PSYCHOMETRIKA—VOL. 49, NO. 1, 11–24
MARCH 1984

THE COMPARISON OF INTERDEPENDENT CORRELATIONS BETWEEN OPTIMAL LINEAR COMPOSITES

JAMES H. STEIGER

UNIVERSITY OF BRITISH COLUMBIA

MICHAEL W. BROWNE

UNIVERSITY OF SOUTH AFRICA

A general procedure is provided for comparing correlation coefficients between optimal linear composites. The procedure allows computationally efficient significance tests on independent or dependent multiple correlations, partial correlations, and canonical correlations, with or without the assumption of multivariate normality. Evidence from some Monte Carlo studies on the effectiveness of the methods is also provided.

Key words: asymptotic distribution theory, correlational significance tests, multiple correlation, partial correlation, canonical correlation.

Asymptotic theory is available for the joint distribution of correlation coefficients, under an assumption of multivariate normality (Hsu, 1949; Olkin & Siotani, 1976; Pearson & Filon, 1898), and without this assumption (Hsu, 1949; Steiger & Hakstian, 1982, 1983). This theory can be used for constructing large sample tests for equality of interdependent correlation coefficients. A review and evaluation of available methods is contained in Steiger (1980a). Asymptotic distribution theory is also available (e.g., Muirhead, 1982; Muirhead & Waternaux, 1980) for certain correlation coefficients (such as multiple correlations, canonical correlations, and partial correlations) between optimal linear composites of variables. Available theory has not been concerned with the comparison of correlation coefficients of optimal composites involving different subsets of variables. This situation could be handled using the delta method, either with analytic derivatives or with numerical derivatives (Lord, 1975). Unfortunately, unless algebraic simplifications are available, the amount of computational effort becomes prohibitive when the numbers of variables involved in the linear composite become large. For example, consider a routine application of the delta method to the comparison of two multiple correlations measured on the same test batteries given to the same subjects on two occasions. With 20 predictor variables and one criterion, an estimated variance-covariance matrix of correlations of order 861×861 would have to be calculated, and 1722 derivatives would have to be evaluated. A massive computational effort would be involved.

This paper shows that standard theory available for simple correlations may be applied directly to certain correlation coefficients between optimal linear composites such as multiple correlations, canonical correlations, partial correlations, and an internal correlation coefficient derived by Venables (1976) as a union-intersection test for sphericity.

This research was supported in part by an operating grant (#67-4640) to the first author from the National Sciences and Engineering Research Council of Canada. The authors would also like to acknowledge the helpful comments and encouragement of Alexander Shapiro, Stanley Nash, and Ingram Olkin.

Requests for reprints and details of a FORTRAN program useful in computing the significance tests described in this article should be sent to James H. Steiger, Department of Psychology, University of British Columbia, Vancouver, B.C., Canada V6T 1Y7.

0033-3123/84/0399-5079$00.75/0

The necessity for much tedious algebra is eliminated, and the resulting formulae are generally quite simple.

Section 2 will review the available asymptotic theory for simple correlation coefficients and functions of a correlation matrix. Section 3 reviews some useful asymptotic theory on testing linear hypotheses. This theory may be applied to obtain simplified tests for certain hypotheses on equality of correlations. In Section 4 the basic theorem providing the asymptotic multivariate distribution of correlation coefficients between optimal composites is derived. Optimal correlations to which this theorem applies are given in Section 5, and a Monte Carlo study demonstrating the usefulness of the results obtained is reported in Section 6.

2. *Asymptotic Distribution of Correlation Coefficients*

Let x_i, x_j, x_k, and x_h be random variables with a multivariate distribution having finite fourth-order moments. Define

$$\mu_i = E(x_i) \tag{2.1}$$

$$\sigma_{ij} = E(x_i - \mu_i)(x_j - \mu_j) \tag{2.2}$$

$$\sigma_{ijkh} = E(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_h - \mu_h) \tag{2.3}$$

$$\rho_{ij} = \sigma_{ij}(\sigma_{ii}\sigma_{jj})^{-1/2} \tag{2.4}$$

$$\rho_{ijkh} = \sigma_{ijkh}(\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{hh})^{-1/2} \tag{2.5}$$

Then consider samples of $N = n + 1$ independent observations on variates x_i, x_j, x_k, and x_h. We define the sample statistics

$$m_i = N^{-1}\sum_{r=1}^{N} x_{ri} \tag{2.6}$$

$$s_{ij} = n^{-1}\sum_{r=1}^{N} (x_{ri} - m_i)(x_{rj} - m_j) \tag{2.7}$$

$$s_{ijkh} = n^{-1}\sum_{r=1}^{N} (x_{ri} - m_i)(x_{rj} - m_j)(x_{rk} - m_k)(x_{rh} - m_h) \tag{2.8}$$

$$z_{ri} = (x_{ri} - m_i)s_{ii}^{-1/2} \tag{2.9}$$

$$r_{ij} = s_{ij}(s_{ii}s_{jj})^{-1/2} = n^{-1}\sum_{r=1}^{N} z_{ri}z_{rj} \tag{2.10}$$

$$r_{ijkh} = s_{ijkh}(s_{ii}s_{jj}s_{kk}s_{hh})^{-1/2} = n^{-1}\sum_{r=1}^{N} z_{ri}z_{rj}z_{rk}z_{rh} \tag{2.11}$$

In the discussion which follows, we shall refer frequently to two Propositions.

Proposition 1. (Rao, 1973, Theorem 6a.2(iii)). Let $\mathbf{s}$ be a $p^* \times 1$ random vector, dependent on n, and $\boldsymbol{\sigma}$ be a vector of constants of the same order, such that $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ has an asymptotic distribution which is MVN($\mathbf{0}$, Υ) (i.e., a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix Υ). Let $\mathbf{c}(\mathbf{s})$ be a $q \times 1$ vector (with elements $c_i(\mathbf{s})$) of functions of the elements of $\mathbf{s}$, which is differentiable at $\mathbf{s} = \boldsymbol{\sigma}$.

Then $n^{1/2}\{\mathbf{c}(\mathbf{s}) - \mathbf{c}(\boldsymbol{\sigma})\}$ has an asymptotic distribution which is MVN($\mathbf{0}$, G), $G = \Delta'\Upsilon\Delta$, where $\Delta' = \partial\mathbf{c}/\partial\mathbf{s}'|_{\mathbf{s}=\boldsymbol{\sigma}}$ is the $p^* \times q$ Jacobian matrix of $c(\mathbf{s})$ evaluated at $\boldsymbol{\sigma}$.

Direct application of Proposition 1 to a consequence of the multivariate central limit theorem leads to the following general result on the asymptotic distribution of elements of a correlation matrix.

