Analysis of Covariance Structures Under Elliptical Distributions

ALEXANDER SHAPIRO and MICHAEL W. BROWNE*

This article examines the adjustment of normal theory methods for the analysis of covariance structures to make them applicable under the class of elliptical distributions. It is shown that if the model satisfies a mild scale invariance condition and the data have an elliptical distribution, the asymptotic covariance matrix of sample covariances has a structure that results in the retention of many of the asymptotic properties of normal theory methods. If a scale adjustment is applied, the likelihood ratio tests of fit have the usual asymptotic chi-squared distributions. Difference tests retain their property of asymptotic independence, and maximum likelihood estimators retain their relative asymptotic efficiency within the class of estimators based on the sample covariance matrix. An adjustment to the asymptotic covariance matrix of normal theory maximum likelihood estimators for elliptical distributions is provided. This adjustment is particularly simple in models for patterned covariance or correlation matrices. These results apply not only to normal theory maximum likelihood methods but also to a class of minimum discrepancy methods. Similar results also apply when certain robust estimators of the covariance matrix are employed.

KEY WORDS: Affine invariant M-estimators; Asymptotic distributions; Maximum likelihood; Minimum discrepancy.

1. INTRODUCTION

A considerable part of classical multivariate analysis is devoted to hypotheses concerning the population covariance matrix, Σ. The associated statistical inference is well developed under the assumption that the sample is drawn from a normally distributed population (e.g., Muirhead 1982). These normal theory methods can, however, be sensitive to deviations from normality and, in particular, to the kurtosis of data distributions. Multivariate distributions that are convenient for investigating the sensitivity of normal theory methods to kurtosis are the elliptical distributions (Chmielewski 1981; Devlin, Gnanadesikan, and Kettenring 1976; Muirhead 1982, sec. 1.5). The elliptical class of distributions incorporates a single additional kurtosis parameter, κ, and contains the multivariate normal distribution as a special case with $\kappa = 0$.

Elliptical distributions have been employed in two general approaches yielding somewhat different results. In one, an $N \times p$ data matrix is regarded as being distributed according to an Np-dimensional elliptical distribution. Elements in different rows of the data matrix are regarded as uncorrelated but *not independent* if $\kappa \neq 0$. Under these conditions certain normal theory likelihood ratio tests remain valid *without correction* (Anderson, Fang, and Hsu 1986; Chmielewski 1980).

The present article will adopt the other approach where rows of the data matrix are regarded as being *independently and identically* distributed according to a p-variate elliptical distribution. Under these assumptions a number of situations were found (Muirhead 1982; Muirhead and Waternaux 1980) where normal theory likelihood ratio tests retain their asymptotic chi-squared distribution if divided by a correction factor dependent on kurtosis. Tyler (1983) gave a class of null hypotheses defined by equality constraints on elements of Σ where these scale corrections for the likelihood ratio test are applicable. A class of structural models $\Sigma = \Sigma(\boldsymbol{\theta})$, where scale corrections for likelihood ratio goodness-of-fit tests are applicable and normal theory maximum likelihood parameter estimators retain their relative asymptotic efficiency within a certain class of estimators, was given in Browne (1982, 1984). Tyler (1982, 1983) also showed that correction factors can be found when the usual sample covariance matrix is replaced by an alternative estimator from the class of M-estimators (Maronna 1976) or by an estimator of Σ that is a maximum likelihood estimator under the assumption of some specific elliptical distribution.

The present article unifies and extends the findings of Tyler (1982, 1983) and Browne (1982, 1984). This is done by showing that their two superficially different sets of conditions on Σ both imply a property of the model that justifies scale corrections to the test statistic. We consider a class of minimum discrepancy test statistics, which includes the previously considered normal theory likelihood ratio statistic, and show that similar scale corrections also apply to difference tests with constrained alternative hypotheses. A new test statistic that does not require a scale correction for the kurtosis of an elliptical distribution is also obtained. The result concerning robustness of the asymptotic efficiency of maximum likelihood estimators given by Browne (1982, 1984) is extended to other discrepancy functions based on the wider class of covariance matrix estimates considered by Tyler (1982, 1983). In addition, we provide a new correction factor of rank 1 to the asymptotic covariance matrix of estimators that is more direct and simpler to apply than the correction factor given in Browne (1982, 1984).

2. PRELIMINARY DISCUSSION AND NOTATION

Let Σ_0 represent the $p \times p$ population covariance matrix. Suppose that we wish to test the null hypothesis that Σ_0 belongs to a given set $\mathcal{P}$ against the unrestricted alternative. The parameter set $\mathcal{P}$ constitutes a subset of the set of positive definite matrices and can be specified in two standard ways. It can be defined by a number of equality constraints

$$\mathcal{P} = \{\Sigma : h_i(\Sigma) = 0, i = 1, \ldots, k\}, \qquad (2.1)$$

* Alexander Shapiro is Professor, Department of Mathematics and Applied Mathematics, and Michael W. Browne is Professor, Department of Statistics, both at the University of South Africa, Pretoria 0001, South Africa. The authors thank an associate editor and the referees for helpful comments.

