Asymptotic properties of statistical inference based on Fisher consistent estimators in the analysis of covariance structures

Yutaka Kano

Department of Mathematical Sciences, College of Engineering, University of Osaka Prefecture, Sakai, Osaka 591, Japan

Abstract

The methods of maximum likelihood (ML) and generalized least squares (GLS) under the normality assumption are often used for inference on covariance structures, and asymptotic properties and robustness of the statistical inference have been extensively studied. In this article, we generalize these results to inference based on Fisher consistent (FC) estimators which include simple least squares (LS) and noniterative estimation methods as well as ML and GLS. Although the LS and noniterative methods do not yield asymptotically efficient estimators under the normality, for small or moderate samples they are often superior to efficient estimators in mean squared error and less often result in so-called improper solutions. This shows that there do exist cases where such inefficient inference should be made rather than ML and GLS. Thus, the extension to be described here is important. Furthermore, a key relation shown from a property of the FC estimators makes the derivation of asymptotics of the inference very easy and comprehensive. The asymptotic efficiency of the MLE within the class of FC estimators is proved under a situation where the fourth-order moments of observations may not be finite.

1. INTRODUCTION

In a covariance structure model, the covariance matrix of a p-dimensional observation $\mathbf{x}$ is represented as

$$\mathrm{Cov}(\mathbf{x}, \mathbf{x}) = \Sigma(\theta),$$

where θ is a q-dimensional structural parameter vector to be estimated. Statistical inference in the analysis of covariance structures is usually made under the normality assumption on observations. That is, observations, $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}$, are assumed to come from a normal population $N(\mu, \Sigma(\theta))$, and then estimation of θ, testing hypotheses regarding θ, and testing goodness-of-fit are made by methods of the maximum likelihood, least squares or others. All softwares for analyzing covariance structures are formed on the basis of normal theory, e.g., LISREL by Jöreskog and Sörbom (1988) and EQS by Bentler (1989). On the other hand, many researchers have pointed out that the main assumption of the normality is questionable in areas of social sciences where these analyses are often utilized. Thus, it is very important to develop statistical theory under general distributions, e.g., ADF methods (Browne 1982, 1984; Bentler and Dijkstra

1985) and to investigate to what extent the normal theory inference remains valid, in other words, robustness to violation of the normality assumption.

In the 1980's, robustness theory has been developed very much. Muirhead and Waternaux (1980) studied a wide range of the likelihood ratio (LR) test statistics under elliptical populations, and showed that some of the LR statistics remain valid if one makes a correction with the multivariate kurtosis η of observations whereas others are not valid at all. Tyler (1983) provided an elegant condition in terms of null hypotheses under which the correction with η makes the LR statistics valid. Browne (1982, 1984) showed that in a covariance structure model, the η correction is possible if the covariance structure is invariant under a constant scaling factor. This result is slightly more general than Tyler's. Shapiro and Browne (1987) unified these results. Bentler (1983) developed a fit-function for models not meeting the invariance assumption. Shapiro (1986, 1987) provided general and useful conditions in terms of the fourth-order moments under which asymptotic efficiency of estimators and chi-squaredness of goodness-of-fit statistics hold and the standard errors of estimators are correct.

Another approach to investigating robustness is to assume independence among latent variates which generate the covariance structure. Amemiya (1985) first developed this situation and showed chi-squaredness of goodness-of-fit statistics in structural and functional relationships for a case when the fourth-order cumulants of error variates are zero, but no distributional assumption on the structural variates is made. In 1985, Amemiya and Anderson showed that the above condition can be weakened for a factor analysis model, in which only independence among latent variables is assumed, and remarkably, the fourth-order moments may not exist. This result is extremely interesting, but the proof was very complicated. A revised version of the paper was published by Amemiya and Anderson (1990). This result and a corresponding result on estimation due to Anderson and Amemiya (1988) for a factor analysis model were extended to those for a linear latent variate model by Anderson (1987, 1989).

On the other hand, Browne-Shapiro have developed similar results under the assumption that the fourth-order moments be finite (Browne 1987; Browne and Shapiro 1987, 1988). In this framework, one can derive asymptotic efficiency of estimators and obtain asymptotic distributions of estimators for variances of latent variables, which are out of the scope of Anderson-Amemiya. Mooijaart and Bentler (1985, 1991) took a close look at factor analysis models, and Satorra and Bentler (1990) extended Brown-Shapiro's to inference based on arbitrary discrepancy functions.

As seen above, the three frameworks, namely ellipticity, independence and independence with finite fourth-order moments, have been developed independently and appear to be essentially different. A notable feature of the paper is to separately discuss distributional assumptions and estimation methods, and we shall describe a key relation (see (2.4) that is derived from (A4)) on estimation methods and distributional assumptions (see (A2), (A3)) such that the robustness property holds. The previous results on the robustness, which look different from one another, are shown to be a special case of the general situation that satisfies the assumptions (A2)–(A4). Thus, the first contribution of this article is to unify the three frameworks described above.

The property (2.4) on estimation methods is derived from Fisher consistency (FC) which are met by almost all of the estimators in the literature. So, the results obtained here are an extension from minimum discrepancy function estimators to FC estimators. Furthermore, introduction of the notion of FC estimators simplifies the derivations of the asymptotic results very much.

Another fruit of the present paper is to establish the asymptotic efficiency of the

MLE even under the case where the fourth-order moments of observations may not be finite. The property derived from the FC estimation makes it possible to discuss the efficiency under such general cases.

In Section 2 we develop the asymptotic distribution of FC estimators under very general distributional assumptions. Sections 3 and 4 provide many examples that satisfy the assumptions on distributions and on estimation methods, respectively. Some specific developments of the paper are described in Section 5.

At the Trento workshop we noticed that Satorra (1989) had presented part of the results of Theorem 1 of this article at the Psychometric meeting. Our results are more general, though the essential idea would be the same. These results were developed independently, and we shall note how Satorra's result is covered with our derivation in Section 5.

2. MAIN RESULTS

Let $\Sigma(\theta)$ be a $p \times p$ matrix-valued function defined on a subset Θ of R^q such that $\Sigma(\theta)$ is positive definite for each $\theta \in \Theta$. We take θ_0, an interior point in Θ, as a true value of θ. We say that a (matrix-valued) function belongs to C^k-class in a domain G if the function is k times continuously differentiable in G, where k is any natural number.

(A1) $\Sigma(\theta)$ belongs to C^2-class in a neighborhood of θ_0, and the matrix of the derivatives,

$$\Delta = \left.\frac{\partial v(\Sigma(\theta))}{\partial \theta'}\right|_{\theta=\theta_0},$$

is of rank q.

Here $v(A)$ denotes a p^*-vector formed from the distinct elements of a $p \times p$ symmetric matrix A, where $p^* = p(p+1)/2$ (see Magnus and Neudecker 1988, page 49 (2)). We shall summarize the vec operator and duplication matrices which will be used in this paper. Define $vec(A)$ by a p^2-vector obtained by stacking all column vectors of A in order. The duplication matrix D_p of $p^2 \times p^*$ is defined as a linear operator such that $D_p v(A) = vec(A)$ for any symmetric matrix A, and then we have $D_p^+ vec(A) = v(A)$ with $D_p^+ = (D_p' D_p)^{-1} D_p'$.

