

Computing Confidence Intervals for Standardized Regression Coefficients

Jeff A. Jones and Niels G. Waller
University of Minnesota–Twin Cities

With fixed predictors, the standard method (Cohen, Cohen, West, & Aiken, 2003, p. 86; Harris, 2001, p. 80; Hays, 1994, p. 709) for computing confidence intervals (CIs) for standardized regression coefficients fails to account for the sampling variability of the criterion standard deviation. With random predictors, this method also fails to account for the sampling variability of the predictor standard deviations. Nevertheless, under some conditions the standard method will produce CIs with accurate coverage rates. To delineate these conditions, we used a Monte Carlo simulation to compute empirical CI coverage rates in samples drawn from 36 populations with a wide range of data characteristics. We also computed the empirical CI coverage rates for 4 alternative methods that have been discussed in the literature: noncentrality interval estimation, the delta method, the percentile bootstrap, and the bias-corrected and accelerated bootstrap. Our results showed that for many data-parameter configurations—for example, sample size, predictor correlations, coefficient of determination (R^2), orientation of β with respect to the eigenvectors of the predictor correlation matrix, R_X —the standard method produced coverage rates that were close to their expected values. However, when population R^2 was large and when β approached the last eigenvector of R_X , then the standard method coverage rates were frequently below the nominal rate (sometimes by a considerable amount). In these conditions, the delta method and the 2 bootstrap procedures were consistently accurate. Results using noncentrality interval estimation were inconsistent. In light of these findings, we recommend that researchers use the delta method to evaluate the sampling variability of standardized regression coefficients.

Keywords: standard errors, multiple regression, delta method, confidence intervals

For over 5 decades, researchers have debated the merits of estimating and interpreting standardized regression coefficients (i.e., coefficients estimated from standardized data; Achen, 1977; Bring, 1994; Gelman, 2008; Greenland, Maclure, Schlesselman, Poole, & Morgenstern, 1991; Hunter & Hamilton, 2002; King, 1985; Tukey, 1954). Defending the practice of standardization, some researchers claim that “results based on correlational analysis are invariably easier to understand than results based on raw scores” (Hunter & Hamilton, 2002, p. 553). Proponents of this view contend that standardization adds meaning by removing arbitrary differences in predictor metrics. Other researchers disagree with this contention and note that “if the original data were meaningless, then the standardized regression coefficients are precisely as meaningless; if standardized coefficients do not add information, they certainly do not add meaning” (King, 1985, p. 671). Voicing an extreme version of this notion, Greenland et al. (1991) claimed that standardization has no place in research “because more sensible, un-

confounded alternatives are always available, [and the] use of traditionally standardized coefficients is never justified and should be avoided in any thoughtful analysis, without exception” (p. 391).

Among the many competing views in this controversy, one thing is certain: In many disciplines, researchers continue to report and interpret standardized regression coefficients in published research. For instance, in a comprehensive review of the 2004 volume of the *Journal of Personality and Social Psychology*, West, Aiken, Wu, and Taylor (2007) found that nearly 67% of the articles that used regression reported standardized regression coefficients (West et al., 2007, p. 578). Other researchers have also noted the tendency of social scientists to report standardized regression coefficients (Bring, 1994). For these reasons, we believe that it may be prudent to take a fresh look at the statistical properties of regression coefficients that have been calculated on standardized data.

In this article we show that the so-called “textbook formula” (Yuan & Chan, 2011) for the standard error of a standardized regression coefficient (Cohen, Cohen, West, & Aiken, 2003, p. 86; Harris, 2001, p. 80; Hays, 1994, p. 709) is incorrect and may lead to inaccurate confidence interval (CI) coverage rates. To show this, in the first section, we review some fundamentals of regression theory and the CI formula for unstandardized coefficients. We then demonstrate why this formula—hereafter called the “standard method”—is inappropriate for standardized coefficients. Next, in the Alternative Methods for Computing CIs for β section, we describe four alternative methods for computing CIs—noncentrality interval estimation (NCIE), the

This article was published Online First September 30, 2013.

Jeff A. Jones and Niels G. Waller, Department of Psychology, University of Minnesota–Twin Cities.

Authorship order was determined by a coin flip. This work was supported in part by the University of Minnesota Supercomputing Institute.

Correspondence concerning this article should be addressed to Niels G. Waller, Department of Psychology, University of Minnesota–Twin Cities, Elliott Hall, 75 East River Road, Minneapolis, MN 55455. E-mail: nwaller@umn.edu

delta method (DEL), the percentile bootstrap (PRC), and the bias-corrected and accelerated bootstrap (BCA)—and show how these methods may be used to overcome the statistical limitations of the standard method.¹ In the Simulation Design section, we describe a simulation study that was designed to answer the following two questions: (a) When will the standard method produce CIs with accurate coverage rates for standardized regression coefficients, and (b) when will the alternative methods produce CIs with accurate coverage rates? To set the stage for our simulations, we review the underlying mathematics of the linear regression model.

The Linear Regression Model

Let \mathbf{y} be an $n \times 1$ random vector of criterion scores and let $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ be an $n \times (p + 1)$ matrix of known constants (with the first column being a vector of ones). Then, the regression of \mathbf{y} on \mathbf{X} can be written

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \quad (1)$$

where \mathbf{b} is a $(p + 1) \times 1$ vector of unstandardized regression coefficients, and \mathbf{e} is an $n \times 1$ vector of random errors. The most common method for estimating \mathbf{b} is by ordinary least squares (OLS), such that

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2)$$

If we assume that each error term has zero expectation,

$$E(e_i) = 0, \quad (3)$$

and constant variance,

$$\text{var}(e_i) = \sigma_e^2, \quad (4)$$

then by the Gauss-Markov theorem (Rencher, 2008, p. 146; for a discussion of the Gauss-Markov theorem with random regressors, see Shaffer, 1991), $\hat{\mathbf{b}}$ is the best linear unbiased estimator (BLUE) of \mathbf{b} . In less formal terms, this means that the OLS estimate is unbiased and has the smallest sampling variability of all linear unbiased estimators of \mathbf{b} . Furthermore, if we assume that the errors are normally distributed with constant variance such that $\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$, where \mathbf{I}_n is an $n \times n$ identity matrix, then CIs for the regression coefficients in (1) can be computed using a central t distribution (Rencher, 2008, pp. 116–117). To illustrate these ideas, note that under the stated assumptions,

$$\text{var}(\hat{\mathbf{b}}) = \sigma_e^2 (\mathbf{X}'\mathbf{X})^{-1}, \quad (5)$$

where $\text{var}(\hat{\mathbf{b}})$ is the variance-covariance matrix of the estimated (unstandardized) regression coefficients (Weisberg, 2005, p. 57). From the above, a $(1 - \alpha)$ CI for the j th regression coefficient is defined by

$$\hat{b}_j \pm \text{se}(\hat{b}_j) \cdot t_{(1-\alpha/2), (n-p-1)}, \quad (6)$$

where α is a user-defined Type I error rate, $\text{se}(\hat{b}_j) = \text{var}(\hat{\mathbf{b}})_{jj}^{1/2}$, and $t_{(1-\alpha/2), (n-p-1)}$ is the $1 - \alpha/2$ quantile of a t distribution with $n - p - 1$ degrees of freedom. As noted previously, we refer to this method of computing CIs as the “standard method.”

The standard method is routinely discussed in introductory textbooks on statistics and regression theory (Cohen et al., 2003, pp. 86–87; Howell, 2010, p. 273); under standard assumptions, this method yields CIs with accurate coverage rates. That is, Equation (6) produces accurate CIs for unstandardized regression coefficients when four conditions are satisfied: (a) predictors are fixed, and (b) errors are normally distributed with (c) expected values of zero and (d) constant variance (Fox, 1997, p. 235). As noted earlier, some of these assumptions may be false in the social sciences where it is commonplace to build regression models with standardized random variables (West et al., 2007). In the following paragraphs, we demonstrate why the standard method provides incorrect coverage rates when predictors are both random and standardized. We begin by reviewing the important—but oftentimes unacknowledged—distinction between fixed and random predictors.

Fixed Versus Random Predictors

In statistical theory, a predictor is “fixed” when all values of the predictor are under the control of the researcher. In other words, the “values of $[\mathbf{X}]$ are treated as “fixed” in the analysis of variance sense, that is, as selected by the investigator rather than sampled from some population of $[\mathbf{X}]$ values” (Cohen et al., 2003, pp. 41–42). In the social sciences, this strong assumption is likely to be violated in many research settings. Indeed, according to Bentler and Lee (1983), “virtually all social science predictor variables except experimental design variables are random rather than fixed” (p. 208). Thus, it is desirable to know which statistical properties of standard regression theory are robust to violations of the fixed predictor assumption.

Interestingly, under assumptions that are considerably stronger than those used with fixed predictors, the aforementioned characterization of regression theory is still valid. That is, with random predictors, $\hat{\mathbf{b}}$ is unbiased under slight modifications of Equations (3) and (4) (Sampson, 1974). Under these modifications, we require each error term to have zero expectation,

$$E(e_i | X_1 = x_{1i}, X_2 = x_{2i}, \dots, X_p = x_{pi}) = 0, \quad (7)$$

and constant variance,

$$\text{var}(e_i | X_1 = x_{1i}, X_2 = x_{2i}, \dots, X_p = x_{pi}) = \sigma_e^2, \quad (8)$$

under all predictor combinations (Fox, 1997, pp. 236–237). In the above expressions, X_j ($j = 1, \dots, p$) denotes a random variable, and x_{ji} denotes its realization. Under these assumptions, the OLS estimator in (2) remains unbiased although it need not be the minimum variance estimator (Graybill, 1976, p. 380). That is, other unbiased estimators of \mathbf{b} may have smaller sampling variances.

Whether $\hat{\mathbf{b}}$ is an unbiased, minimum variance estimator (UMVUE, Casella & Berger, 2001, p. 334) depends on the joint distribution of X_1, X_2, \dots , and X_p . Unfortunately, in most research settings this (population) distribution is unknown (Graybill, 1976, p.

¹ Additional methods for computing the standard error of standardized regression coefficients within covariance structure models have been described by Browne and Du Toit (1992) and Bentler and Lee (1983).

380). Nevertheless, using the assumptions in (7) and (8) we can calculate CIs for \mathbf{b} with random predictors using (6). Graybill (1976, p. 383) proved that CI calculations for random predictors are identical to (6) whenever the conditional distribution of Y is normal and (1) is the correct regression model. Under these assumptions, CIs for unstandardized regression coefficients under both fixed and random predictors are computed with a common formula; thus, we treat all predictors as random variables in the sequel.

Standardizing Regression Coefficients

We have seen that under strong assumptions the distinction between fixed and random predictors does not change the CI formula for unstandardized regression coefficients. In this section, we consider the common practice of standardizing the criterion and predictors prior to fitting a model. Next, we focus on how this practice affects CI coverage rates for standardized regression coefficients, $\boldsymbol{\beta}$.

