British Journal of Mathematical and Statistical Psychology (2018), 71, 167–185
© 2017 The British Psychological Society

The British Psychological Society

www.wileyonlinelibrary.com

Asymptotic confidence intervals for the Pearson correlation via skewness and kurtosis

Anthony J. Bishara[1]*, Jiexiang Li[2] and Thomas Nash[3]

[1]Department of Psychology, College of Charleston, South Carolina, USA
[2]Department of Mathematics, College of Charleston, South Carolina, USA
[3]Department of Computer Science, College of Charleston, South Carolina, USA

When bivariate normality is violated, the default confidence interval of the Pearson correlation can be inaccurate. Two new methods were developed based on the asymptotic sampling distribution of Fisher's z' under the general case where bivariate normality need not be assumed. In Monte Carlo simulations, the most successful of these methods relied on the (Vale & Maurelli, 1983, *Psychometrika*, 48, 465) family to approximate a distribution via the marginal skewness and kurtosis of the sample data. In Simulation 1, this method provided more accurate confidence intervals of the correlation in non-normal data, at least as compared to no adjustment of the Fisher z' interval, or to adjustment via the sample joint moments. In Simulation 2, this approximate distribution method performed favourably relative to common non-parametric bootstrap methods, but its performance was mixed relative to an observed imposed bootstrap and two other robust methods (PM1 and HC4). No method was completely satisfactory. An advantage of the approximate distribution method, though, is that it can be implemented even without access to raw data if sample skewness and kurtosis are reported, making the method particularly useful for meta-analysis. Supporting information includes R code.

1. Introduction

In psychological research, the vast majority of data sets show violations of normality (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013; Micceri, 1989). Such violations of normality can be problematic even for relatively simple statistics, such as the Pearson correlation coefficient, which is one of the more commonly reported statistics in the field (Austin, Scherbaum, & Mahlman, 2002; Goodwin & Goodwin, 1985). When data are not bivariate normal, the sampling distribution used for the Pearson correlation can fail to converge to that used under the assumption of bivariate normality (Duncan & Layard, 1973; Gayen, 1951; Hawkins, 1989). For this reason, non-normality can cause the default confidence interval to fail, producing a coverage rate that does not equal the one intended (e.g., Puth, Neuhäuser, & Ruxton, 2014). The purpose of the present research is to develop and evaluate two new methods for adjusting the confidence interval of the correlation, with the adjustments based on the asymptotic sampling distribution under the general case where bivariate normality (BVN) need not be assumed.

*Correspondence should be addressed to Anthony J. Bishara, Department of Psychology, College of Charleston, 66 George St., Charleston, SC 29424, USA (email: BisharaA@cofc.edu).

DOI:10.1111/bmsp.12113

Construction of the confidence interval for ρ usually relies on the sampling distribution of Fisher's z'. The asymptotic variance of this statistic for the general case, without assuming BVN, has been known for some time (Duncan & Layard, 1973; Hawkins, 1989). However, it has not yet been taken advantage of for the construction of confidence intervals for ρ. We develop two methods, labelled throughout as *adjustment via sample joint moments* and *adjustment via approximate distribution*. The second method, adjustment via approximate distribution, requires only the marginal sample skewness and kurtosis, along with the frequently reported sample correlation and sample size, to construct the confidence interval. Because of these minimal requirements, adjustment via approximate distribution may be useful when reanalysing previously published research, for example, through meta-analysis. Additionally, it will be shown that adjustment via approximate distribution can provide more accurate confidence intervals than no adjustment and the first adjustment method.

2. Confidence interval construction under bivariate normality

The Pearson product moment correlation is defined as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{1/2}}. \tag{1}$$

Because the sampling distribution of r is not necessarily normal, particularly when $\rho \neq 0$, Fisher's (1915, 1921) r-to-z' transformation is typically used in the construction of the confidence interval:

$$z' = 0.5 * \ln\left(\frac{1+r}{1-r}\right). \tag{2}$$

The sampling distribution of z' is approximately normal with a standard error of approximately

$$\sigma_{z'} = (n-3)^{-1/2}. \tag{3}$$

Thus, the 95% confidence interval for $0.5 * \ln\left(\frac{1+\rho}{1-\rho}\right)$ is defined as

$$z' \pm 1.96 \cdot \sigma_{z'}. \tag{4}$$

These confidence interval bounds can then be transformed back to the scale of r using

$$r = \frac{\exp(2z') - 1}{\exp(2z') + 1}. \tag{5}$$

The resulting confidence interval of the correlation is an approximation, but it is quite accurate when BVN is satisfied (Fouladi & Steiger, 2008).

3. Confidence interval construction under non-normality

Under the more general case where BVN might not be satisfied, Hawkins (1989) noted that the asymptotic variance of z' must be multiplied by τ_f^2:

$$\tau_f^2 = \frac{(\mu_{40} + 2\mu_{22} + \mu_{04})\rho^2 - 4(\mu_{31} + \mu_{13})\rho + 4\mu_{22}}{4(1-\rho^2)^2}, \tag{6}$$

where μ_{jk} represents a population joint moment, defined as

$$\mu_{jk} = \mathrm{E}\left[X^j Y^k\right]. \tag{7}$$

Without loss of generality, X and Y are assumed to be standardized ($\mu_{10} = \mu_{01} = 0$, $\mu_{20} = \mu_{02} = 1$). The standard error of z' can then be approximated as $\tilde{\sigma}_{z'}$:

$$\tilde{\sigma}_{z'} = \tau_f (n-3)^{-1/2}. \tag{8}$$

Note that τ_f simplifies to 1 when data are bivariate normal (Hawkins, 1989), so equation (8) simplifies to equation (3) in that case. Additionally, τ_f simplifies to 1 if X and Y are independent because in that case $\rho = 0$ and $\mu_{22} = 1$. In other words, this asymptotic variance term becomes relevant only if data violate BVN and X and Y are dependent.

Unfortunately, this term is a function of the fourth joint population moments, which are typically unknown. Perhaps because of this difficulty, there is not a single example in the literature of equation (6) being used to adjust the confidence interval of the correlation. Here, we develop and evaluate two methods of estimating τ_f^2 based solely on sample information.

3.1. Adjustment via sample joint moments

One simple approach to estimating τ_f^2 is to estimate population joint moments by using the corresponding sample moments, m_{jk}:

$$\hat{\mu}_{jk} = m_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^j y_i^k\right). \tag{9}$$

Likewise, the sample correlation can be used to estimate ρ:

$$\hat{\rho} = r. \tag{10}$$

Then, equations (9) and (10) can be used in equation (6) to estimate τ_f^2 and $\tilde{\sigma}_{z'}$, which can then be used to construct the confidence interval for the correlation, as outlined earlier (see equations 1–5).

This approach has the advantage of being simple and easy to use. However, unless the sample size is extremely large, the higher-order sample joint moments may be unstable estimators of their population counterparts. Thus, in typical samples of psychological data, this estimate of τ_f^2 may be inaccurate, leading to inaccurate confidence intervals. An additional limitation is that the sample joint moments are almost never analysed, let alone

reported in published work, and so adjustment by this method may be impossible without access to the raw data.