Let $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}$ be independent random p-vectors that are not necessarily identically distributed, and define the sample covariance matrix S by

$$S = \frac{1}{n} \sum_{\alpha=1}^{N} (\mathbf{x}^{(\alpha)} - \bar{\mathbf{x}})(\mathbf{x}^{(\alpha)} - \bar{\mathbf{x}})',$$

where $n = N - 1$ and $\bar{\mathbf{x}}$ is the sample mean vector.

Make the following assumptions on the underlined distribution:

(A2) There exist a $p^* \times p^*$ positive definite matrix Γ, a $p \times p$ symmetric matrix D, and a possibly random sequence of θ_n converging to θ_0 in probability such that

$$\sqrt{n}(v(S) - v(\Sigma_n)) \xrightarrow{L} N_{p^*}(0, \Gamma)$$

with $\Sigma_n = \Sigma(\theta_n) + \frac{1}{\sqrt{n}}D$.

Under (A2) we have that $S \xrightarrow{P} \Sigma(\theta_0)$. This assumption is made so as to cover the following three important cases. The first case is a typical situation when we take $\theta_n = \theta_0$ and $D = 0$. Usually Γ is the covariance matrix of $v((\mathbf{x}-\mu)(\mathbf{x}-\mu)')$ with $\mu = \mathrm{E}(\mathbf{x})$. In particular, if the observations are independently and identically distributed according to the multivariate normal distribution, it is known that

$$\Gamma = 2D_p^+(\Sigma(\theta_0) \otimes \Sigma(\theta_0))D_p^{+\prime} \ (= \Gamma_N, \text{ say}). \tag{2.1}$$

The second one is where the data come from distributions under a sequence of local alternatives. This set-up is employed to investigate the local power of test statistics. For testing goodness-of-fit of a given covariance structure, a sequence of the true population covariance matrices, $\Sigma_n = \Sigma(\theta_0) + \frac{1}{\sqrt{n}}D$, under the alternative hypothesis is employed, where we have taken $\theta_n = \theta_0$. The third is a quite interesting case developed by Anderson-Amemiya in which θ_n is random. This case will be described in the next section.

Another assumption is made.

(A3) There exist a scalar $\eta > 0$ and a $q \times q$ symmetric matrix G such that $\Gamma = \eta\Gamma_N + \Delta G \Delta'$.

We shall describe what kinds of distributions meet the above assumptions in the next section. Notice that under a case of $\theta_n = \theta_0$, the normality assumption implies that $\eta = 1$ and $G = 0$. Thus, $\kappa\,(= \eta - 1)$ and G could be regarded as the degree of the nonnormality or the departure from the normal distribution.

Now we shall describe a class of estimators that will be dealt with in this article. Let $\widehat{\theta}_\xi = \widehat{\theta}_\xi(S)$ be a $q \times 1$ vector-valued function defined on the set of $p \times p$ positive definite matrices, where ξ is possibly random and converges to $\bar{\xi}$ in probability. An example of ξ is a random weight matrix in the method of weighted least squares. When ξ is nonrandom or unrelated to the sample, as in unweighted least squares (see Section 4), all assumptions about ξ which follow should be ignored.

Assume that

(A4) $\widehat{\theta}_\xi(S)$ satisfies

$$\widehat{\theta}_\xi(\Sigma(\theta)) = \theta \tag{2.2}$$

in a neighborhood of θ_0, and belongs to C^1-class in a neighborhood of $\Sigma(\theta_0)$. The derivative $\frac{\partial \widehat{\theta}_\xi(S)}{\partial v(S)'}$ is continuous in a neighborhood of $(\Sigma(\theta_0), \bar{\xi})$.

From the requirement of (2.2), the estimator $\widehat{\theta}_\xi(S)$ will be called Fisher consistent (FC) (see, e.g., Rao 1973, page 345). Almost all the estimators suggested in the literature are FC as will be noted in Section 4.

Differentiating both sides of (2.2) in terms of θ and evaluating it at $\theta = \theta_0$, we get

$$\left.\frac{\partial \widehat{\theta}_\xi(S)}{\partial v(S)'}\frac{\partial v(\Sigma(\theta))}{\partial \theta'}\right|_{\theta=\theta_0} = \left.\frac{\partial \widehat{\theta}_\xi(S)}{\partial v(S)'}\right|_{S=\Sigma(\theta_0)} \Delta = I_q.$$

From this consequence and the continuity assumption in (A4), it follows that

$$\lim_{\substack{\xi \to \bar{\xi} \\ S \to \Sigma(\theta_0)}} \frac{\partial \widehat{\theta}_\xi(S)}{\partial v(S)'} = \left.\frac{\partial \widehat{\theta}_{\bar{\xi}}(S)}{\partial v(S)'}\right|_{S=\Sigma(\theta_0)} \left(= \frac{\partial \widehat{\theta}}{\partial \sigma_0'}\ , \quad \text{say} \right) \tag{2.3}$$

and

$$\frac{\partial \widehat{\theta}}{\partial \sigma_0'}\Delta = \lim_{\xi \to \bar{\xi}} \left(\left.\frac{\partial \widehat{\theta}_\xi(S)}{\partial v(S)'}\right|_{S=\Sigma(\theta_0)} \Delta \right) = I_q, \tag{2.4}$$

which is a key relation of this paper.

Notice that the first-order derivatives, $\frac{\partial \widehat{\theta}}{\partial \sigma_0'}$, depend only on estimation methods employed. The relation (2.4) is easily verified for particular estimators. For instance, the MLE under the normality has the derivatives

$$\frac{\partial \widehat{\theta}_{MLE}}{\partial \sigma_0'} = (\Delta' \Gamma_N^{-1} \Delta)^{-1} \Delta' \Gamma_N^{-1}, \tag{2.5}$$

from which (2.4) follows.

The properties in (2.3) and (2.4) greatly simplify the proof of the following theorem which is an extension of the previous results on the asymptotic robustness in a covariance structure model.

Theorem 1. Assume (A1) and (A2). Let $\widehat{\theta}_\xi$ be Fisher consistent, i.e., $\widehat{\theta}_\xi$ satisfies (A4). Then we have

$$\sqrt{n}(\widehat{\theta}_\xi - \theta_n) \xrightarrow{L} N_q \left(\frac{\partial \widehat{\theta}}{\partial \sigma_0'} v(D),\ \frac{\partial \widehat{\theta}}{\partial \sigma_0'} \Gamma \frac{\partial \widehat{\theta}}{\partial \sigma_0'}' \right), \tag{2.6}$$

where $\frac{\partial \widehat{\theta}}{\partial \sigma_0'}$ is defined in (2.3). When (A3) is further assumed, the asymptotic covariance matrix becomes

$$\eta \frac{\partial \widehat{\theta}}{\partial \sigma_0'} \Gamma_N \frac{\partial \widehat{\theta}}{\partial \sigma_0'}' + G \tag{2.7}$$

with Γ_N in (2.1).