Journal of the American Statistical Association
December 1987, Vol. 82, No. 400, Theory and Methods

or it can be defined by a parameterization

$$\mathcal{P} = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} = \mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathbb{S}\}, \qquad (2.2)$$

where $\mathbf{G}(\boldsymbol{\theta})$ is a symmetric matrix valued function of the $q \times 1$ vector $\boldsymbol{\theta}$. We shall suppose throughout that in both cases all functions involved are continuously differentiable. If the constraints (2.1) apply and the gradient vectors $\partial h_i/\partial\boldsymbol{\Sigma}$ ($i = 1, \ldots, k$) evaluated at the point $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ are linearly independent, it follows from the implicit function theorem that a parameterization (2.2) exists, locally at least, with $q = p(p + 1)/2 - k$ parameters. On the other hand, if the set $\mathcal{P}$ is given in the form (2.2), then theoretically it also can be locally represented by equality constraints as in (2.1). In particular situations $\mathcal{P}$ can be defined either in the manner of (2.1) or of (2.2) and, therefore, we shall treat them simultaneously.

The function $\mathbf{G}(\boldsymbol{\theta})$, which relates the parameter vector $\boldsymbol{\theta}$ from the specified set $\mathbb{S}$ to $\boldsymbol{\Sigma}$, is called a structural model. Clearly, the null hypothesis is equivalent to the existence of a $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_0 \in \mathbb{S}$, such that $\mathbf{G}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}_0$. The structural model is said to be linear if $\mathbf{G}(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, that is,

$$\mathbf{G}(\boldsymbol{\theta}) = \theta_1\mathbf{G}_1 + \theta_2\mathbf{G}_2 + \cdots + \theta_q\mathbf{G}_q, \qquad (2.3)$$

where $\mathbf{G}_1, \ldots, \mathbf{G}_q$ are known symmetric matrices. Possibly the simplest linear model, associated with the test for sphericity, is the matrix valued function $\mathbf{G}(\lambda) = \lambda\mathbf{I}$ of the single parameter $\lambda > 0$. Other examples of linear covariance structures are patterned covariance matrices of various types (Mukherjee 1970; Szatrowski 1985) and components-of-covariance models [Bock and Bargmann 1966, cases I and II; Wiley, Schmidt, and Bramble 1973, models (1)–(4)].

An example of a nonlinear covariance structure is provided by the well-known factor analysis model,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}, \qquad (2.4)$$

where $\boldsymbol{\Lambda}$ is the $p \times l$ matrix of factor loadings and $\boldsymbol{\Psi}$ is the diagonal matrix of unique variances (Lawley and Maxwell 1971). Here $\mathbf{G}(\boldsymbol{\theta})$ is given by the expression on the right side in (2.4) with the parameter vector $\boldsymbol{\theta}$ composed of the elements of $\boldsymbol{\Lambda}$ and the diagonal elements of $\boldsymbol{\Psi}$. Notice that the parameter matrix $\boldsymbol{\Lambda}$ in (2.4) is not identified unless additional constraints are imposed. Many other examples of predominantly nonlinear structural models for covariance matrices may be found in Jöreskog (1981) and Browne (1982).

Let $\mathbf{S}$ be the sample covariance matrix, based on a sample of size $n + 1$, giving an unbiased estimate of $\boldsymbol{\Sigma}_0$. When the sample is drawn from a normal population, $\mathbf{S}$ has a Wishart distribution. The usual likelihood ratio test statistic, $-2 \log \lambda$, can be shown (e.g., Jöreskog 1967, p. 457) to be equal to $n\hat{F}$, where

$$\hat{F} = \min_{\boldsymbol{\Sigma} \in \mathcal{P}} F(\mathbf{S}, \boldsymbol{\Sigma}) \qquad (2.5)$$

and

$$F(\mathbf{S}, \boldsymbol{\Sigma}) = \log|\boldsymbol{\Sigma}| - \log|\mathbf{S}| + \operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - p. \qquad (2.6)$$

The function $F(\mathbf{S}, \boldsymbol{\Sigma})$, defined in (2.6), is an example of a discrepancy function. In general, a real-valued function $F(\mathbf{S}, \boldsymbol{\Sigma})$ of two matrix variables is called a *discrepancy function* if it is twice continuously differentiable, nonnegative valued, and $F(\mathbf{S}, \boldsymbol{\Sigma}) = 0$ if and only if $\mathbf{S} = \boldsymbol{\Sigma}$ (compare Browne 1982, p. 81). Various examples of discrepancy functions may be found in Swain (1975) and Browne (1982). The test statistic $n\hat{F}$, where $\hat{F}$ is the minimum (2.5) associated with a chosen discrepancy function F, will be referred to as a *minimum discrepancy test statistic*. The corresponding minimum discrepancy estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is defined as a minimizer of $F(\mathbf{S}, \mathbf{G}(\cdot))$ over the set $\mathbb{S}$.

