Interval Estimates of R^2: An Empirical Investigation of the Influence of Fallible Regressors

Jeffrey D. Kromrey **Melinda R. Hess**
University of South Florida

Recent efforts to improve the analysis of multivariate data have included the use of confidence intervals rather than the more commonplace hypothesis testing. The use of interval estimation in regression analysis not only provides the ability to reject or fail to reject a given hypothesis, it also provides estimates of intervals within which a parameter is expected to reside. This study examines the potential effects of fallible regressors on the precision and accuracy of confidence intervals around R^2 when predictors vary in their reliability. Monte Carlo methods were used to investigate four methods for constructing these intervals around R^2: two percentile approaches based on the asymptotic normality of the distribution of R^2, a Fisher *Z* transformation method, and an interval inversion approach. The factors manipulated in the Monte Carlo study included the population value of R^2, number of regressors, sample size, population distribution shape, regressor intercorrelation, and regressor reliability. Results support the superiority of the interval inversion approach to confidence interval construction. However, as the reliability of the regressors decreased, none of the methods provided accurate intervals.

Recent efforts to improve the analysis of multivariate data have focused on (among other issues) the use of confidence intervals rather than the more commonplace hypothesis testing. In the context of multiple regression, many researchers (e.g. Steiger & Foray, 1992; Algina & Olejnik, 2000; Wilkinson & the APA Task Force, 1999) have provided justifications for the use of confidence intervals contending that they provide more information with better accuracy than the testing of null hypotheses. Of course, when properly applied, a confidence interval approach requires that researchers carefully consider design factors such as adequate sample size and appropriate procedures for sample selection and data collection. The use of interval estimation in regression analysis not only provides the ability to reject or fail to reject a given hypothesis (i.e., if the 1-α confidence band contains the null hypothesized parameter value), it also provides the researcher with estimates of intervals within which a parameter is expected to reside. The recent evolution of using confidence intervals for R^2 has primarily employed the assumption of normal distributions (e.g., Alf & Graf, 1999; Algina, 1999; Algina & Keselman, 1999) although emerging research into the effects of non-normal populations (Kromrey & Hess, 2001) on confidence intervals around R^2 has begun. In all of these investigations, however, there is one more element associated with realistic data that has not yet been addressed, namely the use of regressors that are not perfectly reliable. As such, this study examines the potential effects of fallible regressors on the precision and accuracy of confidence intervals around R^2 when predictors vary in their reliability.

Effects of Random Measurement Error in Regression

Although research on the effects of random measurement errors in regression analysis has a fairly long history (see Pedhazur, 1997, for a brief review) and the effects of measurement errors on the validity of regression analysis can be severe (Cochran, 1968). Jencks et al. (1972) suggested that "The most frequent approach to measurement error is indifference" (p. 330). Despite this apparent indifference in much of the applied research that utilizes regression analysis, the effects of random measurement errors (in either the criterion variable or the regressors) are known to result in a downward bias in the estimation of ρ^2 (Cochran, 1970). In addition, measurement error in the regressors in multiple regression models leads to bias (either positive or negative bias) in the regression coefficients (Cochran, 1968).

Two parameters are of interest when regression is based on predictor variables measured with error. One parameter is the population squared multiple correlation that would have been obtained if the regressors were measured perfectly $\left(\rho^2\right)$. This parameter, which represents a disattenuated multiple correlation, is primarily of interest in explanatory applications of regression, in which researchers are investigating relationships among variables for their theoretical importance. The second parameter of interest is the population squared multiple correlation that would be obtained using the fallible regressors themselves $\left(\rho_*^2\right)$. This parameter is primarily of interest in predictive applications of regression analysis, in which researchers are interested in the predictive power of the regressors as they are measured (i.e., including their measurement error).

Methods of Interval Estimation

Although interval estimates have been infrequently used in drawing inferences about the population squared multiple correlation $\left(\rho^2\right)$, several methods of constructing confidence bands are available. Fisher's (1928) derivation of the density function of R^2 has been implemented by Steiger and Fouladi (1992), using the interval inversion approach. This numerical method evaluates the cumulative distribution function of the sample R^2, given a population value of ρ^2. The method seeks that value of ρ^2 for which the obtained sample R^2 or smaller is expected (for example) 2.5% of the time and 97.5% of the time. These values of ρ^2 provide the endpoints of a 95% confidence band around the sample value of R^2.

Olkin and Finn (1995) provided several methods that were used by Algina (1999) to estimate confidence bands for the squared multiple correlation. The first method uses an estimated variance of R^2, given by

$$\sigma_{R^2}^2 = \frac{4\rho^2\left(1-\rho^2\right)\left(n-k-1\right)^2}{\left(n^2-1\right)\left(n+3\right)}$$

where n = sample size, and
k = number of regressors in the model.

A confidence interval is obtained by substituting the sample R^2 for ρ^2 in the equation (yielding $S_{R^2}^2$) and using $R^2 \pm z_{\alpha/2} S_{R^2}$ to obtain the endpoints of the confidence band.

A second method suggested by Olkin and Finn (1995) provides an estimate of the variance using

$$\dot{\sigma}_{R^2}^2 = \frac{4\rho^2\left(1-\rho^2\right)^2}{n}$$

As with the first method, a sample squared multiple correlation is used to obtain the estimated variance and a confidence band is constructed using the normal distribution.

The third method suggested by Olkin and Finn (1995) uses Fisher's z transformation of the multivariate R to normalize its distribution, resulting in a transformed variable with a variance of 4/n. That is,

$$z^* = \log_e\left(\frac{1+R}{1-R}\right)$$

The endpoints of a confidence band for the z* are given by

$$z^* \pm \frac{2z_{\alpha/2}}{\sqrt{n}}$$

If both endpoints (z_i) of this confidence band are non-negative, the endpoints are transformed to provide endpoints of the confidence band for ρ^2

$$\left[\frac{\exp\left(z_i\right)-1}{\exp\left(z_i\right)+1}\right]^2$$

If the lower endpoint of the confidence band for z* is negative, then the lower endpoint for the band for ρ^2 is set to zero.

The approximations presented by Olkin and Finn (1995) present problems when applied to samples from populations in which the squared multiple correlation is close to zero (Kendall and Stuart, 1977), and an investigation by Lee (1971) suggested that the Fisher transformation method worked poorly unless n was large relative to k. Recent work by Algina (1999) suggests that all of these approximations work poorly in comparison to the inversion method suggested by Steiger and Fouladi.

Previous studies on these estimates have included multivariate normal and non-normal data with perfectly reliable predictors. In reality, predictors in the social sciences have varying degrees of reliability and our intent was to investigate the performance of these confidence bands considering different levels of reliability among predictor variables.

Method

The confidence band estimates were constructed and compared using Monte Carlo methods, in which random samples were generated under known and controlled population conditions. In this Monte Carlo study, samples were generated from multivariate populations and each confidence band estimate was calculated based on each sample.

The Monte Carlo study included six factors in the design. These factors were (a) the true population multiple correlation (with ρ^2 = 0.01, 0.05, 0.10, 0.30, and 0.60), (b) number of regressor variables (with k = 2, 4, and 8), (c) sample sizes (with n = 5*k, 10*k, and 50*k), and (d) population distribution shape (conditions in which each variable evidenced population skewness and kurtosis values, respectively, of 0,0 [i.e., normal distribution]; 1,3; 1.5,5; 2,6; and 0,25), (e) regressor intercorrelation (ρ_{12} = 0.2, 0.4, 0.6, 0.8, 1.0), and (f) regressor reliability (ρ_{xx} = 0.40, 0.60, 0.80, 0.90, and 1.00).

Measurement error was simulated in the data following procedures used by Maxwell, Delaney, and Dill (1984), Jaccard and Wan (1995), and Kromrey and Foster-Johnson (1999). In this method, two normally distributed random variables for each regressor are generated, one of which represents the 'true score' on the regressor, the other representing measurement error. Fallible, observed scores on the regessors were calculated as the sum of the true and error components, consistent with classical measurement theory. The reliabilities of the regressors were controlled by adjusting the error variance relative to the true score variance by:

$$\rho_{xx} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2},$$

where σ_T^2 and σ_E^2 are the true and error variances, respectively, and ρ_{xx} is the reliability.

The research was conducted using SAS/IML version 8.1. Conditions for the study were run under Windows 98. Normally distributed random variables were generated using the RANNOR random number generator in SAS. A different seed value for the random number generator was used in each execution of the program. The program code was verified by hand-checking results from benchmark datasets.

