

Statistical Inference of Moment Structures

Alexander Shapiro

Abstract

The following material is focused on the statistical inference of moment structure models. Although the theory is presented in terms of general moment structures, the main emphasis is on the analysis of covariance structures. Identifiability and the minimum discrepancy function (MDF) approach to statistical analysis (estimation) of such models are discussed. Topics of the large samples theory, in particular, consistency, asymptotic normality of MDF estimators and asymptotic chi-squaredness of MDF test statistics are addressed. Finally, some results addressing asymptotic robustness of the normal theory based MDF statistical inference in the analysis of covariance structures are presented.

1. Introduction

In this paper we discuss statistical inference of moment structures where first and/or second population moments are hypothesized to have a parametric structure. Classical examples of such models are multinomial and covariance structure models. The presented theory is sufficiently general to handle various situations, however the main focus is on covariance structures. Theory and applications of covariance structures were motivated first by the factor analysis model and its various generalizations and later by the development of LISREL models (Jöreskog, 1977, 1981) (see also (Browne, 1982) for a thorough discussion of covariance structure modeling).

2. Moment structures models

In this section we discuss modeling issues in the analysis of moment structures, and, in particular, identifiability of such models. Let $\xi = (\xi_1, \dots, \xi_m)'$ be a vector variable representing a parameter vector of some statistical population. For example, in the analysis of covariance structures, ξ will represent the elements of a $p \times p$ covariance matrix Σ . That is, $\xi := \text{vec}(\Sigma)$, where $\text{vec}(\Sigma)$ denotes the $p^2 \times 1$ vector formed by stacking

elements of Σ . (The notation “:=” means “equal by definition”.) Of course, since matrix Σ is symmetric, vector $\text{vec}(\Sigma)$ has duplicated elements. Therefore an alternative is to consider $\xi := \text{vecs}(\Sigma)$, where $\text{vecs}(\Sigma)$ denotes the $p(p+1)/2 \times 1$ vector formed by stacking (nonduplicated) elements of Σ above and including the diagonal. Note that covariance matrices Σ are positive semidefinite, and hence the corresponding vectors ξ are restricted to a (convex) subset of \mathbb{R}^m . Therefore, we assume that ξ varies in a set $\mathcal{E} \subset \mathbb{R}^m$ representing a *saturated* model for the population vector ξ . In the analysis of covariance structures we have a natural question whether to use vector $\xi = \text{vec}(\Sigma)$ or $\xi = \text{vecs}(\Sigma)$ for the saturated model. Since in both cases the dimension of the corresponding set \mathcal{E} is $p(p+1)/2$, it seems more advantageous to use $\xi := \text{vecs}(\Sigma)$. In that case the set \mathcal{E} has a nonempty interior. (The interior of \mathcal{E} is the set of points $\xi \in \mathcal{E}$ such that \mathcal{E} contains a neighborhood of ξ . For example, the interior of the set of positive semidefinite matrices is formed by its subset of positive definite matrices. A singular positive semidefinite matrix can be viewed as a boundary point of this set \mathcal{E} . A neighborhood of a point $\xi \in \mathbb{R}^m$ is a subset of \mathbb{R}^m containing a ball centered at ξ of a sufficiently small positive radius.) However, for actual calculations it is often more convenient to use $\xi := \text{vec}(\Sigma)$. When dealing with specific applications we will specify a choice of the corresponding vector ξ .

A model for ξ is a subset \mathcal{E}_0 of \mathcal{E} . Of course, this definition is too abstract and one needs a constructive way of defining a model. There are two natural ways for constructing a model, namely either by imposing equations or by a parameterization. The parameterization approach suggests existence of an $m \times 1$ vector valued function $\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \dots, g_m(\boldsymbol{\theta}))$, and a parameter set $\Theta \subset \mathbb{R}^q$, which relates the parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ to ξ . That is,

$$\mathcal{E}_0 := \{\xi \in \mathcal{E}: \xi = \mathbf{g}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}. \quad (2.1)$$

We refer to $\mathbf{g}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, as a *structural* model for ξ . We assume in the subsequent analysis that the mapping $\mathbf{g}(\boldsymbol{\theta})$ is sufficiently smooth. In particular, we always assume that $\mathbf{g}(\boldsymbol{\theta})$ is *twice continuously differentiable*. We associate with mapping $\mathbf{g}(\cdot)$ its $m \times q$ Jacobian matrix $\partial \mathbf{g}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}' = [\partial g_i(\boldsymbol{\theta}) / \partial \theta_j]_{i=1, \dots, m, j=1, \dots, q}$ of partial derivatives and use notation $\mathbf{A}(\boldsymbol{\theta}) := \partial \mathbf{g}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$.

REMARK 1. It should be noted that the same set \mathcal{E}_0 could be represented by different parameterizations in the form (2.1). For example, let \mathcal{E} be the set of all $p \times p$ symmetric positive semidefinite matrices (covariance matrices) and \mathcal{E}_0 be its subset of diagonal matrices with nonnegative diagonal elements. This model can be parameterized by the set $\Theta := \mathbb{R}_+^p$ and mapping $\mathbf{g}(\boldsymbol{\theta}) := \text{diag}(\theta_1, \dots, \theta_p)$. (The set \mathbb{R}_+^p denotes the nonnegative orthant of the space \mathbb{R}^p , i.e., $\mathbb{R}_+^p := \{\boldsymbol{\theta} \in \mathbb{R}^p: \theta_i \geq 0, i = 1, \dots, p\}$.) Alternatively, it can be parameterized by $\Theta := \mathbb{R}^p$ and $\mathbf{g}(\boldsymbol{\theta}) := \text{diag}(\theta_1^2, \dots, \theta_p^2)$. Of course, in applications the considered parameters typically have an interpretation. For instance, in the above example of diagonal covariance matrices, in the first parameterization parameters θ_i represent the corresponding standard deviations while in the second parameterization these are the corresponding variances. Note that this set \mathcal{E}_0 can be also defined by equations by setting the off-diagonal elements of Σ to zero. In the subsequent analysis we mainly deal with the parameterization approach.

1 It is said that model $\mathbf{g}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, is (globally) *identified* at a point $\boldsymbol{\theta}_0 \in \Theta$ if $\boldsymbol{\theta}_0$ 1
 2 is a unique parameter vector corresponding to the value $\boldsymbol{\xi}_0 := \mathbf{g}(\boldsymbol{\theta}_0)$ of the population 2
 3 vector. It is said that the model is locally identified at $\boldsymbol{\theta}_0$ if such uniqueness holds in a 3
 4 neighborhood of $\boldsymbol{\theta}_0$. More formally, we have the following definition. 4
 5

6 DEFINITION 2.1. It is said that structural model $\mathbf{g}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, is *identified (locally* 6
 7 *identified)* at a point $\boldsymbol{\theta}_0 \in \Theta$ if $\mathbf{g}(\boldsymbol{\theta}^*) = \mathbf{g}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}^* \in \Theta$ ($\boldsymbol{\theta}^*$ in a neighborhood of 7
 8 $\boldsymbol{\theta}_0$) implies that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$. 8
 9

10 Of course, (global) identifiability implies local identifiability. A well-known suffi- 10
 11 cient condition for local identifiability of $\boldsymbol{\theta}_0 \in \Theta$ is that the Jacobian matrix $\mathbf{\Delta}(\boldsymbol{\theta}_0)$, of 11
 12 $\mathbf{g}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$, has full column rank q (e.g., Fisher, 1966). In general, this condition is not 12
 13 necessary for local identifiability of $\boldsymbol{\theta}_0$ even if $\boldsymbol{\theta}_0$ is an interior point of Θ . Take, for 13
 14 example, $g(\theta) := \theta^3$, $\theta \in \mathbb{R}$. This model is locally (and globally) identified at $\theta = 0$, 14
 15 while $\partial g(0)/\partial \theta = 0$. This condition becomes necessary and sufficient under the fol- 15
 16 lowing assumption of constant rank regularity which was used by several authors (e.g., 16
 17 Fisher, 1966; Rothenberg, 1971; Wald, 1950). 17
 18

19 DEFINITION 2.2. We say that a point $\boldsymbol{\theta}_0 \in \Theta$ is *locally regular* if the Jacobian matrix 19
 20 $\mathbf{\Delta}(\boldsymbol{\theta})$ has the same rank as $\mathbf{\Delta}(\boldsymbol{\theta}_0)$ for every $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. 20
 21

22 If the mapping $\mathbf{g}(\boldsymbol{\theta})$ is independent of, say, last s parameters $\theta_{q-s+1}, \dots, \theta_q$, then, 22
 23 of course, these parameters are redundant and the model can be viewed as overparame- 23
 24 terized. In that case the rank of the Jacobian matrix $\mathbf{\Delta}(\boldsymbol{\theta})$ is less than or equal to $q - s$ 24
 25 for any $\boldsymbol{\theta}$. In general, it is natural to view the structural model as being (locally) over- 25
 26 parameterized, at a point $\boldsymbol{\theta}_0$ in the interior of Θ , if it can be reduced to the above case 26
 27 by a local transformation (reparameterization). More formally we have the following 27
 28 definition (Shapiro, 1986). 28
 29

30 DEFINITION 2.3. We say that structural model $\mathbf{g}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, is *locally overparame-* 30
 31 *terized*, at an interior point $\boldsymbol{\theta}_0$ of Θ , if the rank r of $\mathbf{\Delta}(\boldsymbol{\theta}_0)$ is less than q and there 31
 32 exists a local diffeomorphism $\boldsymbol{\theta} = \mathbf{h}(\boldsymbol{\gamma})$ such that the composite mapping $\mathbf{g}(\mathbf{h}(\boldsymbol{\gamma}))$ is 32
 33 independent of, say last, $q - r$ coordinates of $\boldsymbol{\gamma}$. 33
 34

35 Mapping $\mathbf{h}(\boldsymbol{\gamma})$ from a neighborhood of $\boldsymbol{\gamma}_0 \in \mathbb{R}^q$ to a neighborhood of $\boldsymbol{\theta}_0$, with 35
 36 $\mathbf{h}(\boldsymbol{\gamma}_0) = \boldsymbol{\theta}_0$, is called local diffeomorphism if it is continuously differentiable, locally 36
 37 one-to-one and its inverse is also continuously differentiable. It can be shown that $\mathbf{h}(\boldsymbol{\gamma})$ 37
 38 is a local diffeomorphism if and only if it is continuously differentiable and the Jaco- 38
 39 bian matrix $\partial \mathbf{h}(\boldsymbol{\gamma}_0)/\partial \boldsymbol{\gamma}$ is nonsingular. We can assume, without loss of generality, that 39
 40 $\boldsymbol{\gamma}_0 = \mathbf{0}$. 40
 41

42 The local diffeomorphism $\mathbf{h}(\boldsymbol{\gamma})$, in the above definition, can be viewed as a local 42
 43 reparameterization of the model. We do not need to construct such a reparameterization 43
 44 explicitly but rather to know about its existence since it gives us an information about 44
 45 a (local) structure of the model. Clearly, if the model is locally overparameterized at 45

1 a point θ_0 , then it is not locally identified at this point. Moreover, in the case of local
 2 overparameterization the set of points θ such that $\mathbf{g}(\theta) = \mathbf{g}(\theta_0)$ forms a smooth
 3 manifold in a neighborhood of θ_0 . Note that the rank of the Jacobian matrix of the composite
 4 mapping $\mathbf{g}(\mathbf{h}(\boldsymbol{\gamma}))$, at $\boldsymbol{\gamma}_0 = \mathbf{0}$, is the same as the rank of $\Delta(\theta_0)$. Therefore, in the
 5 reparameterized model the remaining r coordinates of $\boldsymbol{\gamma}$ are locally identified.