Proof. Noting that $\widehat{\theta}(\Sigma(\theta_n)) = \theta_n$ and $\sqrt{n}\,\{v(\Sigma_n) - v(\Sigma(\theta_n))\} = v(D)$, we have from (2.3), (2.4) and (A2) that

$$\begin{aligned}
\sqrt{n}(\widehat{\theta}_\xi - \theta_n) &= \left.\frac{\partial \widehat{\theta}_\xi(S)}{\partial v(S)'}\right|_{S=\Sigma^*} \sqrt{n}\,\{v(S) - v(\Sigma(\theta_n))\} \\
&= \frac{\partial \widehat{\theta}}{\partial \sigma_0'}\sqrt{n}\,\{v(S) - v(\Sigma_n)\} + \frac{\partial \widehat{\theta}}{\partial \sigma_0'}\sqrt{n}\,\{v(\Sigma_n) - v(\Sigma(\theta_n))\} + o_p(1) \\
&\xrightarrow{L} N_q\left(\frac{\partial \widehat{\theta}}{\partial \sigma_0'}v(D),\ \frac{\partial \widehat{\theta}}{\partial \sigma_0'}\Gamma\frac{\partial \widehat{\theta}}{\partial \sigma_0'}'\right),
\end{aligned}$$

where Σ^* lies between S and $\Sigma(\theta_n)$ and converges to $\Sigma(\theta_0)$ in probability. Under assumption (A3), use of (2.4) shows (2.7). *Q.E.D.*

From now on we suppress ξ and $\bar{\xi}$ for simplicity.

When a sequence of local alternatives is considered, the corresponding result for MDF estimators (see Section 4) was obtained by Bentler and Dijkstra (1985) and Shapiro and Browne (1989).

For the MLE, substitution of (2.5) into (2.7) in Theorem 1 shows

$$\eta(\Delta'\Gamma_N^{-1}\Delta)^{-1} + G. \tag{2.8}$$

This result has been obtained by many authors. In some situations, $\eta = 1$ and some elements of G are equal to zero. No correction of the normal theory results is then needed for estimators of the parameters with the corresponding elements of G zero even under the nonnormality. This has been pointed out by many researchers as well. The result (2.8) is a fruit that was obtained by the recent robustness study. As seen in Theorem 1 and its proof, (A3) and (A4) play an essential role in showing (2.8).

The result in (2.7) enjoys some interesting features. As pointed out before, only η and G reflect the nonnormality, whereas $\frac{\partial \widehat{\theta}}{\partial \sigma_0'}$ depends only on the estimation method. This suggests that the nonnormality and estimation methods independently affect asymptotic distributions of $\widehat{\theta}$. In general, the asymptotic covariance matrix of estimators based on the normality assumption is not true any more for nonnormal distributions. But under the assumption (A3), we can simply correct the normal theory covariance matrix with η and G based on (2.7). One consequence of (2.7) is that one can make a simple correction of the normal theory formula with η and G not only for the MLE but for any other FC estimators as well and that the correction factors η and G in (2.7) is common to all FC estimators. For instance, the η and G needed to correct for the MLE is exactly the same as those to correct for the least squares estimator or noniterative estimators.

For another thing, we may just observe the quantity $\frac{\partial \widehat{\theta}}{\partial \sigma_0'}\Gamma_N\frac{\partial \widehat{\theta}}{\partial \sigma_0'}'$ when studying effect of estimation methods. For instance, for a given underlined distribution, i.e., given η and G, an estimator $\widehat{\theta}$ minimizing $\frac{\partial \widehat{\theta}}{\partial \sigma_0'}\Gamma_N\frac{\partial \widehat{\theta}}{\partial \sigma_0'}'$ minimizes the asymptotic covariance matrix $\frac{\partial \widehat{\theta}}{\partial \sigma_0'}\Gamma\frac{\partial \widehat{\theta}}{\partial \sigma_0'}'$ of $\widehat{\theta}$. As a consequence, asymptotically efficient estimators under the normality assumption are still asymptotically efficient under nonnormal distributions meeting (A2) and (A3).

Now, let us investigate efficiency of FC estimators in more details. Solving the equation in (2.4) for $\frac{\partial\widehat{\theta}}{\partial\sigma_0'}$, we have

$$\frac{\partial\widehat{\theta}}{\partial\sigma_0'} = (\Delta'\Gamma^{-1}\Delta)^{-1}\Delta'\Gamma^{-1} + \Delta^c, \tag{2.9}$$

where Δ^c is a $q \times p^*$ matrix satisfying $\Delta^c\Delta = 0$, and the first term of the right side in (2.9) is a particular solution of the equation (2.4). Using (2.9), the asymptotic covariance matrix in (2.6) has the following lower bound:

$$\begin{aligned}\frac{\partial\widehat{\theta}}{\partial\sigma_0'}\Gamma\frac{\partial\widehat{\theta}}{\partial\sigma_0'}' &= (\Delta'\Gamma^{-1}\Delta)^{-1} + \Delta^c\Gamma(\Delta^c)' \\ &\geq (\Delta'\Gamma^{-1}\Delta)^{-1}.\end{aligned} \tag{2.10}$$

The derivation has been very often used to explore a lower bound of the variances of estimators for structural models, e.g., de Leeuw (1986, page 128) and Dijkstra (1984). It should be noted that the use of the derivation enables us to establish the efficiency property when the fourth-order moments may not be finite.

First, we consider a case when θ_n is nonrandom. In this case, Γ is unrelated to θ_n and the next definition is well-defined. An FC estimator of θ_0 will be called asymptotically efficient within the class of all FC estimators, if the asymptotic covariance matrix of the estimator attains its minimum $(\Delta'\Gamma^{-1}\Delta)^{-1}$.

Assuming (A3), the lower bound in (2.10) becomes $\eta(\Delta'\Gamma_N^{-1}\Delta)^{-1} + G$, which is identical with (2.8). Thus, the MLE is asymptotically efficient under (A1)–(A4) and the assumption that θ_n be nonrandom.

When θ_n is random, the situation is rather complicated because the lower bound may depend on the random sequence θ_n. Here we only consider a case when $\theta_n = [\theta_{1n}', \theta_{2n}']'$ with θ_{1n} $(r \times 1)$ nonrandom and θ_{2n} random, and investigate efficiency of $\widehat{\theta}_1$. Let A_{11} and A^{11} be $r \times r$ principal submatrices of A and A^{-1}, respectively. Under assumptions (A2) and (A3), the asymptotic covariance matrix of $\widehat{\theta}_1$ with $\widehat{\theta} = [\widehat{\theta}_1', \widehat{\theta}_2']'$ is expressed from (2.7) as

$$\eta\left(\frac{\partial\widehat{\theta}}{\partial\sigma_0'}\Gamma_N\frac{\partial\widehat{\theta}}{\partial\sigma_0'}'\right)_{11} + G_{11}. \tag{2.11}$$

Notice that for a given population distribution, (2.11) should not depend on the particular sequence of θ_{2n} since neither should $\sqrt{n}(\widehat{\theta}_1 - \theta_{1n})$, and this is true for any FC estimators. Thus we conclude that η and G_{11} do not depend on the sequence θ_{2n}.