Proposition 2. (Hsu, 1949; Isserlis, 1916; Steiger & Hakstian, 1982). Let $\mathbf{x}$ be a $p \times 1$ random vector having a multivariate distribution with finite fourth-order moments, and population correlation matrix $P = \{\rho_{ij}\}$. Let $\mathbf{r}$ be a random vector of order $p(p-1)/2$ formed from the non-duplicated elements of the sample correlation matrix R, based on $N = n + 1$ observations on $\mathbf{x}$. Let $\boldsymbol{\rho}$ be formed from the nonduplicated elements of P. Define $\mathbf{r}^* = n^{1/2}(\mathbf{r} - \boldsymbol{\rho})$, with elements r_{ij}^*. Then $\mathbf{r}^*$ has an asymptotic distribution which is MVN($\mathbf{0}$, Ψ) with variance-covariance matrix Ψ having typical element $\psi_{ij,kh} = \text{Cov}(r_{ij}^*, r_{kh}^*)$ given by

$$\psi_{ij,kh} = \rho_{ijkh} + \tfrac{1}{4}\rho_{ij}\rho_{kh}(\rho_{iikk} + \rho_{jjkk} + \rho_{iihh} + \rho_{jjhh}) \\ - \tfrac{1}{2}\rho_{ij}(\rho_{iikh} + \rho_{jjkh}) - \tfrac{1}{2}\rho_{kh}(\rho_{ijkk} + \rho_{ijhh}) \tag{2.12}$$

Thus, in general, the asymptotic variances and covariances of correlation coefficients depend on fourth-order moments of the distribution of $\mathbf{x}$. Consequently, an estimate of Ψ would involve estimates of all fourth-order moments.

It is convenient to consider a class of distributions where all elements of Ψ are functions only of the elements of P and of a single kurtosis parameter. This is the class of elliptical distributions (cf. Devlin, Gnanadesikan, & Kettenring, 1976; Muirhead, 1982; Muirhead and Waternaux, 1980).

Let $\mathbf{x}$ be a $p \times 1$ vector variate with expected value $\boldsymbol{\mu}$, covariance matrix Σ, and finite fourth-order moments. We define the multivariate coefficient of relative kurtosis of $\mathbf{x}$ (cf. Browne, 1982), denoted by η_p, to be Mardia's (1970, 1974) coefficient of multivariate kurtosis divided by the corresponding coefficient of multivariate kurtosis for a multivariate normal distribution:

$$\eta_p = E\{(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^2/p(p+2).$$

If the distribution of $\mathbf{x}$ belongs to the elliptical family and $\mathbf{x}_k$, $k = 1, \ldots, p$, is any $k \times 1$ subvector of $\mathbf{x}$, then the relative kurtosis of $\mathbf{x}_k$ is equal to that of $\mathbf{x}$, i.e., $\eta_k = \eta_p$. In particular, all individual elements of $\mathbf{x}$ have the same marginal relative kurtosis. We shall denote the common relative kurtosis of an elliptical distribution by η. This may be estimated by means of an estimate of η_p. An algorithm for obtaining an estimate of $p(p+2)\eta_p$ is given in Mardia and Zemroch (1975).

A convenient property of elliptical distributions is that the standardized fourth-order moments, ρ_{ijkh}, are functions of the correlation coefficients, ρ_{ij}, and the common relative kurtosis coefficient, η:

$$\rho_{ijkh} = \eta(\rho_{ij}\rho_{kh} + \rho_{ik}\rho_{jh} + \rho_{ih}\rho_{jk}) \tag{2.13}$$

Substitution of (2.13) into (2.12) yields the following corollary to Proposition 2.

Corollary 2.1 If the distribution of $\mathbf{x}$ is elliptical with relative kurtosis η, then:

$$\psi_{ij,kh} = \eta\{\tfrac{1}{2}\rho_{ij}\rho_{kh}(\rho_{ik}^2 + \rho_{ih}^2 + \rho_{jk}^2 + \rho_{jh}^2) + \rho_{ik}\rho_{jh} + \rho_{ih}\rho_{jk} \\ - \rho_{ij}(\rho_{jk}\rho_{jh} + \rho_{ik}\rho_{ih}) - \rho_{kh}(\rho_{jk}\rho_{ik} + \rho_{jh}\rho_{ih})\} \tag{2.14}$$

The multivariate normal distribution is an elliptical distribution with relative kurtosis $\eta = 1$. Replacement of η by 1 in (2.14) yields the well known formula for the covariance of correlation coefficients based on observations from a multivariate normal distribution (Hsu, 1949; Olkin & Siotani, 1976; Pearson & Filon, 1898).

3. *Asymptotic Tests for Linear Hypotheses*

Suppose that $\mathbf{r}$ is a $k \times 1$ vector variate and that the asymptotic distribution of $n^{1/2}(\mathbf{r} - \boldsymbol{\rho})$ is MVN($\mathbf{0}$, Ψ). We shall be concerned with tests of linear hypotheses of the form

$$H_0: M\boldsymbol{\rho} = \mathbf{h} \tag{3.1}$$

where M is a specified $g \times k$ matrix of rank g, and $\mathbf{h}$ is a specified $g \times 1$ vector. While, in theory, M and $\mathbf{h}$ can have arbitrarily chosen elements, in the majority of practical situations $\mathbf{h}$ will be a null vector and the rows of M will contain either one $+1$ and one -1 with the remaining elements equal to zero (to test equality of two elements of $\boldsymbol{\rho}$, or will contain one $+1$ with the remaining elements equal to zero to test equality of an element of $\boldsymbol{\rho}$ to zero.

As an example we consider the case where $k = 4$, and we wish to test for equality of the first three elements of $\boldsymbol{\rho}$, and for equality of the last element to zero. Then $\mathbf{h} = \mathbf{0}$, $g = 3$, and a possible choice for M is

$$M = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.2}$$

If $\hat{\Psi}$ is any consistent estimate of Ψ, then the asymptotic distribution under H_0 of the statistic

$$X^2 = n(M\mathbf{r} - \mathbf{h})'(M\hat{\Psi}M')^{-1}(M\mathbf{r} - \mathbf{h}) \tag{3.3}$$

will be chi-square with g degrees of freedom. This test statistic may be used to test H_0 against the general alternative where the elements of $\boldsymbol{\rho}$ are unconstrained.

While the formulation of H_0 given in (3.1) and the test statistic given in (3.3) allow for the testing for equality of elements of $\boldsymbol{\rho}$, this approach does not provide estimates of the equal elements of $\boldsymbol{\rho}$. Such estimates may be obtained by means of a second, mathematically equivalent approach. Equation (3.1) will be true if and only if $\boldsymbol{\rho}$ can be expressed in the form

$$H_0: \boldsymbol{\rho} = A\boldsymbol{\Theta} + \boldsymbol{\rho}^* \tag{3.4}$$

where A is any (conveniently chosen) $k \times q$ matrix of rank $q = k - g$, such that $MA = 0$, $\boldsymbol{\Theta}$ is a $g \times 1$ parameter vector, and $\boldsymbol{\rho}^* = M'(MM')^{-1}\mathbf{h}$ is a known $k \times 1$ vector which is null whenever $\mathbf{h}$ is null.

For example, (3.1) with M given by (3.2) and $\mathbf{h} = \mathbf{0}$ implies, and is implied by (3.4) with

$$A = (1 \quad 1 \quad 1 \quad 0)' \tag{3.5}$$

and $\boldsymbol{\rho}^* = \mathbf{0}$. Since $q = 1$, Θ is a single parameter which represents the unknown common value of the first three elements of $\boldsymbol{\rho}$.