In a population, a standardized regression coefficient, β_j , is an unstandardized regression coefficient that has been scaled as follows (Rencher, 2008, p. 251),

$$\beta_j = b_j \frac{\sigma_{X_j}}{\sigma_Y}, \quad (9)$$

where σ_{X_j} is the standard deviation for the j th predictor, and σ_Y is the standard deviation for the criterion. Estimated coefficients are obtained using sample counterparts,

$$\hat{\beta}_j = \hat{b}_j \frac{s_{X_j}}{s_Y}. \quad (10)$$

As noted earlier, the standard method for constructing CIs for standardized regression coefficients uses incorrect standard errors (see Algina & Moulder, 2001; Kelley & Maxwell, 2003; and Yuan & Chan, 2011). As described in popular textbooks (Cohen et al., 2003, p. 86; Harris, 2001, p. 80; Hays, 1994, p. 709), the (inconsistent and biased) standard error is computed by taking the square root of the j th diagonal element of

$$\sigma_{e_z}^2 (\mathbf{X}'_z \mathbf{X}_z)^{-1}, \quad (11)$$

where subscript z denotes that all observed scores have been standardized (to zero mean and unit variance).² Notice that (11) resembles the covariance matrix for unstandardized coefficients as presented in (5). In essence, the two equations are identical but differ in how the data are scaled. That is, according to the standard method,

$$\text{se}(\hat{\beta}_j) = \frac{\sigma_{X_j}}{\sigma_Y} \text{se}(\hat{b}_j), \quad (12)$$

which, in a given sample, is estimated by

$$\widehat{\text{se}}(\hat{\beta}_j) = \frac{s_{X_j}}{s_Y} \widehat{\text{se}}(\hat{b}_j). \quad (13)$$

Importantly, the above rescaling is correct under the assumption that the estimated, criterion and predictor standard deviations are fixed quantities (otherwise, $\frac{s_{X_j}}{s_Y}$ is a random variable). Of course, this assumption is false because s_Y is always a random variable and

s_{X_j} is often a random variable. Treating this ratio as a fixed quantity “ignores the sampling [variability] in the estimated standard deviations” (Bollen & Stine, 1990, p. 121; see also Algina & Moulder, 2001, p. 648). By ignoring this variability, standard method standard errors are “not consistent even when the explanatory variables are given” (Yuan & Chan, 2011, p. 671).

Using (13) to construct confidence intervals via the standard method (see (6)) is incorrect for a number of reasons, not the least of which is that the sampling distribution of $\hat{\beta}_j$ need not be symmetric. To demonstrate this point we need only consider the sampling distribution of $\hat{\beta}$ in a single predictor regression model. Recall that in simple regression the standardized regression coefficient is equivalent to a Pearson product-moment correlation, r_{XY} . Under this scenario, when X and Y have a bivariate normal distribution then r_{XY} is distributed (Anderson, 2003, p. 125)

$$f(r_{XY}) = \frac{2^{n-2} (1 - \rho_{XY}^2)^{\frac{1}{2}n} (1 - r_{XY}^2)^{\frac{1}{2}(n-3)}}{(n-2)! \pi} \sum_{\alpha=0}^{\infty} \frac{(2\rho_{XY}r_{XY})^{\alpha}}{\alpha!} \times \Gamma^2 \left[\frac{1}{2}(n + \alpha) \right], \quad (14)$$

where ρ_{XY} is the population correlation coefficient, n denotes sample size, α is part of an infinite series that takes on integer values between 0 and infinity, and Γ denotes the gamma function. This equation clearly demonstrates that the sampling distribution of both r_{XY} and $\hat{\beta}_{XY}$ depend on both ρ_{XY} and n . The two plots in Figure 1 illustrate this point.

The left panel in Figure 1 displays the distribution for r_{XY} when $\rho_{XY} = -0.4$ and $n = 40$. The right panel shows the distribution for r_{XY} when $\rho_{XY} = 0.7$ and $n = 50$. Notice that the two distributions are skewed in opposite directions. In both cases, these figures demonstrate why it may be inappropriate to use a (symmetric) central t distribution when constructing a CI for a standardized regression coefficient. In multiple regression, the standardized regression coefficients are weighted functions of a typically small number of predictor-criterion correlations. Thus, the joint distribution of the coefficients also need not be multivariate normal. This is yet another reason why it may be propitious to explore alternatives to the standard method for computing CIs for standardized regression coefficients.

In this article, we consider four alternatives to the standard method (a) noncentrality interval estimation (NCIE; Kelley, 2007; Steiger, 2004; Steiger & Fouladi, 1997), (b) the delta method (DEL; Yuan & Chan, 2011), (c) the percentile bootstrap (PRC; Efron, 1982), and (d) the bias-corrected and accelerated bootstrap (BCA; Efron, 1987). To date, few studies have applied these methods to standardized regression coefficients (although see Yuan & Chan, 2011), and, to our knowledge, no study has evaluated all four methods simultaneously. To remedy this situation, we conducted a comprehensive Monte Carlo study to compare the

² Many textbooks (Cohen et al., 2003, p. 86; Harris, 2001, p. 80; Hays, 1994, p. 709) define the standard error of β_j as $\sqrt{\frac{1-R^2}{(1-R^2_{X_j X_{-j}})(n-p-1)}}$, where $R^2_{X_j X_{-j}}$ is the squared multiple correlation coefficient when X_j is regressed on the remaining predictors (denoted X_{-j}). Note that this equation is equivalent to the square root of the j th diagonal element of (11) when the predictors and criterion are standardized within a sample.

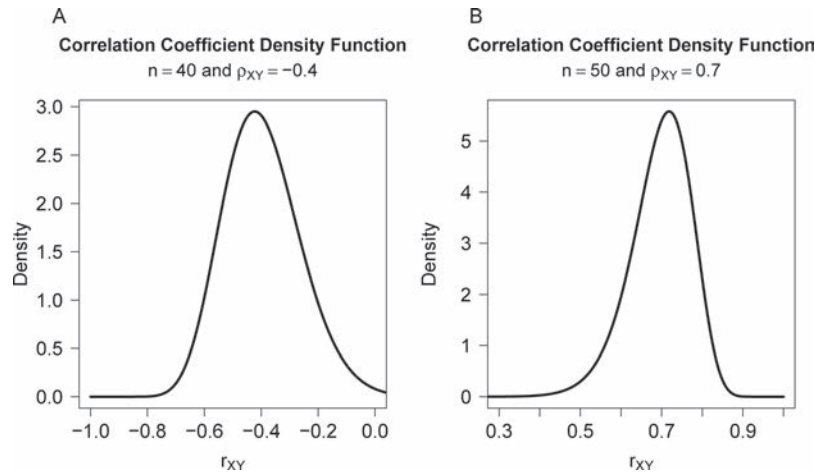


Figure 1. A: This shows the theoretical sampling distribution for r_{XY} when $\rho_{XY} = -0.40$ and $n = 40$. This distribution is right skewed. B: This shows the sampling distribution for r_{XY} when $\rho_{XY} = 0.70$ and $n = 50$. This distribution is left skewed.

empirical performance of the standard method with these alternatives. Before reviewing the results of this study, we briefly review the underlying rationale of the alternative methods. We begin by considering NCIE.

Alternative Methods for Computing CIs for β

Noncentrality Interval Estimation

As a method for computing confidence intervals, NCIE has a long and established history in the field of statistics (see reviews in Cox & Hinkley, 1974; and Kendall & Stuart, 1977). In the social sciences, NCIE was first discussed by Cumming and Finch (2001), Fleishman (1980), Hedges and Olkin (1985), and Steiger (Steiger, 1990; Steiger & Lind, 1980). In a widely cited review of these ideas, Steiger and Fouladi (1997) noted that NCIE relies on two fundamental principles: (a) the confidence interval transformation principle and (b) the inversion confidence interval principle. The underlying logic of these principles can be summarized as follows.

The confidence interval transformation principle characterizes the family of transformations, f , that can be applied to a parameter, θ , such that confidence bounds for $f(\theta)$ are monotonic functions of the confidence bounds for θ . This principle is satisfied whenever the $(1 - \alpha)$ CI about $f(\theta)$ equals

$$f(l_\theta) \leq f(\theta) \leq f(u_\theta),$$

where l_θ and u_θ are the associated $\alpha/2$ and $1 - \alpha/2$ confidence bounds for θ . Fisher's r -to- z transformation is a well-known application of this principle for computing approximate CIs for a correlation coefficient (Cohen et al., 2003, p. 45; Fisher, 1915).

The inversion confidence interval principle (Casella & Berger, 2001; Mood, Graybill, & Boes, 1974; Steiger, 2004) has been described by Steiger and Fouladi (1997) as follows (note that we have modified their notation to be consistent with previously described formulae).

Let $\hat{\theta}$ be the observed value of θ , a random variable having a continuous (cumulative) probability distribution expressible in the form, $F(\hat{\theta}, \lambda) = \Pr(\theta \leq \hat{\theta} | \lambda)$ for some numerical parameter λ . Let $F(\hat{\theta}, \lambda)$ be monotonic, and strictly decreasing in λ , for fixed values

of $\hat{\theta}$. Let λ_l and λ_u be chosen so that $\Pr(\theta \leq \hat{\theta} | \lambda = \lambda_l) = 1 - \alpha/2$ and $\Pr(\theta \leq \hat{\theta} | \lambda = \lambda_u) = \alpha/2$. Then λ_l is a lower $1 - \alpha/2$ confidence limit for λ , λ_u is an upper $1 - \alpha/2$ confidence limit for λ , and the interval with λ_l and λ_u as endpoints is a $1 - \alpha$ confidence interval for λ (Steiger & Fouladi, p. 239)

As described above, the inversion CI principle relies on a noncentrality parameter, λ , and a noncentral cumulative distribution function, $F(\hat{\theta}, \lambda)$ (for discussion of these ideas, see Howell, 2010, p. 232; and Cumming & Finch, 2001). We now review how these notions have been described by Kelley (2007) to construct NCIE CIs for standardized regression coefficients.

Calculation of a CI by NCIE requires knowledge of two non-central cumulative probability distributions. Kelley (2007, p. 19; although see Shieh, 2006) suggests that noncentral t distributions (with separate noncentrality parameters) should be used when computing a CI for a single β . Recall from the sampling theory of an unstandardized regression coefficient, b , we can use a central t distribution to test the null hypothesis $H_0: b = 0$ (Rencher, 2008,

p. 132). When this hypothesis is true, $t = \frac{\hat{b}}{se(\hat{b})}$ has a central t distribution that varies as a function of its degrees of freedom,

$t(df)$. However, when the null hypothesis is false, $t = \frac{\hat{b}}{se(\hat{b})}$ has a

noncentral t distribution, $t(df; \lambda)$ (N. L. Johnson & Welch, 1940, pp. 362–363; Owen, 1968, p. 446; Rencher, 2008, pp. 116–117), with noncentrality parameter, λ , which is dependent on b (the population value of the unstandardized regression coefficient). Specifically,

$$\lambda = \frac{b}{se(\hat{b})} \tag{15}$$

(Rencher, 2008, p. 132). Notice in (15) that the equation for λ resembles that of the t statistic under H_0 with \hat{b} replaced by b . To construct a CI for β_j , Kelley (2007) notes that

$$\lambda_j = \frac{b_j}{\text{se}(\hat{b}_j)} = \frac{\beta_j}{\text{se}(\hat{\beta}_j)}, \quad (16)$$

where β_j is the j th population standardized regression coefficient, and $\text{se}(\hat{\beta}_j)$ is the (previously noted, incorrect) standard error from the standard method. Under this approach,

$$\text{se}(\hat{\beta}_j) = \sqrt{\frac{1 - \rho^2}{(1 - \rho_{X_j X_{-j}}^2)n}}, \quad (17)$$

where $\rho_{X_j X_{-j}}^2$ is the squared multiple correlation coefficient when X_j is regressed on the $j - 1$ remaining predictors. Plugging the above equation into (16) yields

$$\lambda_j = \frac{\beta_j}{\sqrt{\frac{1 - \rho^2}{(1 - \rho_{X_j X_{-j}}^2)n}}}. \quad (18)$$

Kelley (2007, Equation 86) uses the above expression to construct a CI for β_j by way of NCIE. Kelley's method can be summarized as follows.

