11-1-2005

Second-Order Accurate Inference on Simple, Partial, and Multiple Correlations

Robert J. Boik
Montana State University{Bozeman, rjboik@math.montana.edu

Ben Haaland
University of Wisconsin{Madison, haaland@stat.wisc.edu

Invited Articles
Second-Order Accurate Inference on Simple, Partial, and Multiple Correlations

Robert J. Boik
Mathematical Sciences
Montana State University–Bozeman

Ben Haaland
Statistics
University of Wisconsin–Madison

This article develops confidence interval procedures for functions of simple, partial, and squared multiple correlation coefficients. It is assumed that the observed multivariate data represent a random sample from a distribution that possesses finite moments, but there is no requirement that the distribution be normal. The coverage error of conventional one-sided large sample intervals decreases at rate $1/\sqrt{n}$ as n increases, where n is an index of sample size. The coverage error of the proposed intervals decreases at rate $1/n$ as n increases. The results of a simulation study that evaluates the performance of the proposed intervals is reported and the intervals are illustrated on a real data set.

Key words: bootstrap, confidence intervals, Cornish-Fisher expansion, Edgeworth expansion, second-order accuracy

Introduction

Accurate inference procedures for simple, partial, and multiple correlation coefficients depend on accurate approximations to the distributions of estimators of these coefficients. In this article, Edgeworth ex-

Robert Boik (rjboik@math.montana.edu) is Professor of Statistics. His research interests include linear models, multivariate analysis, and large sample methods. Ben Haaland (haaland@stat.wisc.edu) is a Ph.D. student. His research interests include multivariate analysis, rates of convergence, and nonparametric techniques.

pansions for these distributions are derived. The Edgeworth expansions, in turn, are used to construct confidence intervals that are more accurate than conventional large sample intervals.

Denote the $p \times p$ sample correlation matrix based on a random sample of size N from a p-variate distribution by $\mathbf{R}$ and denote the corresponding population correlation matrix by $\boldsymbol{\Delta}$. The $p^2 \times 1$ vectors obtained by stacking the columns of $\mathbf{R}$ and $\boldsymbol{\Delta}$ are denoted by $\mathbf{r}$ and $\boldsymbol{\rho}$, and are obtained by applying the vec operator. That is, $\mathbf{r} \overset{\text{def}}{=} \text{vec}\,\mathbf{R}$ and $\boldsymbol{\rho} \overset{\text{def}}{=} \text{vec}\,\boldsymbol{\Delta}$. The exact joint distribution of the components of $\mathbf{r}$,

when sampling from a multivariate normal distribution, was derived by Fisher (1962), but the distribution is very difficult to use in practice because it is expressed in integral form unless $p = 2$. If sample size is sufficiently large, however, then one may substitute the asymptotic distribution of $\mathbf{r}$, but with some loss of accuracy.

In this article, the big O (pronounced oh) notation (Bishop, Fienberg, & Holland, 1975, §14.2) is used to index the magnitude of a quantity. Let u_n be a quantity that depends on $n = N - r_x$, where r_x is a fixed constant and N is sample size. Then, $u_n = O(n^{-k})$ if $n^k|u_n|$ is bounded as $n \to \infty$. Note that if $u_n = O(n^{-k})$ and $k > 0$, then u_n converges to zero as $n \to \infty$. Also, the rate of convergence to zero is faster for large k than for small k. An approximation to the distribution of a random variable is said to be j^{th}-order accurate if the difference between the approximating cumulative distribution function (cdf) and the exact cdf has magnitude $O(n^{-j/2})$. For example, the cdf of a sample correlation coefficient, r_{ij}, is $F_{r_{ij}}(t) \stackrel{\text{def}}{=} P(r_{ij} \le t)$. Suppose that $\hat{F}_{r_{ij}}$ is an estimator of $F_{r_{ij}}$. Then, $\hat{F}_{r_{ij}}$ is first-order accurate if $|F_{r_{ij}}(t) - \hat{F}_{r_{ij}}(t)| = O(n^{-1/2})$, for all t.

Pearson and Filon (1898) showed that the first-order accurate asymptotic joint distribution of the components of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$, when sampling from a multivariate normal distribution, is itself multivariate normal with mean zero and finite covariance matrix. Pearson and Filon also derived expressions for the components of the asymptotic covariance matrix. A matrix expression for the asymptotic covariance matrix was derived by Nell (1985). Niki and Konishi (1984) derived the Edgeworth and Cornish-Fisher expansions for $\sqrt{n}\left[Z(r) - Z(\rho)\right]$ with error only $O(n^{-9/2})$ when sampling from a bivariate normal

distribution, where Z is Fisher's (1921) Z transformation. If the requirement of multivariate normality is relaxed, then the distributions of r and $\mathbf{r}$ are substantially more complicated. Fortunately, the asymptotic distribution of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ still is multivariate normal with mean zero and finite covariance matrix whenever the parent distribution has finite fourth-order moments. Expressions for the scalar components of the asymptotic covariance matrix of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ when sampling from non-normal distributions were derived by Hsu (1949) and Steiger and Hakstian (1982, 1983). Corresponding matrix expressions were derived by Browne and Shapiro (1986) and Neudecker and Wesselman (1990). The asymptotic bias of $\mathbf{r}$ when sampling from non-normal distributions was obtained by Boik (1998).

In the bivariate case, Cook (1951) obtained scalar expressions for the moments of r with error $O(n^{-5/2})$. These moments could be used to compute the first three cumulants of $\sqrt{n}(r - \rho)$. The first four cumulants and the corresponding Edgeworth expansion for the distribution of of $\sqrt{n}(r - \rho)$ were obtained by Nakagawa and Niki (1992).

Some distributional results for $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ have been obtained under special non-normal conditions. Neudecker (1996), for example, obtained a matrix expression for the asymptotic covariance matrix in the special case of elliptical parent distributions. Also, Yuan and Bentler (1999) gave the asymptotic distribution of correlation and multiple correlation coefficients when the p-vector of random variables can be written as a linear function of $\mathbf{z}v$, where $\mathbf{z}$ is a p-vector of independent components and v is a scalar that is independent of $\mathbf{z}$.

This article focuses on confidence intervals for functions of $\boldsymbol{\Delta}$. Specifically, second-order accurate interval estimators

for simple, partial, and squared multiple correlation coefficients as well as for differences among simple, partial, and squared multiple correlation coefficients are constructed. A confidence interval is said to be j^{th}-order accurate if the difference between the actual coverage and the stated nominal coverage has magnitude $O(n^{-j/2})$. In general, large sample confidence intervals are first-order accurate under mild validity conditions.

First-order accurate confidence intervals for correlation functions are not new. Olkin and Siotani (1976) and Hedges and Olkin (1983) used the delta method (Rao, 1973, §6a.2) together with the asymptotic distribution of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ when sampling from a multivariate normal distribution to derive the asymptotic distribution of partial and multiple correlation coefficient estimators. Olkin and Finn (1995) used these results to obtain explicit expressions for asymptotic standard errors of estimators of simple, partial, and squared multiple correlation coefficients as well as expressions for standard errors of differences among these coefficients. The major contribution of Olkin and Finn (1995) was that they demonstrated how to use existing moment expressions to compute first-order accurate confidence intervals for correlation functions when sampling from multivariate normal distributions.

To avoid complicated expressions for derivatives, Olkin and Finn (1995) gave confidence intervals for partial correlations and their differences only in the special case when the effects of a single variable are partialed out. Similarly, confidence intervals for differences among squared multiple correlation coefficients (estimated using a single sample) are given only for the special cases when either one or two explanatory variables are employed. Graf and Alf (1999) used numerical derivatives to extend the re-

sults of Olkin and Finn to allow conditioning on any set of variables rather than a single variable and to allow squared multiple correlations to be based on an arbitrary number of explanatory variables. Alf and Graf (1999) gave scalar equations for the standard error of the difference between two squared sample multiple correlations. This article also extends the results of Olkin and Finn (1995), but does so using relatively simple and easily computed (with a computer) matrix expressions for the required derivatives. As in the previous articles, this article relies on derivatives of correlation functions. Unlike Olkin and Finn, however, simple, partial, and multiple correlations are treated as functions of the covariance matrix, $\boldsymbol{\Sigma}$, instead of the correlation matrix, $\boldsymbol{\Delta}$. The advantage of the current approach is simplicity of the resulting expansions.

This article also extends the results of Olkin and Finn (1995), Graf and Alf (1999), and Alf and Graf (1999) to be applicable to to non-normal as well as normal distributions. This extension is straightforward. Denote the sample covariance based on a random sample of size N by $\mathbf{S}$. Then, one needs only to replace the asymptotic covariance of vec $\mathbf{S}$ derived under normality with the asymptotic covariance matrix derived under general conditions.

In summary, the contributions of this article are as follows: (a) easily computed matrix expressions for the required derivatives are given (see Theorems 2 and 3), (b) the proposed intervals are asymptotically distribution free (ADF); and (c) the accuracy of the confidence intervals is extended from first-order to second-order. In addition, the proposed interval estimators are illustrated on a real data set and a small simulation study that evaluates the performance of the proposed intervals is conducted.

Before describing the new results, the Olkin and Finn (1995) article will be briefly revisited. It appears that the methods described in Olkin and Finn have begun to be incorporated into statistical practice (34 citations from 1997 to June 2005, most of which are reports of empirical studies). Unfortunately, the derivatives reported in the article contain several errors. The effects of these errors on the examples given by Olkin and Finn are minimal, but their effects in practice are unknown. A list of corrections to these derivatives is given in the next section.

Corrections to Olkin and Finn

Below is a list of corrections to the derivatives reported in Olkin and Finn (1995).

Model B. Components a_1 and a_3 in model B on page 159 should be the following:

$$a_1 = 2\Big[\rho_{01}\left(\rho_{12}^2 - \rho_{13}^2\right) + \rho_{03}\rho_{13} - \rho_{02}\rho_{12}$$

$$+\rho_{12}\rho_{13}\left(\rho_{02}\rho_{13} - \rho_{03}\rho_{12}\right)\Big]$$

$$\div \left[\left(1 - \rho_{12}^2\right)\left(1 - \rho_{13}^2\right)\right]$$

$$\text{and } a_3 = -2\left(\frac{\rho_{03} - \rho_{01}\rho_{13}}{1 - \rho_{13}^2}\right).$$

The nominal 95% confidence interval for $\rho_{0(12)}^2 - \rho_{0(13)}^2$ should be $[-0.19, 0.015]$ rather than $[-0.22, 0.018]$.

Model C. Components a_1 and a_3 in model C on page 160 should be the following:

$$a_1 = \sqrt{\left(1 - \rho_{02}^2\right)\left(1 - \rho_{12}^2\right)} - 1 \text{ and}$$

$$a_3 = \frac{\rho_{02} - \rho_{01}\rho_{12}}{1 - \rho_{12}^2}.$$

The nominal 95% confidence interval for $\rho_{01} - \rho_{01\cdot2}$ should be $[-0.0018, 0.0073]$ rather than $[-0.0017, 0.0080]$.

Model D. Components a_1, a_3, and a_5 in model D on page 161 should be the following:

$$a_1 = \frac{1}{\sqrt{\left(1 - \rho_{02}^2\right)\left(1 - \rho_{12}^2\right)}}$$

$$-\frac{1}{\sqrt{\left(1 - \rho_{03}^2\right)\left(1 - \rho_{13}^2\right)}},$$

$$a_3 = \frac{\rho_{13} - \rho_{01}\rho_{03}}{\left(1 - \rho_{03}^2\right)^{3/2}\left(1 - \rho_{13}^2\right)^{1/2}}, \text{ and}$$

$$a_5 = \frac{\rho_{03} - \rho_{01}\rho_{13}}{\left(1 - \rho_{03}^2\right)^{1/2}\left(1 - \rho_{13}^2\right)^{3/2}}.$$

The nominal 95% confidence interval for $\rho_{01\cdot2} - \rho_{01\cdot3}$ should be $[0.020, 0.054]$ rather than $[0.017, 0.057]$.

Notation and Preliminary Remarks

In some applications, an investigator may have one or more fixed explanatory variables (e.g., group or treatment variables) whose effects must be removed before estimating the correlation matrix. A linear model that is suitable for this purpose is described in this section. The sample covariance matrix, $\mathbf{S}$, is based on the residuals from the fitted model. The confidence interval procedures that are constructed later in this article require estimates of second and third-order moments of $\mathbf{S}$. Expressions for the population moments as well as consistent estimators of these moments are described in this article. Sufficient validity conditions were described in Boik (2005, §5.1).

Linear Model

Denote the response vector for the ith subject by $\mathbf{y}_i$ and let $\mathbf{Y}$ be the $N \times p$ response matrix whose i^{th} row is $\mathbf{y}_i'$. That is,

$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \end{pmatrix}'$. The joint model for the N subjects is

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \qquad (1)$$

where $\mathbf{X}$ is an $N \times q$ matrix of fixed explanatory variables, and $\mathbf{E}$ is an $N \times p$ random matrix. Denote the ith row of $\mathbf{E}$ by $\boldsymbol{\varepsilon}_i'$. It is assumed that $\boldsymbol{\varepsilon}_i$ for $i = 1, \ldots, N$ are independently and identically distributed with mean $\mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$.