A discrepancy function F can be considered as a function of two vector variables $\mathbf{s} = \operatorname{vec}(\mathbf{S})$ and $\boldsymbol{\sigma} = \operatorname{vec}(\boldsymbol{\Sigma})$, where $\operatorname{vec}(\mathbf{S})$ denotes the $p^2 \times 1$ vector formed by stacking columns of $\mathbf{S}$. It can be shown that the second-order Taylor approximation of $F(\mathbf{s}, \boldsymbol{\sigma})$ at the point $(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_0)$, with $\boldsymbol{\sigma}_0 = \operatorname{vec}(\boldsymbol{\Sigma}_0)$, is given by the quadratic function $(\mathbf{s} - \boldsymbol{\sigma})'\mathbf{V}_0(\mathbf{s} - \boldsymbol{\sigma})$ of the vector variables $\mathbf{s}$ and $\boldsymbol{\sigma}$ (Shapiro 1985a). In this quadratic function the $p^2 \times p^2$ weight matrix $\mathbf{V}_0$ is given by the matrix $\frac{1}{2}\partial^2 F/\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'$ of second-order partial derivatives evaluated at $(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_0)$. The asymptotic distributions of $n\hat{F}$ and $\hat{\boldsymbol{\theta}}$ depend on the weight matrix $\mathbf{V}_0$, which is determined by the particular discrepancy function F employed. It will be assumed that the discrepancy function F is *correctly specified for a normal distribution* of the data. This means that the matrix $\mathbf{V}_0 = \boldsymbol{\Gamma}_N^-$ is a generalized inverse of the asymptotic covariance matrix $\boldsymbol{\Gamma}_N$ of $\mathbf{s}^* = n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$ obtained under the assumption of normality. Most available computer packages make use of discrepancy functions that are correctly specified for normal distributions.

3. CORRECTIONS TO THE MINIMUM DISCREPANCY TEST STATISTICS

In this section we study scale corrections to minimum discrepancy test statistics derived from normal theory to enable them to be employed when sampling from a population with a distribution from the elliptical class. We assume throughout that the asymptotic distribution of the elements of the sample covariance matrix $\mathbf{S}$ is multivariate normal with a null mean vector and a certain $p^2 \times p^2$ covariance matrix $\boldsymbol{\Gamma}$. This is ensured by the central limit theorem, assuming existence of fourth-order moments. The basic asymptotic result, employed in the studies of Muirhead and Waternaux (1980), Tyler (1982, 1983), and Browne (1982, 1984), is that $\boldsymbol{\Gamma}$ has the following structure:

$$\boldsymbol{\Gamma} = \alpha\boldsymbol{\Gamma}_N + \beta\boldsymbol{\sigma}_0\boldsymbol{\sigma}_0', \qquad (3.1)$$

where $\alpha = 1 + \kappa$, $\beta = \kappa$, and κ is the kurtosis parameter of a distribution from the elliptical class. Tyler (1982) generalized this result to estimators of $\boldsymbol{\Sigma}_0$ other than the usual sample covariance matrix. He showed that if $\mathbf{S}$ is an estimator of $\boldsymbol{\Sigma}_0$ such that $n^{1/2}(\mathbf{S} - \boldsymbol{\Sigma}_0)$ converges in distribution to a multivariate normal and a certain invariance property is satisfied, then there exist scalars α and β such that the associated covariance matrix $\boldsymbol{\Gamma}$ has the form (3.1).

In particular, estimators with this invariance property include the sample covariance matrix, the maximum likelihood estimator of $\mathbf{\Sigma}_0$ appropriate for any particular distribution from the elliptical class, and the affine invariant M-estimators (Maronna 1976). A discussion of the scalars α and β may be found in Tyler (1982, 1983). We shall assume henceforth that $\mathbf{S}$ is an estimator of $\mathbf{\Sigma}_0$ such that (3.1) holds.

Robustness properties of the test statistic $n\hat{F}$ have been investigated in Shapiro (1985b, 1986). These results will be applied to our situation. Consider the parameter set $\mathcal{P}$ of positive definite matrices $\mathbf{\Sigma}$ defined alternatively by (2.1) or (2.2). We retain the same notation $\mathcal{P}$ for the corresponding set of vectors $\boldsymbol{\sigma} = \text{vec}(\mathbf{\Sigma})$. Under certain regularity conditions, $\mathcal{P}$ is a smooth manifold in a neighborhood of $\boldsymbol{\sigma}_0 \in \mathcal{P}$ and, therefore, has a tangent space T_0 at $\boldsymbol{\sigma}_0$. It is not essential for subsequent understanding for precise definitions to be given of the concepts of "smooth manifold" and "tangent space." Definitions can be found in almost any standard text on differential geometry (e.g., Hirsch 1976). It will be sufficient to bear in mind that T_0 is a linear space generated by vectors orthogonal to the gradients $(\partial/\partial\boldsymbol{\sigma})h_i(\boldsymbol{\sigma}_0)$ $(i = 1, \ldots, k)$ of the constraint functions when $\mathcal{P}$ is defined by (2.1). Alternatively, when $\mathcal{P}$ is defined by (2.2) and $\boldsymbol{\theta}_0$ is an interior point of $\mathbb{S}$, T_0 is a linear space generated by the column vectors of the $p^2 \times q$ Jacobian matrix $\mathbf{\Delta} = (\partial/\partial\boldsymbol{\theta}')g(\boldsymbol{\theta}_0)$, where $\mathbf{g}(\boldsymbol{\theta}) = \text{vec}(G(\boldsymbol{\theta}))$. In the case of a parameterized model with equality constraints on the parameters, where the parameter set $\mathbb{S}$ is defined by a number of equality constraints

$$\mathbb{S} = \{\boldsymbol{\theta} : c_i(\boldsymbol{\theta}) = 0, i = 1, \ldots, m\}, \tag{3.2}$$

then

$$T_0 = \{\boldsymbol{\tau} : \boldsymbol{\tau} = \mathbf{\Delta}\boldsymbol{\gamma}, \mathbf{Z}'\boldsymbol{\gamma} = \mathbf{0}\}, \tag{3.3}$$

where $\mathbf{Z}$ is the $q \times m$ Jacobian matrix $\mathbf{Z} = (\partial/\partial\boldsymbol{\theta})\mathbf{c}(\boldsymbol{\theta}_0)$. A discussion of regularity conditions involved and statistical implications is given in Shapiro (1986).

