The Annals of Statistics
1990, Vol. 18, No. 3, 1453–1463

ASYMPTOTIC CHI-SQUARE TESTS FOR A LARGE CLASS OF FACTOR ANALYSIS MODELS[1]

By Yasuo Amemiya and T. W. Anderson

Iowa State University and Stanford University

Three types of asymptotic χ^2 goodness-of-fit tests derived under the normal assumption have been used widely in factor analysis. Asymptotic behavior of the test statistics is investigated here for the factor analysis model with linearly or nonlinearly restricted factor loadings under weak assumptions on the factor vector and the error vector. In particular the limiting χ^2 result for the three tests is shown to hold for the factor vector, either fixed or random with any distribution having finite second-order moments, and for the error vector with any distribution having finite second-order moments, provided that the components of the error vector are independent, not just uncorrelated. As special cases the result holds for exploratory and confirmatory factor analysis models and for certain nonnormal structural equation (LISREL) models.

1. Introduction. A factor analysis model specifies the structure of the covariance matrix of a random vector. A goodness-of-fit statistic measures the deviation of the sample covariance matrix from the estimated covariance matrix with this structure. If the random vector is normally distributed, a measure such as -2 times the logarithm of the likelihood ratio criterion has a limiting χ^2 distribution. In this paper several such measures are considered, and it is shown that the limiting χ^2 distribution holds under conditions much more general than normality of the observations. The definition of the model is broad enough to include structural equation models and other covariance structures. The impact of these results is that the use of the χ^2 tests computed by the standard computer packages (LISREL, EFAP and EQS, for example) is valid asymptotically for most factor analyses and for a class of structural equation models (LISREL models) even when the data are nonnormal.

Anderson and Amemiya (1988) discussed the estimation problem in factor analysis under general conditions on the distribution of observations. We follow their notation to define the model which will be treated throughout this paper (although our model here is slightly more general). Let the observable p-component random column vector $\mathbf{x}_\alpha$ be written as

$$\mathbf{x}_\alpha = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_\alpha + \mathbf{u}_\alpha, \qquad \alpha = 1, \dots, N, \tag{1.1}$$

Received March 1986; revised May 1989.

[1]Research partly supported by NSF grants DMS 82-19748 and DMS 86-03779.

AMS 1980 *subject classifications*. Primary 62H25; secondary 62F05.

Key words and phrases. Factor analysis, goodness-of-fit test, asymptotic robustness, structural equation model.

where $\boldsymbol{\mu}$ is a p-component vector of parameters, $\boldsymbol{\Lambda}$ is a $p \times k$ matrix of factor loadings, $\mathbf{f}_\alpha$ is a k-component unobservable factor vector, which may contain fixed and/or random components and $\mathbf{u}_\alpha$ is a p-component unobservable random error vector. It is assumed that all $\mathbf{f}_\alpha$'s and $\mathbf{u}_\alpha$'s are uncorrelated and that $\mathscr{E}\mathbf{u}_\alpha = \mathbf{0}$ and $\mathscr{E}\mathbf{u}_\alpha\mathbf{u}'_\alpha = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \operatorname{diag}\{\psi_{11}, \dots, \psi_{pp}\}$ is a $p \times p$ diagonal matrix with diagonal elements ψ_{ii}, $i = 1, \dots, p$.

In (1.1) $\boldsymbol{\Lambda}$ can be replaced by $\boldsymbol{\Lambda}\mathbf{C}$ and $\mathbf{f}_\alpha$ by $\mathbf{C}^{-1}\mathbf{f}_\alpha$ (where $\mathbf{C}$ is nonsingular) to obtain an observationally equivalent model. To eliminate this indeterminacy, restrictions may be imposed on the parameters. In this paper we suppose that

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\lambda}) \tag{1.2}$$

is a specified function of a $q \times 1$ parameter vector $\boldsymbol{\lambda}$. For example, the model identified by specifying certain elements of $\boldsymbol{\Lambda}$ to be 0's and 1's has a loading matrix that is a known linear function of unspecified loadings $\boldsymbol{\lambda}$. Here, however, $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$ may be a nonlinear function. We shall show later that our results derived under parameterization (1.2) hold also under some other parameterizations.

