British Journal of Mathematical and Statistical Psychology (1984), 37, 62–83 *Printed in Great Britain*

Asymptotically distribution-free methods for the analysis of covariance structures

M. W. Browne

Methods for obtaining tests of fit of structural models for covariance matrices and estimator standard errors which are asymptotically distribution free are derived. Modifications to standard normal theory tests and standard errors which make them applicable to the wider class of elliptical distributions are provided. A random sampling experiment to investigate some of the proposed methods is described.

1. Introduction

Many methods for the analysis of covariance matrices which are in frequent use involve an assumption of multivariate normality. The corresponding tests of hypotheses can give misleading results when this assumption is violated. Layard (1972, 1974) pointed this out and proposed a number of asymptotically robust tests for the equality of covariance matrices. One of these was a quadratic form statistic using a general expression for an asymptotic covariance matrix (Layard, 1972, Section 3.2; 1974, p. 462). This provided the initial impetus for the present research where it was found that minimization of an appropriate quadratic form yielded tests of fit of structural models for covariance matrices which were asymptotically insensitive to the distribution of the observations.

A substantial amount of computation was involved and an alternative requiring less computation was sought. Empirical observations on a number of data sets indicated that the ratio of the usual normal theory likelihood ratio test statistic to the newly developed asymptotically distribution-free test statistic tended to be roughly similar in magnitude to the ratio of the sample based estimate of Mardia's (1970, 1974) coefficient of multivariate kurtosis to the corresponding coefficient for a multivariate normal distribution. This suggested a simple 'correction for kurtosis' to the usual likelihood ratio test statistic. It was found that a theoretical justification for this correction could be provided for most models under the assumption of a distribution belonging to the elliptical class (cf. Devlin *et al.*, 1976; Muirhead, 1982). This class of distributions includes the multivariate normal distribution but also contains platykurtic and leptokurtic distributions. Consequently the assumption of a distribution from the elliptical class is substantially less restrictive than the usual assumption of multivariate normality.

The main findings of this research were presented concisely and without proof in a recent survey of methods for the analysis of covariance structures (Browne, 1982). In the present paper the procedures developed are justified theoretically and evaluated empirically. Theoretical aspects are covered in Section 2. The majority of estimators of parameters in structural models for covariance matrices in current use are members of the class of generalized least squares (GLS) estimators. This includes the maximum Wishart likelihood estimators (MWL) in the sense that a generalized least squares discrepancy function can be constructed which will in general attain its minimum at the point defined by the MWL estimator (Browne, 1974, Proposition 6). It will be seen that the test statistics and estimator standard errors associated with the MWL or asymptotically equivalent normal theory GLS estimators are inappropriate for any multivariate distribution whose kurtosis differs from that of the normal distribution. The reason for this is that the matrix of the quadratic form of the GLS discrepancy function then is misspecified. Formulae for the asymptotic covariance matrix of the estimators and for a test statistic with an asymptotic chi-square distribution even when the matrix of the GLS quadratic form is misspecified will be given. Because it is doubtful that any covariance structure is a true representation of reality in most practical situations, the possibility that the model does not hold exactly will be borne in mind. Results which are valid both when the model holds and under moderate lack of fit of the model will be obtained. Correct specification of the matrix of the GLS quadratic form will result in estimators with minimum asymptotic variances. Estimators of this type which are valid for any distribution with finite eighth-order moments are considered in Section 3. Section 4 shows that simple corrections for kurtosis to the test statistic and asymptotic estimator covariance matrix make normal theory results applicable to all other members of the elliptical class of distributions for all models satisfying a mild requirement of scale invariance. This family of models incorporates virtually all models which would be of use in practical situations, including all models appropriate for correlation matrices. The correction proposed for the test statistic satisfies the requirements of Muirhead & Waternaux (1980) and of Muirhead (1982) who have previously derived corrections for normal theory tests of fit of some models satisfying the scale invariance requirement. Implementation of the corrections to the likelihood ratio test statistic and estimator standard errors would involve minor changes to existing programs. In Section 5 a random sampling experiment is described to give a rough impression of the applicability of some of the asymptotic results described previously to moderate samples. General conclusions and suggestions for future research are contained in Section 6.

There has been a fair amount of related research on the sensitivity to non-normality of tests concerning correlation coefficients (e.g. Gayen, 1951; Duncan & Layard, 1973; Kraemer, 1980; Yu & Dunn, 1983). Steiger & Hakstian (1982, 1983) have shown how asymptotically distribution-free tests for correlation coefficients can be constructed using a general expression for the asymptotic distribution of elements of a correlation matrix. These tests are also applicable to correlation coefficients, such as canonical, multiple and partial correlation coefficients, which involve optimal linear composites (Steiger & Browne, 1984). Additional material on asymptotically distribution-free methods may be found in Bentler (1983) and De Leeuw (1983).

2. Theoretical aspects

Suppose that S is the usual unbiased estimator of a $p \times p$ population covariance matrix $\boldsymbol{\Sigma}_0$ obtained from $N = n+1$ independent observations on a $p \times 1$ vector variate $\mathbf{x}$ which has a distribution with finite fourth-order moments. Let $\mathbf{s} = \text{vecs}\,(\mathbf{S})$ and $\boldsymbol{\sigma}_0 = \text{vecs}\,(\boldsymbol{\Sigma}_0)$ where vecs (S) represents a $p^* \times 1$ column vector formed from the $p^* = \frac{1}{2}p(p+1)$ non-duplicated elements $[\mathbf{S}]_{11}, [\mathbf{S}]_{12}, [\mathbf{S}]_{22}, [\mathbf{S}]_{13}, [\mathbf{S}]_{23}, [\mathbf{S}]_{33}, \ldots$ of a $p \times p$ symmetric matrix S (cf. Nel, 1980) and $[S]_{ij}$ represents the element in the ith row and jth column of S. The subscript has been introduced in $\boldsymbol{\Sigma}_0$ to distinguish a specific population covariance matrix from any covariance matrix, $\boldsymbol{\Sigma}$, satisfying a particular model. If $\boldsymbol{\delta}_s = n^{\frac{1}{2}}(\mathbf{s}-\boldsymbol{\sigma}_0)$, the finite sample distribution of $\boldsymbol{\delta}_s$ has a null mean vector and covariance matrix $\mathbf{Y} = \text{cov}\,(\boldsymbol{\delta}_s, \boldsymbol{\delta}_s')$ with typical element (Kendall & Stuart, 1969, Section 13.16)

$$[\Upsilon]_{ij,kl} = \mathrm{cov}\,(n^{\frac{1}{2}}[\mathrm{S}-\Sigma_0]_{ij},\, n^{\frac{1}{2}}[\mathrm{S}-\Sigma_0]_{kl}) = n\,\mathrm{cov}\,([\mathrm{S}-\Sigma_0]_{ij}, [\mathrm{S}-\Sigma_0]_{kl})$$
$$= [\Sigma_0]_{ik}[\Sigma_0]_{jl}+[\Sigma_0]_{il}[\Sigma_0]_{jk}+(n/N)\,\kappa_{ijkl}, \tag{2.1}$$

where κ_{ijkl} is a fourth-order cumulant given by

$$\mathrm{k}_{ijkl} = \sigma_{ijkl}-[\Sigma_0]_{ij}[\Sigma_0]_{kl}-[\Sigma_0]_{ik}[\Sigma_0]_{jl}-[\Sigma_0]_{il}[\Sigma_0]_{jk},$$

with

$$\sigma_{ijkl} = \mathscr{E}(x_i-\xi_i)\,(x_j-\xi_j)\,(x_k-\xi_k)\,(x_l-\xi_l),$$
$$\xi_i = \mathscr{E}(x_i).$$

For consistency of notation with the use of Σ_0 for the specific population covariance matrix under consideration, the subscript o should be introduced into Υ, ξ_i, κ_{ijkl} and σ_{ijkl}. This has been avoided in order to retain some brevity of notation. No confusion will result since Υ, ξ_i, κ_{ijkl} and σ_{ijkl} are not regarded as functions of $\boldsymbol{\gamma}$.

In many situations the exact finite sample distribution of $\boldsymbol{\delta}_s$ is not known but may be approximated by the asymptotic distribution as $n \to \infty$ which is multivariate normal with a null mean vector and covariance matrix $\bar{\Upsilon} = \mathrm{L}\,\mathrm{cov}\,(\boldsymbol{\delta}_s, \boldsymbol{\delta}_s') = \lim_{n\to\infty} \mathrm{cov}\,(\boldsymbol{\delta}_s, \boldsymbol{\delta}_s')$ with typical element

$$[\bar{\Upsilon}]_{ij,kl} = [\Sigma_0]_{ik}[\Sigma_0]_{jl}+[\Sigma_0]_{il}[\Sigma_0]_{jk}+\kappa_{ijkl}$$
$$= \sigma_{ijkl}-[\Sigma_0]_{ij}[\Sigma_0]_{kl}. \tag{2.2}$$

If all fourth-order cumulants are equal to zero, then $\Upsilon = \bar{\Upsilon}$ with typical element

$$[\bar{\Upsilon}]_{ij,kl} = [\Sigma_0]_{ik}[\Sigma_0]_{jl}+[\Sigma_0]_{il}[\Sigma_0]_{jk}. \tag{2.3}$$

If this is the case all marginal kurtosis coefficients as well as Mardia's (1970, 1974) multivariate kurtosis coefficient have the same values as those for a multivariate normal distribution. We shall then say that the multivariate distribution of x 'has no kurtosis'.

A structural model for a covariance matrix is a $p \times p$ symmetric matrix valued function, $\Sigma = \Sigma(\boldsymbol{\gamma})$, of a $q \times 1$ vector $\boldsymbol{\gamma} \in G$ where G is a parameter set contained in $\mathbb{R}^q$. It will sometimes be convenient to express the model in an equivalent vector form, $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\gamma})$, where $\boldsymbol{\sigma}(\boldsymbol{\gamma}) = \mathrm{vecs}\,\{\Sigma(\boldsymbol{\gamma})\}$. The Jacobian matrix of $\boldsymbol{\sigma}(\boldsymbol{\gamma})$ is the $p^* \times q$ matrix valued function of $\boldsymbol{\gamma}$,

$$\Delta = \Delta(\boldsymbol{\gamma}) = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\gamma}'}. \tag{2.4}$$

The model is said to hold if there exists a $\boldsymbol{\gamma}_0 \in G$ such that $\Sigma_0 = \Sigma(\boldsymbol{\gamma}_0)$. Given a sample covariance matrix, S, an estimate $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ may be obtained by minimizing a discrepancy function. This is a scalar valued function $F(\mathrm{S}, \Sigma)$ of two $p \times p$ symmetric matrices S and Σ with the following properties:

(i) $F(\mathrm{S}, \Sigma) \geqslant 0$.

(ii) $F(\mathrm{S}, \Sigma) = 0$ *if and only if* $\Sigma = \mathrm{S}$.

(iii) $F(\mathrm{S}, \Sigma)$ *is a twice continuously differentiable function of* S *and* Σ.

A discrepancy function $F(\mathrm{S}, \Sigma)$ need not be symmetric in S and Σ in that $F(\mathrm{S}, \Sigma)$ need not be equal to $F(\Sigma, \mathrm{S})$. Examples of discrepancy functions are given in (2.7), (2.9) and (2.10).

If the estimate, $\hat{\gamma}$, is obtained by minimising some discrepancy function $F(\mathbf{S}, \boldsymbol{\Sigma})$, then

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\gamma})) = \min_{\gamma \in G} F(\mathbf{S}, \boldsymbol{\Sigma}(\gamma)). \tag{2.5}$$

The reproduced covariance matrix will be denoted by $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\gamma})$.

A more general definition of the population value, γ_0, of the parameter vector, γ, which is still valid if $\boldsymbol{\Sigma}_0$ can only be approximated by the structural model will be convenient. We shall regard γ_0 as the value of γ which minimizes $F(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\gamma))$ and represent $\boldsymbol{\Sigma}(\gamma_0)$ by $\hat{\boldsymbol{\Sigma}}_0$ (which need not be equal to $\boldsymbol{\Sigma}_0$), i.e.

$$\min_{\gamma \in G} F(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\gamma)) = F(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\gamma_0)) = F(\boldsymbol{\Sigma}_0, \hat{\boldsymbol{\Sigma}}_0). \tag{2.6}$$

This definition of γ_0 depends on the particular discrepancy function employed unless the model holds so that $\boldsymbol{\Sigma}(\gamma_0) = \boldsymbol{\Sigma}_0$ and $F(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\gamma_0)) = 0$ for all discrepancy functions which satisfy the requirements specified earlier.