Equation (3.1) implies that the difference between the adjusted covariance matrix $\alpha^{-1}\mathbf{\Gamma}$ and the covariance matrix $\mathbf{\Gamma}_N$ corresponding to the normal case is a rank 1 matrix collinear with $\boldsymbol{\sigma}_0\boldsymbol{\sigma}_0'$. It then follows (Shapiro 1985b, th. 3.2) that under the null hypothesis the adjusted test statistic $\alpha^{-1}n\hat{F}$ is asymptotically chi-squared if and, when $\beta \neq 0$, only if $\boldsymbol{\sigma}_0$ belongs to the tangent space T_0. The corresponding degrees of freedom are given by $p(p + 1)/2 - d$, where d is the dimension of T_0. Furthermore, a similar, noncentral, result can be obtained for a sequence of local alternatives. Let $\mathbf{\Sigma}_{0,n}$ represent a *sequence* of population covariance matrices converging to a matrix $\mathbf{\Sigma}_0$ from the parameter set $\mathcal{P}$. Define

$$F_{0,n} = \min\{F(\mathbf{\Sigma}_{0,n}, \mathbf{X}) : \mathbf{X} \in \mathcal{P}\},$$

which corresponds to the population lack of fit of the model. Suppose that $nF_{0,n}$ tends to a limit δ as $n \to \infty$. Then $\alpha^{-1}n\hat{F}$ has an asymptotic distribution that is noncentral chi-squared if $\boldsymbol{\sigma}_0 \in T_0$. The associated noncentrality parameter is given by $\alpha^{-1}\delta$ (see Steiger, Shapiro, and Browne 1985). If a consistent estimator $\hat{\alpha}$ of the parameter α is available, these asymptotic results still hold when α is replaced by $\hat{\alpha}$.

We summarize our discussion in the following theorem.

Theorem 1. Under a sequence of local alternatives the adjusted test statistic $\hat{\alpha}^{-1}n\hat{F}$ is asymptotically chi-squared with noncentrality parameter $\alpha^{-1}\delta$ and degrees of freedom $p(p + 1)/2 - d$, if and, when $\beta \neq 0$, only if $\boldsymbol{\sigma}_0 \in T_0$.

Verification of the condition $\boldsymbol{\sigma}_0 \in T_0$ requires knowledge of the tangent space, T_0, which in some cases is not readily available. There is, however, an easily verifiable property of $\mathcal{P}$ that will ensure that $\boldsymbol{\sigma}_0 \in T_0$. The set $\mathcal{P}$ is said to be positively homogeneous if $\mathbf{\Sigma} \in \mathcal{P}$ implies that $t\mathbf{\Sigma} \in \mathcal{P}$ for every positive scalar t. A proof of the following lemma is given in the Appendix.

Lemma 1. If the set $\mathcal{P}$ is positively homogeneous, then every $\boldsymbol{\sigma}_0 \in \mathcal{P}$ belongs to the corresponding tangent space T_0.

When $\mathcal{P}$ is defined in the parameterized form (2.2) and the corresponding parameter set $\mathbb{S}$ is given by (3.2) such that (3.3) holds, then $\boldsymbol{\sigma}_0 \in T_0$ means that there exists a vector $\boldsymbol{\zeta}$ such that $\boldsymbol{\sigma}_0 = \mathbf{\Delta}\boldsymbol{\zeta}$ and $\mathbf{Z}'\boldsymbol{\zeta} = \mathbf{0}$.

We now have the following result.

Corollary 1. Under a sequence of local alternatives, $\hat{\alpha}^{-1}n\hat{F}$ is asymptotically chi-squared if the parameter set $\mathcal{P}$ is postively homogeneous.

If the set $\mathcal{P}$ is defined by equality constraints as in (2.1), then it will be positively homogeneous if the constraint functions h_i are positively homogeneous of degree η; that is, the following condition holds.

Condition 1. There exists a constant η such that $h_i(t\mathbf{\Sigma}) = t^{\eta}h_i(\mathbf{\Sigma})$ $(i = 1, \ldots, k)$ for all $t > 0$ and all positive definite matrices $\mathbf{\Sigma}$.

For example, consider the factor analysis model (2.4) with one factor, $l = 1$. Then it is well known (Anderson and Rubin 1956, p. 116) that the corresponding set $\mathcal{P}$ can be defined by equality constraints (tetrad conditions) involving second-order determinants of off-diagonal elements of $\mathbf{\Sigma}$. In this case Condition 1 holds with $\eta = 2$.

Condition 1 with $\eta = 0$, and an asymptotic result similar to that of Corollary 1, are due to Tyler (1983, corollary 1).

When $\mathcal{P}$ is defined in the parameterized form of (2.2), it will be positively homogeneous if and only if the following condition holds.

Condition 2. For every $t > 0$ and $\boldsymbol{\theta} \in \mathbb{S}$ there exists a $\bar{\boldsymbol{\theta}} \in \mathbb{S}$ such that $t\mathbf{G}(\boldsymbol{\theta}) = \mathbf{G}(\bar{\boldsymbol{\theta}})$.

A model that satisfies this condition was said to be "invariant under a constant scaling factor" in Browne (1982, 1984). Clearly this condition is equivalent to positive homogeneity of $\mathcal{P}$. The linear model (2.3), for example, is invariant under a constant scaling factor if no constraints are imposed on the θ_i $(i = 1, \ldots, q)$.