For conditions involving nonnormal population distributions, the nonnormal data were produced by transforming the normal random variates obtained from RANNOR using the technique described by Bradley and Fleisher (1994), and operationalized by Ferron, *Yi*, and Kromrey (1997). In this method, a population correlation matrix, R, with a multivariate non-normal shape is constructed by an iterative process in which large simulated samples (*n* = 100,000) are generated from an approximation of R, $\tilde{R}$. The correlation matrix estimated from this large sample $\left(\hat{R}\right)$ is compared elementwise to R, and the residuals $\left(R-\hat{R}\right)$ are used to adjust the generating matrix $\tilde{R}$. This sequence of large sample generation, matrix estimation, and adjustment of $\tilde{R}$ continues until the process converges. The resulting matrix, $\tilde{R}$, is used to generate correlated non-normal data for the Monte Carlo study.

For each condition investigated in this study, 10,000 samples were generated. The use of 10,000 estimates provides adequate precision for the investigation of the sampling behavior of these confidence bands. For example, 10,000 samples provides a maximum 95% confidence interval width around an observed proportion that is $\pm$ .0098 (Robey & Barcikowski, 1992).

The relative performance of the confidence band estimates was evaluated by a comparison of the confidence band coverage (the proportion of confidence bands that included the population parameter) and the average width of the confidence bands. These indices correspond to statistical bias and estimation precision.

Results

The results were analyzed in terms of the confidence band coverage probabilities for the two parameters of interest, the population squared multiple correlation that would have been obtained with regressors measured without error $\left(\rho^2\right)$, and the squared multiple correlation in the population based upon the fallible regressors actually used $\left(\rho_*^2\right)$. In addition, the widths of the resulting confidence bands were analyzed. With the exception of tables for overall results for different reliability conditions for each of the three parameters of interest, in the interest of space and efficiency, other tables only contain results for the lowest, middle, and highest reliabilities investigated (0.4, 0.8 and 1.0) when three specific factors of the study design (shape, population squared multiple correlation, and sample size) are discussed. Many of the figures provided to illustrate results reflect conditions with the other reliabilities to maximize information. Specific results for reliabilities of 0.6 and 0.9 may be obtained from the authors if desired. Results for the different regressor intercorrelations did not show appreciable differences and are therefore not included in the detailed discussion.

Confidence Band Coverage of ρ^2

Figure 1 presents the distributions of coverage probabilities for the population squared multiple correlation based upon perfectly reliable regressors (i.e., ρ^2) across all conditions in this research. The band coverage for this parameter is rather poor under most conditions using a 95% confidence interval with extremely low coverage for specific conditions. The Steiger and Fouladi method provided the best band coverage overall, and the Olkin and Finn 3 method had notably poorer performance when compared to Olkin and Finn methods 1 and 2. Additional analyses of these coverage probabilities were addressed by considering the average coverage probabilities for differences in reliability, shape, population squared multiple correlation, and sample size relative to the number of regressors.

Reliability. Table 1 presents the distributions of coverage probabilities across all five reliabilities reflected in the conditions in this research. These values indicate the likelihood that ρ^2 will fall within the confidence interval when the fallibility of the regressors is not taken into account. Regardless of the method employed, all improve as reliability improves and the Steiger and Fouladi method consistently outperforms the other three methods, in spite of the number of regressors.

Distribution Shape. In Table 2, the impact of distribution shape on the ability of the different techniques to provide adequate coverage is explored based on the number of regressors as well as the

Figure 1. Distribution of confidence band coverage for ρ^2.

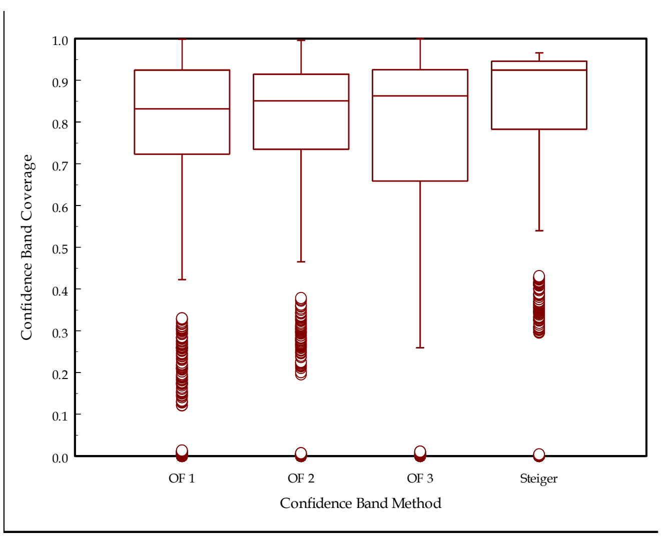

Table 1. Confidence Band Coverage for ρ^2 by Number of Regressors and Measurement Reliability

		Reliability				
	Method	0.4	0.6	0.8	0.9	1.0
2 Regressors	**O&F 1**	0.64	0.71	0.81	0.86	0.87
	O&F 2	0.67	0.74	0.82	0.86	0.86
	O& F 3	0.84	0.83	0.86	0.89	0.88
	S&F	0.71	0.77	0.86	0.91	0.92
4 Regressors	**O&F 1**	0.63	0.70	0.82	0.87	0.88
	O&F 2	0.66	0.72	0.82	0.86	0.86
	O& F 3	0.62	0.68	0.76	0.80	0.80
	S&F	0.63	0.70	0.81	0.88	0.92
8 Regressors	**O&F 1**	0.56	0.67	0.76	0.78	0.81
	O&F 2	0.60	0.68	0.77	0.79	0.80
	O& F 3	0.54	0.62	0.69	0.70	0.72
	S&F	0.55	0.70	0.78	0.88	0.92

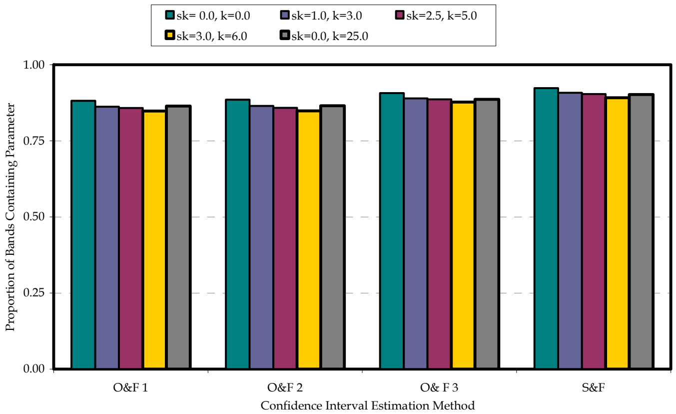

Figure 2. Proportion of confidence bands containing ρ^2 by distribution shape. Two regressors, reliability = 0.9

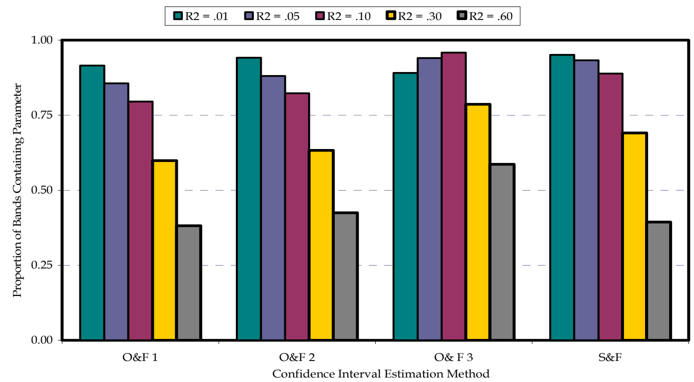

Figure 3. Proportion of confidence bands containing ρ^2 by ρ^2 . Two regressors, reliability = 0.6.