Let t_0 denote the observed test statistic, $t_0 = \frac{\hat{b}}{\text{se}(\hat{b})}$. We desire endpoints λ_l and λ_u such that

$$\Pr(T < t_0 | \lambda = \lambda_l) = 0.975,$$

and

$$\Pr(T < t_0 | \lambda = \lambda_u) = 0.025.$$

A closed form solution for these limits does not exist but can be numerically obtained (Cumming & Finch, 2001; Fidler & Thompson, 2001; Kelley & Lai, 2011). Once obtained, these limits can be used to derive confidence bounds for β_j . From previous considerations,

$$\lambda_l \leq \beta_j \sqrt{\frac{1 - \rho_{X_j X_{-j}}^2}{1 - \rho^2}} \sqrt{n} \leq \lambda_u,$$

and thus

$$\frac{\lambda_l}{\sqrt{n}} \sqrt{\frac{1 - \rho^2}{1 - \rho_{X_j X_{-j}}^2}} \leq \beta_j \leq \frac{\lambda_u}{\sqrt{n}} \sqrt{\frac{1 - \rho^2}{1 - \rho_{X_j X_{-j}}^2}}.$$

Shieh (2006, p. 532, Equation 8) notes that when the predictors are random, the t statistic discussed earlier has a "two-stage" distribution that involves a noncentral t distribution with a random noncentrality parameter (that has a scaled χ^2 distribution). This suggests that NCIE, as outlined above, may be inappropriate for random regressors.

The Delta Method

The delta method is a well-known technique for approximating variances and covariances of nonlinear functions of one or more random variables (Casella & Berger, 2001, p. 243; Greene, 2003, p. 913; Oehlert, 1992). Essentially, this method "linearizes" a nonlinear function by a truncated Taylor Series (for an elementary

introduction to the Taylor Series, see Stein & Barcellos, 1992, p. 624). For many functions, a first-order Taylor Series produces an accurate linear approximation (for other functions, more accurate estimates of sampling variability can be achieved by using a second-order Taylor Series; see Jamshidian & Bentler, 2000, for details). We call the delta method "DEL" when a first-order Taylor Series is used to estimate the sampling variability of standardized regression coefficients.

We now briefly review the underlying logic of the delta method and show how this flexible procedure can be used to derive an asymptotic covariance matrix for a set of estimated standardized regression coefficients. To breathe some life into this discussion, consider a simple linear function: $f(X) = bX + a$, where a and b are real-valued constants. In this example, as is well known, $\text{var}(f(X)) = \text{var}(bX + a) = b^2 \text{var}(X)$ (Evans & Rosenthal, 2004, p. 144). Calculating the variance of a linear function does not require the delta method (although, in this example, the delta method provides the equivalent expression for $\text{var}(f(X))$). However, with nonlinear functions, it may be difficult to obtain a simple expression for the variance of $f(X)$ without first approximating the function by a polynomial. We now demonstrate this idea.

Let \bar{X} be a random variable with mean μ and variance $n^{-1}\sigma^2$, and let $f(\bar{X})$ be an everywhere continuous and differentiable, nonlinear function of \bar{X} . If

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2),$$

where the symbol \xrightarrow{D} implies that, in the limit (i.e., as n increases without bound), $\sqrt{n}(\bar{X} - \mu)$ assumes a normal distribution with mean 0 and variance σ^2 , then

$$\sqrt{n}[f(\bar{X}) - f(\mu)] \xrightarrow{D} N(0, f'(\mu)^2 \sigma^2), \quad (19)$$

where $f'(\mu)$ denotes the first derivative of $f(\bar{X})$ evaluated at μ (Greene, 2003, p. 913). This result follows from the first-order, Taylor Series approximation of $f(\bar{X})$. From Equation (19), the variance of $f(\bar{X})$ is approximated by $\frac{f'(\mu)^2 \sigma^2}{n}$. In applied settings, $f'(\mu)$ would be evaluated at $f'(\bar{x})$ (where \bar{x} is a realization of \bar{X}).

Following similar logic, we now show how the delta method can be used to approximate the covariance matrix of standardized regression coefficients. Using previously defined notation, let

$$\beta_j = b_j \frac{\sigma_{X_j}}{\sigma_Y} = (\Sigma_X^{-1} \sigma_{XY})_j \frac{\sigma_{X_j}}{\sigma_Y}, \quad (20)$$

where Σ_X is a $p \times p$ predictor covariance matrix and σ_{XY} is a $p \times 1$ vector of predictor-criterion covariances. Replacing parameters with statistics,

$$\hat{\beta}_j = (S_X^{-1} s_{XY})_j \frac{s_{X_j}}{s_Y}. \quad (21)$$

Notice in the above expression that $\hat{\beta}_j$ is a nonlinear function of the observed variances and covariances. Thus, as noted by Yuan and Chan (2011), the delta method is an attractive choice for deriving

the asymptotic covariance matrix for the standardized regression coefficients. Following their derivations, let

$$\Sigma_s = \begin{bmatrix} \Sigma_X & \sigma_{XY} \\ \sigma'_{XY} & \sigma_Y^2 \end{bmatrix} \quad (22)$$

denote a $q \times q$ supermatrix of covariances (where $q = p + 1$). Define $\sigma_v = \text{vech}(\Sigma_s)$ and $s_v = \text{vech}(S_s)$ where the $\text{vech}(\cdot)$ operator stacks the columns of the lower triangular portion of a symmetric matrix (Abadir & Magnus, 2005, p. 299). Using the central limit theorem,

$$\sqrt{n}(s_v - \sigma_v) \xrightarrow{D} N(\mathbf{0}, \mathbf{Q}),$$

where \mathbf{Q} is the asymptotic covariance matrix for \sqrt{ns}_v (Yuan & Chan, 2011). By the delta method,

$$\sqrt{n}[f(s_v) - f(\sigma_v)] \xrightarrow{D} N(\mathbf{0}, \mathbf{J}\mathbf{Q}\mathbf{J}'),$$

where $f(\cdot)$ denotes a vector valued function, and \mathbf{J} denotes the Jacobian (Abadir & Magnus, 2005, p. 351) of $f(\cdot)$. Estimation of \mathbf{J} and \mathbf{Q} is beyond the scope of this review but is described in detail in Yuan and Chan (2011; see Equations 13 & 17). For the present discussion, it is sufficient to note that a standardized regression coefficient is a nonlinear function of σ_v . More formally,

$$\beta = f(\sigma_v) = D_X^{1/2} \Sigma_X^{-1} \sigma_{XY} \cdot \frac{1}{\sigma_Y}, \quad (23)$$

where D_X is a diagonal matrix of the predictor variances. Assuming multivariate normal data, Yuan and Chan (2011, p. 676, Equation 20) show that the approximate, asymptotic variance of the j th standardized regression coefficient equals

$$\text{var}(\hat{\beta}_j) = \frac{\sigma_{X_j}^2 c_j \sigma_e^2}{n \cdot \sigma_Y^2} + \frac{b_j^2 [\sigma_{X_j}^2 (b' \Sigma_X b) - \sigma_{X_j}^2 \sigma_e^2 - \sigma_{X_j Y}^2]}{n \cdot \sigma_Y^4}, \quad (24)$$

where σ_e^2 is the error variance from (1), c_j is the j th diagonal entry of Σ_X^{-1} , and $\sigma_{X_j Y}^2$ is the squared covariance between the j th predictor and the criterion. The sample analogue of (24) is

$$\text{v\hat{a}r}(\hat{\beta}_j) = \frac{s_{X_j}^2 \hat{c}_j \hat{\sigma}_e^2}{n \cdot \hat{\sigma}_Y^2} + \frac{\hat{b}_j^2 [s_{X_j}^2 (\hat{b}' S_X \hat{b}) - s_{X_j}^2 \hat{\sigma}_e^2 - s_{X_j Y}^2]}{n \cdot \hat{\sigma}_Y^4}. \quad (25)$$

Using sample estimates, a $(1 - \alpha)$ CI for the j th standardized regression coefficient can be expressed

$$\hat{\beta}_j \pm t_{1-\alpha/2, df} \cdot \widehat{\text{se}}(\hat{\beta}_j),$$

where $\widehat{\text{se}}(\hat{\beta}_j)$ is the standard error computed from DEL. Note that DEL creates a symmetric CI centered at $\hat{\beta}_j$. As noted previously, this feature of DEL may lead to inaccurate CI coverage rates when $\hat{\beta}_j$ has a skewed sampling distribution. In small samples, Yuan and Chan (2011) recommend using $n - 3$ instead of n in (25) to improve accuracy.

An obvious desideratum of DEL is its simplicity. DEL standard errors are easily obtainable using little more than the observed variances/covariances and the unstandardized regression coeffi-

cients. We report R (R Core Team, 2012) code in Appendix B for computing DEL CIs for standardized regression coefficients.

Bootstrap Methods

In this section, we consider two bootstrap methods (Efron & Tibshirani, 1998) that can be used to construct CIs for standardized regression coefficients. Before describing these methods, we note that the term bootstrap has many connotations in statistics (Efron & Tibshirani, 1998), psychology (Meehl, 1995), and literature (Raspe, 2003). For present purposes, we use the term bootstrap to mean a computer-assisted method for generating an empirical sampling distribution for a possibly vector-valued random variable. Elementary introductions to the bootstrap are available in R. Johnson (2001) and Carpenter and Bithell (2000); more advanced treatments are given by Efron (Efron, 1979, 1982, 1987; Efron & Tibshirani, 1998), and Davison and Hinkley (1997). Freedman (1981) provides a discussion of the bootstrap for regression models with unstandardized coefficients. We consider the use of the percentile (PRC) and the bias-corrected and accelerated bootstrap (BCA; Efron & Tibshirani, 1998, pp. 184–188) for standardized regression coefficients.

We first describe the PRC (Efron & Tibshirani, 1998, pp. 168–176). As with all bootstrap methods, this method constructs an empirical sampling distribution for one or more statistics by computing the statistics on random samples (with replacement) from an observed data set. For many normally distributed statistics (e.g., the mean), the quantiles from a bootstrap distribution yield CIs with excellent coverage rates (Efron & Tibshirani, 1998, p. 171). Although few studies have evaluated PRC performance with standardized regression coefficients (although see Bollen & Stine, 1990), pertinent work has been conducted with Pearson product-moment correlations (Beasley et al., 2007; Efron, 1988; Rasmussen, 1987; Sievers, 1996; Strube, 1988). We summarize this work later.

