Asymptotic Theory of Overparameterized Structural Models

ALEXANDER SHAPIRO*

A theory of overparameterized structural models is presented. In such a model some "redundant" parameters are involved; the parameter vector is not identified, and the information matrix is not nonsingular. The minimum discrepancy function (MDF) test statistic is shown to have an asymptotic chi-squared distribution almost everywhere for a wide class of discrepancy functions. Asymptotic distribution properties of the MDF estimators are investigated. The factor analysis model is discussed as an example.

KEY WORDS: Covariance structures; Asymptotic normality; Asymptotic efficiency; Chi-squared test statistic; Generalized inverse; Discrepancy function.

1. INTRODUCTION

In recent years much attention has been paid to covariance or, more generally, to moment structural models appearing in behavioral, educational, biological, and so forth, studies. In an ever-increasing number of publications, methodological and theoretical aspects of moment structure analysis have come under intensive investigation (e.g., see Anderson 1969, 1973; Jöreskog 1970, 1981; Jöreskog and Sörbom 1979; Browne 1974, 1982, 1984; and Shapiro 1983a, 1985a and the references contained therein). In a moment structure, moments (usually of first and second order) are regarded as functions of a $q \times 1$ parameter vector $\boldsymbol{\theta}$ that belongs to a specified parameter space Θ. That is, the elements of the $p \times 1$ mean vector $\boldsymbol{\mu}$ and $p(p + 1)/2$ nonduplicated elements of the $p \times p$ covariance matrix $\boldsymbol{\Sigma}$, arranged in the $m \times 1$ vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)'$, are considered as functions of $\boldsymbol{\theta}$:

$$\xi_i = g_i(\boldsymbol{\theta}), \qquad i = 1, \ldots, m. \tag{1.1}$$

For instance, the well-known factor analysis model assumes that $\boldsymbol{\Sigma}$ is representable in the form

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}, \tag{1.2}$$

where $\boldsymbol{\Lambda}$ is the $p \times k$ matrix of factor loadings and $\boldsymbol{\Psi}$ is the diagonal matrix of the residual variances (e.g., Lawley and Maxwell 1971). Equation (1.2) may be considered as a covariance structural model with the parameter vector $\boldsymbol{\theta}$ composed of the elements of $\boldsymbol{\Lambda}$ and the diagonal elements of $\boldsymbol{\Psi}$ and $\boldsymbol{\xi}$ composed of $p(p + 1)/2$ nonduplicated elements of $\boldsymbol{\Sigma}$. For many other examples of covariance structural models the reader is referred to Jöreskog (1981), Browne (1982), and Bentler (1983). Multinomial parametric models are yet another example of moment structures, where the cell probabilities $\pi_1(\boldsymbol{\theta}), \ldots, \pi_p(\boldsymbol{\theta})$ are some specified functions of the parameter vector $\boldsymbol{\theta}$ (e.g., Rao 1973, sec. 6b).

For a given sample estimate $\hat{\mathbf{x}}$ of the population value $\boldsymbol{\xi}_0$ of $\boldsymbol{\xi}$, parameter estimates are usually obtained by minimizing a function measuring the discrepancy between $\hat{\mathbf{x}}$ and the fitted model $\boldsymbol{\xi} = \mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \ldots, g_m(\boldsymbol{\theta}))'$. Such estimation procedure is closely related to multivariate nonlinear least squares regression. In fact, if the discrepancy function is chosen to be a generalized least squares (GLS) distance function, then the minimum discrepancy function (MDF) estimation can be considered as a particular case of nonlinear regression. Indeed, let

$$\mathbf{y}_t = \mathbf{g}_t(\boldsymbol{\theta}) + \mathbf{e}_t, \qquad t = 1, \ldots, n,$$

be a nonlinear regression model with $\mathbf{y}_t \in \mathbf{R}^m$ and $\mathbf{g}_t(\boldsymbol{\theta}) \equiv \mathbf{g}(\boldsymbol{\theta})$ for $t = 1, \ldots, n$. The corresponding least squares estimators are obtained by minimizing the objective function

$$n^{-1} \sum_{t=1}^{n} (\mathbf{y}_t - \mathbf{g}(\boldsymbol{\theta}))'\mathbf{V}(\mathbf{y}_t - \mathbf{g}(\boldsymbol{\theta})),$$

where $\mathbf{V}$ is an $m \times m$ positive definite matrix [e.g., see a survey paper by Burguete, Gallant, and Souza (1982)]. Since $\mathbf{g}_t$ is independent of t, this objective function is equal to

$$(\bar{\mathbf{y}} - \mathbf{g}(\boldsymbol{\theta}))'\mathbf{V}(\bar{\mathbf{y}} - \mathbf{g}(\boldsymbol{\theta})) + n^{-1} \sum_{t=1}^{n} (\mathbf{y}_t - \bar{\mathbf{y}})'\mathbf{V}(\mathbf{y}_t - \bar{\mathbf{y}}),$$

where $\bar{\mathbf{y}} = (\mathbf{y}_1 + \cdots + \mathbf{y}_n)/n$ is the sample mean vector. Therefore we may instead minimize the function $(\bar{\mathbf{y}} - \mathbf{g}(\boldsymbol{\theta}))'\mathbf{V}(\bar{\mathbf{y}} - \mathbf{g}(\boldsymbol{\theta}))$. Here $\bar{\mathbf{y}}$ is the estimate $\hat{\mathbf{x}}$ and $(\mathbf{x} - \boldsymbol{\xi})'\mathbf{V}(\mathbf{x} - \boldsymbol{\xi})$ is a GLS discrepancy function (cf. Browne 1982, sec. 1.4).

Asymptotic distribution theory of moment structural models is well developed under two basic assumptions—namely, identifiability of the parameter vector $\boldsymbol{\theta}$ and nonsingularity (nondeficiency) of the information (Jacobian) matrix corresponding to the model $\boldsymbol{\xi} = \mathbf{g}(\boldsymbol{\theta})$ (see Browne 1974, 1984; Lee and Bentler 1980; and Shapiro 1983a). The aim of this article is to investigate *overparameterized* models in which some "redundant" parameters are involved. In such models the parameter vector is not identified; neither is the information matrix nonsingular. This situation happens quite often in practical applications. For example, in the factor analysis model (1.2) the covariance matrix $\boldsymbol{\Sigma}$ determines $\boldsymbol{\Lambda}$ up to rotation by a $k \times k$ orthogonal matrix.

A way of dealing with the problem of overparameterization is to restrict the parameter space—that is, to identify the model—by imposing equality constraints. For instance, $k(k - 1)/2$ equality constraints are needed to identify the factor analysis model (1.2). An extensive discussion of various identification conditions in factor analysis is given in Anderson and Rubin (1956, sec. 5). A general treatment of equality-constrained es-

* Alexander Shapiro is Senior Lecturer, Department of Mathematics and Applied Mathematics, University of South Africa, P.O. Box 392, Pretoria 0001, South Africa. The author wishes to express his gratitude to referees and an associate editor for helpful suggestions and correction of a technical error.

Journal of the American Statistical Association
March 1986, Vol. 81, No. 393, Theory and Methods

timation in the case of maximum likelihood was presented by Aitchison and Silvey (1958) and Silvey (1970). The problem with this approach, however, is that usually identification constraints are merely a convenient artifact for obtaining a unique solution and have no clear interpretation. Moreover, often introduction of the identification constraints unnecessarily complicates the calculations.

Here I propose an alternative approach based on techniques borrowed from the theory of generalized inverse matrices. I propose an asymptotic theory of the MDF estimators and the associated test statistic. I will demonstrate that for a wide class of discrepancy functions and structural models the MDF test statistic has an asymptotic chi-squared distribution for almost every population value of $\boldsymbol{\theta}$. The corresponding number of degrees of freedom is given by $m - r$, where r is the so-called *characteristic rank* of the model. Intuitively, r represents the number of "functionally independent" parametric variables and is equal to the rank of the corresponding Jacobian matrix. Asymptotic distribution of the MDF estimators will be given in terms of generalized inverse matrices. In this respect I follow the terminology and notation of Rao and Mitra (1971). Naturally such an asymptotic theory is interrelated with the theory of linear least squares estimation (e.g., Rao 1973, chap. 4).

2. PRELIMINARY DISCUSSION

Let $\boldsymbol{\xi}_0$ be the population value of the $m \times 1$ vector $\boldsymbol{\xi}$. The model (1.1) is said to hold if there exists a vector $\boldsymbol{\theta}_0$ from the parameter space Θ such that

$$\boldsymbol{\xi}_0 = \mathbf{g}(\boldsymbol{\theta}_0). \tag{2.1}$$

It will be supposed throughout that Θ is an *open* and *connected* subset of $\mathbf{R}^q$ and that the functions $g_1(\boldsymbol{\theta}), \ldots, g_m(\boldsymbol{\theta})$ are *twice continuously differentiable* on Θ. We say that $\boldsymbol{\theta}$ is *identified* (*locally identified*) at $\boldsymbol{\theta}_0$ if the vector $\boldsymbol{\theta}_0$ is unique (locally unique)—that is, if $\mathbf{g}(\boldsymbol{\theta}^*) = \mathbf{g}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}^* \in \Theta$ ($\boldsymbol{\theta}^*$ is in a neighborhood of $\boldsymbol{\theta}_0$) implies that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$.