3.2. Adjustment via approximate distribution by skewness and kurtosis

Instead of using sample joint moments, it may be more advantageous to estimate an approximate distribution that the sample appears to be drawn from, and then analytically solve for τ_f^2 based on that distribution's parameters. To do so, a bivariate distribution family must be chosen. For generality, that family must be flexible enough to account for a wide variety of observed distribution shapes. However, it must also be as simple as possible, with a small number of free parameters so that those parameters can be accurately estimated from sample data. To satisfy these constraints, a third-order polynomial family was used (Vale & Maurelli, 1983; see also Fleishman, 1978; Headrick & Sawilowsky, 1999). The third-order polynomial also has the advantage of allowing estimation of distribution parameters using marginal skewness and kurtosis estimates.

In the third-order polynomial family, each variable is a nonlinear combination of standard normal variables, Z_1 and Z_*:

$$X = a_1 + b_1 Z_1 + c_1 Z_1^2 + d_1 Z_1^3, \tag{11}$$

$$Y = a_2 + b_2 Z_* + c_2 Z_*^2 + d_2 Z_*^3. \tag{12}$$

The constants (a_i, b_i, c_i, and d_i) control the particular shape of each marginal. The bivariate standard normal case occurs when $b_i = 1$ and $a_i = c_i = d_i = 0$ for $i = 1, 2$. In order to allow X and Y to be correlated, Z_* is defined as a linear combination of independent standard normal deviates Z_1 and Z_2,

$$Z_* = tZ_1 + \sqrt{1 - t^2} Z_2, \tag{13}$$

where t is the intermediate correlation, whose value may differ somewhat from the original population correlation. The connection between these two correlations will be established later.

The constants in the third-order polynomial family are determined by skewness (γ_1) and kurtosis (γ_2) of X and Y separately, and of ρ. For standardized variables,

$$\gamma_1 = \mu_3, \tag{14}$$

$$\gamma_2 = \mu_4 - 3, \tag{15}$$

where μ_j for X is $E[X^j]$, and for Y is $E[Y^j]$. For the present purposes, skewness and kurtosis can be estimated using the marginal sample skewness (g_1) and kurtosis (g_2) statistics. Specifically, in standardized data:

$$\hat{\gamma}_1 = g_1 = m_3, \tag{16}$$

$$\hat{\gamma}_2 = g_2 = m_4 - 3, \tag{17}$$

where m_j for X is $\frac{1}{n}\sum_{i=1}^{n}\left(x_i^j\right)$ and for Y is $\frac{1}{n}\sum_{i=1}^{n}\left(y_i^j\right)$. Additionally, ρ can be estimated as r, just as in the first adjustment method.

Estimating constants separately for X and Y, let $a = -c$ (which must be the case for $\mu = 0$). Then, the remaining constants can be derived by substituting the skewness and kurtosis values estimated from equations (16) and (17) (for details, see Fleishman, 1978) and solving for the second, third, and fourth moments of equation (11). Specifically, b, c, and d can be estimated by $\hat{b}$, $\hat{c}$ and $\hat{d}$ from the following equations:

$$\hat{b}^2 + 6\hat{b}\hat{d} + 2\hat{c}^2 + 15\hat{d}^2 - 1 = 0, \tag{18}$$

$$2\hat{c}\left(\hat{b}^2 + 24\hat{b}\hat{d} + 105\hat{d}^2 + 2\right) - g_1 = 0, \tag{19}$$

$$24\left[\hat{b}\hat{d} + \hat{c}^2\left(1 + \hat{b}^2 + 28\hat{b}\hat{d}\right) + \hat{d}^2\left(12 + 48\hat{b}\hat{d} + 141\hat{c}^2 + 225\hat{d}^2\right)\right] - g_2 = 0. \tag{20}$$

Then the estimate of the intermediate correlation, $\hat{t}$, can be found by solving

$$\hat{t}\left(\hat{b}_1\hat{b}_2 + 3\hat{b}_1\hat{d}_2 + 3\hat{d}_1\hat{b}_2 + 9\hat{d}_1\hat{d}_2\right) + \hat{t}^2(2\hat{c}_1\hat{c}_2) + \hat{t}^3\left(6\hat{d}_1\hat{d}_2\right) - r = 0, \tag{21}$$

where subscripts 1 and 2 correspond to X and Y, respectively. This equation was arrived at by setting $\rho = \mathrm{E}[XY]$ and using r as an estimate of ρ (see Vale & Maurelli, 1983).

Having estimated the parameters of the third-order polynomial family, an estimate of the population joint moments can then be found analytically. First, substituting equation (13) into equation (11) and (12):

$$\begin{aligned}
\hat{\mu}^*_{jk} &= \mathrm{E}\left[\widehat{X^j Y^k}\right] \\
&= \mathrm{E}\left[\left(\hat{a}_1 + \hat{b}_1 Z_1 + \hat{c}_1 Z_1^2 + \hat{d}_1 Z_1^3\right)^j \left(\hat{a}_2 + \hat{b}_2 Z_* + \hat{c}_2 Z_*^2 + \hat{d}_2 Z_*^3\right)^k\right] \\
&= \mathrm{E}\left[\left(\hat{a}_1 + \hat{b}_1 Z_1 + \hat{c}_1 Z_1^2 + \hat{d}_1 Z_1^3\right)^j \left(\hat{a}_2 + \hat{b}_2\left(\hat{t}Z_1 + \sqrt{1-\hat{t}^2}Z_2\right)\right.\right. \\
&\left.\left. + \hat{c}_2\left(\hat{t}Z_1 + \sqrt{1-\hat{t}^2}Z_2\right)^2 + \hat{d}_2\left(\hat{t}Z_1 + \sqrt{1-\hat{t}^2}Z_2\right)^3\right)^k\right].
\end{aligned} \tag{22}$$

Then the solutions to the specific terms in τ_f^2 ($\hat{\mu}^*_{40}$, $\hat{\mu}^*_{04}$, $\hat{\mu}^*_{22}$, $\hat{\mu}^*_{31}$ and $\hat{\mu}^*_{13}$) can be found by setting j and k appropriately. The appendix provides the exact solutions to these terms.

4. Simulation 1

A Monte Carlo simulation was conducted to evaluate the accuracy of confidence intervals under four different methods: unadjusted; adjustment via sample joint moments; adjustment via approximate distribution;, and ideal adjustment. The last method involved adjusting confidence intervals by using population information to solve τ_f^2. Although the population is typically unknown in actual data analysis, this last method was included as an ideal comparison condition, as a benchmark that any accurate method should approach. In each simulation, each of the four methods was used to construct a 95% confidence interval for the correlation. The primary dependent measure was the coverage rate, that is, the proportion of simulations in which the confidence interval of ρ covered the true value of ρ. An additional dependent measure was confidence interval length, which should approach that of the ideal adjustment as sample size increases.

4.1. Method

4.1.1. Design

A factorial design was used for simulation scenarios. To examine the impact of effect size, three population correlation coefficients were considered: $\rho = 0$, .25, and .50. To examine the effect of various types of non-normality, five population distribution shapes for Y were selected, represented by skewness and kurtosis values (γ_1, γ_2): normal (0, 0), negative kurtosis (0, –1), extreme kurtosis (0, 40), moderate skew and kurtosis (2, 8), and extreme skew and kurtosis (4, 40). Extreme values of skewness and kurtosis were chosen to be slightly beyond the range often seen in psychological studies (see Blanca *et al.*, 2013; Micceri, 1989). There were two distribution combinations: either X was normally distributed or X had the same distribution as Y. Finally, five sample sizes were used: n = 10, 40, 160, 640, and 2,560. The wide range for n was chosen so that the effects of using large-sample theory (Hawkins, 1989) would likely be noticeable. This design (three effect sizes × 5 distribution shapes × 2 distribution combinations × 5 sample sizes) resulted in 150 scenarios.