Under (A2) and (A3), the lower bound in (2.10) becomes

$$\eta(\Delta'\Gamma_N^{-1}\Delta)^{11} + G_{11}, \tag{2.12}$$

which is unrelated to the particular choice of θ_{2n} since so are η and G_{11}. Thus, there is no ambiguity when an FC estimator whose asymptotic covariance is given by (2.12) is called asymptotically efficient within the class of all FC estimators. Again, the lower bound in (2.12) is attained when $\widehat{\theta}$ is the MLE.

Here we shall summarize the results on efficiency.

Theorem 2. Assume (A1) and (A2). Let $\widehat{\theta}$ be Fisher consistent, i.e., $\widehat{\theta}$ satisfies (A4). Then, (i) when θ_n is nonrandom, the lower bound of the asymptotic covariance matrix of $\widehat{\theta}$ within the class of FC estimators is given by (2.10); (ii) when (A3) is met and θ_{1n} with $\theta_n = [\theta'_{1n}, \theta'_{2n}]'$ is nonrandom, the lower bound of the asymptotic covariance matrix of $\widehat{\theta}_1$ is given by (2.12), where $\widehat{\theta} = [\widehat{\theta}'_1, \widehat{\theta}'_2]'$. For either case, the MLE ($\widehat{\theta}$ or $\widehat{\theta}_1$) is asymptotically efficient within the class of FC estimators, provided that (A3) is met.

Notice that the theorem covers the case treated by Anderson-Amemiya where the fourth-order moments of observations may not be finite (see Sections 3 and 4). Theorem 2 states that the MLE, for whole or part of θ with the corresponding θ_n nonrandom, is asymptotically efficient as long as (A1)–(A4) are met. This is also true for FC estimators of which the derivatives are equal to (2.5). An example of such estimators is the GLSE with a weight matrix $\widehat{\Sigma}(\theta_0)$ (see (4.2)).

3. ON DISTRIBUTIONAL ASSUMPTIONS

In this section, we shall describe underlined distributions meeting assumptions (A2) and (A3). Basically, the results stated here are due to Anderson (1987, 1989), Browne and Shapiro (1988), Kano (1990b), and Shapiro and Browne (1987). Throughout this section, we suppress the subscript $_0$, denoting the true value of parameters, for simplicity.

The first distribution introduced here is the elliptical distribution (see, e.g., Muirhead 1982). When the observations $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}$ are independent and identically distributed according to the elliptical distribution, the covariance matrix of $v(S)$ is represented as

$$n \cdot \mathrm{Cov}(v(S), v(S)) = \eta\Gamma_N + (\eta - 1)v(\Sigma(\theta))v(\Sigma(\theta))' + o(1),$$

where η is the rescaled multivariate kurtosis parameter (Mardia 1970). Assuming invariance under a constant scaling factor, we have

$$v(\Sigma(\theta)) = \Delta c$$

for some c. See Shapiro and Browne (1987). Thus, it is seen that the elliptical population meets our requirements (A2) and (A3) with $\theta_n = \theta, G = (\eta - 1)cc'$ and $D = 0$. Here this model will be denoted by (I). Some generalizations of elliptical distribution theory were given by Kano, Berkane, and Bentler (1990, in press).

The next example is a general linear latent variate model which was originally defined by Browne and Shapiro (1988) and slightly modified by Anderson (1987, 1989). The model is described as

$$\mathbf{x}^{(\alpha)} = \mu + \sum_{g=0}^{G+1} \Lambda_g(\lambda)\mathbf{z}_g^{(\alpha)} \qquad (\alpha = 1, \cdots, N),$$

where μ is the general mean, $\Lambda_g(\lambda)$ is of $p \times k_g$ with $r \times 1$ parameter vector λ, and where $\mathbf{z}_g^{(\alpha)}$ $(g = 0, \cdots, G+1)$ are latent variates of $k_g \times 1$ such that $\mathbf{z}_0^{(\alpha)}$ is a fixed-effect parameter vector satisfying

$$\bar{\mathbf{z}}_0 = \frac{1}{N}\sum_{\alpha=1}^{N} \mathbf{z}_0^{(\alpha)} \longrightarrow \mu_z$$

$$\Phi_{00}(n) = \frac{1}{n}\sum_{\alpha=1}^{N} (\mathbf{z}_0^{(\alpha)} - \bar{\mathbf{z}}_0)(\mathbf{z}_0^{(\alpha)} - \bar{\mathbf{z}}_0)' \longrightarrow \Phi_0$$

as $N \to \infty$ for some μ_z and Φ_0 and that $\mathbf{z}_1^{(\alpha)}, \cdots, \mathbf{z}_{G+1}^{(\alpha)}$ are random-effect variates with means zero satisfying

$$\begin{aligned} \mathrm{Cov}(\mathbf{z}_g^{(\alpha)}, \mathbf{z}_h^{(\alpha)}) &= 0 && \text{for } g \neq h \\ \mathrm{Cov}(\mathbf{z}_g^{(\alpha)}, \mathbf{z}_g^{(\alpha)}) &= \Phi_g && \text{for } g = 1, \cdots, G \\ \mathrm{Cov}(\mathbf{z}_{G+1}^{(\alpha)}, \mathbf{z}_{G+1}^{(\alpha)}) &= \Phi_{G+1}(\tau). \end{aligned} \tag{3.1}$$

Then it is seen that

$$S \xrightarrow{P} \lim_{N\to\infty} \mathrm{E}(S) = \sum_{g=0}^{G} \Lambda_g(\lambda)\Phi_g\Lambda_g(\lambda)' + \Lambda_{G+1}(\lambda)\Phi_{G+1}(\tau)\Lambda_{G+1}(\lambda)' \;\; (= \Sigma(\theta), \text{ say})$$

with $\theta = [\lambda', \tau', v(\Phi_0)', \cdots, v(\Phi_G)']'$.

Browne (1987) and Browne and Shapiro (1988) showed that when $\mathbf{z}_g^{(\alpha)}$ $(g = 1, \cdots, G+1)$ are independent and have finite fourth-order moments, and the fourth-order cumulants of $\mathbf{z}_{G+1}^{(\alpha)}$ are equal to zero, then

$$\sqrt{n}\{v(S) - v(\Sigma(\theta))\} \xrightarrow{L} N_{p^*}(0, \Gamma_N + \Delta G \Delta'),$$

where

$$G = \begin{bmatrix} 0 & & & & & \\ & 0 & & & & \\ & & -2D_{k_0}^{+}(\Phi_0 \otimes \Phi_0)D_{k_0}^{+\prime} & & & \\ & & & C_1^* & & \\ & & & & \ddots & \\ & & & & & C_G^* \end{bmatrix} \begin{array}{l} \} \;\; \lambda \\ \} \;\; \tau \\ \} \;\; \Phi_0 \\ \} \;\; \Phi_1 \\ \quad\;\; \vdots \\ \} \;\; \Phi_G \end{array}$$

with $C_g^* = D_{k_g}^{+} C_g D_{k_g}^{+\prime}$ and C_g being the fourth-order cumulant matrix of $\mathbf{z}_g^{(\alpha)}$. This example will be labeled as (II).