Given a sample estimate $\hat{\mathbf{x}}$ of $\boldsymbol{\xi}_0$, one fits the model (1.1) by minimizing the discrepancy between $\hat{\mathbf{x}}$ and $\boldsymbol{\xi} = \mathbf{g}(\boldsymbol{\theta})$, which is measured by means of a certain real-valued function $F(\mathbf{x}, \boldsymbol{\xi})$ of two $m \times 1$ vector variables $\mathbf{x}$ and $\boldsymbol{\xi}$. Following the terminology proposed by Browne (1982, p. 80) I call $F(\mathbf{x}, \boldsymbol{\xi})$ the *discrepancy function*. This term has been used previously by Robertson (1972) in a context of multinomial estimation. The MDF estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is chosen to be a minimizer of the function $F(\hat{\mathbf{x}}, \mathbf{g}(\cdot))$ over the set Θ. The validity of the model is tested by means of the MDF test statistic $n\hat{F}$, where n is the sample size and $\hat{F}$ is the minimal value of the discrepancy function

$$\hat{F} = \min_{\boldsymbol{\theta} \in \Theta} F(\hat{\mathbf{x}}, \mathbf{g}(\boldsymbol{\theta})). \tag{2.2}$$

We assume throughout that the asymptotic distribution of $n^{1/2}(\hat{\mathbf{x}} - \boldsymbol{\xi}_0)$ is multivariate normal with zero mean and a certain covariance matrix $\boldsymbol{\Gamma}$. Usually this assumption can be justified by an application of the central limit theorem. Then it is known that under suitable regularity conditions and for an appropriate choice of F, $\hat{\boldsymbol{\theta}}$ is a consistent estimate of $\boldsymbol{\theta}_0$, $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal, and $n\hat{F}$ has an asymptotic chi-squared distribution with $m - q$ degrees of freedom. Two important regularity conditions have to be met here; notably, $\boldsymbol{\theta}$ must be identified at $\boldsymbol{\theta}_0$ and the $m \times q$ Jacobian matrix $\boldsymbol{\Delta}_0 = (\partial/\partial\boldsymbol{\theta}')\mathbf{g}(\boldsymbol{\theta}_0)$ must be of full rank q. So what happens if these conditions are not satisfied? Typically such a situation occurs when "redundant" parameters are involved. For instance, in the simplest case the mapping $\mathbf{g}(\boldsymbol{\theta})$ can be independent of, say, the last variable θ_q. Then the last column of the Jacobian matrix $\boldsymbol{\Delta}(\boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta}')\mathbf{g}(\boldsymbol{\theta})$ is identically zero and hence the rank of $\boldsymbol{\Delta}(\boldsymbol{\theta})$ is less than q for all $\boldsymbol{\theta}$. Of course more complicated situations leading to the deficient rank of $\boldsymbol{\Delta}$ may happen.

It is well known that nondeficiency of the rank of $\boldsymbol{\Delta}_0$ implies local identifiability of $\boldsymbol{\theta}$. The converse is also true under the following condition of local regularity (see Fisher 1966 and Rothenberg 1971).

Definition 2.1. A point $\boldsymbol{\theta}_0 \in \Theta$ is locally regular if the Jacobian matrix $\boldsymbol{\Delta}(\boldsymbol{\theta})$ has the same rank as $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}(\boldsymbol{\theta}_0)$ for every $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

It should be mentioned that rank deficiency of $\boldsymbol{\Delta}_0$ alone is not sufficient for local nonidentifiability of $\boldsymbol{\theta}$ and that the condition of local regularity is essential here. An illustrative counterexample associated with the factor analysis model can be found (e.g., in Shapiro and Browne 1983; a more sophisticated counterexample is in Shapiro 1985d).

Now I give a precise formulation of the basic concept of (local) overparameterization.

Definition 2.2. The model (1.1) is locally overparameterized at $\boldsymbol{\theta}_0$ if the rank r of $\boldsymbol{\Delta}_0$ is less than q and there exists a local diffeomorphism $\boldsymbol{\theta} = \mathbf{h}(\boldsymbol{\gamma})$ from a neighborhood of $\boldsymbol{\gamma}_0 \in \mathbf{R}^q$ to a neighborhood of $\boldsymbol{\theta}_0$, with $\boldsymbol{\theta}_0 = \mathbf{h}(\boldsymbol{\gamma}_0)$, such that the composite mapping $\mathbf{g}(\mathbf{h}(\boldsymbol{\gamma}))$ is independent of $q - r$ coordinates of $\boldsymbol{\gamma}$—say, the last $q - r$ coordinates—in a neighborhood of $\boldsymbol{\gamma}_0$. (A continuously differentiable mapping from a neighborhood of a q-dimensional vector space into itself is called *local diffeomorphism* if it is locally one-to-one and its inverse is also continuously differentiable.)

The intuitive meaning of local overparameterization is that if $\boldsymbol{\Delta}_0$ has a deficient rank r, then by a suitable *local reparameterization* (i.e., local diffeomorphism) precisely $q - r$ parameters become redundant, since locally the model does not depend on them. Needless to say, local overparameterization implies local nonidentification. Since $\boldsymbol{\theta} = \mathbf{h}(\boldsymbol{\gamma})$ is one-to-one, the minimization with respect to $\boldsymbol{\gamma}$ in a neighborhood of $\boldsymbol{\gamma}_0$ will be equivalent to the original optimization problem. Moreover, $\partial\mathbf{g}/\partial\boldsymbol{\gamma}' = (\partial\mathbf{g}/\partial\boldsymbol{\theta}')(\partial\mathbf{h}/\partial\boldsymbol{\gamma}')$, and since the Jacobian matrix of local diffeomorphism $\mathbf{h}(\boldsymbol{\gamma})$ must be nonsingular near $\boldsymbol{\gamma}_0$, the rank of $\boldsymbol{\Delta}$ is preserved by a local reparameterization.

The following result clarifies a relation between the preceding concepts of local regularity and local overparameterization and is known as the rank theorem (e.g., Dieudonné 1960).

Proposition 2.1. The model is locally overparameterized at $\boldsymbol{\theta}_0$ if and only if the rank r of $\boldsymbol{\Delta}_0$ is less than q and the point $\boldsymbol{\theta}_0$ is locally regular.

So far I have discussed *local* structure of the model (1.1) whereas the optimization procedure requires a *global* knowledge. Let us denote by Ξ the image of the mapping $\mathbf{g}(\boldsymbol{\theta})$; that is, $\Xi = \{\boldsymbol{\xi}:\boldsymbol{\xi} = \mathbf{g}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$. Then an equivalent formulation of (2.1) is that $\boldsymbol{\xi}_0 \in \Xi$. Furthermore, $\hat{F}$ can be expressed as the minimum of $F(\hat{\mathbf{x}}, \cdot)$ over the set Ξ. Therefore, not surprisingly, the asymptotic behavior of MDF estimators is closely related to the analytical structure of Ξ.

Let me discuss two examples. In the first example I consider the model $\mathbf{g}(\theta) = (\theta^2, \theta^2)'$, depending on the parameter $\theta \in \mathbf{R}$. Since $\mathbf{g}(\theta) = \mathbf{g}(-\theta)$, this model is locally but not globally identified at every $\theta \neq 0$. However, local structure of the set Ξ, which is a straight line, is simple and coincides for both values θ and $-\theta$. In the second example I define $\mathbf{g}(\theta) =$ (sin (4 arc tan θ), sin(2 arc tan θ))$'$, $\theta \in \mathbf{R}$. Here the model is locally and globally identified at every θ. The set Ξ has the form of a figure 8, and its local structure at the point $\boldsymbol{\xi} = (0, 0)'$ does not coincide with the local structure provided by the corresponding point $\theta = 0$. To rule out pathological cases such as in the second example, I introduce the following concept of (global) regularity.

Definition 2.3. A point $\boldsymbol{\theta}_0 \in \Theta$ is regular if $\boldsymbol{\theta}_0$ is locally regular and there exist neighborhoods $\mathcal{U}$ and $\mathcal{V}$ of $\boldsymbol{\theta}_0$ and $\boldsymbol{\xi}_0$, respectively, such that $\Xi \cap \mathcal{V} = \mathbf{g}(\mathcal{U})$.