PRC CIs can be constructed by two methods. These methods differ in how data are resampled (Freedman, 1981). In the first method, both criterion and predictors are considered random variables and are resampled together. In the second method, the predictors are presumed to be fixed and thus only model residuals are resampled. Taking these differences into consideration, the regression model's distributional assumptions should guide method choice. In the simulations described below, the criterion and predictors were considered random variables from a multivariate normal distribution. Hence, we used the first method (which resamples criterion and predictor scores) to construct PRC CIs for standardized regression coefficients. We now outline the general approach.

1. Adjoin the criterion and predictor scores to form a supermatrix of size $n \times (p + 1)$.
2. Sample, with replacement, n rows from the supermatrix to create a provisional data set, M_i .
3. Calculate the standardized regression coefficients in M_i and store for later use.
4. Repeat Steps 2 and 3 B times to construct the joint, empirical sampling distribution of the standardized re-

gression coefficients. Efron (1982) recommends $B = 1,000$ when estimating CIs and $B = 200$ when estimating standard errors.

5. Locate the 2.5% and 97.5% quantiles of the marginal, empirical sampling distributions. These values form the lower and upper bounds for the 95% CIs for the standardized regression coefficients.

Unfortunately, PRC CIs can have poor coverage rates for statistics with skewed sampling distributions (see Efron, 1982, Table 10.3). We have already noted (see Figure 1) that the sampling distributions of Pearson product-moment correlations can be skewed. Thus, it is not surprising that PRC performance for correlations has been found lacking (Beasley et al., 2007; Efron, 1988; Rasmussen, 1987; Sievers, 1996; Strube, 1988). From these results, we might conclude that PRC CIs for standardized regression coefficients will also have poor coverage rates. Recognizing this possibility, we included the BCA bootstrap (DiCiccio & Tibshirani, 1987, p. 164; Efron, 1987, p. 172; Efron & Tibshirani, 1998, pp. 184–188) in our simulations because this method was specifically designed to handle asymmetric sampling distributions.

The BCA improves upon PRC accuracy by including two additional parameters in its formula: (a) a bias-correction parameter and (b) an acceleration parameter. In simplistic terms, bias-correction removes so-called median-bias in the bootstrapped estimator (where median-bias is defined as the difference between the original estimator and the median of the bootstrapped distribution), whereas the acceleration parameter recognizes that the standard error of $\hat{\beta}$ is a function of β (just as the sampling distribution of r depends on ρ ; see (14)). A more detailed treatment of these corrections may be found in Efron (1987, see also Efron & Tibshirani, 1998). Both theoretical and empirical considerations suggest that these corrections should improve CI coverage rates over those obtained from PRC.

Simulation Design

Our Monte Carlo study was designed to explore the relative performance of the standard method and four alternative methods for computing CIs for standardized regression coefficients. Four factors were varied: (a) R^2 , the model coefficient of determination, (b) κ , the condition number of \mathbf{R}_X (the predictor correlation matrix), where $\kappa = \sqrt{\frac{\lambda_1}{\lambda_p}}$ (and λ_i denotes the i th ordered eigenvalue of \mathbf{R}_X), (c) β_v , the orientation of β with respect to the eigenvectors (V) of \mathbf{R}_X , and (d) n , the sample size. All simulations were conducted in the R programming language (R Core Team, 2012).

Population data were generated to be consistent with $R^2 = \{0.2, 0.5, 0.8\}$, $p = 6$, and $\kappa = \{3, 6, 9\}$. Using these parameters, population correlation matrices (\mathbf{R}_X) were constructed with an algorithm by Marsaglia and Olkin (1984; Jones, 2010). Predictor multicollinearity was controlled by varying κ (Cohen et al., 2003, p. 424, suggest that $\kappa > 30$ indicates severe multicollinearity).

A novel aspect of our design is the explicit consideration of β_v , the orientation of β with respect to the eigenvectors of \mathbf{R}_X . To better understand the value of this factor, consider the geometry of standardized regression coefficients. As recently outlined in

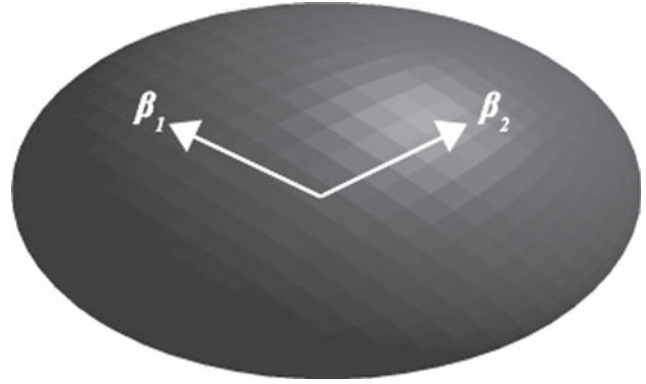


Figure 2. The ellipsoid of all possible regression coefficient vectors that satisfy the constraint $\beta_i' \mathbf{R}_X \beta_i = R^2$. Notice that β_1 and β_2 , like all vectors, terminate on the surface of the ellipsoid.

Waller and Jones (2011), for fixed values of \mathbf{R}_X and R^2 , there are an infinite number of β_i such that

$$\beta_i' \mathbf{R}_X \beta_i = R^2, \tag{26}$$

where β_i denotes the i th vector of standardized regression coefficients ($i = 1, \dots, \infty$) from among the infinite set of vectors that satisfy (26). Note that under the stated constraints, the coefficient vectors that satisfy (26) define a p -dimensional hyper-ellipsoid (Waller & Jones, 2011). Figure 2 illustrates this idea.

The ellipsoid in Figure 2 is composed of the terminal points of an infinite number of coefficient vectors that satisfy (26) for a given $\{\mathbf{R}_X, R^2\}$. Two coefficient vectors, β_1 and β_2 , have been highlighted to illustrate this idea. Note that because each β_i is paired with a distinct \hat{y}_i , an ellipsoid is the geometrical representation of all possible regression models for a set of fixed predictors and a constant R^2 (Waller & Jones, 2011). In the simulations reported below, we focused on four β_i in a plane that was spanned by the first and last eigenvectors of \mathbf{R}_X . Specifically, β_1 was collinear with the first eigenvector, β_2 was 30° from the first eigenvector (offset in the direction of the last eigenvector), β_3 was 60° from the first eigenvector, and β_4 was collinear with the last eigenvector. These points were chosen to relate our findings to those from previous research (Waller, 2011; Waller & Jones, 2010).

To make the choice of β_i more concrete, consider the eigen decomposition of a two-predictor correlation matrix with $r_{12} = .5$,

$$\mathbf{R}_X = \mathbf{V} \mathbf{\Lambda} \mathbf{V}',$$

where V is a $p \times p$ orthogonal matrix of eigenvectors

$$\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2] = \begin{bmatrix} .707 & -.707 \\ .707 & .707 \end{bmatrix}$$

and $\mathbf{\Lambda}$ is a 2×2 diagonal matrix of eigenvalues of \mathbf{R}_X

$$\mathbf{\Lambda} = \begin{bmatrix} 1.5 & 0 \\ 0 & .5 \end{bmatrix}.$$

We desire four vectors of standardized regression coefficients with particular orientations with \mathbf{v}_1 and \mathbf{v}_2 . Specifically, for the aforementioned angles,

$$\begin{aligned}
 \beta_1 &= [.258 \ .258] \\
 \beta_2 &= [.104 \ .386] \\
 \beta_3 &= [-.134 \ .499] \\
 \beta_4 &= [-.477 \ .477].
 \end{aligned}
 \tag{27}$$

Note that each vector of regression coefficients yields $R^2 = .2$ when substituted into (26). Also note that each vector points in a different direction in the plane. To better understand this point, recall that

$$\cos(\theta) = \frac{\mathbf{a}'\mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|},$$

for vectors \mathbf{a} and \mathbf{b} , where the $\|\cdot\|$ operator denotes the vector norm (i.e., the vector length) and θ denotes the $\angle(\mathbf{a}, \mathbf{b})$ (the inscribed angle between vectors \mathbf{a} and \mathbf{b}). Expressed in degrees, the desired angle (θ) can be computed as follows,

$$\theta = \text{acos}\left(\cos\left(\frac{\mathbf{a}'\mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|}\right)\right) \cdot \frac{180}{\pi},
 \tag{28}$$

where $\text{acos}(\cdot)$ denotes the inverse cosine function (for a review of these geometrical ideas, see Fox, 2009). Applying this equation to the above data yields $\angle(\beta_1, v_1) = 0^\circ$ (β_1 is collinear with v_1), $\angle(\beta_2, v_1) = 30^\circ$, $\angle(\beta_3, v_1) = 60^\circ$, and $\angle(\beta_4, v_1) = 90^\circ$ (β_4 is collinear with v_2).

Returning to our simulation design, combinations of the aforementioned factors produced 36 multivariate normal populations. From each population, we drew 5,000 samples of $n = \{50, 100, 500\}$. In each sample we computed CIs for the standardized regression coefficients using the standard method, NCIE, DEL, PRC, and BCA.³ The bootstrap CIs were computed using $B = 10,000$ bootstrap replications.⁴ For all methods, we computed the empirical CI coverage rates for 5,000 samples. See Appendix A for a step-by-step guide (and R code) that describes how we generated data for the simulations.

Results

Our results are summarized in two parts. In the first part, we summarize our findings for the standard method. We next summarize our results for the alternative methods.

We begin by considering regression models with $R^2 = \{.20, .50, .80\}$, $\kappa = 3$, and β_i oriented at $0^\circ, 30^\circ, 60^\circ$, and 90° from the first eigenvector of R_X . For these models, the regression coefficients reside in six dimensional space (i.e., \mathbb{R}^6 ; we are assuming throughout that R_X has full column rank). Recall that we constrained β_i to lie in the (v_1, v_6) plane (i.e., all β_i were in the plane defined by the first, v_1 , and last, v_6 , eigenvectors of R_X). All regression coefficients for these models are reported in Table 1.

Using Marsaglia and Olkin's (1984) algorithm for generating population correlation matrices with user-defined eigenvalues, our predictor correlation matrix had an $|\bar{r}| \approx .16$ with a $\text{sd}(|r|) \approx .10$ and eigenvalues of $\{1.48, 1.18, 1.14, 1.11, 0.92, 0.16\}$. The complete matrix is reproduced below.

Table 1
Standardized Regression Coefficients for Models With $R^2 = \{.20, .50, .80\}$ and $\kappa = 3$

Location of β_i	$\beta_{i[1]}$ ^a	$\beta_{i[2]}$	$\beta_{i[3]}$	$\beta_{i[4]}$	$\beta_{i[5]}$	$\beta_{i[6]}$
$R^2 = .20$						
$\beta_i \ v_1^b$	-0.086	0.066	-0.104	-0.078	0.291	-0.148
$\angle(\beta_i, v_1) = 30^\circ$	-0.004	0.014	-0.041	0.007	0.411	-0.054
$\angle(\beta_i, v_1) = 60^\circ$	0.137	-0.077	0.071	0.153	0.583	0.113
$\beta_i \ v_6$	0.423	-0.268	0.322	0.440	0.662	0.483
$R^2 = .50$						
$\beta_i \ v_1^b$	-0.135	0.104	-0.164	-0.123	0.461	-0.234
$\angle(\beta_i, v_1) = 30^\circ$	-0.007	0.022	-0.065	0.011	0.650	-0.086
$\angle(\beta_i, v_1) = 60^\circ$	0.217	-0.121	0.112	0.242	0.922	0.179
$\beta_i \ v_6$	0.669	-0.423	0.508	0.696	1.047	0.763
$R^2 = .80$						
$\beta_i \ v_1^b$	-0.171	0.132	-0.208	-0.155	0.583	-0.296
$\angle(\beta_i, v_1) = 30^\circ$	-0.008	0.028	-0.082	0.014	0.823	-0.109
$\angle(\beta_i, v_1) = 60^\circ$	0.275	-0.153	0.142	0.306	1.167	0.226
$\beta_i \ v_6$	0.846	-0.535	0.643	0.880	1.324	0.965

^a $\beta_{i[j]}$ denotes the j th element of β_i . ^b The notation $\mathbf{a} \| \mathbf{b}$ denotes that vectors \mathbf{a} and \mathbf{b} are parallel.