Inference will be based on the unbiased sample covariance matrix

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

where $\mathbf{x} = (1/N)\Sigma_{\alpha=1}^{N}\mathbf{x}_\alpha$ and $n = N - 1$. We shall study the goodness-of-fit test statistics, which were originally proposed under the normality of $\mathbf{f}_\alpha$ and $\mathbf{u}_\alpha$, under a large class of distributional assumptions on $\mathbf{f}_\alpha$ and $\mathbf{u}_\alpha$. To define the test statistics, suppose for a moment that all components of $\mathbf{f}_\alpha$ are random, all $\mathbf{f}_\alpha$ and $\mathbf{u}_\alpha$ are independent, $\mathbf{f}_\alpha \sim N_k(\mathbf{0}, \boldsymbol{\Phi})$ and $\mathbf{u}_\alpha \sim N_p(\mathbf{0}, \boldsymbol{\Psi})$. Then,

$$n\mathbf{S} \sim W_p[\boldsymbol{\Sigma}(\boldsymbol{\theta}), n], \tag{1.3}$$

where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\boldsymbol{\lambda})\boldsymbol{\Phi}\boldsymbol{\Lambda}'(\boldsymbol{\lambda}) + \boldsymbol{\Psi} \tag{1.4}$$

and $\boldsymbol{\theta} = (\boldsymbol{\lambda}', [\operatorname{vech} \boldsymbol{\Phi}]', \boldsymbol{\psi}')'$; $\operatorname{vech} \boldsymbol{\Phi}$ is a column vector of the $\frac{1}{2}k(k+1)$ functionally independent components of $\boldsymbol{\Phi}$ and $\boldsymbol{\psi} = (\psi_{11}, \dots, \psi_{pp})'$. The computation of each goodness-of-fit test statistic involves the minimization of an objective function of $\boldsymbol{\theta}$ over the parameter space. For this purpose we define the parameter space for $\boldsymbol{\theta}$ as $\Omega = \Omega_\lambda \times \Omega_\phi \times \Omega_\psi$, where $\Omega_\lambda \subseteq \mathbb{R}^q$, Ω_ϕ consists of $\operatorname{vech} \boldsymbol{\Phi}$ such that $\boldsymbol{\Phi}$ is nonnegative definite and Ω_ψ consists of $\boldsymbol{\psi}$ with nonnegative components.

If (1.3) holds, the maximum Wishart likelihood estimator $\hat{\boldsymbol{\theta}}_1 = (\hat{\boldsymbol{\lambda}}_1, [\operatorname{vech} \hat{\boldsymbol{\Phi}}_1]', \hat{\boldsymbol{\psi}}_1')'$ is the vector $\boldsymbol{\theta} \in \Omega$ that minimizes

$$\begin{aligned} L_1[\boldsymbol{\Sigma}(\theta), \mathbf{S}] &= \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \log|\mathbf{S}| + \operatorname{tr} \mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - p \\ &= \operatorname{tr} \mathbf{S}\boldsymbol{\Sigma}^{-1}(\theta) - \log|\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})| - p, \end{aligned} \tag{1.5}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is defined by (1.4). The discrepancy function $L_1[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]$ is $-(2/n)$ times the logarithm of the Wishart likelihood function plus terms not depending on $\boldsymbol{\theta}$. Anderson and Amemiya (1988) discussed asymptotic properties of $\hat{\boldsymbol{\theta}}_1$. One goodness-of-fit statistic is $G_1 = nL_1[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_1), \mathbf{S}]$, which is -2 times the logarithm of the Wishart likelihood ratio criterion for testing the null hypothesis of (1.4) against the alternative hypothesis of unrestricted $\boldsymbol{\Sigma}$. For a certain class of nonlinear functions $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$, the standard computer packages for the structural equation model, such as LISREL, compute the maximum Wishart likelihood estimator $\hat{\boldsymbol{\theta}}_1$ and the χ^2 goodness-of-fit test statistic G_1. [The likelihood ratio test statistic based on $\mathbf{x}_\alpha$, $\alpha = 1, \dots, N$ (not just on $\mathbf{S}$) has the form of G_1 with n replaced by N in the expression of G_1, the divisor in $\mathbf{S}$ and the computation of $\hat{\boldsymbol{\theta}}_1$. Our asymptotic results for G_1 hold also for such a statistic based on the $\mathbf{x}_\alpha$'s.]

An alternative approach to estimation is generalized least squares. [See Browne (1974), for example.] Let $\hat{\boldsymbol{\theta}}_2 \in \Omega$ minimize the discrepancy function

$$(1.6) \quad L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}] = \frac{1}{2}\operatorname{tr}\{[\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \mathbf{S}]\mathbf{S}^{-1}\}^2 = \frac{1}{2}\operatorname{tr}[\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{S}^{-1} - \mathbf{I}]^2.$$

The corresponding goodness-of-fit statistic is $G_2 = nL_2[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2), \mathbf{S}]$.