In other words, regularity of $\boldsymbol{\theta}_0$ ensures that local structure of Ξ near $\boldsymbol{\xi}_0$ is provided by the mapping $\mathbf{g}(\boldsymbol{\theta})$ defined in a neighborhood of $\boldsymbol{\theta}_0$. Furthermore, the result of Proposition 2.1 shows that near $\boldsymbol{\theta}_0$ the model depends on exactly r "functionally independent" parameters. Consequently Ξ has at $\boldsymbol{\xi}_0$ the tangent space of dimension r. This tangent space is spanned by the column vectors of the matrix $\boldsymbol{\Delta}_0$. (For precise definitions of such relevant concepts from differential geometry as smooth manifold and tangent space, the reader is referred to Hirsch 1976.)

A natural question that arises from the preceding discussion is: How frequent are the regular points? It is not difficult to show that if the functions $g_i(\boldsymbol{\theta})$, $i = 1, \ldots, m$, are *analytic*, then almost every point of Θ is locally regular with the same rank r of the Jacobian matrix $\boldsymbol{\Delta}(\boldsymbol{\theta})$ (e.g., Shapiro 1983b). (A function is analytic on an open set if it can be expanded in power series in a neighborhood of every point of this set.) A much stronger result is that if the functions $g_i(\boldsymbol{\theta})$ are analytic on some compact set containing the space Θ [or when Θ is unbounded, $\mathbf{g}(\boldsymbol{\theta}) \to \infty$ as $\boldsymbol{\theta} \to \infty$], then almost every point of Θ is regular. This result is a consequence of some deep properties of analytic mappings (e.g., Federer 1969, sec. 3.4).

We see that an integer r that represents the rank of $\boldsymbol{\Delta}(\boldsymbol{\theta})$ almost everywhere is associated with a model defined by analytic functions. I call this number r the *characteristic rank* of the model; whenever talking about the characteristic rank, we may assume that the mapping $\mathbf{g}(\boldsymbol{\theta})$ is analytic. It follows from the preceding discussion that either $r = q$ and hence the model is *locally identified almost everywhere* or $r < q$ and then the model is *locally overparameterized almost everywhere*.

3. ASYMPTOTIC DISTRIBUTION OF THE MDF TEST STATISTIC

In this section I study asymptotic distribution of the MDF test statistic $n\hat{F}$. I give necessary and sufficient conditions for $n\hat{F}$ to have asymptotic chi-squared distribution. It will be shown that for a wide class of appropriately chosen discrepancy functions, $n\hat{F}$ is asymptotically chi-squared with $m - r$ degrees of freedom for every regular value of $\boldsymbol{\theta}_0 \in \Theta$.

We may suppose that the discrepancy function satisfies the following conditions (Browne 1982, p. 81):

1. $F(\mathbf{x}, \boldsymbol{\xi}) \geq 0$ for all $\mathbf{x}, \boldsymbol{\xi}$.
2. $F(\mathbf{x}, \boldsymbol{\xi}) = 0$ if and only if $\mathbf{x} = \boldsymbol{\xi}$.
3. F is twice continuously differentiable in $\mathbf{x}$ and $\boldsymbol{\xi}$.

The discrepancy functions currently in use in the analysis of covariance structures, maximum likelihood, and generalized least squares both satisfy conditions 1–3. For other examples of discrepancy functions in covariance structures satisfying conditions 1–3, see Swain (1975) and Browne (1984).

It can be shown that if conditions 1–3 hold, then the second-order Taylor approximation of F at $(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ is given by $(\mathbf{x} - \boldsymbol{\xi})'\mathbf{V}_0(\mathbf{x} - \boldsymbol{\xi})$. The $m \times m$ symmetric matrix $2\mathbf{V}_0$ can be defined alternatively by the Hessian matrices $\partial^2 F/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'$, $-\partial^2 F/\partial\mathbf{x}\partial\boldsymbol{\xi}'$, or $\partial^2 F/\partial\mathbf{x}\partial\mathbf{x}'$, calculated at the point $(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ (Shapiro 1985b). This result indicates that any discrepancy function satisfying conditions 1–3 has local structure of generalized least squares. Note that conditions 1 and 2 imply that function $F(\boldsymbol{\xi}_0, \cdot)$ attains its minimal value of zero at the point $\boldsymbol{\xi}_0$ and hence the matrix $\mathbf{V}_0$ is nonnegative definite.

For a given $\mathbf{x}$, I denote by $\boldsymbol{\xi}^*(\mathbf{x})$ a minimizer of $F(\mathbf{x}, \cdot)$ over the set Ξ and by $F_{\min}(\mathbf{x})$ the corresponding minimum

$$F_{\min}(\mathbf{x}) = \inf_{\boldsymbol{\xi} \in \Xi} F(\mathbf{x}, \boldsymbol{\xi}) = \inf_{\boldsymbol{\theta} \in \Theta} F(\mathbf{x}, \mathbf{g}(\boldsymbol{\theta})).$$

Because $\hat{F} = F_{\min}(\hat{\mathbf{x}})$, the asymptotic behavior of the MDF test statistic is closely related to analytical properties of the function $F_{\min}(\mathbf{x})$. Note that $\boldsymbol{\xi}_0$ is the unique minimizer of $F(\boldsymbol{\xi}_0, \cdot)$ and hence $\boldsymbol{\xi}^*(\boldsymbol{\xi}_0) = \boldsymbol{\xi}_0$. Strictly speaking, however, conditions 1–3 are not sufficient for continuity of the mapping $\boldsymbol{\xi}^*(\cdot)$ and hence consistency of the MDF estimator $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}^*(\hat{\mathbf{x}})$. (The MDF estimator $\hat{\boldsymbol{\xi}}$ will be discussed in detail in the next section.) In this respect we need to impose the following technical condition.

4. There exist positive constants M and ε such that $F(\mathbf{x}, \boldsymbol{\xi}) \geq \varepsilon$ whenever $\|\mathbf{x} - \boldsymbol{\xi}\| \geq M$.

Condition (4) prevents $\boldsymbol{\xi}^*(\mathbf{x})$ from going to infinity as $\mathbf{x}$ approaches $\boldsymbol{\xi}_0$. Together with 1–3, this implies that $\boldsymbol{\xi}^*(\mathbf{x})$ tends to $\boldsymbol{\xi}_0$ as $\mathbf{x}$ tends to $\boldsymbol{\xi}_0$. For more details and a discussion, see Shapiro (1983a, 1984). Usually $F(\mathbf{x}, \boldsymbol{\xi})$ tends to infinity as $\|\mathbf{x} - \boldsymbol{\xi}\| \to \infty$, which implies condition 4, so condition 4 is a mild restriction on the discrepancy function.

Now consider the problem of minimizing $F(\mathbf{x}, \cdot)$ over Ξ. As I have mentioned earlier, second-order approximation to F is given by the quadratic function $(\mathbf{x} - \boldsymbol{\xi})'\mathbf{V}(\mathbf{x} - \boldsymbol{\xi})$ and the set Ξ is approximated by the tangent space given by the column space of the matrix $\boldsymbol{\Delta}$. (In order to simplify notation, I drop the null subscript and write $\boldsymbol{\Delta}$ and $\mathbf{V}$ instead of $\boldsymbol{\Delta}_0$ and $\mathbf{V}_0$.) Then it is possible to show that the first-order approximation of $\boldsymbol{\xi}^*(\cdot)$ at $\boldsymbol{\xi}_0$ is given by the orthogonal (with respect to the seminorm $\|\mathbf{x}\| = (\mathbf{x}'\mathbf{V}\mathbf{x})^{1/2}$) projection onto the tangent space to Ξ. A formal proof of this fact is given in Abatzoglou (1978) and Shapiro (1985c), and for a detailed discussion of orthogonal projections the reader is referred to Rao and Mitra (1971, chap.

5). The seminorm $(\mathbf{x}'\mathbf{V}\mathbf{x})^{1/2}$ corresponds to the inner product $(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{V}\mathbf{y}$ and becomes a norm if the matrix $\mathbf{V}$ is positive definite. I need here the less restrictive assumption that $\mathbf{V}$ is positive definite on the linear space spanned by the columns of $\boldsymbol{\Delta}$. Since $\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}$ is nonnegative definite, this is equivalent to the condition rank $(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})$ = rank $(\boldsymbol{\Delta})$. Of course, if $\mathbf{V}$ is positive definite, then this condition holds automatically.

An explicit expression for the orthogonal projection matrix is $\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-}\boldsymbol{\Delta}'\mathbf{V}$ (Rao and Mitra 1971, p. 111). This expression gives the Jacobian matrix of $\boldsymbol{\xi}^*(\cdot)$ at $\boldsymbol{\xi}_0$. Then second-order derivatives of $F_{\min}$ can be easily calculated by the chain rule (see Shapiro 1985c). I summarize this discussion in the following proposition.