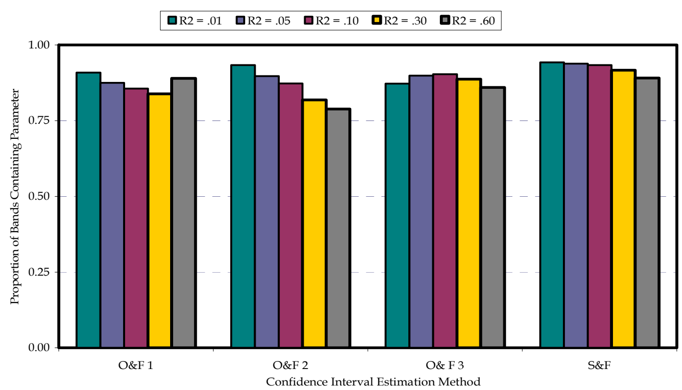

Figure 4. Proportion of confidence bands containing ρ^2 by ρ^2 . Two regressors, reliability = 1.0

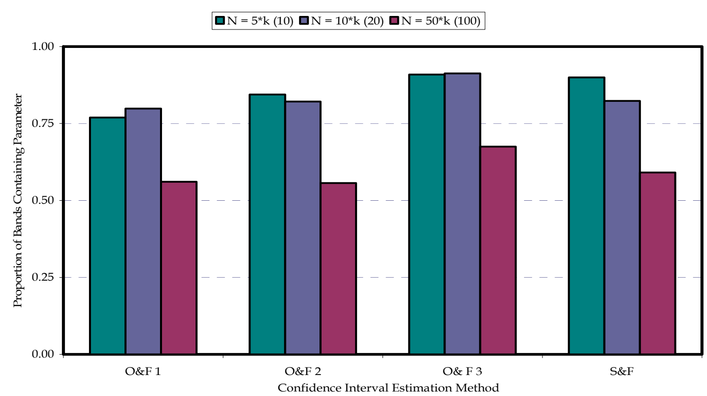

Figure 5. Proportion of confidence bands containing ρ^2 by sample size . Two regressors, reliability = 0.6.

reliability. An examination of these data as well as representation of coverage for 2 regressors with a reliability of 0.9 in Figure 2 indicates that the distribution shape has very little effect on the confidence band coverage within each method and that none of the methods provide very good coverage when ρ^2 is estimated.

Coverage is especially poor when the number of regressors is high (k=8) and reliability low (r = 0.4) with the Steiger and Fouladi method providing a consistent coverage probability of 0.55 at the low end and Olkin and Finn method 2 providing an almost consistent rate at approximately 0.60 across shapes. No method was consistently better across shapes with ρ^2, although the Steiger and Fouladi method did provide notably better coverage in a few isolated instances, i.e., when k = 8, r = 1.0 Steiger and Fouladi averaged a 0.91 coverage rate compared to the next best method, Olkin & Finn 1 which had a 0.80 coverage rate. Such instances of such clear superior performance were few.

Population Squared Multiple Correlation. When coverage was examined as a function of the population squared multiple correlation, there was a notable difference in how well the different methods performed. Table 3 provides estimated coverage for all conditions with reliabilities of 0.4, 0.8, and 1.0. Figures 3 and 4 illustrate these results for two specific conditions with 2 regressors (r = 0.6 and r = 1.0, respectively). When reliability is 0.6, Olkin and Finn method 3 tends to perform slightly better for higher values of ρ^2, however, as reliability increases, this slight superiority diminishes.

Table 2. Confidence Band Coverage for ρ^2 by Number of Regressors, Measurement Reliability and Distribution Shape.

			Distribution Shape				
	Reliability	Method	Sk = 0.0 Kurt = 0.0	Sk = 1.0 Kurt = 3.0	Sk = 1.5 Kurt = 5.0	Sk = 2.0 Kurt = 6.0	Sk = 0.0 Kurt = 25.0
2 Regressors	0.4	O&F 1	0.64	0.64	0.63	0.64	0.63
		O&F 2	0.68	0.68	0.67	0.68	0.67
		O& F 3	0.83	0.84	0.84	0.84	0.84
		S&F	0.71	0.71	0.70	0.71	0.70
	0.8	O&F 1	0.83	0.82	0.81	0.80	0.81
		O&F 2	0.84	0.83	0.82	0.81	0.82
		O& F 3	0.87	0.87	0.86	0.86	0.86
		S&F	0.87	0.86	0.85	0.85	0.85
	1.0	O&F 1	0.90	0.88	0.86	0.85	0.88
		O&F 2	0.89	0.87	0.85	0.83	0.87
		O& F 3	0.91	0.89	0.87	0.86	0.89
		S&F	0.95	0.93	0.92	0.90	0.93
4 Regressors	0.4	O&F 1	0.63	0.61	0.63	0.63	0.66
		O&F 2	0.66	0.64	0.66	0.66	0.69
		O& F 3	0.62	0.62	0.62	0.63	0.63
		S&F	0.62	0.60	0.63	0.63	0.66
	0.8	O&F 1	0.83	0.82	0.81	0.80	0.82
		O&F 2	0.83	0.81	0.81	0.80	0.82
		O& F 3	0.77	0.77	0.76	0.74	0.74
		S&F	0.82	0.81	0.81	0.80	0.82
	1.0	O&F 1	0.91	0.89	0.88	0.86	0.88
		O&F 2	0.89	0.86	0.85	0.83	0.87
		O& F 3	0.84	0.83	0.80	0.78	0.78
		S&F	0.95	0.93	0.91	0.90	0.92
8 Regressors	0.4	O&F 1	0.56	0.56	0.56	0.56	0.56
		O&F 2	0.60	0.60	0.60	0.60	0.59
		O& F 3	0.54	0.54	0.54	0.54	0.54
		S&F	0.55	0.55	0.55	0.55	0.55
	0.8	O&F 1	0.77	0.76	0.75	0.75	0.75
		O&F 2	0.78	0.77	0.76	0.76	0.76
		O& F 3	0.70	0.69	0.68	0.68	0.67
		S&F	0.79	0.78	0.78	0.77	0.77
	1.0	O&F 1	0.84	0.82	0.81	0.80	0.80
		O&F 2	0.83	0.81	0.79	0.78	0.79
		O& F 3	0.75	0.73	0.72	0.71	0.71
		S&F	0.95	0.93	0.92	0.90	0.91

Again, the poorest results were evident when low reliability and large numbers of regressors were involved. Effectiveness of the different methods varied. The Steiger and Fouladi method tended to do better when larger numbers of regressors were considered and reliability was 0.8 or 1.0 regardless of the population squared multiple correlation coefficient. Olkin & Finn 3, performed better than the others when reliability was low. When considering a ρ^2 of 0.6, O&F3 had a coverage of 0.60 when reliability was 0.4 with two regressors, compared with 0.31 by O&F 2, 0.27 by S & F, and 0.25 by O&F 1.

Table 3. Confidence Band Coverage for ρ^2 by Number of Regressors, Measurement Reliability and ρ^2.

			Population Squared Multiple Correlation				
	Reliability	Method	$\rho^2 = 0.01$	$\rho^2 = 0.05$	$\rho^2 = 0.1$	$\rho^2 = 0.3$	$\rho^2 = 0.6$
2 Regressors	0.4	O&F 1	0.92	0.83	0.73	0.46	0.25
		O&F 2	0.94	0.86	0.76	0.51	0.31
		O& F 3	0.90	0.95	0.97	0.77	0.60
		S&F	0.95	0.91	0.83	0.57	0.27
	0.8	O&F 1	0.91	0.87	0.84	0.77	0.66
		O&F 2	0.94	0.90	0.87	0.78	0.63
		O& F 3	0.88	0.92	0.94	0.88	0.69
		S&F	0.95	0.94	0.93	0.85	0.61
	1.0	O&F 1	0.91	0.88	0.86	0.84	0.89
		O&F 2	0.93	0.90	0.87	0.82	0.79
		O& F 3	0.87	0.90	0.90	0.89	0.86
		S&F	0.94	0.94	0.93	0.92	0.89
4 Regressors	0.4	O&F 1	0.89	0.85	0.74	0.48	0.19
		O&F 2	0.93	0.87	0.75	0.51	0.22
		O& F 3	0.66	0.88	0.81	0.58	0.20
		S&F	0.95	0.87	0.74	0.45	0.12
	0.8	O&F 1	0.89	0.88	0.87	0.80	0.64
		O&F 2	0.92	0.91	0.88	0.79	0.59
		O& F 3	0.64	0.86	0.89	0.83	0.57
		S&F	0.95	0.93	0.92	0.79	0.50
	1.0	O&F 1	0.88	0.87	0.87	0.87	0.92
		O&F 2	0.92	0.89	0.87	0.83	0.80
		O& F 3	0.63	0.83	0.86	0.86	0.84
		S&F	0.94	0.94	0.93	0.91	0.89
8 Regressors	0.4	O&F 1	0.76	0.79	0.69	0.46	0.12
		O&F 2	0.84	0.83	0.70	0.49	0.13
		O& F 3	0.56	0.80	0.72	0.53	0.11
		S&F	0.94	0.80	0.64	0.32	0.04
	0.8	O&F 1	0.73	0.81	0.84	0.80	0.60
		O&F 2	0.82	0.85	0.86	0.77	0.54
		O& F 3	0.52	0.77	0.84	0.79	0.52
		S&F	0.95	0.93	0.90	0.72	0.40
	1.0	O&F 1	0.77	0.77	0.80	0.86	0.92
		O&F 2	0.81	0.81	0.81	0.81	0.78
		O& F 3	0.71	0.70	0.77	0.82	0.82
		S&F	0.94	0.94	0.93	0.91	0.88

Sample Size. Table 4 provides coverage results by sample size. Coverage varied again across reliabilities and regressors, although there was little difference in coverage for conditions with 2 or 4 regressors under similar conditions; however, 8 regressor conditions typically showed a decrease.