The generalized least squares estimator of $\boldsymbol{\Theta}$ is

$$\hat{\boldsymbol{\Theta}} = (A'\hat{\Psi}^{-1}A)^{-1}A'\hat{\Psi}^{-1}(\mathbf{r} - \boldsymbol{\rho}^*) \tag{3.6}$$

and the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})$ is MVN($\mathbf{0}$, $\{A'\Psi^{-1}A\}^{-1}$). When using (3.6), one may calculate the test statistic of (3.3) conveniently using the mathematically equivalent form

$$X^2 = n(\mathbf{r} - \hat{\boldsymbol{\rho}})'\hat{\Psi}^{-1}(\mathbf{r} - \hat{\boldsymbol{\rho}}) \tag{3.7}$$

where

$$\hat{\boldsymbol{\rho}} = A\hat{\boldsymbol{\Theta}} + \boldsymbol{\rho}^*. \tag{3.8}$$

The results reviewed here, with $k \leq \frac{1}{2}p(p-1)$ may be employed to test for patterns of equality of elements of a single $p \times p$ correlation matrix (Steiger, 1980a, 1980b, Note 2) or of elements of several independently estimated correlation matrices (Steiger, Note 1). In the case where $k = \frac{1}{2}p(p-1)$ so that $\boldsymbol{\rho}$ consists of all nonduplicated elements of the population correlation matrix and Ψ has the special form of (2.14) the generalized least squares estimates (3.6) and test statistic (3.7) may be expressed in alternative forms which are more efficient for computational purposes (Browne, 1977).

Any consistent estimate, $\hat{\Psi}$, of Ψ may be employed. It has been found (Browne, 1977; Steiger, 1980a,b) that, under the assumption of multivariate normality when $\eta = 1$ in (2.14), it is preferable to replace the ρ_{ij} in (2.14) by ordinary least squares estimates which are elements of the vector

$$\tilde{\boldsymbol{\rho}} = A[(A'A)^{-1}A'\mathbf{r}] + \boldsymbol{\rho}^* \tag{3.9}$$

rather than to use (2.10). When (2.12) was used in the Monte Carlo experiments of Section 6, the ρ_{ijkh} were estimated using formulae corresponding to (2.11), and the ρ_{ij} using formulae corresponding to (3.9).

4. *Correlations between Optimal Linear Composites*

In this section we demonstrate why, and how, the method of Section 3 may be used to test pattern hypotheses on a number of well-known correlational statistics, including multiple correlations, partial correlations, and canonical correlations. First we will need

Proposition 3. Suppose that $c(\mathbf{s}, \mathbf{b})$ is a differentiable scalar valued function of two vector valued arguments $\mathbf{s}$ and $\mathbf{b}$, and that $\dot{\mathbf{b}}(\mathbf{s})$ is a vector valued function of $\mathbf{s}$ which yields a stationary point of $c(\mathbf{s}, \mathbf{b})$ with respect to $\mathbf{b}$ given $\mathbf{s}$. That is:

$$\left.\frac{\partial c(\mathbf{s}, \mathbf{b})}{\partial \mathbf{b}}\right|_{\mathbf{b} = \dot{\mathbf{b}}(\mathbf{s})} = \mathbf{0}$$

for all $\mathbf{s}$ in $N(\boldsymbol{\sigma})$, a neighborhood of a given point $\boldsymbol{\sigma}$. Consider the composite function $c^*(\mathbf{s}) = c[\mathbf{s}, \dot{\mathbf{b}}(\mathbf{s})]$. If $\dot{\mathbf{b}}(\mathbf{s})$ is differentiable at $\mathbf{s} = \boldsymbol{\sigma}$ then $c^*(\mathbf{s})$ is differentiable at $\mathbf{s} = \boldsymbol{\sigma}$, and

$$\frac{\partial c^*(\mathbf{s})}{\partial \mathbf{s}} = \left.\frac{\partial c(\mathbf{s}, \mathbf{b})}{\partial \mathbf{s}}\right|_{\mathbf{b} = \dot{\mathbf{b}}(\mathbf{s})} \tag{4.1}$$

Proof. By the chain rule,

$$\frac{\partial c^*(\mathbf{s})}{\partial \mathbf{s}} = \left.\frac{\partial c(\mathbf{s}, \mathbf{b})}{\partial \mathbf{s}} + \frac{\partial \dot{\mathbf{b}}'(\mathbf{s})}{\partial \mathbf{s}} \frac{\partial c(\mathbf{s}, \mathbf{b})}{\partial \mathbf{b}}\right|_{\mathbf{b} = \dot{\mathbf{b}}(\mathbf{s})}$$

$$= \left.\frac{\partial c(\mathbf{s}, \mathbf{b})}{\partial \mathbf{s}}\right|_{\mathbf{b} = \dot{\mathbf{b}}(\mathbf{s})} + \mathbf{0} \qquad \square$$

Result (4.1) has been applied in the development of "nested" (Ross, 1970) algorithms for obtaining maximum likelihood estimates (e.g., Browne, 1979, p. 212; Jöreskog, 1967, p. 450). In these applications a likelihood function $c(\mathbf{s}, \mathbf{b})$ was to be maximised with respect to *both* $\mathbf{s}$ and $\mathbf{b}$. An example is that of unrestricted factor analysis, where $\mathbf{s}$ represents unique variances, and $\mathbf{b}$ factor loadings (e.g., Jöreskog, 1967). In the present paper, result (4.1) will be used in another manner, which is similar to that in Shapiro [1983]. Here $\mathbf{s}$

will be a vector variate representing sample variances and covariances, and $c(\mathbf{s}, \mathbf{b})$ will be optimised with respect to $\mathbf{b}$ alone to provide weights for a linear composite of variables. Proposition 3 will then be applied to derive the asymptotic distribution of $c^*(\mathbf{s})$ rather than to provide a computational algorithm.

Proposition 4. Let the asymptotic distribution of $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ be multivariate normal, with a null mean vector and covariance matrix Υ. Suppose that $\mathbf{c}(\mathbf{s}, \mathbf{b})$ is a differentiable $q \times 1$ vector valued function of $\mathbf{s}$ and $\mathbf{b}$. The vector valued function $\acute{\mathbf{b}}(\mathbf{s})$ yields a stationary point of $\mathbf{c}(\mathbf{s}, \mathbf{b})$ with respect to $\mathbf{b}$ given $\mathbf{s}$, i.e.,

$$\left.\frac{\partial \mathbf{c}'(\mathbf{s}, \mathbf{b})}{\partial \mathbf{b}}\right|_{\mathbf{b} = \acute{\mathbf{b}}(\mathbf{s})} = \mathbf{0}, \qquad \forall \mathbf{s} \varepsilon N(\boldsymbol{\sigma}).$$

Let $\mathbf{c}^*(\mathbf{s}) = \mathbf{c}[\mathbf{s}, \acute{\mathbf{b}}(\mathbf{s})]$ and let $\boldsymbol{\beta} = \acute{\mathbf{b}}(\boldsymbol{\sigma})$, so that $\mathbf{c}^*(\boldsymbol{\sigma}) = \mathbf{c}(\boldsymbol{\sigma}, \boldsymbol{\beta}) = \boldsymbol{\kappa}$, say. If $\acute{\mathbf{b}}(\mathbf{s})$ is differentiable at $\mathbf{s} = \boldsymbol{\sigma}$, then:

(a) The $p^* \times q$ Jacobian matrix $\Delta^{*\prime} = \partial \mathbf{c}^*(\boldsymbol{\sigma})/\partial \mathbf{s}'$ of $\mathbf{c}^*(\mathbf{s})$ at $\boldsymbol{\sigma}$ is equal to the $p^* \times q$ Jacobian matrix $\Delta' = \partial \mathbf{c}(\boldsymbol{\sigma}, \boldsymbol{\beta})/\partial \mathbf{s}'$ of $\mathbf{c}(\mathbf{s}, \boldsymbol{\beta})$ at $\boldsymbol{\sigma}$.