The next model, denoted by (II)$'$, is a slight generalization due to Kano (1990b). Let $\mathbf{y}_g^{(\alpha)}$ $(g = 1, \cdots, G+1)$ be latent variates satisfying the above requirements on $\mathbf{z}_g^{(\alpha)}$ (i.e., (3.1), independence, finite fourth-order moments, and zero fourth-order cumulants of $\mathbf{z}_{G+1}^{(\alpha)}$), and define

$$\mathbf{z}_g^{(\alpha)} = \mathbf{y}_g^{(\alpha)}/\epsilon^{(\alpha)} \qquad (g = 1, \cdots, G+1),$$

where $\epsilon^{(\alpha)}$ is a random variable independent of $\mathbf{y}_g^{(\alpha)}$'s such that $\mathrm{E}\left\{\left(\epsilon^{(\alpha)}\right)^{-2}\right\} = 1$ and $\mathrm{E}\left\{\left(\epsilon^{(\alpha)}\right)^{-4}\right\} = \eta$. Form a linear latent variate model

$$\mathbf{x}^{(\alpha)} = \mu + \sum_{g=1}^{G+1} \Lambda_g(\lambda)\mathbf{z}_g^{(\alpha)} \qquad (\alpha = 1, \cdots, N).$$

Here, $\mathbf{z}_g^{(\alpha)}$'s are permitted to be dependent through $\epsilon^{(\alpha)}$ (but uncorrelated) and the fixed-effect latent variate $\mathbf{z}_0^{(\alpha)}$ is not included in the model. Notice that this model includes elliptical distributions. Assume further that $\Phi_{G+1}(\tau)$ is invariant under a constant scaling factor, and then we get

$$v(\Phi_{G+1}(\tau)) = \frac{\partial v(\Phi_{G+1}(\tau))}{\partial \tau'} c$$

for some c. Under this model,

$$\sqrt{n}\{v(S) - v(\Sigma(\theta))\} \xrightarrow{L} N_{p^*}(0, \eta\Gamma_N + \Delta G \Delta'),$$

where $G = \eta G_1 + (\eta - 1)G_2$ and

$$G_1 = \begin{array}{c} \left[\begin{array}{ccccc} 0 & & & & \\ & 0 & & & \\ & & C_1^* & & \\ & & & \ddots & \\ & & & & C_G^* \end{array}\right] \end{array} \begin{array}{l} \} \; \lambda \\ \} \; \tau \\ \} \; \Phi_1 \\ \quad \vdots \\ \} \; \Phi_G \end{array}$$

and

$$G_2 = \left[\begin{array}{c} 0 \\ c \\ v(\Phi_1) \\ \vdots \\ v(\Phi_G) \end{array}\right] \left[\begin{array}{c} 0 \\ c \\ v(\Phi_1) \\ \vdots \\ v(\Phi_G) \end{array}\right]' \begin{array}{l} \} \; \lambda \\ \} \; \tau \\ \} \; \Phi_1 \\ \quad \vdots \\ \} \; \Phi_G \end{array} .$$

Here C_g^* is the matrix of the fourth-order cumulants of $\mathbf{y}_g^{(\alpha)}$ defined in the same manner as in the model (II).

In the model (II), Anderson (1987, 1989) relaxed the condition for the fourth-order moments of $\mathbf{z}_g^{(\alpha)}$ $(g = 1, \cdots, G)$ to be finite, in other words, he only assumed that $\mathbf{z}_g^{(\alpha)}$ $(g = 1, \cdots, G+1)$ are independent and the fourth-order cumulants of $\mathbf{z}_{G+1}^{(\alpha)}$ are equal to zero. This model will be labeled as (III). Let

$$\Phi_{gh}(n) = \frac{1}{n}\sum_{\alpha=1}^{N}(\mathbf{z}_g^{(\alpha)} - \bar{\mathbf{z}}_g)(\mathbf{z}_h^{(\alpha)} - \bar{\mathbf{z}}_h)' \quad (g, h = 0, 1, \cdots, G+1),$$

and define

$$\theta_n = [\lambda', \tau', v(\Phi_{00}(n))', \cdots, v(\Phi_{GG}(n))']'.$$

Notice that θ_n is random. Then we can show that

$$\sqrt{n}\{v(S) - v(\Sigma(\theta_n))\} \xrightarrow{L} N_{p^*}(0, \Gamma_N + \Delta G \Delta'),$$

where

$$G = \begin{array}{c} \left[\begin{array}{ccccc} 0 & & & & \\ & 0 & & & \\ & & -2D_{k_0}^{+}(\Phi_0 \otimes \Phi_0)D_{k_0}^{+\prime} & & \\ & & & \ddots & \\ & & & & -2D_{k_G}^{+}(\Phi_G \otimes \Phi_G)D_{k_G}^{+\prime} \end{array}\right] \end{array} \begin{array}{l} \} \quad \lambda \\ \} \quad \tau \\ \} \quad \Phi_0 \\ \quad\;\; \vdots \\ \} \quad \Phi_G \end{array} . \qquad (3.2)$$

A variant of this model similar to (II)$'$ can be defined in the same manner.

4. FISHER CONSISTENCY

In this section, we shall show that almost all the estimators suggested in the literature are Fisher consistent (FC).

In the analysis of covariance structures, methods of the maximum likelihood (ML) (Lawley 1940) and the generalized least squares (GLS) (Browne 1974) are typical and most frequently used. The estimators, MLE and GLSE, are obtained, respectively, by

$$\min_{\theta \in \Theta} \{\log|\Sigma(\theta)| - \log|S| + \mathrm{tr}(\Sigma(\theta)^{-1}S) - p\} \qquad (4.1)$$

and

$$\min_{\theta \in \Theta} \frac{1}{2}\mathrm{tr}[\{(S - \Sigma(\theta))V^{-1}\}^2], \qquad (4.2)$$

where S is the sample covariance matrix and V is a $p \times p$ possibly random weight matrix converging in probability to some positive definite matrix $\bar{V}$. Jöreskog and Goldberger (1972) took $V = S$ to result in an estimator that is called the weighted least squares estimator (WLSE), whereas $V = I_p$ and $V = \mathrm{Diag}(S)$ are used for the unweighted and simple least squares estimators (ULSE, SLSE)(see, e.g., Krane and McDonald 1978). Browne (1982, 1984) and Shapiro (1985) treated the extended generalized least squares estimator obtained via

$$\min_{\theta\in\Theta}\{v(S)-v(\Sigma(\theta))\}'W^{-1}\{v(S)-v(\Sigma(\theta))\} \tag{4.3}$$

with W a $p^*\times p^*$ weight matrix converging to $\bar{W}$. Browne (1982) called these functions to be minimized a discrepancy function and estimators formed through a discrepancy function a minimum discrepancy function (MDF) estimator. Discrepancy functions may depend on random weight matrices such as V and W in (4.2) and (4.3). Here we write them as ξ, and assume that ξ converges to $\bar{\xi}$ in probability. A general discrepancy function will be denoted by $F_\xi(S,\Sigma)$, which must satisfy

$$\begin{array}{rl} \text{i)} & F_\xi(S,\Sigma)\geq 0,\\ \text{ii)} & F_\xi(S,\Sigma)=0 \text{ if and only if } S=\Sigma,\\ \text{iii)} & F_\xi(S,\Sigma) \text{ is continuous in } S \text{ and } \Sigma. \end{array} \tag{4.4}$$