6 A relation between the concepts of local regularity and local overparameterization is
 7 clarified by the following result known as the Rank Theorem (Fisher, 1966).

8
 9 **PROPOSITION 2.1.** *Let θ_0 be an interior point of the parameter set Θ . Then the fol-*
 10 *lowing holds.*

- 11 (i) *Suppose that θ_0 is locally regular. Then the model is locally identified at θ_0 if and*
 12 *only if the rank r of $\Delta(\theta_0)$ is equal to q , otherwise if $r < q$, then the model is*
 13 *locally overparameterized at θ_0 .*
 14 (ii) *Conversely, if the model is locally overparameterized at θ_0 , then $r < q$ and the*
 15 *point θ_0 is locally regular.*

16
 17 The above results are not very useful for verification of (local) identifiability at an
 18 individual point θ_0 . For one thing the population value of the parameter vector usually
 19 is unknown, and even if the value θ_0 is specified, it is not possible to calculate the
 20 rank of the corresponding Jacobian matrix numerically because of the round off errors.
 21 However, one can approach the identifiability problem from a generic point of view.
 22 Suppose that the mapping $\mathbf{g}(\cdot)$ is analytic, i.e., every coordinate function $g_i(\cdot)$ can be
 23 expanded into a power series in a neighborhood of every point $\theta \in \Theta$. Suppose also
 24 that the set Θ is connected. Let ι be an index set of rows and columns of $\Delta(\theta)$ defining
 25 its squared submatrix and $v_\iota(\theta)$ be the determinant of that submatrix. Clearly there
 26 is only a finite number of such submatrices and hence the corresponding determinant
 27 functions. Since $\mathbf{g}(\cdot)$ is analytic, every such determinant function $v_\iota(\theta)$ is also analytic.
 28 Consequently, either $v_\iota(\theta)$ is identically zero for all $\theta \in \Theta$, or $v_\iota(\theta) \neq 0$ for almost
 29 every $\theta \in \Theta$. (The “almost every” statement here can be understood in the sense that
 30 it holds for all θ in Θ except on a subset of Θ of Lebesgue measure zero.) These
 31 arguments lead to the following result.

32
 33 **PROPOSITION 2.2.** *Suppose that the mapping $\mathbf{g}(\cdot)$ is analytic and the set Θ is con-*
 34 *connected. Then almost every point of Θ is locally regular with the same rank r of the*
 35 *Jacobian matrix $\Delta(\theta)$, and, moreover, the rank of $\Delta(\theta)$ is less than or equal to r for all*
 36 *$\theta \in \Theta$.*

37
 38 By the above proposition we have that with the model, defined by analytic mapping
 39 $\mathbf{g}(\theta)$, is associated an integer r equal to the rank of $\Delta(\theta)$ almost everywhere. We refer
 40 to this number r as the *characteristic rank* of the model, and whenever talking about
 41 the characteristic rank we assume that the mapping $\mathbf{g}(\theta)$ is analytic and the parameter
 42 set Θ is connected (the concept of characteristic rank was introduced in Shapiro (1986)).
 43 It follows from the preceding discussion that either $r = q$ in which case the model
 44 is locally identified almost everywhere, or $r < q$ in which case the model is locally
 45 overparameterized almost everywhere.

We say that the model is identified (locally identified) in the *generic sense* if it is identified (locally identified) at almost every $\theta \in \Theta$ (Shapiro, 1985a). By the above analysis we have that the model is locally identified in the generic sense if and only if its characteristic rank is equal q . Note that the characteristic rank is always less than or equal to the dimension m of the saturated model.

In situations where the model is (locally) overparameterized, the usual practice is to restrict the parameter space by imposing constraints. According to Definition 2.3, if the model is locally overparameterized, at a point θ_0 , then it can be reparameterized such that the reparameterized model locally does not depend on the last $q - r$ coordinates $\gamma_{r+1}, \dots, \gamma_q$. Consequently by imposing the constraints $\gamma_i = 0, i = r + 1, \dots, q$, the reparameterized model becomes locally identified at $\gamma_0 = \mathbf{0}$ while its image space \mathcal{E}_0 is not changed. For the original model this is equivalent to imposing (locally) the identifiability constraints $c_i(\theta) = 0, i = r + 1, \dots, q$, where $\gamma = c(\theta)$ is the inverse of the mapping $\theta = h(\gamma)$.

EXAMPLE 2.1. Consider the factor analysis model:

$$\Sigma = \Lambda \Lambda' + \Psi, \tag{2.2}$$

which relates the $p \times p$ covariance Σ to the $p \times k$ matrix Λ of factor loadings and the $p \times p$ diagonal matrix Ψ of the residual variances. The corresponding parameter vector θ is composed here from the elements of matrix Λ and diagonal elements of Ψ , and hence has dimension $q = pk + p$. Note that the diagonal elements of the matrix Ψ should be nonnegative while there are no restrictions on the elements of Λ .

By substituting ΛT for Λ , where T is an arbitrary $k \times k$ orthogonal matrix, we end up with the same matrix Σ (this is the so-called indeterminacy of the factor analysis model). Since the dimension of the (smooth) manifold of $k \times k$ orthogonal matrices is $k(k - 1)/2$ and the dimension of the space of $p \times p$ symmetric matrices is $p(p + 1)/2$, it is possible to show that the characteristic rank r of the factor analysis model (2.2) is

$$r = \min\{pk + p - k(k - 1)/2, p(p + 1)/2\}. \tag{2.3}$$

It follows that for $k > 1$ the model (2.2) is locally overparameterized. A way of dealing with this is to reduce the number of parameters given by matrix Λ by setting $k(k - 1)/2$ appropriate elements of Λ to zero (Jennrich, 1987). Then the question of global (local) identifiability of the factor analysis model is reduced to the global (local) identifiability of the diagonal matrix Ψ (Anderson and Rubin, 1956). We have that a necessary condition for *generic* local identifiability of Ψ is that $pk + p - k(k - 1)/2$ is less than or equal to $p(p + 1)/2$, which is equivalent to $(p - k)(p - k + 1)/2 \geq p$, and in turn is equivalent to $k \leq \phi(p)$, where

$$\phi(p) := \frac{2p + 1 - \sqrt{8p + 1}}{2}. \tag{2.4}$$

The above function $\phi(p)$ corresponds to the so-called Ledermann bound (Ledermann, 1937). In the present case we have that $k \leq \phi(p)$ is a necessary and sufficient condition for local identifiability of the diagonal matrix Ψ , of the factor analysis model, in the *generic* sense (Shapiro, 1985a).

1 It is more difficult to establish (global) identifiability of a considered model. We can
 2 also approach the (global) identifiability problem from the generic point of view. Of
 3 course, if a model is not locally identified it cannot be globally identified. Therefore,
 4 $k \leq \phi(p)$ is a necessary condition for (global) identifiability of the factor analysis
 5 model in the generic sense. It is known that matrix Ψ , in the factor analysis model (2.2),
 6 is globally identified in the generic sense if and only if $k < \phi(p)$ (Bekker and ten Berge,
 7 1997).
 8
 9

10 3. Minimum discrepancy function estimation approach 10

11
 12 Let $\xi_0 \in \mathcal{E}$ be a population value of the parameter vector of the saturated model. Recall
 13 that we refer to a subset \mathcal{E}_0 of \mathcal{E} as a model for ξ . Unless stated otherwise it will be
 14 assumed that the model is structural, i.e., the set \mathcal{E}_0 is given in the parametric form
 15 (2.1). It is said that the model holds if $\xi_0 \in \mathcal{E}_0$. Clearly this means that there exists
 16 $\theta_0 \in \Theta$ such that $\xi_0 = \mathbf{g}(\theta_0)$. If the model is identified at θ_0 , then this vector θ_0 is
 17 defined uniquely. In that case we refer to θ_0 as the *population* value of the parameter
 18 vector θ .
 19

20 Suppose that we are given an estimator $\hat{\xi}$ of ξ_0 , based on a sample of size n . We will
 21 be interested then in testing the hypothesis $H_0: \xi_0 \in \mathcal{E}_0$, and consequently in estimation
 22 of the population value of the parameter θ . Consider the setting of the covariance struc-
 23 tures with Σ being the covariance matrix of $p \times 1$ random vector X , and let $\Sigma = \Sigma(\theta)$
 24 be an associated structural model. Let X_1, \dots, X_n be an iid (independent identically
 25 distributed) random sample drawn from a considered population. Then the standard es-
 26 timator of the population value Σ_0 of the covariance matrix is the sample covariance
 27 matrix
 28

$$29 \quad S := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})', \quad (3.1)$$

30 where $\bar{X} := n^{-1} \sum_{i=1}^n X_i$. Suppose, further, that the population distribution is (multi-
 31 variate) normal with mean vector μ_0 and covariance matrix Σ_0 . Then the corresponding
 32 log-likelihood function (up to a constant independent of the parameters) is
 33

$$34 \quad \ell(\mu, \Sigma) = -\frac{1}{2}n \ln |\Sigma| - \frac{1}{2} \operatorname{tr} \left(\Sigma^{-1} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)' \right). \quad (3.2)$$

35
 36 (By $|A|$ and $\operatorname{tr}(A)$ we denote the determinant and the trace, respectively, of a (square)
 37 matrix A .) The maximum likelihood (ML) estimator of μ_0 is \bar{X} and the ML estimator
 38 of Σ_0 , for the saturated model, is
 39

$$40 \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' = \frac{n-1}{n} S. \quad (3.3)$$

41
 42 Of course, for reasonably large values of n , the ML estimator $\hat{\Sigma}$ is “almost” equal to
 43 the unbiased estimator S . Therefore, with some abuse of notation, we use $S = \hat{\Sigma}$ as the
 44 estimator of the population covariance matrix.
 45

It follows that two times log-likelihood ratio statistic for testing the null hypothesis $\Sigma_0 = \Sigma(\theta_0)$ is given by $n\widehat{F}$, where

$$\widehat{F} := \min_{\theta \in \Theta} F_{ML}(S, \Sigma(\theta)), \tag{3.4}$$

with $F_{ML}(\cdot, \cdot)$ being a function of two (matrix valued) variables defined by

$$F_{ML}(S, \Sigma) := \ln |\Sigma| - \ln |S| + \text{tr}(S\Sigma^{-1}) - p. \tag{3.5}$$

The corresponding ML estimator $\hat{\theta}$ of θ_0 is a minimizer of $F_{ML}(S, \Sigma(\theta))$ over $\theta \in \Theta$, i.e.,

$$\hat{\theta} \in \arg \min_{\theta \in \Theta} F_{ML}(S, \Sigma(\theta)). \tag{3.6}$$

(By $\arg \min\{f(\theta) : \theta \in \Theta\}$ we denote the set of all minimizers of $f(\theta)$ over $\theta \in \Theta$; this set can be empty or contain more than one element.) Let us note at this point that the estimator $\hat{\theta}$, defined in (3.6), can be calculated whether the model holds or not. We will discuss implications of this later.