4.1.2. Confidence interval construction

The unadjusted and adjusted via sample joint moments methods were described earlier.

For adjustment via approximate distribution, seven parameters needed to be estimated from each simulated data set: $\hat{b}_1$, $\hat{c}_1$, $\hat{d}_1$, $\hat{b}_2$, $\hat{c}_2$, $\hat{d}_2$, $\hat{t}$. The first six parameters were estimated through a simplex algorithm. The last parameter, t, was estimated separately in a one-parameter search algorithm that combined successive parabolic interpolation and golden section search.

Specifically, starting parameters values were entered into the left-hand sides of equations (18–20) for the observed X data. Next, the squared results of each expression were summed together, and the sum of squares was minimized via a simplex optimization algorithm (Nelder & Mead, 1965) using the optim() function from the R programming language (R Core Team, 2014). If the resulting parameters led to a sum of squares greater than the tolerance of 0.0001, then a random set of starting parameter values was generated, and the process was repeated again, up to five times if needed. Some combinations of g_1, g_2, and r are not solvable in the polynomial families (see Headrick, 2002, 2010), which is often the case when these statistics are far from 0. Because of this

issue, if five random starting parameter sets failed to achieve the tolerance threshold, then g_1 and g_2 were adjusted towards 0 by subtracting 1% of their original values. This process (five random starting parameter vectors and then an adjustment of g_1 and g_2) was repeated until the tolerance threshold was achieved. The same approach was taken to estimate parameters for Y. To solve for $\hat{t}$, a one-parameter search was conducted via the optimize() function in R, attempting to minimize the expression on the left-hand side of equation (21). If the squared result exceeded the tolerance threshold, r was adjusted towards 0 by subtracting 1% of its original value, and the process was repeated until the tolerance threshold was reached. Adjusting g_1, g_2, and r towards 0 will typically lead to a more conservative estimate of τ_f^2 (closer to 1). In other words, it should result in a confidence interval length more similar to that of the unadjusted method.

Note that bias adjustments for ρ (e.g., Fisher, 1915; Olkin & Pratt, 1958) were avoided because such adjustments were derived under the assumption of BVN. When that assumption is wrong, such adjustments can sometimes increase bias (Bishara & Hittner, 2015).

The last method, ideal adjustment, relied on population information to solve τ_f^2. In particular, ρ was treated as known and used in equation (6). Relevant values of μ_{jk} were not easily solvable in the generating algorithm, so close approximations were achieved by generating a pseudo-population consisting of 5 million observations.

4.1.3. Simulation

In each scenario, there were 10,000 simulations. With this number of simulations, the simulation margin of error for coverage probability was less than ± 0.01 in each scenario.

Data were generated by a fifth-order polynomial (Headrick, 2002), described in detail in the supporting information. This generating method was chosen because it is more precise than the third-order polynomial in terms of generating intended values of skewness and kurtosis (Olvera Astivia & Zumbo, 2015). Additionally, generating from a slightly different family than the third-order polynomial provides a modest stress test for the adjustment via approximate distribution method used here.

Simulations were conducted in R (R Core Team, 2014), with different scenarios distributed across different cores of a high-performance computing cluster. The supporting information contains two sections of R code: a short section for quickly applying the two adjustment methods developed here to new data and a longer section for simulation replication purposes.

4.2. Results and discussion

The commonly used unadjusted method was accurate in scenarios where BVN was satisfied, producing 95% confidence intervals with observed coverage probability of approximately .95 (results ranged from .946 to .954 across scenarios). However, in scenarios where BVN was violated, unadjusted coverage probability was often lower (ranging from .609 to .956). The scenario that produced the lowest coverage had the largest n (2,560), the largest population correlation (.5), and extreme skew and kurtosis in both variables. Adjustment via sample joint moments fared only slightly better for scenarios that violated BVN (.677 to .953). Furthermore, the sample joint moments method sometimes failed to achieve .95 coverage even when BVN was satisfied (.857 to .954). This may have resulted from high variance in the sample joint moment estimates, particularly when n was small. In contrast, the coverage probability for adjustment via

approximate distribution was approximately .95 when BVN was satisfied (.947 to .967), and never fell below .88 when it was not satisfied (.883 to .975). As a basis for comparison, the coverage probability for ideal adjustment ranged from .916 to .983.

As shown in Table 1, as n increased, the unadjusted method's coverage failed to converge to .95. This likely occurred because equation (3) fails to converge to equation (8) as n approaches infinity if $\tau_f^2 \neq 1$, which is often the case when BVN is violated. In contrast, all adjustment methods approached .95 coverage as n increased. Adjustment via approximate distribution averaged approximately .95 coverage even with the smallest sample size. As shown in Table 2, adjustment via approximate distribution also produced confidence interval lengths that were most similar to those of the ideal adjustment. Other methods sometimes had shorter confidence interval lengths, but those shorter lengths occurred with less than .95 coverage.

The effects of population shape and correlation can be seen in Table 3. As shown there, the unadjusted method produced low coverage mainly when both X and Y were non-normal (bottom half). If only Y was non-normal (top half), unadjusted coverage fell slightly below .95 only for $\rho = .5$, $\gamma_2 = 40$ scenarios. The coverage of the adjustment via approximate distribution method was always closer to .95 than that of adjustment via sample joint moments. The ideal adjustment produced clear overcoverage in some situations, which may be due to the presence of small-n scenarios that would cause an asymptotic adjustment to be inaccurate.

5. Simulation 2

Simulation 1 suggested that the adjustment via approximate distribution method could improve confidence interval performance relative to no adjustment or to adjustment via the sample joint moments. However, it is unclear whether the

Table 1. Mean coverage probability as a function of sample size in Simulation 1

n	Unadjusted	Adj. via sample joint moments	Adj. via approx. distribution	Ideal adj.
10	.933	.828	.958	.953
40	.929	.891	.950	.954
160	.921	.922	.948	.954
640	.915	.938	.950	.952
2,560	.910	.945	.951	.951

Note. Adj. = adjusted; approx. = approximate.

Table 2. Mean confidence interval length as a function of sample size in Simulation 1

n	Unadjusted	Adj. via sample joint moments	Adj. via approx. distribution	Ideal adj.
10	1.070	0.812	1.123	1.172
40	0.547	0.490	0.586	0.618
160	0.276	0.276	0.302	0.315
640	0.139	0.148	0.156	0.158
2,560	0.069	0.077	0.080	0.080

Note. Adj.=adjusted, approx.=approximate.