The components of $\boldsymbol{\Sigma}$ are denoted by σ_{ij} for $i = 1, \ldots, p$ and $j = 1, \ldots, p$. Note that the variance of the ith variable is σ_{ii}, not σ_{ii}^2, and the correlation between the ith and j^{th} variables is $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. To represent the matrix of correlations as a function of $\boldsymbol{\Sigma}$, a notation for diagonal matrices is needed. Suppose that $\mathbf{M}$ is a $q \times q$ matrix with positive diagonal components $m_{11}, m_{22}, \ldots, m_{qq}$ and that b is a real number. Then the diagonal matrix with diagonal components $m_{11}^b, m_{22}^b, \ldots, m_{qq}^b$ is denoted by $(\mathbf{M})_D^b$. That is,

$$(\mathbf{M})_D^b = \sum_{i=1}^{q} \mathbf{e}_i^q m_{ii}^b \mathbf{e}_i^{q\prime}, \qquad (2)$$

where $\mathbf{e}_i^q$ is the ith column of $\mathbf{I}_q$. Using this notation, the $p \times p$ correlation matrix is

$$\boldsymbol{\Delta} = (\boldsymbol{\Sigma})_D^{-\frac{1}{2}} \, \boldsymbol{\Sigma} \, (\boldsymbol{\Sigma})_D^{-\frac{1}{2}} .$$

Estimators of simple, partial, and multiple correlation coefficients will be based on the sample covariance matrix, namely

$$\mathbf{S} = n^{-1} \mathbf{Y}' \mathbf{A} \mathbf{Y}, \qquad (3)$$

where $\mathbf{A} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$, $(\;)^{-}$ denotes an arbitrary generalized inverse, $n = N - r_x$, and $r_x = \mathrm{rank}(\mathbf{X})$. It is readily shown that if the rows of $\mathbf{Y}$ have finite variance, then $\mathbf{S}$ is unbiased for $\boldsymbol{\Sigma}$.

Moments of the Sample Covariance Matrix

If fourth-order moments of $\mathbf{Y}$ are finite, then the central limit theorem ensures that

$$\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_{22,\infty}), \text{ where} \quad (4)$$

$$\boldsymbol{\Omega}_{22,\infty} = \boldsymbol{\Upsilon}_{22} - \boldsymbol{\sigma}\boldsymbol{\sigma}', \quad \boldsymbol{\Upsilon}_{22} = \mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \otimes \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'),$$

$\mathbf{s} \stackrel{\text{def}}{=} \mathrm{vec}\,\mathbf{S}$, $\boldsymbol{\sigma} \stackrel{\text{def}}{=} \mathrm{vec}\,\boldsymbol{\Sigma}$, and $\boldsymbol{\varepsilon}_i$ is the ith row of $\mathbf{E}$ in (1). Boik (1998, 2005, eq. 12) showed that if fourth-order moments of $\mathbf{Y}$ are finite, then the finite-sample variance of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ is

$$\boldsymbol{\Omega}_{22,n} \stackrel{\text{def}}{=} \mathrm{Var}\left[\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right] \qquad (5)$$

$$= \frac{c_1}{n}\left(\boldsymbol{\Upsilon}_{22} - \boldsymbol{\sigma}\boldsymbol{\sigma}'\right) + \left(1 - \frac{c_1}{n}\right) 2\mathbf{N}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}),$$

where $c_1 = \sum_{i=1}^{N} a_{ii}^2$, a_{ij} is the ij^{th} component of $\mathbf{A}$ in (3), $\boldsymbol{\Upsilon}_{22}$ is defined in (4), $\mathbf{N}_p = \left(\mathbf{I}_{p^2} + \mathbf{I}_{(p,p)}\right)/2$, and $\mathbf{I}_{(a,b)}$ is the commutation matrix (MacRae, 1974). Magnus and Neudecker (1979, 1999 §3.7) denoted $\mathbf{I}_{(a,b)}$ by $\mathbf{K}_{b,a}$. If data are sampled from a multivariate normal distribution, then the finite-sample variance and an unbiased estimator of the finite-sample variance of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ are

$$\boldsymbol{\Omega}_{22,n} = 2\mathbf{N}_p\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right) \text{ and}$$

$$\widehat{\boldsymbol{\Omega}}_{22,n} = \frac{n^2}{(n-1)(n+2)} 2\mathbf{N}_p\left(\mathbf{S} \otimes \mathbf{S}\right)$$

$$-\frac{2n}{(n-1)(n+2)}\mathbf{ss}',$$

respectively. From Boik (1998, 2005, eq. 13), an unbiased estimator of $\boldsymbol{\Omega}_{22,n}$ under general conditions is

$$\widehat{\boldsymbol{\Omega}}_{22,n} = a_1 \widetilde{\boldsymbol{\Upsilon}}_{22} + a_2 2\mathbf{N}_p\left(\mathbf{S} \otimes \mathbf{S}\right) \qquad (6)$$

$$+a_3 \mathbf{ss}', \text{ where}$$

$$a_1 = \frac{n^2 c_1}{d}, \quad a_2 = -\frac{n^2(c_1^2 - nc_2)}{(n-1)d},$$

$$a_3 = -\frac{n\left[2nc_2 + (n-3)c_1^2\right]}{(n-1)d},$$

$$c_2 = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^4, \; d = n(n+2)c_2 - 3c_1^2,$$

$$\widetilde{\boldsymbol{\Upsilon}}_{22} = n^{-1}\sum_{i=1}^{N}\left(\widetilde{\varepsilon}_i\widetilde{\varepsilon}_i' \otimes \widetilde{\varepsilon}_i\widetilde{\varepsilon}_i'\right),$$

c_1 is defined in (5), and $\widetilde{\varepsilon}_i = \mathbf{Y}'\mathbf{A}\mathbf{e}_i^N$ is the p-vector of residuals from the fitted model for subject i.

Two additional covariance-related quantities are needed in this article, namely

$$\boldsymbol{\Omega}_{42,n} \tag{7}$$

$$\stackrel{\text{def}}{=} \text{Cov}\left[\sqrt{n}\,(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}), \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]$$

$$= n\text{E}\left[(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\mathbf{s} - \boldsymbol{\sigma})'\right] \text{ and }$$

$$\boldsymbol{\Omega}_{222,n} \stackrel{\text{def}}{=} n^{\frac{3}{2}}\text{E}\left[(\mathbf{s} - \boldsymbol{\sigma})(\mathbf{s} - \boldsymbol{\sigma})' \otimes (\mathbf{s} - \boldsymbol{\sigma})\right],$$

where $\widehat{\boldsymbol{\omega}} = \text{vec}\,\widehat{\boldsymbol{\Omega}}_{22,n}$ and $\boldsymbol{\omega} = \text{vec}\,\boldsymbol{\Omega}_{22,n}$. Using the results of Boik (2005), it can be shown that if sixth order moments are finite and the column space generated by $\mathbf{X}$ in (1) contains the n-vector of ones, then $\boldsymbol{\Omega}_{42,n}$ and $\sqrt{n}\boldsymbol{\Omega}_{222,n}$ can be written as follows:

$$\boldsymbol{\Omega}_{42,n} \tag{8}$$

$$= \boldsymbol{\Upsilon}_{42} - 2\mathbf{N}_{p^2}\left(\boldsymbol{\Omega}_{22,n} \otimes \boldsymbol{\sigma}\right) - \left(\boldsymbol{\Upsilon}_{22} \otimes \boldsymbol{\sigma}\right)$$

$$-4\mathbf{N}_{p^2}\left(\mathbf{N}_p \otimes \mathbf{I}_{p^2}\right)\left(\boldsymbol{\Upsilon}_{21}' \otimes \text{vec}\,\boldsymbol{\Upsilon}_{21}\right)$$

$$+O\left(n^{-1}\right) \text{ and } \sqrt{n}\boldsymbol{\Omega}_{222,n}$$

$$= \boldsymbol{\Omega}_{42,n} - \left(\boldsymbol{\Upsilon}_{21} \otimes \boldsymbol{\Upsilon}_{21}\right)2\mathbf{N}_p + O\left(n^{-1}\right),$$

where $\boldsymbol{\Upsilon}_{42} = \text{E}\left(\varepsilon_i \otimes \varepsilon_i \otimes \varepsilon_i\varepsilon_i' \otimes \varepsilon_i\varepsilon_i'\right),$

and $\boldsymbol{\Upsilon}_{21} = \text{E}\left(\varepsilon_i \otimes \varepsilon_i\varepsilon_i'\right).$

If the data are a random sample from a multivariate normal distribution, then to order $O(n^{-1})$, $\boldsymbol{\Omega}_{42,n}$ and $\sqrt{n}\boldsymbol{\Omega}_{222,n}$ simplify to

$$\boldsymbol{\Omega}_{42,n} \tag{9}$$

$$= \left(2\mathbf{N}_p \otimes 2\mathbf{N}_p\right)\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\sigma} \otimes \boldsymbol{\Sigma}\right)2\mathbf{N}_p$$

$$\text{and } \sqrt{n}\boldsymbol{\Omega}_{222,n} = \boldsymbol{\Omega}_{42,n}.$$

It will be seen later in this article that the proposed inference procedures depend on $\boldsymbol{\Omega}_{222,n}$ and $\boldsymbol{\Omega}_{42,n}$ only through functions that can be written as $\left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)'\text{vec}\,\boldsymbol{\Omega}_{222,n}$ and $\left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)'\text{vec}\,\boldsymbol{\Omega}_{42,n}$, where $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ is a $p^2 \times 1$ vector that satisfies $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\sigma} = 0$. Theorem 1 gives consistent estimators of the required quantities. The results in Theorem 1 follow from the consistency of $\widehat{\boldsymbol{\Omega}}_{42,n}$ and $\widehat{\boldsymbol{\Omega}}_{222,n}$ described by Boik (2005).

Theorem 1. *Suppose that $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ is a $p^2 \times 1$ vector function of $\boldsymbol{\Sigma}$ that satisfies $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\sigma} = 0$. Then, the required sixth-order quantities can be consistently estimated as*

$$\left(\dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}}\right)'\text{vec}\,\widehat{\boldsymbol{\Omega}}_{42,n} =$$

$$n^{-1}\sum_{i=1}^{N}\left[\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\left(\widetilde{\varepsilon}_i \otimes \widetilde{\varepsilon}_i\right)\right]^3 - 4\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Upsilon}}_{21}\dot{\boldsymbol{\Psi}}_{\mathbf{s}}\widehat{\boldsymbol{\Upsilon}}_{21}'\dot{\boldsymbol{\psi}}_{\mathbf{s}}$$

and $\sqrt{n}\left(\dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}}\right)'\text{vec}\,\widehat{\boldsymbol{\Omega}}_{222,n}$

$$= \left(\dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}}\right)'\text{vec}\,\widehat{\boldsymbol{\Omega}}_{42,n}$$

$$-2\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Upsilon}}_{21}\dot{\boldsymbol{\Psi}}_{\mathbf{s}}\widehat{\boldsymbol{\Upsilon}}_{21}'\dot{\boldsymbol{\psi}}_{\mathbf{s}},$$

$$\text{where } \widehat{\boldsymbol{\Upsilon}}_{21} = n^{-1} \sum_{i=1}^{N} \left(\widetilde{\boldsymbol{\varepsilon}}_i \otimes \widetilde{\boldsymbol{\varepsilon}}_i \widetilde{\boldsymbol{\varepsilon}}_i' \right),$$

$\dot{\boldsymbol{\psi}}_{\mathbf{s}}$ is a consistent estimator of $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$, $\dot{\boldsymbol{\Psi}}_{\mathbf{s}}$ is a $p \times p$ matrix that satisfies $\mathrm{vec}\, \dot{\boldsymbol{\Psi}}_{\mathbf{s}} = \dot{\boldsymbol{\psi}}_{\mathbf{s}}$, and $\widetilde{\boldsymbol{\varepsilon}}_i$ is defined in (6). If the rows of $\mathbf{Y}$ have a multivariate normal distribution, then the required sixth-order quantities can be consistently estimated as

$$\left(\dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \right)' \mathrm{vec}\, \widehat{\boldsymbol{\Omega}}_{42,n} \qquad (10)$$

$$= \sqrt{n} \left(\dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}} \right)' \mathrm{vec}\, \widehat{\boldsymbol{\Omega}}_{222,n}$$

$$= 8 \,\mathrm{trace} \left[\left(\mathbf{S} \dot{\boldsymbol{\Psi}}_{\mathbf{s}} \right)^3 \right]. \qquad \square$$

Specific structures for $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ and $\dot{\boldsymbol{\psi}}_{\mathbf{s}}$ are described below.

Derivatives

The main theoretical tool used in this article is the Taylor series expansion of random quantities. The first-order linear expansion is the basis of the delta method. To construct second-order accurate interval estimators, it is necessary to include the quadratic term in the expansion as well. Accordingly, both first and second derivatives of correlation functions are required. Derivatives of partial correlations as well as squared multiple correlations are given in this article.

Partial Correlations

Suppose that an investigator is interested in the correlation between variables i and j, controlling for a subset of variables indexed by $\mathbf{k} = \begin{pmatrix} k_1 & k_2 & \cdots & k_q \end{pmatrix}'$. This partial correlation, denoted as $\rho_{ij\cdot\mathbf{k}}$, is the off diagonal component in the correlation matrix $\boldsymbol{\Delta}_{ij\cdot\mathbf{k}}$, where

$$\boldsymbol{\Delta}_{ij\cdot\mathbf{k}} = \begin{pmatrix} 1 & \rho_{ij\cdot\mathbf{k}} \\ \rho_{ij\cdot\mathbf{k}} & 1 \end{pmatrix} \qquad (11)$$

$$= \left(\boldsymbol{\Sigma}_{ij\cdot\mathbf{k}} \right)_D^{-\frac{1}{2}} \boldsymbol{\Sigma}_{ij\cdot\mathbf{k}} \left(\boldsymbol{\Sigma}_{ij\cdot\mathbf{k}} \right)_D^{-\frac{1}{2}}, \text{ where}$$

$$\boldsymbol{\Sigma}_{ij\cdot\mathbf{k}} = \begin{pmatrix} \sigma_{ii\cdot\mathbf{k}} & \sigma_{ij\cdot\mathbf{k}} \\ \sigma_{ij\cdot\mathbf{k}} & \sigma_{jj\cdot\mathbf{k}} \end{pmatrix}$$

$$= \mathbf{E}_{ij}' \left[\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{H}_k\boldsymbol{\Sigma} \right] \mathbf{E}_{ij}, \quad \mathbf{E}_{ij} = \begin{pmatrix} \mathbf{e}_i^p & \mathbf{e}_j^p \end{pmatrix},$$

$$\mathbf{E}_{\mathbf{k}} = \begin{pmatrix} \mathbf{e}_{k_1}^p & \mathbf{e}_{k_2}^p & \cdots & \mathbf{e}_{k_q}^p \end{pmatrix},$$

$$\mathbf{H}_k = \mathbf{E}_{\mathbf{k}} \left(\mathbf{E}_{\mathbf{k}}' \boldsymbol{\Sigma}\mathbf{E}_{\mathbf{k}} \right)^{-1} \mathbf{E}_{\mathbf{k}}',$$

and the remaining terms are defined in (2). It is required that neither i nor j be a component of $\mathbf{k}$; otherwise $\rho_{ij\cdot\mathbf{k}} = 0$ and no inference is necessary. The parameter $\rho_{ij\cdot\mathbf{k}}$ can be estimated by $r_{ij\cdot\mathbf{k}}$, where $r_{ij\cdot\mathbf{k}}$ is obtained by substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$ in (11).