Following Browne (1982), we say that a function $F(x, \xi)$, of two vector variables $x, \xi \in \mathcal{E}$, is a discrepancy function if it satisfies the following conditions:

- (i) $F(x, \xi) \geq 0$ for all $x, \xi \in \mathcal{E}$,
- (ii) $F(x, \xi) = 0$ if and only if $x = \xi$,
- (iii) $F(x, \xi)$ is twice continuously differentiable jointly in x and ξ .

Let $g(\theta)$, $\theta \in \Theta$, be a structural model being considered. Given an estimator $\hat{\xi}$ of ξ_0 and a discrepancy function $F(x, \xi)$, we refer to the statistic $n\widehat{F}$, where

$$\widehat{F} := \min_{\theta \in \Theta} F(\hat{\xi}, g(\theta)), \tag{3.7}$$

as the minimum discrepancy function (MDF) test statistic, and to

$$\hat{\theta} \in \arg \min_{\theta \in \Theta} F(\hat{\xi}, g(\theta)) \tag{3.8}$$

as the MDF estimator.

The function $F_{ML}(S, \Sigma)$ defined in (3.5), considered as a function of $s = \text{vec}(S)$ and $\sigma = \text{vec}(\Sigma)$, is an example of a discrepancy function. It is referred to as the maximum likelihood (ML) discrepancy function. Another popular choice of the discrepancy function in the analysis of covariance structures is

$$F_{GLS}(S, \Sigma) := \frac{1}{2} \text{tr}[(S - \Sigma)S^{-1}(S - \Sigma)S^{-1}]. \tag{3.9}$$

We refer to a function of the form

$$F(x, \xi) := (x - \xi)'[V(x)](x - \xi), \quad x, \xi \in \mathcal{E}, \tag{3.10}$$

as a generalized least squares (GLS) discrepancy function. Here $V(x)$ is an $m \times m$ symmetric matrix valued function of $x \in \mathcal{E}$. We assume that for any $x \in \mathcal{E}$, the corresponding matrix $V(x)$ is positive definite, and hence conditions (i) and (ii) hold, and that $V(x)$ is twice continuously differentiable, and hence condition (iii) is satisfied.

The $F_{\text{GLS}}(\mathbf{S}, \boldsymbol{\Sigma})$ function, defined in (3.9), is a particular example of GLS discrepancy functions with the weight matrix $\mathbf{V}(\mathbf{s}) = \frac{1}{2}\mathbf{S}^{-1} \otimes \mathbf{S}^{-1}$ (by \otimes we denote the Kronecker product of matrices).

We have the following basic result about the structure of discrepancy functions (Shapiro, 1985b).

PROPOSITION 3.1. *Let $F(\mathbf{x}, \boldsymbol{\xi})$ be a discrepancy function satisfying conditions (i)–(iii). Then there exists a continuous $m \times m$ symmetric matrix valued function $\mathbf{V}(\mathbf{x}, \boldsymbol{\xi})$ such that*

$$F(\mathbf{x}, \boldsymbol{\xi}) = (\mathbf{x} - \boldsymbol{\xi})' [\mathbf{V}(\mathbf{x}, \boldsymbol{\xi})] (\mathbf{x} - \boldsymbol{\xi}) \quad (3.11)$$

for all $\mathbf{x}, \boldsymbol{\xi} \in \mathcal{E}$.

The above result shows that any discrepancy function can be represented in a form of an “almost” GLS function. A difference between the representation (3.11) and the general form (3.10) of GLS discrepancy functions is that the weight matrix in (3.11) can also depend on $\boldsymbol{\xi}$ as well as on \mathbf{x} .

Let $\boldsymbol{\xi}_0 \in \mathcal{E}$ be a given (say the population) value of vector $\boldsymbol{\xi}$. Consider matrix $\mathbf{V}_0 := \mathbf{V}(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ associated with the matrix valued function $\mathbf{V}(\cdot, \cdot)$ of representation (3.11). We can then write

$$F(\mathbf{x}, \boldsymbol{\xi}) = (\mathbf{x} - \boldsymbol{\xi})' \mathbf{V}_0 (\mathbf{x} - \boldsymbol{\xi}) + \mathbf{r}(\mathbf{x}, \boldsymbol{\xi}), \quad (3.12)$$

where $\mathbf{r}(\mathbf{x}, \boldsymbol{\xi}) := (\mathbf{x} - \boldsymbol{\xi})' [\mathbf{V}(\mathbf{x}, \boldsymbol{\xi}) - \mathbf{V}_0] (\mathbf{x} - \boldsymbol{\xi})$. We have that

$$|\mathbf{r}(\mathbf{x}, \boldsymbol{\xi})| \leq \|\mathbf{x} - \boldsymbol{\xi}\|^2 \|\mathbf{V}(\mathbf{x}, \boldsymbol{\xi}) - \mathbf{V}_0\|,$$

and $\mathbf{V}(\mathbf{x}, \boldsymbol{\xi})$ tends to \mathbf{V}_0 as $\mathbf{x} \rightarrow \boldsymbol{\xi}_0$ and $\boldsymbol{\xi} \rightarrow \boldsymbol{\xi}_0$. Consequently, for $(\mathbf{x}, \boldsymbol{\xi})$ near $(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ the remainder term $\mathbf{r}(\mathbf{x}, \boldsymbol{\xi})$ in (3.12) is of order

$$\mathbf{r}(\mathbf{x}, \boldsymbol{\xi}) = o(\|\mathbf{x} - \boldsymbol{\xi}_0\|^2 + \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2).$$

The notation $o(x)$ means that $o(x)$ is a function of x such that $o(x)/x$ tends to zero as $x \rightarrow 0$. For a sequence X_n of random variables the notation $X_n = o_p(a_n)$ means that X_n/a_n converges in probability to zero. In particular, $X_n = o_p(1)$ means that X_n converges in probability to zero.

This can be compared with a Taylor expansion of $F(\mathbf{x}, \boldsymbol{\xi})$ at $(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$. We have that $F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0) = 0$, and since $F(\cdot, \boldsymbol{\xi}_0)$ attains its minimum (of zero) at $\mathbf{x} = \boldsymbol{\xi}_0$, we have that $\partial F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)/\partial \mathbf{x} = 0$, and similarly $\partial F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)/\partial \boldsymbol{\xi} = 0$. It follows that the second-order Taylor expansion of $F(\mathbf{x}, \boldsymbol{\xi})$ at $(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ can be written as follows

$$F(\mathbf{x}, \boldsymbol{\xi}) = \frac{1}{2}(\mathbf{x} - \boldsymbol{\xi}_0)' \mathbf{H}_{xx} (\mathbf{x} - \boldsymbol{\xi}_0) + \frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\xi}_0)' \mathbf{H}_{\xi\xi} (\boldsymbol{\xi} - \boldsymbol{\xi}_0) + (\mathbf{x} - \boldsymbol{\xi}_0)' \mathbf{H}_{x\xi} (\boldsymbol{\xi} - \boldsymbol{\xi}_0) + o(\|\mathbf{x} - \boldsymbol{\xi}_0\|^2 + \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2), \quad (3.13)$$

where $\mathbf{H}_{xx} := \partial^2 F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)/\partial \mathbf{x} \partial \mathbf{x}'$, $\mathbf{H}_{\xi\xi} := \partial^2 F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)/\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'$, $\mathbf{H}_{x\xi} := \partial^2 F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)/\partial \mathbf{x} \partial \boldsymbol{\xi}'$ are the corresponding Hessian matrices of second order partial derivatives. By

comparing (3.12) and (3.13), we obtain that

$$\frac{\partial^2 F(\xi_0, \xi_0)}{\partial \mathbf{x} \partial \mathbf{x}'} = \frac{\partial^2 F(\xi_0, \xi_0)}{\partial \xi \partial \xi'} = -\frac{\partial^2 F(\xi_0, \xi_0)}{\partial \mathbf{x} \partial \xi'} = 2V_0. \tag{3.14}$$

Both discrepancy functions $F_{ML}(s, \sigma)$ and $F_{GLS}(s, \sigma)$, defined in (3.5) and (3.9), respectively, have the same Hessian matrix

$$\frac{\partial^2 F_{ML}(\sigma_0, \sigma_0)}{\partial s \partial s'} = \frac{\partial^2 F_{GLS}(\sigma_0, \sigma_0)}{\partial s \partial s'} = \Sigma_0^{-1} \otimes \Sigma_0^{-1}, \tag{3.15}$$

and hence the same second-order Taylor approximation at (σ_0, σ_0) .

REMARK 2. For technical reasons we also assume the following condition for discrepancy functions.

(iv) For any (fixed) $\bar{\mathbf{x}} \in \mathcal{E}$, $F(\mathbf{x}, \xi)$ tends to infinity as $\mathbf{x} \rightarrow \bar{\mathbf{x}}$ and $\|\xi\| \rightarrow \infty$.

It is not difficult to verify that the ML discrepancy function, defined in (3.5), and the GLS discrepancy functions satisfy this condition.

4. Consistency of MDF estimators

Let $\hat{\xi} = \hat{\xi}_n$ be a given estimator, based on a sample of size n , of the population value $\xi_0 \in \mathcal{E}$ of the parameter vector of the saturated model (we use the subscript n , in the notation $\hat{\xi}_n$, to emphasize that the estimator is a function of a sample of size n being considered). It is said that the estimator $\hat{\xi}_n$ is *consistent* if it converges with probability one (w.p.1) to ξ_0 as $n \rightarrow \infty$. For example, by the (strong) Law of Large Numbers we have that the sample covariance matrix S converges to Σ_0 w.p.1 as $n \rightarrow \infty$. For this to hold we only need to assume that the population distribution has finite second-order moments, and hence the covariance matrix Σ_0 does exist, and that the corresponding random sample is iid.