At low reliabilities, O&F3 again did slightly better than the others, although it quickly lost ground as reliability increased. All methods performed very poorly when sample sizes were large and reliabilities were less than 0.9. For the reliability of 0.6 with two regressors, O&F 3 outperformed the others, as illustrated in Figure 5. However, this method was typically outperformed by the other Olkin and Finn methods as well as the Steiger and Fouladi methods under higher reliabilities as evident by Figure 6 which illustrates results for two regressors under perfect reliability conditions. At a sample size of 10*k for k = 4 and a reliability of 0.8, the O&F 3 method had a coverage rate of 0.81 while the others had approximately a 90% coverage rate.

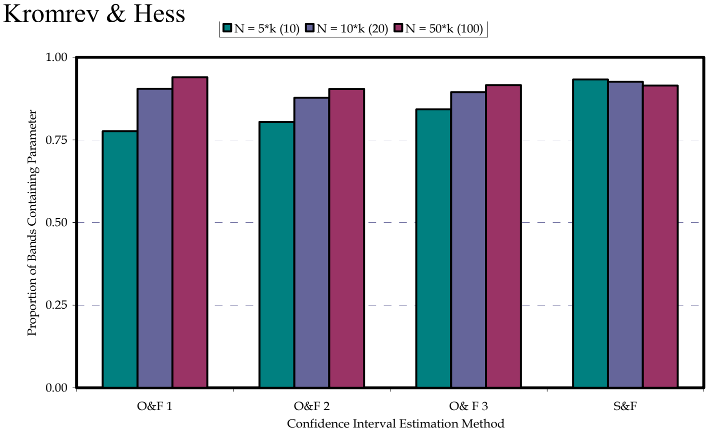

Figure 6. Proportion of confidence bands containing ρ_*^2 by sample size. Two regressors, reliability = 1.0

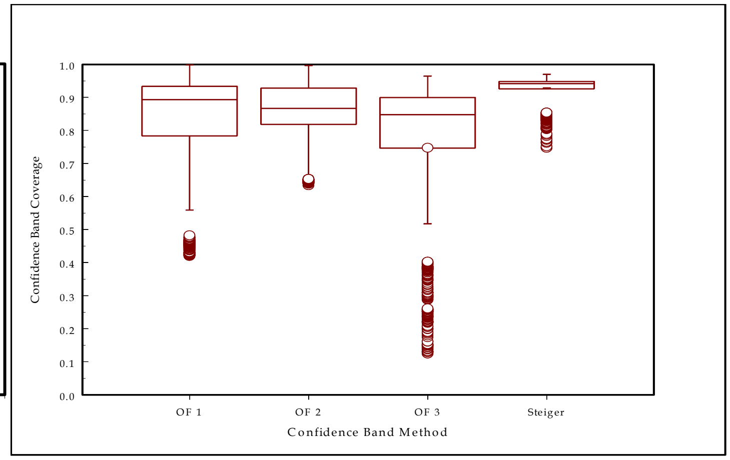

Figure 7. Distribution of confidence band coverage for ρ_*^2

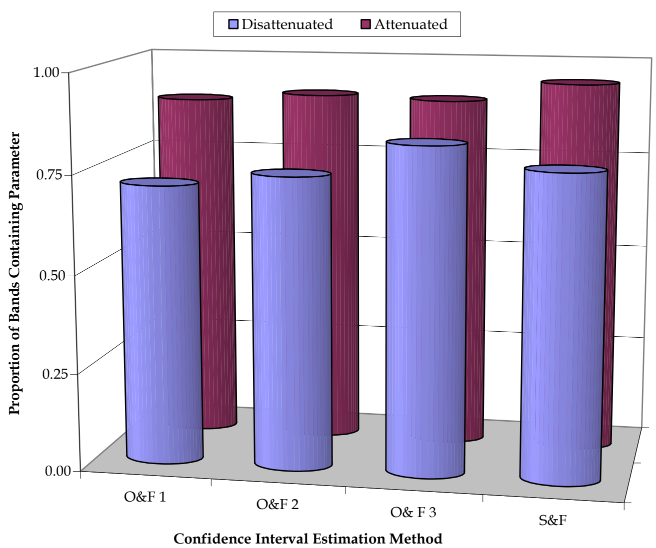

Figure 8. Comparison of the Proportion of Confidence Bands Containing ρ_*^2, k=2, r = 0.6, sk = 0.0 & kurt = 0.0.

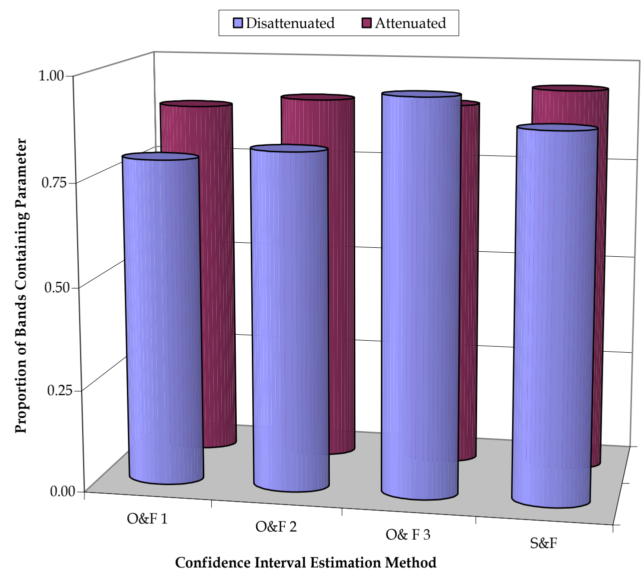

Figure 9. Comparison of the Proportion of ConfidenceBands Containing ρ_*^2, k=2, r = 0.6, and ρ_*^2 = 0.1.

Confidence Band Coverage of ρ_*^2

Figure 7 presents the distributions of coverage probabilities for the population squared multiple correlation when the fallibility of the regressors was taken into account (i.e., ρ_*^2). The use of 95% confidence bands based upon the Steiger and Fouladi method continued to provide the best band coverage overall. The Olkin and Finn method 2 had the next best results, although Figure 7 clearly shows distinctly poorer performance than the Steiger and Fouladi method. Olkin and Finn methods 1 and 3 present many conditions with poor band coverage, with method 3 providing extremely low coverage for some conditions. Additional analyses of these coverage probabilities were addressed by considering the average coverage probabilities for each factor in the Monte Carlo study design

Reliability. Table 5 presents the distributions of coverage probabilities for the parameter ρ_*^2 by number of regressors and reliability. When one compares and contrasts the values in Table 1 with the values in Table 5, it is readily apparent how coverage is improved when the fallibility of the regressors is taken into account. Once again, the Steiger and Fouladi method outperforms the other three under all conditions, with coverage ranging from 92% to 94% throughout the conditions. The other three

Table 4. Confidence Band Coverage for ρ^2 by Number of Regressors, Measurement Reliability and Sample Size.

	Reliability	Method	Sample Size N=5*k	N=10*k	N=50*k
2 Regressors	0.4	O&F 1	0.74	0.73	0.44
		O&F 2	0.83	0.75	0.44
		O& F 3	0.92	0.92	0.67
		S&F	0.87	0.76	0.49
	0.8	O&F 1	0.79	0.88	0.77
		O&F 2	0.85	0.89	0.74
		O& F 3	0.89	0.92	0.78
		S&F	0.93	0.90	0.74
	1.0	O&F 1	0.78	0.90	0.94
		O&F 2	0.80	0.88	0.90
		O& F 3	0.84	0.89	0.92
		S&F	0.93	0.93	0.91
4 Regressors	0.4	O&F 1	0.74	0.72	0.43
		O&F 2	0.82	0.73	0.44
		O& F 3	0.71	0.67	0.49
		S&F	0.80	0.68	0.41
	0.8	O&F 1	0.81	0.90	0.74
		O&F 2	0.85	0.89	0.70
		O& F 3	0.76	0.81	0.71
		S&F	0.81	0.90	0.74
	1.0	O&F 1	0.78	0.91	0.95
		O&F 2	0.79	0.88	0.91
		O& F 3	0.71	0.80	0.89
		S&F	0.93	0.92	0.91
8 Regressors	0.4	O&F 1	0.64	0.65	0.40
		O&F 2	0.72	0.67	0.40
		O& F 3	0.60	0.59	0.43
		S&F	0.72	0.60	0.32
	0.8	O&F 1	0.71	0.85	0.71
		O&F 2	0.78	0.84	0.68
		O& F 3	0.66	0.73	0.66
		S&F	0.89	0.82	0.63
	1.0	O&F 1	0.65	0.85	0.94
		O&F 2	0.69	0.82	0.90
		O& F 3	0.60	0.72	0.85
		S&F	0.93	0.92	0.91

performed similarly when conditions only called for two regressors, however, the Olkin and Finn 3 quickly fell behind in effectiveness as the number of regressors increased, i.e., when k = 4 and r = 0.4, the confidence band coverage probabilities for the methods, from best to worst were: (1) 0.94 for S&F, (2) 0.90 for O&F 2, (3) 0.88 for O&F 1, and (4) 0.69 for O&F 3. When the number of regressors increased to 8, the Steiger and Fouladi method maintained its performance in the mid 90's, however, all of the other three fell even more in their coverage rates, with O&F3 continuing to have much poorer performance than the other two Olkin and Finn methods which consistently had similar performance (Table 5).