Theorem 2 gives first and second derivatives of $\rho_{ij\cdot\mathbf{k}}$ with respect to $\boldsymbol{\sigma}$. The derivatives are given without proof because they follow from established matrix calculus results (Graham, 1981).

Theorem 2. *The first two derivatives of $\rho_{ij\cdot\mathbf{k}}$ with respect to $\boldsymbol{\sigma}$ are*

$$\dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}} \stackrel{\text{def}}{=} \frac{\partial \rho_{ij\cdot\mathbf{k}}}{\partial \boldsymbol{\sigma}} = \mathbf{N}_p(\boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}})$$

$$- \left[(\boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}}) + (\boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}}) \right] \rho_{ij\cdot\mathbf{k}}/2 \ and$$

$$\ddot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}} \stackrel{\text{def}}{=} \frac{\partial^2 \rho_{ij\cdot\mathbf{k}}}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}}$$

$$= \mathbf{N}_{p^2} \left(\boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \right) \rho_{ij\cdot\mathbf{k}}/2$$

$$+ \Big[\left(\boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \right)$$

$$+ \left(\boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \right) \Big] 3\rho_{ij\cdot\mathbf{k}}/4$$

$$- \left(\mathbf{N}_p \otimes \mathbf{N}_p \right) \mathbf{N}_{p^2} \Big[\left(\boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{j\cdot\mathbf{k}} \right)$$

$$+ \left(\boldsymbol{\gamma}_{j\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \otimes \boldsymbol{\gamma}_{i\cdot\mathbf{k}} \right) \Big]$$

$$-2 \left(\mathbf{N}_p \otimes \mathbf{N}_p \right) \left(\mathbf{I}_p \otimes \mathbf{h}_k \otimes \mathbf{I}_p \right) \dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}, \ \ where$$

$$\boldsymbol{\gamma}_{i\cdot\mathbf{k}} = \left[\mathbf{I}_p - \mathbf{H}_k \boldsymbol{\Sigma} \right] \mathbf{e}_i^p / \sqrt{\sigma_{ii\cdot\mathbf{k}}},$$

$$\mathbf{h}_k = \mathrm{vec} \left(\mathbf{H}_k \right),$$

$\mathbf{H}_k$ *is defined in* (11), $\mathbf{N}_p$ *is defined in* (5), *and* $\sigma_{ii\cdot\mathbf{k}}$ *is defined in* (11). *If* $\mathbf{k}$ *is empty, then* $\rho_{ij\cdot\mathbf{k}}$ *becomes the simple correlation* ρ_{ij} *and* $\boldsymbol{\gamma}_{i\cdot\mathbf{k}}$ *becomes* $\boldsymbol{\gamma}_i = \mathbf{e}_i^p / \sqrt{\sigma_{ii}}$. □

Multiple Correlations

Suppose that an investigator has interest in the multiple correlation between variable i and variables indexed by $\mathbf{k} = \begin{pmatrix} k_1 & k_2 & \cdots & k_q \end{pmatrix}'$. The squared multiple correlation, denoted as $\rho_{i(\mathbf{k})}^2$, can be written as

$$\rho_{i(\mathbf{k})}^2 = \mathbf{e}_i^{p\prime} \boldsymbol{\Sigma} \mathbf{H}_k \boldsymbol{\Sigma} \mathbf{e}_i^p / \sigma_{ii}, \qquad (12)$$

where $\mathbf{e}_i^p$ is defined in (2), and $\mathbf{H}_k$ is defined in (11). It is required that i not be a component of $\mathbf{k}$; otherwise $\rho_{i(\mathbf{k})}^2 = 1$ and no inference is necessary. The parameter $\rho_{i(\mathbf{k})}^2$ can be estimated by $r_{i(\mathbf{k})}^2$, where $r_{i(\mathbf{k})}^2$ is obtained by substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$ in (12).

Theorem 3 gives first and second derivatives of $\rho_{i(\mathbf{k})}^2$ with respect to $\boldsymbol{\sigma}$. The derivatives are given without proof because they follow from established matrix calculus results (Graham, 1981).

Theorem 3. *The first two derivatives of* $\rho_{i(\mathbf{k})}^2$ *with respect to* $\boldsymbol{\sigma}$ *are*

$$\dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2 \stackrel{\mathrm{def}}{=} \frac{\partial \rho_{i(\mathbf{k})}^2}{\partial \boldsymbol{\sigma}} = \left(\boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \right) \left(1 - \rho_{i(\mathbf{k})}^2 \right)$$

$$- \left(\boldsymbol{\gamma}_{i(\mathbf{k})} \otimes \boldsymbol{\gamma}_{i(\mathbf{k})} \right) \ \ and$$

$$\ddot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2 \stackrel{\mathrm{def}}{=} \frac{\partial^2 \rho_{i(\mathbf{k})}^2}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}}$$

$$= 2\mathbf{N}_{p^2} \left(\boldsymbol{\gamma}_{i(\mathbf{k})} \otimes \boldsymbol{\gamma}_{i(\mathbf{k})} \otimes \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \right)$$

$$+ 2 \left(\mathbf{N}_p \otimes \mathbf{N}_p \right) \left(\boldsymbol{\gamma}_{i(\mathbf{k})} \otimes \mathbf{h}_k \otimes \boldsymbol{\gamma}_{i(\mathbf{k})} \right)$$

$$- 2 \left(\boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \right) \left(1 - \rho_{i(\mathbf{k})}^2 \right),$$

$$where \ \boldsymbol{\gamma}_i = \mathbf{e}_i^p / \sqrt{\sigma_{ii}},$$

$$\boldsymbol{\gamma}_{i(\mathbf{k})} = \left(\mathbf{I}_p - \mathbf{H}_k \boldsymbol{\Sigma} \right) \mathbf{e}_i^p / \sqrt{\sigma_{ii}},$$

$\mathbf{H}_k$ *is defined in* (11), *and* $\mathbf{h}_k$ *is defined in Theorem 2.* □

Derivatives of Differences

In practice, a linear function of correlation coefficients rather than a correlation coefficient itself could be of interest. Let ψ be a linear function of simple, partial, and/or squared multiple correlation coefficients. For example, if the difference between a simple and a partial correlation is of interest, then $\psi = \rho_{ij} - \rho_{ij\cdot\mathbf{k}}$, where $\mathbf{k}$ is a non-empty vector of index values. By properties of derivatives, the derivative of a linear function is the the linear function of the derivatives. Accordingly, Theorems 2 and 3 can be used to obtain the first two derivatives of any linear function, ψ, with respect to $\boldsymbol{\sigma}$.

Denote the first two derivatives of ψ with respect to $\boldsymbol{\sigma}$ by $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ and $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$. That is,

$$\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \stackrel{\mathrm{def}}{=} \frac{\partial \psi}{\partial \boldsymbol{\sigma}} \ and \ \ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \stackrel{\mathrm{def}}{=} \frac{\partial^2 \psi}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}}, \quad (13)$$

where ψ is a linear function of simple, partial, and/or squared multiple correlation coefficients. The specific structure of these derivatives for individual correlation coefficients as well as for linear functions that correspond to extensions of the first four models in Olkin and Finn (1995) are listed below. The fifth model in Olkin and Finn is discussed separately later in this article.

1. Single Partial Correlation Coefficient. If $\psi = \rho_{ij\cdot\mathbf{k}}$, then $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}$, and $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \ddot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}$, where the derivatives are given in Theorem 2. If $\mathbf{k}$ is empty, then ψ becomes $\psi = \rho_{ij}$.

2. Difference Between Partial Correlation Coefficients. Denote the set that consists of the components of $\mathbf{k}$ by $\{\mathbf{k}\}$. If $\psi = \rho_{ij\cdot\mathbf{k}_1} - \rho_{tu\cdot\mathbf{k}_0}$, then $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}_1} - \dot{\boldsymbol{\rho}}_{tu\cdot\mathbf{k}_0}$, and $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \ddot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}_1} - \ddot{\boldsymbol{\rho}}_{tu\cdot\mathbf{k}_0}$, where $\mathbf{k}_1$ and/or $\mathbf{k}_0$ could be empty, $\{i,j\} = \{t,u\} \implies \{\mathbf{k}_0\} \neq \{\mathbf{k}_1\}$, and the derivatives are given in Theorem 2. This difference is an extension of Olkin and Finn's Models C and D. Examples include $\psi = \rho_{ij\cdot\mathbf{k}_1} - \rho_{ij\cdot\mathbf{k}_0}$, $\psi = \rho_{ij} - \rho_{ij\cdot\mathbf{k}}$, $\psi = \rho_{ij\cdot\mathbf{k}} - \rho_{tu\cdot\mathbf{k}}$, and $\psi = \rho_{ij} - \rho_{tu}$.

3. Single Squared Multiple Correlation Coefficient. If $\psi = \rho_{i(\mathbf{k})}^2$, then $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2$, and $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \ddot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2$, where the derivatives are given in Theorem 3.

4. Difference Between Squared Multiple Correlation Coefficients. If $\psi = \rho_{i(\mathbf{k}_1)}^2 - \rho_{j(\mathbf{k}_0)}^2$, then $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \dot{\boldsymbol{\rho}}_{i(\mathbf{k}_1)}^2 - \dot{\boldsymbol{\rho}}_{j(\mathbf{k}_0)}^2$ and $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} = \ddot{\boldsymbol{\rho}}_{i(\mathbf{k}_1)}^2 - \ddot{\boldsymbol{\rho}}_{j(\mathbf{k}_0)}^2$, where $i = j \implies \{\mathbf{k}_0\} \neq \{\mathbf{k}_1\}$, and the derivatives are given in Theorem 3. This difference is an extension Olkin and Finn's Models A and B.

Second-Order Accurate Confidence Intervals

Let ψ be a linear function of simple, partial, or squared multiple correlation coefficients and denote the estimator obtained by substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$ by $\hat{\psi}$. Define σ_ψ^2 as $\sigma_\psi^2 \stackrel{\text{def}}{=} n\,\mathrm{Var}(\hat{\psi})$. An application of the delta method reveals that

$$\sigma_\psi^2 = \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \boldsymbol{\Omega}_{22,n} \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} + O\left(n^{-1}\right), \qquad (14)$$

where $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ is defined in (13) and $\boldsymbol{\Omega}_{22,n}$ is defined in (5). The central limit theorem ensures that

$$\mathrm{P}\left(\widehat{\psi} \leq u\right) = \Phi\left(\frac{\sqrt{n}\,(u - \psi)}{\sigma_\psi}\right) + O\left(n^{-\frac{1}{2}}\right),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

The proposed confidence intervals are based on the asymptotically pivotal quantity

$$T = \frac{\sqrt{n}\left(\widehat{\psi} - \psi\right)}{\widehat{\sigma}_\psi}, \qquad (15)$$

where $\hat{\sigma}_\psi^2$ is a consistent estimator of σ_ψ^2. The quantity, $\dot{\boldsymbol{\psi}}_{\mathbf{s}}' \widehat{\boldsymbol{\Omega}}_{22,n} \dot{\boldsymbol{\psi}}_{\mathbf{s}}$, for example, is consistent for σ_ψ^2, where $\widehat{\boldsymbol{\Omega}}_{22,n}$ is defined in (6) and $\dot{\boldsymbol{\psi}}_{\mathbf{s}}$ is $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ after substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$. It follows from the central limit Theorem and Slutsky's Theorem (Casella & Berger, 2002, Theorem 5.5.17) that $T \sim \mathrm{N}(0,1)$ to first-order accuracy. Inverting the cdf of T yields first-order accurate one-sided lower and upper intervals:

$$(\widehat{\psi} - \widehat{\sigma}_\psi z_{1-\alpha}/\sqrt{n},\ \infty) \text{ and } (-\infty,\ \widehat{\psi} - \widehat{\sigma}_\psi z_\alpha/\sqrt{n})$$

respectively, where z_α is the 100α percentile of the standard normal distribution. The standard normal critical value z_α can be replaced by the t critical value, $t_{\alpha,n}$, without affecting the asymptotic properties of the intervals.

Edgeworth Expansion of the Distribution of T

To construct a second-order accurate approximation to the distribution of T, it is necessary to obtain the mean (bias), variance, and skewness of T. These quantities and estimators of these quantities are summarized in Theorem 4. A proof is sketched in the Appendix.