Proposition 3.1. Suppose that the point $\boldsymbol{\theta}_0$ is regular, the discrepancy function F satisfies conditions 1–4, and rank $(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})$ = rank $(\boldsymbol{\Delta})$. Then (a) the minimizer $\boldsymbol{\xi}^*(\mathbf{x})$ is uniquely defined and continuously differentiable in a neighborhood of $\boldsymbol{\xi}_0$, with

$$(\partial/\partial\mathbf{x}')\boldsymbol{\xi}^*(\mathbf{x}_0) = \boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-}\boldsymbol{\Delta}'\mathbf{V}, \tag{3.1}$$

and (b) the min function $F_{\min}(\mathbf{x})$ is twice continuously differentiable at $\boldsymbol{\xi}_0$ with the first-order partial derivatives vanishing at $\boldsymbol{\xi}_0$ and the (half) Hessian matrix of second-order partial derivatives $\mathbf{U} = \frac{1}{2}(\partial^2/\partial\mathbf{x}\partial\mathbf{x}')F_{\min}(\boldsymbol{\xi}_0)$ is given by

$$\mathbf{U} = \mathbf{V} - \mathbf{V}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-}\boldsymbol{\Delta}'\mathbf{V}. \tag{3.2}$$

If the matrix $\mathbf{V}$ is nonsingular, then the right-side expression in (3.2) can be simplified as follows: Let $\boldsymbol{\Phi}$ be an orthogonal complement of $\boldsymbol{\Delta}$; that is, $\boldsymbol{\Phi}$ is an $m \times (m - r)$ matrix of full rank $m - r$ such that $\boldsymbol{\Phi}'\boldsymbol{\Delta} = \mathbf{0}$. Then by a known matrix identity (e.g., Rao 1973, p. 77),

$$\mathbf{U} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'. \tag{3.3}$$

The second part of Proposition 3.1 and the delta theorem (e.g., Rao 1973, p. 388) immediately imply that

$$n\hat{F} = [n^{1/2}(\hat{\mathbf{x}} - \boldsymbol{\xi}_0)]'\mathbf{U}[n^{1/2}(\hat{\mathbf{x}} - \boldsymbol{\xi}_0)] + o_p(1),$$

and hence the asymptotic distribution of the MDF test statistic $n\hat{F}$ is the same as the distribution of the quadratic form $\mathbf{z}'\mathbf{U}\mathbf{z}$ with $\mathbf{z} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$. Distributions of quadratic forms are well discussed. It is known that $\mathbf{z}'\mathbf{U}\mathbf{z}$ has (central) chi-squared distribution if and only if the following condition holds (Rao and Mitra 1971, p. 171):

$$\boldsymbol{\Gamma}\mathbf{U}\boldsymbol{\Gamma}\mathbf{U}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\mathbf{U}\boldsymbol{\Gamma}, \tag{3.4}$$

in which case the number ν of degrees of freedom is

$$\nu = \operatorname{tr} \mathbf{U}\boldsymbol{\Gamma}. \tag{3.5}$$

If the matrix $\mathbf{V}$ is nonsingular, then we can use an equivalent expression (3.3) for the matrix $\mathbf{U}$ and (3.4) becomes

$$\boldsymbol{\Gamma}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma}.$$

This equation holds if (and when $\boldsymbol{\Gamma}$ is nonsingular only if)

$$\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi} = \boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Phi}. \tag{3.6}$$

In particular it holds if $\mathbf{V} = \boldsymbol{\Gamma}^{-1}$. This fact has been widely exploited for constructing MDF test statistics with an asymptotic chi-squared distribution (see Browne 1974, 1984). Equation (3.6) shows, however, that a considerably broader class of discrepancy functions has this property. One can verify that condition (3.6) is equivalent to the following expression for the matrix $\mathbf{V}$:

$$\mathbf{V} = (\boldsymbol{\Gamma} + \boldsymbol{\Delta}\mathbf{C}' + \mathbf{C}\boldsymbol{\Delta}')^{-1}, \tag{3.7}$$

where $\mathbf{C}$ is an arbitrary $m \times q$ matrix.

Under condition (3.6) we can express the number of degrees of freedom ν as

$$\nu = \operatorname{tr} \mathbf{U}\boldsymbol{\Gamma} = \operatorname{tr} \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma} = \operatorname{tr}(\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{\Gamma}\boldsymbol{\Phi} = m - \operatorname{rank}(\boldsymbol{\Delta}).$$

I came to the following result.

Proposition 3.2. Let the matrices $\mathbf{V}$ and $\boldsymbol{\Gamma}$ be nonsingular and the point $\boldsymbol{\theta}_0$ be regular. Then the MDF test statistic has asymptotic chi-squared distribution if and only if condition (3.6), or equivalently condition (3.7), holds. In the last case the number of degrees of freedom is given by m − rank $(\boldsymbol{\Delta})$.

Structural models one uses in practice are usually defined by "well behaved" analytic functions (often these functions are simply polynomials). Then almost every population point $\boldsymbol{\theta}_0$ is regular and Proposition 3.2 implies that for an appropriately chosen discrepancy function, the MDF test statistic is chi-squared with $m - r$ degrees of freedom almost everywhere. It is worthwhile to note that I do not claim and it is possibly not correct that the model is (globally) identified almost everywhere. For instance, in the first example discussed in Section 2, the model is (globally) nonidentified at every $\theta \neq 0$. The set Ξ, however, has a simple "smooth" structure almost everywhere. Such behavior is typical for the model defined by analytic functions.

Still what happens at nonregular points? As I have mentioned earlier, it is possible for $\boldsymbol{\Delta}_0$ to have deficient rank r less than q and yet identify the model at $\boldsymbol{\theta}_0$. To some extent this case (with $r = q - 1$) has been discussed by Sargan (1983). Often this situation occurs when the model is reparameterized in order to avoid boundary solutions. For instance, the inequality constraints $\theta_i \geq 0$ $(i = 1, \ldots, q)$ can be released by the reparameterization $\theta_i \to \theta_i^2$. If the population vector $\boldsymbol{\theta}_0$ has, say, s zero coordinates, however, then the Jacobian matrix $\boldsymbol{\Delta}_0$ corresponding to the reparameterized model has deficient rank $r \leq q - s$. A detailed study of the boundary case is given in Shapiro (1985a). It is shown that if $\boldsymbol{\theta}_0$ is a boundary point of Θ, then the MDF test statistic has asymptotic distribution that is a mixture of chi-squared distributions.

It is remarkable that for an appropriately chosen discrepancy function, the asymptotic distribution of $n\hat{F}$ can be overestimated by the chi-squared distribution regardless of any regularity conditions. The precise statement is given in the following proposition.

Proposition 3.3. Let the discrepancy function satisfy conditions 1–3, and let condition (3.6) hold. Then

$$\limsup_{n\to\infty} \Pr\{n\hat{F} \geq c\} \leq \Pr\{\chi^2_\nu \geq c\},$$

where χ^2_ν is a chi-squared variable with $\nu = m$ − rank $(\boldsymbol{\Delta}_0)$ degrees of freedom.

Proposition 3.3 shows that upper tail probabilities of the MDF test statistic are always asymptoticaly overestimated by the corresponding chi-squared distribution. The result follows immediately from the algebraic discussion above and the fact that

$$F_{\min}(\mathbf{x}) \le (\mathbf{x} - \boldsymbol{\xi}_0)'\mathbf{U}(\mathbf{x} - \boldsymbol{\xi}_0) + o(\|\mathbf{x} - \boldsymbol{\xi}_0\|^2).$$

A formal proof of the last inequality can be found in Shapiro (1985c, theorem 3.1 and corollary 3.1).

4. ASYMPTOTIC DISTRIBUTION OF THE MDF ESTIMATORS

In this section I discuss asymptotic distribution of the MDF estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\xi}}$. The main difficulty here is that if the model is overparameterized, then the estimator $\hat{\boldsymbol{\theta}}$, as well as the population value $\boldsymbol{\theta}_0$, is not uniquely defined. Therefore it is natural to take a representative and to consider a class of possible estimators $\hat{\boldsymbol{\theta}}$. As we shall see, this corresponds to a different choice of the generalized inverse. But let us first consider the estimator $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}^*(\hat{\mathbf{x}})$, which in covariance structures represents the so-called reproduced covariance matrix. This estimator is unique and Proposition 3.1(a) together with the delta theorem implies the following result.

Proposition 4.1. Suppose that assumptions of Proposition 3.1 hold. Then $\hat{\boldsymbol{\xi}}$ is a consistent estimator of $\boldsymbol{\xi}_0$ and $n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$ has asymptotically normal distribution with zero mean and covariance matrix $\mathbf{P\Gamma P}'$, where $\mathbf{P}$ is the projection matrix $\mathbf{P} = \boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V}$.