(b) $n^{1/2}\{c^*(\mathbf{s}) - \boldsymbol{\kappa}\}$ and $n^{1/2}\{\mathbf{c}(\mathbf{s}, \boldsymbol{\beta}) - \boldsymbol{\kappa}\}$ have the same asymptotic distribution.

Proof. Application of Proposition 3 shows that each column of Δ^* is equal to the corresponding column of Δ so that part (a) follows. Part (b) follows immediately from Proposition 1. □

We shall be concentrating on a situation where $\mathbf{b}$ is partitioned into q subvectors, $\mathbf{b} = (\mathbf{b}_1', \mathbf{b}_2', \ldots, \mathbf{b}_q')'$, and the ith element of $\mathbf{c}(\mathbf{s}, \mathbf{b})$ depends only on $\mathbf{b}_i$: i.e., $c_i(\mathbf{s}, \mathbf{b}) = c_i(\mathbf{s}, \mathbf{b}_i)$. Similarly, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_q')'$ where $\boldsymbol{\beta}_i = \acute{\mathbf{b}}_i(\boldsymbol{\sigma})$.

Proposition 4(b) will be used to obtain the asymptotic distribution of $\mathbf{c}^*(\mathbf{s})$ in situations where the asymptotic distribution of $\mathbf{c}(\mathbf{s}, \boldsymbol{\beta})$ is already known. This obviates tedious algebra (necessary when implementing the delta method) involved in deriving the Jacobian matrix $\Delta^{*\prime}$ of $\mathbf{c}^*(\mathbf{s})$.

We shall be concerned with situations where the vector variate $\mathbf{s}$ is formed from the $p^* = \frac{1}{2}p(p + 1)$ nonduplicated elements of a $p \times p$ sample covariance matrix, S, based on $n + 1$ independent observations on $\mathbf{x}$, and $\boldsymbol{\sigma}$ is formed from the corresponding elements of the population covariance matrix Σ. The elements $c_i(\mathbf{s}, \mathbf{b}_i)$ of $\mathbf{c}(\mathbf{s}, \mathbf{b})$ will be functions of the general form

$$c_i(\mathbf{s}, \mathbf{b}_i) = \mathbf{g}_{1i}' S \mathbf{g}_{2i} \{(\mathbf{g}_{1i}' S \mathbf{g}_{1i})(\mathbf{g}_{2i}' S \mathbf{g}_{2i})\}^{-1/2} \tag{4.2}$$

where $\mathbf{g}_{1i} = \dot{\mathbf{g}}_{1i}(\mathbf{b}_i)$ and $\mathbf{g}_{2i} = \dot{\mathbf{g}}_{2i}(\mathbf{b}_i)$ are two $p \times 1$ vectors whose elements are equal either to elements of $\mathbf{b}_i$ or to the constants 0 or 1. Specific examples are given in Section 5. It is clear from (4.2) that $c_i(\mathbf{s}, \mathbf{b}_i)$ represents the correlation coefficient between $\mathbf{g}_{1i}'\mathbf{x}$ and $\mathbf{g}_{2i}'\mathbf{x}$. Sample optimal weights are represented by $\hat{\boldsymbol{\beta}}_i = \acute{\mathbf{b}}_i(\mathbf{s})$. Substitution of $\hat{\boldsymbol{\gamma}}_{1i} = \dot{\mathbf{g}}_{1i}(\hat{\boldsymbol{\beta}}_i)$ and $\hat{\boldsymbol{\gamma}}_{2i} = \dot{\mathbf{g}}_{2i}(\hat{\boldsymbol{\beta}}_i)$ into (4.2) gives the sample optimal correlation coefficient $c_i^*(\mathbf{s}) = c_i(\mathbf{s}, \hat{\boldsymbol{\beta}}_i)$. This is the sample correlation coefficient between $\hat{y}_{1i} = \hat{\boldsymbol{\gamma}}_{1i}'\mathbf{x}$ and $\hat{y}_{2i} = \hat{\boldsymbol{\gamma}}_{2i}'\mathbf{x}$, i.e.,

$$c_i^*(\mathbf{s}) = c_i(\mathbf{s}, \hat{\boldsymbol{\beta}}_i) = r(\hat{y}_{1i}, \hat{y}_{2i}) \tag{4.3}$$

Corresponding population optimal weights are given by $\boldsymbol{\beta}_i = \acute{\mathbf{b}}_i(\boldsymbol{\sigma})$ with $\boldsymbol{\gamma}_{1i} = \dot{\mathbf{g}}_{1i}(\boldsymbol{\beta}_i)$ and $\boldsymbol{\gamma}_{2i} = \dot{\mathbf{g}}_{2i}(\boldsymbol{\beta}_i)$ yielding a population optimal correlation coefficient $c_1^*(\boldsymbol{\sigma}) = c_1(\boldsymbol{\sigma}, \boldsymbol{\beta}_1)$, the population correlation coefficient between $y_{1i} = \boldsymbol{\gamma}_{1i}'\mathbf{x}$ and $y_{2i} = \boldsymbol{\gamma}_{2i}'\mathbf{x}$.

The asymptotic joint distribution of the sample optimal correlation coefficients, $c_i^*(\mathbf{s})$, is required for testing for equality of the population optimal correlation coefficients, $c_i^*(\boldsymbol{\sigma})$, using the methods of Section 3. Since the sample optimal weights, $\hat{\boldsymbol{\beta}}_i$, are stochastic and vary from sample to sample, the exact distribution of $\mathbf{c}^*(\mathbf{s})$ will not be easy to obtain. It is known, however, from Proposition 4 that the asymptotic distribution of $\mathbf{c}^*(\mathbf{s}) = \mathbf{c}(\mathbf{s}, \hat{\boldsymbol{\beta}})$ is the same as that of $\mathbf{c}(\mathbf{s}, \boldsymbol{\beta})$. Since the elements, $c_i(\mathbf{s}, \boldsymbol{\beta})$ of $\mathbf{c}(\mathbf{s}, \boldsymbol{\beta})$ are simple correlation coefficients between the linear composites y_{1i} and y_{2i},

$$c_i(\mathbf{s}, \boldsymbol{\beta}) = r(y_{1i}, y_{2i}), \tag{4.4}$$

their joint asymptotic distribution is known from Proposition 2 or Corollary 2.1. This will involve population correlation coefficients and population standardised fourth order moments of the y_{1i} and y_{2i}. These cannot be estimated directly from corresponding sample correlation coefficients and standardized fourth order moments of the y_{1i} and y_{2i}, since the β_{ii} are not known (and the y_{1i} and y_{2i} consequently cannot be calculated). Adequate indirect estimates are available, however, since the sample correlation coefficients and standardized fourth order moments of the $\hat{y}_{1i}$ and $\hat{y}_{2i}$ can be computed and will provide consistent estimates of the population correlation coefficients and standardized fourth order moments of the y_{1i} and y_{2i}.