Theorem 4. *If ψ is a function of a single covariance matrix, $\mathbf{\Sigma}$, and $\hat{\psi}$ is the same function of a single sample covariance matrix, $\mathbf{S}$, then the first three cumulants of T in (15) can be written as follows:*

$$\mathrm{E}(T) = \frac{\kappa_1}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right),$$

$$\mathrm{Var}(T) = 1 + O\left(n^{-1}\right),$$

and $\mathrm{E}\left[\left(T - \frac{\kappa_1}{\sqrt{n}}\right)^3\right] = \frac{\kappa_3}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right),$

where $\kappa_1 = \dfrac{m_1}{\sigma_\psi} - \dfrac{m_{11}}{2\sigma_\psi^3},$

$$\kappa_3 = \frac{m_3}{\sigma_\psi^3} - 3\frac{m_{11}}{\sigma_\psi^3},$$

$$\frac{m_1}{\sqrt{n}} = \mathrm{E}\left[\sqrt{n}\left(\hat{\psi} - \psi\right)\right] + O\left(n^{-\frac{3}{2}}\right),$$

$$\frac{m_3}{\sqrt{n}}$$

$$= \mathrm{E}\left\{\left[\sqrt{n}\left(\hat{\psi} - \psi\right) - \frac{m_1}{\sqrt{n}}\right]^3\right\} + O\left(n^{-\frac{3}{2}}\right),$$

and m_{11}

$$= n\,\mathrm{Cov}\left(\hat{\psi} - \psi, \hat{\sigma}_\psi^2 - \sigma_\psi^2\right) + O\left(n^{-1}\right).$$

Specific solutions for m_1, m_3, and m_{11} are

$$m_1 = \frac{1}{2}\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\,\mathrm{vec}\,\boldsymbol{\Omega}_{22,n},$$

$$m_3 = \sqrt{n}\left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)'\mathrm{vec}\,\boldsymbol{\Omega}_{222,n}$$

$$+3\left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\Omega}_{22,n} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\Omega}_{22,n}\right)\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}},\ \text{and}$$

$$m_{11} = \left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)'\mathrm{vec}\,\boldsymbol{\Omega}_{42,n}$$

$$+2\left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\Omega}_{22,n} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\Omega}_{22,n}\right)\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}},$$

where $\boldsymbol{\Omega}_{42,n}$ *and* $\boldsymbol{\Omega}_{222,n}$ *are described in* (7) *and* (8), *and* $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ *and* $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ *are defined in* (13). *Furthermore,* $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\sigma} = \mathbf{0}$ *as required in Theorem 1 and κ_1 and κ_2 can be estimated consistently by*

$$\widehat{\kappa}_1 = \frac{\widehat{m}_1}{\widehat{\sigma}_\psi} - \frac{\widehat{m}_{11}}{2\widehat{\sigma}_\psi^3},\ \ \text{and}\ \ \widehat{\kappa}_3 = \frac{\widehat{m}_3}{\widehat{\sigma}_\psi^3} - 3\frac{\widehat{m}_{11}}{\widehat{\sigma}_\psi^3},$$

where $\widehat{\sigma}_\psi^2 = \dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Omega}}_{22,n}\dot{\boldsymbol{\psi}}_{\mathbf{s}} + \dfrac{U}{n},$

$$\widehat{m}_1 = \frac{1}{2}\ddot{\boldsymbol{\psi}}_{\mathbf{s}}'\,\mathrm{vec}\,\widehat{\boldsymbol{\Omega}}_{22,n},$$

$$\widehat{m}_3 = n^{-1}\sum_{i=1}^{N}\left[\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\left(\widetilde{\boldsymbol{\varepsilon}}_i \otimes \widetilde{\boldsymbol{\varepsilon}}_i\right)\right]^3$$

$$-6\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Upsilon}}_{21}\dot{\boldsymbol{\Psi}}_{\mathbf{s}}\widehat{\boldsymbol{\Upsilon}}_{21}'\dot{\boldsymbol{\psi}}_{\mathbf{s}}$$

$$+3\left(\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Omega}}_{22,n} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Omega}}_{22,n}\right)\ddot{\boldsymbol{\psi}}_{\mathbf{s}},$$

$$\widehat{m}_{11} = n^{-1}\sum_{i=1}^{N}\left[\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\left(\widetilde{\boldsymbol{\varepsilon}}_i \otimes \widetilde{\boldsymbol{\varepsilon}}_i\right)\right]^3$$

$$-4\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Upsilon}}_{21}\dot{\boldsymbol{\Psi}}_{\mathbf{s}}\widehat{\boldsymbol{\Upsilon}}_{21}'\dot{\boldsymbol{\psi}}_{\mathbf{s}}$$

$$+2\left(\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Omega}}_{22,n} \otimes \dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Omega}}_{22,n}\right)\ddot{\boldsymbol{\psi}}_{\mathbf{s}},$$

$\dot{\boldsymbol{\psi}}_{\mathbf{s}}$ *and* $\ddot{\boldsymbol{\psi}}_{\mathbf{s}}$ *are* $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ *and* $\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ *after substituting* $\mathbf{S}$ *for* $\mathbf{\Sigma}$, U *is any* $O_p(1)$ *random variable, and the remaining terms are defined in Theorem 1.* $\square$

Asymptotically, the choice of U in Theorem 4 is arbitrary, because $\lim_{n\to 0} U/n = 0$ and the asymptotic properties of the intervals are not affected by U/n. Nonetheless, the finite-sample properties of the intervals are affected by the choice of U. From experience, it appears that $\dot{\boldsymbol{\psi}}_{\mathbf{s}}'\widehat{\boldsymbol{\Omega}}_{22,n}\dot{\boldsymbol{\psi}}_{\mathbf{s}}$ often

underestimates the variance of $\sqrt{n}(\hat{\psi} - \psi)$. Accordingly, choosing U to be a positive quantity might be beneficial. In this article, the quantity U will be chosen by using the second order Taylor series expansion of $\hat{\psi}$ under the condition that $\dot{\psi}_{\boldsymbol{\sigma}} = \mathbf{0}$. Specifically, the second-order Taylor series expansion of $\hat{\psi}$ is

$$\sqrt{n}\left(\widehat{\psi} - \psi\right) = \dot{\psi}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$$

$$+\frac{1}{2\sqrt{n}}\ddot{\psi}_{\boldsymbol{\sigma}}'\left[\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \otimes \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]$$

$$+O_p\left(n^{-1}\right).$$

If $\dot{\psi}_{\boldsymbol{\sigma}} = \mathbf{0}$, then the variance of $\sqrt{n}(\hat{\psi} - \psi)$ is

$$\mathrm{Var}\left[\sqrt{n}\left(\widehat{\psi} - \psi\right)\right]$$

$$= \frac{1}{2n}\ddot{\psi}_{\boldsymbol{\sigma}}'\left(\boldsymbol{\Omega}_{22,n} \otimes \boldsymbol{\Omega}_{22,n}\right)\ddot{\psi}_{\boldsymbol{\sigma}} + O\left(n^{-2}\right).$$

Accordingly, U will be chosen as $U = \ddot{\psi}_{\mathbf{s}}'\left(\widehat{\boldsymbol{\Omega}}_{22,n} \otimes \widehat{\boldsymbol{\Omega}}_{22,n}\right)\ddot{\psi}_{\mathbf{s}}/2$. For this choice, the quantity U/n represents an estimate of one of the two leading terms in the $O(n^{-1})$ component of (14).

The primary device for constructing the interval estimator is a second-order Edgeworth expansion of the distribution of T. This expansion is summarized in Theorem 5. A proof is sketched in the Appendix.

Theorem 5. *The probability density function (pdf) and the cdf of T in (15) can be written as*

$$f_T(t) =$$

$$\varphi(t)\left[1 + \frac{\kappa_1 t}{\sqrt{n}} + \frac{\kappa_3\left(t^3 - 3t\right)}{6\sqrt{n}}\right] + O\left(n^{-1}\right)$$

$$and \ F_T(t) \stackrel{\text{def}}{=} P(T \le t) = \Phi(t)$$

$$-\varphi(t)\left[\frac{\kappa_1}{\sqrt{n}} + \frac{\kappa_3\left(t^2 - 1\right)}{6\sqrt{n}}\right] + O\left(n^{-1}\right)$$

respectively, where $\kappa_1/\sqrt{n}$ and $\kappa_3/\sqrt{n}$ are the mean and skewness of T (see Theorem 4), and $\varphi(t)$ and $\Phi(t)$ are the standard normal pdf and cdf. $\square$

Proposed Interval Estimators

The expansion of F_T in Theorem 5 can be used to express the percentiles of T in terms of those of a standard normal random variable and to obtain a polynomial transformation of T that has distribution $N(0, 1)$ up to $O(n^{-1})$. These results, known as Cornish-Fisher expansions (Pace and Salvan, 1997, §10.6), are summarized in Corollary 5.1.

Corollary 5.1. *Denote the 100α percentile of T by t_α. Then,*

$$t_\alpha = z_\alpha + \frac{\kappa_1}{\sqrt{n}} + \frac{\kappa_3(z_\alpha^2 - 1)}{6\sqrt{n}} + O\left(n^{-1}\right) \ and$$

$$t_\alpha = \widehat{t}_\alpha + O_p\left(n^{-1}\right),$$

where z_α is the 100α percentile of the $N(0, 1)$ distribution,

$$\widehat{t}_\alpha = z_\alpha + \widehat{\kappa}_1/\sqrt{n} + \widehat{\kappa}_3(z_\alpha^2 - 1)/(6\sqrt{n}),$$

and $\widehat{\kappa}_1$ and $\widehat{\kappa}_3$ are consistent estimators of κ_1 and κ_3 in Theorem 4. The normal critical value, z_α, can be replaced by the t critical value, $t_{\alpha,n}$, without affecting the second-order property of $\hat{t}_\alpha$. Also, define T_1 and $\hat{T}_1$ as

$$T_1 \stackrel{\text{def}}{=} T - \frac{\kappa_1}{\sqrt{n}} - \frac{\kappa_3(T^2 - 1)}{6\sqrt{n}} \ and$$

$$\widehat{T}_1 \stackrel{\text{def}}{=} T - \frac{\widehat{\kappa}_1}{\sqrt{n}} - \frac{\widehat{\kappa}_3(T^2 - 1)}{6\sqrt{n}}.$$

Then, $P(T_1 \le t) = \Phi(t) + O\left(n^{-1}\right)$ and $P(\hat{T}_1 \le t) = \Phi(t) + O\left(n^{-1}\right)$. $\square$

A straightforward application of Corollary 5.1 yields second-order accurate intervals for ψ. Note that $\hat{t}_\alpha$ in Corollary 5.1 is a second-order accurate percentile of the distribution of T. Accordingly, $(\hat{\psi} - \hat{t}_{1-\alpha}\hat{\sigma}_\psi/\sqrt{n}, \infty)$ is a second-order accurate $100(1-\alpha)\%$ lower confidence interval for ψ and $(-\infty, \hat{\psi} - \hat{t}_\alpha\hat{\sigma}_\psi/\sqrt{n})$ is a second-order accurate $100(1-\alpha)\%$ upper confidence interval for ψ. One drawback of these intervals, however, is that $\hat{t}_\alpha$ is a quadratic function of z_α and, therefore, is not monotonic in α. The same issue exists if $\hat{T}_1$ is used as an asymptotic pivotal quantity. It is not monotone in T and, therefore, its cdf cannot be inverted for arbitrary α. Methods for correcting this non-monotone deficiency are summarized in Corollaries 5.2 and 5.3. The method in Corollary 5.2 is based on Hall's (1992) cubic transformation, whereas the method in Corollary 5.3 is based on the exponential transformation described by Boik (2006).

Corollary 5.2. *Define* $\hat{T}_2$ *as*

$$\widehat{T}_2 \stackrel{\text{def}}{=} T - \frac{\widehat{\kappa}_1}{\sqrt{n}} - \frac{\widehat{\kappa}_3(T^2-1)}{6\sqrt{n}} + \frac{\widehat{\kappa}_3^2 T^3}{108n}.$$

Then, $\hat{T}_2 = \hat{T}_1 + O_p(n^{-1})$ *and* $\hat{T}_2$ *is monotone in* T. *Inverting the cdf of* $\hat{T}_2$ *yields the following lower and upper* $100(1-\alpha)\%$ *second-order accurate confidence intervals:*

$$\left(\widehat{\psi} - \frac{\widehat{\sigma}_\psi \hat{t}_{2,1-\alpha}}{\sqrt{n}}, \infty\right) \ and$$

$$\left(-\infty, \widehat{\psi} - \frac{\widehat{\sigma}_\psi \hat{t}_{2,\alpha}}{\sqrt{n}}\right),$$

where $\hat{t}_{2,\alpha} = \dfrac{\widehat{\kappa}_1}{\sqrt{n}} + \dfrac{6\sqrt{n}}{\widehat{\kappa}_3} \times$

$$\left\{ 1 - \left[1 + \frac{\widehat{\kappa}_3}{2\sqrt{n}}\left(\frac{\widehat{\kappa}_3}{6\sqrt{n}} - z_\alpha \right) \right]^{\frac{1}{3}} \right\}. \quad \square$$

Corollary 5.3. *Define* $\hat{T}_3$ *as*

$$\widehat{T}_3 \stackrel{\text{def}}{=} T - \frac{\widehat{\kappa}_1}{\sqrt{n}} - \frac{\widehat{\kappa}_3\left(T^2 e^{-\frac{\hat{d}T^2}{2}} - 1\right)}{6\sqrt{n}}, \ where$$

$$\widehat{d} = \frac{\widehat{\kappa}_3^2\left(31 - 7\sqrt{17}\right)}{72n} e^{-\frac{(5-\sqrt{17})}{2}}.$$

Then, $\hat{T}_3 = \hat{T}_1 + O_p(n^{-3/2})$, *and is monotone in* T. *Inverting the cdf of* $\hat{T}_3$ *yields the following lower and upper* $100(1-\alpha)\%$ *second-order accurate confidence intervals:*

$$\left(\widehat{\psi} - \frac{\widehat{\sigma}_\psi \hat{t}_{3,1-\alpha}}{\sqrt{n}}, \infty\right) \ and$$

$$\left(-\infty, \widehat{\psi} - \frac{\widehat{\sigma}_\psi \hat{t}_{3,\alpha}}{\sqrt{n}}\right),$$

where $\hat{t}_{3,\alpha}$ *is the solution to* $\hat{T}_3 = z_\alpha$ *or* $\hat{T}_3 = t_{\alpha,n}$ *for* T. *The solution can be computed using the modified Newton method described in the Appendix.* $\square$

The intervals described in Corollaries 5.2 and 5.3 require consistent estimators of κ_1 and κ_3, the mean and skewness of T. If ψ is a function of a single covariance matrix, then the estimators described in Theorem 4 can be used. In some situations, investigators are interested in comparing correlation functions based on two covariance matrices. For example, if a comparison between squared multiple correlation coefficients from two populations is of interest, then one could define ψ as $\psi = \rho_{i(\mathbf{k})}^2(\mathbf{\Sigma}_1) - \rho_{i(\mathbf{k})}^2(\mathbf{\Sigma}_2)$, where $\rho_{i(\mathbf{k})}^2(\mathbf{\Sigma}_u)$ is $\rho_{i(\mathbf{k})}^2$ computed on $\mathbf{\Sigma}_u$, for $u = 1, 2$. Of course, the comparison need not be restricted to squared multiple correlations. One could, for example, define ψ as $\psi = \rho_{ij\cdot\mathbf{k}}(\mathbf{\Sigma}_1) - \rho_{ij\cdot\mathbf{k}}(\mathbf{\Sigma}_2)$ or $\psi = \rho_{ij\cdot\mathbf{k}}(\mathbf{\Sigma}_1) - \rho_{tu\cdot\mathbf{k}_0}(\mathbf{\Sigma}_2)$. These comparisons are extensions of Olkin and Finn's (1995) Model E. The mean and skewness

as well as estimators of the mean and skewness of T for comparisons such as these are given in Theorem 6. A proof is sketched in the Appendix.