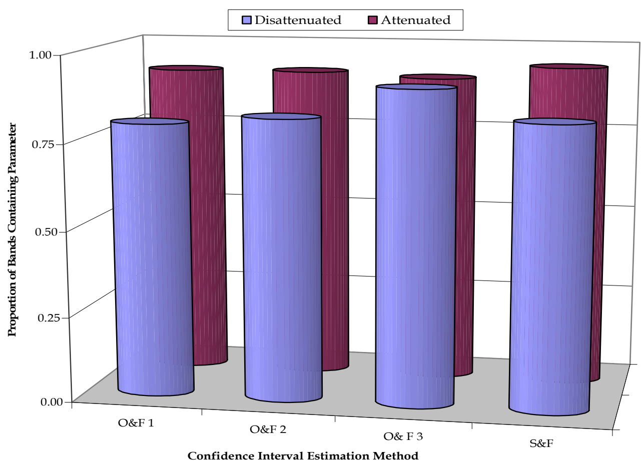

Figure 10. Comparison of the Proportion of Confidence Bands Containing ρ^2 and ρ_*^2 k=2, r = 0.6, N = 10*k (20).

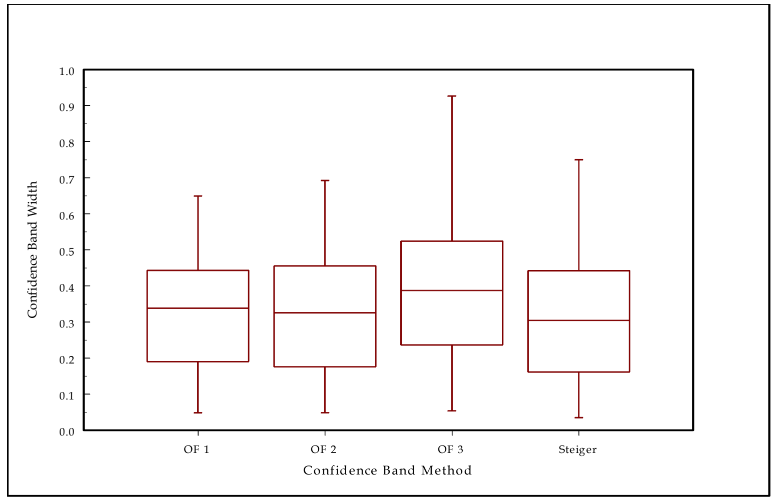

Figure 11. Distribution of Confidence Band Widths

Distribution Shape. Confidence band coverage probabilities for ρ_*^2 as a function of distribution shape is shown in Table 6. As with the confidence coverage of ρ^2, shape does not seem to have an appreciable effect on coverage of this parameter. However, once again, the Steiger and Fouladi method is superior to the other two under all conditions. The Olkin and Finn 3 method consistently provides the poorest performance. Figure 8 illustrates the improved coverage by all methods for one condition which is fairly representative of all conditions. When error in the regressors is accounted for, all methods have coverage's much closer to 95% although Steiger and Fouladi is still superior. Olkin and Finn 3 shows the least amount of improvement although it had the better coverage when regressor fallibility was not considered.

Population Squared Multiple Correlation. Table 7 and Figure 9 present coverage probabilities when considering differences in the population squared multiple correlation. Once again the Steiger and Fouladi method has consistently superior results when measurement error is accounted for. Mean coverages for this method range from 0.91 to 0.95. The three Olkin and Finn methods have similar performance when reliabilities are high; However, all three show dramatic drops in coverage when reliability falls to 0.4 and the number of regressors is greater than 2, with the Olkin and Finn method 3 performing very poorly even for high reliabilities under conditions containing eight regressors. When reliability is 0.8 and the number of regressors is 8, O&F 3 shows a coverage of only 45%, compared to 71% for O&F 1, 80% for O&F 2, and 95% for S&F. Once again, coverage is improved dramatically when results from ρ_*^2 are compared with ρ^2 as illustrated in Figure 9, at least for three of the four methods. Olkin and Finn 3, the least consistent of the methods actually shows comparable coverage whether ρ_*^2 or ρ^2 is considered under these conditions.

Sample Size. As sample size increases, so does the coverage for ρ_*^2 as clearly shown in Table 8. Steiger and Fouladi remains fairly consistent across sample size, regardless of the number of regressors or the reliability. However, Olkin and Finn 3, while somewhat satisfactory for conditions with large samples and a small number of regressors (0.93 when n= 50*k, r = 0.8, and k = 2), drops quite a bit with smaller sample sizes and a large number of regressors (0.57 when n = 5*k. r = 0.8, and k = 8). Olkin and Finn 1 and 2 also show relatively poor performance for small samples sizes (0.77 and 0.81, respectively when n = 5*k, r = 0.8, and k = 4). Once again, coverage is clearly improved across sample sizes when ρ_*^2 is estimated, rather than ρ^2. Figure 10 shows how all methods are close to 0.95 coverage when rho^2 is attenuated, with marked improvement for the Olkin and Finn 1 and 2 methods as well as the Steiger and Fouladi method. Interestingly, there was very little improvement for the Olkin and Finn 3 method.

Table 5.*Confidence Band Coverage for ρ_*^2 by Measurement Reliability.*

		Reliability				
	Method	0.4	0.6	0.8	0.9	1.0
2 Regressors	O&F 1	0.89	0.88	0.87	0.88	0.87
	O&F 2	0.92	0.90	0.88	0.87	0.86
	O& F 3	0.88	0.89	0.89	0.89	0.88
	S&F	0.94	0.94	0.93	0.93	0.92
4 Regressors	O&F 1	0.88	0.88	0.88	0.88	0.88
	O&F 2	0.90	0.89	0.88	0.87	0.86
	O& F 3	0.69	0.77	0.79	0.80	0.80
	S&F	0.94	0.94	0.93	0.93	0.92
8 Regressors	O&F 1	0.76	0.79	0.80	0.81	0.81
	O&F 2	0.81	0.81	0.81	0.80	0.80
	O& F 3	0.60	0.67	0.70	0.71	0.72
	S&F	0.94	0.94	0.93	0.93	0.92

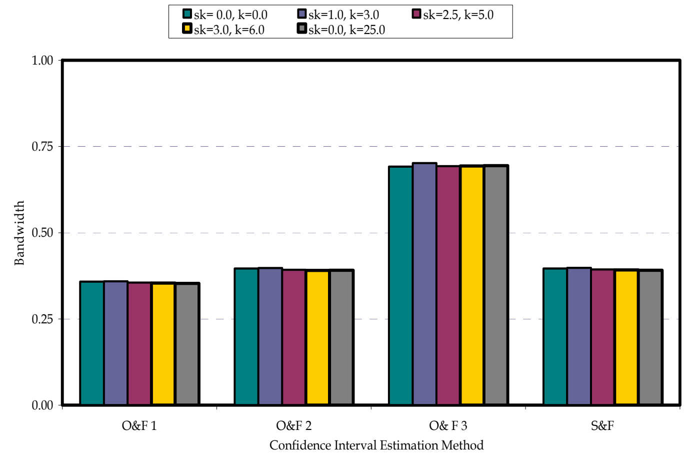

Figure 12. Width of confidence band by distribution shape. Two regressors, reliability = 0.6.

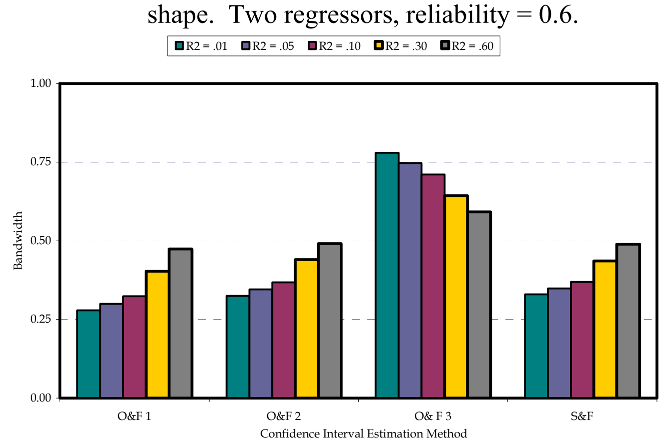

Figure 13. Width of confidence band by ρ^2. Two regressors, reliability = 0.6.