R_X

$$= \begin{bmatrix} 1.000 & 0.127 & -0.066 & -0.134 & -0.283 & -0.107 \\ 0.127 & 1.000 & -0.061 & 0.076 & 0.181 & 0.074 \\ -0.066 & -0.061 & 1.000 & -0.092 & -0.252 & -0.102 \\ -0.134 & 0.076 & -0.092 & 1.000 & -0.290 & -0.144 \\ -0.283 & 0.181 & -0.252 & -0.290 & 1.000 & -0.366 \\ -0.107 & 0.074 & -0.102 & -0.144 & -0.366 & 1.000 \end{bmatrix}.
 \tag{29}$$

Recall that standard-method standard errors are inconsistent (Yuan & Chan, 2011). In essence, this implies that if the method performs poorly in large sample sizes, it is likely to perform poorly in smaller samples sizes, all else being equal. Recognizing this point, in Table 2 we summarize the results for the $n = 500$ condition, only. The entries in Table 2 represent discrepancies between the empirical and expected (95%) CI coverage rates for the 12 populations in which $\kappa = 3$. Each entry summarizes the results from 5,000 samples and can be interpreted as follows. Negative numbers denote CIs that (on average) are too narrow, whereas positive numbers denote CIs that (on average) are too wide. Because percentages must lie within $[0, 100]$, values greater than 5% are not possible.

³ A reviewer suggested that we also compute standard errors for the standardized regression coefficients using the RAMONA (Browne, Mels, & Cowan, 1994) program for mean and covariance structures. In the early stages of writing this article, we wrote an R function that implements the RAMONA approach using equations that are described in Browne and Du Toit (1992). In all cases studied, the RAMONA standard errors agreed with the Yuan and Chan (2011) estimates to six or more decimal places. Moreover, at a recent meeting of the Society for Multivariate Experimental Psychology (2012), Michael Browne acknowledged that the two methods should produce equivalent results. Thus, our DEL results elucidate the performance of RAMONA under the studied conditions.

⁴ The bootstrap replications were run in parallel using 200 processors on a university super computer. Parallelization was accomplished with the `snow` (Tierney, Rossini, Li, & Sevcikova, 2012) and `Rmpi` (Yu, 2012) packages in R (R Core Team, 2012).

Table 2

Discrepancies in the Empirical and Expected CI Coverage Rates for the Standard Method for $R^2 = \{.20, .50, .80\}$, $\kappa = 3$, and $n = 500$

Location of β_i	$\beta_{i[1]}$ ^a	$\beta_{i[2]}$	$\beta_{i[3]}$	$\beta_{i[4]}$	$\beta_{i[5]}$	$\beta_{i[6]}$
$R^2 = .20$						
$\beta_i \parallel \mathbf{v}_1$ ^b	-0.02 ^c	0.14	-0.06	-0.40	0.72	0.26
$\angle(\beta_i, \mathbf{v}_1) = 30^\circ$	0.28	-0.18	-0.34	-0.22	0.52	0.20
$\angle(\beta_i, \mathbf{v}_1) = 60^\circ$	-0.18	-0.14	0.22	0.40	1.42	0.26
$\beta_i \parallel \mathbf{v}_6$	1.14	0.12	0.28	0.70	1.38	1.26
$R^2 = .50$						
$\beta_i \parallel \mathbf{v}_1$	-0.58	0.02	0.52	-0.14	0.84	0.06
$\angle(\beta_i, \mathbf{v}_1) = 30^\circ$	0.10	-0.70	-0.42	-0.18	1.70	-0.38
$\angle(\beta_i, \mathbf{v}_1) = 60^\circ$	-0.28	0.12	0.36	-0.78	2.92	-0.20
$\beta_i \parallel \mathbf{v}_6$	-0.08	-0.12	-0.28	0.32	0.00	0.02
$R^2 = .80$						
$\beta_i \parallel \mathbf{v}_1$	-0.26	-0.66	-0.88	-0.04	1.22	-1.00
$\angle(\beta_i, \mathbf{v}_1) = 30^\circ$	0.34	-0.22	0.06	-0.22	2.60	0.10
$\angle(\beta_i, \mathbf{v}_1) = 60^\circ$	-2.36	-0.56	-1.06	-1.38	3.38	-1.08
$\beta_i \parallel \mathbf{v}_6$	-12.40	-7.32	-9.88	-14.60	-19.64	-15.86

Note. CI = confidence interval.

^a $\beta_{i[j]}$ denotes the j th element of β_i . ^b The notation $a \parallel b$ denotes that vectors a and b are parallel. ^c Tabled values denote the difference between the observed and expected coverage rates (expected = 95%); negative numbers denote CIs that are too narrow, positive numbers denote CIs that are too wide.

The findings in Table 2 suggest that the (theoretically incorrect) standard method provides relatively accurate coverage rates when β_i is collinear with the first eigenvector of \mathbf{R}_X . Stated differently, the standard method works surprisingly well when the expected criterion is highly correlated with the first principal component of the predictors (because the eigenvectors of \mathbf{R}_X equal the principal component scoring weights for standardized predictors). Our findings also suggest that the accuracy of standard method CIs is affected by the interaction of $\beta_{\mathbf{v}}$ (orientation of the standardized regression coefficients) and model strength (as measured by R^2). We tested this conjecture by single-replicate analyses of variance (using the three-way interaction term as the *MSE* estimate) on the coverage rates for each coefficient. These analyses were conducted using the data from all sample sizes. To measure effect sizes we computed ω^2 (Hays, 1994, p. 499).

Overall, the results for each coefficient were highly similar (all results available upon request) and are well summarized by the findings for the sixth coefficient. For β_6 we found $\omega_{R^2}^2 = .25$, $\omega_{\beta_{\mathbf{v}}}^2 = .20$, and $\omega_{R^2 \times \beta_{\mathbf{v}}}^2 = .54$; all other effects were essentially zero. In aggregate, these findings suggest that when R^2 is low and the predicted criterion scores are highly correlated with the first principal component of the regressors, then the standard method produces CIs with accurate coverage rates. However, when R^2 is high and the predicted criterion scores are highly correlated with the last principal component of the regressors, then the standard method produces CIs with inaccurate coverage rates. For instance, the last row of Table 2 corresponds to a condition in which the correlation between the criterion and the last principal component of the regressors is .89. This condition is known to produce suppression effects (Waller, 2011).

To better understand the above results, let us consider the CIs for the fifth coefficient of β_i in greater detail. Locate column 5 in Table 2, and read the entries in rows 11 and 12. As shown in Table 1, these numbers denote the coverage rates for $\beta_{60^\circ[5]} = 1.17$ and $\beta_{90^\circ[5]} = 1.32$. Importantly, these relatively large coefficients have coverage rates with the largest amounts of positive and negative

bias. Notice that although the coefficients are similar in size (1.17 vs. 1.32), their associated coverage rates are considerably different (with coverage rate discrepancies of 3.38 and -19.64). This discrepancy demonstrates that regression coefficient size is only one of several factors that determine coverage rate accuracy. Orientation of the regression coefficients, $\beta_{\mathbf{v}}$, is another important determinant. As we now demonstrate, still other determinants can be gleaned from the underlying equation for DEL.

For the reader's convenience, we reproduce Equation (24) (the formula for DEL) below.

$$\text{var}(\hat{\beta}_j) = \frac{\sigma_{X_j}^2 c_j \sigma_e^2}{n \cdot \sigma_Y^2} + \frac{b_j^2 [\sigma_{X_j}^2 (\mathbf{b}' \Sigma_X \mathbf{b}) - \sigma_{X_j}^2 \sigma_e^2 - \sigma_{X_j Y}^2]}{n \cdot \sigma_Y^4}.$$

From the above, after recalling that c_j denotes the j th diagonal entry of Σ_X^{-1} , it is easily shown that the squared standard error from the standard method can be written,

$$\frac{\sigma_{X_j}^2 c_j \sigma_e^2}{n \cdot \sigma_Y^2} = \frac{\sigma_{X_j}^2}{\sigma_Y^2} \cdot \frac{\sigma_e^2 \Sigma_X^{-1}}{n} = \frac{\sigma_{X_j}^2}{\sigma_Y^2} \cdot \text{var}(\hat{b}_j). \quad (30)$$

When expressed in this form, it is apparent that the standard method CI for a standardized regression coefficient will be biased whenever the second term in (24) is nonzero. For example, when the second term is positive (and ignored), the standard method standard error is too small. When the second term is negative (and ignored), the standard method standard error is too large. This observation goes a long way in explaining the findings in Table 2. For instance, returning to $\beta_{60^\circ[5]}$, we find that the first and second terms of (24) equal 0.001 and -0.0003. By ignoring the latter value, the standard method estimates of $\text{var}(\hat{\beta}_j)$ is positively biased by 43%. Similar logic explains why the standard method estimate of $\text{var}(\hat{\beta}_{90^\circ[5]})$ is negatively biased by 67%.

These results demonstrate that the bias in the standard method is elucidated by the equation for DEL. Yuan and Chan (2011) recognized this point and stated that the standard method "bias

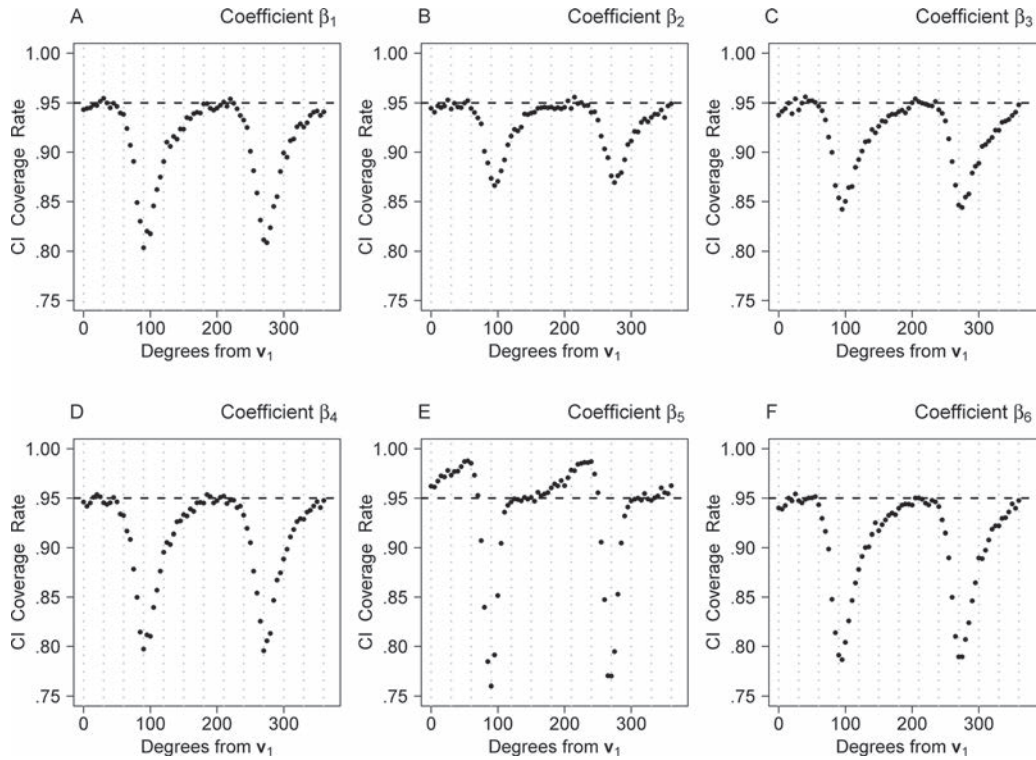


Figure 3. A–F: These display the empirical, standard method confidence interval (CI) coverage rates for the six elements of 72 population regression coefficient vectors.

depends on the value of β_j , Σ_X , and σ_e^2 ” (p. 676; Yuan and Chan’s notation has been modified to be consistent with Equation (24)) and that “the bias is positive when σ_e^2 is large and negative when σ_e^2 is small” (p. 676). However, as shown in Figure 3, the role of σ_e^2 in determining the bias of the standard method is less straightforward.