Another discrepancy function is

$$(1.7) \quad L_3[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}] = \frac{1}{2}\operatorname{tr}\{[\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \mathbf{S}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\}^2 = \frac{n}{2}\operatorname{tr}[\mathbf{I} - \mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})]^2.$$

The estimator $\hat{\boldsymbol{\theta}}_3$ minimizes $L_3[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]$, but we define $G_3 = nL_3[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_1), \mathbf{S}]$ because it is more commonly used. [This statistic may be obtained by iteratively reweighted least squares, as presented by Lee and Jennrich (1979), for example.]

The statistics G_1, G_2 and G_3 have been used for assessing the fit of a model with a particular structure $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$ and for determining the appropriate number of factors k. If $(\mathbf{f}'_\alpha, \mathbf{u}'_\alpha)'$ is normally distributed, that is, if (1.3) holds, then under the correct model specification, as $n \to \infty$, each of G_1, G_2 and G_3 converges in distribution to a χ^2 random variable with degrees of freedom $d = \frac{1}{2}p(p+1) - \frac{1}{2}k(k+1) - p - q$, where q is the dimension of $\boldsymbol{\lambda}$ in $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$. See Browne (1974), for example. Because the G_i's are readily computable by existing packages, we may consider using the G_i's as the goodness-of-fit test statistics for nonnormal data also, when the $(\mathbf{f}'_\alpha, \mathbf{u}'_\alpha)$'s are not considered to be normally distributed. This paper justifies that use.

A limited literature exists for the study of the asymptotic behavior of the G_i's when $\mathbf{u}_\alpha$'s are normal but $\mathbf{f}_\alpha$'s are not normal. For a class of functional and structural relationship models containing the factor analysis model as a special case, Amemiya (1985) discussed asymptotic properties of the statistics corresponding to G_1 and G_3. For factor analysis, Amemiya's (1985) result is that under the correct model specification, as $n \to \infty$, each of G_1 and G_3 converges to a χ^2_d random variable if the $\mathbf{f}_\alpha$'s are either fixed or independently and identically distributed (i.i.d.) with finite second-order moments and if $\mathbf{u}_\alpha$'s are i.i.d. normal random variables. Browne (1987) showed the χ^2_d result for G_1 and G_2 under a condition similar to, but stronger than, the condition in Amemiya's (1985). In this paper we present general conditions under which the limiting χ^2 result for the G_i's hold. In particular, we shall show that under the correct model specification G_1, G_2 and G_3 converge to a χ_d^2 random variable if $\mathbf{f}_\alpha$'s are fixed or i.i.d. with any distribution having finite second-order moments, if $\mathbf{u}_\alpha$'s are i.i.d. with any distribution having finite second-order moments and if the p-components of $\mathbf{u}_\alpha$ are independent, not just uncorrelated. Using an approach different from ours, stimulated by Amemiya and Anderson (1985), Browne and Shapiro (1988) showed the limiting χ_d^2 result for G_1 and G_2 for a broad class of nonnormal linear structure models. For the factor analysis model, their conditions on $\mathbf{f}_\alpha$ and $\mathbf{u}_\alpha$ are stronger than ours.

Anderson and Amemiya (1988) showed that the asymptotic covariance matrix of $\hat{\boldsymbol{\lambda}}_1$, the MLE of the loading parameter, is common to a large class of nonnormal $(\mathbf{f}'_\alpha, \mathbf{u}'_\alpha)$. Here we shall show that for the same class of nonnormal $(\mathbf{f}'_\alpha, \mathbf{u}'_\alpha)$, G_1, G_2 and G_3 converge in distribution to a χ_d^2 random variable under the correctly specified model. The approach underlying developments in this paper is the same as that used in the earlier paper. Because G_1 and G_3 are functions of $\hat{\boldsymbol{\theta}}_1$, we use a result for $\hat{\boldsymbol{\theta}}_1$ derived in the earlier paper. A similar result for $\hat{\boldsymbol{\theta}}_2$ appearing in G_2 is derived here. The results in the earlier paper are dependent on the identifying restriction of the type (1.2). Here, we develop results for the test statistics G_i's which hold also under restrictions other than (1.2).

More references for statistical inference in factor analysis as well as a description of these results are available in Anderson (1984a, 1984b) and Anderson and Amemiya (1988). The generalized least squares method was first applied to the factor analysis model by Jöreskog and Goldberger (1972). The method was later applied to more general covariance structure models, for example, by Browne (1974, 1982), Bentler (1983) and Shapiro (1983).