Confidence Band Width

Figure 11 presents the distributions of confidence band widths across all conditions in this research. This figure suggests that the Olkin and Finn method 3 produced slightly larger bands than the other methods. Additional analyses of these coverage probabilities were addressed by considering the average band widths for each factor in the Monte Carlo study design.

Reliability. Table 9 contains the bandwidths under different reliability conditions. In most cases, bandwidth increased as reliability did with the exception of the Olkin and Finn 3 method. This method showed either decreasing or approximately consistent bandwidths as reliability increased. When conditions contained 2 regressors, the bandwidth decreased by 0.11 when reliability changed from 0.4 to 1.0. Under conditions containing 4 and 8 regressors, bandwidths only varied 0.02 between different reliabilities. However, these isolated cases of width constriction and consistency are not of great consequence as they are either much larger than those provided by the other three methods (when k=2) or about the same as the other three methods (when k = 4 or k = 8). Steiger and Fouladi outperformed the other methods for conditions with a large number of regressors (width = 0.20 when k = 8, r = 0.4 compared to 0.24 or 0.25 for the others) and has widths that are either smaller than, or similar to, those constructed by the other methods for conditions with fewer regressors.

Distribution Shape. The confidence band widths were not appreciably related to the population distribution shape (Table 10). Across all distribution shapes, the average width of confidence bands constructed by Olkin and Finn method 3 were larger than those of the other methods, and the Olkin and Finn method 1 and Steiger and Fouladi approaches produced the smallest bands. The tendency for bandwidths constructed by Olkin and Finn 3 to be larger than the others is especially evident for conditions with only two regressors (Figure 12). As the number of regressors increased, the widths resulting from this method were very similar to those obtained by the other two Olkin and Finn methods, and usually just slightly larger than those constructed by the Steiger and Fouladi method.

Population Squared Multiple Correlation. The relationship between average band width and ρ^2 is presented in Table 11. All methods typically showed progressively larger bands as the value of the parameter increased, although there were some exceptions. As the number of regressors increased, the bandwidths tended to tighten a bit. For low values of ρ^2, the Steiger method was typically the best performer. When k=8, r = 0.4 and ρ^2 = 0.01, the Steiger method had a mean bandwidth of 0.16 while Olkin and Finn methods 1, 2, and 3, and widths of 0.20, 0.22, and 0.23 respectively. However, as ρ^2 increased (specifically 0.3 and 0.6), Steiger tended to provide bandwidths slightly larger than those produced by Olkin and Finn methods 2 and 3.

For conditions with a small number of regressors (k = 2), bands tended to get smaller when the Olkin and Finn 3 method was employed (Figure 13 is an example of this phenomena). However, when one takes into account the large bandwidth for smaller values of ρ^2 under these k = 2 conditions, this constriction does not provide any notable benefit compared to the other methods.

Sample Size. As anticipated, all of the methods produced smaller confidence intervals with larger samples (Table 12). The Olkin and Finn 1 and Steiger and Fouladi methods produced the smallest bands with small samples, but with samples of 10 or 50 times the number of regressors (10*k or 50*k) that occurred in conditions with either 4 or 8 regressors (k = 4 or k = 8), the difference in the confidence interval widths was negligible. However, when the number of regressors was small, k = 2, the bandwidths resulting from the Olkin and Finn method 3 were wider (Figure 14).

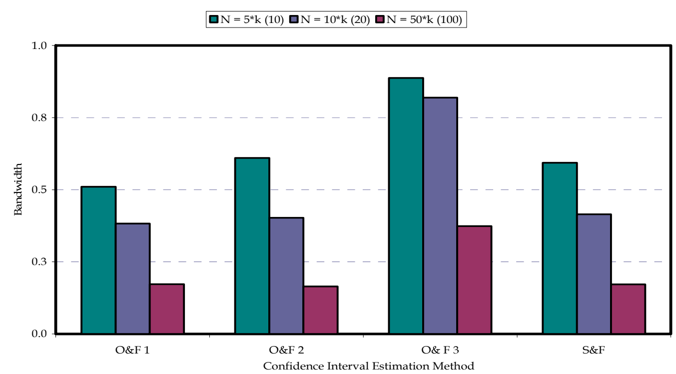

Figure 14. Width of confidence band by sample size. Two regressors, reliability = 0.6.

Conclusions

The general superiority of the Steiger and Fouladi approach to confidence band construction is evident in these results. This approach provided consistently more accurate confidence intervals across most of the conditions examined. Further, these more accurate bands were obtained without substantially increasing the confidence band width. Such results are consistent with a previous comparison of these methods with normal and non-normal data measured without error (Kromrey & Hess, 2001).

However, notable differences were obtained between the accuracy of the confidence bands for the two parameters of interest in this study. The population squared multiple correlation obtained with fallible regressors $\left(\rho_*^2\right)$ was accurately estimated with the Steiger and Fouladi bands, but the parameter that would be realized if the regressors were measured without error $\left(\rho^2\right)$ was poorly estimated in most conditions. The difference in these parameters and the differential success in estimating them, suggests that it is incumbent upon researchers to remain cognizant of the distinction. Concern for the psychometric characteristics and the consequences of poor reliability needs to become a part of the variable selection process in regression applications. We concur with others that unreliable regressors can and should be avoided (cf. Cohen & Cohen, 1983).

Obtaining accurate estimates of the parameter ρ^2 in the context of fallible regressors may require a disattenuation of the sample R^2 as a part of the confidence band construction. For example, Fuller and Hidirouglou (1978) described a correction for attenuation as a sample covariance matrix that is modified using reliabilities or error estimates obtained from a source independent of the sample covariance matrix.

However, the sampling distribution of a disattenuated ρ^2 will differ from that of the sample R^2, and the interval inversion approach of Steiger and Fouladi will need to be adjusted for these sampling characteristics. Further research in this direction is certainly in order.

It is becoming increasingly critical in today's high stakes educational environment to ensure that measurement methods and statistical analyses are as accurate and informative as possible. A continued reliance solely on the testing of null hypotheses in conjunction with unrealistic assumptions (i.e., normality; perfectly reliable predictors) limits the amount of information that we may glean from available information on students, teachers, methods and the myriad of other elements that compose educational systems. Although hypothesis testing provides probabilistic information about the accuracy of rejecting a null hypothesis, the proper development and use of confidence intervals for correlation applications under less than perfect conditions will allow us to estimate the bounds within the parameter is expected to reside with increased accuracy. As a result of this expected increase in precision, practitioners will be better equipped to make critical decisions with greater confidence.

Table 6. Confidence Band Coverage for ρ_*^2 by Number of Regressors, Reliability & Distribution Shape.

			Distribution Shape				
	Reliability	Method	Sk = 0.0 Kurt = 0.0	Sk = 1.0 Kurt = 3.0	Sk = 1.5 Kurt = 5.0	Sk = 2.0 Kurt = 6.0	Sk = 0.0 Kurt = 25.0
2 Regressors	0.4	O&F 1	0.90	0.89	0.90	0.89	0.89
		O&F 2	0.92	0.92	0.92	0.91	0.91
		O& F 3	0.89	0.89	0.89	0.88	0.88
		S&F	0.95	0.95	0.94	0.94	0.94
	0.8	O&F 1	0.89	0.88	0.87	0.86	0.88
		O&F 2	0.90	0.88	0.87	0.87	0.88
		O& F 3	0.91	0.90	0.89	0.88	0.89
		S&F	0.95	0.94	0.93	0.92	0.93
	1.0	O&F 1	0.90	0.88	0.86	0.85	0.88
		O&F 2	0.89	0.87	0.85	0.83	0.87
		O& F 3	0.91	0.89	0.87	0.86	0.89
		S&F	0.95	0.93	0.92	0.90	0.93
4 Regressors	0.4	O&F 1	0.88	0.88	0.88	0.88	0.87
		O&F 2	0.91	0.90	0.90	0.90	0.90
		O& F 3	0.70	0.72	0.70	0.69	0.66
		S&F	0.95	0.95	0.94	0.94	0.93
	0.8	O&F 1	0.90	0.88	0.87	0.86	0.87
		O&F 2	0.90	0.88	0.87	0.86	0.88
		O& F 3	0.81	0.81	0.79	0.78	0.77
		S&F	0.95	0.94	0.92	0.92	0.92
	1.0	O&F 1	0.91	0.89	0.88	0.86	0.88
		O&F 2	0.89	0.86	0.85	0.83	0.87
		O& F 3	0.84	0.83	0.80	0.78	0.78
		S&F	0.95	0.93	0.91	0.90	0.92
8 Regressors	0.4	O&F 1	0.77	0.76	0.76	0.76	0.75
		O&F 2	0.82	0.81	0.81	0.81	0.80
		O& F 3	0.61	0.60	0.60	0.60	0.60
		S&F	0.95	0.94	0.94	0.94	0.93
	0.8	O&F 1	0.82	0.80	0.80	0.79	0.79
		O&F 2	0.83	0.81	0.80	0.80	0.80
		O& F 3	0.72	0.71	0.70	0.69	0.69
		S&F	0.95	0.93	0.92	0.91	0.91
	1.0	O&F 1	0.84	0.82	0.81	0.80	0.80
		O&F 2	0.83	0.81	0.79	0.78	0.79
		O& F 3	0.75	0.73	0.72	0.71	0.71
		S&F	0.95	0.93	0.92	0.90	0.91