This paper will be concerned predominantly with quadratic form discrepancy functions of the type

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\gamma) \mid \mathbf{U}) = (\mathbf{s} - \boldsymbol{\sigma}(\gamma))' \mathbf{U}^{-1} (\mathbf{s} - \boldsymbol{\sigma}(\gamma)), \tag{2.7}$$

where $\mathbf{U}$ is a $p^* \times p^*$ positive definite matrix. In many applications $\mathbf{U}$ is a stochastic matrix which converges in probability to a positive definite matrix $\bar{\mathbf{U}}$ as $n \to \infty$. Sometimes, however, $\mathbf{U} = \bar{\mathbf{U}}$ may be a fixed matrix such as the identity matrix. It will be seen that generalized least squares (GLS) estimators obtained by minimizing $F(\mathbf{S}, \boldsymbol{\Sigma}(\gamma) \mid \mathbf{U})$ have minimum asymptotic variances if $\bar{\mathbf{U}} = \bar{\mathbf{Y}}$. If $\bar{\mathbf{U}} \neq \bar{\mathbf{Y}}$ the matrix $\mathbf{U}$ of the quadratic form discrepancy function (2.6) will be said to be misspecified. This is a situation which can easily occur in practical situations and which will be considered explicitly here.

If the distribution of $\mathbf{x}$ is assumed to have no kurtosis, so that (2.3) holds, then it is desirable for a typical element of $\mathbf{U}$ to have the corresponding form

$$[\mathbf{U}]_{ij,kl} = [\mathbf{V}]_{ik}[\mathbf{V}]_{jl} + [\mathbf{V}]_{il}[\mathbf{V}]_{jk}, \tag{2.8}$$

where $\mathbf{V}$ is a $p \times p$ positive definite stochastic matrix which converges in probability to a positive definite matrix $\bar{\mathbf{V}}$ as $n \to \infty$. If (2.8) holds, then the quadratic form discrepancy function $F(\mathbf{S}, \boldsymbol{\Sigma} \mid \mathbf{U})$ in (2.7) is equal (Browne, 1974) to

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\gamma) \mid \mathbf{V}) = \tfrac{1}{2} \operatorname{tr} [(\mathbf{S} - \boldsymbol{\Sigma}(\gamma)) \mathbf{V}^{-1}]^2, \tag{2.9}$$

which is more easily computed. A possible choice of $\mathbf{V}$ is $\mathbf{V} = \mathbf{S}$, so that $\bar{\mathbf{V}} = \boldsymbol{\Sigma}_0$. If the distribution of $\mathbf{x}$ has no kurtosis this implies that $\bar{\mathbf{U}} = \bar{\mathbf{Y}}$.

Under the assumption of a multivariate normal distribution for $\mathbf{x}$ or, equivalently, a Wishart distribution for $\mathbf{S}$, the Maximum Wishart Likelihood (MWL) estimate, $\hat{\gamma}_{\mathrm{W}}$, may be obtained by minimizing

$$F_{\mathrm{W}}(\mathbf{S}, \boldsymbol{\Sigma}(\gamma)) = \ln |\boldsymbol{\Sigma}(\gamma)| - \ln |\mathbf{S}| + \operatorname{tr} [\mathbf{S}\{\boldsymbol{\Sigma}(\gamma)\}^{-1}] - p. \tag{2.10}$$

Although (2.10) is not a quadratic form discrepancy function, $\hat{\gamma}_{\mathrm{W}}$ may be regarded as a generalized least squares estimate because it will in general also minimize the quadratic form discrepancy function $F(\mathbf{S}, \boldsymbol{\Sigma}(\gamma) \mid \mathbf{V} = \boldsymbol{\Sigma}(\hat{\gamma}_{\mathrm{W}}))$ (cf. Browne, 1974, Proposition 6).

Asymptotic distribution theory for GLS estimators and associated test statistics will be provided under the assumption of some regularity conditions which will now be specified.

(R1) $\bar{\Upsilon}$ *is positive definite*
Remark. If the distribution of $\mathbf{x}$ has no kurtosis, (R1) is equivalent to the condition that Σ_0 be positive definite.

(R2) $F(\Sigma_0, \Sigma(\gamma))$ *has a unique minimum on* G *at* $\gamma = \gamma_0$.
Remark. In the terminology of Shapiro (1983, Definition 2.1) γ is 'conditionally identified' with respect to $F(\cdot,\cdot)$ at γ_0 given Σ_0. If $F(\Sigma_0, \Sigma(\gamma_0)) = 0$, conditional identification implies identification in the usual sense (i.e. $\Sigma(\gamma_*) = \Sigma(\gamma_0), \gamma_* \in G$, implies that $\gamma_* = \gamma_0$).

(R3) γ_0 *is an interior point of* G.

(R4) $\Delta_0 = \Delta(\gamma_0)$ *is of full column rank* q.

(R5) $\| \Sigma_0 - \Sigma(\gamma_0) \|$ *is* $O(n^{-\frac{1}{2}})$.
Remark. A regularity condition similar to (R5) is generally assumed in the derivative of non-central chi-square asymptotic distributions for test statistics (cf. Stroud, 1972; Kendall & Stuart, 1979, p. 247). Condition (R5) assumes that systematic errors due to lack of fit of the model to the population covariance matrix are not large relative to random sampling errors in $\mathbf{S}$. A sequence of population covariance matrices converging in the limit to a matrix satisfying the model is implied (cf. Stroud, 1972). Clearly (R5) is always satisfied if the structural model holds (i.e. $\Sigma_0 = \Sigma(\gamma_0)$).
Methods provided by Shapiro (1983, Theorems 5.4, 5.6) could be employed to derive asymptotic distributions without the imposition of (R5). The resulting formulae would be more complicated and would involve eighth-order moments of the distribution of $\mathbf{x}$.

(R6) *The parameter set* G *is closed and bounded*
Remark. It is not always easy to decide whether it is reasonable to assume (R6) in a specific situation. Because of regularity condition (R7) and property (iii) of a discrepancy function, (R6) may be assumed (cf. Shapiro, 1983, Definition 2.2; Shapiro, 1984) if (but not only if) the following more easily understood condition involving both the discrepancy function and the model is satisfied:

(R6*) Given any $\mathbf{S}$, $F(\mathbf{S}, \Sigma) \to \infty$ if $\| \Sigma \| = \{\mathrm{tr}(\Sigma^2)\}^{\frac{1}{2}} \to \infty$ and $\| \Sigma(\gamma) \| \to \infty$ if $\| \gamma \| = \{\gamma'\gamma\}^{\frac{1}{2}} \to \infty$.

(R7) $\Delta(\gamma)$ *and, consequently,* $\Sigma(\gamma)$ *are continuous functions of* γ
We first investigate the consistency of minimum discrepancy function estimators. Use of property (iii) of a discrepancy function and regularity conditions (R2), (R6) and (R7) will show that the conditions of Theorem 5.1 of Shapiro (1983) are satisfied. This implies the following result.

Proposition 1. A minimum discrepancy function estimator defined in (2.5) is a consistent estimator for γ_0, as defined by (2.6). ■

Together with (R7) and (R5), Proposition 1 implies that $\| \Sigma(\hat{\gamma}_W) - \Sigma_0 \|$ converges in probability to zero as $n \to \infty$ so that the MWL estimator is equivalent to a GLS estimator with

$$[\bar{\mathbf{U}}]_{ij,kl} = [\mathbf{\Sigma}_0]_{ik}[\mathbf{\Sigma}_0]_{jl} + [\mathbf{\Sigma}_0]_{il}[\mathbf{\Sigma}_0]_{jk}. \tag{2.11}$$

The asymptotic distribution of GLS estimators is now given.

Proposition 2 (cf. Browne, 1982 (1.6.4)). If $\hat{\gamma}$ is a GLS estimator obtained by minimizing $F(\mathbf{S}, \mathbf{\Sigma}(\gamma)\,|\,\mathbf{U})$ in (2.7) then the asymptotic distribution of $\hat{\boldsymbol{\delta}}_\gamma = n^{\frac{1}{2}}(\hat{\gamma} - \gamma_0)$ is multivariate normal with a null mean vector and covariance matrix

$$\mathrm{L}\operatorname{cov}(\hat{\boldsymbol{\delta}}_\gamma \hat{\boldsymbol{\delta}}_\gamma') = \{\mathbf{\Delta}_0' \bar{\mathbf{U}}^{-1} \mathbf{\Delta}_0'\}^{-1} \mathbf{\Delta}_0' \bar{\mathbf{U}}^{-1} \bar{\mathbf{Y}} \bar{\mathbf{U}}^{-1} \mathbf{\Delta}_0 \{\mathbf{\Delta}_0' \bar{\mathbf{U}}^{-1} \mathbf{\Delta}_0\}^{-1}, \tag{2.12a}$$

where $\mathbf{\Delta}_0 = \mathbf{\Delta}(\gamma_0)$. Equivalently

$$\mathrm{L}\operatorname{cov}(\hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\gamma') = \{\mathbf{H}(\bar{\mathbf{U}}^{-1}, \mathbf{\Delta}_0)\}^{-1} \mathbf{H}(\bar{\mathbf{U}}^{-1} \mathbf{Y} \bar{\mathbf{U}}^{-1}, \mathbf{\Delta}_0) \{\mathbf{H}(\bar{\mathbf{U}}^{-1}, \mathbf{\Delta}_0)\}^{-1}, \tag{2.12b}$$

where $\mathbf{H}(\cdot,\cdot)$ is a $q \times q$ matrix quadratic form defined by

$$\mathbf{H}(\mathbf{U}^{-1}, \mathbf{\Delta}) = \mathbf{\Delta}'\mathbf{U}^{-1}\mathbf{\Delta}. \tag{2.13}$$

Proof

Let $\mathbf{g}(\gamma\,|\,\mathbf{S}, \mathbf{U})$ be the $q \times 1$ negative gradient of $\frac{1}{2}F(\mathbf{S}, \mathbf{\Sigma}(\gamma)\,|\,\mathbf{U})$:

$$\mathbf{g}(\gamma\,|\,\mathbf{S}, \mathbf{U}) = \mathbf{\Delta}'(\gamma)\,\mathbf{U}^{-1}\{\mathbf{s} - \boldsymbol{\sigma}(\gamma)\}. \tag{2.14}$$

Since $\hat{\gamma}$ minimizes $\frac{1}{2}F(\mathbf{S}, \mathbf{\Sigma}(\gamma)\,|\,\mathbf{U})$ it follows from (R3), (2.5) and Proposition 1 that (with probability arbitrarily close to 1 for sufficiently large n)

$$\mathbf{g}(\hat{\gamma}\,|\,\mathbf{S}, \mathbf{U}) = \hat{\mathbf{\Delta}}'\mathbf{U}^{-1}[\mathbf{s} - \boldsymbol{\sigma}_0 + \{\boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}}_0\} - \{\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_0\}] = 0, \tag{2.15}$$

where $\boldsymbol{\sigma}_0 = \operatorname{vecs}(\mathbf{\Sigma}_0)$, $\hat{\boldsymbol{\sigma}}_0 = \operatorname{vecs}(\mathbf{\Sigma}(\gamma_0))$, $\hat{\boldsymbol{\sigma}} = \operatorname{vecs}(\mathbf{\Sigma}(\hat{\gamma}))$ and $\hat{\mathbf{\Delta}} = \mathbf{\Delta}(\hat{\gamma})$.