The six panels in Figure 3 display the empirical CI coverage rates for 72 population regression vectors. Each panel shows the results for a single coefficient (i.e., 3A shows the results for the coefficient that is associated with the first regressor; 3B shows the results for the coefficient that is associated with the second regressor; etc.). All models used the predictor correlation matrix in (29), and all models had a $R^2 = .80$. The population coefficient vectors (i.e., the vectors of standardized regression coefficients) were constrained to lie in the (v_1, v_6) plane. The first vector, β_{0° , was collinear with v_1 ; then β_{0° was rotated in 5° increments to produce the remaining 71 vectors (note that $72 \times 5^\circ = 360^\circ$). From each population, we drew 5,000 samples of size $n = 500$ and computed standard method CIs.

While reviewing the six panels in Figure 3, recall that the residual variances of the population regression models were held constant at $\sigma_e^2 = .20$. This common residual variance is relatively small (due to the scaling of the population predictors and criterion, the residual variance must lie between 0 and 1), and most panels show findings that are consistent with Yuan and Chan’s (2011) claim. Namely, when σ_e^2 is small, the standard method bias is negative. However, in Figure 3E we see that

under these same conditions the standard method bias can be positive. For instance, the coverage rates for the fifth element of the 72 regression vectors (shown in 3E) range from a high of 99% to a low of 76%. Thus, when σ_e^2 is small, the standard method bias can be either negative or positive. Moreover, the sign and magnitude of the bias is related to the orientation of the standardized regression coefficients with respect to the eigenvectors of the predictor correlation matrix.

Before leaving this section, one more aspect of Figure 3 merits comment. To wit, although it is difficult to discern from these plots, our results demonstrate that coefficient size, like σ_e^2 , is not a sufficient indicator of the sign and magnitude of the standard method bias. Again, focusing on Figure 3E, our results showed that when $\angle(\beta, v_1) = 70^\circ$, $\beta_{70^\circ[5]} = 1.32$, and $r_{X_5Y} = .78$, the CI coverage rate was 95%. However, when $\angle(\beta, v_1) = 90^\circ$ (i.e., when the regression coefficients are orthogonal to the first eigenvector of the predictor correlation matrix), $\beta_{90^\circ[5]} = 1.32$, and $r_{X_5Y} = .22$, the CI coverage rate was only 76%. Importantly, at both positions, the regression coefficient for the fifth predictor is 1.32 (rounded to two decimal places). Nevertheless, the coverage rates for these coefficients are surprisingly different. This result can also be understood by returning to Equation (24). A close inspection of this formula reveals that standard method bias is also a function of the squared predictor-criterion validity ($\sigma_{X_jY}^2$). Consequently, equally sized regression coefficients can have different CI coverage rates when their predictor-criterion validities differ.

Table 3
 Discrepancies in the Empirical and Expected CI Coverage Rates for All Methods for $\kappa = 3$, and $N = 500$

Location of β_i	Method									
	SM		NCIE		DEL		PRC		BCA	
	$\beta_{i[5]}^a$	$\beta_{i[6]}$	$\beta_{i[5]}$	$\beta_{i[6]}$	$\beta_{i[5]}$	$\beta_{i[6]}$	$\beta_{i[5]}$	$\beta_{i[6]}$	$\beta_{i[5]}$	$\beta_{i[6]}$
$R^2 = .20$										
$\beta_i \parallel \mathbf{v}_1^b$	0.72 ^c	0.26	0.32	-0.10	-0.32	-0.36	0.30	0.16	0.24	0.12
$\angle(\beta_i, \mathbf{v}_1) = 30^\circ$	0.52	0.20	1.12	0.22	-0.48	0.02	-0.42	-0.08	-0.38	0.00
$\angle(\beta_i, \mathbf{v}_1) = 60^\circ$	1.42	0.26	2.26	0.16	-0.04	0.30	-0.30	-0.22	-0.36	-0.22
$\beta_i \parallel \mathbf{v}_6$	1.38	1.26	2.46	1.70	0.00	-0.22	-0.26	-0.52	-0.46	-0.52
$R^2 = .80$										
$\beta_i \parallel \mathbf{v}_1$	1.22	-1.00	3.20	0.62	0.20	0.34	-0.30	0.46	-0.40	0.34
$\angle(\beta_i, \mathbf{v}_1) = 30^\circ$	2.60	0.10	4.66	0.26	0.30	0.14	-0.20	-0.36	-0.34	-0.32
$\angle(\beta_i, \mathbf{v}_1) = 60^\circ$	3.38	-1.08	5.00	-0.32	0.08	-0.36	0.44	-0.28	0.36	-0.32
$\beta_i \parallel \mathbf{v}_6$	-19.64	-15.86	-0.74	-0.56	0.14	0.28	-0.02	-0.82	-0.12	-0.88

Note. CI = confidence interval; SM = standard method; NCIE = noncentral interval estimation; DEL = delta method; PRC = percentile bootstrap; BCA = bias-corrected and accelerated bootstrap.

^a $\beta_{i[j]}$ denotes the j th element of β_i . ^b The notation $a \parallel b$ denotes that vectors a and b are parallel. ^c Tabled values denote the difference between the observed and expected coverage rates (expected = 95%); negative numbers denote CIs that are too narrow, positive numbers denote CIs that are too wide.

Having reviewed the performance of the standard method for computing CIs for standardized regression coefficients, we now turn our attention to the performance of NCIE, DEL, PRC, and BCA.

Our complete design considered models with $R^2 = \{.20, .50, .80\}$, $\kappa = \{3, 6, 9\}$, and $n = \{50, 100, 500\}$. For these conditions, $\beta_V = \{0^\circ, 30^\circ, 60^\circ, 90^\circ\}$ from \mathbf{v}_1 in the $(\mathbf{v}_1, \mathbf{v}_6)$ plane. To conserve space, Table 3 summarizes the results for $R^2 = \{.20, .80\}$, $\kappa = \{3\}$, $n = \{500\}$, and the last two coefficients of β_i (i.e., $\beta_{i[5]}$ and $\beta_{i[6]}$; all results are available upon request). These results demonstrate the important point that a CI method can produce coverage rates with disparate accuracies for different coefficients within the same model. For instance, notice that when $\beta_V = 60^\circ$ and $R^2 = .80$, NCIE yields a coverage rate of 100% (positive bias of $\frac{100}{95} - 1) \% = 5.3\%$) for the first coefficient and a coverage rate of 95.04% (positive bias of .04%) for the second coefficient. Also notice that the alternative methods frequently outperformed the standard method. Among the alternatives, DEL produced the most accurate coverage rates overall. This was especially true when the regression coefficients moved toward the last eigenvector of \mathbf{R}_X . That is, DEL was especially effective when the predicted criterion scores were highly correlated with the *last* principal component of the standardized regressors. The two bootstrap methods also performed well. Thus, because of its computational efficiency we recommend that researchers use DEL to compute CIs for standardized regression coefficients unless data are markedly nonnormal (Micceri, 1989). When data are markedly skewed, or otherwise nonnormal, the bootstrap procedures may be preferable (especially in larger samples). Future research will need to explore these recommendations more carefully.

Discussion

With fixed predictors, the so-called “textbook” method (Yuan & Chan, 2011) for computing confidence intervals for standardized regression coefficients fails to account for the sampling variability of the criterion standard deviation. With random predictors, this method also fails to account for the sampling variability of the

predictor standard deviations. With these points in mind, we conducted a Monte Carlo study to address two questions. First, when will the textbook method (Yuan & Chan, 2011; see also Harris, 2001, p. 80; Hays, 1994, p. 709; Kelley & Maxwell, 2003) for computing CIs for standardized regression coefficients produce accurate coverage rates? Second, when will the noncentrality interval estimation method (NCIE), the delta method (DEL), the percentile bootstrap method (PRC), and the bias-corrected and accelerated bootstrap (BCA) method produce CIs for standardized coefficients with accurate coverage rates?

With respect to the first question, our findings suggest that the accuracy of the textbook method—which we called the “standard method” in this article—depends on three factors: (a) the strength of the regression model as quantified by R^2 , (b) the orientation of the standardized regression coefficients in \mathbb{R}^p (p -dimensional Real space) and, (c) the interaction of these factors. Specifically, accuracy was high when β_i was near the first eigenvector of \mathbf{R}_X (the predictor correlation matrix). This suggests that the standard method will produce accurate coverage rates when the predicted criterion scores are highly correlated with the first principal component of the predictors. However, when R^2 is high and β_i is closely aligned with the last eigenvector of \mathbf{R}_X then the performance of the standard method can be disturbingly erratic. Interestingly, in some cases the bias of the standard method was strongly negative. In one sense, this finding supports an oft cited result that standard errors from covariance structure models that have been fit to correlation matrices are often wrongly computed in popular software (see Cudeck, 1989, p. 323). In another sense, they seem to conflict with the view that the incorrectly computed standard errors will generally be too big (Kelley & Maxwell, 2003; MacCallum & Austin, 2000, p. 217). Our results demonstrate that the standard method can produce standard errors that are (a) highly accurate, (b) too big, or (c) too small. Thus, we suggest that researchers would be well advised not to conclude that standard method standard errors yield conservative inferential results.

Moving on to the second question, our results suggest that DEL (Yuan & Chan, 2011) is the best choice among the alternatives for computing CIs for standardized regression coefficients. Although

its performance suffered slightly in small samples, DEL bias was never more than 2.3%. On the other hand, the bootstrap procedures (Percentile and BCA methods) also worked well in all but the smallest samples. We therefore also recommend the PRC or BCA bootstrap for computing CIs for standardized regression coefficients when computational cost is not an issue. We do not recommend NCIE for standardized regression models with random predictors due to its erratic performance.