Table 7. Confidence Band Coverage for ρ_*^2 by Number of Regressors, Measurement Reliability & ρ^2.

	Reliability	Method	Population Squared Multiple Correlation $\rho^2 = 0.01$	$\rho^2 = 0.05$	$\rho^2 = 0.1$	$\rho^2 = 0.3$	$\rho^2 = 0.6$
2 Regressors	0.4	O&F 1	0.92	0.91	0.90	0.88	0.86
		O&F 2	0.95	0.94	0.93	0.90	0.87
		O& F 3	0.85	0.87	0.89	0.91	0.91
		S&F	0.95	0.95	0.95	0.94	0.94
	0.8	O&F 1	0.91	0.89	0.87	0.84	0.85
		O&F 2	0.94	0.91	0.89	0.84	0.82
		O& F 3	0.87	0.90	0.91	0.90	0.89
		S&F	0.95	0.94	0.94	0.93	0.91
	1.0	O&F 1	0.91	0.87	0.86	0.84	0.89
		O&F 2	0.93	0.90	0.87	0.82	0.79
		O& F 3	0.87	0.90	0.90	0.89	0.86
		S&F	0.94	0.94	0.93	0.92	0.89
4 Regressors	0.4	O&F 1	0.88	0.88	0.88	0.87	0.87
		O&F 2	0.92	0.92	0.91	0.89	0.87
		O& F 3	0.38	0.64	0.75	0.85	0.87
		S&F	0.95	0.95	0.95	0.94	0.93
	0.8	O&F 1	0.88	0.87	0.87	0.87	0.89
		O&F 2	0.92	0.90	0.89	0.85	0.83
		O& F 3	0.56	0.80	0.85	0.87	0.87
		S&F	0.95	0.94	0.94	0.92	0.90
	1.0	O&F 1	0.88	0.87	0.87	0.87	0.92
		O&F 2	0.92	0.89	0.88	0.83	0.80
		O& F 3	0.62	0.82	0.86	0.86	0.85
		S&F	0.94	0.94	0.93	0.91	0.89
8 Regressors	0.4	O&F 1	0.70	0.72	0.75	0.79	0.83
		O&F 2	0.80	0.80	0.81	0.82	0.82
		O& F 3	0.31	0.52	0.62	0.75	0.81
		S&F	0.95	0.95	0.94	0.93	0.92
	0.8	O&F 1	0.71	0.76	0.79	0.84	0.89
		O&F 2	0.80	0.81	0.82	0.82	0.80
		O& F 3	0.45	0.67	0.75	0.82	0.83
		S&F	0.95	0.94	0.94	0.92	0.89
	1.0	O&F 1	0.77	0.77	0.80	0.86	0.92
		O&F 2	0.81	0.81	0.81	0.81	0.78
		O& F 3	0.71	0.70	0.77	0.82	0.82
		S&F	0.94	0.94	0.93	0.91	0.88

Table 8. Confidence Band Coverage for ρ_*^2 by Number of Regressors, Measurement Reliability and Sample Size.

	Reliability	Method	Sample Size N=5*k	N=10*k	N=50*k
2 Regressors	0.4	O&F 1	0.79	0.94	0.95
		O&F 2	0.86	0.94	0.95
		O& F 3	0.82	0.89	0.94
		S&F	0.95	0.95	0.94
	0.8	O&F 1	0.78	0.91	0.94
		O&F 2	0.83	0.90	0.92
		O& F 3	0.85	0.91	0.93
		S&F	0.94	0.94	0.92
	1.0	O&F 1	0.78	0.90	0.94
		O&F 2	0.80	0.88	0.90
		O& F 3	0.84	0.89	0.92
		S&F	0.93	0.93	0.91
4 Regressors	0.4	O&F 1	0.75	0.92	0.96
		O&F 2	0.83	0.92	0.95
		O& F 3	0.55	0.68	0.85
		S&F	0.95	0.94	0.94
	0.8	O&F 1	0.77	0.91	0.95
		O&F 2	0.81	0.90	0.92
		O& F 3	0.69	0.79	0.89
		S&F	0.94	0.93	0.92
	1.0	O&F 1	0.78	0.91	0.95
		O&F 2	0.79	0.88	0.91
		O& F 3	0.71	0.80	0.89
		S&F	0.93	0.92	0.91
8 Regressors	0.4	O&F 1	0.54	0.81	0.93
		O&F 2	0.68	0.83	0.92
		O& F 3	0.43	0.58	0.79
		S&F	0.94	0.94	0.93
	0.8	O&F 1	0.63	0.84	0.93
		O&F 2	0.70	0.83	0.90
		O& F 3	0.57	0.70	0.85
		S&F	0.93	0.93	0.92
	1.0	O&F 1	0.65	0.85	0.94
		O&F 2	0.69	0.82	0.90
		O& F 3	0.60	0.72	0.85
		S&F	0.93	0.92	0.91

Table 9. Width of Confidence Band by Measurement Reliability.

	Reliability					
	Method	0.4	0.6	0.8	0.9	1.0
2 Regressors	O&F 1	0.32	0.36	0.39	0.40	0.41
	O&F 2	0.36	0.39	0.41	0.42	0.41
	O& F 3	0.73	0.69	0.66	0.64	0.62
	S&F	0.36	0.39	0.42	0.43	0.43
4 Regressors	O&F 1	0.29	0.32	0.34	0.35	0.35
	O&F 2	0.31	0.33	0.34	0.34	0.36
	O& F 3	0.32	0.33	0.34	0.34	0.33
	S&F	0.27	0.30	0.32	0.33	0.33
8 Regressors	O&F 1	0.24	0.25	0.27	0.26	0.27
	O&F 2	0.25	0.25	0.27	0.27	0.26
	O& F 3	0.25	0.26	0.26	0.27	0.26
	S&F	0.20	0.23	0.24	0.25	0.24

References

Algina, J. (1999). A comparison of methods for constructing confidence intervals for the squared multiple correlation coefficient. *Multivariate Behavioral Research*, *34*, 493-504.

Bradley, D. R. & Fleisher, C. L. (1994). Generating multivariate data from nonnormal distributions: Mihal and Barrett revisited. *Behavior Research Methods, Instruments, and Computers*, *26*, 156-166.

Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics*, *10*, 637 – 666.

Cochran, W. G. (1970). Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association*, *65*, 22 – 34.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ferron, J., Yi, Q. & Kromrey, J.D. (1997). NNCORR: A SAS/IML program for generating nonnormal correlated data. *Applied Psychological Measurement*, *21*, 64.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society*, *121*, 654-673.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

Fuller, W. A. & Hidirouglou, M.A. (1978). Regression estimation after correcting for attenuation. *Journal of the American Statistical Association*, *73*, 99-104

Harlow, L. L., Mulaik, S. A. & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

Jaccard, J & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348-357.

Jencks, C., and others (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

Kendall, M. G. & Stuart, A. (1977). *The advanced theory of statistics* (4th Ed., Vol. 2). London: Griffin.

Table 10. Width of Confidence Band by Number of Regressors, Reliability and Distribution Shape.