Theorem 6. *Let ψ_i be a correlation function computed on $\boldsymbol{\Sigma}_i$ for $i = 1, 2$ and let $\hat{\psi}_i$ be ψ_i computed on $\mathbf{S}_i$, where $\mathbf{S}_i$ is a sample covariance matrix based on n_i degrees of freedom from sample i, and samples 1 and 2 are independent. Define T as*

$$T = \frac{\sqrt{n_1 n_2}\left(\widehat{\psi} - \psi\right)}{\sqrt{\hat{\delta}}}, \ where$$

$$\widehat{\psi} = \widehat{\psi}_1 - \widehat{\psi}_2, \quad \psi = \psi_1 - \psi_2,$$

$\hat{\delta} = n_2\hat{\sigma}_1^2 + n_1\hat{\sigma}_2^2$, and $\hat{\sigma}_i^2$, $\dot{\boldsymbol{\psi}}_{\mathbf{s}_i}$, and $\ddot{\boldsymbol{\psi}}_{\mathbf{s}_i}$ are $\hat{\sigma}_\psi^2$, $\dot{\boldsymbol{\psi}}_{\mathbf{s}}$, and $\ddot{\boldsymbol{\psi}}_{\mathbf{s}}$ based on sample i. Let $m_{1;i}$, $m_{3;i}$, and $m_{11;i}$ be m_1, m_3, and m_{11} of Theorem 4 for population i. Then, the first three cumulants of T are

$$\mathrm{E}(T) = \frac{\kappa_1^*}{\sqrt{\delta}} + O\left(n^{-\frac{3}{2}}\right),$$

$$\mathrm{Var}(T) = 1 + O\left(n^{-1}\right), \ and$$

$$\mathrm{E}\left[T - \mathrm{E}(T)\right]^3 = \frac{\kappa_3^*}{\sqrt{\delta}} + O\left(n^{-\frac{3}{2}}\right),$$

$$where \ \delta = n_2\sigma_1^2 + n_1\sigma_2^2,$$

$$\kappa_1^* = \frac{n_2 m_{1;1} - n_1 m_{1;2}}{\sqrt{n_1 n_2}} - \frac{n_2^{3/2} m_{11;1}}{2\delta\sqrt{n_1}}$$

$$+ \frac{n_1^{3/2} m_{11;2}}{2\delta\sqrt{n_2}},$$

$$\kappa_3^* = \frac{n_2^{3/2} m_{3;1}}{\delta\sqrt{n_1}} - \frac{n_1^{3/2} m_{3;2}}{\delta\sqrt{n_2}} - 3\frac{n_2^{5/2}\sigma_1^2 m_{11;1}}{\delta^2\sqrt{n_1}}$$

$$+ 3\frac{n_1^{5/2}\sigma_2^2 m_{11;2}}{\delta^2\sqrt{n_2}},$$

and $n = \min(n_1, n_2)$. Furthermore, consistent estimators of κ_1^* and κ_3^* are obtained by substituting $\hat{\sigma}_i^2$, $\hat{m}_{1;i}$, $\hat{m}_{3;i}$, and $\hat{m}_{11;i}$ for σ_i^2, $m_{1;i}$, $m_{3;i}$, and $m_{11;i}$, respectively. Estimators of $m_{1;i}$, $m_{3;i}$, and $m_{11;i}$ are given by $\hat{m}_1$, $\hat{m}_3$, and $\hat{m}_{11}$ in Theorem 4 computed on sample i. □

Second-order accurate confidence intervals for ψ in Theorem 6 can be computed by substituting $\hat{\kappa}_1^*/\sqrt{\hat{\delta}}$ and $\hat{\kappa}_3^*/\sqrt{\hat{\delta}}$ for $\hat{\kappa}_1/\sqrt{n}$ and $\hat{\kappa}_3/\sqrt{n}$ in Corollaries 5.2 and 5.3.

Simulation Studies

The finite sample properties of the proposed intervals were evaluated by computing the intervals on samples from normal, normal mixture, and affine lognormal distributions. Let $\mathbf{C}$ be a $p \times p$ nonsingular matrix and denote a randomly selected p-vector from the $\mathrm{N}(\mathbf{0}, \mathbf{I}_p)$ distribution as $\mathbf{z}$. Denote a randomly selected p-vector from a normal, normal mixture, or affine lognormal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma} = \mathbf{CC}'$ as $\mathbf{y}$. Then $\mathbf{z}$ can be transformed to $\mathbf{y}$ by

$$\mathbf{y} = \mathbf{Cz}, \quad \mathbf{y} = \mathbf{Cz}\frac{[U + k(1 - U)]}{\sqrt{\theta + k^2(1 - \theta)}}, \ and$$

$$\mathbf{y} = \mathbf{C}\left(e^{\odot \mathbf{z}m} - \mathbf{1}_p e^{m^2/2}\right)\frac{1}{\sqrt{e^{m^2}\left(e^{m^2} - 1\right)}}$$

for the normal, normal mixture, and affine lognormal distributions, respectively; where U is a Bernoulli(θ) random variable distributed independently of $\mathbf{z}$, k and m are scalar constants, and $e^{\odot \mathbf{z}m}$ is a p-vector whose i^{th} component is $e^{z_i m}$.

The parameter values for the normal mixture were $\theta = 0.7$ and $k = 3$. For this choice, marginal kurtosis is 3.488 and multivariate kurtosis (Bilodeau and Brenner, 1999, §13.2) is 1.1626. The normal

mixture is a member of the class of elliptically contoured distributions. Accordingly, asymptotic standard errors based on normal theory are smaller than the correct standard errors by a factor of $1/\sqrt{2.1626}$ and coverage of one-sided nominal $100(1-\alpha)\%$ normal theory intervals converges to $\Phi(z_{1-\alpha}/\sqrt{2.1626})$, where $z_{1-\alpha}$ is the $100(1-\alpha)$ percentile of the standard normal distribution. For $\alpha = 0.05$, coverage of normal theory intervals converges to 0.868 rather than 0.95.

The parameter value for the affine lognormal distribution was $m = 0.8326$. For this choice, the marginal skewness and kurtosis for each component of the lognormal random vector $e^{\odot \mathbf{z}m}$ is 4.00 and 38. This distribution is not a member of the class of elliptically contoured distributions.

Generation of Covariance Matrices

To examine the performance of intervals for $\psi = \rho_{ij} - \rho_{ij\cdot\mathbf{k}}$, ten 4-dimensional covariance matrices were constructed such that $\rho_{ij} \in \mathcal{L}_1$, $\rho_{ij\cdot\mathbf{k}} \in \mathcal{L}_1$, and $\rho_{ij} \leq \rho_{ij\cdot\mathbf{k}}$, where $\mathcal{L}_1 = \{\sqrt{1/6}, \sqrt{1/3}, \sqrt{2/3}, \sqrt{5/6}\}$. The covariance matrices were constructed as follows:

$$\mathbf{\Sigma} = \mathbf{CC}', \text{ where } \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & v & w & 1 \end{pmatrix},$$

$$w = \frac{-1 \pm \sqrt{\rho_{ij\cdot\mathbf{k}}^2 \left(1 - \rho_{ij\cdot\mathbf{k}}^2\right)}}{1 - 2\rho_{ij\cdot\mathbf{k}}^2},$$

$$v = \frac{-(2+w) \pm 2\sqrt{h}}{1 - 4\rho_{ij}^2},$$

$$h = 2\rho_{ij}^2(3 + 2w + w^2 - 4\rho_{ij}^2 - 2\rho_{ij}^2 w^2),$$

$i = 3$, $j = 4$, and $\mathbf{k} = \begin{pmatrix} 1 & 2 \end{pmatrix}'$.

To examine the performance of intervals for $\psi = \rho_{i(\mathbf{k})}^2 - \rho_{j(\mathbf{k})}^2$, ten 4-dimensional

covariance matrices were constructed such that $\rho_{i(\mathbf{k})}^2 \in \mathcal{L}_2$, $\rho_{j(\mathbf{k})}^2 \in \mathcal{L}_2$, and $\rho_{j(\mathbf{k})}^2 \leq \rho_{i(\mathbf{k})}^2$, where $\mathcal{L}_2 = \{0.2, 0.4, 0.6, 0.8\}$. The covariance matrices were constructed as follows:

$$\mathbf{\Sigma} = \mathbf{CC}', \text{ where } \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ v & v & 1 & 0 \\ w & w & 1 & 1 \end{pmatrix},$$

$$v = \sqrt{\frac{\rho_{i(\mathbf{k})}^2}{2\left(1 - \rho_{i(\mathbf{k})}^2\right)}}, \quad w = \sqrt{\frac{\rho_{j(\mathbf{k})}^2}{1 - \rho_{j(\mathbf{k})}^2}},$$

$i = 3$, $j = 4$, and $\mathbf{k} = \begin{pmatrix} 1 & 2 \end{pmatrix}'$.

For each covariance matrix, 5,000 realizations of the $N \times 4$ matrix $\mathbf{Y}$ were sampled from each of the three distributions (normal, normal mixture, affine lognormal) and one-sided nominal 95% confidence intervals for ψ were computed. The simulation study was repeated for sample sizes $N \in \{25, 50, 100, 200\}$.

Results

Each of Figures 1–3 displays four panels of 16 plots each. The upper two panels display the coverage of normal theory and ADF confidence intervals for $\psi = \rho_{ij} - \rho_{ij\cdot\mathbf{k}}$. Coverage of one-sided intervals for $\psi = \rho_{i(\mathbf{k})}^2 - \rho_{j(\mathbf{k})}^2$ is displayed in the lower two panels. The 10 plots in the upper triangle of each panel display the coverage of upper one-sided intervals and the 10 plots in the lower triangle display the coverage of lower one-sided intervals. The four plots on the diagonal in each panel display coverage for both lower and upper intervals.

In the upper two panels, row and column labels correspond to ρ_{ij}^2 and $\rho_{ij\cdot\mathbf{k}}^2$, respectively, for upper intervals and to $\rho_{ij\cdot\mathbf{k}}^2$ and ρ_{ij}^2, respectively, for lower intervals. In the lower two panels, row and column labels correspond to $\rho_{j(\mathbf{k})}^2$ and $\rho_{i(\mathbf{k})}^2$, respectively, for upper intervals and to $\rho_{i(\mathbf{k})}^2$ and

$\rho^2_{j(\mathbf{k})}$, respectively, for lower intervals. In all panels, plot symbols are assigned as follows:

Interval	Symbol
First-Order Lower	$\times$
First-Order Upper	$*$
Second-Order Lower	$\square$
Second-Order Upper	$\bigcirc$

Second-order intervals were based on the exponential method described in Corollary 5.3. Within each plot, the coverage of intervals based on sample sizes of 25, 50, 100, and 200 is plotted from left to right. Figures 2 and 3 display coverage of intervals when sampling from normal mixtures and from affine lognormal distributions.