2. Main results. The statistics G_1, G_2 and G_3 are functions of the sample covariance matrix $\mathbf{S}$, which in turn is a function of the unobservable sample covariance matrices of $\mathbf{f}_\alpha$ and $\mathbf{u}_\alpha$:

$$\begin{aligned}\mathbf{S} &= \frac{1}{n}\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})' \\ &= \frac{1}{n}\sum_{\alpha=1}^{N}\left[\boldsymbol{\Lambda}(\boldsymbol{\lambda}_0)(\mathbf{f}_\alpha - \bar{\mathbf{f}}) + (\mathbf{u}_\alpha - \bar{\mathbf{u}})\right]\left[\boldsymbol{\Lambda}(\boldsymbol{\lambda}_0)(\mathbf{f}_\alpha - \bar{\mathbf{f}}) + (\mathbf{u}_\alpha - \bar{\mathbf{u}})\right]' \qquad (2.1)\\ &= \boldsymbol{\Lambda}(\boldsymbol{\lambda}_0)\boldsymbol{\Phi}(n)\boldsymbol{\Lambda}'(\boldsymbol{\lambda}_0) + \boldsymbol{\Lambda}(\boldsymbol{\lambda}_0)\boldsymbol{\Gamma}(n) + \boldsymbol{\Gamma}'(n)\boldsymbol{\Lambda}'(\boldsymbol{\lambda}_0) + \boldsymbol{\Psi}(n),\end{aligned}$$

where $\boldsymbol{\lambda}_0$ is the true value of $\boldsymbol{\lambda}$ (the value of $\boldsymbol{\lambda}$ in the population sampled) and the (unobservable) sample covariances of the factors and errors are

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

$$\boldsymbol{\Psi}(n) = \frac{1}{n}\sum_{\alpha=1}^{N}(\mathbf{u}_\alpha - \bar{\mathbf{u}})(\mathbf{u}_\alpha - \bar{\mathbf{u}})',$$

$$\boldsymbol{\Gamma}(n) = \frac{1}{n}\sum_{\alpha=1}^{N}(\mathbf{f}_\alpha - \bar{\mathbf{f}})(\mathbf{u}_\alpha - \bar{\mathbf{u}})',$$

$\bar{\mathbf{f}} = (1/N)\Sigma_{\alpha=1}^{N}\mathbf{f}_\alpha$, $\bar{\mathbf{u}} = (1/N)\Sigma_{\alpha=1}^{N}\mathbf{u}_\alpha$. We need to separate the diagonal elements and the off-diagonal elements of $\mathbf{\Psi}(n)$: Let $\boldsymbol{\psi}(n) = [\psi_{11}(n), \dots, \psi_{pp}(n)]'$ be the $p \times 1$ vector consisting of the diagonal elements of $\mathbf{\Psi}(n)$ and let $\boldsymbol{\psi}_b(n)$ be the $\frac{1}{2}p(p-1) \times 1$ vector listing the off-diagonal elements of $\mathbf{\Psi}(n)$. We present assumptions on these quantities which will be used later.

ASSUMPTION 1. (a) $\mathbf{\Phi}(n) \to_p \mathbf{\Phi}_0$ positive definite.
(b) $\boldsymbol{\psi}(n) \to_p \psi_0$, where each component of $\boldsymbol{\psi}_0$ is positive.

Throughout this section the model (1.1) holds with the parameterization (1.2). We define $\boldsymbol{\theta}_0 = [\boldsymbol{\lambda}'_0, (\operatorname{vech}\mathbf{\Phi}_0)', \boldsymbol{\psi}'_0]'$, where $\mathbf{\Phi}_0$ and $\boldsymbol{\psi}_0$ are given in Assumption 1. The vector $\boldsymbol{\theta}_0$ is the limiting true value of $\boldsymbol{\theta}$. In order to estimate $\boldsymbol{\theta}$, the family of matrices of $\mathbf{\Sigma}(\boldsymbol{\theta})$ must have the property that any $\mathbf{\Sigma}(\boldsymbol{\theta})$ corresponds to exactly one value of $\boldsymbol{\theta}$; that is, the model must be identified. In Assumption 2(a) the inverse of $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; Rao (1973) calls this condition *strong identifiability*. [Assumption 2(a) is slightly different from the identification condition (v) in Anderson and Amemiya (1988); the condition here is easier to understand and verify.] Let $\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}}$, $\|\mathbf{X}\| = \sqrt{\operatorname{tr}\mathbf{X}'\mathbf{X}} = \sqrt{(\operatorname{vec}\mathbf{X})'\operatorname{vec}\mathbf{X}}$ and $\operatorname{vec}\mathbf{A} = (\mathbf{a}'_1, \dots, \mathbf{a}'_m)'$ for $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_m)$.

ASSUMPTION 2. (a) Given $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\|\mathbf{\Sigma}(\boldsymbol{\theta}) - \mathbf{\Sigma}(\boldsymbol{\theta}_0)\| < \delta$$

implies $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \varepsilon$.
(b) The population vector $\boldsymbol{\lambda}_0$ is in the interior of Ω_λ.
(c) $\mathbf{\Lambda}(\boldsymbol{\lambda})$ is twice continuously differentiable in a neighborhood of $\boldsymbol{\lambda}_0$ and

$$\left.\frac{\partial \operatorname{vec}\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

is of full column rank.