Use of (R3), (R7) and the Mean-Value Theorem shows that

$$\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_0 = \tilde{\mathbf{\Delta}}\{\hat{\gamma} - \gamma_0\}, \tag{2.16}$$

where $\tilde{\mathbf{\Delta}}$ is a $p^* \times q$ matrix of derivatives of $\boldsymbol{\sigma}(\gamma)$ with respect to γ evaluated at points on the line segment joining $\hat{\gamma}$ and γ_0. Substitution of (2.16) into (2.15) and use of Proposition 1, (R7), (R4) and (R5) shows that

$$n^{\frac{1}{2}}\{\hat{\gamma} - \gamma_0\} \stackrel{\mathrm{a}}{=} n^{\frac{1}{2}}\{\mathbf{\Delta}_0' \bar{\mathbf{U}}^{-1} \mathbf{\Delta}_0\}^{-1} \mathbf{\Delta}_0' \bar{\mathbf{U}}^{-1} \boldsymbol{\delta}_s + n^{\frac{1}{2}}\, \mathbf{g}(\gamma_0\,|\,\mathbf{\Sigma}_0, \bar{\mathbf{U}}),$$

where $\stackrel{\mathrm{a}}{=}$ stands for 'is asymptotically equivalent to' and implies that the difference between the left- and right-hand sides converges in probability to zero as $n \to \infty$. But $\mathbf{g}(\gamma_0\,|\,\mathbf{\Sigma}_0, \bar{\mathbf{U}}) = 0$ from (R3) and (2.6). The proposition then follows from the asymptotic multivariate normal distribution of $\boldsymbol{\delta}_s$. ■

Proposition 2 implies that the multivariate normal distribution with mean γ_0 and covariance matrix $n^{-1}\,\mathrm{L}\operatorname{cov}(\hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\gamma')$ will provide an approximation to the distribution of $\hat{\gamma}$.

If the matrix $\mathbf{U}$ is correctly specified (i.e. $\bar{\mathbf{U}} = \dot{\mathbf{Y}}$) then (2.12) simplifies considerably.

Corollary 2.1 (cf. Browne, 1982 (1.6.7)). If the matrix $\mathbf{U}$ of the discrepancy function $F(\mathbf{S}, \mathbf{\Sigma}(\gamma)\,|\,\mathbf{U})$ is correctly specified then the covariance matrix of the limiting multivariate normal distribution of $\hat{\boldsymbol{\delta}}_\gamma$ is given by:

$$\mathrm{L\,cov}\,(\hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\gamma') = \{\boldsymbol{\Delta}_0' \bar{\boldsymbol{\Upsilon}}^{-1} \boldsymbol{\Delta}_0\}^{-1} = \{\mathbf{H}(\bar{\boldsymbol{\Upsilon}}^{-1}, \boldsymbol{\Delta}_0)\}^{-1}. \qquad \blacksquare \qquad (2.17)$$

Not only does correct specification of $\mathbf{U}$ result in a simpler expression for $\mathrm{L\,cov}\,(\hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\gamma')$ but it yields estimators with smaller asymptotic variances.

Proposition 3. If $\bar{\mathbf{U}} = \bar{\boldsymbol{\Upsilon}}$ then $\hat{\boldsymbol{\gamma}}$ is asymptotically efficient within the restricted class of estimators minimizing discrepancy functions of the form of $F(\mathbf{S}, \boldsymbol{\Sigma} \mid \mathbf{U})$ in (2.7) in the sense that the asymptotic variances of all its elements attain lower bounds for asymptotic variances of estimators in this class (see Browne, 1974, Proposition 3). $\blacksquare$

These GLS estimators with minimum asymptotic variances will be referred to as 'best' generalized least squares (BGLS) estimators. The asymptotic covariance matrix is given by (2.17). Note that when the distribution of $\mathbf{x}$ is not normal the BGLS estimators will *not necessarily* be asymptotically efficient within the class of all possible estimators of $\boldsymbol{\gamma}_0$ in that (2.17) will not necessarily represent the Cramèr–Rao minimum variance bound (e.g. Silvey, 1970, Section 2.12).

We note that the term 'best' is used in a very restricted sense with respect to a specific asymptotic property which possibly may not carry over to finite samples. It is possible that other estimators may have other properties which render them superior to BGLS estimators for practical applications involving samples of moderate size.

If the distribution of $\mathbf{x}$ has no kurtosis then all matrices $\mathbf{U}$ whose elements satisfy (2.8) with $\bar{\mathbf{V}} = \boldsymbol{\Sigma}_0$ will yield BGLS estimators. The discrepancy function may then be expressed in the form of (2.9) and typical elements of $\mathbf{H}(\mathbf{U}^{-1}, \boldsymbol{\Delta})$ in (2.13) and $\mathbf{g}(\boldsymbol{\gamma} \mid \mathbf{S}, \mathbf{U})$ in (2.14) may be expressed in the more computationally efficient forms:

$$[\mathbf{H}(\mathbf{U}^{-1}, \boldsymbol{\Delta})]_{ij} = \tfrac{1}{2}\,\mathrm{tr}\left[\mathbf{V}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}\mathbf{V}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_j}\right], \qquad (2.18)$$

$$[\mathbf{g}(\boldsymbol{\gamma} \mid \mathbf{S}, \mathbf{U})]_i = \tfrac{1}{2}\,\mathrm{tr}\left[\mathbf{V}^{-1}\{\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\gamma})\}\mathbf{V}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}\right]. \qquad (2.19)$$

If the distribution of $\mathbf{x}$ has no kurtosis, the class of BGLS estimators includes a GLS estimator with $\mathbf{V} = \mathbf{S}$ and the MWL estimator where $\mathbf{V} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_W)$. Note, however, that if the assumption that $\mathbf{x}$ has no kurtosis is made incorrectly then the $\mathbf{U}$ of (2.8) is misspecified and the resulting estimators are no longer BGLS estimators.

We shall now consider a test statistic which can be used in conjunction with a class of estimators which includes, but is not restricted to, the class of GLS estimators.

Proposition 4 (cf. Browne, 1982 (1.7.14)). Suppose that $\hat{\boldsymbol{\gamma}}$ is an estimator with the property that, as $n \to \infty$, $\hat{\boldsymbol{\delta}}_\gamma = n^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotic distribution with a null mean vector and a finite covariance matrix. (This property applies to any consistent estimator of $\boldsymbol{\gamma}_0$ which is a differentiable function of $\mathbf{S}$ at $\mathbf{S} = \boldsymbol{\Sigma}_0$, and, under present regularity conditions, to MWL and all GLS estimators (cf. Proposition 2).)

Let $\boldsymbol{\Delta}_c = \boldsymbol{\Delta}_c(\boldsymbol{\gamma})$ be a $p^* \times (p^* - q)$ matrix valued function of $\boldsymbol{\gamma}$, such that the rank of $\boldsymbol{\Delta}_c$ is $(p^* - q)$ and $\boldsymbol{\Delta}_c' \boldsymbol{\Delta} = 0$ when $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\gamma})$ in (2.4). Also let

$$\hat{\mathbf{e}} = \mathbf{s} - \hat{\boldsymbol{\sigma}} = \mathrm{vecs}\,(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}})), \quad \mathbf{e}_0 = \boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}}_0 = \mathrm{vecs}\,(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)),$$

$$\hat{\boldsymbol{\Delta}}_c = \boldsymbol{\Delta}_c(\hat{\boldsymbol{\gamma}}) \quad \text{and} \quad \boldsymbol{\Delta}_{0c} = \boldsymbol{\Delta}_c(\boldsymbol{\gamma}_0).$$

If the positive definite matrix $\hat{\mathbf{\Upsilon}}$ is a consistent estimator of $\bar{\mathbf{\Upsilon}}$, then the asymptotic distribution of the quadratic form statistic

$$c(\hat{\gamma}, \mathbf{S}, \hat{\mathbf{\Upsilon}}) = n\hat{\mathbf{e}}'\hat{\mathbf{\Delta}}_c\{\hat{\mathbf{\Delta}}_c'\hat{\mathbf{\Upsilon}}\hat{\mathbf{\Delta}}_c\}^{-1}\hat{\mathbf{\Delta}}_c'\hat{\mathbf{e}} \tag{2.20a}$$

as n tends to infinity is non-central chi-square with p^*-q degrees of freedom and non-centrality parameter

$$\lambda = n\mathbf{e}_0'\mathbf{\Delta}_{0c}\{\mathbf{\Delta}_{0c}'\bar{\mathbf{\Upsilon}}\mathbf{\Delta}_{0c}\}^{-1}\mathbf{\Delta}_{0c}'\mathbf{e}_0. \tag{2.21a}$$

Equivalent expressions for the quadratic form statistic and non-centrality parameter are

$$\begin{aligned} c(\hat{\gamma}, \mathbf{S}, \hat{\mathbf{\Upsilon}}) &= n\hat{\mathbf{e}}'[\hat{\mathbf{\Upsilon}}^{-1} - \hat{\mathbf{\Upsilon}}^{-1}\hat{\mathbf{\Delta}}\{\hat{\mathbf{\Delta}}'\hat{\mathbf{\Upsilon}}^{-1}\hat{\mathbf{\Delta}}\}^{-1}\hat{\mathbf{\Delta}}'\hat{\mathbf{\Upsilon}}^{-1}]\hat{\mathbf{e}} \\ &= n\{\hat{\mathbf{e}}'\hat{\mathbf{\Upsilon}}^{-1}\hat{\mathbf{e}} - \hat{\mathbf{g}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{g}}\}, \end{aligned} \tag{2.20b}$$

where $\hat{\mathbf{g}} = \mathbf{g}(\hat{\gamma}\,|\,\mathbf{S}, \hat{\mathbf{\Upsilon}})$ in (2.14), $\hat{\mathbf{H}} = \mathbf{H}(\hat{\mathbf{\Upsilon}}^{-1}, \hat{\mathbf{\Delta}})$ in (2.13), and

$$\begin{aligned} \lambda &= n\mathbf{e}_0'[\bar{\mathbf{\Upsilon}}^{-1} - \bar{\mathbf{\Upsilon}}^{-1}\mathbf{\Delta}_0\{\mathbf{\Delta}_0'\bar{\mathbf{\Upsilon}}^{-1}\mathbf{\Delta}_0\}^{-1}\mathbf{\Delta}_0'\bar{\mathbf{\Upsilon}}^{-1}]\mathbf{e}_0 \\ &= n\{\mathbf{e}_0'\bar{\mathbf{\Upsilon}}^{-1}\mathbf{e}_0 - \mathbf{g}_0'\mathbf{H}_0^{-1}\mathbf{g}_0\}, \end{aligned} \tag{2.21b}$$

where $\mathbf{g}_0 = \mathbf{g}(\gamma_0\,|\,\mathbf{\Sigma}_0, \bar{\mathbf{\Upsilon}})$ in (2.14), $\mathbf{H}_0 = \mathbf{H}(\bar{\mathbf{\Upsilon}}^{-1}, \mathbf{\Delta}_0)$ in (2.13).

Proof

Use of (2.16) shows that

$$\begin{aligned} n^{\frac{1}{2}}\hat{\mathbf{\Delta}}_c'\hat{\mathbf{e}} &= n^{\frac{1}{2}}\hat{\mathbf{\Delta}}_c'[\mathbf{s} - \boldsymbol{\sigma}_0 + \{\boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}}_0\} - \{\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_0\}] \\ &= \hat{\mathbf{\Delta}}_c'\{\boldsymbol{\delta}_s + n^{\frac{1}{2}}\mathbf{e}_0\} - \hat{\mathbf{\Delta}}_c'\tilde{\mathbf{\Delta}}\hat{\boldsymbol{\delta}}_\gamma. \end{aligned}$$

Since $\hat{\boldsymbol{\delta}}_\gamma$ remains bounded in probability as $n \to \infty$, use of (R7) shows that

$$\hat{\mathbf{\Delta}}_c'\tilde{\mathbf{\Delta}}\hat{\boldsymbol{\delta}}_\gamma \overset{\text{a}}{=} \mathbf{\Delta}_{0c}'\mathbf{\Delta}_0\hat{\boldsymbol{\delta}}_\gamma = \mathbf{0}.$$

Since $n^{\frac{1}{2}}\mathbf{e}_0$ remains bounded as $n \to \infty$ because of (R5),

$$n^{\frac{1}{2}}\hat{\mathbf{\Delta}}_c'\hat{\mathbf{e}} \overset{\text{a}}{=} \mathbf{\Delta}_{0c}'\{\boldsymbol{\delta}_s + n^{\frac{1}{2}}\mathbf{e}_0\} = \mathbf{z}, \quad \text{say},$$

has an asymptotic multivariate normal distribution with mean vector $n^{\frac{1}{2}}\mathbf{\Delta}_{0c}'\mathbf{e}_0$ and $(p^*-q)\times(p^*-q)$ covariance matrix $\mathbf{\Delta}_{0c}'\bar{\mathbf{\Upsilon}}\mathbf{\Delta}_{0c}$ which is non-singular because of (R1). Therefore

$$c(\hat{\gamma}, \mathbf{S}, \hat{\mathbf{\Upsilon}}) \overset{\text{a}}{=} \mathbf{z}'\{\mathbf{\Delta}_{0c}'\bar{\mathbf{\Upsilon}}\mathbf{\Delta}_{0c}\}^{-1}\mathbf{z}$$

has an asymptotic non-central chi-square distribution with N^*-q degrees of freedom and non-centrality parameter λ in (2.21a). The equivalence of (2.20b) to (2.20a) and (2.21b) to (2.21a) follows from Khatri (1966, Lemma 1).