To obtain asymptotic distributions of MDF estimators, we assume

(A5-1) $F_\xi(S,\Sigma)$ belongs to C^2-class in a neighborhood of $(\Sigma(\theta_0),\Sigma(\theta_0))$.

Under (4.4) and (A5-1), Shapiro (1985) showed that

$$F_\xi(S,\Sigma)=\{v(S)-v(\Sigma)\}'U\{v(S)-v(\Sigma)\}+o(\|S-\Sigma\|^2)$$

in a neighborhood of $(\Sigma(\theta_0),\Sigma(\theta_0))$ for some nonnegative definite matrix U. As a consequence,

$$\left.\frac{\partial}{\partial v(\Sigma)}F_\xi(S,\Sigma)\right|_{S=\Sigma=\Sigma(\theta_0)}=0$$

and

$$\begin{aligned} U&=\left.\frac{\partial^2}{\partial v(\Sigma)\partial v(\Sigma)'}F_\xi(S,\Sigma)\right|_{S=\Sigma=\Sigma(\theta_0)}\\ &=-\left.\frac{\partial^2}{\partial v(S)\partial v(\Sigma)'}F_\xi(S,\Sigma)\right|_{S=\Sigma=\Sigma(\theta_0)}. \end{aligned}$$

Now we further assume

(A5-2) $\frac{\partial}{\partial v(\Sigma)}F_\xi(S,\Sigma)$ belongs to C^1-class in a neighborhood of $(\Sigma(\theta_0),\Sigma(\theta_0),\bar{\xi})$, and the two equal matrices

$$\left.\frac{\partial^2}{\partial v(\Sigma)\partial v(\Sigma)'}F_{\bar{\xi}}(S,\Sigma)\right|_{S=\Sigma=\Sigma(\theta_0)} \quad\text{and}\quad -\left.\frac{\partial^2}{\partial v(S)\partial v(\Sigma)'}F_{\bar{\xi}}(S,\Sigma)\right|_{S=\Sigma=\Sigma(\theta_0)}$$

are positive definite. We write the matrices as $\bar{H}^{-1}$.

These assumptions (A5-1) and (A5-2) are regularity conditions which are satisfied by most discrepancy functions, of course, including (4.1)–(4.3) with $\bar{H}=2D_p^+(\Sigma(\theta_0)\otimes$

$\Sigma(\theta_0))D_p^{+\prime}$, $2D_p^+(\bar{V}\otimes\bar{V})D_p^{+\prime}$, and $\bar{W}$, respectively. We simply write (A5-1) and (A5-2) as (A5).

Shapiro (1983) and Kano (1983, 1986) have shown that MDF estimators are consistent under mild regularity conditions including

(A6) θ_0 is strongly identifiable, i.e., for any $\epsilon > 0$, there exists $\delta > 0$ such that if $\|\Sigma(\theta) - \Sigma(\theta_0)\| < \delta$ and $\theta \in \Theta$, then $\|\theta - \theta_0\| < \epsilon$.

Thus, for large sample sizes, the estimator $\widehat{\theta}$ is a solution of

$$\frac{\partial}{\partial\theta}F_\xi(S,\,\Sigma(\theta)) = 0,$$

or

$$\Delta(\theta)'\frac{\partial}{\partial v(\Sigma)}F_\xi(S,\,\Sigma(\theta)) = 0, \tag{4.5}$$

where $\Delta(\theta) = \frac{\partial v(\Sigma(\theta))}{\partial\theta'}$. Under (A1) and (A5), the left side in (4.5) is a C^1-class function of $(S,\,\xi,\,\theta)$ in a neighborhood of $(\Sigma(\theta_0),\,\bar{\xi},\,\theta_0)$, and takes zero at $(\Sigma(\theta_0),\,\bar{\xi},\,\theta_0)$. Furthermore, from (A5-2) and (4.5), the Jacobian matrix of its derivatives in terms of θ evaluated at $(\Sigma(\theta_0),\,\bar{\xi},\,\theta_0)$ is $\Delta'\bar{H}^{-1}\Delta$, which is nonsingular. Thus, the implicit function theorem shows that the solution $\widehat{\theta}$ of (4.5) belongs to C^1-class in a neighborhood of $(\Sigma(\theta_0),\,\bar{\xi})$, and that the derivatives of $\widehat{\theta}$ are given as

$$\frac{\partial\widehat{\theta}}{\partial\sigma_0'} = (\Delta'\bar{H}^{-1}\Delta)^{-1}\Delta'\bar{H}^{-1}. \tag{4.6}$$

Fisher consistency of $\widehat{\theta}$ follows from the identifiability in (A.6) and the property ii) in (4.4). Thus, the MDF estimators meeting (A1) and (A5) satisfy (A4).

Let θ be decomposed into $[\theta_1',\,\theta_2']'$, and assume that an FC estimator $\widehat{\theta}_2$ is given. Let an estimator $\widehat{\theta}_1$ be determined by a solution to

$$\min_{\theta_1} F(S,\,\Sigma(\theta_1,\widehat{\theta}_2)). \tag{4.7}$$

This estimation method is also shown to lead to an FC estimator $\widehat{\theta} = [\widehat{\theta}_1',\,\widehat{\theta}_2']'$ under the same assumptions. A useful formula of the asymptotic covariance matrix of such estimators was given by Parke (1986).

An alternative approach to estimation is to use an inverse function $g(\Sigma)$ of $\Sigma = \Sigma(\theta)$, which may not, of course, be unique. Under (A1), the inverse function, defined on the set of $p\times p$ positive definite matrices, belongs to C^1-class and satisfies $g(\Sigma) = \theta$ on the image of $\Sigma(\theta)$, i.e., $g(\Sigma(\theta)) = \theta$. Thus, the estimator $g(S)$ of θ is obviously FC.

5. SPECIFIC CONSEQUENCES

Here we shall describe more specifically what we have developed in this article. To make derivations more comprehensive, we take a factor analysis model. But the results will be easily extended to a general linear latent variate model.