The crucial idea in Anderson and Amemiya (1988), which was suggested by Anderson and Rubin (1956), is obtaining an asymptotic expansion of $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}(n)$ instead of $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}(n)$ is the hybrid vector

$$\boldsymbol{\theta}(n) = \{\boldsymbol{\lambda}'_0, [\operatorname{vech}\mathbf{\Phi}(n)]', \boldsymbol{\psi}(n)\}'.$$

They showed that the leading term in $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}(n)$ is a linear function of

$$\boldsymbol{\xi}(n) = \{[\operatorname{vec}\mathbf{\Gamma}(n)]', \boldsymbol{\psi}'_b(n)\}'$$

and is free of $\mathbf{\Phi}(n)$ and $\boldsymbol{\psi}(n)$. In this paper we derive the same expansion for $\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}(n)$. Using the expansions of $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}(n)$ and $\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}(n)$, we shall show that each of G_1, G_2 and G_3 is approximately a quadratic form in $\boldsymbol{\xi}(n)$. When the asymptotic covariance matrix of this vector is of a certain form (related to the matrix of the quadratic form), the statistics have a common asymptotic χ^2 distribution.

ASSUMPTION 3. $\xi(n) = O_p(1/\sqrt{n})$.

LEMMA 1. *Assumptions* 1, 2 *and* 3 *imply*

$$\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}(n) = \mathbf{C}_1(\boldsymbol{\theta}_0)\operatorname{vec}\{\mathbf{S} - \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)]\} + o_p\left(\frac{1}{\sqrt{n}}\right), \qquad i = 1, 2, \tag{2.2}$$

where $\mathbf{C}_1(\boldsymbol{\theta}_0)$ *is a matrix depending only on* $\boldsymbol{\theta}_0$.

PROOF. Anderson and Amemiya (1988) proved (2.2) for $\hat{\boldsymbol{\theta}}_1$ and $\boldsymbol{\Lambda}(\boldsymbol{\lambda}) = \mathbf{a} + A\boldsymbol{\lambda}$ under Assumptions 1, 2 and 3 and argued further that the result was true for any $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$ satisfying Assumption 2(c). We present here the proof for $\hat{\boldsymbol{\theta}}_2$.

Assumptions 1 and 3 imply $\mathbf{S} \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and

$$0 \le L_2\big[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2), \mathbf{S}\big] \le L_2\big[\boldsymbol{\Sigma}(\boldsymbol{\theta}_0), \mathbf{S}\big] \to_p 0.$$

Hence, $\mathbf{S}^{-1/2}[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)]\mathbf{S}^{-1/2} \to_p \mathbf{0}$ and $\boldsymbol{\Sigma}(\hat{\theta}_2) \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. Thus by Assumption 2(a), $\hat{\boldsymbol{\theta}}_2 \to_p \boldsymbol{\theta}_0$. Hence the probability that $\hat{\boldsymbol{\theta}}_2$ satisfies the derivative conditions tends to 1 as $n \to \infty$. By Assumption 2(c), the first and second derivatives of $L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]$ with respect to $\boldsymbol{\theta}$ exist and are continuous functions of $\boldsymbol{\theta}$ and $\mathbf{S}$. Thus, using the Taylor expansion around $\boldsymbol{\theta}(n)$, we have

$$\begin{aligned} o_p\left(\frac{1}{n}\right) &= \frac{\partial L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]}{\partial \boldsymbol{\theta}}\bigg|_{\hat{\boldsymbol{\theta}}_2} \\ &= \frac{\partial L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}(n)} + \frac{\partial^2 L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}^*} \big[\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}(n)\big], \end{aligned}$$

where $\boldsymbol{\theta}^*$ is on the line segment between $\hat{\boldsymbol{\theta}}_2$ and $\boldsymbol{\theta}(n)$. Direct computation shows

$$\begin{aligned} &\frac{\partial L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}(n)} \\ &\quad = -\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}(n)}\right]'(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\operatorname{vec}\{\mathbf{S} - \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)]\} \\ &\quad = -\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}_0}\right]'\big[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\big]\operatorname{vec}\{\mathbf{S} - \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)]\} \\ &\qquad + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &\quad = O_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

because

$$\begin{aligned}\mathbf{S} - \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)] &= \boldsymbol{\Lambda}(\boldsymbol{\lambda}_0)\boldsymbol{\Gamma}(n) + \boldsymbol{\Gamma}'(n)\boldsymbol{\Lambda}'(\boldsymbol{\lambda}_0) + \boldsymbol{\Psi}(n) - \operatorname{diag}\{\boldsymbol{\psi}(n)\}\\ &= O_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{2.3}$$