Therefore, the following procedure may be used to test for equality of elements of $\mathbf{c}^*(\boldsymbol{\sigma})$. Sample optimal weights are calculated, and the $p \times k$ matrices $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are formed with the $\hat{\boldsymbol{\gamma}}_{1i}$ and $\hat{\boldsymbol{\gamma}}_{2i}$ as columns. Then the $2k \times 2k$ sample correlation matrix

$$R^* = \begin{bmatrix} R_{11}^* & R_{12}^* \\ R_{21}^* & R_{22}^* \end{bmatrix} \tag{4.5}$$

of

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_1' \mathbf{x} \\ \hat{\Gamma}_2' \mathbf{x} \end{bmatrix} \tag{4.6}$$

and the standardized fourth order sample moments for $\hat{\mathbf{y}}$ are calculated. The diagonal elements of the $k \times k$ nonsymmetric matrix R_{12}^* are the sample optimal correlations, or elements of $\mathbf{c}^*(\mathbf{s})$. Methods reviewed in Section 3 may then be applied to $\mathbf{c}^*(\mathbf{s})$ to test for equality of elements of $\mathbf{c}^*(\boldsymbol{\sigma})$. The nondiagonal elements of R_{12}^*, R_{11}^*, and R_{22}^* and the standardized fourth order sample moments of $\hat{\mathbf{y}}$ are employed to obtain the estimated covariance matrix, $\hat{\Psi}$, of $\mathbf{c}^*(\mathbf{s})$ using Proposition 2 or Corollary 2.1. Thus, equality of optimal correlations is tested in exactly the same manner as that of simple correlations by replacing the original observations in $\mathbf{x}$ by the composite scores $\hat{\mathbf{y}}$. Existing computer programs [e.g., Steiger, 1979] for testing correlational pattern hypotheses may be employed, though one must remember that *only* pattern hypotheses on population correlations corresponding to diagonal elements of R_{12}^* may be investigated.

5. *Optimal Correlation Coefficients*

In this section we shall review some types of optimal correlation coefficients whose equality can be tested using the approach described in Section 4. Formulae for the sample optimal correlation coefficients will be given. The corresponding population optimal correlation coefficients are obtained by merely replacing the sample covariance matrix, S, by the population covariance matrix, Σ, in the relevant formulae.

For brevity of notation the subscript "i" used previously to distinguish elements of the vector, $\mathbf{c}^*(\mathbf{s})$, of optimal correlation coefficients will be omitted, and $c^*(\mathbf{s})$ will represent a single optimal correlation coefficient. In each case the correlation coefficient optimised to yield $c^*(\mathbf{s})$ is of the form (cf. (4.2))

$$c(\mathbf{s}, \mathbf{b}) = \mathbf{g}_1' S \mathbf{g}_2 \{(\mathbf{g}_1' S \mathbf{g}_1)(\mathbf{g}_2' S \mathbf{g}_2)\}^{-1/2} \tag{5.1}$$

where the two $p \times 1$ vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ are formed from elements of a weight vector $\mathbf{b}$

(usually partitioned as $\mathbf{b} = [\mathbf{b}_1', \mathbf{b}_2']'$), supplemented by constant elements equal to zero or one. The gradient $\partial c(\mathbf{s}, \mathbf{b})/\partial \mathbf{b}$ will be equal to zero when $\mathbf{b}$ assumes an "optimal" value $\hat{\boldsymbol{\beta}}$ and the "optimal correlation coefficient" is $c^*(\mathbf{s}) = c(\mathbf{s}, \hat{\boldsymbol{\beta}})$. In each case we shall define $\mathbf{g}_1$, $\mathbf{g}_2$, and provide $c(\mathbf{s}, \mathbf{b})$, $\partial c(\mathbf{s}, \mathbf{b})/\partial \mathbf{b}$, $\hat{\boldsymbol{\beta}}$, $\hat{y}_1$, and $\hat{y}_2$.

(a) *Multiple Correlations.* Consider, as an example, the multiple correlation of x_1 on $\mathbf{x}_2 = (x_2, x_3, \ldots, x_p)'$. Let

$$S = \begin{bmatrix} s_{11} & \mathbf{s}_{21}' \\ \mathbf{s}_{21} & S_{22} \end{bmatrix}$$

and $\mathbf{g}_1' = (1, \mathbf{0}')$, $\mathbf{g}_2' = (0, \mathbf{b}')$. Then

$$c(\mathbf{s}, \mathbf{b}) = \mathbf{b}'\mathbf{s}_{21}\{s_{11}\mathbf{b}'S_{22}\mathbf{b}\}^{-1/2} \tag{5.2}$$

and the gradient is

$$\partial c(\mathbf{s}, \mathbf{b})/\partial \mathbf{b} = \{\mathbf{s}_{21} - S_{22}\mathbf{b}(\mathbf{b}'\mathbf{s}_{21}/\mathbf{b}'S_{22}\mathbf{b})\}\{s_{11}\mathbf{b}'S_{22}\mathbf{b}\}^{-1/2} \tag{5.3}$$

which is equal to zero at

$$\mathbf{b} = \hat{\boldsymbol{\beta}} = S_{22}^{-1}\mathbf{s}_{21} \tag{5.4}$$

This yields the multiple correlation coefficient

$$c^*(\mathbf{s}) = c(\mathbf{s}, \hat{\boldsymbol{\beta}}) = \{\mathbf{s}_{21}'S_{22}^{-1}\mathbf{s}_{21}/s_{11}\}^{1/2}, \tag{5.5}$$

which is the maximum of $c(\mathbf{s}, \mathbf{b})$ with respect to $\mathbf{b}$. All requirements of Proposition 4 are satisfied, except when $\boldsymbol{\sigma}_{21} = \mathbf{0}$, since $c(\mathbf{s}, \mathbf{b})$ does not exist at $\mathbf{b} = \boldsymbol{\beta} = \mathbf{0}$. Consequently, the methods of Section 4 do not apply if the population multiple correlation coefficient is zero.

The linear composites employed when applying the methods of Section 4 are $\hat{y}_1 = x_1$, the criterion score, and $\hat{y}_2 = \hat{\boldsymbol{\beta}}'\mathbf{x}_2$. Note that $\hat{y}_2$ differs only by an additive constant from the usual predicted criterion score in multiple regression. Since the correlation coefficients in R^* will not be affected by a change of location of $\hat{y}_2$, the usual predicted criterion score may be employed for $\hat{y}_2$.