Now let $\mathcal{P}^*$ be a subset of $\mathcal{P}$ and suppose that we wish to test the null hypothesis $\Sigma_0 \in \mathcal{P}^*$ against the alternative $\Sigma_0 \in \mathcal{P}$. For instance, if $\mathcal{P}$ is defined in the parameterized form (2.2), then $\mathcal{P}^*$ can be obtained by imposing a number of additional equality constraints on the parameter vector $\boldsymbol{\theta}$ as in (3.2). This hypothesis may be tested by means of the difference test statistic $n(\bar{F} - \hat{F})$, where

$$\bar{F} = \min_{\Sigma \in \mathcal{P}^*} F(\mathbf{S}, \Sigma).$$

Notice that if the discrepancy function F is defined by (2.6), then the difference test statistic is the usual likelihood ratio test statistic of normal theory with a restricted alternative hypothesis.

Under a sequence of local alternatives, $\Sigma_{0,n} \rightarrow \Sigma_0 \in \mathcal{P}^*$, we define

$$F^*_{0,n} = \min\{F(\Sigma_{0,n}, \mathbf{X}) : \mathbf{X} \in \mathcal{P}^*\}$$

and suppose that $nF^*_{0,n}$ tends to a limit δ^* as $n \rightarrow \infty$. We assume that $\mathcal{P}^*$ is a smooth manifold and hence has a tangent space T^*_0 at the point Σ_0. The tangent space, T^*_0, is a subspace of T_0 of dimension d^*, $d^* \leq d = \dim T_0$, and may be defined in a similar manner. Then the following result is implied by a general robustness theory of moment structures (Shapiro 1985b, sec. 4) and the preceding discussion.

Theorem 2. If $\boldsymbol{\sigma}_0 \in T^*_0$, then under a sequence of local alternatives: (i) the adjusted difference test statistic $\hat{\alpha}^{-1}n(\bar{F} - \hat{F})$ is asymptotically chi-squared with noncentrality parameter $\alpha^{-1}(\delta^* - \delta)$ and $d - d^*$ degrees of freedom, and (ii) the test statistics $\hat{\alpha}^{-1}n\hat{F}$ and $\hat{\alpha}^{-1}n(\bar{F} - \hat{F})$ are asymptotically independent.

Our previous discussion implies that if the set $\mathcal{P}^*$ is positively homogeneous, then $\boldsymbol{\sigma}_0 \in T^*_0$ and the following result consequently holds.

Corollary 2. If $\mathcal{P}^*$ is positively homogeneous, then properties (i) and (ii) of Theorem 2 hold.

Another interesting result follows. Properties (i) and (ii) of Theorem 2 and Theorem 1 show that the ratio statistic

$$[(\bar{F} - \hat{F})\{p(p + 1)/2 - d\}]/\{\hat{F}(d - d^*)\} \quad (3.4)$$

has an asymptotic doubly noncentral F distribution.

Corollary 3. If $\boldsymbol{\sigma}_0 \in T^*_0$, then under a sequence of local alternatives the ratio statistic of (3.4) has an asymptotic doubly noncentral F distribution with noncentrality parameters $\alpha^{-1}(\delta^* - \delta)$ and $\alpha^{-1}\delta$ and with $d - d^*$ and $p(p + 1)/2 - d$ degrees of freedom.

Note that the ratio statistic is independent of α and asymptotically has a central F distribution under the null hypothesis. This makes it possible to test the null hypothesis without obtaining an estimate of the parameter α. In general, however, this results in loss of the asymptotic local power compared with the test based on the difference test statistic $\hat{\alpha}^{-1}n(\bar{F} - \hat{F})$ with a known or well-estimated parameter α. Since an F distribution becomes approximately chi-squared for large values of the second number of degrees of freedom, this loss of power may be negligible for large values of the number $p(p + 1)/2 - d$.

The asymptotic local power of the minimum discrepancy test statistics discussed in this section depends on the noncentrality parameters $\alpha^{-1}\delta$ and $\alpha^{-1}(\delta^* - \delta)$ and consequently is affected by the parameter α. Therefore, it is advantageous to make use of an estimator $\mathbf{S}$ of Σ_0 with a small value of α, as one can then expect a greater power of the minimum discrepancy test statistic. As we shall see, a similar phenomenon occurs in the estimation of parameters. An interesting study of interrelationships between the parameters α corresponding to various estimators $\mathbf{S}$ may be found in Tyler (1983, sec. 5).

4. ROBUSTNESS PROPERTIES OF MINIMUM DISCREPANCY ESTIMATORS

In this section we suppose that $\mathcal{P}$ is defined in the parametric manner of (2.2) and study asymptotic properties of the minimum discrepancy estimator $\hat{\boldsymbol{\theta}}$. For the sake of simplicity it will be assumed that $\boldsymbol{\theta}_0$ is an interior point of $\mathbb{S}$ and that the Jacobian matrix $\boldsymbol{\Delta}$ is of full column rank, q. Under mild regularity conditions, we have that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ and $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal (Shapiro 1983) with null mean vector and a covariance matrix that will be denoted by $\boldsymbol{\Pi}$.