Now we consider the MDF estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$. I denote by $\boldsymbol{\theta}^*(\mathbf{x})$ a minimizer of $F(\mathbf{x}, \mathbf{g}(\cdot))$ over Θ. Clearly $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^*(\hat{\mathbf{x}})$. If the model is overparameterized, then the minimizer $\boldsymbol{\theta}^*$, and hence the estimator $\hat{\boldsymbol{\theta}}$, is not unique. We always have that $\boldsymbol{\xi}^*(\mathbf{x}) = \mathbf{g}(\boldsymbol{\theta}^*(\mathbf{x}))$, however. Suppose for the moment that the minimizer $\boldsymbol{\theta}^*$ exists and is a differentiable function of $\mathbf{x}$. Then differentiating the last equation at $\mathbf{x} = \boldsymbol{\xi}_0$, we obtain that the $q \times m$ Jacobian matrix $\mathbf{J}$ of $\boldsymbol{\theta}^*(\mathbf{x})$ at $\boldsymbol{\xi}_0$ must satisfy

$$\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V} = \boldsymbol{\Delta}\mathbf{J}. \quad (4.1)$$

I remind the reader that the left-side expression in (4.1) represents the Jacobian matrix $\partial\boldsymbol{\xi}^*/\partial\mathbf{x}'$ at $\mathbf{x} = \boldsymbol{\xi}_0$. Premultiplying both sides of (4.1) with $\boldsymbol{\Delta}'\mathbf{V}$ and using matrix identity

$$(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}' = \boldsymbol{\Delta}'$$

(e.g., Rao and Mitra 1971, lemma 2.2.6(c)), we obtain that $\mathbf{J}$ must satisfy

$$\boldsymbol{\Delta}'\mathbf{V} = (\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})\mathbf{J}. \quad (4.2)$$

It is not difficult to show that if the point $\boldsymbol{\theta}_0$ is regular, then the converse is also true. Namely, for every matrix $\mathbf{J}$ satisfying (4.2), there exists the minimizer $\boldsymbol{\theta}^*(\mathbf{x})$ in a neighborhood of $\boldsymbol{\theta}_0$ such that $(\partial/\partial\mathbf{x}')\boldsymbol{\theta}^*(\boldsymbol{\xi}_0) = \mathbf{J}$. (The proof of this fact is given in Appendix A.) Because of the delta theorem, the covariance matrix $\boldsymbol{\Pi}$ associated with the estimator $\hat{\boldsymbol{\theta}}$ is given by $\boldsymbol{\Pi} = \mathbf{J\Gamma J}'$. Consequently Equation (4.2) implies that

$$(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})\boldsymbol{\Pi}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}) = \boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Gamma}\mathbf{V}\boldsymbol{\Delta}. \quad (4.3)$$

I came to the following result.

Proposition 4.2. Suppose that regularity assumptions of Proposition 3.1 hold. Then for every nonnegative definite matrix $\boldsymbol{\Pi}$ satisfying (4.3), there exists a consistent MDF estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ such that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ has asymptotically normal distribution with zero mean and covariance matrix $\boldsymbol{\Pi}$.

A matrix $\mathbf{J}$ satisfying Equation (4.2) is called a $\mathbf{V}$ least squares g inverse of $\boldsymbol{\Delta}$ (Rao and Mitra 1971, p. 49). One choice of $\mathbf{J}$ is $\mathbf{J} = (\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V}$. In this case the covariance matrix $\boldsymbol{\Pi}$ becomes

$$\boldsymbol{\Pi} = (\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Gamma}\mathbf{V}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-. \quad (4.4)$$

The general solution is more complicated and is given by

$$\mathbf{J} = (\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V} + [\mathbf{I} - (\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}]\mathbf{Q}, \quad (4.5)$$

where $\mathbf{Q}$ is an arbitrary $q \times m$ matrix (Rao and Mitra 1971, p. 49).

If $\mathbf{V}$ is nonsingular, then $(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^-\boldsymbol{\Delta}'\mathbf{V}$ is a reflexive g inverse of $\boldsymbol{\Delta}$ and hence has the same rank as $\boldsymbol{\Delta}$ (Rao and Mitra 1971, theorem 3.2.2 and lemma 2.5.1). It can be shown that the corresponding expression in the right side of (4.4) represents the class of covariance matrices that can be obtained under identification constraints. In other words, for every choice of the generalized inverse of $\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}$, there exist $q - r$ identification constraints such that the corresponding covariance matrix is given by the right side of (4.4) and vice versa. A proof of this result is given in Appendix B.

Sometimes a certain function $\eta = f(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ is of interest rather than the parameter vector $\boldsymbol{\theta}$ itself. The new variable η is uniquely defined if it is a function of $\boldsymbol{\xi}$; that is, $f(\boldsymbol{\theta})$ can be represented as a composite function $f(\boldsymbol{\theta}) = t(\boldsymbol{\xi}) = t(\mathbf{g}(\boldsymbol{\theta}))$. The converse is also true at least locally. By analogy with linear models (e.g., Searle 1971, p. 180), I call such a parameter η an *estimable* parameter. Now consider the estimator $\hat{\eta} = f(\hat{\boldsymbol{\theta}})$ of $\eta_0 = f(\boldsymbol{\theta}_0)$. Then from Proposition 4.2 and the delta theorem, $n^{1/2}(\hat{\eta} - \eta_0)$ is asymptotically normal with zero mean and variance $\mathbf{b}'\boldsymbol{\Pi}\mathbf{b}$, where $\boldsymbol{\Pi}$ satisfies (4.3) and $\mathbf{b}$ is the gradient vector $\mathbf{b} = (\partial/\partial\boldsymbol{\theta})f(\boldsymbol{\theta}_0)$. If η is (locally) unique, then f can be represented as the composite function and $\mathbf{b} = \boldsymbol{\Delta}'\mathbf{a}$, where $\mathbf{a}$ is the gradient vector $\mathbf{a} = \partial t/\partial\mathbf{g}$. Hence in this case $\mathbf{b}$ belongs to the linear space $\mathcal{M}(\boldsymbol{\Delta}')$ generated by the column vectors of $\boldsymbol{\Delta}'$. Then $\mathbf{b}'\boldsymbol{\Pi}\mathbf{b}$ is independent of a particular solution of Equation (4.3) and $\boldsymbol{\Pi}$ can be taken in the form (4.4) with an arbitrary choice of the generalized inverse of $\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}$.

Often the function $f(\boldsymbol{\theta})$ is chosen to be a coordinate function of $\boldsymbol{\theta}$. In other words, we are interested in vector $\boldsymbol{\theta}_1$ of, say, the first s coordinates of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_1 = (\theta_1, \ldots, \theta_s)'$, whereas the last $q - s$ coordinates $\boldsymbol{\theta}_2 = (\theta_{s+1}, \ldots, \theta_q)'$ are considered as nuisance parameters. Let $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$ and $\mathbf{J}' = (\mathbf{J}_1', \mathbf{J}_2')$ be the corresponding partition of the matrices $\boldsymbol{\Delta}$ and $\mathbf{J}$, respectively. Then using formulas for the generalized inverse of a partitioned matrix we obtain

$$\mathbf{J}_1 = (\boldsymbol{\Delta}_1'\mathbf{U}_2\boldsymbol{\Delta}_1)^-\boldsymbol{\Delta}_1'\mathbf{U}_2, \quad (4.6)$$

where

$$\mathbf{U}_2 = \mathbf{V} - \mathbf{V}\boldsymbol{\Delta}_2(\boldsymbol{\Delta}_2'\mathbf{V}\boldsymbol{\Delta}_2)^-\boldsymbol{\Delta}_2'\mathbf{V} \quad (4.7)$$

(e.g., Pringle and Rayner 1971, sec. 3.3). Consequently the covariance matrix $\boldsymbol{\Pi}_{11}$ corresponding to the estimator $\hat{\boldsymbol{\theta}}_1$ must satisfy the equation

$$(\boldsymbol{\Delta}_1'\mathbf{U}_2\boldsymbol{\Delta}_1)\boldsymbol{\Pi}_{11}(\boldsymbol{\Delta}_1'\mathbf{U}_2\boldsymbol{\Delta}_1) = \boldsymbol{\Delta}_1'\mathbf{U}_2\boldsymbol{\Gamma}\mathbf{U}_2\boldsymbol{\Delta}_1. \quad (4.8)$$

Note that for nonsingular $\mathbf{V}$ the matrix $\mathbf{U}_2$ can be given in an equivalent form as follows (compare with expression (3.3)):

$$\mathbf{U}_2 = \mathbf{\Phi}_2(\mathbf{\Phi}_2'\mathbf{V}^{-1}\mathbf{\Phi}_2)^{-1}\mathbf{\Phi}_2', \tag{4.9}$$

where $\mathbf{\Phi}_2$ is an orthogonal complement of Δ_2.