Table 3. Coverage probability as a function of distribution shape and population correlation in Simulation 1

X distrib.	Pop. corr. (ρ)	*Y* skewness (γ_1)	*Y* kurtosis (γ_2)	Unadjusted	Adj. via sample joint moments	Adj. via approximate distribution	Ideal adj.
Normal	0	0	−1	.949	.926	.951	.949
		0	0	.951	.926	.953	.951
		0	40	.950	.897	.955	.949
		2	8	.951	.921	.956	.951
		4	40	.950	.902	.958	.950
	.25	0	−1	.951	.929	.954	.951
		0	0	.950	.927	.953	.950
		0	40	.950	.901	.958	.955
		2	8	.952	.920	.956	.951
		4	40	.949	.900	.961	.954
	.50	0	−1	.953	.928	.959	.949
		0	0	.950	.925	.953	.950
		0	40	.939	.901	.959	.970
		2	8	.952	.916	.956	.948
		4	40	.939	.898	.961	.970
Same as *Y*	0	0	−1	.948	.931	.951	.948
		0	0	.949	.924	.952	.949
		0	40	.937	.912	.955	.937
		2	8	.949	.909	.956	.949
		4	40	.942	.885	.961	.943
	.25	0	−1	.948	.931	.954	.949
		0	0	.950	.926	.952	.950
		0	40	.856	.845	.940	.960
		2	8	.915	.904	.952	.954
		4	40	.823	.839	.941	.960
	.50	0	−1	.946	.931	.955	.951
		0	0	.950	.926	.953	.950
		0	40	.724	.835	.912	.968
		2	8	.879	.899	.949	.958
		4	40	.700	.832	.919	.966

Note. Distrib. = distribution; pop. = population; corr. = correlation; unadj. = unadjusted; adj. = adjusted.

approximate distribution method can improve upon pre-existing methods intended to address non-normality. In Simulation 2, six additional methods were considered: the nonparametric percentile bootstrap, the nonparametric bootstrap with bias correction and acceleration (BCa), the observed imposed (OI) bootstrap, the OI bootstrap with BCa, PM1, and HC4. The two nonparametric bootstrap methods have a long history of use for correlations (Efron, 1988; Lunneborg, 1985; Rasmussen, 1987; Strube, 1988), and so any newly proposed method should at the very least be able to improve upon them. The OI bootstrap (Beasley, DeShea, Toothaker, Mendoza, Bard, & Rodgers, 2007) involves resampling from n^2 rather than n pairs of

observations, leading to a somewhat smoother resampling distribution. The OI bootstrap with BCa is often more accurate than traditional non-parametric bootstrap procedures for non-normal data (Bishara & Hittner, 2017). Estimators intended to address heteroscedasticity sometimes also alleviate problems stemming from normality violations (Wilcox, 2012), so we considered two such estimators: PM1 (Wilcox, 1996, 2012) and HC4 (Cribari-Neto, 2004).

5.1. Method

5.1.1. Nonparametric bootstrap with percentile interval

For this method, n pairs of X and Y were sampled with replacement, and then the bootstrap correlation, r_{b}^{*}, was recorded. This procedure was repeated $B = 10{,}000$ times per simulation. Let the ascending rank-ordered r_{b}^{*} values be $r_{(1),}^{*}\ r_{(2)}^{*},\ \ldots,\ r_{(10{,}000)}^{*}$. The 95% percentile interval was defined as $\left(r_{(.025\cdot B)}^{*},\ r_{(.975\cdot B)}^{*}\right)$.

5.1.2. Nonparametric bootstrap with BCa

BCa intervals were defined as $\left(r_{(\alpha_1\cdot B)}^{*},\ r_{(\alpha_2\cdot B)}^{*}\right)$. Specifically,

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(.025)}{1 - \hat{a}\left(\hat{z}_0 + \Phi^{-1}(.025)\right)}\right), \tag{23}$$

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(.975)}{1 - \hat{a}\left(\hat{z}_0 + \Phi^{-1}(.975)\right)}\right), \tag{24}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The bias correction term, $\hat{z}_0$, is a function of the proportion of $r_{\mathrm{b}}^{*} < r$,

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{r_{\mathrm{b}}^{*} < r\}}{B}\right). \tag{25}$$

The acceleration term, $\hat{a}$, is a function of the jackknife estimates of r. Let r_{-j}^{*} be the jth jackknife correlation, that is, the correlation computed without the jth pair (x_j, y_j). Then the estimate of acceleration is

$$\hat{a} = \frac{\sum_{j=1}^{n}\left(\bar{r}_{-}^{*} - r_{-j}^{*}\right)^3}{6\{\sum_{j=1}^{n}\left(\bar{r}_{-}^{*} - r_{-j}^{*}\right)^2\}^{\ 3/2}}, \tag{26}$$

where $\bar{r}_{-}^{*}$ is the mean jackknife correlation (see Efron & Tibshirani, 1993).

5.1.3. OI bootstrap with percentile intervals

A preliminary sampling frame was constructed with all n^2 combinations of observed x and y: $\{(x_1, y_1), (x_1, y_2), \ldots (x_2, y_1), \ldots (x_{\mathrm{n}}, y_{\mathrm{n}})\}$. Let x' and y' be standardized vectors of all

possible pairs. The correlation between x' and y' will necessarily be 0. To reimpose the observed correlation (r), the final sampling frame was x' and y'', where

$$y'' = r \cdot x' + (\sqrt{1 - r^2}) \cdot y'. \tag{27}$$

Then n pairs of (x'_j, y''_j) were sampled with replacement. Then the construction of the percentile confidence interval proceeded as with the non-parametric bootstrap.

5.1.4. OI bootstrap with BCa

Following Beasley *et al.* (2007), the jackknife correlations were calculated for all n^2 pairs in the final sampling frame. All other details followed (equations 23–26), except with r_b^* calculated from the final sampling frame from the OI method.

5.1.5. PM1

Sometimes called the modified percentile bootstrap method (Wilcox, 2012), PM1 is identical to the nonparametric bootstrap with percentile interval method described earlier, except that the percentiles are modified based on the sample size. Specifically, for $B = 599$, the 95% confidence interval is $(r^*_{(a)}, r^*_{(c)})$, where

$$(a, c) = \begin{cases} (7, 593), & \text{if } n < 40, \\ (8, 592), & \text{if } 40 \leq n < 80, \\ (11, 588), & \text{if } 80 \leq n < 180, \\ (14, 585), & \text{if } 180 \leq n < 250, \\ (15, 584), & \text{if } 250 \leq n. \end{cases} \tag{28}$$

Because $B = 10{,}000$ here, we used the rounded value of (a, c)•10,000/599 (see also the PERAD method in Wilcox, 1996).