Let $F(\mathbf{x}, \xi)$ be a chosen discrepancy function satisfying conditions (i)–(iv) specified in the previous section, and $\bar{\xi} = \bar{\xi}(\mathbf{x})$ be an optimal solution of the minimization problem:

$$\min_{\xi \in \mathcal{E}_0} F(\mathbf{x}, \xi). \tag{4.1}$$

Note that since $F(\mathbf{x}, \cdot)$ is continuous and because of condition (iv), such a minimizer always exists (provided that the set \mathcal{E}_0 is closed), although it may be not unique. Define

$$\hat{\xi}_n^* := \bar{\xi}(\hat{\xi}_n) \quad \text{and} \quad \xi^* := \bar{\xi}(\xi_0). \tag{4.2}$$

That is, $\hat{\xi}_n^*$ and ξ^* are minimizers of $F(\hat{\xi}_n, \cdot)$ and $F(\xi_0, \cdot)$, respectively, over \mathcal{E}_0 . It could be noted that if the model holds, i.e., $\xi_0 \in \mathcal{E}_0$, then the minimizer ξ^* coincides with ξ_0 and is unique (because of the properties (i) and (ii) of the discrepancy function). It is possible to show that if the minimizer ξ^* is unique, then the function $\bar{\xi}(\mathbf{x})$ is continuous at $\mathbf{x} = \xi_0$, i.e., $\bar{\xi}(\mathbf{x}) \rightarrow \xi^*$ as $\mathbf{x} \rightarrow \xi_0$. Together with consistency of $\hat{\xi}_n$ this implies the following result (Shapiro, 1984).

1 PROPOSITION 4.1. *Suppose that the discrepancy function satisfies conditions (i)–(iv)*
 2 *and $\hat{\xi}_n$ is a consistent estimator of ξ_0 . Then $\hat{\xi}_n^*$ converges to ξ^* w.p.1 as $n \rightarrow \infty$,*
 3 *provided that the minimizer ξ^* is unique. In particular, if $\xi_0 \in \Xi_0$, then $\hat{\xi}_n^* \rightarrow \xi_0$ w.p.1*
 4 *as $n \rightarrow \infty$.*

5
 6 Similar analysis can be applied to studying consistency of the MDF estimators of the
 7 parameter vectors in Θ . For a given $x \in \mathcal{E}$ consider the optimization (minimization)
 8 problem:

$$10 \quad \min_{\theta \in \Theta} F(x, g(\theta)). \quad (4.3)$$

11
 12 Recall that the MDF estimator $\hat{\theta}_n$ is an optimal solution of problem (4.3) for $x = \hat{\xi}_n$.
 13 Let θ^* be an optimal solution of (4.3) for $x = \xi_0$, i.e.,

$$14 \quad \theta^* \in \arg \min_{\theta \in \Theta} F(\xi_0, g(\theta)).$$

15
 16 Of course, if $\xi_0 = g(\theta_0)$ for some $\theta_0 \in \Theta$ (i.e., the model holds), then θ_0 is an optimal
 17 solution of (4.3) for $x = \xi_0$, and we can take $\theta^* = \theta_0$. The optimal values of problems
 18 (4.1) and (4.3) are equal to each other and there is a one-to-one correspondence between
 19 the sets of optimal solutions of problems (4.1) and (4.3). That is, if $\hat{\xi}$ is an optimal
 20 solution of (4.3), then $\hat{\xi} = g(\hat{\theta})$ is an optimal solution of (4.1), and conversely if $\hat{\xi}$ is an
 21 optimal solution of (4.1) and $\hat{\theta} \in \Theta$ is a corresponding point of Θ , then $\hat{\theta}$ is an optimal
 22 solution of (4.3). The relation between $\hat{\xi}$ and $\hat{\theta}$ is defined by the equation $\hat{\xi} = g(\hat{\theta})$. If
 23 the model is identified at θ , then the equation $\xi = g(\theta)$ defines the point θ uniquely.

24 It follows that, under the assumptions of Proposition 4.1, $\hat{\theta}_n$ is a consistent estima-
 25 tor of θ^* , if the inverse of the mapping $g(\cdot)$ is continuous at θ^* , i.e., if the following
 26 condition holds:

$$27 \quad g(\theta_n) \rightarrow g(\theta^*), \quad \text{for some sequence } \{\theta_n\} \subset \Theta,$$

$$28 \quad \text{implies that } \theta_n \rightarrow \theta^*. \quad (4.4)$$

29
 30 Note that the above condition (4.4) can only hold if the model is identified at θ^* . This
 31 leads to the following result (Kano, 1986; Shapiro, 1984).
 32

33
 34 PROPOSITION 4.2. *Suppose that the discrepancy function satisfies conditions (i)–(iv),*
 35 *$\hat{\xi}_n$ is a consistent estimator of ξ_0 , and for $x = \xi_0$ problem (4.3) has unique optimal*
 36 *solution θ^* and condition (4.4) holds. Then $\hat{\theta}_n$ converges to θ^* w.p.1 as $n \rightarrow \infty$. In*
 37 *particular, if $\xi_0 = g(\theta_0)$, for some $\theta_0 \in \Theta$ (i.e., the model holds), then $\theta_0 = \theta^*$ and the*
 38 *MDF estimator $\hat{\theta}_n$ is a consistent estimator of θ_0 .*

39
 40 Note that uniqueness of the optimal solution θ^* implies uniqueness of the corre-
 41 sponding optimal solution ξ^* . Converse of that also holds if the model is identified
 42 at θ^* . As it was mentioned above, identifiability of θ^* is a necessary condition for the
 43 property (4.4) to hold. It is also sufficient if the set Θ is compact (i.e., bounded and
 44 closed). For a noncompact set Θ , condition (4.4) prevents the MDF estimator from
 45 escaping to infinity.

The above proposition shows that if the model holds, then under mild regularity conditions (in particular, identifiability of the model at θ_0) the MDF estimator $\hat{\theta}_n$ converges w.p.1 to the true (population) value θ_0 of the parameter vector. On the other hand, if the model does not hold, then $\hat{\theta}_n$ converges to an optimal solution θ^* of the problem (4.3). It could be noted that if the model does not hold, then such an optimal solution depends on a particular choice of the discrepancy function.

As we can see uniqueness of the (population) minimizer θ^* is crucial for convergence of the MDF estimator $\hat{\theta}_n$. If the model holds, then $\theta^* = \theta_0$ and uniqueness of θ_0 is equivalent to identifiability of the model at θ_0 . Now if a point $\theta = \theta(x)$ is an optimal solution of problem (4.3) and is an interior point of the set Θ , then it satisfies the necessary optimality condition

$$\frac{\partial F(x, g(\theta))}{\partial \theta} = \mathbf{0}. \tag{4.5}$$

This condition can be viewed as a system of (nonlinear) equations. Consider a point θ_0 in the interior of the set Θ and let $\xi_0 := g(\theta_0)$. By linearizing (4.5) at $x = \xi_0$ and $\theta = \theta_0$, we obtain the following (linear) system of equations:

$$\left[\frac{\partial^2 F(\xi_0, g(\theta_0))}{\partial \theta \partial x'} \right] (x - \xi_0) + \left[\frac{\partial^2 F(\xi_0, g(\theta_0))}{\partial \theta \partial \theta'} \right] (\theta - \theta_0) = \mathbf{0}. \tag{4.6}$$

Note that by (3.14) we have that

$$\frac{\partial^2 F(\xi_0, g(\theta_0))}{\partial \theta \partial x'} = -2\Delta_0' V_0 \quad \text{and} \quad \frac{\partial^2 F(\xi_0, g(\theta_0))}{\partial \theta \partial \theta'} = 2\Delta_0' V_0 \Delta_0, \tag{4.7}$$

where $\Delta_0 := \Delta(\theta_0)$. Since the matrix V_0 is positive definite, we have that the matrix $\Delta_0' V_0 \Delta_0$ is nonsingular iff the Jacobian matrix Δ_0 has full column rank q (recall that this is a sufficient condition for identifiability of θ_0). It follows then by the Implicit Function Theorem that:

If Δ_0 has full column rank q , then for all x sufficiently close to ξ_0 the system (4.5) has a unique solution $\bar{\theta} = \bar{\theta}(x)$ in a neighborhood of θ_0 , and $\bar{\theta}$ is the unique optimal solution of problem (4.3) in that neighborhood of θ_0 .

The above is a local result. It implies that, under the specified conditions, if the estimator $\hat{\xi}_n$ is sufficiently close to a point $\xi_0 = g(\theta_0)$ satisfying the model, i.e., the fit is good enough, then the corresponding MDF estimator $\hat{\theta}_n$ is unique, and can be obtained by solving Eq. (4.5) with $x = \hat{\xi}_n$, in a neighborhood of the point $\theta_0 \in \Theta$. Of course, in practice it is impossible to say a priori when “sufficiently close” is close enough for the above to hold.

5. Asymptotic analysis of the MDF estimation procedure

In this section we discuss a basic theory of asymptotics of the MDF estimation procedure. We assume that the discrepancy function considered satisfies conditions (i)–(iv)

1 specified in Section 3. We also assume that the estimator $\hat{\xi}_n$ is *asymptotically normal*.
 2 That is, we assume that the sequence

$$3 \quad \mathbf{Z}_n := n^{1/2}(\hat{\xi}_n - \xi_0), \quad 4$$

5 of random vectors, with ξ_0 being the population value of the parameter vector $\xi \in \Xi$,
 6 converges in distribution to multivariate normal with mean vector zero and covariance
 7 matrix Γ , i.e., $\mathbf{Z}_n \Rightarrow N(\mathbf{0}, \Gamma)$ (by “ \Rightarrow ” we denote convergence in distribution). For
 8 example, in the analysis of covariance structures we have that vector $s := \text{vec}(\mathbf{S})$,
 9 associated with the sample covariance matrix \mathbf{S} , is asymptotically normal. This follows
 10 from the Central Limit Theorem provided that the population distribution has fourth-
 11 order moments and the sample is iid. Moreover, if the population distribution is normal,
 12 then $\Gamma = \Gamma_N$, where

$$13 \quad \Gamma_N := 2\mathbf{M}_p(\Sigma_0 \otimes \Sigma_0) \quad 14 \quad (5.1)$$