■

An algorithm which may be employed to compute an orthogonal completion $\hat{\mathbf{\Delta}}_c$ for (2.20a) is described in Section 2.2.5.3 of Gill *et al.* (1981). Alternatively, computation of $\hat{\mathbf{\Delta}}_c$ may be avoided entirely if (2.20b) is used. If the assumption that the distribution of $\mathbf{x}$ has no kurtosis is valid, then $\hat{\mathbf{g}}$ and $\hat{\mathbf{H}}$ in (2.20b) may be obtained using versions of (2.18) and (2.19).

A considerable simplification of the statistic $c(\hat{\gamma}, S, \hat{\Upsilon})$ is possible if $\hat{\gamma}$ is a GLS estimator with a correctly specified matrix $\mathbf{U}$ (i.e. $\bar{\mathbf{U}} = \bar{\Upsilon}$).

Corollary 4.1. If $\mathbf{U}$ is correctly specified the asymptotic distribution of $nF(S, \Sigma(\hat{\gamma}) \mid \mathbf{U})$ is non-central chi-square with $p^* - q$ degrees of freedom and non-centrality parameter $nF(\Sigma_0, \Sigma(\gamma_0) \mid \bar{\Upsilon})$.

Proof

If $\mathbf{U}$ is correctly specified then $\hat{\Upsilon} = \mathbf{U}$ is permissible and (2.20*b*) shows that $c(\hat{\gamma}, S, \mathbf{U}) = nF(S, \Sigma(\hat{\gamma}) \mid \mathbf{U})$ since $\hat{\mathbf{g}} = \mathbf{g}(\hat{\gamma} \mid S, \hat{\Upsilon}) = \mathbf{g}(\hat{\gamma} \mid S, \mathbf{U}) = \mathbf{0}$. Similarly it follows from (2.21*b*) that $\lambda = nF(\Sigma_0, \Sigma(\gamma_0) \mid \bar{\mathbf{U}})$ since $\mathbf{g}_0 = \mathbf{g}(\gamma_0 \mid \Sigma_0, \bar{\mathbf{U}}) = \mathbf{0}$. ■

The second term in (2.20*b*) may be regarded as a correction term for a non-efficient estimator. Consider a situation where the model holds (i.e. $\Sigma_0 = \Sigma(\gamma_0)$). Let $\gamma = \hat{\gamma}_a$ minimize the discrepancy function $F(S, \Sigma(\gamma) \mid \hat{\Upsilon})$ so that $\hat{\gamma}_a$ is a BGLS estimator of γ_0. Suppose that $\hat{\gamma}_b$ is any other estimator which satisfies the condition of Proposition 4. By the definition of $\hat{\gamma}_a$ we have that $F(S, \Sigma(\hat{\gamma}_b) \mid \hat{\Upsilon}) \geqslant F(S, \Sigma(\hat{\gamma}_a) \mid \hat{\Upsilon})$. Now the statistic in (2.20*b*) may be expressed as

$$c(\hat{\gamma}_b, S, \hat{\Upsilon}) = n\{F(S, \Sigma(\hat{\gamma}_b) \mid \hat{\Upsilon}) - d(\hat{\gamma}_b)\},$$

where the correction term $d(\hat{\gamma}_b)$ for a non-efficient estimator is given by

$$d(\hat{\gamma}_b) = \hat{\mathbf{e}}_b' \hat{\Upsilon}^{-1} \hat{\Delta}_b \{\hat{\Delta}_b' \hat{\Upsilon}^{-1} \hat{\Delta}_b\}^{-1} \hat{\Delta}_b' \hat{\Upsilon}^{-1} \hat{\mathbf{e}}_b \geqslant 0$$

with $\hat{\mathbf{e}}_b = \mathbf{s} - \boldsymbol{\sigma}(\hat{\gamma}_b)$, $\hat{\Delta}_b = \Delta(\hat{\gamma}_b)$. If the model is linear then $\boldsymbol{\sigma}(\gamma) = \boldsymbol{\kappa} + \Delta_0 \gamma$ where $\boldsymbol{\kappa}$ and $\Delta_0 = \Delta(\gamma_0) = \hat{\Delta}_b$ are constant and it is easily shown that $c(\hat{\gamma}_b, S, \hat{\Upsilon}) = nF(S, \Sigma(\hat{\gamma}_a) \mid \hat{\Upsilon})$. Thus $c(\hat{\gamma}_b, S, \hat{\Upsilon})$ may be regarded as an approximation for $nF(S, \Sigma(\hat{\gamma}_a) \mid \hat{\Upsilon})$ based on an approximation of the structural model $\boldsymbol{\sigma}(\gamma)$ by a linear model in the vicinity of $\hat{\gamma}_a$ and $\hat{\gamma}_b$. This approximation improves as n tends to infinity since $\hat{\gamma}_b$ and $\hat{\gamma}_a$ converge in probability to the same point γ_0.

The statistics given in Proposition 4 and Corollary 4.1 will have asymptotic (central) chi-square distributions when $\Sigma_0 = \Sigma(\gamma_0)$ so that $\mathbf{e}_0 = \mathbf{0}$ and $\lambda = 0$. Consequently they may be employed to test the null hypothesis that the model holds. The non-central distribution is also of interest, since a model is often an approximation for reality and measures of goodness of approximation are required.

Bentler (1983, pp. 22–23) and De Leeuw (1983, pp. 118–119) have recently adapted improvement methods to the analysis of covariance and correlation structures. A second asymptotically more efficient estimator is derived from an initial estimator satisfying the conditions of Proposition 4 using a single Gauss–Newton iteration step. This approach could be used to provide an entirely different derivation of (2.20*b*) (or Browne, 1982 (1.7.14)). The test statistic employed by Bentler (1983, equations (3) and (6)), however, differs slightly from (2.20*b*) and requires a little additional computation.

Most available results in the analysis of covariance structures have been obtained under the assumption of a multivariate normal distribution for $\mathbf{x}$. Formulae for the asymptotic estimator covariance matrix are based on an inverse information matrix equivalent to (2.17) with $\bar{\Upsilon}$ defined by (2.3). The equivalent form in (2.18) is often employed. Test statistics are either of the form $nF(S, \Sigma(\hat{\gamma}) \mid \mathbf{V})$ where $\bar{\mathbf{V}} = \Sigma_0$ or of the form $nF_W(S, \Sigma(\hat{\gamma}_W))$ which is asymptotically equivalent (cf. Browne, 1974, Proposition 7). Application of Corollary 4.1 then requires that $\bar{\Upsilon}$ be as in (2.3). Consequently the usual asymptotic formulae for the estimator covariance matrix and the asymptotic chi-square distribution of the usual test statistics derived under normality assumptions are incorrect if the fourth-order cumulants of the distribution of $\mathbf{x}$ involved in (2.2) are not all zero since $\mathbf{U}$ is then misspecified.

Results presented here have been derived without the assumption of any equality constraints on γ. They may easily be adapted to situations where equality constraints are imposed using the reparameterization approach described in Lee & Bentler (1980, pp. 133–135). Adaptations of (2.12*b*) and (2.20*b*) for equality constraints are given in formulae (1.6.18) and (1.7.19) of Browne (1982).

3. Asymptotically distribution-free BGLS estimators

An estimator, $\hat{\gamma}$, which satisfies the conditions of Proposition 3 in general may be obtained by defining a $\mathbf{U}$ which is a consistent estimator of $\bar{\mathbf{Y}}$. This may be accomplished by substituting sample moments for the population moments in (2.2).

Let $\mathbf{x}_r = (x_{1r}, x_{2r}, \ldots, x_{ir}, \ldots, x_{pr})'$ represent the rth observation, $r = 1, \ldots, N$, on $\mathbf{x}$ and let

$$\bar{x}_i = N^{-1} \sum_{r=1}^{N} x_{ir}, \tag{3.1}$$

$$w_{ij} = N^{-1} \sum_{r=1}^{N} (x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j) = \frac{n}{N}[\mathbf{S}]_{ij}, \tag{3.2}$$

$$w_{ijkl} = N^{-1} \sum_{r=1}^{N} (x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j)(x_{kr}-\bar{x}_k)(x_{lr}-\bar{x}_l). \tag{3.3}$$

In order for all sample fourth-order moments about the mean, w_{ijkl}, to be consistent estimators of the σ_{ijkl}, an additional mild assumption about the distribution of $\mathbf{x}$ is required; namely, that all eighth-order moments be finite. This assumption also ensures that w_{ij} will be a consistent estimator of $[\mathbf{\Sigma}_0]_{ij}$ so that

$$[\mathbf{U}]_{ij,kl} = w_{ijkl} - w_{ij} w_{kl} \tag{3.4}$$

will be a consistent estimator of $[\bar{\mathbf{Y}}]_{ij,kl}$ in (2.2). It can be seen that (3.4) represents the sample covariance between the product variables $(x_{ri}-\bar{x}_i)(x_{rj}-\bar{x}_j)$ and $(x_{rk}-\bar{x}_k)(x_{rl}-\bar{x}_l)$ with means w_{ij} and w_{kl}. Consequently the matrix $\mathbf{U}$ defined by (3.4) will be positive definite with probability 1 provided that N is greater than p^* and $\bar{\mathbf{Y}}$ is positive definite (i.e. (R1) holds).

Because of force of habit, due to the general use of the unbiased estimator, $[\mathbf{S}]_{ij}$, for $[\mathbf{\Sigma}_0]_{ij}$ instead of the biased estimator, w_{ij}, the divisions by N instead of by $(N-1)$ implicit in (3.4) may cause some unease. An unbiased estimator of the finite sample covariance matrix $\mathbf{Y}$ in (2.1) may be obtained from the k-statistics (cf. Kendall & Stuart, 1969, chapter 13)

$$\begin{aligned} k_{ijkl} = N^2\{(N+1)w_{ijkl} - (N-1)(w_{ij}w_{kl} + w_{ik}w_{jl} + w_{il}w_{jk})\}/ \\ \{(N-1)(N-2)(N-3)\}, \end{aligned} \tag{3.5}$$

$$k_{ij} = N w_{ij}/(N-1) = [\mathbf{S}]_{ij}. \tag{3.6}$$

Use of the expected values

$$\mathscr{E}(k_{ijkl}) = \kappa_{ijkl},$$

$$\mathscr{E}(k_{ij}k_{kl}) = \kappa_{ij}\kappa_{kl} + (N-1)^{-1}(\kappa_{ik}\kappa_{jl} + \kappa_{il}\kappa_{jk}) + N^{-1}\kappa_{ijkl},$$

with $\kappa_{ij} = [\Sigma_0]_{ij}$, will show that the expected value of

$$[\mathbf{U}]_{ij,kl} = (N-1)^2\{N(N+1)\}^{-1} k_{ijkl} + (N-1)^2\{(N-2)(N+1)\}^{-1}\{k_{ik}k_{jl}+k_{il}k_{jk}\} \\ -2(N-1)\{(N-2)(N+1)\}^{-1} k_{ij}k_{kl} \qquad (3.7)$$

is equal to $[\mathbf{Y}]_{ij,kl}$ in (2.1). Substitution of (3.5) and (3.6) into (3.7) yields the following expression,

$$[\mathbf{U}]_{ij,kl} = N(N-1)\{(N-2)(N-3)\}^{-1}\{w_{ijkl}-w_{ij}w_{kl}\} \\ -N\{(N-2)(N-3)\}^{-1}\{w_{ik}w_{jl}+w_{il}w_{jk}-2(N-1)^{-1}w_{ij}w_{kl}\} \qquad (3.8)$$

for a typical element of an unbiased estimator, $\mathbf{U}$, of $\mathbf{Y}$. While this matrix could conceivably not be positive definite, it is unlikely that this would be the case in any reasonable application with N substantially greater than p^* since the second term in (3.8) is of order N^{-1}.