Equation (4.8) can be applied straightforwardly without going to the intermediate step of identification constraints. I demonstrate advantages of this approach in Section 6 while investigating an example of the factor analysis model. Note that $\mathbf{\Pi}_{11}$ is defined uniquely by (4.8) if and only if the matrix $\mathbf{\Delta}_1'\mathbf{U}_2\mathbf{\Delta}_1$ is nonsingular. This in turn is equivalent to the condition that the coordinate vectors $\mathbf{e}_i = \partial\theta_i/\partial\boldsymbol{\theta}$ $(i = 1, \ldots, s)$ belong to the space $\mathfrak{M}(\mathbf{\Delta}')$. As I have mentioned earlier, this is a necessary condition for local identification of θ_i $(i = 1, \ldots, s)$.

5. ASYMPTOTIC EFFICIENCY

In this section I address the question of how to choose the discrepancy function in order to obtain the "best" MDF estimators in the sense of having minimum asymptotic variance. As we shall see, such an asymptotic efficiency is guaranteed if, but not only if, $\mathbf{V} = \alpha\mathbf{\Gamma}^{-1}$ $(\alpha > 0)$ (cf. Browne 1974, proposition 3). I show that a more general result holds. It will be demonstrated that if

$$\mathbf{V} = (\mathbf{\Gamma} + \mathbf{\Delta}\mathbf{D}\mathbf{\Delta}')^{-1}, \tag{5.1}$$

where $\mathbf{D}$ is an arbitrary $q \times q$ symmetric matrix, then the MDF estimators are *asymptotically efficient* and the MDF test statistic is *asymptotically chi-squared.*

The basic condition for asymptotic efficiency is given in the following proposition (for the proof of this result, see Appendix C).

Proposition 5.1. Let the matrices $\mathbf{V}$ and $\mathbf{\Gamma}$ be nonsingular. Then for any nonnegative definite matrix $\mathbf{\Pi}$ satisfying (4.3) and all vectors $\mathbf{b} \in \mathfrak{M}(\mathbf{\Delta}')$, the following inequality holds:

$$\mathbf{b}'\mathbf{\Pi}\mathbf{b} \geq \mathbf{b}'(\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta})^{-}\mathbf{b}. \tag{5.2}$$

It should be mentioned that, since $\mathbf{b} \in \mathfrak{M}(\mathbf{\Delta}')$, the left and right sides of (5.2) are independent of a particular choice of $\mathbf{\Pi}$ satisfying (4.3) and of the generalized inverse, respectively. One may note that $\mathbf{b}'\mathbf{\Pi}\mathbf{b}$ represents the asymptotic variance associated with the estimator η, which is a (uniquely defined) function of $\hat{\boldsymbol{\theta}}$. Inequality (5.2) indicates that the minimum of this variance is attained if and only if

$$\mathbf{\Pi} = (\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta})^{-} \tag{5.3}$$

is a solution of Equation (4.3). In other words, the MDF estimators are asymptotically efficient if and only if the following equality holds:

$$\mathbf{\Delta}'\mathbf{V}\mathbf{\Delta}(\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta})^{-}\mathbf{\Delta}'\mathbf{V}\mathbf{\Delta} = \mathbf{\Delta}'\mathbf{V}\mathbf{\Gamma}\mathbf{V}\mathbf{\Delta}. \tag{5.4}$$

Note that the left side of (5.4) is independent of a particular choice of the generalized inverse.

As I have mentioned earlier, my asymptotic theory is closely related to the least-squares theory of linear models. In fact on some occasions we solve the same matrix equations. Therefore we can apply straightforwardly some results from linear theory in order to obtain the corresponding asymptotics. In particular, in the case of nonsingular $\mathbf{\Gamma}$, the general solution of (5.4) can be written (in my notation) as follows (see Rao and Mitra 1971, p. 160, corollaries 1 and 2):

$$\mathbf{V}^{-1} = \alpha\mathbf{\Gamma} + \mathbf{\Delta}\mathbf{D}\mathbf{\Delta}' + \mathbf{\Gamma}\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}'\mathbf{\Gamma}, \tag{5.5}$$

where $\mathbf{D}$ and $\mathbf{Q}$ are arbitrary symmetric matrices and α is an arbitrary positive scalar. Clearly (5.5) and (3.6) (i.e., asymptotic efficiency and asymptotic chi-squaredness) hold together if and only if condition (5.1) is satisfied.

Under condition (5.5) the class of matrices $\mathbf{\Pi}$ satisfying (4.3) coincides with the class of generalized inverse matrices of $\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta}$. In the case of *identification constraints,* the corresponding covariance matrix $\mathbf{\Pi}$ must satisfy the additional condition of rank $(\mathbf{\Pi})$ = rank $(\mathbf{\Delta})$. Therefore these matrices are given by the nonnegative definite *reflexive* g-inverse matrices of $\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta}$.

Finally, let $\boldsymbol{\theta}$ be partitioned $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$ and suppose that (5.3) holds. Then the covariance matrix $\mathbf{\Pi}_{11}$ associated with the estimator $\hat{\boldsymbol{\theta}}_1$ can be calculated using formulas for the generalized inverse of a partitioned matrix. That is, $\mathbf{\Pi}_{11}$ is given by the nonnegative definite g-inverse matrices of the matrix

$$\mathbf{\Delta}_1'\mathbf{\Gamma}^{-1}\mathbf{\Delta}_1 - \mathbf{\Delta}_1'\mathbf{\Gamma}^{-1}\mathbf{\Delta}_2(\mathbf{\Delta}_2'\mathbf{\Gamma}^{-1}\mathbf{\Delta}_2)^{-}\mathbf{\Delta}_2'\mathbf{\Gamma}^{-1}\mathbf{\Delta}_1. \tag{5.6}$$

An equivalent expression for the matrix in (5.6) is

$$\mathbf{\Delta}_1'\mathbf{\Phi}_2(\mathbf{\Phi}_2'\mathbf{\Gamma}\mathbf{\Phi}_2)^{-1}\mathbf{\Phi}_2'\mathbf{\Delta}_1, \tag{5.7}$$

where $\mathbf{\Phi}_2$ is an orthogonal complement of $\mathbf{\Delta}_2$.

6. FACTOR ANALYSIS MODEL

In this section I briefly discuss applications of the developed theory to the factor analysis model (1.2). The characteristic rank r of the factor analysis model can be calculated to be $r = p + pk - k(k - 1)/2$ if this integer is less than or equal to $m = p(p + 1)/2$ and $r = m$ otherwise (see Shapiro 1982, 1983b). It follows that under suitable regularity conditions and for an appropriately chosen discrepancy function, the MDF test statistic has asymptotic chi-squared distribution with $m - r = [(p - k)^2 - (p + k)]/2$ degrees of freedom. Previously this result has been obtained essentially under two regularity conditions—namely, (global) identification of $\mathbf{\Psi}_0$ and nonsingularity of the corresponding information matrix (see Lawley and Maxwell 1971 and Jöreskog and Goldberger 1972). My derivations imply that this result holds *generically* in the sense that the MDF test statistic is asymptotically chi-squared for almost every value of $\mathbf{\Lambda}_0$ and $\mathbf{\Psi}_0$. It is not claimed that $\mathbf{\Psi}$ is globally identified almost everywhere. Instead, I assert that the image set Ξ is a smooth manifold in a neighborhood of $\boldsymbol{\xi}_0$ for almost every $\mathbf{\Lambda}_0$ and $\mathbf{\Psi}_0$. (A subset of $\mathbf{R}^m$ is called a smooth manifold of dimension r if locally it has the differential structure of the Euclidean space $\mathbf{R}^r$. For a precise definition see, e.g., Hirsch 1976, p. 12.)

Now let us consider MDF estimators of $\mathbf{\Lambda}_0$ and $\mathbf{\Psi}_0$. Naturally the parameter vector consists of two parts associated with $\mathbf{\Lambda}$ and $\mathbf{\Psi}$, and I treat them separately. Let $\mathbf{S}$ be the $p \times p$ sample covariance matrix giving an unbiased estimate of $\mathbf{\Sigma}$. I denote by vec($\mathbf{S}$) and vecps($\mathbf{S}$) column vectors formed from the elements of $\mathbf{S}$ taken columnwise and from the nonduplicated elements, respectively. Since the matrix $\mathbf{S}$ is symmetric, the $m \times 1$ vector $\mathbf{s}$ = vecps($\mathbf{S}$) can be expressed in terms of the $p^2 \times$

1 vector $\mathbf{c} = \text{vec}(\mathbf{S})$ as follows: $\mathbf{s} = \mathbf{K}_p'\mathbf{c}$, where matrix $\mathbf{K}_p$ is of order $p^2 \times m$. We also need the $p^2 \times p$ matrix $\mathbf{H}_p$ defined by

$$\text{diag}(\mathbf{S}) = \mathbf{H}_p'\text{vec}(\mathbf{S}).$$

For a detailed discussion of the transition matrices $\mathbf{K}_p$ and $\mathbf{H}_p$ and the associated matrix calculus, see Browne (1974) and Nel (1980, sec. 6.1). I also define $\boldsymbol{\lambda}_0 = \text{vec}(\Lambda_0)$, $\boldsymbol{\psi}_0 = \text{diag}(\boldsymbol{\Psi}_0)$, and consider the corresponding MDF estimators $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\psi}}$.