15 with \mathbf{M}_p being an $p^2 \times p^2$ symmetric idempotent matrix of rank $p(p+1)/2$ with
 16 element in row ij and column kl given by $\mathbf{M}_p(ij, kl) = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$ (Browne,
 17 1974) (here $\delta_{ik} = 1$ if $i = k$, and $\delta_{ik} = 0$ if $i \neq k$). It follows that matrix Γ_N also has
 18 rank $p(p+1)/2$, provided that the covariance matrix Σ_0 is nonsingular. We assume that
 19 the (asymptotic) covariance matrix Γ , of \mathbf{Z}_n , has the *maximal rank*, which in the case
 20 of covariance structures is $p(p+1)/2$. It follows then that the linear space generated
 21 by columns of the Jacobian matrix $\Delta(\theta)$ is contained in the linear space generated by
 22 columns of Γ .
 23

24 Denote by $\vartheta(x)$ the optimal value of problem (4.1), i.e., $\vartheta(x) := \inf_{\xi \in \Xi_0} F(x, \xi)$.
 25 Recall that the optimal values of problems (4.1) and (4.3) are the same, and hence $\vartheta(x)$
 26 is also the optimal value of problem (4.3), i.e., we can write $\vartheta(x) = \inf_{\theta \in \Theta} F(x, \mathbf{g}(\theta))$.
 27 Denote by $\bar{\theta}(x)$ an optimal solution of problem (4.3), i.e. $\bar{\theta}(x) \in \arg \min_{\theta \in \Theta} F(x, \mathbf{g}(\theta))$.
 28 By the definitions, we have that $\hat{F} = \vartheta(\hat{\xi}_n)$ and $\hat{\theta}_n = \bar{\theta}(\hat{\xi}_n)$. Therefore it should not
 29 be surprising that asymptotic properties of the MDF test statistics and estimators are
 30 closely related to analytical properties of functions $\vartheta(\cdot)$ and $\bar{\theta}(\cdot)$.
 31

32 Suppose that the model holds, i.e., $\xi_0 \in \Xi_0$ or equivalently $\xi_0 = \mathbf{g}(\theta_0)$ for some
 33 $\theta_0 \in \Theta$. The second-order Taylor approximation of the discrepancy function, at the
 34 point $(x, \xi) = (\xi_0, \xi_0)$, can be written in the form (3.12). Suppose, further, that the set
 35 Ξ_0 can be approximated at the point ξ_0 by a cone $\mathcal{T} \subset \mathbb{R}^m$ in the following sense:

$$36 \quad \text{dist}(\xi_0 + z, \Xi_0) = o(\|z\|), \quad z \in \mathcal{T}, \quad 36 \quad (5.2)$$

$$37 \quad \text{dist}(\xi - \xi_0, \mathcal{T}) = o(\|\xi - \xi_0\|), \quad \xi \in \Xi_0. \quad 37 \quad (5.3)$$

38
 39 This definition of cone approximation goes back to Chernoff (1954). (By $\text{dist}(x, A) :=$
 40 $\inf_{z \in A} \|x - z\|$ we denote the distance from a point $x \in \mathbb{R}^m$ to a set $A \subset \mathbb{R}^m$. A set
 41 $\mathcal{T} \subset \mathbb{R}^m$ is said to be a *cone* if for any $z \in \mathcal{T}$ and $t \geq 0$ it follows that $tz \in \mathcal{T}$.)
 42

43 In particular, if Ξ_0 is a smooth manifold near ξ_0 , then it is approximated at ξ_0 by a
 44 *linear space* referred to as its tangent space at ξ_0 . Suppose that θ_0 is an *interior* point of
 45 Θ and θ_0 is locally regular (see Definition 2.2). Denote $\Delta_0 := \Delta(\theta_0)$. Then the image
 $\mathbf{g}(\mathcal{N}) := \{\xi: \xi = \mathbf{g}(\theta), \theta \in \mathcal{N}\}$ of the set Θ restricted to a neighborhood $\mathcal{N} \subset \Theta$ of

θ_0 , is a smooth manifold with the tangent space at ξ_0 given by

$$\mathcal{T} = \{\zeta = \Delta_0 \beta: \beta \in \mathbb{R}^q\}. \tag{5.4}$$

Of course, $\mathbf{g}(\mathcal{N})$ is a subset of \mathcal{E}_0 , restricted to a neighborhood of ξ_0 . The asymptotic analysis is local in nature. Therefore, there is no loss of generality here by restricting the set Θ to a neighborhood of θ_0 .

DEFINITION 5.1. A point $\theta_0 \in \Theta$ is said to be *regular* if θ_0 is locally regular and there exist a neighborhood \mathcal{V} of $\xi_0 = \mathbf{g}(\theta_0)$ and a neighborhood $\mathcal{N} \subset \Theta$ of θ_0 such that $\mathcal{E}_0 \cap \mathcal{V} = \mathbf{g}(\mathcal{N})$.

In other words, regularity of θ_0 ensures that local structure of \mathcal{E}_0 near ξ_0 is provided by the mapping $\mathbf{g}(\theta)$ defined in a neighborhood of θ_0 . Regularity of θ_0 implies that \mathcal{E}_0 is a smooth manifold near ξ_0 and is approximated at ξ_0 by its tangent space \mathcal{T} of the form (5.4).

In particular, if condition (4.4) holds and $\Delta(\theta)$ has full column rank q for all θ in a neighborhood of θ_0 (i.e., point θ_0 is locally regular of rank q), then \mathcal{E}_0 is a smooth manifold near ξ_0 and its tangent space at ξ_0 is given by (5.4). Note that $\mathbf{g}(\mathcal{N})$ is a smooth manifold even if the rank of the Jacobian matrix Δ_0 is less than q provided that the local regularity condition holds. Note also that if the tangent space \mathcal{T} is given in the form (5.4), then its dimension, $\dim(\mathcal{T})$, is equal to the rank of the Jacobian matrix Δ_0 , i.e., $\dim(\mathcal{T}) = \text{rank}(\Delta_0)$.

REMARK 3. Let us remark at this point that if the point θ_0 is a *boundary* point of Θ , then under certain regularity conditions the set Θ can be approximated at θ_0 by a cone $\mathcal{C} \subset \mathbb{R}^q$, rather than a linear space, and consequently \mathcal{E}_0 can be approximated by the cone

$$\mathcal{T} = \{\zeta = \Delta_0 \beta: \beta \in \mathcal{C}\}. \tag{5.5}$$

Later we will discuss implications of this to the asymptotics of the MDF estimators (see Section 5.4).

The above discussion suggests the following approximation of the optimal value function $\vartheta(\mathbf{x})$ near ξ_0 (Shapiro, 1985c, Lemma 3.1).

PROPOSITION 5.1. Suppose that the set \mathcal{E}_0 can be approximated at $\xi_0 \in \mathcal{E}_0$ by a cone $\mathcal{T} \subset \mathbb{R}^m$. Then

$$\vartheta(\xi_0 + z) = \min_{\zeta \in \mathcal{T}} (z - \zeta)' V_0 (z - \zeta) + o(\|z\|^2). \tag{5.6}$$

Suppose, further, that the cone \mathcal{T} actually is a linear space, i.e., \mathcal{E}_0 can be approximated at $\xi_0 \in \mathcal{E}_0$ by a linear space $\mathcal{T} \subset \mathbb{R}^m$. Then the main (first) term in the right-hand side of (5.6) is a quadratic function of z and can be written as $z' Q z$ for some symmetric positive semidefinite matrix Q . In particular, if the space \mathcal{T} is given in the form (5.4),

1 then this term can be written as

$$2 \min_{\beta \in \mathbb{R}^q} (z - \Delta_0 \beta)' V_0 (z - \Delta_0 \beta) = z' Q z, \quad (5.7)$$

3 where $Q = V_0 - V_0 \Delta_0 (\Delta_0' V_0 \Delta_0)^{-1} \Delta_0' V_0$ (by A^{-} we denote a generalized inverse of
4 matrix A). It is also possible to write this matrix in the form

$$5 Q = \Delta_c (\Delta_c' V_0^{-1} \Delta_c)^{-1} \Delta_c', \quad (5.8)$$

6 where Δ_c is an orthogonal complement of Δ_0 , i.e., Δ_c is an $m \times (m - \text{rank}(\Delta_0))$
7 matrix of full column rank such that $\Delta_c' \Delta_0 = 0$. This follows by the standard theory
8 of linear models (see, e.g., (Seber, 1977, Sections 3.6 and 3.8)). Note that $\text{rank}(Q) =$
9 $m - \text{rank}(\Delta_0)$.

10 We already discussed continuity properties of $\bar{\theta}(\cdot)$ in Section 4. Suppose that the
11 model is (globally) identified at θ_0 , and hence $\bar{\theta}(\xi_0) = \theta_0$ and is defined uniquely.
12 Then under mild regularity conditions we have that $\bar{\theta}(\cdot)$ is continuous at ξ_0 . Moreover,
13 we have the following result (Shapiro, 1985c, Lemma 3.1).

14 PROPOSITION 5.2. *Suppose that the set Θ can be approximated at $\theta_0 \in \Theta$ by a convex*
15 *cone $\mathcal{C} \subset \mathbb{R}^q$, the Jacobian matrix Δ_0 has full column rank q and $\bar{\theta}(\cdot)$ is continuous at*
16 *$\xi_0 = g(\theta_0)$. Then*

$$17 \bar{\theta}(\xi_0 + z) = \theta_0 + \bar{\beta}(z) + o(\|z\|), \quad (5.9)$$

18 where $\bar{\beta}(z)$ is the optimal solution of the problem

$$19 \min_{\beta \in \mathcal{C}} (z - \Delta_0 \beta)' V_0 (z - \Delta_0 \beta). \quad (5.10)$$

20 Note that since the approximating cone \mathcal{C} is assumed to be *convex*, $\text{rank}(\Delta_0) = q$
21 and the matrix V_0 is positive definite, the minimizer $\bar{\beta}(z)$ is unique for any $z \in \mathbb{R}^m$. If,
22 moreover, the point θ_0 is an interior point of Θ and hence $\mathcal{C} = \mathbb{R}^q$, then $\bar{\beta}(\cdot)$ is a linear
23 function and can be written explicitly as

$$24 \bar{\beta}(z) = (\Delta_0' V_0 \Delta_0)^{-1} \Delta_0' V_0 z. \quad (5.11)$$

25 Now if $\xi_0 \notin \Xi_0$, then the analysis becomes considerably more involved. It will
26 be beyond the scope of this paper to give a detailed description of such theory. We
27 refer the interested reader to Bonnans and Shapiro (2000) for a thorough development
28 of that theory. We give below some, relatively simple, results which will be relevant
29 for the statistical inference. In the optimization literature the following result, giving a
30 first order approximation of the optimal value function, is often referred to as Danskin
31 Theorem (Danskin, 1967).

