefore, using the standard error for examining the significance of the estimated indirect effect (Sobel test) may not be appropriate. It is proposed that the BC bootstrap confidence intervals, which can be easily obtained from Amos, should be used in examining the significance of the mediation effect. The proposed procedure provides adequate power in detecting a medium to large mediation effect even with a sample size of 100. When the effect size is small, a sample size of 500 provides adequate power for examining the mediation effect.

Because the procedures for testing suppression effects are very similar to those for testing mediation effects, this study also examines the suppression (negative indirect) effect. Results show that the characteristics of suppression effects are very similar to those of mediation effect. However, interpretation of the estimated parameters is less straightforward when suppression effect exists. Because suppressor significantly affects the estimated direct effect of the predictor on the criterion, it is recommended that researchers also consider the possible existence of suppressor while they are considering the existence of control variables.

Finally, the procedures with Amos proposed in this study generate the BC confidence intervals not only for the indirect effects but also for other estimated parameters such as the direct and total effects, factor loadings, uniqueness errors, and elements of the ϕ and ψ matrices. These confidence intervals can be useful also when the distribution of estimated parameters deviates from normality.

Notes

1. The uniqueness variance in structural equation modeling (SEM) (δ and ε) is actually composed of two parts, specific variance and errors of measurement. Specific variance is a consistent and reliable component of the observed score that is not captured by the latent variable (Bollen, 1989, p. 220). Unfortunately, the size of specific variance is rarely known and is therefore usually assumed to be zero (Alwin & Jackson, 1979; Bollen, 1989). Correction for the uniqueness variance is the same as correction for measurement errors when specific variance is zero.

2. Velicer (1978) further classified suppressor into classical suppressor, negative suppressor and reciprocal suppressor by examining the magnitude and direction of r_{y2} and r_{12}.

3. To examine the presence of mediation with SEM, Kelloway (1995) and Holmbeck (1997) have suggested examining the significance of $\hat{\beta}_{21}$ and $\hat{\beta}_{Y2.1}$ in Model 1B by comparing several nested structural equation models. However, both approaches fail to estimate the magnitude of the mediation effect, that is, $\hat{\beta}_{21}\hat{\beta}_{Y2.1}$ or $\hat{\beta}_{Y1} - \hat{\beta}_{Y1.2}$.

4. Percentages are the magnitude of biases divided by the true value of the mediation and suppression effect. For example, when $\beta_{21} = 0.59$, $\beta_{Y2.1} = -0.59$, and $\beta_{Y1.2} = 0.2$, the true suppression effect is 0.3481. The average size of bias is 0.0564, which is 16.2% of 0.3481.

5. Three are the percentages from the Goodman σ method when both β_{21} and $\beta_{Y2.1}$ equal to 0.14. The percentages are larger when $\beta_{Y1.2} = 0.2$. Another significant difference is from the Jackknife σ method, when β_{21} equals to 0.14, $\beta_{Y2.1}$ equals to -0.14, and sample size equals to 100. 9.83% of the cases with true values of mediating effect fall to the left of the 95% confidence intervals when $\beta_{Y1.2} = 0$ and 6.17% when $\beta_{Y1.2} = 0.2$.

References

Alwin, D. F., & Jackson, D. J. (1979). Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological methodology* (pp. 68-119). San Francisco: Jossey-Bass.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SmallWaters.

Aroian, L. A. (1944). The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics, 18*, 265-271.

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*, 1173-1182.

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

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

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Brown, S. P., Ganesan, S., & Challagalla, G. (2001). Self-efficacy as a moderator of information-seeking effectiveness. *Journal of Applied Psychology*, *86*, 1043-1051.

Burton, J. P., Lee, T. W., & Holtom, B. C. (2002). The influence of motivation to attend, ability to attend, and organizational commitment on different types of absence behaviors. *Journal of Managerial Issues*, *14*, 181-197.

Chatman, J. A., & Flynn, F. J. (2001). The influence of demographic heterogeneity on the emergence and consequences of cooperative norms in work teams. *Academy of Management Journal*, *44*, 956-974.

Cheloha, R. S., & Farr, J. L. (1980). Absenteeism, job involvement, and job satisfaction in an organizational setting. *Journal of Applied Psychology, 65*, 467-473.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/Correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*, 35-46.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Gleser, L. J. (1996). Comment on ''Bootstrapping confidence intervals'' by T. J. DiCiccio and B. Efron. *Statistical Science, 11*, 219-221.

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

Henik, A., & Tzelgov, J. (1985). Control for halo error: A multiple regression approach. *Journal of Applied Psychology*, *70*, 577-580.

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, *65*, 599-610.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of consulting and clinical psychology, 62,* 429-440.

Jöreskog K., & Sörbom D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International, Inc.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.

Kelloway, E. K. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior*, *16*, 215-224.

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

Lane, P.J., Salk, J.E., & Lyles, M.A. (2001). Absorptive capacity, learning, and performance in international joint ventures. *Strategic Management Journal*, 22, 1139-1161.

Lester, S.W., Meglino, B.M., & Korsgaard, M.A. (2002). The antecedents and consequences of group potency: A longitudinal investigation of newly formed work groups. *Academy of Management Journal*, 45, 352-368.

Maassen, G. H. & Bakker, A. B. (2001). Suppressor variables in path models: Definitions and interpretations. *Sociological Methods & Research*, *30*, 241-270.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediating effects in prevention studies. *Evaluation Review, 17*, 144-158.

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

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

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41-62.

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

Roberson, Q. M., Moye, N. A., & Locke, E. A. (1999). Identifying a missing link between participation and satisfaction: the mediating role of procedural justice perceptions. *Journal of Applied Psychology, 84*, 585-593.

Schaubroeck, J., & Lam, S. S. K. (2002). How similarity to peers and supervisor influences organizational advancement in different cultures. *Academy of Management Journal*, *45*, 1120-1136.

Seibert, S. E., Kraimer, M. L., & Crant, J. M. (2001). What do proactive people do? A longitudinal model linking proactive personality and career success. *Personnel Psychology, 54*, 845-874.

Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal, 44*, 219-237.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312.

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55*, 337-352.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, 109, 524-536.

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, *38*, 953-958.

Whitener, E. M. (2001). Do "high commitment" human resource practices affect employee commitment? A cross-level analysis using hierarchical linear modeling. *Journal of Management*, 27, 515-535.

Zellars, K. L., Tepper, B. J., & Duffy, M. K. (2002). Abusive supervision and subordinates' organizational citizenship behavior. *Journal of Applied Psychology*, *87*, 1068-1076.

Gordon W. Cheung received his PhD in management in 1993 from Virginia Polytechnic Institute and State University. He is currently a professor in the department of management at the Chinese University of Hong Kong. His main research interest is SEM, specifically on fit measures, measurement equivalence, and second-order models. He was the chair of the research methods division of the Academy of Management in 2006–2007.

Rebecca S. Lau is a PhD student in the department of management in Pamplin College of Business at Virginia Polytechnic Institute and State University. She received her master's degree at the Chinese University of Hong Kong under the supervision of Professor Gordon Cheung. Her research interests include conflict management and organizational justice.