It can be shown that

$$\boldsymbol{\Pi} \geq (\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-}\boldsymbol{\Delta})^{-1} \quad (4.1)$$

in the Loewner sense of inequality; that is, the difference between the left side and right side in (4.1) is nonnegative definite. Moreover, the lower bound given by the right side of (4.1) is attained for an appropriate choice of the discrepancy function F given a particular estimator $\mathbf{S}$ of Σ_0. Of course, one can improve this lower bound by starting with a better estimator $\mathbf{S}$ having a smaller asymptotic covariance matrix $\boldsymbol{\Gamma}$. We shall say that the estimator $\hat{\boldsymbol{\theta}}$ *is asymptotically efficient within the class of minimum discrepancy estimators based on* $\mathbf{S}$ if $\boldsymbol{\Pi}$ is equal to $(\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-}\boldsymbol{\Delta})^{-1}$. Note that the matrix $\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-}\boldsymbol{\Delta}$ is independent of a particular choice of the generalized inverse $\boldsymbol{\Gamma}^{-}$.

Necessary and sufficient conditions for asymptotic efficiency are given in Shapiro (1985b; 1986). Applied to our situation, these conditions immediately imply the following result.

Theorem 3. The estimator $\hat{\boldsymbol{\theta}}$ is asymptotically efficient within the class of minimum discrepancy estimators if and, when $\beta \neq 0$, only if $\boldsymbol{\sigma}_0 \in T_0$.

It is interesting to note that in the situation considered in the present article conditions for asymptotic chi-squaredness of $\hat{\alpha}^{-1}n\hat{F}$ and asymptotic efficiency of $\hat{\boldsymbol{\theta}}$ are the same. This property does not hold in general (compare Shapiro 1985b).

It follows from (3.1) that if $\boldsymbol{\sigma}_0 \in T_0$ and consequently $\boldsymbol{\sigma}_0 = \boldsymbol{\Delta\zeta}$ for some vector $\boldsymbol{\zeta}$, then

$$\boldsymbol{\Gamma} = \alpha\boldsymbol{\Gamma}_N + \beta\boldsymbol{\Delta\zeta\zeta}'\boldsymbol{\Delta}'.$$

It can then be shown (Shapiro 1985b) that

$$(\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-}\boldsymbol{\Delta})^{-1} = \alpha(\boldsymbol{\Delta}'\boldsymbol{\Gamma}_N^{-}\boldsymbol{\Delta})^{-1} + \beta\boldsymbol{\zeta\zeta}'.$$

The matrix

$$\Pi_N = (\Delta'\Gamma_N^{-}\Delta)^{-1}$$

is the asymptotic covariance matrix of $n^{1/2}(\hat{\theta} - \theta_0)$ when the underlying distribution is normal, so $\beta\zeta\zeta'$ gives a simple correction term of rank 1 to the normal theory asymptotic result for a distribution from the elliptical class.

Theorem 4. If $\sigma_0 \in T_0$, then the asymptotic covariance matrix Π of $n^{1/2}(\hat{\theta} - \theta_0)$ is given by

$$\Pi = \alpha\Pi_N + \beta\zeta\zeta'.$$

This expression is more direct than a corresponding expression given in Browne (1984, proposition 5) in that it provides a rank 1 correction to Π_N rather than to Π_N^{-1}.

The vector ζ may be obtained by solving the linear equations $\Delta\zeta = \sigma$ that are consistent when $\mathcal{G}$ is positively homogeneous. In some cases ζ is immediately available. For example, if the model is linear as in (2.3), it is clear that $\zeta = \theta_0$ and we obtain the following result.

Corollary 4. If the model is linear, the estimator $\hat{\theta}$ is asymptotically efficient within the class of minimum discrepancy estimators and the associated covariance matrix is given by

$$\Pi = \alpha\Pi_N + \beta\theta_0\theta_0'.$$

This result is applicable to models for patterned covariance matrices (Mukherjee 1970; Szatrowski 1985) and to component-of-covariance models [Wiley et al. 1973, models (1)–(4)]. All asymptotic variances and covariances are obtained simply from normal theory results.

There are situations where the parameter vector θ can be partitioned into two parts $\theta' = (\theta_1', \theta_2')$ such that the positive homogeneity of $\mathcal{G}$ is ensured by θ_2 alone. That is, for every $\theta' = (\theta_1', \theta_2') \in \mathbb{S}$ and $t > 0$ there exists a $\bar{\theta}' = (\bar{\theta}_1', \bar{\theta}_2') \in \mathbb{S}$ such that $t\mathbf{G}(\theta) = \mathbf{G}(\bar{\theta})$ and $\theta_1 = \bar{\theta}_1$. We shall then say that θ_1 *is free of a constant scaling factor.* In this situation there exists a vector ξ such that $\sigma_0 = \Delta_2\xi$, where Δ_2 is the Jacobian matrix corresponding to θ_2. This implies that the vector ζ is given by $\zeta' = (\mathbf{0}', \xi')$ and, consequently, the covariance matrix Π has the following relationship with the normal theory covariance matrix Π_N. It is assumed that Π and Π_N are partitioned conformably with θ.

Corollary 5. If θ_1 is free of a constant scaling factor, then

$$\Pi_{11} = \alpha\Pi_{11N}, \qquad \Pi_{12} = \alpha\Pi_{12N},$$

and

$$\Pi_{22} = \alpha\Pi_{22N} + \beta\xi\xi'.$$

Corollary 5 shows that when θ_1 is free of a constant scaling factor the submatrices Π_{11} and Π_{12} of the asymptotic covariance matrix Π may be obtained by simply scaling the corresponding submatrices of the normal theory asymptotic covariance matrix Π_N. In these situations adjustments to asymptotic standard errors of some, but not all, parameter estimates are simple.