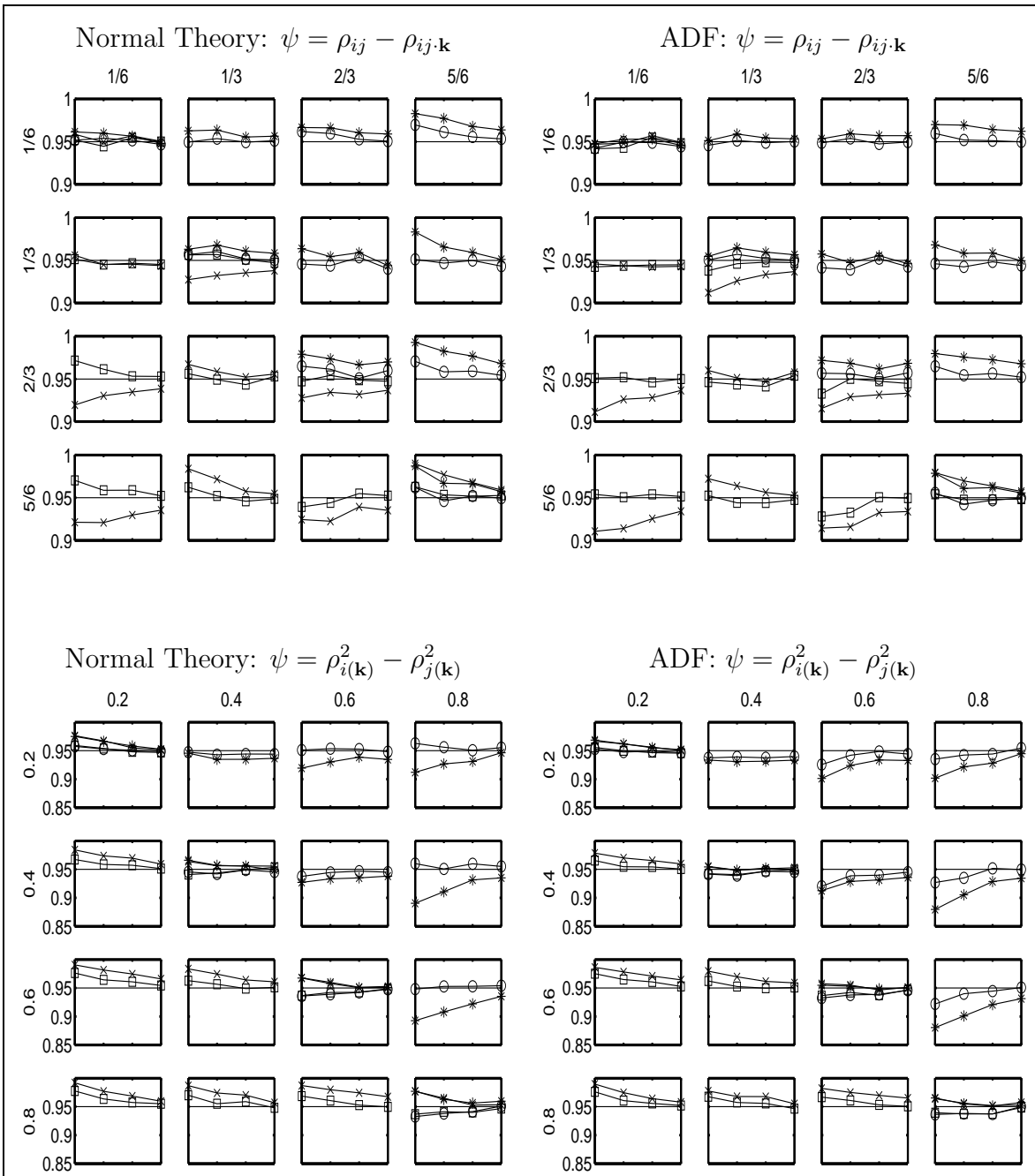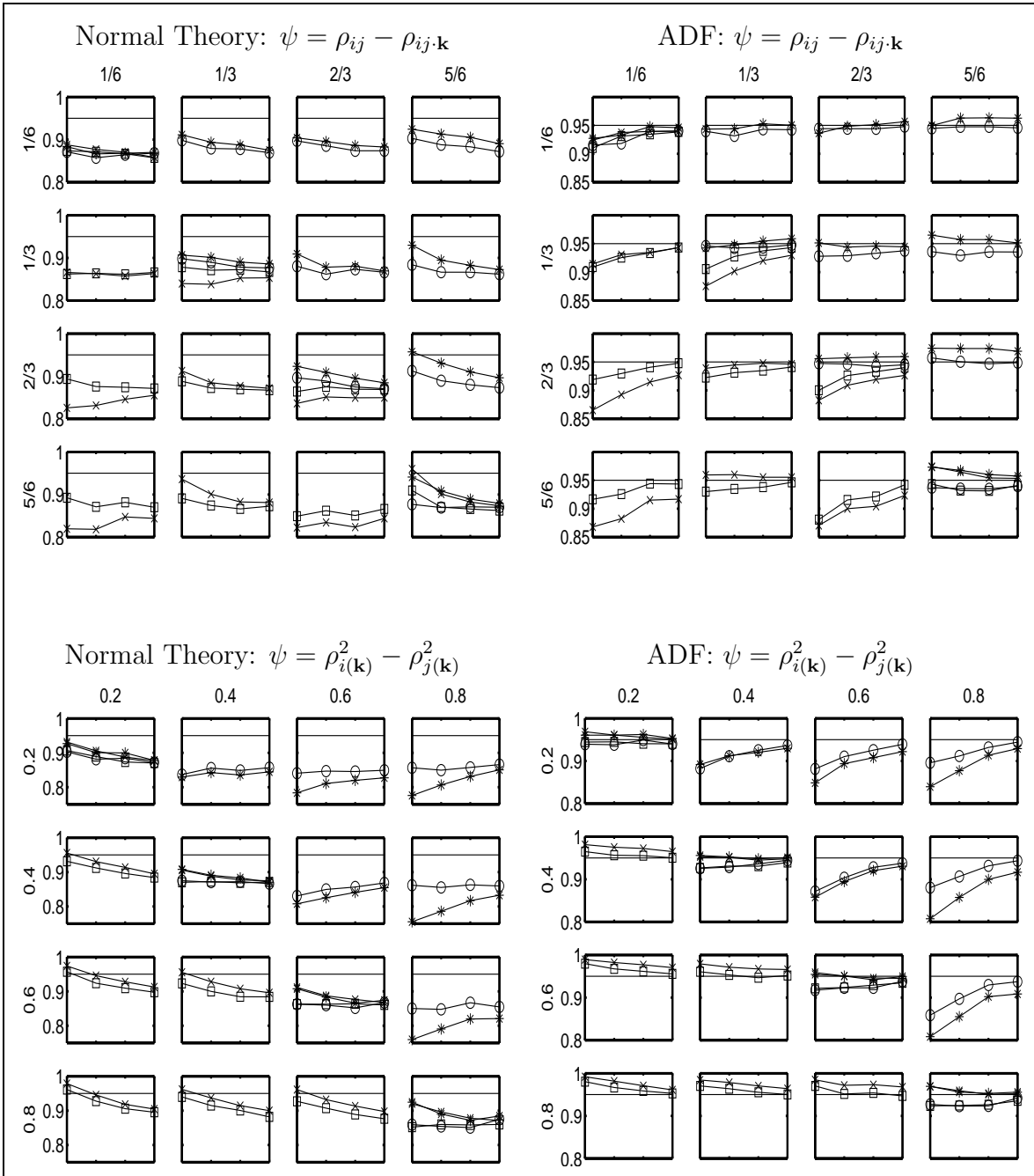Figure 1: Coverage when Sampling from Normal Distributions

Figure 2: Coverage when Sampling from Normal Mixture Distributions

It is apparent in Figure 1 that the coverage of second-order intervals is superior to that of first-order intervals when sampling from multivariate normal distributions. Furthermore, the second-order ADF intervals perform nearly as well as do the second-order intervals that are explicitly based on normal theory.

Figure 2 confirms that normal theory based intervals perform poorly when sampling from normal mixtures. As expected from theory, coverage approaches 0.86 as N increases in each plot. The ADF intervals fare much better. Also, the coverage of second-order accurate ADF intervals is equal to or superior to that of first-order

Figure 3: Coverage when Sampling from Affine Lognormal Distributions
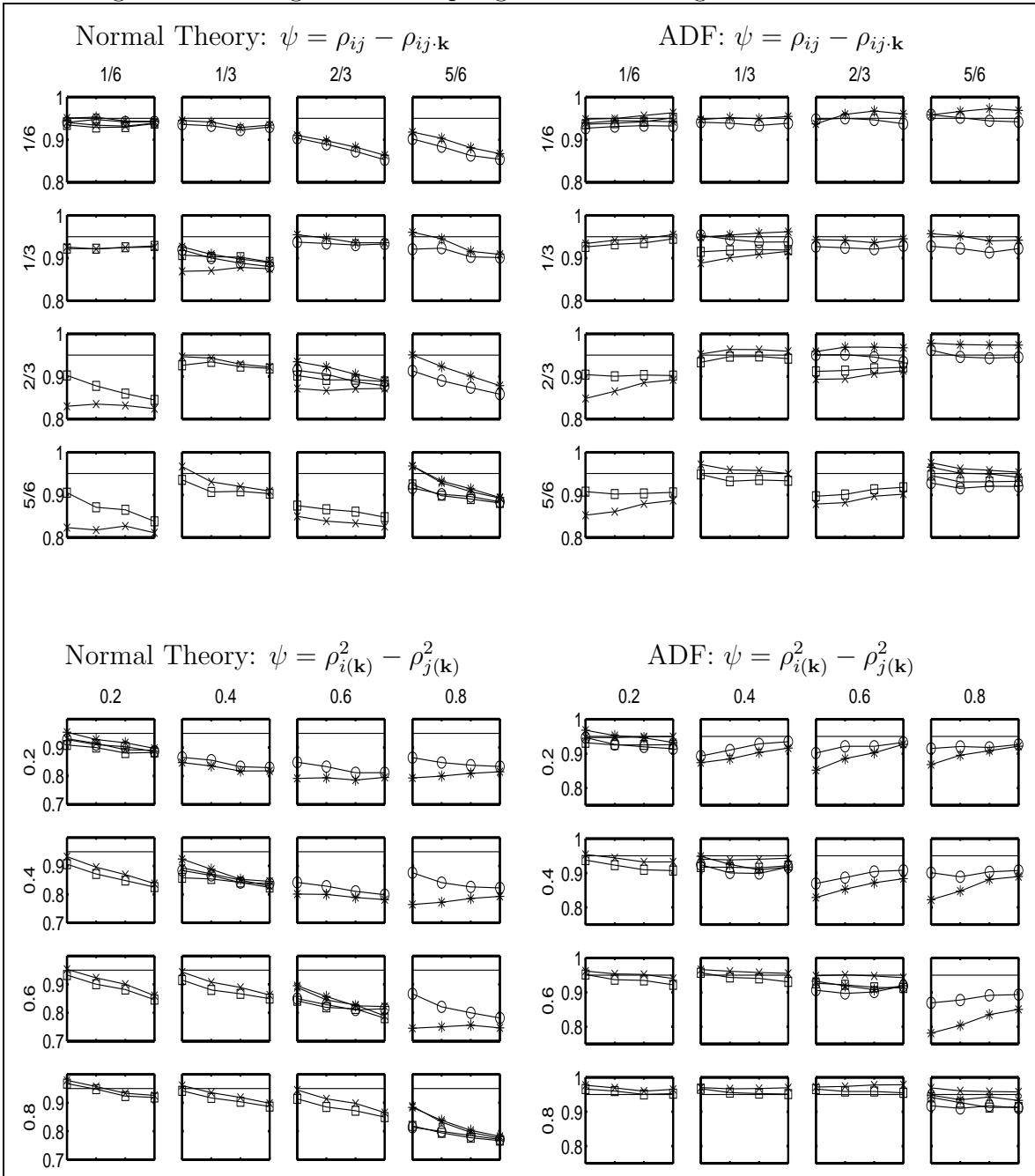
accurate ADF intervals whenever the coverage of first-order accurate intervals departs substantially from 0.95. See, for example, plot $(4, 1)$ in the upper right-hand panel of Figure 2. First-order accurate ADF intervals perform well if the bias of $\hat{\psi}$ is small and the distribution of $\hat{\psi}$ is nearly symmetric. In these cases, second-order accurate intervals cannot improve on first-order accurate intervals and can even perform slightly worse if sample size is small. The performance degradation is due to the difficulty of estimating higher-order moments from small samples. Variability in these estimates induces variability in the interval endpoints. Nonetheless, Figure 2 reveals

that the second-order ADF intervals never performed much worse and sometimes performed much better than first-order accurate intervals.

Figure 3 displays coverage of one-sided intervals when sampling from affine lognormal distributions. The performance of normal theory intervals is quite poor. The performance of first-order ADF intervals varies depending on the covariance matrix and coverage is as low as 0.78; see plot $(3, 4)$ in the lower right-hand panel. Figure 3 also reveals that second-order ADF intervals never performed much worse and some-

times performed much better than first-order accurate intervals.

Illustration

Heller, Judge, and Watson (2002) conducted a survey of university employees to examine relationships among personality traits, job satisfaction, and life satisfaction. They obtained six broad personality trait measures, three job satisfaction measures, and three life satisfaction measures on each of $N = 134$ subjects. A subset of their data from a random sample of 15 subjects is listed in Table 1.

Table 1: Sample of Size $N = 15$ from Heller

	Variable							
	1	2	3	4	5	6	7	8
Case	N	E	C	CSE	PA	NA	LSO	JSO
1	2.75	2.50	3.92	2.44	3.80	2.30	3.00	3.25
1	1.83	4.42	4.25	2.55	4.10	1.30	4.20	4.63
3	2.00	3.25	4.83	2.92	4.00	1.40	2.60	3.88
4	2.17	2.50	4.17	2.53	3.80	1.50	3.60	3.88
5	2.92	3.50	4.08	1.96	3.10	2.20	4.00	3.25
6	1.58	3.67	3.67	2.89	4.00	1.30	4.00	4.00
7	2.00	3.67	4.00	2.68	4.20	1.50	4.00	4.38
8	1.58	4.17	4.75	2.90	4.00	1.30	2.20	2.38
9	2.58	3.17	4.08	2.34	2.80	1.30	3.40	4.38
10	2.92	3.00	3.92	2.21	3.40	1.50	3.60	3.13
11	2.33	2.08	4.67	2.17	2.40	1.70	3.00	3.88
12	2.25	3.67	4.92	2.74	3.90	1.60	3.80	3.00
13	2.08	3.33	3.50	2.39	3.30	1.60	4.00	3.88
14	3.92	2.42	3.00	1.60	2.50	3.10	3.60	4.13
15	1.67	4.17	4.58	2.93	4.30	1.60	4.00	3.63

Notes: N = Neuroticism, E = Extroversion,
C = Conscientiousness, CSE = Core Self-Evaluations,
PA = Positive Affectivity, NA = Negative Affectivity,
LSO = Life Satisfaction—Significant Other Report
JSO = Job Satisfaction—Significant Other Report.

Appreciation is expressed to Danny Heller who provided the raw data for this example. They expected that job and life satisfaction would be positively related, and

that this relationship, in part, would be due to the influence of stable personality traits. Specifically, they hypothesized that (a) zero-order correlation coefficients be-

tween job and life satisfaction would be positive, and that (b) partial correlation coefficients between job and life satisfaction, controlling for personality traits, would be smaller than the zero-order coefficients.