As with all Monte Carlo studies, the generalizability of our findings is limited by the constraints of our design. Nevertheless, we believe that our results provide useful guidance to researchers wishing to assess the statistical fidelity of standardized regression coefficients. For this purpose, we recommend the delta method (Yuan & Chan, 2011) or a nonparametric bootstrap (Efron & Tibshirani, 1998). We provide R code in Appendix B for the delta method. The R boot package (Canty & Ripley, 2011) includes suitable functions for the bootstrap.

References

- Abadir, K. M., & Magnus, J. R. (2005). *Matrix algebra*. New York, NY: Cambridge University Press.
- Achen, C. H. (1977). Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science*, 21, 805–815. doi:10.2307/2110737
- Algina, J., & Moulder, B. C. (2001). Sample sizes for confidence intervals on the increase in the squared multiple correlation coefficient. *Educational and Psychological Measurement*, 61, 633–649. doi:10.1177/00131640121971400
- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). Hoboken, NJ: Wiley.
- Beasley, W. H., Deshea, L., Toothaker, L. E., Mendoza, J. L., Bard, D. E., & Rodgers, J. L. (2007). Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods*, 12, 414–433. doi:10.1037/1082-989X.12.4.414
- Bentler, P. M., & Lee, S. Y. (1983). Covariance structures under polynomial constraints: Applications to correlations and alpha-type structural models. *Journal of Educational Statistics*, 8, 207–222. doi:10.2307/1164760
- Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 20, 115–140. doi:10.2307/271084
- Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, 48, 209–213. doi:10.2307/2684719
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.
- Browne, M. W., & Du Toit, S. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research*, 27, 269–300. doi:10.1177/0049124192021002005
- Browne, M. W., Mels, G., & Cowan, M. (1994). *Path analysis: RAMONA: SYSTAT for DOS advanced applications* (Version 6, pp. 167–224) [Computer software]. Evanston, IL: SYSTAT.
- Canty, A., & Ripley, B. (2011). *boot: Bootstrap R (S-plus) functions: R package (Version 1.3–9)*. Retrieved from http://cran.us.r-project.org/src/contrib/boot_1.3-9.tar.gz
- Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19, 1141–1164. doi:10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F
- Casella, G., & Berger, R. L. (2001). *Statistical inference*. Belmont, CA: Wadsworth.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Cox, D., & Hinkley, D. (1974). *Theoretical statistics* (2nd ed.). Boca Raton, FL: Chapman and Hall.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317–327. doi:10.1037/0033-2909.105.2.317
- Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and non-central distributions. *Educational and Psychological Measurement*, 61, 532–574. doi:10.1177/00131640121971374
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, England: Cambridge University Press.
- DiCiccio, T., & Tibshirani, R. (1987). Bootstrap confidence intervals and bootstrap approximations. *Journal of the American Statistical Association*, 82, 163–170. doi:10.2307/2289143
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26. doi:10.1214/aos/1176344552
- Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171–185. doi:10.2307/2289144
- Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, 104, 293–296. doi:10.1037/0033-2909.104.2.293
- Efron, B., & Tibshirani, R. (1998). *An introduction to the bootstrap*. Boca Raton, FL: Chapman and Hall.
- Evans, M. J., & Rosenthal, J. S. (2004). *Probability and statistics: The science of uncertainty*. New York, NY: W. H. Freeman.
- Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random-effects effect sizes. *Educational and Psychological Measurement*, 61, 575–604. doi:10.1177/00131640121971383
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, 10, 507–521. doi:10.1093/biomet/10.4.507
- Fleishman, A. E. (1980). Confidence intervals for correlation ratios. *Educational and Psychological Measurement*, 40, 659–670. doi:10.1177/001316448004000309
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.
- Fox, J. (2009). *A mathematical primer for social statistics*. Thousand Oaks, CA: Sage.
- Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9, 1218–1228. doi:10.1214/aos/1176345638
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27, 2865–2873. doi:10.1002/sim.3107
- Graybill, F. A. (1976). *Theory and application of the linear model*. Belmont, CA: Wadsworth.
- Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Greenland, S., Maclure, M., Schlesselman, J. J., Poole, C., & Morgenstern, H. (1991). Standardized regression coefficients: A further critique and review of some alternatives. *Epidemiology*, 2, 387–392. doi:10.1097/00001648-199109000-00015
- Harris, R. J. (2001). *A primer on multivariate statistics* (3rd ed.). Mahwah, NJ: Erlbaum.
- Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace College.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.
- Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Wadsworth Cengage.

- Hunter, J. E., & Hamilton, M. A. (2002). The advantages of using standardized scores in causal analysis. *Human Communication Research, 28*, 552–561. doi:10.1111/j.1468-2958.2002.tb00823.x
- Jamshidian, M., & Bentler, P. M. (2000). Improved standard errors of standardized parameters in covariance structure models: Implications for construct explication. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 73–94). Dordrecht, the Netherlands: Kluwer Academic.
- Johnson, N. L., & Welch, B. L. (1940). Applications of the non-central t-distribution. *Biometrika, 3/4*, 362–389. doi:10.2307/2332616
- Johnson, R. (2001). An introduction to the bootstrap. *Teaching Statistics, 23*, 49–54. doi:10.1111/1467-9639.00050
- Jones, J. A. (2010). GenCorr: An R routine to generate correlation matrices from a user-defined eigenvalue structure. *Applied Psychological Measurement, 34*, 68–69. doi:10.1177/0146621609339727
- Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20*, 1–24.
- Kelley, K., & Lai, K. (2011). MBESS: R package (Version 3.2.1). Retrieved from <http://CRAN.R-project.org/package=MBESS>.
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8*, 305–321. doi:10.1037/1082-989X.8.3.305
- Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics* (Vol. 2). New York, NY: Macmillan.
- King, G. (1985). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30*, 666–687. doi:10.2307/2111095
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201–226. doi: 10.1146/annurev.psych.51.1.201
- Marsaglia, G., & Olkin, I. (1984). Generating correlation matrices. *SIAM Journal on Scientific and Statistical Computing, 5*, 470–475. doi: 10.1137/0905034
- Meehl, P. E. (1995). Bootstrap taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–275. doi: 10.1037/0003-066X.50.4.266
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156–166. doi:10.1037/0033-2909.105.1.156
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. New York, NY: McGraw-Hill.
- Oehlert, G. W. (1992). A note on the delta method. *The American Statistician, 46*, 27–29. doi:10.2307/2684406
- Owen, D. B. (1968). A survey of properties and applications of the noncentral t-distribution. *Technometrics, 10*, 445–478. doi:10.2307/1267101
- Rasmussen, J. L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches. *Psychological Bulletin, 101*, 136–139. doi: 10.1037/0033-2909.101.1.136
- Raspe, R. E. (2003). *The surprising adventures of Baron Munchhausen*. Rockville, MD: Wildside Press.
- R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rencher, A. C. (2008). *Linear models in statistics* (2nd ed.). Hoboken, NJ: Wiley.
- Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association, 69*, 682–689. doi:10.2307/2286002
- Shaffer, J. P. (1991). The Gauss-Markov theorem and random regressors. *The American Statistician, 45*, 269–273. doi:10.2307/2684451
- Shieh, G. (2006). Exact interval estimation, power, calculation, and sample size determination in normal correlation analysis. *Psychometrika, 71*, 529–540. doi:10.1007/s11336-04-1221-6
- Sievers, W. (1996). Standard and bootstrap confidence intervals for the correlation coefficient. *British Journal of Mathematical and Statistical Psychology, 49*, 381–396. doi:10.1007/s11336-04-1221-6
- Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180. doi:10.1207/s15327906mbr2502_4
- Steiger, J. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164–182. doi:10.1037/1082-989X.9.2.164
- Steiger, J., & Fouladi, R. (1997). Noncentrality interval estimation and the evaluation of statistical methods. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221–257). Mahwah, NJ: Erlbaum.
- Steiger, J. H., & Lind, J. C. (1980). *Statistically-based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Stein, S., & Barcellos, A. (1992). *Calculus and analytic geometry* (5th ed.). McGraw-Hill.
- Strube, M. J. (1988). Bootstrap type I error rates for the correlation coefficient: An examination of alternate procedures. *Psychological Bulletin, 104*, 290–292. doi:10.1037/0033-2909.104.2.290
- Tierney, L., Rossini, A. J., Li, N., & Sevcikova, H. (2012). snow: Simple network of workstations: R package (Version 0.3–10). Retrieved from <http://CRAN.R-project.org/package=snow>
- Tukey, J. W. (1954). Causation, regression, and path analysis. In O. Kempthorne, T. A. Bancroft, J. W. Gowen, & J. L. Lush (Eds.), *Statistics and mathematics in biology* (pp. 35–66). Ames: Iowa State College Press.
- Waller, N. G. (2011). The geometry of enhancement in multiple regression. *Psychometrika, 76*, 634–649. doi:10.1007/s11336-011-9220-x
- Waller, N. G., & Jones, J. A. (2010). Correlation weights in multiple regression. *Psychometrika, 75*, 58–69. doi:10.1007/s11336-009-9127-y
- Waller, N. G., & Jones, J. A. (2011). Investigating the performance of alternate regression weights by studying all possible criteria in regression models with a fixed set of predictors. *Psychometrika, 76*, 410–439. doi:10.1007/s11336-011-9209-5
- Weisberg, S. (2005). *Applied linear regression*. Hoboken, NJ: Wiley.
- West, S. G., Aiken, L. S., Wu, W., & Taylor, A. B. (2007). Multiple regression: Applications of the basics and beyond in personality research. In R. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 573–601). New York, NY: Guilford Press.
- Yu, H. (2002). Rmpi: Parallel statistical computing in R. *R News, 2*, 10–14.
- Yuan, K. H., & Chan, W. (2011). Biases and standard errors of standardized regression coefficients. *Psychometrika, 76*, 670–690. doi:10.1007/s11336-011-9224-6

(Appendices follow)

Appendix A

Data Generation

In this section, we describe how the data were generated in our simulation study, and we report R code (R Core Team, 2012) that implements our method.

1. Select κ , the desired condition number for \mathbf{R}_X , and p the desired order of \mathbf{R}_X . We chose $\kappa = 3$ and $p = 6$ for the example described below.
2. To insure that \mathbf{R}_X is positive definite, the eigenvalues of \mathbf{R}_X must be greater than zero and sum to p . For $\kappa = 3$, the first (λ_1) and last (λ_p) eigenvalues can equal 1.8 and 0.2 (other choices are possible).
3. Generate and append $p - 2$ random numbers $\lambda_p < x_i < \lambda_1$, ($i = 1 \dots p - 2$) to the two values from Step 2.
4. Sort and scale the p numbers from Step 3 to sum to p . These are the eigenvalues of \mathbf{R}_X .
5. With the eigenvalues from Step 4, generate \mathbf{R}_X using Method III by Marsaglia and Olkin (1984). Jones (2010) has created an R function for this purpose.
6. Generate a vector of standardized regression coefficients, $\boldsymbol{\beta}$. These coefficients must yield a desired R^2 . In the example described below, we set $R^2 = .5$. For a given choice of \mathbf{R}_X and R^2 , there are an infinite number of choices for $\boldsymbol{\beta}$ (Waller & Jones, 2011). In our regression data, we simulated a criterion by controlling the orientation of $\boldsymbol{\beta}$ with respect to the eigenvectors of \mathbf{R}_X . Each $\boldsymbol{\beta}$ corresponds to a different predicted criterion. Regression coefficient vectors in our simulation study were constrained to lie in a plane defined by the first and last eigenvectors of \mathbf{R}_X . In the example described below, we generate a vector that is 30° from the first eigenvector of \mathbf{R}_X , 60° from the last eigenvector of \mathbf{R}_X , and 90° from the remaining eigenvectors. For this case, $\boldsymbol{\beta}$ is a linear combination of the first and last eigenvectors of \mathbf{R}_X .
7. For a given $\boldsymbol{\beta}$ and \mathbf{R}_X , compute the vector of predictor-criterion correlations: $\mathbf{r}_{XY} = \mathbf{R}_X \boldsymbol{\beta}$.
8. Construct a super correlation matrix, \mathbf{R}_S , from \mathbf{R}_X and \mathbf{r}_{XY} .
9. Sample multivariate normal data from a population with correlation structure \mathbf{R}_S . In R, this is easily accomplished with the `mvrnorm()` function.