Clearly the matrices with typical elements (3.4) and (3.8) are asymptotically equivalent and both satisfy the requirement of Proposition 3 that $\bar{\mathbf{U}} = \bar{\mathbf{Y}}$.

If $\mathbf{U}$ is to be computed either using (3.4) or (3.8) the w_{ijkl} and w_{ij} are required. Du Toit (1979, equation (7.5.21)) has pointed out that the number of non-duplicated dth-order moments involving p variates is given by $\binom{p+d-1}{d}$. Consequently the number of non-duplicated w_{ijkl} to be computed for $\mathbf{U}$ is $p(p+1)(p+2)(p+3)/24$ which will be considerably less than $p^*(p^*+1)/2$, the number of non-duplicated elements of $\mathbf{U}$. The ratio is 0·46 for $p = 10$, 0·40 for $p = 20$ and tends to 1/3 from above as p increases without bound.

The BGLS estimate, $\hat{\gamma}$, is obtained by minimizing

$$F(\mathbf{S}, \Sigma(\gamma) \mid \mathbf{U}) = (\mathbf{s}-\sigma(\gamma))'\mathbf{U}^{-1}(\mathbf{s}-\sigma(\gamma)) \qquad (3.9)$$

with respect to γ, where $\mathbf{U}$ is defined by either (3.4) or (3.8).

In the author's computer programs the Jennrich & Sampson (1968) modification of the Gauss–Newton method was employed. If $\hat{\gamma}_t$ represents the tth successive approximation for the estimate, the new approximation is defined by

$$\hat{\gamma}_{t+1} = \hat{\gamma}_t + \alpha_t \mathbf{H}_t^{-1}\mathbf{g}_t, \qquad (3.10)$$

where

$$\mathbf{H}_t = \mathbf{H}(\mathbf{U}^{-1}, \Delta(\hat{\gamma}_t)), \quad \text{cf. (2.13)},$$

$$\mathbf{g}_t = \mathbf{g}(\hat{\gamma}_t \mid \mathbf{S}, \mathbf{U}), \quad \text{cf. (2.14)},$$

and $0 < \alpha_t \leqslant 1$ is a stepsize parameter chosen so as to ensure that $F(\mathbf{S}, \Sigma(\hat{\gamma}_{t+1}) \mid \mathbf{U}) < F(\mathbf{S}, \Sigma(\hat{\gamma}_t) \mid \mathbf{U})$ with $\alpha_t = 1$ in the majority of cases. The inversion of the large matrix $\mathbf{U}$ was avoided. Initially the lower triangular (Choleski) square root $\mathbf{U}^{\frac{1}{2}}$ of $\mathbf{U}$ was computed (i.e. $\mathbf{U} = \mathbf{U}^{\frac{1}{2}}\mathbf{U}^{\frac{1}{2}\prime}$). On each iteration the equations,

$$\mathbf{U}^{\frac{1}{2}}\Delta_t^* = \Delta(\hat{\gamma}_t),$$

$$\mathbf{U}^{\frac{1}{2}}\mathbf{e}_t^* = (\mathbf{s}-\sigma(\hat{\gamma}_t))$$

were solved for the $p^* \times q$ matrix Δ_t^* and $\mathbf{p}^* \times 1$ vector $\mathbf{e}_t^*$ by forward substitution. Then $F(\mathbf{S}, \Sigma(\hat{\gamma}_t) \mid \mathbf{U})$, $\mathbf{H}_t$ and $\mathbf{g}_t$ were obtained from

$$F(\mathbf{S}, \Sigma(\hat{\gamma}_t) \mid \mathbf{U}) = \mathbf{e}_t^{*\prime}\mathbf{e}_t^*, \qquad (3.11)$$

$$\mathbf{H}_t = \Delta_t^{*\prime}\Delta_t^*, \qquad (3.12)$$

$$\mathbf{g}_t = \Delta_t^{*\prime}\mathbf{e}_t^* \qquad (3.13)$$

and the increment $\mathbf{H}_t^{-1}\mathbf{g}_t$ was obtained by means of the stepwise regression procedure described by Jennrich & Sampson (1968). A convergence criterion based on the largest absolute residual cosine (Dennis, 1977, p. 273; Browne, 1982, p. 105) was employed.

This algorithm was found to converge satisfactorily in the majority of instances. After convergence $n^{-1}\{\mathbf{H}(\hat{\gamma})\}^{-1}$ provides an estimate of the covariance matrix of $\hat{\gamma}$ (Corollary 2.1) and $nF(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\gamma}) \mid \mathbf{U})$ provides a test of the null hypothesis that the model holds (Corollary 4.1).

4. Corrections for kurtosis under the assumption of an elliptical distribution

The distribution-free BGLS estimators of Section 3 involve the algebraic manipulation and retention within the computer of the $p^* \times p^*$ symmetric matrix $\mathbf{U}$. If the number of variables is $p = 20$ then $p^* = 210$ and $\mathbf{U}$ has 22 155 non-duplicated elements. Consequently the computation of BGLS estimates will tend to become infeasible as p approaches 20.

The MWL estimates are more easily computed than the distribution-free BGLS estimates and computer programs for computing them are generally available. It will be shown here that if the structural model $\boldsymbol{\Sigma}(\gamma)$ has a certain property of scale invariance then the maximum Wishart likelihood estimators are BGLS estimators, not only if $\mathbf{x}$ has a multivariate normal distribution but also if the distribution of $\mathbf{x}$ belongs to the larger class of elliptical distributions. Corrections for kurtosis for the test statistic and estimator covariance matrix will be provided.

A structural *model*, $\boldsymbol{\Sigma}(\gamma)$, for a covariance matrix will be said to be *invariant under a constant scaling factor* (ICSF) if given any $\gamma \in G$ and any positive scalar, α^2, there exists a $\gamma^* \in G$ such that $\boldsymbol{\Sigma}(\gamma^*) = \alpha^2 \boldsymbol{\Sigma}(\gamma)$. Nearly all models for covariance matrices in current use are ICSF. Amongst the few exceptions are models which require certain elements of $\boldsymbol{\Sigma}$ to have fixed *non-zero* values and restricted factor analysis models for $\boldsymbol{\Sigma}$ with some fixed *non-zero* factor loadings and fixed factor variances. These are seldom employed in practice because of the difficulty of prespecifying non-zero parameter values.

Detailed information on the class of elliptical distribution is contained in Devlin *et al.* (1976) and Muirhead (1982, Sections 1.5 and 1.6). We shall be concerned only with the subclass of elliptical distributions with a finite non-singular covariance matrix, $\boldsymbol{\Sigma}_0$. If $\boldsymbol{\xi}$ is the mean vector, the density function of $\mathbf{x}$ then is of the form

$$f(\mathbf{x}) = c_p |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g(\{\mathbf{x}-\boldsymbol{\xi}\}' \boldsymbol{\Sigma}_0^{-1} \{\mathbf{x}-\boldsymbol{\xi}\}),$$

where c_p is a normalizing constant dependent on p and $g(\cdot)$ is a non-negative function. Contours of constant probability then are ellipsoids.

We shall show that if $\mathbf{x}$ has some elliptical distribution then $\boldsymbol{\Upsilon}$ has the convenient property of depending only on $\boldsymbol{\Sigma}_0$ and a single kurtosis parameter. The coefficient of relative kurtosis, η_p, of a p-variate distribution, will be defined to be Mardia's (1970, 1974) coefficient of multivariate kurtosis divided by the corresponding coefficient of multivariate kurtosis for a multivariate normal distribution:

$$\eta_p = \mathscr{E}(\{\mathbf{x}-\boldsymbol{\xi}\}' \boldsymbol{\Sigma}_0^{-1} \{\mathbf{x}-\boldsymbol{\xi}\})^2 / \{p(p+2)\}. \tag{4.1}$$

If $\mathbf{x}_k$ is any $k \times 1$ subvector of $\mathbf{x}$ and the distribution of $\mathbf{x}$ belongs to the elliptical class, 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, $\eta_1 = \eta_p$. The common relative kurtosis of an elliptical distribution will be denoted by η.

If $\mathbf{x}$ has some elliptical distribution then (cf. Muirhead, 1982, p. 41) all fourth-order cumulants satisfy the equations

$$\kappa_{ijkl} = (\eta - 1)\{[\Sigma_0]_{ij}[\Sigma_0]_{kl} + [\Sigma_0]_{ik}[\Sigma_0]_{jl} + [\Sigma_0]_{il}[\Sigma_0]_{jk}\}.$$

Substitution into (2.1) shows that the covariance matrix, $\mathbf{Y}$, of the unknown finite sample distribution of $\boldsymbol{\delta}_s = n^{\frac{1}{2}}(\mathbf{s} - \boldsymbol{\sigma}_0)$ has the typical element

$$[\mathbf{Y}]_{ij,kl} = \{\eta - N^{-1}(\eta - 1)\}\{[\Sigma_0]_{ik}[\Sigma_0]_{jl} + [\Sigma_0]_{il}[\Sigma_0]_{jk}\} + N^{-1}(N-1)(\eta - 1)[\Sigma_0]_{ij}[\Sigma_0]_{kl}. \tag{4.2}$$

(The term $-N^{-1}(\eta - 1)$ was omitted in error from formula (1.5.26) of Browne (1982). No other results therein were affected since the omission had no effect on the limiting covariance matrix, $\bar{\mathbf{Y}}$.)

The multivariate normal distribution is an elliptical distribution with a relative kurtosis of unity. If η is set equal to 1, (4.2) reduces to (2.3).

Letting N tend to infinity in (4.2) shows that the covariance matrix, $\bar{\mathbf{Y}}$, of the multivariate normal asymptotic distribution of $\boldsymbol{\delta}_s$ has a typical element

$$[\bar{\mathbf{Y}}]_{ij,kl} = \eta\{[\Sigma_0]_{ik}[\Sigma_0]_{jl} + [\Sigma_0]_{il}[\Sigma_0]_{jk}\} + (\eta - 1)\{[\Sigma_0]_{ij}[\Sigma_0]_{kl}\} \tag{4.3}$$

which depends only on the common relative kurtosis coefficient, η, of $\mathbf{x}$ and on elements of Σ.

A consistent estimator of η may be obtained by rescaling Mardia's (1970, 1974) sample measure of multivariate kurtosis to yield

$$\hat{\eta} = \hat{\eta}_p = N^{-1} \sum_{r=1}^{N} \{(\mathbf{x}_r - \bar{\mathbf{x}})'\mathbf{W}^{-1}(\mathbf{x}_r - \bar{\mathbf{x}})\}^2/\{p(p+2)\} \tag{4.4}$$

with typical elements of $\bar{\mathbf{x}}$ and $\mathbf{W}$ given in (3.1) and (3.2). An appropriate algorithm is described by Mardia & Zemroch (1975).

If $\mathbf{V}$ is a consistent estimator of Σ_0, then the matrix $\mathbf{U}$ with typical element

$$[\mathbf{U}]_{ij,kl} = \hat{\eta}\{[\mathbf{V}]_{ik}[\mathbf{V}]_{jl} + [\mathbf{V}]_{il}[\mathbf{V}]_{jk}\} + (\hat{\eta} - 1)\{[\mathbf{V}]_{ij}[\mathbf{V}]_{kl}\}, \tag{4.5}$$

will be a consistent estimator of $\mathbf{Y}$ and therefore yield a BGLS estimator, $\hat{\boldsymbol{\gamma}}$, whenever $\mathbf{x}$ has a distribution which belongs to the elliptical class. In order to simplify the quadratic form $F(\mathbf{S}, \Sigma(\boldsymbol{\gamma}) \mid \mathbf{U})$ it will be convenient to express (4.5) in matrix notation.