Under mild assumptions it follows from the central limit theorem that $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ is asymptotically normal with zero mean and a certain $m \times m$ covariance matrix $\boldsymbol{\Gamma}$. In particular, if the sample is drawn from a normally distributed population, then

$$\boldsymbol{\Gamma} = 2\mathbf{K}_p'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{K}_p \tag{6.1}$$

(see Browne 1974). I suppose that the $m \times m$ matrix $\mathbf{V}$ associated with the discrepancy function is of the form

$$\mathbf{V} = \tfrac{1}{2}\mathbf{K}_p^-(\mathbf{B} \otimes \mathbf{B})\mathbf{K}_p^{-\prime}, \tag{6.2}$$

where $\mathbf{B}$ is a $p \times p$ positive definite matrix and $\mathbf{K}_p^- = (\mathbf{K}_p'\mathbf{K}_p)^{-1}\mathbf{K}_p'$. Such a choice of $\mathbf{V}$ is relevant for the GLS and maximum likelihood discrepancy functions (see Browne 1974 for details). Note that if $\mathbf{B} = \boldsymbol{\Sigma}^{-1}$, then $\mathbf{V}$ is the inverse of the covariance matrix $\boldsymbol{\Gamma}$ given in (6.1).

For the rest of this section I suppose that the characteristic rank r is less than m; hence the number $m - r$, of degrees of freedom, is positive. I also suppose that the point $(\boldsymbol{\lambda}_0, \boldsymbol{\psi}_0)$ is regular. As I have mentioned earlier, the parameter vector $\boldsymbol{\lambda}$ is locally unidentified whenever k is greater than one, whereas $\boldsymbol{\psi}$ is locally identified almost everywhere. A thorough discussion of the identification problem in factor analysis is given in Anderson and Rubin (1956) and more recently in Shapiro (1985d).

The MDF estimator $\hat{\boldsymbol{\psi}}$ is uniquely defined near $\boldsymbol{\psi}_0$, and $n^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ is asymptotically normal. In order to calculate the corresponding covariance matrix, let us consider the $m \times pk$ and $m \times p$ Jacobian matrices $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ of vecps$(\boldsymbol{\Lambda}\boldsymbol{\Lambda}')$ and vecps$(\boldsymbol{\Psi})$, respectively. It can be seen that $\boldsymbol{\Delta}_2$ is given by $\boldsymbol{\Delta}_2 = \mathbf{K}_p'\mathbf{H}_p$. It is also not difficult to calculate the Jacobian matrix $\boldsymbol{\Delta}_1$ (see Shapiro 1983a, p. 73). At this stage, however, we are rather interested in its orthogonal complement $\boldsymbol{\Phi}_1$:

$$\boldsymbol{\Phi}_1 = \mathbf{K}_p^-(\mathbf{F} \otimes \mathbf{F})\mathbf{K}_{p-k},$$

where $\mathbf{F}$ is an orthogonal complement of Λ_0 (Shapiro 1983a, p. 73). After some calculations, we obtain the matrix $\mathbf{U}_1 = \boldsymbol{\Phi}_1(\boldsymbol{\Phi}_1'\mathbf{V}^{-1}\boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Phi}_1'$ in the form

$$\mathbf{U}_1 = \tfrac{1}{2}\mathbf{K}_p^-(\mathbf{Z} \otimes \mathbf{Z})\mathbf{K}_p^{-\prime},$$

with $\mathbf{Z} = \mathbf{F}(\mathbf{F}'\mathbf{B}^{-1}\mathbf{F})^{-1}\mathbf{F}'$. It follows that $\boldsymbol{\Delta}_2'\mathbf{U}_1\boldsymbol{\Delta}_2 = \tfrac{1}{2}\mathbf{Z}^{(2)}$, where $\mathbf{Z}^{(2)}$ denotes the matrix whose elements are squares of the corresponding elements of $\mathbf{Z}$. Then from (4.8) and (4.9) we obtain that the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ is given by

$$(\mathbf{Z}^{(2)})^{-1}\mathbf{H}_p'(\mathbf{Z} \otimes \mathbf{Z})\mathbf{K}_p^{-\prime}\boldsymbol{\Gamma}\mathbf{K}_p^-(\mathbf{Z} \otimes \mathbf{Z})\mathbf{H}_p(\mathbf{Z}^{(2)})^{-1} \tag{6.3}$$

(cf. Shapiro 1983a, p. 75). Under normality assumption (6.1), this covariance matrix becomes

$$2(\mathbf{Z}^{(2)})^{-1}(\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z})^{(2)}(\mathbf{Z}^{(2)})^{-1}. \tag{6.4}$$

In particular, in the case of asymptotic efficiency (e.g., $\mathbf{B} = \boldsymbol{\Sigma}^{-1}$) we obtain this covariance matrix in the form $2(\mathbf{Z}^{(2)})^{-1}$, where $\mathbf{Z} = \mathbf{F}(\mathbf{F}'\boldsymbol{\Sigma}\mathbf{F})^{-1}\mathbf{F}'$.

Possible covariance matrices associated with the estimator $\hat{\boldsymbol{\lambda}}$ are defined by Equation (4.8). Using Equation (4.7) one can obtain the required matrix $\mathbf{U}_2$ in the form

$$\mathbf{U}_2 = \tfrac{1}{2}\mathbf{K}_p^-[\mathbf{B} \otimes \mathbf{B} - (\mathbf{B} \otimes \mathbf{B})\mathbf{H}_p(\mathbf{B}^{(2)})^{-1}\mathbf{H}_p'(\mathbf{B} \otimes \mathbf{B})]\mathbf{K}_p^{-\prime}. \tag{6.5}$$

In the case of asymptotic efficiency (e.g., $\mathbf{B} = \boldsymbol{\Sigma}^{-1}$) it follows from (5.6) that the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0)$ is given by a nonnegative definite g-inverse matrix of $\boldsymbol{\Delta}_1'\mathbf{U}_2\boldsymbol{\Delta}_1$, where $\mathbf{U}_2$ is given by the right side of (6.5) with $\mathbf{B} = \boldsymbol{\Sigma}^{-1}$. One approach to identifying $\hat{\boldsymbol{\lambda}}$ is by imposing identification constraints (e.g., see Lawley and Maxwell 1971). This leads to the nonnegative definite reflexive g-inverse matrix of $\boldsymbol{\Delta}_1'\mathbf{U}_2\boldsymbol{\Delta}_1$, which of course depends on constraints imposed. Another widely used method of selecting $\hat{\boldsymbol{\lambda}}$ is to optimize a certain objective function on the set of all permissible $\hat{\boldsymbol{\lambda}}$. It has been shown by Jennrich (1974) that this method is equivalent to imposing identification constraints, giving the necessary optimality conditions for the corresponding minimizer.

APPENDIX A

Let the point $\boldsymbol{\theta}_0$ be regular, and consider a matrix $\mathbf{J}$ satisfying Equation (4.2). We want to show that there exists a minimizer $\boldsymbol{\theta}^*(\mathbf{x})$ such that $\boldsymbol{\theta}^*(\mathbf{x}) \to \boldsymbol{\theta}_0$ as $\mathbf{x} \to \boldsymbol{\xi}_0$ and $(\partial/\partial\mathbf{x}')\boldsymbol{\theta}^*(\boldsymbol{\xi}_0) = \mathbf{J}$. First we note that Equation (4.2) is invariant under local reparameterization (local diffeomorphism). Indeed, replacing $\boldsymbol{\theta}$ with $\boldsymbol{\gamma} = \mathbf{h}^{-1}(\boldsymbol{\theta})$, we have to replace $\boldsymbol{\Delta}$ with $\boldsymbol{\Delta}H$ and $\mathbf{J}$ with $\partial\boldsymbol{\gamma}^*/\partial\mathbf{x}' = \mathbf{H}^{-1}\mathbf{J}$, where $\mathbf{H}$ is the Jacobian matrix of $\mathbf{h}$ at $\boldsymbol{\gamma}_0$. Clearly this does not change Equation (4.2). Therefore we can partition $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$, with $\boldsymbol{\theta}_1 \in \mathbf{R}^r$ and $\boldsymbol{\theta}_2 \in \mathbf{R}^{q-r}$, and suppose without loss of generality that $\mathbf{g}(\boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}_2$. The corresponding partition of $\boldsymbol{\Delta}$ and $\mathbf{J}$ is $(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$ and $\mathbf{J}' = (\mathbf{J}_1', \mathbf{J}_2')$, respectively, with $\boldsymbol{\Delta}_2(\boldsymbol{\theta})$ identically zero. We have that $\boldsymbol{\theta}_1^*(\mathbf{x})$ is locally unique, and by the standard theory equation $(\boldsymbol{\Delta}_1'\mathbf{V}\boldsymbol{\Delta}_1)\mathbf{J}_1 = \boldsymbol{\Delta}_1'\mathbf{V}$ holds. On the other hand, $\boldsymbol{\theta}_2^*(\mathbf{x})$ and then $\mathbf{J}_2$ can be arbitrary. Noting that this is precisely the class of matrices $\mathbf{J}$ given by Equation (4.2), we complete the proof.