We shall consider some examples. Let

$$\mathbf{G}(\theta) = \rho_1\mathbf{G}_1(\gamma) + \cdots + \rho_k\mathbf{G}_k(\gamma), \tag{4.2}$$

where the $\mathbf{G}_i(\gamma)$ are symmetric matrix valued functions of a vector γ. Here the parameter vector θ is given by $\theta' = (\gamma', \rho')$ with $\rho' = (\rho_1, \ldots, \rho_k)'$. Clearly, γ is free of a constant scaling factor, so the asymptotic covariance matrix associated with γ is simply α times the corresponding normal theory asymptotic covariance matrix. Situations of this type occur in scaled component-of-covariance models [Wiley et al. 1973, models (5)–(8)] when one of the scaling factors γ_i is fixed for identification purposes.

Now consider the expression of a correlation structure in the form of a covariance structure. Let

$$\mathbf{G}(\theta) = \mathbf{D}\mathbf{G}^*(\gamma)\mathbf{D}, \tag{4.3}$$

where $\mathbf{D}$ is a diagonal matrix representing standard deviations and $\mathbf{G}^*(\gamma)$ is a symmetric matrix valued function of γ such that $\text{diag}\{\mathbf{G}^*(\gamma)\} = \mathbf{I}$, which represents a correlation structure. Here θ consists of the vector γ and the vector of diagonal elements of $\mathbf{D}$. The parameter vector γ of the correlation structure is free of a constant scaling factor, so the corresponding asymptotic covariances are a simple scaling of the normal theory asymptotic covariances. Particular examples of models of this type are patterned correlation matrices (compare Szatrowski 1985, p. 640), where the elements of γ are correlation coefficients, and the factor analysis model treated as a correlation structure, where the elements of γ are standardized factor loadings (compare Lawley and Maxwell 1971, tables 5.3 and 7.10).

Corrections of a nature similar to those employed in correlation structures apply in scaled component-of-covariance models when an error variance is fixed for identification purposes (Bock and Bargmann 1966, p. 519, case III).

In the immediately preceding examples, the additive adjustment of Theorem 4 has been null for some of the parameter estimators. We now consider the example of the factor analysis covariance structure (2.4) where the additive adjustment for asymptotic variances of some parameter estimators is not null but is a simple function of some parameters. Here we have that $\mathbf{G}(\theta) = \mathbf{G}_1(\lambda) + \mathbf{G}_2(\psi)$, where $\lambda = \text{vec}(\Lambda)$, $\psi = \text{diag}(\Psi)$ and $\mathbf{G}_1(\lambda) = \Lambda\Lambda'$, $\mathbf{G}_2(\psi) = \Psi$. Accordingly, $\sigma_0 = \sigma_0^{(1)} + \sigma_0^{(2)}$, where $\sigma_0^{(1)} = \text{vec}(\Lambda_0\Lambda_0')$ and $\sigma_0^{(2)} = \text{vec}(\Psi_0)$. Since both $\mathbf{G}_1(\lambda)$ and $\mathbf{G}_2(\psi)$ represent models that are invariant under a constant scaling factor, we obtain that $\sigma_0^{(1)} = \Delta_1\zeta_1$ and $\sigma_0^{(2)} = \Delta_2\zeta_2$, where Δ_1 and Δ_2 are the Jacobian matrices associated with $\mathbf{G}_1$ and $\mathbf{G}_2$, respectively. Moreover, since $\mathbf{G}_2(\psi)$ is linear it follows that $\zeta_2 = \psi_0$. We then have that the asymptotic covariance matrix associated with the unique variance estimator $\hat{\psi}$ is simply the sum of α times the normal theory covariance matrix and the rank 1 matrix $\beta\psi_0\psi_0'$.

5. DISCUSSION

The results we have obtained have definite practical implications. Nearly all models employed in practice are invariant under a constant scaling factor. For example, all applications reported in Jöreskog (1981) and in Browne (1982, sec. 2.9) were of models with this property. Most computer programs for fitting structural models for covariance matrices involve an assumption of multivariate normality. The minimum discrepancy test statistic can then easily be adapted to the less restrictive assumption of a distribution from the elliptical class by dividing it by an estimate of α. If a test with a restricted alternative hypothesis is of interest, the ratio test statistic (3.4) may be employed in conjunction with an F distribution, as it is robust within the class of elliptical distributions and does not require an estimate of α. As we have mentioned, the possible loss of asymptotic power compared with the difference test statistic is negligible for large values of the second number $p(p + 1)/2 - d$ of degrees of freedom. This test would be particularly convenient if some estimate of $\boldsymbol{\Sigma}_0$ other than the sample covariance matrix is employed and an estimate of α is difficult to obtain. Frequently the correlation structure is of primary interest, and the standard errors of estimators appropriate for an elliptical distribution may be obtained from the normal theory standard errors output from a computer program simply by multiplying them by an estimate of $\alpha^{1/2}$.