the diagonal elements of $\boldsymbol{\Psi}(n) - \operatorname{diag}\{\boldsymbol{\psi}(n)\}$ are zeroes and the off-diagonal elements are elements of $\boldsymbol{\psi}_b(n)$. [Because $\operatorname{plim}_{n\to\infty}\mathbf{S} = \operatorname{plim}_{n\to\infty}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is positive definite by Assumption 1, $\mathbf{S}^{-1}$ and $\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}_2)$ exist with probability approaching 1.] The proof is completed by

$$\begin{aligned}\left.\frac{\partial^2 L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}^*} &\to_p \left.\frac{\partial^2 L_2[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)]}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\\ &= \left[\left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\right]'\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\right]\left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\\ &= \mathbf{H}(\boldsymbol{\theta}_0),\end{aligned}$$

say; $\mathbf{H}(\boldsymbol{\theta}_0)$ is positive definite by Assumption 2(c). □

LEMMA 2. *Assumptions* 1, 2 *and* 3 *imply*

$$G_i = n\boldsymbol{\xi}'(n)\mathbf{P}(\boldsymbol{\theta}_0)\boldsymbol{\xi}(n) + o_p(1), \qquad i = 1, 2, 3,$$

where $\mathbf{P}(\boldsymbol{\theta}_0)$ *is a symmetric nonstochastic matrix depending only on* $\boldsymbol{\theta}_0$.

PROOF. For $G_2 = nL_2[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2), \mathbf{S}]$

$$\begin{aligned}G_2 &= \frac{n}{2}\{\operatorname{vec}[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)]\}'(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\operatorname{vec}[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)]\\ &= \frac{n}{2}\{\operatorname{vec}[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)]\}'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)]\operatorname{vec}[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)] + o_p(1).\end{aligned}$$

By (2.3), $\operatorname{vec}\{\mathbf{S} - \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)]\} = \mathbf{C}_2(\boldsymbol{\theta}_0)\boldsymbol{\xi}(n)$, where $\mathbf{C}_2(\boldsymbol{\theta}_0)$ depends only on $\boldsymbol{\theta}_0$. Hence, by Lemma 1,

$$\begin{aligned}\operatorname{vec}[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)] &= \operatorname{vec}\{\mathbf{S} - \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)] + \boldsymbol{\Sigma}[\boldsymbol{\theta}(n)] - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)\}\\ &= \mathbf{C}_3(\boldsymbol{\theta}_0)\boldsymbol{\xi}(n) + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where

$$\mathbf{C}_3(\boldsymbol{\theta}_0) = \mathbf{C}_2(\boldsymbol{\theta}_0) - \left[\left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\theta_0)}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\right]\mathbf{C}_1(\boldsymbol{\theta}_0)\mathbf{C}_2(\boldsymbol{\theta}_0).$$

Hence, for $i = 2$,

$$G_2 = \frac{n}{2}\boldsymbol{\xi}'(n)\mathbf{C}_3'(\boldsymbol{\theta}_0)[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)]\mathbf{C}_3(\boldsymbol{\theta}_0)\boldsymbol{\xi}(n) + o_p(1). \tag{2.4}$$

For $i = 3$, the proof is similar. For $i = 1$ we use the Taylor expansion

$$G_1 = n\left\{\operatorname{tr}\left[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_1)\right]\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}_1) - \log\left|\mathbf{I} + \left[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_1)\right]\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}_1)\right|\right\}$$

$$= \frac{n}{2}\operatorname{tr}\left\{\left[\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_1)\right]\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}_1)\right\}^2 + o_p(1)$$

to reduce the proof to the case of G_3. □

If the $(\mathbf{f}'_\alpha, \mathbf{u}'_\alpha)$'s are normally distributed, the covariance matrix of $\sqrt{n}\,\boldsymbol{\xi}(n)$ depends only on $\boldsymbol{\Phi}_0$ and $\boldsymbol{\psi}_0$, say $\mathbf{W}_0(\boldsymbol{\theta}_0)$. In this case by the usual asymptotic theory for likelihood ratio criteria, $G_1 \to_L \chi^2_d$, where $d = \frac{1}{2}p(p+1) - q - \frac{1}{2}k(k+1) - p$. We, therefore, shall replace Assumption 3 by the following more restrictive assumption on the limiting behavior of $\boldsymbol{\xi}(n)$.