32 PROPOSITION 5.3. *Let \mathfrak{S} be the set of optimal solutions of problem (4.1) for $x = \xi_0$.*
33 *Then*

$$34 \vartheta(\xi_0 + z) = \vartheta(\xi_0) + \min_{\xi \in \mathfrak{S}} g'_\xi z + o(\|z\|), \quad (5.12)$$

where $\mathfrak{g}_\xi := \partial F(\xi_0, \xi)/\partial \mathbf{x}$. In particular, if problem (4.1) has unique optimal solution ξ^* for $\mathbf{x} = \xi_0$, then

$$\frac{\partial \vartheta(\xi_0)}{\partial \mathbf{x}} = \frac{\partial F(\xi_0, \xi^*)}{\partial \mathbf{x}}. \tag{5.13}$$

Of course, if $\xi_0 \in \Xi_0$, then problem (4.1) has unique optimal solution $\xi^* = \xi_0$ and hence $\partial \vartheta(\xi_0)/\partial \mathbf{x} = 0$. The following result is a consequence of the Implicit Function Theorem (Shapiro, 1983, Theorem 4.2).

PROPOSITION 5.4. *Suppose that:*

- (i) for $\mathbf{x} = \xi_0$ problem (4.3) has unique optimal solution θ^* ,
- (ii) the point θ^* is an interior point of Θ ,
- (iii) $\bar{\theta}(\cdot)$ is continuous at θ^* ,
- (iv) the Hessian matrix $\mathbf{H}_{\theta\theta} := \partial^2 F(\xi_0, \mathbf{g}(\theta^*))/\partial \theta \partial \theta'$ is nonsingular.

Then $\bar{\theta}(\cdot)$ is continuously differentiable and $\vartheta(\cdot)$ is twice continuously differentiable at ξ_0 , and

$$\frac{\partial \bar{\theta}(\xi_0)}{\partial \mathbf{x}} = -\mathbf{H}_{\theta\theta}^{-1} \mathbf{H}_{\theta\mathbf{x}}, \tag{5.14}$$

$$\frac{\partial^2 \vartheta(\xi_0)}{\partial \mathbf{x} \partial \mathbf{x}'} = \mathbf{H}_{\mathbf{x}\mathbf{x}} - \mathbf{H}'_{\theta\mathbf{x}} \mathbf{H}_{\theta\theta}^{-1} \mathbf{H}_{\theta\mathbf{x}}, \tag{5.15}$$

where $\mathbf{H}_{\theta\mathbf{x}} := \partial^2 F(\xi_0, \mathbf{g}(\theta^*))/\partial \theta \partial \mathbf{x}'$ and $\mathbf{H}_{\mathbf{x}\mathbf{x}} := \partial^2 F(\xi_0, \mathbf{g}(\theta^*))/\partial \mathbf{x} \partial \mathbf{x}'$.

REMARK 4. If the model holds, and hence $\theta^* = \theta_0$, then $\mathbf{H}_{\theta\theta} = 2\Delta_0' V_0 \Delta_0$, $\mathbf{H}_{\theta\mathbf{x}} = -2\Delta_0' V_0$ and $\mathbf{H}_{\mathbf{x}\mathbf{x}} = 2V_0$ (compare with (4.7)). In that case formula (5.14) gives the same derivatives as (5.9) and (5.11), and (5.15) is equivalent to (5.7), and these formulas involve only first-order derivatives (i.e., the Jacobian matrix) of $\mathbf{g}(\cdot)$. On the other hand, if the model does not hold, and hence $\theta^* \neq \theta_0$, then these derivatives involve *second-order* derivatives of $\mathbf{g}(\cdot)$. Note also that the Hessian matrix $\mathbf{H}_{\theta\theta}$ can be nonsingular only if the Jacobian matrix $\Delta(\theta^*)$ has full column rank q and hence the model is locally identified at θ^* .

5.1. Asymptotics of MDF test statistics

Suppose that the model holds, i.e., $\xi_0 \in \Xi_0$. Since $\widehat{F} = \vartheta(\widehat{\xi}_n)$, we obtain from the approximation (5.6) the following asymptotic expansion of the MDF test statistic (under the null hypothesis):

$$n\widehat{F} = \min_{\zeta \in \mathcal{I}} (\mathbf{Z}_n - \zeta)' V_0 (\mathbf{Z}_n - \zeta) + o_p(1). \tag{5.16}$$

Recall that $\mathbf{Z}_n \Rightarrow N(\mathbf{0}, \Gamma)$. It follows that

$$n\widehat{F} \Rightarrow \min_{\zeta \in \mathcal{I}} (\mathbf{Z} - \zeta)' V_0 (\mathbf{Z} - \zeta), \tag{5.17}$$

where \mathbf{Z} is a random vector having normal distribution $N(\mathbf{0}, \mathbf{\Gamma})$. The optimal value of the right-hand side of (5.17) is a quadratic function of \mathbf{Z} . Under certain conditions this quadratic function $\mathbf{Z}'\mathbf{Q}\mathbf{Z}$ has a chi-square distribution (see, e.g., (Seber, 1977, Section 2.4)). In particular, this holds if $\mathbf{V}_0 = \mathbf{\Gamma}^{-1}$. As it was discussed earlier, nonsingularity of the covariance matrix $\mathbf{\Gamma}$ depends on a choice of the space where the saturated model is defined. In applications it is often convenient to take a larger space in which case $\mathbf{\Gamma}$ becomes singular. It is said that the discrepancy function is *correctly specified* if \mathbf{V}_0 is equal to a generalized inverse of $\mathbf{\Gamma}$, that is, $\mathbf{\Gamma}\mathbf{V}_0\mathbf{\Gamma} = \mathbf{\Gamma}$. Of course, if $\mathbf{\Gamma}$ is nonsingular, then this is the same as $\mathbf{V}_0 = \mathbf{\Gamma}^{-1}$. As it was mentioned earlier we assume that the asymptotic covariance matrix $\mathbf{\Gamma}$ has the maximal rank, e.g., in the case of covariance structures we assume that $\text{rank}(\mathbf{\Gamma}) = p(p+1)/2$. It follows then that each column vector of the Jacobian matrix $\mathbf{\Delta}(\boldsymbol{\theta})$ is contained in the linear space generated by columns of $\mathbf{\Gamma}$.

We have the following result (Browne, 1982; Shapiro, 1986) giving asymptotics of the null distribution of the MDF test statistic. Recall Definition 5.1 of a regular point.

THEOREM 5.1. *Suppose that the model holds, the discrepancy function is correctly specified and the point $\boldsymbol{\theta}_0$ is regular (and hence the set \mathcal{E}_0 is approximated at $\boldsymbol{\xi}_0 = \mathbf{g}(\boldsymbol{\theta}_0)$ by a linear space \mathcal{T} of the form (5.4)). Then the MDF test statistic $n\widehat{F}$ converges in distribution to a (central) chi-square with*

$$v = \text{rank}(\mathbf{\Gamma}) - \dim(\mathcal{T}) = \text{rank}(\mathbf{\Gamma}) - \text{rank}(\mathbf{\Delta}_0) \quad (5.18)$$

degrees of freedom.

Suppose that the mapping $\mathbf{g}(\cdot)$ is analytic and let r be the *characteristic rank* of the model. The above results imply that if the discrepancy function is correctly specified, then under the null hypothesis, generically, the MDF test statistic $n\widehat{F}$ has asymptotically a chi-square distribution, with $v = \text{rank}(\mathbf{\Gamma}) - r$ degrees of freedom. Recall that “generically” means that this holds for almost every population value $\boldsymbol{\theta}_0 \in \Theta$ of the parameter vector. For example, consider the setting of covariance structures and suppose that the population distribution is normal. Then the covariance matrix $\mathbf{\Gamma}$ can be written in the form (5.1), has rank $p(p+1)/2$ and matrix $\mathbf{V}_0 := \frac{1}{2}\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}$ is its generalized inverse. It follows that for a normally distributed population, both discrepancy functions F_{ML} and F_{GLS} , defined in (3.5) and (3.9), respectively, are correctly specified. Therefore, we have that:

Under the null hypothesis, generically, the MDF test statistics, associated with F_{ML} and F_{GLS} , are asymptotically chi-square distributed, with $v = p(p+1)/2 - r$ degrees of freedom, provided that the population distribution is normal.

It is possible to extend this basic result in various directions. Suppose now that the model does not hold, i.e., $\boldsymbol{\xi}_0 \notin \mathcal{E}_0$. Let $\boldsymbol{\xi}^*$ be a minimizer of $F(\boldsymbol{\xi}_0, \cdot)$ over \mathcal{E}_0 , i.e., $\boldsymbol{\xi}^*$ is an optimal solution of problem (4.1) for $\mathbf{x} = \boldsymbol{\xi}_0$. Since the model does not hold, we have here that $\boldsymbol{\xi}^* \neq \boldsymbol{\xi}_0$. Suppose, however, that the population value $\boldsymbol{\xi}_0$ is close to the model set \mathcal{E}_0 , i.e., there is no big difference between $\boldsymbol{\xi}^*$ and $\boldsymbol{\xi}_0$. We can employ approximation (5.6) at the point $\boldsymbol{\xi}^*$, instead of $\boldsymbol{\xi}_0$, by taking $\mathbf{V}_0 := \mathbf{V}(\boldsymbol{\xi}^*, \boldsymbol{\xi}^*)$,

i.e., making the second-order Taylor expansion of the discrepancy function at the point $(\mathbf{x}, \boldsymbol{\xi}) = (\boldsymbol{\xi}^*, \boldsymbol{\xi}^*)$, and using the tangent space \mathcal{T} at $\boldsymbol{\xi}^*$. We obtain the following approximation of the MDF statistic:

$$n\widehat{F} = \min_{\boldsymbol{\zeta} \in \mathcal{T}} (\mathbf{Z}_n^* - \boldsymbol{\zeta})' \mathbf{V}_0 (\mathbf{Z}_n^* - \boldsymbol{\zeta}) + o(\|\mathbf{Z}_n^*\|^2), \tag{5.19}$$

where

$$\mathbf{Z}_n^* := n^{1/2}(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}^*) = \underbrace{n^{1/2}(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)}_{\mathbf{Z}_n} + \underbrace{n^{1/2}(\boldsymbol{\xi}_0 - \boldsymbol{\xi}^*)}_{\boldsymbol{\mu}_n}.$$