If the sample covariance matrix is employed to estimate $\boldsymbol{\Sigma}_0$, as is usually the case, a suitable estimate of α may be obtained by dividing Mardia's (1970) coefficient of multivariate kurtosis by $p(p + 2)$ (Browne 1982, 1984). On the other hand, there is good reason to depart from standard practice and employ an alternative estimate of $\boldsymbol{\Sigma}_0$ that satisfies Tyler's requirement and yields a smaller value of α if a nonnormal distribution from the elliptical class is considered appropriate for one's data. Since the power of tests of fit of the model is related to α^{-1}, these tests will then have more power.

We have seen that the asymptotic variances of the parameter estimators depend not only on α but also on β. Tyler (1983, sec. 4) pointed out that the parameter β can be arbitrarily large for some elliptical distributions even if the appropriate maximum likelihood estimator of $\boldsymbol{\Sigma}_0$ is employed. Consequently, there could be situations where the fit of a model can be tested with reasonable power but parameter estimators are unacceptably imprecise.

APPENDIX: PROOF OF LEMMA 1

Consider the curve $\boldsymbol{\sigma}(t) = \boldsymbol{\sigma}_0 + t\boldsymbol{\sigma}_0$, starting from the point $\boldsymbol{\sigma}_0$, as t varies from 0 to a positive constant ε. Since $\mathcal{S}$ is positively homogeneous, we have that $\mathcal{S}$ contains $\boldsymbol{\sigma}(t)$ and, consequently, the tangent vector $\partial\boldsymbol{\sigma}(t)/\partial t|t = 0$ belongs to T_0. Clearly, this tangent vector is $\boldsymbol{\sigma}_0$ and thus $\boldsymbol{\sigma}_0 \in T_0$.

[*Received December 1985. Revised April 1987.*]

REFERENCES

Anderson, T. W., Fang, K. T., and Hsu, H. (1986), "Maximum Likelihood Estimates and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," *Canadian Journal of Statistics*, 14, 55–59.

Anderson, T. W., and Rubin, H. (1956), "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (vol. 5), ed. J. Neyman, Berkeley, CA: University of California Press, pp. 111–150.

Bock, R. D., and Bargmann, R. E. (1966), "Analysis of Covariance Structures," *Psychometrika*, 31, 211–252.

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

——— (1984), "Asymptotically Distribution-Free Methods in the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.

Chmielewski, M. A. (1980), "Invariant Scale Matrix Hypothesis Tests Under Elliptical Symmetry," *Journal of Multivariate Analysis*, 10, 343–350.

——— (1981), "Elliptically Symmetric Distributions: A Review and Bibliography," *International Statistical Review*, 49, 67–74.

Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1976), "Some Multivariate Applications of Elliptical Distributions," in *Essays in Probability and Statistics*, ed. S. Ikeda, Tokyo: Shinko Tsusho, pp. 365–395.

Hirsch, M. W. (1976), *Differential Topology*, New York: Springer-Verlag.

Jöreskog, K. G. (1967), "Some Contributions to Maximum Likelihood Factor Analysis," *Psychometrika*, 32, 443–482.

——— (1981), "Analysis of Covariance Structures" (with discussion), *Scandinavian Journal of Statistics*, 8, 65–92.

Lawley, D. N., and Maxwell, A. E. (1971), *Factor Analysis as a Statistical Method*, New York: American Elsevier.

Mardia, K. V. (1970), "Measures of Multivariate Skewness and Kurtosis With Applications," *Biometrika*, 57, 519–530.

Maronna, R. A. (1976), "Robust *M*-Estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 4, 51–67.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, New York: John Wiley.

Muirhead, R. J., and Waternaux, C. M. (1980), "Asymptotic Distributions in Canonical Correlation Analysis and Other Multivariate Procedures for Nonnormal Populations," *Biometrika*, 67, 31–43.

Mukherjee, B. N. (1970), "Likelihood Ratio Tests of Statistical Hypotheses Associated With Patterned Covariance Matrices in Psychology," *British Journal of Mathematical and Statistical Psychology*, 23, 89–120.

Shapiro, A. (1983), "Asymptotic Distribution Theory in the Analysis of Covariance Structures (A Unified Approach)," *South African Statistical Journal*, 17, 33–81.

——— (1985a), "Asymptotic Equivalence of Minimum Discrepancy Estimators to GLS Estimators," *South African Statistical Journal*, 19, 73–81.

——— (1985b), "Robustness Properties of the MDF Analysis of Moment Structures," Research Report 9/85, University of South Africa, Dept. of Mathematics, Pretoria, South Africa.

——— (1986), "Asymptotic Theory of Overparameterized Structural Models," *Journal of the American Statistical Association*, 81, 142–149.

Steiger, J. H., Shapiro, A., and Browne, M. W. (1985), "On the Multivariate Asymptotic Distribution of Sequential Chi-Square Statistics," *Psychometrika*, 50, 253–264.

Swain, A. J. (1975), "A Class of Factor Analysis Estimation Procedures With Common Asymptotic Properties," *Psychometrika*, 40, 315–335.

Szatrowski, T. H. (1985), "Patterned Covariances," in *Encyclopedia of Statistical Sciences* (vol. 6), eds. S. Kotz and N. L. Johnson, New York: John Wiley, pp. 638–641.

Tyler, D. E. (1982), "Radial Estimates and the Test for Sphericity," *Biometrika*, 69, 429–436.

——— (1983), "Robustness and Efficiency Properties of Scatter Matrices," *Biometrika*, 70, 411–420.

Wiley, D. E., Schmidt, W. H., and Bramble, W. J. (1973), "Studies of a Class of Covariance Structure Models," *Journal of the American Statistical Association*, 68, 317–323.