5.1.6. HC4

First consider the linear regression model $y = X\beta + \varepsilon$ in the usual matrix notation comprising all n observations, with dependent variable y, regressor x with coefficient vector β and error term ε. It is typically assumed that the errors have zero mean and variance $\text{var}[\varepsilon] = \Omega$. Under suitable regularity conditions the coefficients β can be consistently estimated by ordinary least squares, giving the well-known OLS estimator:

$$\hat{\beta} = (X'X)^{-1}X'y, \tag{29}$$

$$\hat{e} = (I - H)y = (I - X(X'X)^{-1}X')y, \tag{30}$$

where I is the n-dimensional identity matrix and H is usually called the hat matrix. The covariance matrix Ψ is usually written as

$$\Psi = \mathrm{var}[\hat{\beta}] = (X'X)^{-1}X'\Omega X(X'X)^{-1}. \tag{31}$$

In the classical linear model independent and homoscedastic errors with variance σ^2 are assumed, yielding $\Omega = \sigma^2 I$ and $\Psi = \sigma^2(X'X)^{-1}$, which can be consistently estimated by plugging in the usual OLS estimator, $\hat{\sigma}^2 = \frac{1}{n-2}\sum \hat{\varepsilon}_i^2$. But if independence and/or homoscedasticity are violated, inference based on this estimator will be biased. Heteroscedasticity-consistent (HC) estimators address this problem using estimates for Ω which are consistent in the presence of heteroscedasticity (Cribari-Neto, 2004). Specifically, HC4 estimators are constructed by plugging an estimate $\hat{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_n)$ into equation (31) with $\omega_i = \frac{\hat{\varepsilon}_i^2}{(1-h_i)^{\delta_i}}$, where $h_i = H_{ii}$ are the diagonal elements of the hat matrix, $\bar{h}$ is their mean and $\delta_i = \min\{4, h_i/\bar{h}\}$. Among the HC estimators, HC4 provides the best performance in small samples, especially in the presence of influential observations (Zeileis, 2004). To construct confidence intervals of ρ, x and y were standardized. The interval then consisted of $\hat{\beta}_1 \pm t_{(\alpha/2,n-2)}\sqrt{\mathrm{var}[\hat{\beta}_1]}$, with $\mathrm{var}[\hat{\beta}_1]$ calculated using $\hat{\Omega}$ as described above (see Wilcox, 2012).[1] The length of this interval can exceed 2, especially in small samples. To avoid penalizing the HC4 method, confidence interval lengths >2 were treated as 2 when calculating mean length.

5.1.7. Additional methods

We also considered several hybrid methods involving a combination of bootstrapping and the asymptotic variance of z' in the general case (equations 6–8). Only one of these methods improved upon the original: coverage of adjustment via sample joint moments was improved by using bootstrap estimates of the joint moments rather than the observed sample joint moments. Unfortunately, even in that case, observed coverage was still less accurate than that of the adjustment via approximate distribution. Because of such results, the hybrid methods are not reported further here.

5.1.8. Simulation

Bootstrapping methods required more CPU time and memory than did other methods. This problem necessitated design simplifications. Simulation 2 excluded the middle effect size (ρ = .25). The maximum n was set to 1,280 rather than 2,560. Each scenario had 2,000 rather than 10,000 simulations, leading to a simulation margin of error for coverage probability of less than ± 0.022 in each scenario. All other details followed those of Simulation 1.

5.2. Results and discussion

Figure 1 shows a summary of the coverage probability results. As shown in the first four columns, the pattern from Simulation 1 was generally replicated. Adjustment via approximate distribution produced observed coverage rates close to .95, and close to the ideal adjustment, an adjustment made assuming perfect knowledge of the population's distributional form. Among the bootstrap methods in the next four columns of Figure 1, the OI bootstrap with BCa tended to produce observed coverage closest .95. As shown in

[1] R code adapted from Rallfun-v30.txt, see http://dornsife.usc.edu/labs/rwilcox/software.

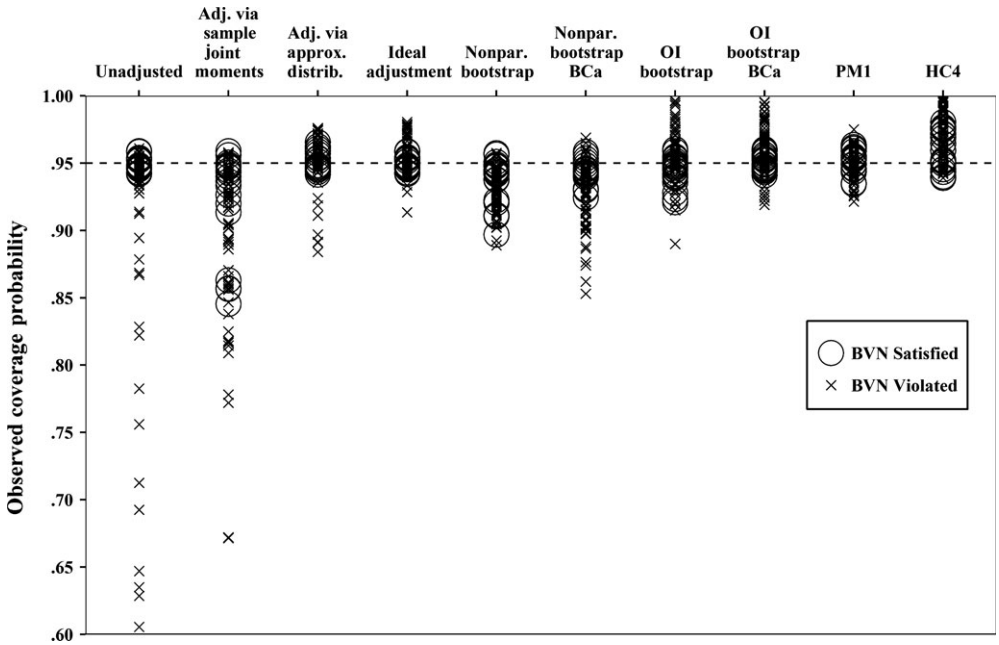

Figure 1. In Simulation 2, observed coverage probabilities are shown with circles for each scenario where bivariate normality was satisfied, and with crosses for each scenario in which it was violated. Adj. = adjusted; approx. distrib. = approximate distribution; nonpar. = nonparametric; BCa = bias correction and acceleration; OI = observed imposed; PM = percentile modified; HC = heteroscedasticity consistent; BVN = bivariate normality.

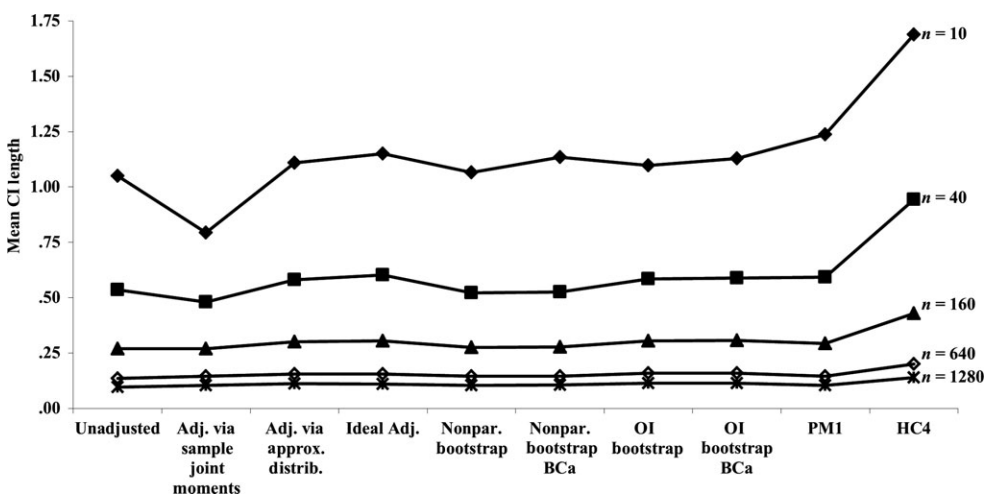

Figure 2. In Simulation 2, mean confidence interval (CI) lengths (upper bound – lower bound). Adj. = adjustment; approx. distrib. = approximate distribution; nonpar. = nonparametric; BCa = bias correction and acceleration; OI = observed imposed; PM = percentile modified; HC = heteroscedasticity consistent.

the next column, the PM1 method tended to get closest to .95 coverage in the largest number of scenarios. Finally, HC4 typically produced at least .95 coverage, but the coverage also noticeably exceeded .95 in several scenarios, suggesting that the HC4 method may have led to confidence intervals that were unnecessarily long. As shown in Figure 2, the mean confidence interval length was indeed larger for the HC4 method as compared to all other methods, including the ideal adjustment method. Thus, HC4 involved a trade-off: the benefit of higher coverage probability was accompanied by the cost of longer, less precise intervals. PM1 also had relatively long confidence intervals, but only for the $n = 10$ scenarios, and not to the extreme degree that HC4 did.