Let $\mathbf{v}^+ = \mathrm{vec}(\mathbf{V})$ represent a $p^2 \times 1$ vector formed by stacking the columns of the symmetric matrix $\mathbf{V}$, so that $\mathbf{v}^+$ has some duplicated elements. We shall employ a $p^2 \times p^*$ transition matrix, $\mathbf{K}_p$ (cf. Browne, 1974, Section 2; Nel, 1980, Section 6.1) with typical element

$$[\mathbf{K}_p]_{ij,gh} = \tfrac{1}{2}(\delta_{ig}\delta_{jh} + \delta_{ih}\delta_{jg}), \quad i \leqslant p, \quad j \leqslant p, \quad g \leqslant h \leqslant p,$$

where δ_{ig} is the Kronecker delta.

Then $\mathbf{U}$ in (4.5) may be expressed as

$$\mathbf{U} = 2\hat{\eta}\mathbf{K}_p'(\mathbf{V} \otimes \mathbf{V})\mathbf{K}_p + \mathbf{K}_p'\mathbf{v}^+(\hat{\eta} - 1)\mathbf{v}^{+\prime}\mathbf{K}_p, \tag{4.6}$$

where $\mathbf{V} \otimes \mathbf{V}$ represents a Kronecker product. Use of the binomial inverse theorem (e.g. Press, 1972, p. 23) and properties of the transition matrix, $\mathbf{K}_p$ (Browne, 1974, Section 2, formulae (5), (13), (16)) shows that

$$\mathbf{U}^{-1} = \tfrac{1}{2}\hat{\eta}^{-1}\mathbf{K}_p^{-}\left\{(\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) - \mathrm{vec}(\mathbf{V}^{-1})\, b\, \mathrm{vec}'(\mathbf{V}^{-1})\right\}\mathbf{K}_p^{-\prime}, \tag{4.7}$$

where

$$b = b(\hat{\eta}, p) = (\hat{\eta}-1)/\{(p+2)\hat{\eta}-p\} \tag{4.8}$$

and $\mathbf{K}_p^- = (\mathbf{K}_p'\mathbf{K}_p)^{-1}\mathbf{K}_p'$. It then follows (cf. Browne, 1974, formulae (8), (16)) that

$$\begin{aligned} F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \mathbf{U}) &= \{\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\gamma})\}'\mathbf{U}^{-1}\{\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\gamma})\} \\ &= \tfrac{1}{2}\hat{\eta}^{-1}\left\{\mathrm{tr}\,[\{\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\gamma})\}\,\mathbf{V}^{-1}]^2 - b\,\mathrm{tr}^2\,[\{\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\gamma})\}\,\mathbf{V}^{-1}]\right\} \\ &= F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \mathbf{V}), \quad \text{say.} \end{aligned} \tag{4.9}$$

Note that $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \mathbf{V})$ reduces to the normal theory quadratic form discrepancy function $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \mathbf{V})$ in (2.9) when $\hat{\eta} = 1$.

Minimization of $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \mathbf{V})$ will yield a BGLS estimator under the assumption that the distribution of $\mathbf{x}$ belongs to the elliptical class, provided that $\mathbf{V}$ is a consistent estimator of $\boldsymbol{\Sigma}_0$.

A necessary condition for a minimum is that the derivatives

$$\begin{aligned} \frac{\partial}{\partial \gamma_i} F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \mathbf{V}) = -\hat{\eta}^{-1}\Bigg\{&\mathrm{tr}\left[\mathbf{V}^{-1}\{\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\gamma})\}\,\mathbf{V}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}\right] \\ &- b\,\mathrm{tr}\,[\{\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\gamma})\}\,\mathbf{V}^{-1}]\,\mathrm{tr}\left[\mathbf{V}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}\right]\Bigg\}, \\ i = 1, \ldots, q \end{aligned} \tag{4.10}$$

be zero at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$.

Suppose that the model holds. It follows from Proposition 1 that $\hat{\boldsymbol{\gamma}}_{\mathrm{W}}$ which minimizes $F_{\mathrm{W}}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))$ converges in probability to $\boldsymbol{\gamma}_0$. Since $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ is continuous from (R7) it follows that $\hat{\boldsymbol{\Sigma}}_{\mathrm{W}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}})$ is a consistent estimator of $\boldsymbol{\Sigma}_0$, and therefore $\mathbf{V} = \hat{\boldsymbol{\Sigma}}_{\mathrm{W}}$ in (4.9) and (4.10) is a suitable choice for $\mathbf{V}$.

Since $\hat{\boldsymbol{\gamma}}_{\mathrm{W}}$ minimizes $F_{\mathrm{W}}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))$ we have

$$\frac{\partial}{\gamma_i} F(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}})) = -\,\mathrm{tr}\left[\hat{\boldsymbol{\Sigma}}_{\mathrm{W}}^{-1}\{\mathbf{S}-\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}})\}\,\hat{\boldsymbol{\Sigma}}_{\mathrm{W}}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}})\right] = 0 \tag{4.11}$$

for $i = 1, \ldots, q$. Also, whenever $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ is ICSF,

$$\mathrm{tr}\,[\{\mathbf{S}-\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}})\}\,\hat{\boldsymbol{\Sigma}}_{\mathrm{W}}^{-1}] = \mathrm{tr}\,[\mathbf{S}\hat{\boldsymbol{\Sigma}}_{\mathrm{W}}^{-1}] - p = 0 \tag{4.12}$$

from Browne (1974, Proposition 8).

Use of (4.11) and (4.12) shows that all partial derivatives in (4.10) are zero at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}_{\mathrm{W}}$ so that $\hat{\boldsymbol{\gamma}}_{\mathrm{W}}$ is a stationary point of $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\mathrm{W}})$. It can be shown that the Hessian of $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\mathrm{W}})$ evaluated at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}_{\mathrm{W}}$ converges in probability to a positive definite matrix. Consequently, while it is conceivable that $\hat{\boldsymbol{\gamma}}_{\mathrm{W}}$ could be a stationary point other than a minimum of $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}) \,|\, \hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\mathrm{W}})$ in a small sample, the probability of this happening tends to zero as n tends to infinity. Consequently the maximum likelihood estimator, $\hat{\boldsymbol{\gamma}}_{\mathrm{W}}$, has the asymptotic properties of a BGLS estimator, not only if $\mathbf{x}$ has a multivariate normal distribution, but also whenever the distribution of $\mathbf{x}$ belongs to the elliptical class.

With use of (4.12) it can be seen from (4.9) and (2.9) that at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}_{\mathrm{W}}$,

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}}) \,|\, \hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\mathrm{W}}) = \hat{\eta}^{-1} F(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}}) \,|\, \hat{\boldsymbol{\Sigma}}_{\mathrm{W}}). \tag{4.13}$$

Also (Browne, 1974, Proposition 7) the quadratic form statistic is asymptotically equivalent to the Wishart likelihood ratio statistic,

$$nF(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_W) \mid \hat{\boldsymbol{\Sigma}}_W) \overset{a}{=} nF_W(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_W)). \tag{4.14}$$

The following proposition consequently holds:

Proposition 5 (cf. Browne, 1982, p. 92, pp. 97–98). Suppose that the distribution of $\mathbf{x}$ belongs to the elliptical class with relative kurtosis, η, and $p \times p$ covariance matrix $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$. If the model $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ is invariant under a constant scaling factor and $\hat{\boldsymbol{\gamma}}_W$ is the maximum Wishart likelihood estimator of $\boldsymbol{\gamma}_0$, then the asymptotic distribution of $\hat{\boldsymbol{\delta}}_\gamma = n^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}_W - \boldsymbol{\gamma}_0)$ is multivariate normal with a null mean vector and covariance matrix

$$\mathrm{L\,cov}\,(\hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\gamma') = \{\mathbf{H}(\boldsymbol{\gamma}_0, \eta)\}^{-1}$$

in which a typical element of $\mathbf{H}(\boldsymbol{\gamma}_0, \eta)$ is given by

$$[\mathbf{H}(\boldsymbol{\gamma}_0, \eta)]_{ij} = \tfrac{1}{2}\eta^{-1}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_i}\boldsymbol{\Sigma}_0^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_j}\right] - b_0\,\mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_i}\right]\mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_j}\right]\right\}, \tag{4.15}$$

where $b_0 = b(\eta, p) = (\eta - 1)/\{(p+2)\eta - p\}$ and the symmetric matrices $\partial\boldsymbol{\Sigma}/\partial\gamma_i$ of partial derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ are evaluated at $\boldsymbol{\gamma}_0$.

Also, the two test statistics

$$n\hat{\eta}^{-1}F(\mathbf{S}, \hat{\boldsymbol{\Sigma}}_W \mid \mathbf{V} = \hat{\boldsymbol{\Sigma}}_W) = \tfrac{1}{2}n\hat{\eta}^{-1}\,\mathrm{tr}\,[(\mathbf{S} - \hat{\boldsymbol{\Sigma}}_W)\hat{\boldsymbol{\Sigma}}_W^{-1}]^2, \tag{4.16}$$

and

$$n\hat{\eta}^{-1}F_W(\mathbf{S}, \hat{\boldsymbol{\Sigma}}_W) = n\hat{\eta}^{-1}\{\ln|\hat{\boldsymbol{\Sigma}}_W| - \ln|\mathbf{S}|\} \tag{4.17}$$

with $\hat{\boldsymbol{\Sigma}}_W = \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_W)$, are asymptotically equivalent and have an asymptotic central chi-square distribution with $p^* - q$ degrees of freedom. ■

This proposition provides convenient 'corrections for kurtosis' which modify standard results derived under normality assumptions to make them appropriate for the larger class of elliptical distributions. The usual likelihood ratio test statistic provided by most computer programs, or the alternative quadratic form statistic, is divided by an estimate (4.4) of relative kurtosis in (4.17) and (4.16). The modification to the estimator covariance matrix is a little less simple but its nature is apparent from the fact that when $\eta = 1$, so that $b_0 = 0$, (4.15) reduces to the usual normal theory information matrix.

Some special cases of the rescaling in (4.17) are known. Muirhead (1982, Theorems 8.3.10 and 11.2.10) has shown that it applies to the normal theory likelihood ratio tests for sphericity ($\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$) and for uncorrelated sets of variables ($\boldsymbol{\Sigma}$ is a block diagonal matrix). These two covariance structures are ICSF. Muirhead & Waternaux (1980) have shown that this rescaling also applies to a well-known normal theory test for a specified number of non-zero canonical correlations. This test is also the normal theory likelihood ratio test (Browne, 1979) for the inter-battery factor analysis model which is ICSF. The multivariate relative kurtosis estimator (4.4) for the single kurtosis parameter of any elliptical multivariate distribution was not employed in these specific instances. Muirhead (1982, Theorem 8.4.12) also obtained the asymptotic distribution, under the assumption of an elliptical distribution, of the normal theory likelihood ratio test statistic for the hypothesis that Σ is a known matrix, which is not ICSF. In this case the simple correction of (4.17) was inadequate.

The asymptotic distribution of elements of a sample correlation matrix under the assumption of a distribution of $\mathbf{x}$ from the elliptical class is given in Steiger & Browne (1984).

It is not anticipated that members of the class of elliptical distributions will provide perfect representations for the distributions of data which occur in practice in the social sciences. The hope is, however, that a member of the wider class of elliptical distributions will provide a closer approximation in practical situations than does the multivariate normal distribution. Most models for covariance matrices in current use are ICSF, and the adaptation of computer programs for these models to cater for the class of elliptical distributions, instead of the multivariate normal distribution alone, will involve minimal changes.

5. A random sampling experiment

Results derived in previous sections are all concerned with asymptotic distributions. In order to obtain a rough preliminary impression of their adequacy as approximations with finite sample sizes, a small random sampling experiment was carried out.