The factor analysis model is defined as

$$\mathbf{x}^{(\alpha)} = \mu + \Lambda(\lambda)\mathbf{f}^{(\alpha)} + \mathbf{u}^{(\alpha)},$$

where $\Lambda(\lambda)$ is a $p \times k$ matrix of factor loadings ($k < p$) and where $\mathbf{f}^{(\alpha)}$ and $\mathbf{u}^{(\alpha)}$ are latent random k- and p-vectors of the common and unique factors such that

$$\mathrm{E}(\mathbf{f}^{(\alpha)}) = 0, \quad \mathrm{E}(\mathbf{u}^{(\alpha)}) = 0$$
$$\mathrm{Cov}(\mathbf{f}^{(\alpha)}, \mathbf{f}^{(\alpha)}) = \Phi, \ \ \mathrm{Cov}(\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\alpha)}) = \Psi, \ \ \text{and} \ \ \mathrm{Cov}(\mathbf{f}^{(\alpha)}, \mathbf{u}^{(\alpha)}) = 0$$

with $\Psi = \mathrm{diag}(\psi_1, \cdots, \psi_p)$ a diagonal matrix (see, e.g., Lawley and Maxwell 1971). Here λ is an $r \times 1$ parameter vector. Then we have

$$\mathrm{Cov}(\mathbf{x}^{(\alpha)}, \mathbf{x}^{(\alpha)}) = \Lambda(\lambda)\Phi\Lambda(\lambda)' + \Psi = \Sigma(\theta).$$

The parameter vector to be estimated is $\theta = [\lambda', v(\Phi)', \psi_1, \cdots, \psi_p]'$ and $q = r + \frac{1}{2}k(k+1) + p$.

Assume first that $\mathbf{f}^{(\alpha)}$ and $U_1^{(\alpha)}, \cdots, U_p^{(\alpha)}$, the components of $\mathbf{u}^{(\alpha)}$, are mutually independent of one another. This is the case labeled as (III) in Section 3, since (A2) and (A3) are satisfied with $D = 0$, $\eta = 1$, and $\theta_n = [\lambda', v(\Phi(n))', \psi_1(n), \cdots, \psi_p(n)]'$, where

$$\Phi(n) = \frac{1}{n} \sum_{\alpha=1}^{N} (\mathbf{f}^{(\alpha)} - \bar{\mathbf{f}})(\mathbf{f}^{(\alpha)} - \bar{\mathbf{f}})'$$

and

$$\psi_i(n) = \frac{1}{n} \sum_{\alpha=1}^{N} (U_i^{(\alpha)} - \bar{U}_i)^2 \quad (i = 1, \cdots, p).$$

Thus, it is seen that $G_{11} = 0$ in view of (3.2), and it follows from Theorem 1 that for any FC estimators satisfying (A4),

$$\sqrt{n}(\widehat{\lambda} - \lambda) \xrightarrow{L} N_r \left(0, \left(\frac{\partial \widehat{\theta}}{\partial \sigma_0'} \Gamma_N \frac{\partial \widehat{\theta}}{\partial \sigma_0'}' \right)_{11} \right). \tag{5.1}$$

The asymptotic distribution of any FC estimator $\widehat{\lambda}$ under the general distributions is exactly the same as that under the normal, indicating robustness of $\widehat{\lambda}$. Satorra (1989) stated, without proofs, that (5.1) holds when the fourth-order moments are finite.

For MDF estimators, (5.1) becomes

$$\sqrt{n}(\widehat{\lambda}_{MDF} - \lambda) \xrightarrow{L} N_r \left(0, \ \left((\Delta'\bar{H}^{-1}\Delta)^{-1}\Delta'\bar{H}^{-1}\Gamma_N\bar{H}^{-1}\Delta(\Delta'\bar{H}^{-1}\Delta)^{-1})\right)_{11}\right)$$

in view of (4.6). For the MLE defined by (4.1) and its equivalences,

$$\sqrt{n}(\widehat{\lambda}_{MLE} - \lambda) \xrightarrow{L} N_r(0,\ (\Delta'\Gamma_N^{-1}\Delta)^{11}) \tag{5.2}$$

since $\bar{H} = \Gamma_N$. The result in (5.2) was obtained by Anderson (1987) and Anderson and Amemiya (1988).

Ihara and Kano (1986) proposed an estimator of Ψ that is an explicit function of S. Let S be partitioned into

$$S = \begin{array}{cc} \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & S_{22} & S_{23} & S_{24} \\ s_{31} & S_{32} & S_{33} & S_{34} \\ s_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} & \begin{array}{l} \}\quad 1 \\ \}\quad m \\ \}\quad m \\ \}\quad p-2m-1 \end{array} \end{array}.$$

Ihara and Kano then constituted an closed form estimator as follows:

$$\widehat{\psi}_1 = s_{11} - s_{12}S_{32}^{-1}s_{31}.$$

Interchanging observed variables appropriately, we get $\widehat{\psi}_i$ for $i = 2, \cdots, p$ in the same manner, and define $\widehat{\Psi} = \mathrm{diag}(\widehat{\psi}_1, \cdots, \widehat{\psi}_p)$. Take $m = k$, and then this estimator is obviously FC. The rest of the parameters, λ and Φ, are estimated by such methods that the resulting estimators are FC, e.g., by the method in (4.7). Notice that the (asymptotic) distribution of $\widehat{\Psi}$ can not be obtained because the fourth-order moments of the observations do not exist. Since the estimator is FC and the distributional assumptions (A2) and (A3) are met, Theorem 1 applies. Therefore, the estimator $\widehat{\lambda}$ is asymptotically normal, and (5.1) remains true, though $\widehat{\lambda}$ and $\widehat{\Psi}$ do correlate and the distribution of $\widehat{\Psi}$ is not known.

If we take $m > k$, $\widehat{\theta}$ is not FC any more (but weakly consistent, see Kano 1990a). Hence, none of the results developed here is true. Indeed, the asymptotic normality does not hold (see Kano 1991).

Hägglund (1982) developed FABIN methods in which three noniterative estimators were formed for a certain linear structure $\Lambda(\lambda)$. These estimators are verified to be FC. Hence, the formula for the asymptotic covariance matrix of $\widehat{\lambda}$ given by Hägglund remains valid not only for the normal case but for the case treated in this section as well. Satorra and Bentler (1991) studied robustness of goodness-of-fit statistics with the FABIN estimators.

We have shown that most estimators $\widehat{\lambda}$ are asymptotically normal. The next question to be raised is which estimator is the best of them, in other words, efficiency of estimators. This issue has been discussed only when the fourth-order moments of the observations are finite.

Recall that in the factor analysis model, (A1)–(A3) are satisfied and λ in θ_n is nonrandom. Thus Theorem 2 applies, so that the MLE $\widehat{\lambda}_{MLE}$ is asymptotically efficient. It should be emphasized that the fourth-order moments of the observations may not exist here. We would mention that the introduction of the notion of FC estimators as well as assumptions (A2) and (A3) enables us to observe the efficiency of FC estimators even though the fourth-order moments of the observations are infinite.