Table 2: Confidence Intervals Computed on Heller's Data

ψ	Quantity	Normal Theory		ADF	
		$N = 15$	$N = 134$	$N = 15$	$N = 134$
	$r_{7,8}$	0.476	0.477	0.476	0.477
	$r_{7,8\cdot\mathbf{k}}$	0.410	0.398	0.410	0.398
	$\hat{\sigma}_\psi$	0.889	0.473	0.966	0.562
	$\hat{\omega}_1$	-0.086	-0.821	-0.281	-1.597
	$\hat{\omega}_3$	-0.183	-2.171	-0.082	-3.597
$\rho_{7,8} - \rho_{7,8\cdot\mathbf{k}}$	$\hat{t}_{3,0.05}$	-1.803	-1.797	-1.845	-1.937
	$\hat{t}_{3,0.95}$	1.722	1.543	1.679	1.459
	$L(1)$	-0.353	0.110	-0.389	-0.001
	$U(1)$	0.484	0.147	0.520	0.160
	$L(2)$	-0.344	0.016	-0.368	0.008
	$U(2)$	0.494	0.153	0.542	0.174
	$t_{0.05,n}$	-1.761	-1.656	-1.761	-1.656
	$t_{0.95,n}$	1.761	1.656	1.761	1.656
	$r^2_{7(\mathbf{k})}$	0.468	0.249	0.468	0.249
	$r^2_{8(\mathbf{k})}$	0.263	0.109	0.263	0.109
	$\hat{\sigma}_\psi$	1.191	0.818	1.343	0.822
	$\hat{\omega}_1$	-0.269	-0.110	-0.347	-0.212
	$\hat{\omega}_3$	0.095	-0.010	-0.147	-0.124
$\rho^2_{7(\mathbf{k})} - \rho^2_{8(\mathbf{k})}$	$\hat{t}_{3,0.05}$	-1.984	-1.771	-2.203	-1.924
	$\hat{t}_{3,0.95}$	1.512	1.543	1.392	1.423
	$L(1)$	-0.355	0.023	-0.426	0.023
	$U(1)$	0.766	0.258	0.838	0.259
	$L(2)$	-0.276	0.031	-0.294	0.039
	$U(2)$	0.837	0.266	0.996	0.278

$L(1)$ and $U(1)$ are lower and upper limits of first-order accurate intervals.
$L(2)$ and $U(2)$ are lower and upper limits of second-order accurate intervals.

One way to test hypothesis (b) is to compute a confidence interval for the difference between the simple and partial correlation coefficients. The upper portion of Table 2 displays first- and second-order accurate intervals for $\rho_{7,8} - \rho_{7,8\cdot\mathbf{k}}$, where $\mathbf{k}$ consists of indices $1, \ldots, 6$. This interval was not examined by Heller, Judge, and Watson, but is comparable to intervals that were examined. Intervals based on the subsample of $N = 15$ cases and intervals based on the full sample of $N = 134$ cases are reported in Table 2. Results of intermediate calculations also are given to aid investigators who wish to verify accuracy of computer routines. Endpoints of second-

order accurate ADF intervals are somewhat shifted to the right compared to first-order accurate ADF intervals both for $N = 15$ and for $N = 134$. Notice that the first-order accurate interval based on N=134 contains zero whereas the second-order accurate interval does not contain zero.

The lower portion of Table 2 displays intermediate computations and intervals for $\rho_{7(\mathbf{k})}^2 - \rho_{8(\mathbf{k})}^2$ both for the subsample and for the full sample. Again, the second-order intervals are shifted to the right compared to the first-order accurate intervals.

Concluding Remarks

Some caution must be exercised when interpreting the proposed intervals computed on nonnormal data. If the population is multivariate normal, then all relationships among variables are linear and the traditional correlation coefficients adequately summarize the existing relationships. If the population is nonnormal, however, then relationships among variables need not be linear. Traditional correlation coefficients still summarize the linear relationships, but they do not reflect nonlinear components of the existing relationships.

In some cases, confidence interval procedures can be improved by applying a normalizing transformation of the correlation estimator. Let W_n be a statistic computed on a random sample of size N, where $n = N - r$ and r is a fixed positive integer. It is assumed that the moments of W_n are finite at least to order three. Denote the mean and variance of W_n by μ_W and σ_W^2, respectively. A normalizing transformation of W_n, if it exists, is a function of W_n, say $g(W_n)$, chosen such that the skewness of $\sqrt{n}\,[g(W_n) - g(\mu_W)]$ has magnitude $O(n^{-3/2})$. Konishi (1981) showed that if data are sampled from a multivariate nor-

mal distribution, then

$$Z_{\rho_{ij}}(r_{ij}) = \frac{1}{2}\ln\left(\frac{1+r_{ij}}{1-r_{ij}}\right) \text{ and}$$

$$Z_{\rho_{i(\mathbf{k})}^2}(r_{i(\mathbf{k})}^2) = \frac{1}{2}\ln\left(\frac{1+\sqrt{r_{i(\mathbf{k})}^2}}{1-\sqrt{r_{i(\mathbf{k})}^2}}\right)$$

are normalizing transformations for r_{ij} and $r_{i(\mathbf{k})}^2$, respectively. Also,

$$Z_{\rho_{ij\cdot\mathbf{k}}}(r_{ij\cdot\mathbf{k}}) = \frac{1}{2}\ln\left(\frac{1+r_{ij\cdot\mathbf{k}}}{1-r_{ij\cdot\mathbf{k}}}\right)$$

is a normalizing transformation for $r_{ij\cdot\mathbf{k}}$. If data are sampled from multivariate normal distributions, then conventional confidence intervals for $\rho_{ij\cdot\mathbf{k}}$ or $\rho_{i(\mathbf{k})}^2$ that are based on the above normalizing transformations are only first-order accurate, however, because the normalizing transformation corrects for skewness but not for bias.

Derivatives of functions of $\rho_{ij\cdot\mathbf{k}}$ or $\rho_{i(\mathbf{k})}^2$ are readily obtained by one or more applications of the chain rule. For simple and partial correlations,

$$\frac{\partial\, Z_{\rho_{ij\cdot\mathbf{k}}}(\rho_{ij\cdot\mathbf{k}})}{\partial\,\boldsymbol{\sigma}} = \frac{\dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}}{1-\rho_{ij\cdot\mathbf{k}}^2} \text{ and}$$

$$\frac{\partial^2 Z_{\rho_{ij\cdot\mathbf{k}}}(\rho_{ij\cdot\mathbf{k}})}{\partial\,\boldsymbol{\sigma}\otimes\partial\,\boldsymbol{\sigma}} = \frac{\ddot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}}{1-\rho_{ij\cdot\mathbf{k}}^2}$$

$$+\frac{\left(\dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}\otimes\dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}\right)2\rho_{ij\cdot\mathbf{k}}}{\left(1-\rho_{ij\cdot\mathbf{k}}^2\right)^2}.$$

Also, for squared multiple correlation coefficients,

$$\frac{\partial\, Z_{\rho_{i(\mathbf{k})}^2}(\rho_{i(\mathbf{k})}^2)}{\partial\,\boldsymbol{\sigma}} = \frac{\dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2}{2\sqrt{\rho_{i(\mathbf{k})}^2}\left(1-\rho_{i(\mathbf{k})}^2\right)} \text{ and}$$

$$\frac{\partial^2 Z_{\rho_{i(\mathbf{k})}^2}(\rho_{i(\mathbf{k})}^2)}{\partial\,\boldsymbol{\sigma}\otimes\partial\,\boldsymbol{\sigma}} = \frac{\ddot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2}{2\sqrt{\rho_{i(\mathbf{k})}^2}\left(1-\rho_{i(\mathbf{k})}^2\right)}$$

$$+\frac{\left(\dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2 \otimes \dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2\right)\left(3\rho_{i(\mathbf{k})}^2 - 1\right)}{4\left(\rho_{i(\mathbf{k})}^2\right)^{\frac{3}{2}}\left(1 - \rho_{i(\mathbf{k})}^2\right)^2}.$$

Employing the above transformations when data have not been sampled from multivariate normal distributions does not necessarily reduce skewness. Nonetheless, their use can be advantageous because the endpoints of the confidence intervals obtained by back-transforming the intervals for $Z_{\rho_{ij\cdot\mathbf{k}}}$ are guaranteed to be in $(-1, 1)$ and endpoints obtained by back-transforming the intervals for $Z_{\rho_{i(\mathbf{k})}^2}$ are guaranteed to be in $(0, 1)$. Intervals computed directly on the correlation coefficients need not satisfy this property. In small samples, it might be better to transform $r_{i(\mathbf{k})}^2$ as

$$Z^*(r_{i(\mathbf{k})}^2) = \frac{1}{2}\ln\left(\frac{\sqrt{r_{i(\mathbf{k})}^2}}{1 - \sqrt{r_{i(\mathbf{k})}^2}}\right)$$

because this transformation maps $(0, 1)$ to $(-\infty, \infty)$, whereas the normalizing transformation for $r_{i(\mathbf{k})}^2$ maps $(0, 1)$ to $(0, \infty)$.

Fisher's Z transformation also can be employed when ψ is a difference between correlation coefficients. For example, suppose that $\psi = (\rho_{ij} - \rho_{ij\cdot\mathbf{k}})/2$. The divisor 2 is used so that $\psi \in (-1, 1)$. Then, $\hat{\psi}$ can be transformed by

$$Z_\psi(\widehat{\psi}) = \frac{1}{2}\ln\left(\frac{1 + \widehat{\psi}}{1 - \widehat{\psi}}\right).$$

Even if skewness is not reduced by this transformation, the endpoints of the back-transformed interval will be in $(-1, 1)$. The derivatives of this function are

$$\frac{\partial\, Z_\psi(\psi)}{\partial\,\boldsymbol{\sigma}} = \frac{\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}}{1 - \psi^2} \text{ and}$$

$$\frac{\partial^2 Z_\psi(\psi)}{\partial\,\boldsymbol{\sigma}\otimes\partial\,\boldsymbol{\sigma}} = \frac{\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}}{1 - \psi^2} + \frac{\left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\otimes\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)2\psi}{\left(1 - \psi^2\right)^2}.$$

The simulation study also examined properties of first- and second- order intervals that are based on Fisher's Z transformation. The results are very similar to those displayed in Figures 1–3. The performance of the second-order intervals was comparable to that of first-order intervals whenever the first-order intervals did not perform too badly. The second-order intervals improved on the first-order intervals in cases where coverage of the first-order intervals deviated substantially from $1 - \alpha$.

An investigator might prefer to bootstrap correlation functions and thereby dispense with the requirement of explicitly estimating the bias, variance, and skewness of the relevant sampling distributions. One-sided percentile bootstrap intervals for functions of simple, partial, and multiple correlations, however, are only first-order accurate. Often these intervals have poor coverage and can be inferior to normal theory intervals even when multivariate normality is violated (Rasmussen, 1987, 1988; Strube, 1988). Also see Efron (1988) for a discussion of this issue. Second-order accurate confidence intervals can be obtained by bootstrapping the asymptotic pivotal quantity T in (15); i.e., percentile-t intervals. Hall, Martin, and Schucany (1989) examined percentile, percentile-t, transformed percentile-t, and coverage-corrected iterated bootstrap intervals for zero-order correlation coefficients. Fisher's transformation was employed for the transformed percentile-t intervals. They found that in very small samples ($N \le 20$), the iterated intervals as well as the percentile-t intervals worked well. Percentile-t intervals that employed the jackknife to estimate the standard error of the (transformed) sample correlation coefficient were superior to those that employed the delta method. One could

employ jackknife estimates of bias, variance, and skewness in the proposed second-order accurate intervals. This issue remains to be explored.

Abramovitch and Singh (1985) and Zhou and Gao (2000) showed that bootstrapping corrected asymptotic pivotal quantities such as $\hat{T}_2$ in Corollary 5.2 or $\hat{T}_3$ in Corollary 5.3 to obtain quantiles of their respective distributions yields third-order accurate confidence intervals. This bootstrap adjustment procedure was examined for the conditions corresponding to plot $(3, 4)$ in the lower right-hand panel of Figure 3. A sample size of $N = 25$ was employed. The results are summarized in Table 3. The coverage values in the first two lines of Table 3 are those plotted in Figure 3. The remaining lines are based on an additional 2,000 samples. For each of these samples, 1,000 bootstrap samples were drawn and the bootstrap sampling distributions of T and $\hat{T}_3$ were approximated. Quantiles of these distributions were substituted for $t_{\alpha,n}$ in the computation of confidence intervals. Table 3 demonstrates that bootstrapping T yields second-order accurate confidence intervals with coverage similar to those based on $\hat{T}_3$. Bootstrapping $\hat{T}_3$ slightly improved the coverage of the upper interval, but over adjusted the endpoint of the lower interval.

Table 3: Coverage of Bootstrap-Adjusted Confidence Intervals for $\psi = \rho_{i(\mathbf{k})}^2 - \rho_{i(\mathbf{k})}^2$ when Sampling from a Lognormal Distribution

Pivotal Quantity	Bootstrap Adjusted	Coverage Lower Intervals	Coverage Upper Intervals	Number of Samples
T	No	0.972	0.781	5000
$\widehat{T}_3$	No	0.966	0.869	5000
T	No	0.968	0.786	2000
$\widehat{T}_3$	No	0.959	0.877	2000
T	Yes	0.930	0.869	2000
$\widehat{T}_3$	Yes	0.934	0.886	2000

The intervals constructed in this article have reasonable coverage properties under most of the conditions that were examined. In some conditions, however, the coverage of the second-order accurate intervals is less accurate than desired, even though it represents a substantial improvement over coverage of first-order accurate intervals. A bootstrap adjustment to the second-order accurate intervals might improve coverage slightly, but this issue needs to be examined more thoroughly.

1 References

Alf, E.F., & Graf, R.G. (1999). Asymptotic confidence limits for the difference between two squared multiple correlations: A simplified approach. *Psychological Methods*, *4*, 70–75.

Abramovitch, L., & Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *The Annals of Statistics, 13*, 116–132.

Bilodeau, M., & Brenner, D. (1999). *Theory of multivariate statistics.* New York: Springer.

Bishop, Y.M.M., Fienberg, S.F., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge: MIT Press.