(b) *Partial Correlations.* As an example we shall consider the partial correlation between x_1 and x_2 eliminating the effect of $\mathbf{x}_3 = (x_3, x_4, \ldots, x_p)'$. Let

$$S = \begin{bmatrix} s_{11} & s_{12} & \mathbf{s}_{31}' \\ s_{21} & s_{22} & \mathbf{s}_{32}' \\ \mathbf{s}_{31} & \mathbf{s}_{32} & S_{33} \end{bmatrix}$$

and $\mathbf{g}_1' = (1, 0, -\mathbf{b}_1')$, $\mathbf{g}_2' = (0, 1, -\mathbf{b}_2')$, with $\mathbf{b}' = (\mathbf{b}_1', \mathbf{b}_2')$. Then

$$c(\mathbf{s}, \mathbf{b}) = v_{12}(v_{11}v_{22})^{-1/2} \tag{5.6}$$

where $v_{jk} = s_{jk} - \mathbf{b}_j'\mathbf{s}_{3k} - \mathbf{b}_k'\mathbf{s}_{3j} + \mathbf{b}_j'S_{33}\mathbf{b}_k$; $j = 1, 2$; $k = 1, 2$ so that the gradient is given by

$$\partial c(\mathbf{s}, \mathbf{b})/\partial \mathbf{b}_j = \{S_{33}\mathbf{b}_m - \mathbf{s}_{3m} - v_{jj}^{-1}v_{jm}(S_{33}\mathbf{b}_j - \mathbf{s}_{3j})\}/v_{mm}^{1/2}; \quad j = 1, 2; \quad m = 3 - j \tag{5.7}$$

which is equal to zero at

$$\mathbf{b}_j = \hat{\boldsymbol{\beta}}_j = S_{33}^{-1}\mathbf{s}_{3j}; \qquad j = 1, 2 \tag{5.8}$$

This is a saddle point of $c(\mathbf{s}, \mathbf{b})$ and yields the partial correlation coefficient

$$c^*(\mathbf{s}) = c(\mathbf{s}, \hat{\boldsymbol{\beta}}) = (s_{12} - \mathbf{s}_{32}'S_{33}^{-1}\mathbf{s}_{32})/\{(s_{11} - \mathbf{s}_{31}'S_{33}^{-1}\mathbf{s}_{31})(s_{11} - \mathbf{s}_{31}'S_{33}^{-1}\mathbf{s}_{31})\}^{1/2} \tag{5.9}$$

All requirements of Proposition 4 are satisfied whenever the population partial correlation $c^*(\boldsymbol{\sigma})$ exists.

The linear composites employed when applying the methods of Section 4 are $\hat{y}_j = x_j - \hat{\boldsymbol{\beta}}_j' \mathbf{x}_3$, $j = 1, 2$, which differ only by an additive constant from, and may be replaced by, the usual regression residuals.

(c) *Canonical Correlations.* Let $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$ where $\mathbf{x}_1$ has p_1 elements and $\mathbf{x}_2$ has p_2 elements, and let

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

with $\mathbf{b}' = (\mathbf{b}_1', \mathbf{b}_2')$.

The correlation coefficient between $\mathbf{b}_1'\mathbf{x}_1$ and $\mathbf{b}_2'\mathbf{x}_2$ is given by substituting $\mathbf{g}_1' = (\mathbf{b}_1', \mathbf{0}')$, $\mathbf{g}_2' = (\mathbf{0}', \mathbf{b}_2')$ into (5.1) to yield

$$c(\mathbf{s}, \mathbf{b}) = \mathbf{b}_1' S_{12} \mathbf{b}_2 / \{(\mathbf{b}_1' S_{11} \mathbf{b}_1)(\mathbf{b}_2' S_{22} \mathbf{b}_2)\}^{1/2} \tag{5.10}$$

The gradient is given by

$$\begin{aligned} \partial c(\mathbf{s}, \mathbf{b})/\partial \mathbf{b}_j &= (\mathbf{b}_j' S_{jj} \mathbf{b}_j)^{-1/2} (\mathbf{b}_m' S_{mm} \mathbf{b}_m)^{-1/2} \\ &\times \{S_{jm} \mathbf{b}_m - (\mathbf{b}_j' S_{jj} \mathbf{b}_j)^{-1} (\mathbf{b}_j' S_{jm} \mathbf{b}_m) S_{jj} \mathbf{b}_j\}; \qquad j = 1, 2;\ m = 3 - j \end{aligned} \tag{5.11}$$

Let $\hat{\lambda}_k$ be the kth largest characteristic root of the matrix $S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}$ $\{k \le \text{Min}\ (p_1, p_2)\}$ and let $\mathbf{u}_k$ be the corresponding characteristic vector standardized so that $\mathbf{u}_k' S_{kk} \mathbf{u}_k = 1$.

The gradient is equal to zero at

$$\mathbf{b}_1 = \hat{\boldsymbol{\beta}}_1 = \mathbf{u}_k. \qquad k = 1, 2, \ldots, \text{Min}\ (p_1, p_2) \tag{5.12a}$$

$$\mathbf{b}_2 = \hat{\boldsymbol{\beta}}_2 = S_{22}^{-1} S_{21} \hat{\boldsymbol{\beta}}_1 \hat{\lambda}_k^{-1/2} \tag{5.12b}$$

and the corresponding value of $c(\mathbf{s}, \mathbf{b})$ is the kth canonical correlation,

$$c^*(\mathbf{s}) = c(\mathbf{s}, \hat{\boldsymbol{\beta}}) = \hat{\lambda}_k^{-1/2} \tag{5.13}$$

Thus, $c(\mathbf{s}, \mathbf{b})$ has Min (p_1, p_2) stationary points, and the stationary values are the canonical correlations. The requirements of Proposition 4 are satisfied if and only if the population canonical correlation under consideration is distinct (since a characteristic vector is differentiable if and only if the corresponding characteristic root is distinct). Note that this implies that the methods of Section 4 do not apply if all canonical correlations are zero. While the methods of Section 4 are applicable to all canonical correlations, in most practical situations the largest canonical correlation will be of interest.

The required linear composites are $\hat{y}_1 = \hat{\boldsymbol{\beta}}_1' \mathbf{x}_1$ and $\hat{y}_2 = \hat{\boldsymbol{\beta}}_2' \mathbf{x}_2$, but may be replaced by the usual canonical scores since this only involves a change of location.