Recall that it is assumed that $\mathbf{Z}_n \Rightarrow N(\mathbf{0}, \boldsymbol{\Gamma})$. On the other hand, as n tends to infinity, the “deterministic” part $\boldsymbol{\mu}_n := n^{1/2}(\boldsymbol{\xi}_0 - \boldsymbol{\xi}^*)$ of \mathbf{Z}_n^* grows indefinitely. However, the quadratic approximation, given by the right-hand side of (5.19), could be reasonable if the “stochastic” part \mathbf{Z}_n is bigger than the “deterministic” part $\boldsymbol{\mu}_n$ (we will discuss this in more details later). In order to formulate this in a mathematically rigorous way, we make the so-called assumption of a *sequence of local alternatives*. That is, we assume that there is a sequence $\boldsymbol{\xi}_0 = \boldsymbol{\xi}_{0,n}$ of population values (local alternatives) converging to a point $\boldsymbol{\xi}^* \in \Xi_0$ such that $n^{1/2}(\boldsymbol{\xi}_{0,n} - \boldsymbol{\xi}^*)$ converges to a (deterministic) vector $\boldsymbol{\mu}$ (this assumption is often referred to as *Pitman drift*). It follows then that $\mathbf{Z}_n^* \Rightarrow N(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ and the remainder term $o(\|\mathbf{Z}_n^*\|^2)$ in (5.19) converges in probability to zero. If, moreover, $\mathbf{V}_0 = \boldsymbol{\Gamma}^-$, then the quadratic term in (5.19) converges in distribution to noncentral chi-square with the same degrees of freedom ν and the noncentrality parameter

$$\delta = \min_{\boldsymbol{\zeta} \in \mathcal{T}} (\boldsymbol{\mu} - \boldsymbol{\zeta})' \mathbf{V}_0 (\boldsymbol{\mu} - \boldsymbol{\zeta}). \tag{5.20}$$

Moreover, by (5.6) we have that

$$\delta = \lim_{n \rightarrow \infty} \left[n \min_{\boldsymbol{\xi} \in \Xi_0} F(\boldsymbol{\xi}_{0,n}, \boldsymbol{\xi}) \right]. \tag{5.21}$$

This leads to the following result (Shapiro, 1983, Theorem 5.5; Steiger et al., 1985).

THEOREM 5.2. *Suppose that the assumption of a sequence of local alternatives (Pitman drift) holds, the discrepancy function is correctly specified and the set Ξ_0 is approximated at $\boldsymbol{\xi}^* = \mathbf{g}(\boldsymbol{\theta}^*)$ by a linear space \mathcal{T} generated by the columns of the matrix $\boldsymbol{\Delta}^* = \boldsymbol{\Delta}(\boldsymbol{\theta}^*)$. Then the MDF test statistic $n\widehat{F}$ converges in distribution to a noncentral chi-square with $\nu = \text{rank}(\boldsymbol{\Gamma}) - \dim(\mathcal{T})$ degrees of freedom and the noncentrality parameter δ given in (5.20) or, equivalently, (5.21).*

From the practical point of view it is important to understand when the noncentral chi-square distribution gives a reasonable approximation of the true distribution of the MDF test statistics. By the analysis of Section 4, we have that \widehat{F} converges w.p.1 to the value

$$F^* := \min_{\boldsymbol{\xi} \in \Xi_0} F(\boldsymbol{\xi}_0, \boldsymbol{\xi}) = F(\boldsymbol{\xi}_0, \boldsymbol{\xi}^*). \tag{5.22}$$

Recall that $\vartheta(x)$ denotes the optimal value of problem (4.1), and hence $\widehat{F} = \vartheta(\widehat{\xi}_n)$ and $F^* = \vartheta(\xi_0)$. Suppose that ξ^* is the *unique* minimizer of $F(\xi_0, \cdot)$ over Ξ_0 . Then we have by Danskin theorem (see Proposition 5.3) that $\partial\vartheta(\xi_0)/\partial x = \partial F(\xi_0, \xi^*)/\partial x$. It follows that

$$n^{1/2}(\widehat{F} - F^*) = \mathfrak{g}'_0 Z_n + o_p(1), \quad (5.23)$$

where $\mathfrak{g}_0 := \frac{\partial F(x, \xi^*)}{\partial x} \Big|_{x=\xi_0}$, which in turn implies the following asymptotic result (Shapiro, 1983, Theorem 5.3).

THEOREM 5.3. *Suppose that ξ^* is the unique minimizer of $F(\xi_0, \cdot)$ over Ξ_0 . Then*

$$n^{1/2}(\widehat{F} - F^*) \Rightarrow N(0, \mathfrak{g}'_0 \Gamma \mathfrak{g}_0). \quad (5.24)$$

If the model holds, then $F^* = 0$ and $\mathfrak{g}_0 = \mathbf{0}$. In that case the asymptotic result (5.24) degenerates into the trivial statement that $n^{1/2}\widehat{F}$ converges in probability to zero. And, indeed, as it was discussed above, under the null hypothesis one needs to scale \widehat{F} by the factor of n , instead of $n^{1/2}$, in order to get meaningful asymptotics. However, as the distance between the population value ξ_0 and the model set Ξ_0 becomes larger, the noncentral chi-square distribution approximation deteriorates and the normal distribution, with mean F^* and variance $n^{-1}\mathfrak{g}'_0 \Gamma \mathfrak{g}_0$, could become a better approximation of the distribution of \widehat{F} . The noncentral chi-square approximation is based on the distribution of the quadratic form

$$\begin{aligned} \min_{\zeta \in \mathcal{T}} (Z_n^* - \zeta)' V_0 (Z_n^* - \zeta) \\ = Z_n^{*'} Q Z_n^* = Z_n' Q Z_n + \mu_n' Q \mu_n + 2\mu_n' Q Z_n. \end{aligned} \quad (5.25)$$

Recall that $Z_n^* = Z_n + \mu_n$ and $\mu_n = n^{1/2}(\xi_0 - \xi^*)$. The first term, in the right-hand side of (5.25), has approximately a central chi-square distribution with $\nu = \text{rank}(\Gamma) - \dim(\mathcal{T})$ degrees of freedom. Suppose that ξ_0 is close to Ξ_0 . By (5.6), the second term in the right-hand side of (5.25) can be approximated as follows

$$\mu_n' Q \mu_n = n \min_{\zeta \in \mathcal{T}} (\xi_0 - \xi^* - \zeta)' V_0 (\xi_0 - \xi^* - \zeta) \approx n F^*. \quad (5.26)$$

Recall that, by (5.21), $n F^*$ is approximately equal to the noncentrality parameter δ . We also have that

$$\begin{aligned} \mathfrak{g}_0 &= \frac{\partial F(x, \xi^*)}{\partial x} \Big|_{x=\xi_0} \approx \frac{\partial (x - \xi^*)' V_0 (x - \xi^*)}{\partial x} \Big|_{x=\xi_0} \\ &= 2V_0(\xi_0 - \xi^*). \end{aligned} \quad (5.27)$$

Moreover, by the first-order optimality conditions, the gradient vector \mathfrak{g}_0 is orthogonal to the space \mathcal{T} , and hence $V_0(\xi_0 - \xi^*) \approx Q(\xi_0 - \xi^*)$.

It follows that the sum of the second and third terms in the right-hand side of (5.25) has approximately a normal distribution with mean $n F^*$ and variance $n \mathfrak{g}'_0 \Gamma \mathfrak{g}_0$. Therefore, for ξ_0 close to Ξ_0 the difference between the noncentral chi-square and normal approximations, given in Theorems 5.2 and 5.3, respectively, is the first term in the

right-hand side of (5.25). The expected value and the variance of a chi-square random variable with ν degrees of freedom is equal to ν and 2ν , respectively. Therefore, when the number of degrees of freedom $\nu = \text{rank}(\Gamma) - \text{dim}(T)$ is bigger or comparable with the noncentrality parameter $\delta \approx nF^*$, the noncentral chi-square approximation, which is based on a second-order Taylor expansion at the point ξ^* , should be better than the normal approximation, which is based on a first-order approximation at ξ_0 . On the other hand, if δ is significantly bigger than ν , then the first term in the right-hand side of (5.25) becomes negligible and the normal approximation could be reasonable. This is in agreement with the property that a noncentral chi-square distribution, with ν degrees of freedom and noncentrality parameter δ , becomes approximately normal if δ is much bigger than ν . In such a case the normal approximation can be used to construct a confidence interval, for F^* , of the form $\widehat{F} \pm \kappa \widehat{\sigma}_F$. Here $\widehat{\sigma}_F$ is an estimate of the standard deviation of \widehat{F} and κ is a critical value. Recall that the expected value and variance of a noncentral chi-square random variable, with ν degrees of freedom and noncentrality parameter δ , is $\nu + \delta$ and $2\nu + 4\delta$, respectively. Therefore, the normal approximation could be reasonable if

$$\frac{n\widehat{F} - \nu}{\sqrt{4n\widehat{F} + 2\nu}} \geq \kappa, \tag{5.28}$$

where κ is a critical value, say $\kappa = 3$.

Let us finally remark that from a theoretical point of view one can obtain a better approximation of the distribution of the MDF test statistic by using a second-order Taylor expansion of the optimal value function at the population point ξ_0 . The corresponding first- and second-order derivatives are given in (5.13) and (5.15), respectively, provided that the optimal solution ξ^* is unique. Note, however, that in practical applications this will require an accurate estimation of the corresponding first- and second-order derivatives which could be a problem.

5.2. Nested models

Suppose now that we have two models $\mathcal{E}_1 \subset \mathcal{E}$ and $\mathcal{E}_2 \subset \mathcal{E}$ for the same parameter vector ξ . It is said that the second model is *nested*, within the first model, if \mathcal{E}_2 is a subset of \mathcal{E}_1 , i.e., $\mathcal{E}_2 \subset \mathcal{E}_1$. We refer to the models associated with the sets \mathcal{E}_1 and \mathcal{E}_2 as *full* and *restricted* models, respectively. If \mathcal{E}_1 is given in the parametric form

$$\mathcal{E}_1 := \{ \xi \in \mathcal{E} : \xi = g(\theta), \theta \in \Theta_1 \}, \tag{5.29}$$

i.e., the full model is structural, then it is natural to define a nested model by restricting the parameter space Θ_1 to a subset Θ_2 . Typically the subset $\Theta_2 \subset \Theta_1$ is defined by imposing constraints on the parameter vector θ . In this section we discuss asymptotics of the MDF test statistics $n\widehat{F}_i, i = 1, 2$, where

$$\widehat{F}_i := \min_{\xi \in \mathcal{E}_i} F(\widehat{\xi}_n, \xi). \tag{5.30}$$

Suppose that the population value $\xi_0 \in \mathcal{E}_2$, i.e., the restricted model holds. Suppose, further, that the sets \mathcal{E}_1 and \mathcal{E}_2 are approximated at ξ_0 by linear spaces \mathcal{T}_1 and \mathcal{T}_2 ,