			Distribution Shape				
	Reliability	Method	Sk = 0.0 Kurt = 0.0	Sk = 1.0 Kurt = 3.0	Sk = 1.5 Kurt = 5.0	Sk = 2.0 Kurt = 6.0	Sk = 0.0 Kurt = 25.0
2 Regressors	0.4	O&F 1	0.32	0.32	0.32	0.32	0.32
		O&F 2	0.36	0.36	0.36	0.36	0.36
		O& F 3	0.72	0.73	0.72	0.73	0.73
		S&F	0.37	0.36	0.36	0.36	0.36
	0.8	O&F 1	0.39	0.39	0.39	0.38	0.39
		O&F 2	0.42	0.42	0.41	0.41	0.42
		O& F 3	0.66	0.66	0.66	0.66	0.66
		S&F	0.42	0.42	0.42	0.41	0.42
	1.0	O&F 1	0.41	0.41	0.41	0.40	0.41
		O&F 2	0.42	0.42	0.41	0.40	0.42
		O& F 3	0.62	0.62	0.62	0.62	0.62
		S&F	0.43	0.43	0.43	0.42	0.43
4 Regressors	0.4	O&F 1	0.29	0.29	0.29	0.29	0.28
		O&F 2	0.31	0.31	0.31	0.31	0.30
		O& F 3	0.32	0.32	0.32	0.32	0.31
		S&F	0.27	0.28	0.27	0.27	0.26
	0.8	O&F 1	0.34	0.34	0.34	0.34	0.33
		O&F 2	0.35	0.35	0.34	0.35	0.33
		O& F 3	0.35	0.35	0.34	0.34	0.33
		S&F	0.33	0.33	0.33	0.33	0.31
	1.0	O&F 1	0.35	0.35	0.35	0.35	0.34
		O&F 2	0.34	0.34	0.33	0.33	0.33
		O& F 3	0.34	0.33	0.33	0.33	0.33
		S&F	0.33	0.33	0.33	0.33	0.32
8 Regressors	0.4	O&F 1	0.24	0.24	0.24	0.24	0.24
		O&F 2	0.25	0.25	0.25	0.25	0.25
		O& F 3	0.25	0.25	0.25	0.25	0.25
		S&F	0.20	0.20	0.20	0.20	0.20
	0.8	O&F 1	0.27	0.27	0.27	0.27	0.27
		O&F 2	0.27	0.27	0.27	0.27	0.27
		O& F 3	0.26	0.26	0.26	0.26	0.26
		S&F	0.24	0.24	0.24	0.24	0.24
	1.0	O&F 1	0.27	0.27	0.27	0.27	0.27
		O&F 2	0.26	0.26	0.26	0.26	0.26
		O& F 3	0.26	0.26	0.26	0.25	0.26
		S&F	0.25	0.24	0.24	0.24	0.24

Kromrey, J. & Foster-Johnson, L. (1999, April). *Bias, Type I Error Control and Statistical Power in Multiple Regression: An Empirical Comparison of OLS and Errors-in-Variables Regression Algorithms.* Paper presented at the annual conference of the American Educational Research Association, New Orleans, LA.

Kromrey, J. & Hess, M. (2001, April). *Interval estimates of* R^2*: An empirical comparison of accuracy and precision under violations of the normality assumption.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Lee, Y. (1971). Some results on the sampling distribution of the multiple correlation coefficient. *Journal of the Royal Statistical Society*, *B*, *33*, 117-130.

Table 11. Width of Confidence Band by Number of Regressors, Measurement Reliability and ρ ^{2}.

	Reliability	Method	$\rho^2 = 0.01$	$\rho^2 = 0.05$	$\rho^2 = 0.1$	$\rho^2 = 0.3$	$\rho^2 = 0.6$
2 Regressors	0.4	O&F 1	0.27	0.29	0.30	0.34	0.39
		O&F 2	0.32	0.33	0.34	0.38	0.43
		O& F 3	0.78	0.77	0.75	0.69	0.64
		S&F	0.32	0.34	0.35	0.38	0.43
	0.8	O&F 1	0.28	0.32	0.35	0.46	0.53
		O&F 2	0.33	0.36	0.39	0.48	0.51
		O& F 3	0.77	0.72	0.67	0.60	0.53
		S&F	0.33	0.36	0.39	0.48	0.53
	1.0	O&F 1	0.29	0.33	0.39	0.51	0.52
		O&F 2	0.33	0.38	0.42	0.51	0.43
		O& F 3	0.77	0.69	0.64	0.56	0.44
		S&F	0.34	0.38	0.42	0.51	0.49
4 Regressors	0.4	O&F 1	0.24	0.25	0.27	0.31	0.35
		O&F 2	0.27	0.28	0.29	0.33	0.37
		O& F 3	0.29	0.29	0.30	0.33	0.36
		S&F	0.23	0.24	0.25	0.30	0.34
	0.8	O&F 1	0.25	0.28	0.32	0.40	0.44
		O&F 2	0.27	0.31	0.34	0.40	0.39
		O& F 3	0.29	0.31	0.34	0.38	0.38
		S&F	0.23	0.27	0.31	0.39	0.42
	1.0	O&F 1	0.25	0.30	0.33	0.43	0.43
		O&F 2	0.27	0.32	0.35	0.41	0.32
		O& F 3	0.29	0.32	0.35	0.39	0.32
		S&F	0.23	0.28	0.32	0.42	0.38
8 Regressors	0.4	O&F 1	0.20	0.21	0.22	0.26	0.29
		O&F 2	0.22	0.23	0.24	0.27	0.30
		O& F 3	0.23	0.23	0.24	0.26	0.29
		S&F	0.16	0.17	0.18	0.23	0.28
	0.8	O&F 1	0.21	0.23	0.26	0.31	0.34
		O&F 2	0.23	0.25	0.27	0.31	0.28
		O& F 3	0.23	0.25	0.26	0.29	0.28
		S&F	0.16	0.20	0.23	0.31	0.32
	1.0	O&F 1	0.24	0.24	0.27	0.33	0.32
		O&F 2	0.26	0.26	0.28	0.30	0.23
		O& F 3	0.25	0.25	0.27	0.30	0.23
		S&F	0.21	0.21	0.25	0.32	0.27

Column group header (over the ρ^2 columns): Population Squared Multiple Correlation

Maxwell, S. E., Delaney, H. D., & Dill, C. A. (1984). Another look at ANCOVA versus blocking. *Psychological Bulletin*, *95*, 136-147.

Olkin, I. & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin*, *188,* 155-164.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd Ed.). Fort Worth: Harcourt Brace.

Robey, R. R. & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, *45*, 283-288.

Steiger, J. H. & Fouladi, R. T. (1992). R2: A computer program for interval estimation, power calculation, and hypothesis testing for the squared multiple correlation. *Behavior Research, Methods, Instruments, and Computers*, *4*, 581-582.

Table 12. Width of Confidence Band by Number of Regressors, Reliability and Sample Size.

	Reliability	Method	Sample Size N=5*k	N=10*k	N=50*k
2 Regressors	0.4	O&F 1	0.49	0.34	0.13
		O&F 2	0.59	0.37	0.13
		O& F 3	0.91	0.86	0.41
		S&F	0.57	0.38	0.14
	0.8	O&F 1	0.53	0.43	0.21
		O&F 2	0.62	0.43	0.19
		O& F 3	0.86	0.77	0.35
		S&F	0.62	0.44	0.20
	1.0	O&F 1	0.53	0.45	0.23
		O&F 2	0.61	0.44	0.20
		O& F 3	0.81	0.73	0.32
		S&F	0.63	0.46	0.20
4 Regressors	0.4	O&F 1	0.45	0.30	0.11
		O&F 2	0.51	0.31	0.11
		O& F 3	0.50	0.33	0.11
		S&F	0.43	0.28	0.10
	0.8	O&F 1	0.49	0.37	0.16
		O&F 2	0.53	0.36	0.15
		O& F 3	0.51	0.36	0.15
		S&F	0.49	0.34	0.14
	1.0	O&F 1	0.49	0.38	0.18
		O&F 2	0.51	0.35	0.15
		O& F 3	0.50	0.36	0.15
		S&F	0.50	0.34	0.15
8 Regressors	0.4	O&F 1	0.37	0.25	0.09
		O&F 2	0.42	0.26	0.09
		O& F 3	0.40	0.26	0.09
		S&F	0.33	0.21	0.08
	0.8	O&F 1	0.40	0.29	0.12
		O&F 2	0.42	0.28	0.11
		O& F 3	0.40	0.27	0.11
		S&F	0.37	0.25	0.10
	1.0	O&F 1	0.40	0.30	0.13
		O&F 2	0.40	0.27	0.11
		O& F 3	0.39	0.27	0.11
		S&F	0.38	0.25	0.10

Steiger, J. H. & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In Harlow, L. L., Mulaik, S. A. & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum, p. 221-257.

Wilkinson & APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, p. 594-604.

Send correspondence to: Jeffrey D Kromrey, Ph.D.
University of South Florida
Email: kromrey@tempest.coedu.usf.edu