The population covariance matrix, Σ_0, had all diagonal elements equal to 1 and all non-diagonal elements equal to 0·5, with the number of variables, p, equal to 8. Sets of 20 sample covariance matrices, $\mathbf{S}$, based on samples of size $N = 500$, were generated from each of two distributions. The first was the multivariate normal distribution while the second was a rescaled multivariate chi-square distribution (Krishnaiah & Rao, 1961; Johnson & Kotz, 1972, Section 40.3) with two degrees of freedom and all marginal relative kurtosis coefficients equal to 3, generated as follows. Let $\mathbf{y}_1 = (y_{11}, \ldots, y_{81})'$ and $\mathbf{y}_2 = (y_{12}, \ldots, y_{82})'$ be generated independently from a multivariate normal distribution with zero means, unit variances and all covariances equal to $0{\cdot}5^{\frac{1}{2}}$. Then the vector $\mathbf{x}$ with a typical element defined by $x_i = 2^{-1}\{y_{i1}{}^2 + y_{i2}{}^2 - 2\}$ has the rescaled multivariate chi-squared distribution with zero means, unit variances, and all covariances equal to 0·5. Examples of a sample covariance matrix from each of these two sets are given in Browne (1982, Tables 1 and 2).

Two models were fitted to each of the 40 sample covariance matrices. One was the intraclass correlation model,

$$\Sigma = \mathbf{1}\varphi\mathbf{1}' + \psi\mathbf{I} \tag{5.1}$$

with two parameters (i.e. $\gamma' = (\varphi, \psi)$) and 34 degrees of freedom. After estimates $\hat{\varphi}$ and $\hat{\psi}$ had been obtained the intraclass correlation estimate $\hat{\rho} = \hat{\varphi}/(\hat{\varphi} + \hat{\psi})$ was also obtained. Population parameter values were $\varphi_0 = 0{\cdot}5$, $\psi_0 = 0{\cdot}5$ and $\rho_0 = 0{\cdot}5$. The other model employed was the factor analysis model with one factor, *regarded as a correlation structure*:

$$\Sigma = \mathbf{D}_\zeta(\boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{D}_\nu)\mathbf{D}_\zeta, \tag{5.2}$$

where each diagonal element ν_{ii} of $\mathbf{D}_\nu$ is regarded as a function of the corresponding factor loading, λ_i:

$$\nu_{ii} = (1 - \lambda_i^2). \tag{5.3}$$

There were 16 parameters (i.e. $\gamma' = (\zeta', \lambda')$) and 20 degrees of freedom. Since the analyses were carried out on sample correlation matrices, the estimates $\hat{\zeta}_i$ were stochastically scaled and were not recorded. The population value of the factor loading vector was $\lambda_0 = 0{\cdot}5^{\frac{1}{2}}\mathbf{1}$ (where $0{\cdot}5^{\frac{1}{2}} \approx 0{\cdot}71$).

Estimates of population parameters were obtained from each of the 40 sample correlation matrices by two methods. Firstly, $F(\mathbf{S}, \Sigma(\gamma) \mid \mathbf{U})$ in (2.7), with $\mathbf{U}$ the unbiased estimate of Υ given by (3.8), was minimized to give asymptotically distribution-free 'best' (ADFB) generalized least squares estimates. Note that, with $N = 500$, the $\mathbf{U}$ defined by (3.4) would give essentially the same results. Secondly, $F_W(\mathbf{S}, \Sigma(\gamma))$ in (2.10) was minimized to give MWL estimates. The asymptotically distribution-free (ADF) test statistic $nF(\mathbf{S}, \Sigma(\hat{\gamma}) \mid \mathbf{U})$, the Wishart likelihood ratio (WLR) test statistic $nF_W(\mathbf{S}, \Sigma(\hat{\gamma}_W))$, the corrected WLR test statistic (CWLR) $n\hat{\eta}^{-1} F_W(\mathbf{S}, \hat{\Sigma}_W)$ in (4.17) and the corrected quadratic form (CQF) test statistic $n\hat{\eta}^{-1} F(\mathbf{S}, \hat{\Sigma}_W \mid \hat{\Sigma}_W)$ in (4.16) were all obtained. Strictly the corrections for kurtosis in the CWLR and CQF test statistics are not applicable to the multivariate chi-square data since the marginal chi-square distributions are skew and the distribution therefore does not belong to the elliptical class. Both models employed are ICSF, however, and the multivariate chi-square distribution does have equal marginal kurtosis coefficients, so that the corrections for kurtosis were applied to gain an impression of their robustness under mild deviations from underlying assumptions.

Results concerning the distributions of the various test statistics for the intraclass correlation model are shown in Table 1 both for the 20 multivariate normal samples

Table 1. Intraclass correlation model: Test statistics (d.f. = 34)

	Normal				Multivariate χ^2			
	ADF	WLR	CWLR	CQF	ADF	WLR	CWLR	CQF
$0 \leqslant 0{\cdot}25$	8	6	6	6	9	20	9	8
$0{\cdot}25 \leqslant 0{\cdot}50$	6	5	6	4	5	0	4	5
$0{\cdot}50 \leqslant 0{\cdot}75$	3	5	5	6	3	0	5	3
$0{\cdot}75 \leqslant 1{\cdot}0$	3	4	3	4	3	0	2	4
$P\{\chi^2(3)\ \text{Fit}\}$	0·31	0·94	0·75	0·85	0·19	0·00	0·16	0·42
$0 \leqslant 0{\cdot}05$	3	2	2	2	3	18	2	2
Mean (p)	0·36	0·47	0·46	0·47	0·35	0·02	0·38	0·38
		Mean ($\hat{\eta}$) = 0·99				Mean ($\hat{\eta}$) = 1·93		
		SD ($\hat{\eta}$) = 0·01				SD ($\hat{\eta}$) = 0·14		

and the 20 multivariate chi-square samples. The first four rows show the frequency distributions of the upper tail probabilities determined from a central chi-square distribution with 34 degrees of freedom. Each cell has an expected value of 5 if this chi-square distribution is appropriate. The fifth row shows the upper tail probability for the usual chi-square goodness-of-fit test (Kendall & Stuart, 1979, chapter 30) based on the four preceding entries. In the sixth row is shown the number of times the test statistic resulted in a rejection of the intraclass correlation model at the 5 per cent level while the last row shows the mean upper tail probability. The means and standard deviations of the sample relative kurtosis coefficients, $\hat{\eta}$, are also shown.

Since only 20 replications were employed, only marked deviations from the chi-square distribution for the ADF, WLR, CWLR and CQF statistics would be detected by the goodness-of-fit test. The only significant result was when the uncorrected WLR statistic was misapplied to the leptokurtic multivariate chi-square samples. Eighteen of the 20 WLR statistics rejected the true null hypothesis at the 5 per cent level and the average value of the upper tail probabilities was 0·02. This clearly illustrates the danger of applying the usual WLR statistics without first investigating the kurtosis of the observations.

It is interesting that the corrections for kurtosis in CWLR and CQF yielded very similar results to the more computationally expensive ADF test statistic in the multivariate chi-square samples. In the multivariate normal samples CWLR and CQF gave similar results to WLR so that the corrections for kurtosis had no discernible harmful effects.

In Table 2 similar information is presented concerning the behaviour of the ADF, WLR, CWLR and CQF tests when employed to test the fit of the factor analysis model. Although the WLR test when misapplied to the multivariate chi-square samples again rejected a correct null hypothesis at the 5 per cent level in a large number of samples, this tendency was less extreme than in Table 1. Noting that 34 degrees of freedom are associated with the intraclass correlation model and 20 with the factor analysis model, it seems plausible that the sensitivity of the WLR test to kurtosis increases as the number of degrees of freedom for the model increases.

Table 2. Factor analysis model: Test statistics (d.f. = 20)

	Normal				Multivariate χ^2			
	ADF	WLR	CWLR	CQF	ADF	WLR	CWLR	CQF
$0 \leqslant 0{\cdot}25$	4	4	4	2	7	15	3	4
$0{\cdot}25 \leqslant 0{\cdot}50$	6	5	6	8	4	3	4	4
$0{\cdot}50 \leqslant 0{\cdot}75$	9	7	6	6	3	0	5	3
$0{\cdot}75 \leqslant 1{\cdot}0$	1	4	4	4	6	1	8	9
$P\{\chi^2(3)\ \mathrm{Fit}\}$	0·08	0·75	0·85	0·26	0·57	0·00	0·42	0·22
$0 \leqslant 0{\cdot}05$	1	1	1	1	2	11	1	1
Mean (p)	0·46	0·50	0·50	0·50	0·45	0·13	0·60	0·60

In Table 3 information is presented concerning the distributions of estimates of parameters of both models in the multivariate normal samples. Corresponding information for the multivariate chi-square samples is given in Table 4. In an attempt to avoid faulty conclusions because of the fairly small number of replications, 95 per cent confidence intervals on means (e.g. Pearson & Hartley, 1958, Section 5.5) and standard deviations (e.g. Pearson & Hartley, 1958, Section 4.3) are given instead of point estimates. For example, the first row gives a 95 per cent confidence interval for the mean $\mu(\hat{\varphi})$ of the distribution of the particular estimator $\hat{\varphi}$ (defined in the column) of the population parameter $\varphi_0 = 0{\cdot}5$ of the intraclass correlation model. In the second row a 95 per cent confidence interval on the standard deviation, $\sigma(\hat{\varphi})$, of the distribution of $\hat{\varphi}$ is given. From *each sample* an *estimate*, $\hat{\sigma}(\hat{\varphi})$, of $\sigma(\hat{\varphi})$ was obtained from the square root of a diagonal element of $n^{-1}\{\mathbf{H}(\mathbf{U}^{-1}, \hat{\mathbf{\Delta}})\}$ (cf. 2.17). A 95 per cent confidence interval on the mean, $\mu(\hat{\sigma}(\hat{\varphi}))$, of the distribution of the sample standard error estimator, $\hat{\sigma}(\hat{\varphi})$, is given in the third row. In this row, the element in the first column (MWL) corresponds to the matrix defined in (2.18) with $\mathbf{V} = \hat{\mathbf{\Sigma}}_{\mathrm{W}}$, the second element (CMWL) corresponds to the correction for kurtosis, $\mathbf{H}(\hat{\boldsymbol{\gamma}}_{\mathrm{W}}, \hat{\eta})$, of (4.15) and the third element (ADFB) corresponds

Table 3. Parameter estimate means and SDs, 95 per cent confidence intervals, normal distribution

	MWL	CMWL	ADFB
$\mu(\hat{\varphi})$	0·49, 0·51		0·47, 0·50
$\sigma(\hat{\varphi})$	0·020, 0·039		0·020, 0·038
$\mu(\hat{\sigma}(\hat{\varphi}))$	0·030, 0·041	0·035, 0·036	0·033, 0·035
$\mu(\hat{\psi})$	0·49, 0·50		0·48, 0·49
$\sigma(\hat{\psi})$	0·009, 0·018		0·010, 0·019
$\mu(\hat{\sigma}(\hat{\psi}))$	0·012, 0·012	0·012, 0·012	0·011, 0·011
$\mu(\hat{\rho})$	0·49, 0·51		0·49, 0·51
$\mu(\hat{\lambda}_1)$	0·69, 0·72		0·70, 0·73
$\sigma(\hat{\lambda}_1)$	0·024, 0·045		0·024, 0·046
$\mu(\hat{\sigma}(\hat{\lambda}_1))$	0·025, 0·026	0·024, 0·026	0·024, 0·026
$\mu(\hat{\lambda}_8)$	0·70, 0·72		0·71, 0·72
$\sigma(\hat{\lambda}_8)$	0·010, 0·020		0·012, 0·024
$\mu(\hat{\sigma}(\hat{\lambda}_8))$	0·025, 0·026	0·025, 0·026	0·024, 0·026
$\bar{\mu}(\hat{\lambda}_i)$	0·71		0·71
$\tilde{\sigma}(\hat{\lambda}_i)$	0·024		0·025
$\bar{\mu}(\hat{\sigma}(\hat{\lambda}_i))$	0·025	0·025	0·025

to $\mathbf{H}(\mathbf{U}^{-1}, \hat{\mathbf{\Delta}})$ in (2.13) with $\mathbf{U}$ given by (3.8). Note that the second column (CMWL) only has terms for the kurtosis corrected standard error estimates of the MWL estimators. If the confidence interval on $\mu(\hat{\sigma}(\hat{\varphi}))$ does not overlap the confidence interval on $\sigma(\hat{\varphi})$ this suggests that $\hat{\sigma}(\hat{\varphi})$ is an inappropriate estimator for $\sigma(\hat{\varphi})$ (cf. Table 4, column MWL).