The following code illustrates how to implement the above 9 steps in the R programming language. This code will generate 100 observations that are consistent with the aforementioned predictor and predictor-criterion correlations.

R Code for Generating Data

```
> set.seed(670)
> # Step 1
> kappa <- 3; p <- 6
> # Step 2
> L1 <- 1.8; L2 <- .2
> # Step 3
> tmp.ivals <- c(L1, runif(4,.2,1.8), L2)
> # Step 4
> (ivals <- sort(tmp.ivals*(p/sum(tmp.ivals)),dec=TRUE))
[1] 1.9912683 1.1967666 1.0804206 1.0133657 0.4969268 0.2212520
>
```

(Appendices continue)

```

> # Check condition number.
> sqrt(evals[1]/evals[6])
[1] 3
> # Step 5
> # Use Marsaglia and Olkin's (1984) Method III
> # to construct a correlation matrix with the above eigenvalues.
> # R function GenCorr (Jones, 2010) can be
> # used for this purpose.
> # Rx <- GenCorr(evals)
> round(Rx,3)
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.000 0.369 -0.440 0.161 0.108 0.325
[2,] 0.369 1.000 0.079 -0.003 0.111 0.061
[3,] -0.440 0.079 1.000 0.071 0.060 -0.503
[4,] 0.161 -0.003 0.071 1.000 -0.054 0.338
[5,] 0.108 0.111 0.060 -0.054 1.000 0.264
[6,] 0.325 0.061 -0.503 0.338 0.264 1.000
> eigen(Rx)$values
[1] 1.9912683 1.1967666 1.0804206 1.0133657 0.4969268 0.2212520

> # Step 6
> # The following function is used in Step 6.
> # Convert degrees to radians.
> d2r <- function(deg) pi/180 * deg
>
> # Desired coefficient of determination.
> Rsq <- .5
> # Desired number of degrees from
> # 1st eigenvector of Rx.
> deg <- 30
> rad <- d2r(30)
> # Compute eigenvectors of Rx.
> V <- eigen(Rx)$vectors

> # Create a vector 30 degrees between v1 and v6.
> tmp.b <- cos(rad)*V[,1] + sin(rad)*V[,6]
> # Scale beta to have desired Rsq.
> beta <- sqrt(Rsq/as.numeric(t(tmp.b) %*% Rx %*% tmp.b))*tmp.b
> t(beta)%*% Rx %*% beta
      [,1]
[1,] 0.5

> # Step 7
> # Compute the predictor-criterion correlations.
> rxy <- Rx %*% beta

> # Step 8
> # Construct super correlation matrix R.S.

```

(Appendices continue)

```

> (round(R.S <- rbind(cbind(Rx,rxy), c(rxy,1)),3))
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 1.000 0.369 -0.440 0.161 0.108 0.325 -0.511
[2,] 0.369 1.000 0.079 -0.003 0.111 0.061 -0.230
[3,] -0.440 0.079 1.000 0.071 0.060 -0.503 0.513
[4,] 0.161 -0.003 0.071 1.000 -0.054 0.338 -0.256
[5,] 0.108 0.111 0.060 -0.054 1.000 0.264 -0.206
[6,] 0.325 0.061 -0.503 0.338 0.264 1.000 -0.528
[7,] -0.511 -0.230 0.513 -0.256 -0.206 -0.528 1.000
>
> # Step 9
> # Generate data from R.S.
> library(MASS)
> mvrnorm(n=100,mu=rep(0,7), Sigma=R.S)

```

Appendix B

R Function to Compute Normal-Theory Confidence Intervals for Standardized Regression Coefficients

The following function uses the delta method to construct confidence intervals for standardized regression coefficients using a method described in Yuan and Chan (2011).

```

#~~~~~#
# Arguments:
# X - An (n x p) matrix of predictor scores
# y - An (n x 1) vector of criterion scores
# cov.x - covariance matrix for predictors
# cov.xy - vector of covariances between
#         predictors and criterion
# var.y - criterion variance
# Nobs - number of observations
# alpha - desired Type I error rate
#         default = .05
# digits - number of significant digits to
#         print; default = 3
#
# This function accepts either (1) raw data, or (2)
# covariances and sample size.
#
# Output
# CIs - confidence intervals for standardized
#       regression coefficients
#~~~~~#

DEL <- function(X = NULL, y = NULL,
               cov.x = NULL, cov.xy = NULL,
               var.y = NULL, Nobs = NULL,
               alpha = .05, digits = 3) {

# vech function
vech <- function(x) t(x[!upper.tri(x)])

# Transition or Duplicator Matrix
Dn <- function(x){
  mat <- diag(x)
  index <- seq(x * (x+1) / 2)
  mat[lower.tri(mat, TRUE)] <- index
  mat[upper.tri(mat)] <- t(mat)[upper.tri(mat)]
  outer(c(mat), index, function(x, y) ifelse(x == y, 1, 0))
}

```

(Appendices continue)


```
#####
#~~~~~ Error Checking ~~~~~#
if(is.null(X) & !is.null(y))
  stop("\n y is not defined\n Need to specify both X and y\n")
if(!is.null(X) & is.null(y))
  stop("\n X is not defined\n Need to specify both X and y\n")
if(is.null(X) & is.null(y)) {
  if(is.null(cov.x) | is.null(cov.xy) | is.null(var.y) | is.null(Nobs))
    stop("\nYou need to specify covariances and sample size\n")
  scov <- rbind(cbind(cov.x, cov.xy), c(cov.xy, var.y))
  N <- Nobs
  p <- nrow(cov.x)

} else {
  scov <- cov(cbind(X, y))
  N <- length(y)
  p <- ncol(X)
}

# Create covariance matrix of covariances under normality
# See Browne (1984) Eqn 4.6

Kp.lft <- solve(t(Dn(p + 1)) %*% Dn(p + 1)) %*% t(Dn(p + 1))
cov.cov <- 2 * Kp.lft %*% (scov %x% scov) %*% t(Kp.lft)
param <- c(vech(scov))
ncovs <- length(param)

# Find the vector element numbers for the variances of X
v.x.pl <- c(1, rep(0, p - 1))
for(i in 2:p) v.x.pl[i] <- v.x.pl[i - 1] + p - (i - 2)

# Store covariances and variances to use in the derivatives

cx <- scov[1:p, 1:p]
cxy <- scov[1:p, p+1]
vy <- scov[p+1, p+1]

sx <- sqrt(diag(cx))
sy <- sqrt(vy)
bu <- solve(cx) %*% cxy
ncx <- length(vech(cx))

# Derivatives of standardized regression coefficients wrt the covariances.
# These are based on Yuan & Chan's (2011) equation 13

db <- matrix(0, p, ncovs)
V <- matrix(0, p, ncx)
V[as.matrix(cbind(1:p, v.x.pl))] <- 1

db[, 1:ncx] <- (diag(c(solve(diag(2 * sx * sy)) %*% bu)) %*% V -
  diag(sx / sy) %*% (t(bu) %*% solve(cx)) %*% Dn(p))

db[, (ncx+1):(ncx+p)] <- diag(sx / sy) %*% solve(cx)
db[,ncovs] <- -diag(sx / (2 * sy^3)) %*% bu
```

(Appendices continue)

```

# Re-order the derivatives

cx.nms <- matrix(0, p, p)
cxy.nms <- c(rep(0, p), "var_y")

for(i in 1:p) for(j in 1:p) cx.nms[i, j] <- paste("cov_x", i, "x", j, sep='')
for(i in 1:p) cxy.nms[i] <- paste("cov_x", i, "y", sep='')

old.ord <- c(vech(cx.nms), cxy.nms)
new.ord <- vech(rbind(cbind(cx.nms, cxy.nms[1:p]), c(cxy.nms)))

db <- db[, match(new.ord, old.ord)]

# Compute covariance matrix of standardized
# regression coefficients using the Delta Method

DEL.cmat <- db %%% cov.cov %%% t(db) / (N - 3)
b.nms <- NULL

for(i in 1:p) b.nms[i] <- paste("beta_", i, sep='')
rownames(DEL.cmat) <- colnames(DEL.cmat) <- b.nms

# compute standard errors and confidence intervals

DELse <- sqrt(diag(DEL.cmat))
CIs <- as.data.frame(matrix(0, p, 3))
colnames(CIs) <- c("lbound", "estimate", "ubound")
for(i in 1:p) rownames(CIs)[i] <- paste("beta_", i, sep='')

tc <- qt(alpha / 2, N - p - 1, lower = F)
beta <- diag(sx) %%% bu * sy^-1
for(i in 1:p) CIs[i,] <- c(beta[i] - tc * DELse[i], beta[i], beta[i]
+ tc * DELse[i])

cat("\n", 100 * (1 - alpha),
    "% CIs for Standardized Regression Coefficients:\n\n", sep='')
print(round(CIs,digits))
invisible(list(cov.mat=DEL.cmat,SEs=DELse,alpha=alpha,CIs=CIs))
}

# Example data from Yuan & Chan (2011).
cov.x <- matrix(c(3507.1691, 471.2058, 510.5430,
                 471.2058, 333.2295, 150.9121,
                 510.5430, 150.9121, 554.4386),3,3)
cov.xy <- matrix(c(454.5990, 128.6488, 138.0184),3,1)
var.y <- 101.4382
Nobs <- 46
out <- DEL(cov.x=cov.x,cov.xy=cov.xy,var.y=var.y,Nobs=Nobs)

# expected output

```

(Appendices continue)

```
# 95% CIs for Standardized Regression Coefficients:
#      lbound estimate ubound
# beta_1 0.337      0.495 0.653
# beta_2 0.231      0.391 0.552
# beta_3 0.107      0.263 0.419

# print standard errors
# round(out$SEs,3)
# beta_1 beta_2 beta_3
# 0.078 0.080 0.077
# Compare these values to Yuan & Chan (2011, p. 681) Table 1.
# Note: beta_1 = Narticle, beta_2 = Pctgrant, beta_3 = Pctsupp.
```

Received January 22, 2012
Revision received April 20, 2013
Accepted April 23, 2013 ■

Correction to Jones and Waller (2013)

In the article “Computing Confidence Intervals for Standardized Regression Coefficients” by Jeff A. Jones and Niels G. Waller (*Psychological Methods*, 2013, Vol. 18, No. 4, pp. 435–453), there were multiple production errors within the article that were not the fault of the authors. The online version has been replaced, and a corrected version of the printed issue has been reissued.

DOI: 10.1037/a0036247