(d) *An internal correlation coefficient.* This optimal correlation coefficient was derived by Venables (1976) as a union-intersection test statistic for sphericity. Let $\mathbf{g}_1 = \mathbf{b}_1$ and $\mathbf{g}_2 = \mathbf{b}_2$ in (4.2) so that

$$c(\mathbf{s}, \mathbf{b}) = \mathbf{b}_1' S \mathbf{b}_2 / \{(\mathbf{b}_1' S \mathbf{b}_1) \mathbf{b}_2' S \mathbf{b}_2)\}^{1/2} \tag{5.14}$$

This correlation coefficient is maximized subject to the restriction $\mathbf{b}_1'\mathbf{b}_2 = 0$, so that Proposition 4 will be applied to the Lagrangian function

$$f(\mathbf{s}, \mathbf{b}, m) = c(\mathbf{s}, \mathbf{b}) - m\mathbf{b}_1'\mathbf{b}_2 \tag{5.15}$$

rather than to $c(\mathbf{s}, \mathbf{b})$ [cf. Shapiro, 1983, Sections 4 and 6]. The gradient is given by:

$$\partial f/\partial \mathbf{b}_1 = \{(\mathbf{b}_1' S\mathbf{b}_1)(\mathbf{b}_2' S\mathbf{b}_2)\}^{-1/2}\{S\mathbf{b}_2 - (\mathbf{b}_1' S\mathbf{b}_1)^{-1}(\mathbf{b}_1' S\mathbf{b}_2)S\mathbf{b}_1\} - m\mathbf{b}_2 \tag{5.16a}$$

$$\partial f/\partial \mathbf{b}_2 = \{(\mathbf{b}_1' S\mathbf{b}_1)(\mathbf{b}_2' S\mathbf{b}_2)\}^{-1/2}\{S\mathbf{b}_1 - (\mathbf{b}_2' S\mathbf{b}_2)^{-1}(\mathbf{b}_1' S\mathbf{b}_2)S\mathbf{b}_2\} - m\mathbf{b}_1 \tag{5.16b}$$

$$\partial f/\partial m = \mathbf{b}_1'\mathbf{b}_2 \tag{5.16c}$$

Let $\hat{\lambda}_1$ and $\hat{\lambda}_p$ be two characteristic roots of S, and let $\mathbf{u}_1$ and $\mathbf{u}_p$ be the corresponding characteristic vectors standardized to unit length. The gradient is equal to the null vector if

$$\mathbf{b}_1 = \hat{\boldsymbol{\beta}}_1 = \mathbf{u}_1 + \mathbf{u}_p \tag{5.17a}$$

$$\mathbf{b}_2 = \hat{\boldsymbol{\beta}}_2 = \mathbf{u}_1 - \mathbf{u}_p \tag{5.17b}$$

$$m = \hat{\mu} = 2\hat{\lambda}_1\hat{\lambda}_p/(\hat{\lambda}_1 + \hat{\lambda}_p)^2 \tag{5.17c}$$

and the corresponding function value is

$$c^*(\mathbf{s}) = f(\mathbf{s}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\mu}}) = c(\mathbf{s}, \hat{\boldsymbol{\beta}}) = (\hat{\lambda}_1 - \hat{\lambda}_p)/(\hat{\lambda}_1 + \hat{\lambda}_p) \tag{5.18}$$

In order to maximize $c^*(\mathbf{s})$, $\hat{\lambda}_1$ and $\hat{\lambda}_p$ are chosen to be the largest and smallest characteristic roots of S.

The requirements of Proposition 4 are satisfied if and only if both the largest and smallest characteristic roots of Σ are distinct. Consequently, the methods of Section 4 do not apply if the population internal correlation coefficient $c^*(\boldsymbol{\sigma})$ is zero.

The linear composites employed are $\hat{y}_1 = \mathbf{u}_1'\mathbf{x} + \mathbf{u}_p'\mathbf{x}$ and $\hat{y}_2 = \mathbf{u}_1'\mathbf{x} - \mathbf{u}_p'\mathbf{x}$.

A different internal correlation coefficient has been proposed by Schuenemeyer and Bargmann (1978) as a union-intersection test statistic for independence. The covariance matrix, S is replaced throughout by the correlation matrix, R. The methods of Section 4 do not apply to this coefficient, since replacement of S by R in (5.14), and fixing $\mathbf{b}_1$ and $\mathbf{b}_2$ does not yield a statistic with the distribution of a simple correlation coefficient.

6. *Some Monte Carlo Evidence*

The statistics we have described are easy to compute, given that a program capable of testing for equality of correlations is available. One must keep in mind, however, that these are *large sample* statistics, and that, for asymptotically distribution free (ADF) tests, estimation of 4th order standardized moments is required. The r_{ijkh} will have large standard errors at small sample sizes, and this may lead to unstable estimates of the $\psi_{ij,kh}$ under some conditions. Hence, these statistics must be used with considerable caution when N is not truly large.

However, we also emphasize that, although caution is necessary, the above concerns are not cause for rejecting this approach. One must remember that the estimates of the $\psi_{ij,kh}$ are rather complex combinations of the r_{ijkh}, and that this recombination, which is compounded in effect when a chi-square quadratic form statistic of equation (3.3) or (3.7) is calculated, may lead to considerable "smoothing" of the performance of the test statistic. Given the asymptotic properties, the "smoothing" must, of course, occur eventually, and the question of interest is whether it occurs at reasonably small sample sizes. One way of approaching this question is through Monte Carlo simulation studies. Unfortunately, these studies, when applied to methods as complex and flexible as ours, have serious limitations, and are (if results are positive, at least) seldom conclusive. There are many hypotheses of interest, many parameters which might be varied, and the simulation (though the statistics are inexpensive in individual applications) requires a considerable investment in computer time, even when only a few special cases are examined.

Here we will make no pretense of being exhaustive in our assessment; indeed, we will be highly selective. Our purpose is simply to alleviate justifiable concerns with some relevant evidence. It should be remembered that when the assumption of an elliptical or multivariate normal distribution is tenable, the approach we recommend may be used with (2.16), and the ρ_{ijkh} of (2.12) need not be estimated.

We now present some brief Monte Carlo evidence assessing the Type I error rate performance of ADF tests on multiple and partial correlations. We examine the performance of the tests under two parent distributions, the multivariate normal and the lognormal. The lognormal, as has been demonstrated on several occasions (see, e.g., Duncan & Layard, 1973; Steiger, Note 2) produces, because of its high kurtosis, extremely poor performance on a variety of Normal Theory correlation tests. It is a severe test of any ADF procedure.

In the studies which follow, multivariate normal distributions were generated by (a) taking uniform independent variates produced by a linear congruential random number generator, (b) using the rectangle wedge-tail algorithm (Knuth, 1969) to obtain independent standard normal variates, and (c) linearly recombining these variates using a Cholesky factor of the desired correlation matrix. Lognormal random numbers with the desired correlational structure were produced in much the same way, except that, in stage (c) above, a Cholesky factor of a "predistorted" correlation matrix with elements $\rho_{ij}^* = \ln[\rho_{ij}(e-1)+1]$, with ρ_{ij} the desired correlations, was used. Then, a stage (d) was added in which the normal random numbers generated in stage (c) were transformed via $y = \exp(-x)$, yielding, by well-known theory, marginally lognormal variates with the desired correlations ρ_{ij}. In all conditions there were 500 Monte Carlo replications. In all cases, tests were based on a sample size of 150. We now describe the two main simulation conditions, and the data obtained.

(A) *Equality of Dependent Multiple Correlations.* A criterion and a set of predictors were each measured twice. The two multiple correlations were then compared for equality, using the ADF procedure described in Sections 4 and 5. Specifically, the criterion and predicted scores from each regression analysis were treated as 4 variables, and the hypothesis of equal multiple correlations was tested as a correlational pattern hypothesis of the form $\rho_{12} = \rho_{34}$, using the methods of Section 3. In this case the vector $\boldsymbol{\rho}$ in Equations (3.1) and (3.4) contains just two elements, while $M = (1, -1)$ and $A' = (1, 1)$.