Corresponding intervals are given for $\hat{\psi}$, the intraclass correlation estimate, $\hat{\rho}$, the first factor loading estimate, $\hat{\lambda}_1$, and the last factor loading estimate, $\hat{\lambda}_8$. Since the population covariance matrix is invariant under any reordering of variables, all factor loadings will have the same distribution. The fourteenth row, $\bar{\mu}(\hat{\lambda}_i)$, and last row $\bar{\mu}(\hat{\sigma}(\lambda_i))$ were obtained by averaging means for the first and last factor loadings. The fifteenth row, $\tilde{\sigma}(\hat{\lambda}_i)$, was obtained by taking the square root of the average variance for these two factor loadings.

Examination of Table 3 shows that, for the multivariate normal samples, both the MWL and ADFB estimates did not show any marked bias although a significant but negligible bias was apparent in the ADFB $\hat{\psi}$. There was a lack of overlap of the MWL $\sigma(\hat{\lambda}_8)$ interval with either the MWL or CMWL $\mu(\hat{\sigma}(\hat{\lambda}_8))$ intervals but this can probably be disregarded in view of the fact that the corresponding intervals for $\hat{\lambda}_1$ did overlap and the MWL $\tilde{\sigma}(\hat{\lambda}_i)$ was very close to both the MWL and CMWL $\bar{\mu}(\hat{\sigma}(\hat{\lambda}_i))$. For the multivariate normal samples, therefore, the MWL and ADFB parameter estimates and the MWL, CMWL and ADFB standard error estimates appeared quite satisfactory.

In Table 4, it can be seen that, for the multivariate chi-square samples, the inappropriate MWL standard error estimates were obviously biased and did not generally yield confidence intervals which overlapped with those for the standard errors estimated. This was rectified by the correction for kurtosis and the ADFB standard error estimates also appeared reasonable. For example, the MWL $\mu(\hat{\sigma}(\hat{\varphi}))$ interval does *not* overlap the MWL $\sigma(\hat{\varphi})$ interval but the CMWL $\mu(\hat{\sigma}(\hat{\varphi}))$ interval does overlap the MWL $\sigma(\hat{\varphi})$ interval and the ADFB $\mu(\hat{\sigma}(\hat{\varphi}))$ interval overlaps the ADFB $\sigma(\hat{\varphi})$ interval.

Table 4. Parameter estimate means and SDs, 95 per cent confidence intervals, multivariate χ^2 distribution

	MWL	CMWL	ADFB
$\mu(\hat{\varphi})$	0·48, 0·55		0·32, 0·37
$\sigma(\hat{\varphi})$	0·057, 0·110		0·045, 0·087
$\mu(\hat{\sigma}(\hat{\varphi}))$	0·035, 0·039	0·051, 0·060	0·038, 0·047
$\mu(\hat{\psi})$	0·49 , 0·51		0·41, 0·43
$\sigma(\hat{\psi})$	0·018, 0·034		0·014, 0·027
$\mu(\hat{\sigma}(\hat{\psi}))$	0·012, 0·012	0·026, 0·028	0·018, 0·020
$\mu(\hat{\rho})$	0·49, 0·52		0·43, 0·47
$\mu(\hat{\lambda}_1)$	0·70, 0·73		0·68, 0·72
$\sigma(\hat{\lambda}_1)$	0·029, 0·056		0·029, 0·057
$\mu(\hat{\sigma}(\hat{\lambda}_1))$	0·024, 0·026	0·033, 0·036	0·034, 0·039
$\mu(\hat{\lambda}_8)$	0·71, 0·74		0·70, 0·73
$\sigma(\hat{\lambda}_8)$	0·025, 0·049		0·028, 0·054
$\mu(\hat{\sigma}(\hat{\lambda}_8))$	0·023, 0·026	0·033, 0·035	0·032, 0·039
$\bar{\mu}(\hat{\lambda}_i)$	0·72		0·71
$\bar{\sigma}(\hat{\lambda}_i)$	0·036		0·038
$\bar{\mu}(\hat{\sigma}(\hat{\lambda}_i))$	0·025	0·034	0·036

While all MWL parameter estimators in Table 4 appear reasonable, the ADFB estimators of $\varphi_0 = 0{\cdot}5$ and $\psi_0 = 0{\cdot}5$ are clearly and unacceptably biased despite the fact that the model is linear in the parameters. The corresponding MWL estimates are linear functions of the elements of **S** (Wilks, 1946) and are consequently unbiased. It seems that the ADFB estimators of parameters in covariance structures tend to underestimate and this tendency is accentuated when there is a fair amount of random error in **S** due to small samples, or as is the case here, large values of η.

When parameters in a correlation structure are estimated this tendency is less noticeable and the bias is absorbed by the standard deviation estimates $\hat{\mathbf{D}}_\zeta$ (see (5.2)). The ADFB factor loading estimates in Table 4 are quite acceptable although the ADFB intraclass correlation $\hat{\rho}$ is still biased, but less so than $\hat{\varphi}$ and $\hat{\psi}$.

Practical applications of ADF methods are reported in Browne (1982, Section 2.9) and in Huba & Bentler (1983).

6. Conclusions

It is clear from the asymptotic theory of Section 2 that the usual tests of fit of a model and standard errors for parameter estimators derived under the assumption of multivariate normality should not be employed if the distribution of the observations is kurtose. This conclusion is supported by the results of the random sampling experiment of Section 5, where the application of the normal theory likelihood ratio test to samples from a leptokurtic distribution resulted in a too frequent rejection of a true null hypothesis. The asymptotic theory suggests that the opposite would be true if the normal theory likelihood ratio test were to be applied to samples from a platykurtic distribution and a true null hypothesis would be rejected less frequently than is specified by the significance level employed. Standard errors for parameter estimators derived under the assumption of a multivariate normal distribution also are incorrect if the data have a kurtose distribution.

One way of tackling this problem is to make use of generalized least squares estimators, obtained by minimizing the quadratic form discrepancy function

$F(\mathbf{S}; \boldsymbol{\Sigma}(\boldsymbol{\gamma}) | \mathbf{U})$ in (2.7), where $\mathbf{U}$ is a consistent estimator of the limiting covariance matrix, $\bar{\mathbf{Y}}$, of $n^{\frac{1}{2}}(\mathbf{s}-\boldsymbol{\sigma}_0)$. These estimators are asymptotically 'best' in the sense of having minimum variance and the limiting distribution of $nF(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}) | \mathbf{U})$ under the null hypothesis is chi-square for any distribution of the observations (with finite eighth-order moments). In the random sampling experiment the test statistic and standard error estimates appeared reasonably satisfactory. Estimates of parameters in a *covariance* structure appeared to be unacceptably biased. Unacceptable bias in estimates of parameters in a *correlation* structure was not found. A disadvantage of this approach is the amount of computation involved.

Another approach is to employ the usual normal theory maximum likelihood estimates and apply the corrections for kurtosis to the test statistics and standard errors described in Section 4. These corrections are intended for distributions from the elliptical class only, but performed well when applied to the non-elliptical multivariate chi-square distribution in the random sampling experiment. Further random sampling experiments in which these corrections are applied when the distribution of the observations has widely differing marginal kurtosis coefficients would be of interest. An attempt should be made to find out whether there are circumstances where the corrections for kurtosis give *poorer* results than the usual unmodified normal theory results.

A further approach has been investigated theoretically but has not yet been tried out empirically. This is to make use of convenient estimates, such as normal theory maximum likelihood or ordinary least squares, and use Propositions 4 and 2 to obtain an asymptotically distribution-free test statistic and standard errors. This approach still has the disadvantage of requiring an estimate of the large matrix $\bar{\mathbf{Y}}$, but this is employed once only, instead of in an iterative algorithm. Furthermore, the problem of excessive bias, mentioned earlier, could be avoided. A random sampling investigation would be of interest.

Acknowledgements

Part of the research reported here was carried out in 1980 at the Center for the Study of Adolescent Drug Abuse Etiologies of the University of California, Los Angeles under support of grant number DAO1070 from the National Institute on Drug Abuse. I am indebted to Peter M. Bentler and George J. Huba for providing research facilities and to J. O. Horowitz for help with the random sampling experiment of Section 5.

Many helpful discussions with Alexander Shapiro are gratefully acknowledged.

References

Bentler, P. M. (1983). Simultaneous equation systems as moment structure models: With an introduction to latent variable models. *Journal of Econometrics*, **22**, 13–42.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal* **8**, 1–24. (Reprinted in D. J. Aigner & A. S. Goldberger (eds), *Latent Variables in Socio-economic Models*. Amsterdam: North Holland, 1977.)

Browne, M. W. (1979). The maximum likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology*, **32**, 75–86.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (ed.), *Topics in Applied Multivariate Analysis*, pp. 72–141. Cambridge: Cambridge University Press.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics*, **22**, 113–137.

Dennis, J. E. (1977). Non-linear least squares and equations. In D. Jacobs (ed.), *The State of the Art in Numerical Analysis*, pp. 269–312. London: Academic 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*, pp. 365–395. 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.
Du Toit, S. H. C. (1979). The analysis of growth curves. Unpublished PhD thesis. Pretoria: University of South Africa.
Gayen, A. K. (1951). The frequency distribution of the product moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, **38**, 219–247.
Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimisation*. London: Academic Press.
Huba, G. J. & Bentler, P. M. (1983). Test of a drug use causal model using asymptotically distribution free methods. *Journal of Drug Education*, **13**, 3–14.
Jennrich, R. I. & Sampson, P. F. (1968). Application of stepwise regression to non-linear estimation. *Technometrics*, **10**, 63–72.
Johnson, N. L. & Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.
Kendall, M. G. & Stuart, A. (1969). *The Advanced Theory of Statistics*, vol. 1, 3rd ed. London: Griffin.
Kendall, M. G. & Stuart, A. (1979). *The Advanced Theory of Statistics*, vol. 2, 4th ed. London: Griffin.
Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics*, **18**, 75-86.
Kraemer, H. C. (1980). Robustness of the distribution theory of the product moment correlation coefficient. *Journal of Educational Statistics*, **5**, 115–128.
Krishnaiah, P. R. & Rao, M. M. (1961). Remarks on a multivariate gamma distribution. *American Mathematical Monthly*, **68**, 342–346.
Layard, M. W. J. (1972). Large sample tests for the equality of two covariance matrices. *Annals of Mathematical Statistics*, **43**, 123–141.
Layard, M. W. J. (1974). A Monte Carlo comparison of tests for equality of covariance matrices. *Biometrika*, **61**, 461–465.
Lee, S. Y. & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal*, **14**, 121–136.
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 and kurtosis in testing normality and robustness studies. *Sankhyā, 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. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.
Muirhead, R. J. & Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, **67**, 31–43.
Nel, D. G. (1980). On matrix differentiation in statistics. *South African Statistical Journal*, **14**, 137–193.
Pearson, E. S. & Hartley, H. O. (1958). *Biometrika Tables for Statisticians*, vol. 1. Cambridge: Cambridge University Press.
Press, S. J. (1972). *Applied Multivariate Analysis*. New York: Holt, Rinehart & Winston.
Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal*, **17**, 33–81.
Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology*, **37**, (in press).
Silvey, S. D. (1970). *Statistical Inference*. Baltimore: Penguin.
Steiger, J. H. & Browne, M. W. (1984). The comparison of interdependent correlations between optimal linear composites. *Psychometrika* (in press).
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.
Stroud, T. W. (1972). Fixed alternatives and Wald's formulation of the noncentral asymptotic behaviour of the likelihood ratio statistic. *Annals of Mathematical Statistics*, **43**, 447–454.
Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics*, **17**, 257–281.
Yu, M. C. & Dunn, O. J. (1983). Robust tests for the equality of two correlation coefficients: A Monte Carlo study. *Educational and Psychological Measurement*, **42**, 987–1004.

Received 28 June 1983; revised version received 9 December 1983

Requests for reprints should be addressed to M. W. Browne, Department of Statistics, University of South Africa, PO Box 392, Pretoria 0001, South Africa.