Overall, the results suggested that the approximate distribution, OI bootstrap with BCa, PM1, and HC4 methods were helpful in addressing non-normality. Figure 3 shows the coverage probabilities of these four methods in more detail. As shown there, the approximate distribution method had the advantage of generally converging towards .95 coverage as n increased, but the disadvantage of sometimes converging from below, and thus producing undercoverage with small ns. The OI bootstrap with BCa and HC4 methods had the advantage of generally producing .95 or above coverage, but the disadvantage of not necessarily converging towards .95 as n increased, or at least not with the sample sizes examined here. PM1 was generally closer to .95 coverage than the other methods, though it tended to be slightly lower than .95 in the $n = 640$ scenarios. This pattern suggests that PM1 might be further improved by modifying equation (28).

6. GENERAL DISCUSSION

Violations of BVN can lead to extremely inaccurate confidence intervals when the default method (unadjusted) is used. We evaluated two new methods to adjust the confidence

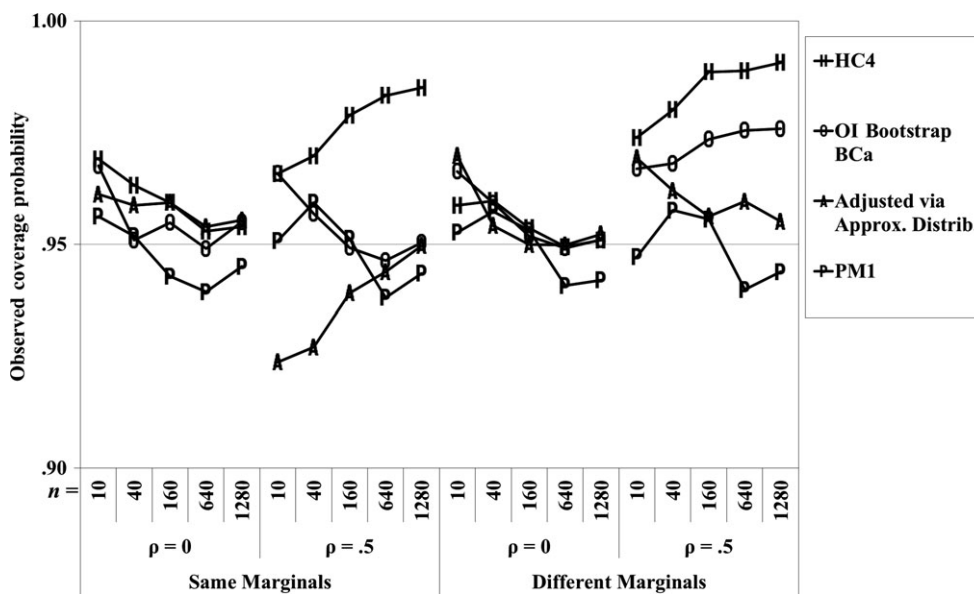

Figure 3. In Simulation 2, mean observed coverage probability of the four most promising methods. Mean coverage is shown as a function of n, ρ, and whether the marginal distributions (of X and Y) had the same form. HC=heteroscedasticity consistent, OI = observed imposed; BCa = bias correction and acceleration; approx. distrib. = approximate distribution; PM = percentile modified.

interval based only on sample information. As shown in Simulations 1 and 2, coverage was somewhat improved by using the simple method of substituting sample joint moments for their population counterparts within the Hawkins (1989) equation. Coverage was more improved, though, by the approximate distribution method, where a distributional form was approximated and then population joint moments could be analytically derived via that form's parameters. Importantly, this approximate distribution method was accurate even when BVN was satisfied, which suggests that overuse of this method would do little harm.

Adjustments to confidence intervals were most needed when ρ was large and both variables were non-normal. Conversely, if data were BVN, or X and Y were independent (with $\rho = 0$), then the unadjusted method produced accurate coverage. The reason for this pattern is that either BVN or independence is sufficient for τ_f to simplify to 1, which causes the default r-to-z' confidence interval to be accurate, or at least in large samples. The implication of this pattern is that, in a hypothesis testing setting, the Type I error probability will converge to alpha even when BVN is violated. The convergence rate may be quite slow, as inflated Type I error rates due to non-normality can occur with sample sizes as large as 160 (Bishara & Hittner, 2012). More generally, this pattern highlights the idea that large samples will only mitigate the effect of non-normality under particular circumstances.

Use of the approximate distribution method requires only a few pieces of information: the observed correlation, the sample size, and skewness and kurtosis for each marginal. More regular reporting of marginal skewness and kurtosis values would allow this adjustment method to be used as part of meta-analysis because access to the raw data would not be necessary. In the present work, skewness and kurtosis were examined via g_1 and g_2, but other estimates (e.g., G_1, b_1) would have similar values (see Joanes & Gill, 1998), or at least so long as n was moderately large. Regardless of the particular estimates chosen, the literature would be better served if at least some estimate of skewness and kurtosis were more regularly reported (see also Hopkins & Weeks, 1990), especially given the difficulty of obtaining raw data from published articles (Wicherts, Bakker, & Molenaar, 2011).

Here, a third-order polynomial family (Vale & Maurelli, 1983) was used to approximate a distribution. This approach may be useful when skewness and kurtosis are available, but not the raw data; or when there are no compelling reasons to expect a more specific distribution. In applications where a more specific bivariate distribution family is plausible, our general approach could be adapted to that family. To be effective, such an adaptation would require both an algorithm to estimate the distribution's parameters and an analytical solution for the relevant $\hat{\mu}^*_{jk}$ values.

The sample joint moments and approximate distribution methods do have limitations. First, both methods require the bivariate distribution to have finite fourth moments, as these values are needed for the calculation of τ_f^2. Second, regarding the approximate distribution method, the third-order polynomial family is diverse but not completely general. For example, distributions in this family cannot have negative kurtosis. Despite this limitation, the third-order polynomial family still performed well here even when kurtosis was negative, possibly due to the negligible effects of negative kurtosis on τ_f^2. Perhaps a more consequential limitation is that the third-order polynomial family cannot have absolute skewness above 4.4 or kurtosis above 43.4 (see Headrick, 2010, pp. 20–21). Values just slightly beyond these boundaries might be tolerated by the iterative algorithm in our code, which adjusts observed skewness and kurtosis towards 0 until a solution is found. However, values markedly higher than these boundaries may lead to insufficiently

long confidence intervals and observed coverage rates lower than .95. The limitations of the third-order polynomial family might be even more apparent in data with tail dependency (Foldnes & Grønneberg, 2015), which is often an important characteristic to model in chaotic systems, such as financial markets. Third, the present simulations examined the effect of BVN violations; violations of homoscedasticity are less likely to be accommodated by the sample joint moments or approximate distribution methods, and such violations would be better addressed via PM1 or HC4. Finally, there were some scenarios where the approximate distribution method led to low coverage probability (albeit not as low as unadjusted methods), and where existing methods, especially PM1, achieved better coverage.