APPENDIX B

Consider a set of $k = q - r$ identification constraints $c_i(\boldsymbol{\theta}) = 0$ ($i = 1, \ldots, k$), with $\mathbf{c}(\boldsymbol{\theta}_0) = \mathbf{0}$, and the corresponding Jacobian matrix $\mathbf{J}$ of the restricted minimizer $\boldsymbol{\theta}^*(\mathbf{x})$. Then $\mathbf{J}$ is equal to $\mathbf{A}\boldsymbol{\Delta}'\mathbf{V}$, where $\mathbf{A}$ is the $q \times q$ upper left block of the matrix

$$\begin{bmatrix} \boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta} & \mathbf{B} \\ \mathbf{B} & \mathbf{0} \end{bmatrix}^{-1}, \tag{B.1}$$

where $\mathbf{B}$ is the $q \times k$ Jacobian matrix of $\mathbf{c}$ at $\boldsymbol{\theta}_0$ (e.g., Shapiro 1983a, p. 54). Of course we suppose that the identification constraints are chosen in such a way that the block matrix in (B.1) is invertible. It can be shown that $\mathbf{A}$ is a (reflexive and symmetric) generalized inverse of $\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}$ (Blattner 1962) and hence (4.4) follows. A closed-form expression for $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{W}(\mathbf{W}'\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}\mathbf{W})^{-1}\mathbf{W}', \tag{B.2}$$

where $\mathbf{W}$ is a $q \times r$ orthogonal complement of $\mathbf{B}$.

On the other hand, let $\mathbf{K}$ be a generalized inverse of $\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta}$. Without loss of generality we can suppose that $\boldsymbol{\Delta}$ is partitioned $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$

with $\Delta_2 = 0$. Then

$$K = \begin{bmatrix} (\Delta_1'V\Delta_1)^{-1} & C \\ E & D \end{bmatrix}, \quad (B.3)$$

where **C**, **D**, and **E** are arbitrary matrices (e.g., Pringle and Rayner 1971, p. 9). Furthermore, the matrix $K\Delta'V$ is independent of a particular choice of **C** and **D**. Now it is easy to verify that for the given **E** and appropriately chosen **C** and **D**, matrix **W** in the right side of (B.2) can be chosen in such a way that $A = K$. Since the matrix **B** in (B.1), and then **W**, is arbitrary, we complete the proof.

APPENDIX C: PROOF OF PROPOSITION 5.1

Since $b \in \mathfrak{M}(\Delta')$, $b'\Pi b$ is invariant under the choice of Π satisfying (4.3). Therefore we can suppose without loss of generality that Π is given by

$$\Pi = A\Delta'V\Gamma V\Delta A', \quad (C.1)$$

where **A** is a generalized inverse of $\Delta'V\Delta$. We have that

$$b'[A\Delta'V - (\Delta'\Gamma^{-1}\Delta)^{-}\Delta'\Gamma^{-1}] \times \Gamma[A\Delta'V - (\Delta'\Gamma^{-1}\Delta)^{-}\Delta'\Gamma^{-1}]'b \geq 0. \quad (C.2)$$

Since $b \in \mathfrak{M}(\Delta')$, we also have (Rao and Mitra 1971, p. 24)

$$(\Delta'V\Delta)(\Delta'V\Delta)^{-}b = b \quad (C.3)$$

and

$$(\Delta'\Gamma^{-1}\Delta)(\Delta'\Gamma^{-1}\Delta)^{-}b = b. \quad (C.4)$$

Then the inequality (C.2), together with (C.1), (C.3), and (C.4), implies the inequality (5.2) of Proposition 5.1.

[Received May 1983. Revised July 1985.]

REFERENCES

Abatzoglou, T. J. (1978), "The Minimum Norm Projection on C^2-Manifolds in R^n," *Transactions of the American Mathematical Society*, 243, 115–122.

Aitchison, J., and Silvey, D. C. (1958), "Maximum Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29, 813–828.

Anderson, T. W. (1969), "Statistical Inference for Covariance Matrices With Linear Structure," in *Multivariate Analysis* (Vol. 2), ed. P. R. Krishnaiah, New York: Academic Press, pp. 55–66.

——— (1973), "Asymptotically Efficient Estimation of Covariance Matrices With Linear Structure," *The Annals of Statistics*, 1, 135–141.

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.

Bentler, P. M. (1983), "Simultaneous Equation Systems as Moment Structure Models," *Journal of Econometrics*, 22, 13–42.

Blattner, J. W. (1962), "Bordered Matrices," *SIAM Journal of Applied Mathematics*, 10, 528–536.

Browne, M. W. (1974), "Generalized Least Squares Estimators in the Analysis of Covariance Structures," *South African Statistical Journal*, 8, 1–24.

——— (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 for the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.

Burguete, J. F., Gallant, A. R., and Souza, G. (1982), "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews*, 1, 151–190.

Dieudonné, J. (1960), *Foundations of Modern Analysis*, New York: Academic Press.

Federer, H. (1969), *Geometric Measure Theory*, New York: Springer-Verlag.

Fisher, F. M. (1966), *The Identification Problem in Econometrics*, New York: McGraw-Hill.

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

Jennrich, R. I. (1974), "Simplified Formulae for Standard Errors in Maximum Likelihood Factor Analysis," *British Journal of Mathematical and Statistical Psychology*, 27, 122–131.

Jöreskog, K. G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239–251.

——— (1981), "Analysis of Covariance Structures," *Scandinavian Journal of Statistics*, 8, 65–83.

Jöreskog, K. G., and Goldberger, A. S. (1972), "Factor Analysis by Generalized Least-Squares," *Psychometrika*, 37, 243–260.

Jöreskog, K. G., and Sörbom, D. (1979), *Advances in Factor Analysis and Structural Equation Models*, Cambridge, MA: Abt Books.

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

Lee, S. Y., and Bentler, P. M. (1980), "Some Asymptotic Properties of Constrained Generalized Least Squares Estimation in Covariance Structure Models," *South African Statistical Journal*, 14, 121–136.

Nel, D. G. (1980), "On Matrix Differentiation in Statistics," *South African Statistical Journal*, 14, 137–193.

Pringle, R. M., and Rayner, A. A. (1971), *Generalized Inverse Matrices With Applications to Statistics*, New York: Griffin's Statistical Monographs and Courses.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, New York: John Wiley.

Rao, C. R., and Mitra, S. K. (1971), *Generalized Inverse of Matrices and Its Applications*, New York: John Wiley.

Robertson, C. A. (1972), "On Minimum Discrepancy Estimators," *Sankhyā*, Ser. A, 34, 133–144.

Rothenberg, T. J. (1971), "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Sargan, J. D. (1983), "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

Searle, S. R. (1971), *Linear Models*, New York: John Wiley.

Shapiro, A. (1982), "Rank-Reducibility of a Symmetric Matrix and Sampling Theory of Minimum Trace Factor Analysis," *Psychometrika*, 47, 187–199.

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

——— (1983b), "On Local Identifiability in Structural Models," research report, University of South Africa, Dept. of Mathematics and Applied Mathematics.

——— (1984), "A Note on the Consistency of Estimators in the Analysis of Moment Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 84–88.

——— (1985a), "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints," *Biometrika*, 72, 133–144.

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

——— (1985c), "Second-Order Derivatives of Extremal-Value Functions and Optimality Conditions for Semi-Infinite Programs," *Mathematics of Operations Research*, 10, 207–219.

——— (1985d), "Identifiability of Factor Analysis, Some Results and Open Problems," in *Linear Algebra and Its Applications* (Vol. 70), eds. I. Olkin, C. R. Rao, and G. P. H. Styan, Amsterdam: Elsevier/North-Holland.

Shapiro, A., and Browe, M. W. (1983), "On the Investigation of Local Identifiability: A Counterexample," *Psychometrika*, 48, 303–304.

Silvey, S. D. (1970), *Statistical Inference*, Baltimore: Penguin Books.

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