For simplicity, all off-diagonal correlations in the population correlation matrices were set to a common value.

A $2 \times 2 \times 3$ (Distribution $\times$ Number of Predictors $\times$ Level of Correlation) factorial design was employed, in which there were two distributions (MVN and Lognormal), 2 sizes of predictor set (2 and 5 variables) and 3 levels of correlation (all off-diagonal elements of P were either .3, .6, or .9).

Empirical Type I error rates were tabulated for a number of nominal cut-off points. For simplicity, we present data from $p = .05$ only, since data from other points are essentially redundant. The results in Table 1 are, basically, quite encouraging, especially when one remembers that, if the true Type I error rate is .050, the standard error of our estimates is about .007. In this context, we see that the results for the normal distribution are, essentially, "right on" while those for the lognormal show a slight, but reasonably tolerable tendency toward excessive rejections.

(B) *Equality of* ρ_{12} *and* $\rho_{12.3}$. To test this hypothesis, the variables $x_{1.3}$ and $x_{2.3}$ were calculated as regression residuals from the sample linear regression, and were input along with x_1 and x_2 as four variables to the ADF procedure. The null hypothesis was then tested as a hypothesis of the form $\rho_{12} = \rho_{34}$.

The hypothesis was tested on 15 different population correlation matrices, repre-

TABLE 1.

Empirical Type I Error Rates (Nominal p = .05)
Test Comparing 2 Dependent Multiple Correlations

Distribution	Correlation	Number of Predictors 2	5
Normal	.3	.054	.050
	.6	.050	.046
	.9	.046	.046
Lognormal	.3	.062	.048
	.6	.062	.070
	.9	.054	.074

senting a variety of different values. The conditions, and the resulting empirical Type I error rates (nominal $p = .05$), are summarized in Table 2.

Overall, the test procedure performs well with both normal and lognormal data. A notable exception seems to be condition 1. In this condition, r_{12} and $r_{12.3}$ are highly correlated, and so the variance-covariance matrix of r_{12} and $r_{12.3}$ is very nearly singular. This may well lead to an unstable estimate of the variance-covariance matrix of these two statistics in this case.

TABLE 2.

Empirical Type I Error Rates (p=.05)
Test for Comparing ρ_{12} with $\rho_{12.3}$

Condition	ρ_{13}	ρ_{23}	$\rho_{12}(=\rho_{12.3})$	Normal	Lognormal
1	.10	.30	.5901	.018	.018
2	.10	.50	.3615	.048	.046
3	.10	.70	.2418	.054	.060
4	.10	.90	.1589	.038	.048
5	.10	.95	.1378	.066	.086
6	.30	.50	.8627	.060	.050
7	.30	.70	.6588	.038	.054
8	.30	.90	.4622	.054	.054
9	.30	.95	.4059	.062	.060
10	.50	.70	.9173	.042	.044
11	.50	.90	.7229	.054	.048
12	.50	.95	.6511	.062	.050
13	.70	.90	.9148	.048	.042
14	.70	.95	.8558	.036	.038
15	.90	.95	.9897	.048	.034

7. *Conclusions*

Results have been presented which lead to a variety of interesting asymptotic significance tests on multiple, partial, and canonical correlations. The technique involves input of ordinary predicted scores, residual scores, and/or canonical variate scores into existing programs for testing pattern hypotheses on correlations. The technique allows tests on functions of a correlation matrix without extensive analytic or numerical differentiation, since it does not require use of the multivariate delta theorem. Preliminary Monte Carlo results indicated that two simple tests for (a) comparing two multiple correlations, and (b) comparing a correlation with a partial correlation worked rather well for either normal or lognormal data. Additional Monte Carlo research should be directed toward particular applications of special interest.

REFERENCE NOTES

1. Steiger, J. H. (1980). *K*-sample pattern hypotheses on correlation matrices by the method of generalized least squares. University of British Columbia, Institute of Applied Mathematics and Statistics Research Bulletin 80-2.
2. Steiger, J. H. (June 2, 1982). A robust large-sample procedure for comparing dependent correlations. Paper presented at the annual Spring Meeting of the Psychometric Society.

REFERENCES

Browne, M. W. (1977). The analysis of patterned correlation matrices by generalized least squares. *British Journal of Mathematical and Statistical Psychology, 30*, 113–124.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis.* Cambridge: Cambridge University Press.

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1976). Some multivariate applications of elliptical distributions. In S. Ideka (Ed.), *Essays in probability and statistics.* Tokyo: Shinko Tsusho.

Duncan, G. T., & Layard, M. W. J. (1973). A Monte Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika, 60*, 551–558.

Hsu, P. L. (1949). The limiting distribution of functions of sample means and application to testing hypotheses. *Proceedings of the First Berkely Symposium on Mathematical Statistics and Probability*, 359–402.

Isserlis, L. (1916). On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika, 11*, 185–190.

Knuth, D. E. (1969). *The art of computer programming. Vol. 2.* Reading, Mass.: Addison-Wesley.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443–482.

Lord, F. M. (1975). Automated hypothesis tests and standard errors for nonstandard problems. *The American Statistician, 29*, 56–59.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness to testing normality and to robustness studies. *Sankhya*, Series B, *36*, 115–128.

Mardia, K. V., & Zemroch, P. J. (1975). Algorithm AS84: Measures of multivariate skewness and kurtosis. *Applied Statistics, 24*, 262–264.

Muirhead, R. J., & Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal distributions. *Biometrika, 67*, 31–43.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* New York: Wiley.

Olkin, I., & Siotani, M. (1976). Asymptotic distribution of functions of a correlation matrix. Technical Report No. 6, Laboratory for Quantitative Research in Education. Stanford University, 1964. Reprinted in S. Ideka (Ed.) *Essays in probability and statistics*, Tokyo, Shinko Tsusho.

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

Rao, C. R. (1973). *Linear statistical inference and its applications* (*2nd Ed.*). New York: Wiley.

Ross, G. J. S. (1970). The efficient use of function minimisation in nonlinear maximum likelihood estimation. *Applied Statistics, 19*, 205–221.

Schuenemeyer, J. H., & Bargmann, R. E. (1978). Maximum eccentricity as a union-intersection test statistic in multivariate analysis. *Journal of Multivariate Analysis, 8*, 268–273.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal, 17*, 33–81.

Steiger, J. H. (1979). MULTICORR: A computer program for fast, accurate, small-sample testing of correlational pattern hypotheses. *Educational and Psychological Measurement, 39*, 677–680.

Steiger, J. H. (1980a). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.

Steiger, J. H. (1980b). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research, 15*, 335–352.

Steiger, J. H., & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and 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 and Statistical Psychology, 36*, 157.

Venables, W. (1976). Some implications of the union-intersection principle for tests of sphericity. *Journal of Multivariate Analysis, 6*, 175–190.

Manuscript received 3/14/83
First revision received 7/11/83
Final version received 8/18/83