Assuming further that the fourth-order moments of the latent variates exist, we can obtain the asymptotic normality of $\widehat{\Phi}$ and $\widehat{\Psi}$ as well as $\widehat{\lambda}$. This is the case labeled as (II) in Section 3. We have from Theorem 1 that

$$\sqrt{n}\left(\begin{bmatrix}\widehat{\lambda}\\ v(\widehat{\Phi})\\ \widehat{\psi}_1\\ \vdots\\ \widehat{\psi}_p\end{bmatrix}-\begin{bmatrix}\lambda\\ v(\Phi)\\ \psi_1\\ \vdots\\ \psi_p\end{bmatrix}\right)\xrightarrow{L} N_q\left(0,\ \frac{\partial\widehat{\theta}}{\partial\sigma_0'}\Gamma_N\frac{\partial\widehat{\theta}}{\partial\sigma_0'}'+G\right) \tag{5.3}$$

with

$$G=\begin{bmatrix}0 & & & & \\ & D_k^{+}C_\Phi D_k^{+\prime} & & & \\ & & C_{\psi_1} & & \\ & & & \ddots & \\ & & & & C_{\psi_p}\end{bmatrix}$$

for any FC estimators, where C_Φ is the fourth-order cumulant matrix of $\mathbf{f}^{(\alpha)}$, and so on.

By (4.6), the asymptotic covariance matrix of MDF estimators is

$$(\Delta'\bar{H}^{-1}\Delta)^{-1}\Delta'\bar{H}^{-1}\Gamma_N\bar{H}^{-1}\Delta(\Delta'\bar{H}^{-1}\Delta)^{-1}+G,$$

which was first obtained by Satorra and Bentler (1990), whereas that of the MLE is

$$(\Delta'\Gamma_N^{-1}\Delta)^{-1}+G$$

in view of (2.5). This is the result due to Browne and Shapiro (1988) and Mooijaart and Bentler (1991).

ACKNOWLEDGMENT

The author would like to thank Professors A. Satorra of the University of Pompeu Fabra, Barcelona and T. K. Dijkstra of the University of Groningen for showing their previous works on Fisher consistency.

REFERENCES

Amemiya, Y. (1985). On the goodness-of-fit tests for linear structural relationships. *Technical Report No. 10*, Econometric Workshop, Stanford Univ.

Amemiya, Y. and Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Ann. Statist.*, **18**, 1453–1463.

Anderson, T. W. (1987). Multivariate linear relations. In *Proc. 2nd Inter. Tampere Con. Statist.*, (T. Pukkila and S. Puntanen, Eds), pp. 9–36, University of Tampere, Finland.

Anderson, T. W. (1989). Linear latent variable models and covariance structures. *J. Econometrics*, **41**, 91–119.

Anderson, T. W. and Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *Ann. Statist.*, **16**, 759–771.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika*, **48**, 493–517.

Bentler, P. M. (1989). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Los Angeles.

Bentler, P. M. and Dijkstra, T. K. (1985). Efficient estimation via linearization in structural models. In *Multivariate Analysis* **VI**, (P. R. Krishnaiah, Ed.), pp. 9–42, North Holland, Amsterdam, The Netherlands.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statist. J.*, **8**, 1–24.

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

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British J. Math. Statist. Psychol.*, **37**, 62–83.

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika*, **74**, 375–384.

Browne, M. and Shapiro, A. (1987). Adjustments for kurtosis in factor analysis with elliptically distributed errors. *J. Roy. Statist. Soc.* Ser. **B 49**, 346–352.

Browne, M. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British J. Math. Statist. Psychol.*, **41**, 193–208.

de Leeuw, J. (1986). Model selection in multinormal experiments. In *On Model Uncertainty and Its Statistical Implications*, (T. K. Dijkstra, Ed.), pp. 118–138. Springer–Verlag, Berlin, German.

Dijkstra, T. K. (1984). Personal Communications.

Hägglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika*, **47**, 209–222.

Ihara, M. and Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, **51**, 563–566.

Jöreskog, K. G. and Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika*, **37**, 243–260.

Jöreskog, K. G. and Sörbom, M. (1988). *LISREL 7, A Guide to the Program and Applications.* SPSS, Chicago.

Kano, Y. (1983). Consistency of estimators in factor analysis. *J. Japan Statist. Soc.*, **13**, 137–144.

Kano, Y. (1986). Conditions on consistency of estimators in covariance structure model. *J. Japan Statist. Soc.*, **16**, 75–80.

Kano, Y. (1990a). Noniterative estimation and the choice of the number of factors in exploratory factor analysis. *Psychometrika*, **55**, 277–291.

Kano, Y. (1990b). A simple adjustment of the normal theory inference for a wide class of distributions in linear latent variate models. (unpublished)

Kano, Y. (1991). The asymptotic distribution of a noniterative estimator in exploratory factor analysis. *Ann. Statist.*, **19**, 272–282.

Kano, Y., Berkane, M. and Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika*, **77**, 575–585.

Kano, Y., Berkane, M. and Bentler, P. M. (in press). Statistical inference based on pseudo maximum likelihood estimators in elliptical populations. *J. Amer. Statist. Assoc.*

Krane, W. A. and McDonald, R. P. (1978). Scale invariance and the factor analysis of correlation matrices. *British J. Math. Statist. Psychol.*, **31**, 218–228.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinb.*, **A 60**, 64–82.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method.* 2nd ed., Butterworths, London.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

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

Mooijaart, A. and Bentler, P. M. (1985). The weight matrix in asymptotic distribution-free methods. *British J. Math. Statist. Psychol.*, **38**, 190–196.

Mooijaart, A. and Bentler, P. M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica*, **49**, 159–171.

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

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

Parke, W. R. (1986). Pseudo maximum likelihood estimation: The asymptotic distribution. *Ann. Statist.*, **14**, 355–357.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* 2nd ed., Wiley, New York.

Satorra, A. (1989, July). Robustness of inferences based on the NT assumption: IV estimation. Paper presented at the Sixth European Meeting of the Psychometric Society. Leuven, Belgium.

Satorra, A. and Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Comp. Statist. Data Anal.*, **10**, 235–249.

Satorra, A. and Bentler, P. M. (1991). Goodness-of-fit test under IV estimation: Asymptotic robustness of a NT test statistic. In *Applied Stochastic Models and Data Analysis*, (R. Gutierrez and M. J. Valderrama, Eds.), pp. 555–567. World Scientific, London.

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

Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statist. J.*, **19**, 73–81.

Shapiro, A. (1986). Asymptotic theory of overparametrized structural models. *J. Amer. Statist. Assoc.*, **81**, 142–149.

Shapiro, A. (1987). Robustness properties of the MDF analysis of moment structures. *South African Statist. J.*, **21**, 39–62.

Shapiro, A. and Browne, M. (1987). Analysis of covariance structures under elliptical distributions. *J. Amer. Statist. Assoc.*, **82**, 1092–1097.

Shapiro, A. and Browne, M. (1989). On the asymptotic bias of estimators under parameter drift. *Statist. Prob. Lett.*, **7**, 221–224.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, **70**, 411–420.