ASSUMPTION 4. As $n \to \infty$,

$$\sqrt{n}\,\boldsymbol{\xi}(n) \to_L \mathbf{N}[\mathbf{0}, \mathbf{W}_0(\boldsymbol{\theta}_0)],$$

where $\mathbf{W}_0(\boldsymbol{\theta}_0)$ is the covariance matrix of $\sqrt{n}\,\boldsymbol{\xi}(n)$ computed for the case of normality.

THEOREM 1. *Assumptions* 1, 2 *and* 4 *imply*

$$G_i \to_L \chi^2_d, \qquad i = 1, 2, 3, \tag{2.5}$$

where $d = \frac{1}{2}p(p+1) - q - \frac{1}{2}k(k+1) - p$.

PROOF. Lemma 2 implies that limiting distribution of G_i depends only on the limiting distribution of $\sqrt{n}\,\boldsymbol{\xi}(n)$. However, the limiting normal distribution of $\sqrt{n}\,\boldsymbol{\xi}(n)$ depends only on the covariance matrix. So it has the χ^2-distribution if the limiting covariance of $\sqrt{n}\,\boldsymbol{\xi}(n)$ is the same as the normal case. □

ASSUMPTION 5. The two sequences $\{\mathbf{f}_\alpha\}$ and $\{\mathbf{u}_\alpha\}$ are independent, the p-components $u_{i\alpha}$, $i = 1, \dots, p$, of $\mathbf{u}_\alpha$ are independent and $u_{i\alpha}$, $\alpha = 1, \dots, n$, are independently identically distributed with mean zero and positive variance ψ_{0i}, the ith component of $\boldsymbol{\psi}_0$, $i = 1, \dots, p$.

ASSUMPTION 6. As $n \to \infty$, $\boldsymbol{\Phi}(n) \to \boldsymbol{\Phi}_0$ a.s., where $\boldsymbol{\Phi}_0$ is positive definite.

COROLLARY 1. *Assumptions* 2, 5 *and* 6 *imply* (2.5).

PROOF. Assumptions 5 and 6 imply Assumptions 1 and 4. See the proof of Corollary 1 (for estimators) in Anderson and Amemiya (1988). Thus the result follows from Theorem 1. □

3. Discussion. The impact of Theorem 1 and Corollary 1 on factor analysis with structured factor loading matrix $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$ is that the three goodness-of-fit test statistics G_1, G_2 and G_3 have the asserted asymptotic significance level under the correct model specification for possibly nonnormal $(\mathbf{f}'_\alpha, \mathbf{u}'_\alpha)'$. Although the asymptotic distribution of $\mathbf{S}$ may depend on the existence of fourth-order moments, the limiting distribution of G_1, G_2 and G_3 requires no restriction on the third- or higher-order moments of $\mathbf{f}_\alpha$ and $\mathbf{u}_\alpha$, not even their existence. (In the earlier paper examples of nonnormal models were given.) The conditions of Corollary 1 are identical to the conditions under which the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_0)$, $i = 1, 2$, is the same as that under the normality. Anderson and Amemiya (1988) discussed the generality of these conditions. They also showed that the results for the factor analysis model with parameterization (1.2) apply to a certain class of structural equation (LISREL) models.

In the literature two kinds of factor analysis are distinguished depending on the nature of the restrictions imposed to remove the indeterminacy in (1.1). In exploratory (unrestricted) factor analysis the restrictions are imposed only for uniqueness of the parameters. In confirmatory (restricted) factor analysis the investigator uses prior knowledge about the variables to formulate a hypothesis imposing restrictions on the parameters, such as certain factor loadings being 0. The model may be restricted in the sense that the number of restrictions may exceed that required for identification [Jöreskog (1969)]. If the restrictions are placed only on $\boldsymbol{\Lambda}$ as in (1.2), then the results of Section 2 apply to both exploratory and confirmatory factor analyses.

The assumption that restrictions are placed only on the factor loading matrix $\boldsymbol{\Lambda}$ can be relaxed to some extent without altering the results of Theorem 1 and Corollary 1. Suppose that for the model with random $\mathbf{f}_\alpha$, one parameterization, denoted by A, places restrictions on the factor loading matrix $\boldsymbol{\Lambda}$, the factor covariance matrix $\boldsymbol{\Phi}$ and the error variance vector $\boldsymbol{\psi}$. Suppose also that parameterization B places restrictions only on $\boldsymbol{\Lambda}$ and that there is a one-to-one relationship between the two sets of possible $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ under parameterizations A and B. Then, our results hold for G_1, G_2 and G_3 computed under parameterization A with restrictions on $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\psi}$. This is because $\boldsymbol{\theta}$ enters the log Wishart likelihood (1.5) and the generalized least squares criterion (1.6) only through $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, because the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_1$ or the generalized least square estimator $\hat{\boldsymbol{\theta}}_2$, enters G_1, G_2 and G_3 only through $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_1)$ or $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_2)$, and because the values of G_1, G_2 and G_3 computed under the two parameterizations are the same. For example, a commonly used confirmatory factor analysis with random factors assumes that certain loadings are 0 and factor variances are 1. Such a model is equivalent to the model where the same factor loadings are 0, one of the nonzero factor loadings in each column of $\boldsymbol{\Lambda}$ is set to be 1, and factor variances are unrestricted. Thus, our results hold for such a confirmatory factor analysis model with restrictions on factor variances.