Overall, we have developed two new methods for adjusting the confidence interval of a correlation for the general case where BVN is not assumed, methods based on the asymptotic distribution of Fisher's z' (Hawkins, 1989). Whereas the sample joint moments method was effective mainly when the sample size was extremely large, in the thousands, the approximate distribution method was effective for a wider range of sample sizes. The primary advantage of this approximate distribution method is that it can be implemented even without access to raw data if sample skewness and kurtosis are reported, making the method especially useful for meta-analysis. We have provided R code to allow application of either method (see supporting information).

Acknowledgements

We thank William Beasley, James Hittner, and James Young for helpful feedback on this project. We also thank Callum Brill and the School of Sciences and Mathematics for help with and access to the Daito high-performance computing cluster.

References

Austin, J. T., Scherbaum, C. A., & Mahlman, R. A. (2002). History of research methods in industrial and organizational psychology: Measurement, design, analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 3–33). Malden, MA: Blackwell.

Beasley, W. H., DeShea, L., Toothaker, L. E., Mendoza, J. L., Bard, D. E., & Rodgers, J. (2007). Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods*, *12*(4), 414–433. https://doi.org/10.1037/1082-989X.12.4.414

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with non-normal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, *17*, 399–417. https://doi.org/10.1037/a0028087

Bishara, A. J., & Hittner, J. B. (2015). Reducing bias and error in the correlation coefficient due to nonnormality. *Educational and Psychological Measurement*, *75*, 785–804. https://doi.org/10.1177/0013164414557639

Bishara, A. J., & Hittner, J. B. (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods*, *49*, 294–309. https://doi.org/10.3758/s13428-016-0702-8

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*, 78–84. https://doi.org/10.1027/1614-2241/a000057

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, *45*(2), 215–233. https://doi.org/10.1016/S0167-9473(02)00366-3

Duncan, G. T., & Layard, M. W. J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika*, *60*, 551–558. https://doi.org/10.1093/biomet/60.3.551

Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, *104*, 293–296. https://doi.org/10.1037/0033-2909.104.2.293

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

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*, 507–521. https://doi.org/10.2307/2331838

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 3–32.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*(4), 521–532. https://doi.org/10.1007/BF02293811

Foldnes, N., & Grønneberg, S. (2015). How general is the Vale-Maurelli simulation approach? *Psychometrika*, *80*, 1066–1083. https://doi.org/10.1007/s11336-014-9414-0

Fouladi, R. T., & Steiger, J. H. (2008). The Fisher transform of the Pearson product moment correlation coefficient and its square: Cumulants, moments, and applications. *Communications in Statistics: Simulation and Computation*, *37*, 928–944. https://doi.org/10.1080/03610910801943735

Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, *38*, 219–247. https://doi.org/10.1093/biomet/38.1-2.219

Goodwin, L. D., & Goodwin, W. L. (1985). An analysis of statistical techniques used in the Journal of Educational Psychology, 1979–1983. *Educational Psychologist*, *20*(1), 13–21. https://doi.org/10.1207/s15326985ep2001_3

Hawkins, D. L. (1989). Using U statistics to derive the asymptotic distribution of Fisher's Z statistic. *American Statistician*, *43*, 235–237. https://doi.org/10.2307/2685369

Headrick, T. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics & Data Analysis*, *40*, 685–711. https://doi.org/10.1016/S0167-9473(02)00072-5

Headrick, T. C. (2010). *Statistical simulation: Power method polynomials and other transformations*. Boca Raton, FL: Chapman & Hall.

Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika*, *64*(1), 25–35. https://doi.org/10.1007/BF02294317

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, *50*, 717–729. https://doi.org/10.1177/0013164490504001

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician*, *47*, 183–189. https://doi.org/10.1111/1467-9884.00122

Lunneborg, C. E. (1985). Estimating the correlation coefficient: The bootstrap approach. *Psychological Bulletin*, *98*, 209–215. https://doi.org/10.1037/0033-2909.98.1.209

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics*, *29*, 201–211.

Olvera Astivia, O. L., & Zumbo, B. D. (2015). A cautionary note on the use of the Vale and Maurelli method to generate multivariate, nonnormal data for simulation purposes. *Educational and Psychological Measurement*, *75*, 541–567. https://doi.org/10.1177/0013164414548894

Puth, M., Neuhäuser, M., & Ruxton, G. D. (2014). Effective use of Pearson's product-moment correlation coefficient. *Animal Behaviour*, *93*, 183–189. https://doi.org/10.1016/j.anbehav.2014.05.003

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rasmussen, J. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches. *Psychological Bulletin*, *101*, 136–139. https://doi.org/10.1037/0033-2909.101.1.136

Strube, M. (1988). Bootstrap Type I error rates for the correlation coefficient: An examination of alternate procedures. *Psychological Bulletin*, *104*(2), 290–292. https://doi.org/10.1037/0033-2909.104.2.290

Stuart, A., & Ord, K. (1994). *Kendall's advanced theory of statistics, Volume I: Distribution theory*. London, UK: Arnold.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*(3), 465–471. https://doi.org/10.1007/BF02293687

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, *6*(11), e26828. https://doi.org/10.1371/journal.pone.0026828

Wilcox, R. R. (1996). Confidence intervals for the slope of a regression line when the error term has nonconstant variance. *Computational Statistics & Data Analysis*, *22*, 89–98. https://doi.org/10.1016/0167-9473(95)00038-0

Wilcox, R. R. (2012). *Modern statistics for the social and behavioral sciences: A practical introduction*. Boca Raton, FL: Taylor & Francis.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*, 1–17. https://doi.org/10.18637/jss.v011.i10

Received 29 October 2015; revised version received 19 June 2017

Supporting Information

The following supporting information may be found in the online edition of the article:

Data S1. Non-normal data simulation method.
Data S2. R code for application.
Data S3. R code for replication.
Data S4. Constants file needed for R code for replication.