Boik, R.J. (1998). A local parameterization of orthogonal and semi-orthogonal matrices with applications. *Journal of Multivariate Analysis*, *67*, 244–276.

Boik, R.J. (2005). Second-order accurate inference on eigenvalues of covariance and correlation matrices. *Journal of Multivariate Analysis*, *96*, 136–171.

Boik, R.J. (2006). Accurate confidence intervals in regression analyses of non-normal data. *Annals of the Institute of Statistical Mathematics*, accepted.

Browne, M.W. & Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra & its Applications*, *82*, 169–176.

Casella, G. & Berger, R.L. (2002). *Statistical inference* (Second Ed.). Pacific Grove CA: Duxbury.

Cook, M.B. (1951). Two applications of bivariate *k*-statistics. *Biometrika, 38*, 368–376.

Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, *104*, 293–296.

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

Fisher, R.A. (1962). The simultaneous distribution of correlation coefficients. *Sankhyā, A, 24*, 1–8.

Graf, R.G., & Alf, E.A., Jr. (1999). Correlations redux: Asymptotic confidence limits for partial and squared multiple correlations. *Applied Psychological Measurement, 23*, 116–119.

Graham, A. (1981). *Kronecker products and matrix differentiation with applications.* Chichester: Ellis Horwood Limited.

Hall, P. (1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B, 54*, 221–228.

Hall, P., Martin, M.A., & Schucany, W.R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *Journal of Statistical Computation and Simulation, 33*, 161–172.

Hedges, L.V., & Olkin, I. (1983). Joint distribution of some indices based on correlation coefficients. In S. Karlin, T. Amemiya, & L. A. Goodman (Eds.), *Studies in econometrics, time series, and multivariate analysis* (pp. 437–454). New York: Academic Press.

Heller, D., Judge, T.A., & Watson, D. (2002). The confounding role of personality and trait affectivity in the relationship between job and life satisfaction. *Journal of Organizational Behavior, 23*, 815–835.

Hsu, P.L. (1949). The limiting distribution of functions of sample means and applications to testing hypotheses. In J. Neyman (Ed.), *Proceedings of the Berkeley Symposium on Mathematical Statistics & Probability* (pp. 359–402). Berkeley: University of California Press.

Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis. *Biometrika*, *68*, 647–651.

MacRae, E.C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *Annals of Statistics*, *2*, 337–346.

Magnus, J.R., & Neudecker, H. (1979). The commutation matrix: some properties and applications. *Annals of Statistics*. *7*, 381–394.

Magnus, J.R., & Neudecker, H. (1999). Matrix differential calculus with applications in statistics and econometrics (Rev. ed.). Chichester: John Wiley & Sons.

Nakagawa, S., & Niki, N. (1992). Distribution of the sample correlation coefficient for nonnormal populations. *Journal of the Japanese Society of Computational Statistics*, *5*, 1–19.

Neudecker, H. (1996). The asymptotic variance matrices of the sample correlation matrix in elliptical and normal situations and their proportionality. *Linear Algebra & its Applications*, *237/238*, 127–132.

Neudecker, H. & Wesselman, A.M. (1990). The asymptotic variance matrix of the sample correlation matrix. *Linear Algebra & its Applications 127*, 589–599.

Niki, N., & Konishi, S. (1986). Higher-order asymptotic expansions for the distribution of the sample correlation coefficient. *Communications in Statistics—Simulation and Computation*, *13*, 169–182.

Olkin, I., & Finn, J.D. (1995). Correlations redux. *Psychological Bulletin*, *118*, 155–164.

Olkin, I., & Siotani, M. (1976). Asymptotic distributions of functions of a correlation matrix. In S. Ikeda, T. Hayakawa, H. Hudimoto, M. Okamoto, S. Siotani, & S. Yamamoto (Eds.), *Essays in probability and Statistics* (pp. 235–251). Tokyo: Shinko Tsusho.

Pace, L., & Salvan, A. (1997). *Principles of statistical inference*. Singapore: World Scientific.

Pearson, K. & Filon, L.N.G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London, A 191*, 229–311.

Rao, C.R. (1973). *Linear statistical inference and its applications* (Second ed.). New York: John Wiley & Sons.

Rasmussen, J.L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches, *Psychological Bulletin*, *101*, 136–139.

Rasmussen, J.L. (1988). "Bootstrap confidence intervals: Good or Bad": Comments on Efron (1988) and Strube (1988) and further evaluation. *Psychological Bulletin*, *104*, 297–299.

Steiger, J.H. & Hakstian, A.R. (1982). The asymptotic distribution of the elements of a correlation matrix: Theory and application. *British Journal of Mathematical & Statistical Psychology*, *35*, 208–215.

Steiger, J.H. & Hakstian, A.R. (1983). A historical note on the asymptotic distribution of correlations. *British Journal of Mathematical & Statistical Psychology, 36*, 157.

Strube, M.J. (1988). Bootstrap type I error rates for the correlation coefficient: An examination of alternate procedures. *Psychological Bulletin, 104*, 290–292.

Zhou, X.-H., & Gao, S. (2000). One-sided confidence intervals for means of positively skewed distributions. *The American Statistician, 54*, 100–104.

Appendix

Proof of Theorem 4

Taylor series expansions for $\sqrt{n}\left(\widehat{\psi} - \psi\right)$, $\left[\sqrt{n}\left(\widehat{\psi} - \psi\right)\right]^2$, and $\left[\sqrt{n}\left(\widehat{\psi} - \psi\right)\right]^3$ around $\mathbf{s} = \boldsymbol{\sigma}$ can be written as follows.

$$\sqrt{n}\left(\widehat{\psi} - \psi\right) = \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$$

$$+\frac{1}{2\sqrt{n}}\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \left[\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \otimes \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]$$

$$+O_p\left(n^{-1}\right),$$

$$\left[\sqrt{n}\left(\widehat{\psi} - \psi\right)\right]^2 = \left[\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]^2$$

$$+O_p\left(n^{-\frac{1}{2}}\right), \text{ and}$$

$$\left[\sqrt{n}\left(\widehat{\psi} - \psi\right)\right]^3 = \left[\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]^3$$

$$+\frac{3}{2\sqrt{n}}\left[\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]^2$$

$$\times\ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \left[\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \otimes \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right] + O\left(n^{-1}\right).$$

Furthermore, expanding $\sqrt{n}(\hat{\sigma}_{\psi}^2 - \sigma_{\psi}^2)$ around $\mathbf{s} = \boldsymbol{\sigma}$ reveals that

$$\sqrt{n}(\hat{\sigma}_{\psi}^2 - \sigma_{\psi}^2) = \left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)' \sqrt{n}\left(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}\right)+$$

$$2\left[\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})' \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\Omega}_{22,n}\right] \ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} + O_p\left(n^{-\frac{1}{2}}\right),$$

where $\boldsymbol{\omega} = \text{vec}\,\boldsymbol{\Omega}_{22,n}$. The claimed result is obtained by taking expectations, using $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\sigma} = 0$, and collecting terms of like order.

To verify that $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}'\boldsymbol{\sigma} = 0$, note that $\rho_{ij\cdot\mathbf{k}}$ and $\rho_{i(\mathbf{k})}^2$ are scale invariant functions of $\boldsymbol{\Sigma}$. That is, $\rho_{ij\cdot\mathbf{k}}(\boldsymbol{\Sigma}) = \rho_{ij\cdot\mathbf{k}}(\alpha\boldsymbol{\Sigma})$ and $\rho_{i(\mathbf{k})}^2(\boldsymbol{\Sigma}) = \rho_{i(\mathbf{k})}^2(\alpha\boldsymbol{\Sigma})$, where α is any positive scalar constant. It follows that the derivatives of $\rho_{ij\cdot\mathbf{k}}(\alpha\boldsymbol{\Sigma})$ and $\rho_{i(\mathbf{k})}^2(\alpha\boldsymbol{\Sigma})$ with respect to α each are zero. Using the chain rule,

$$0 = \frac{\partial \rho_{ij\cdot\mathbf{k}}(\alpha\boldsymbol{\Sigma})}{\partial \alpha} = \left(\frac{\partial \alpha\boldsymbol{\sigma}'}{\partial \alpha}\right)\left(\frac{\partial \rho_{ij\cdot\mathbf{k}}}{\partial \boldsymbol{\sigma}\alpha}\right)$$

$$= \boldsymbol{\sigma}'\dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}\left(\frac{1}{\alpha}\right) \text{ and}$$

$$0 = \frac{\partial \rho_{i(\mathbf{k})}^2(\alpha\boldsymbol{\Sigma})}{\partial \alpha} = \left(\frac{\partial \alpha\boldsymbol{\sigma}'}{\partial \alpha}\right)\left(\frac{\partial \rho_{i(\mathbf{k})}^2}{\partial \boldsymbol{\sigma}\alpha}\right)$$

$$= \boldsymbol{\sigma}'\dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2\left(\frac{1}{\alpha}\right).$$

The result follows because $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ is a linear function of $\dot{\boldsymbol{\rho}}_{ij\cdot\mathbf{k}}$ and/or $\ddot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2$ terms.

Proof of Theorem 5

Expand T as

$$T = \frac{\sqrt{n}(\widehat{\psi} - \psi)}{\sigma_{\psi}}\times$$

$$\left[1 - \frac{1}{2\sqrt{n}\sigma_{\psi}^2}\sqrt{n}(\widehat{\sigma}_{\psi}^2 - \sigma_{\psi}^2) + O_p\left(n^{-1}\right)\right],$$

and then take expectations of T, T^2, and T^3 to verify the claimed expressions for the mean, variance, and skewness of T. The sizes of the various remainder terms differ because the expectation of $O_p(1)$ terms that are even-order in $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ or $\sqrt{n}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ have magnitude $O(1)$ whereas the expectation of $O_p(1)$ terms that are odd-order in $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ or $\sqrt{n}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ have magnitude $O\left(n^{-1/2}\right)$. It follows that the cumulant generating function of T can be expanded as follows:

$$C_T(u) = \ln\left\{\mathrm{E}\left[e^{iuT}\right]\right\}$$

$$= \frac{\kappa_1\, iu}{\sqrt{n}} - \frac{u^2}{2} + \frac{\kappa_3\,(iu)^3}{6\sqrt{n}} + O\left(n^{-1}\right).$$

Inverting the characteristic function, $\exp\{C_T(u)\}$, yields the density function, f_T, and integrating f_T yields the distribution function, F_T.

Inversion of Exponential Functions
 This section describes a Modified Newton algorithm for finding the value of $\hat{t}_{3,\alpha}$ that satisfies $h(\hat{t}_{3,\alpha}) = 0$, where

$$h(\hat{t}_{3,\alpha})$$

$$= \hat{t}_{3,\alpha} - \frac{\widehat{\kappa}_1}{\sqrt{n}} - \frac{\widehat{\kappa}_3\left(\hat{t}_{3,\alpha}^2\, e^{-\hat{d}\hat{t}_{3,\alpha}^2/2} - 1\right)}{6\sqrt{n}} - z_\alpha,$$

where $\hat{d}$ is defined in Corollary 5.3. The solution is unique because $h(\hat{t}_{3,\alpha})$ is a monotonic function of $\hat{t}_{3,\alpha}$ by construction. An initial guess for $\hat{t}_{3,\alpha}$ is $\hat{t}_{3,\alpha,0} = \hat{t}_\alpha$, where $\hat{t}_\alpha$ is defined in Corollary 5.2. At iteration $i + 1$ of the algorithm, the value of $\hat{t}_{3,\alpha}$ is

$$\hat{t}_{3,\alpha,i+1} = \hat{t}_{3,\alpha,i} - u_i, \text{ where}$$

$$u_i = \begin{cases} -\hat{t}_{3,\alpha,i} & \text{if } \dfrac{h(\hat{t}_{3,\alpha,i})}{\hat{t}_{3,\alpha,i}\, h^{(1)}(\hat{t}_{3,\alpha,i})} < -1, \\[3mm] \dfrac{\hat{t}_{3,\alpha,i}}{2} & \text{if } \dfrac{h(\hat{t}_{3,\alpha,i})}{\hat{t}_{3,\alpha,i}\, h^{(1)}(\hat{t}_{3,\alpha,i})} > \dfrac{1}{2}, \\[3mm] \dfrac{h(\hat{t}_{3,\alpha,i})}{h^{(1)}(\hat{t}_{3,\alpha,i})} & \text{otherwise,} \end{cases}$$

and $h^{(1)}(\hat{t}) = 1 - \dfrac{\widehat{\kappa}_3\,\hat{t}\,(2 - \hat{d}\hat{t}^2)e^{-\hat{d}\hat{t}^2/2}}{6\sqrt{n}}$.

Proof of Theorem 6
 Define Z_i as

$$Z_i \stackrel{\mathrm{def}}{=} \frac{\sqrt{n_i}\left(\widehat{\psi}_i - \psi_i\right)}{\sigma_i},$$

where all terms are defined in Theorem 6. Then, the first-order Taylor series expansion of T in Theorem 6 around $\hat{\sigma}_i^2 = \sigma_i^2$ for $i = 1, 2$ is

$$T = \frac{g^{-\frac{1}{2}}\sigma_1 Z_1 - \sigma_2 Z_2}{(g^{-1}\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}} \times$$

$$\left[1 - \frac{V_1}{2\sqrt{n_1}} + \frac{V_2}{2\sqrt{n_2}} + O_p\left(n^{-1}\right)\right], \text{ where}$$

$$g = \frac{n_1}{n_2}, \quad n = \min(n_1, n_2),$$

$$V_1 = \frac{\sqrt{n_1}(\hat{\sigma}_1^2 - \sigma_1^2)}{\sigma_1^2 + g\sigma_2^2}, \text{ and}$$

$$V_2 = \frac{\sqrt{n_2}(\hat{\sigma}_2^2 - \sigma_2^2)}{g^{-1}\sigma_1^2 + \sigma_2^2}.$$

The claimed results are obtained by using Theorem 4 to obtain expectations of T, $[T - \mathrm{E}(T)]^2$, and $[T - \mathrm{E}(T)]^3$.