In exploratory (unrestricted) factor analysis with random $\mathbf{f}_\alpha$, there exist several equivalent parameterizations. One such parameterization assumes that $\boldsymbol{\Lambda}$ is of the form $(\boldsymbol{\Lambda}'_1, \mathbf{I}_k)'$ and leaves $\boldsymbol{\Phi}$ unrestricted; Theorem 1 and Corollary 1 directly apply to G_1, G_2 and G_3 under this parameterization. The set of possible $\boldsymbol{\Sigma}$ under this parameterization with $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}'_1, \mathbf{I}_k)'$ is the same as that under any other parameterization of exploratory factor analysis, for example, the parameterization assuming that $\mathbf{\Phi} = \mathbf{I}_k$ and $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ is diagonal. Thus, the values of G_1, G_2 and G_3 do not depend on which of the equivalent parameterizations for exploratory factor analysis is used in the computation. Hence, our results hold for G_1, G_2 and G_3 computed under any parameterization for exploratory factor analysis.

For exploratory factor analysis, stronger results than Lemma 2, Theorem 1 and Corollary 1 hold. In that case, the common leading term in the expansions of the G_i's in Lemma 2 is a quadratic form only in $\boldsymbol{\psi}_b(n)$ [free of $\mathbf{\Gamma}(n)$] and Assumption 4 in Theorem 1 can be replaced by the condition that $\sqrt{n}\,\boldsymbol{\psi}_b(n)$ has the same limiting normal distribution as that for the case with normal $\mathbf{u}_\alpha$. Corollary 1 holds for exploratory factor analysis replacing the independence of $\{\mathbf{f}_\alpha\}$ and $\{\mathbf{u}_\alpha\}$ in Assumption 5 by $\mathbf{\Gamma}(n) = O_p(1/\sqrt{n})$. See Amemiya and Anderson (1985).

It is possible to extend our results to cover goodness-of-fit test statistics other than G_1, G_2 and G_3. Using $L_i[\mathbf{\Sigma}(\boldsymbol{\theta}), \mathbf{S}]$ and $\hat{\boldsymbol{\theta}}_i$, $i = 1, 2, 3$, defined in (1.5), (1.6) and (1.7), we can consider nine test statistics $nL_i[\mathbf{\Sigma}(\boldsymbol{\theta}_j), \mathbf{S}]$, $i = 1, 2, 3$, $j = 1, 2, 3$. Lemma 2, Theorem 1 and Corollary 1 hold for all nine statistics. The derivations follow the steps of the proof given here for G_1, G_2 and G_3. We have concentrated on G_1, G_2 and G_3 because these three are commonly used goodness-of-fit statistics.

REFERENCES

AMEMIYA, Y. (1985). On the goodness-of-fit tests for linear statistical relationships. Technical Report No. 10, Econometric Workshop, Stanford Univ. (prepared under NSF Grant DMS 82-19748).

AMEMIYA, Y. and ANDERSON, T. W. (1985). Asymptotic chi-square tests for a large class of factor analysis models. Technical Report No. 13, Econometric Workshop, Stanford Univ. (prepared under NSF Grant DMS 82-19748).

ANDERSON, T. W. (1984a). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.

ANDERSON, T. W. (1984b). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

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

ANDERSON, T. W. and RUBIN, H. (1956). Statistical inference in factor analysis. *Proc. Third Berkeley Symp. Math. Statist. Probab.* (J. Neyman, ed.) **5** 111–150. Univ. Calif. Press, Berkeley, Calif.

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

BROWNE, M. W. (1974). Generalized least squares estimator 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.) 72–141. Cambridge Univ. Press, Cambridge, England.

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

BROWNE, M. W. and SHAPIRO, A. (1988). Robustness of normal theory methods in the analysis of linear latent variable models. *British J. Math. Statist. Psych.* **41** 193–208.

JÖRESKOG, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34** 183–202.

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

LEE, S. Y. and JENNRICH, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika* **44** 99–113.

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

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

DEPARTMENT OF STATISTICS
SNEDECOR HALL
IOWA STATE UNIVERSITY
AMES, IOWA 50011

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305