Appendix 1: Solutions to $\hat{\mu}^*_{jk}$ for the third-order polynomial family

The third-order polynomial joint moments $\hat{\mu}^*_{jk}$ can be found for particular values of j and k. Two facts aid simplification. First, because Z_1 and Z_2 are independent, $E[Z_1 \cdot Z_2] = E[Z_1] \cdot E[Z_2]$. Second, because Z_1 and Z_2 are standard normal, all of their moments are known (see Stuart & Ord, 1994):

$$E[Z^j] = \begin{cases} 0, & \text{if } j \text{ is odd,} \\ \frac{j!}{(j/2)! \cdot 2^{j/2}}, & \text{if } j \text{ is even.} \end{cases} \tag{A1}$$

Using these facts with equation (22) produces the following solutions:

$$\begin{aligned}\hat{\mu}^*_{40} = {} & 3\hat{b}_1^4 + 60\hat{b}_1^2\hat{c}_1^2 + 60\hat{c}_1^4 + 60\hat{b}_1^3\hat{d}_1 + 936\hat{b}_1\hat{c}_1^2\hat{d}_1 + 630\hat{b}_1^2\hat{d}_1^2 \\ & + 4500\hat{c}_1^2\hat{d}_1^2 + 3780\hat{b}_1\hat{d}_1^3 + 10395\hat{d}_1^4,\end{aligned} \tag{A2}$$

$$\begin{aligned}\hat{\mu}^*_{22} = {} & \hat{b}_1^2\hat{b}_2^2 + 2\hat{t}^2\hat{b}_1^2\hat{b}_2^2 + 2\hat{b}_2^2\hat{c}_1^2 + 8\hat{t}^2\hat{b}_2^2\hat{c}_1^2 + 16\hat{t}\hat{b}_1\hat{b}_2\hat{c}_1\hat{c}_2 + \\ & 24\hat{t}^3\hat{b}_1\hat{b}_2\hat{c}_1\hat{c}_2 + 2\hat{b}_1^2\hat{c}_2^2 + 8\hat{t}^2\hat{b}_1^2\hat{c}_2^2 + 4\hat{c}_1^2\hat{c}_2^2 + 32\hat{t}^2\hat{c}_1^2\hat{c}_2^2 + \\ & 24\hat{t}^4\hat{c}_1^2\hat{c}_2^2 + 6\hat{b}_1\hat{b}_2^2\hat{d}_1 + 24\hat{t}^2\hat{b}_1\hat{b}_2^2\hat{d}_1 + 96\hat{t}\hat{b}_2\hat{c}_1\hat{c}_2\hat{d}_1 + \\ & 216\hat{t}^3\hat{b}_2\hat{c}_1\hat{c}_2\hat{d}_1 + 12\hat{b}_1\hat{c}_2^2\hat{d}_1 + 96\hat{t}^2\hat{b}_1\hat{c}_2^2\hat{d}_1 + 48\hat{t}^4\hat{b}_1\hat{c}_2^2\hat{d}_1 + \\ & 15\hat{b}_2^2\hat{d}_1^2 + 90\hat{t}^2\hat{b}_2^2\hat{d}_1^2 + 30\hat{c}_2^2\hat{d}_1^2 + 360\hat{t}^2\hat{c}_2^2\hat{d}_1^2 + 360\hat{t}^4\hat{c}_2^2\hat{d}_1^2 + \\ & 6\hat{b}_1^2\hat{b}_2\hat{d}_2 + 24\hat{t}^2\hat{b}_1^2\hat{b}_2\hat{d}_2 + 12\hat{b}_2\hat{c}_1^2\hat{d}_2 + 96\hat{t}^2\hat{b}_2\hat{c}_1^2\hat{d}_2 + \\ & 48\hat{t}^4\hat{b}_2\hat{c}_1^2\hat{d}_2 + 96\hat{t}\hat{b}_1\hat{c}_1\hat{c}_2\hat{d}_2 + 216\hat{t}^3\hat{b}_1\hat{c}_1\hat{c}_2\hat{d}_2 + \\ & 36\hat{b}_1\hat{b}_2\hat{d}_1\hat{d}_2 + 288\hat{t}^2\hat{b}_1\hat{b}_2\hat{d}_1\hat{d}_2 + 96\hat{t}^4\hat{b}_1\hat{b}_2\hat{d}_1\hat{d}_2 + \\ & 576\hat{t}\hat{c}_1\hat{c}_2\hat{d}_1\hat{d}_2 + 1944\hat{t}^3\hat{c}_1\hat{c}_2\hat{d}_1\hat{d}_2 + 480\hat{t}^5\hat{c}_1\hat{c}_2\hat{d}_1\hat{d}_2 + \\ & 90\hat{b}_2\hat{d}_1^2\hat{d}_2 + 1080\hat{t}^2\hat{b}_2\hat{d}_1^2\hat{d}_2 + 720\hat{t}^4\hat{b}_2\hat{d}_1^2\hat{d}_2 + 15\hat{b}_1^2\hat{d}_2^2 + \\ & 90\hat{t}^2\hat{b}_1^2\hat{d}_2^2 + 30\hat{c}_1^2\hat{d}_2^2 + 360\hat{t}^2\hat{c}_1^2\hat{d}_2^2 + 360\hat{t}^4\hat{c}_1^2\hat{d}_2^2 + \\ & 90\hat{b}_1\hat{d}_1\hat{d}_2^2 + 1080\hat{t}^2\hat{b}_1\hat{d}_1\hat{d}_2^2 + 720\hat{t}^4\hat{b}_1\hat{d}_1\hat{d}_2^2 + \\ & 225\hat{d}_1^2\hat{d}_2^2 + 4050\hat{t}^2\hat{d}_1^2\hat{d}_2^2 + 5400\hat{t}^4\hat{d}_1^2\hat{d}_2^2 + 720\hat{t}^6\hat{d}_1^2\hat{d}_2^2,\end{aligned} \tag{A3}$$

$$\begin{aligned}\hat{\mu}^*_{31} = {} & 3\hat{t}\hat{b}_1^3\hat{b}_2 + 30\hat{t}\hat{b}_1\hat{b}_2\hat{c}_1^2 + 30\hat{t}^2\hat{b}_1^2\hat{c}_1\hat{c}_2 + 60\hat{t}^2\hat{c}_1^3\hat{c}_2 + 45\hat{t}\hat{b}_1^2\hat{b}_2\hat{d}_1 + \\ & 234\hat{t}\hat{b}_2\hat{c}_1^2\hat{d}_1 + 468\hat{t}^2\hat{b}_1\hat{c}_1\hat{c}_2\hat{d}_1 + 315\hat{t}\hat{b}_1\hat{b}_2\hat{d}_1^2 + 2250\hat{t}^2\hat{c}_1\hat{c}_2\hat{d}_1^2 + \\ & 945\hat{t}\hat{b}_2\hat{d}_1^3 + 9\hat{t}\hat{b}_1^3\hat{d}_2 + 6\hat{t}^3\hat{b}_1^3\hat{d}_2 + 90\hat{t}\hat{b}_1\hat{c}_1^2\hat{d}_2 + 144\hat{t}^3\hat{b}_1\hat{c}_1^2\hat{d}_2 + \\ & 135\hat{t}\hat{b}_1^2\hat{d}_1\hat{d}_2 + 180\hat{t}^3\hat{b}_1^2\hat{d}_1\hat{d}_2 + 702\hat{t}\hat{c}_1^2\hat{d}_1\hat{d}_2 + 1548\hat{t}^3\hat{c}_1^2\hat{d}_1\hat{d}_2 + \\ & 945\hat{t}\hat{b}_1\hat{d}_1^2\hat{d}_2 + 1890\hat{t}^3\hat{b}_1\hat{d}_1^2\hat{d}_2 + 2835\hat{t}\hat{d}_1^3\hat{d}_2 + 7560\hat{t}^3\hat{d}_1^3\hat{d}_2.\end{aligned} \tag{A4}$$

For the remaining joint moments, $\hat{\mu}^*_{04}$ is identical to $\hat{\mu}^*_{40}$ and $\hat{\mu}^*_{13}$ is identical to $\hat{\mu}^*_{31}$, except with the subscripts on the right sides of the equations reversed (1 replaced with 2, and vice-versa).