1 respectively. This holds if both \mathcal{E}_1 and \mathcal{E}_2 are smooth manifolds near ξ_0 with respective
 2 tangent spaces \mathcal{T}_1 and \mathcal{T}_2 . Note that $\mathcal{T}_2 \subset \mathcal{T}_1$ since $\mathcal{E}_2 \subset \mathcal{E}_1$. We have then (see (5.16))
 3 that

$$4 \quad n\widehat{F}_i = \min_{\xi \in \mathcal{T}_i} (\mathbf{Z}_n - \xi)' \mathbf{V}_0 (\mathbf{Z}_n - \xi) + o_p(1), \quad i = 1, 2. \quad (5.31)$$

5
 6 Suppose, further, that the discrepancy function is *correctly specified*. Then by the analy-
 7 sis of the previous section we have that $n\widehat{F}_i$, $i = 1, 2$, converges in distribution to a
 8 (central) chi-square with $\nu_i = \text{rank}(\mathbf{\Gamma}) - \dim(\mathcal{T}_i)$ degrees of freedom. Moreover, it
 9 follows from the representation (5.31) that $n\widehat{F}_1$ and $n\widehat{F}_2 - n\widehat{F}_1$ are asymptotically in-
 10 dependent of each other. The corresponding arguments are analogous to derivations of
 11 the statistical inference of linear constraints in the theory of linear models (e.g., [Seber,](#)
 12 [1977](#), Section 4.5.1). This can be extended to the setting of a sequence of local alterna-
 13 tives, where there is a sequence $\xi_{0,n}$ of population values converging to a point $\xi^* \in \mathcal{E}_2$
 14 such that the following limits exist

$$15 \quad \delta_i = \lim_{n \rightarrow \infty} \left[n \min_{\xi \in \mathcal{E}_i} F(\xi_{0,n}, \xi) \right], \quad i = 1, 2. \quad (5.32)$$

16
 17 Then the following asymptotic results hold ([Steiger et al., 1985](#)).

18
 19
 20
 21 **THEOREM 5.4.** *Suppose that the assumption of a sequence of local alternatives holds,*
 22 *the discrepancy function is correctly specified and the sets \mathcal{E}_i , $i = 1, 2$, are approxi-*
 23 *imated at $\xi^* \in \mathcal{E}_2$ by respective linear spaces \mathcal{T}_i . Then the following holds:*

- 24 (i) *the MDF test statistics $n\widehat{F}_i$ converge in distribution to noncentral chi-square with*
 25 *respective degrees of freedom $\nu_i = \text{rank}(\mathbf{\Gamma}) - \dim(\mathcal{T}_i)$ and noncentrality parameter*
 26 *δ_i given in (5.32),*
 27 (ii) *the statistic $n\widehat{F}_2 - n\widehat{F}_1$ converges in distribution to a noncentral chi-square with*
 28 *$\nu_2 - \nu_1$ degrees of freedom and noncentrality parameter $\delta_2 - \delta_1$,*
 29 (iii) *the statistics $n\widehat{F}_1$ and $n\widehat{F}_2 - n\widehat{F}_1$ are asymptotically independent of each other,*
 30 (iv) *the ratio statistic $((\widehat{F}_2 - \widehat{F}_1)/(\nu_2 - \nu_1))/(\widehat{F}_1/\nu_1)$ converges in distribution to dou-*
 31 *bly noncentral F -distribution with noncentrality parameters $\delta_2 - \delta_1$ and δ_1 and*
 32 *with $\nu_2 - \nu_1$ and ν_1 degrees of freedom.*

33
 34 It is straightforward to extend the above result to a sequence of nested models. Also
 35 we have that $n\widehat{F}_2 = n\widehat{F}_1 + (n\widehat{F}_2 - n\widehat{F}_1)$. Recall that the variance of a noncentral
 36 chi-square random variable with ν degrees of freedom and noncentrality parameter δ
 37 is $2\nu + 4\delta$. Therefore, under the assumptions of the above theorem, the asymptotic
 38 covariance between $n\widehat{F}_1$ and $n\widehat{F}_2$ is equal to the asymptotic variance of $n\widehat{F}_1$, which is
 39 equal to $2\nu_1 + 4\delta_1$. Consequently, the asymptotic correlation between the MDF statistics
 40 $n\widehat{F}_1$ and $n\widehat{F}_2$ is equal to $\sqrt{(\nu_1 + 2\delta_1)/(\nu_2 + 2\delta_2)}$ ([Steiger et al., 1985](#)).

41 5.3. Asymptotics of MDF estimators

42
 43 In this section we discuss asymptotics of the MDF estimator $\hat{\theta}_n$. Suppose that the model
 44 holds and $\hat{\theta}_n$ is a consistent estimator of the population value $\theta_0 \in \Theta$ (see Section 4
 45

and Proposition 4.2 in particular). Since $\hat{\theta}_n = \bar{\theta}(\hat{\xi}_n)$, we have by (5.9) that, under the assumptions of Proposition 5.2,

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \bar{\beta}(Z_n) + o_p(1). \tag{5.33}$$

Recall that $\bar{\beta}(z)$ is the optimal solution of (5.10) and note that $\bar{\beta}(\cdot)$ is positively homogeneous, i.e., $\bar{\beta}(tz) = t\bar{\beta}(z)$ for any z and $t \geq 0$. This leads to the following asymptotics of the MDF estimator (Browne, 1974; Shapiro, 1983).

THEOREM 5.5. *Suppose that the model holds, $\hat{\theta}_n$ is a consistent estimator of θ_0 , the set Θ is approximated at θ_0 by a convex cone \mathcal{C} and $\text{rank}(\Delta_0) = q$. Then $n^{1/2}(\hat{\theta}_n - \theta_0) \Rightarrow \bar{\beta}(Z)$, where $Z \sim N(0, \Gamma)$. If, furthermore, θ_0 is an interior point of Θ , and hence $\mathcal{C} = \mathbb{R}^q$, then $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to normal with mean vector zero and covariance matrix*

$$\Pi = (\Delta'_0 V_0 \Delta_0)^{-1} \Delta'_0 V_0 \Gamma V_0 \Delta_0 (\Delta'_0 V_0 \Delta_0)^{-1}. \tag{5.34}$$

Moreover, if the discrepancy function is correctly specified, then $\Pi = (\Delta'_0 V_0 \Delta_0)^{-1}$.

In particular, if in the setting of covariance structures the population distribution is normal and the employed discrepancy function is normal-theory correctly specified, then the asymptotic covariance matrix of $n^{1/2}(\hat{\theta}_n - \theta_0)$ can be written as

$$\Pi_N = 2[\Delta'_0 (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \Delta_0]^{-1}. \tag{5.35}$$

Note that since it is assumed that the asymptotic covariance matrix Γ has the maximal rank, and hence the linear space generated by columns of Δ_0 is included in the linear space generated by columns of Γ , we have here that matrix $\Delta'_0 \Gamma^{-1} \Delta_0$ is independent of a particular choice of the generalized inverse Γ^{-1} and is positive definite. In particular, if the discrepancy function is correctly specified, then $\Delta'_0 \Gamma^{-1} \Delta_0 = \Delta'_0 V_0 \Delta_0$. It is possible to show (Browne, 1974, Proposition 3) that the inequality

$$\Pi \geq (\Delta'_0 \Gamma^{-1} \Delta_0)^{-1} \tag{5.36}$$

always holds. (For $q \times q$ symmetric matrices A and B the inequality $A \geq B$ is understood in the Loewner sense, i.e., that matrix $A - B$ is positive semidefinite.) Basically this is the Gauss–Markov theorem. That is, for a correctly specified discrepancy function, the asymptotic covariance matrix of the corresponding MDF estimator attains its lower bound given by the right-hand side of (5.36). Therefore, for a correctly specified discrepancy function the corresponding MDF estimator is *asymptotically efficient* within the class of MDF estimators.

The above asymptotics of MDF estimators were derived under the assumption of identifiability of the model. If the model is overparameterized, then it does not make sense to talk about distribution of the MDF estimators since these estimators are not uniquely defined. However, even in the case of overparameterization some of the parameters could be defined uniquely. Therefore it makes sense to consider the following concept of *estimable functions* borrowed from the theory of linear models (e.g., Seber, 1977, Section 3.8.2).

1 DEFINITION 5.2. Consider a continuously differentiable function $a(\boldsymbol{\theta})$. We say that
 2 $\alpha = a(\boldsymbol{\theta})$ is an *estimable function* (of the parameter vector $\boldsymbol{\theta}$) if $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ and $\mathbf{g}(\boldsymbol{\theta}_1) =$
 3 $\mathbf{g}(\boldsymbol{\theta}_2)$ imply that $a(\boldsymbol{\theta}_1) = a(\boldsymbol{\theta}_2)$. If this holds in a neighborhood of a point $\boldsymbol{\theta}_0 \in \Theta$, we
 4 say α is *locally estimable*, near $\boldsymbol{\theta}_0$.

5
 6 In the analysis of covariance structures the above concept of estimable functions was
 7 discussed in Shapiro (1986, p. 146). By using local reparameterization (see Proposi-
 8 tion 2.1) it is possible to show the following.

9 If $\boldsymbol{\theta}_0$ is a locally regular interior point of Θ , then a parameter $\alpha = a(\boldsymbol{\theta})$ is locally
 10 estimable, near $\boldsymbol{\theta}_0$, iff vector $\partial a(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ belongs to the linear space generated by
 11 rows of $\mathbf{A}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

12 Consider now vector $\boldsymbol{\alpha}(\boldsymbol{\theta}) = (a_1(\boldsymbol{\theta}), \dots, a_s(\boldsymbol{\theta}))$ of locally estimable parameters,
 13 near a population point $\boldsymbol{\theta}_0$. Suppose that $\boldsymbol{\theta}_0$ is locally regular and let $\boldsymbol{\alpha}_0 := \boldsymbol{\alpha}(\boldsymbol{\theta}_0)$ and
 14 $\hat{\boldsymbol{\alpha}}_n := \boldsymbol{\alpha}(\hat{\boldsymbol{\theta}}_n)$. Note that, by local estimability of $\boldsymbol{\alpha}$, the estimator $\hat{\boldsymbol{\alpha}}_n$ is defined uniquely
 15 for $\hat{\boldsymbol{\theta}}_n$ sufficiently close to $\boldsymbol{\theta}_0$. We have then that $n^{1/2}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)$ converges in distribution
 16 to normal with mean vector zero and covariance matrix

$$17 \mathbf{A}_0(\mathbf{A}'_0 \mathbf{V}_0 \mathbf{A}_0)^{-1} \mathbf{A}'_0 \mathbf{V}_0 \boldsymbol{\Gamma} \mathbf{V}_0 \mathbf{A}_0 (\mathbf{A}'_0 \mathbf{V}_0 \mathbf{A}_0)^{-1} \mathbf{A}'_0, \quad (5.37)$$

18 where $\mathbf{A}_0 := \partial \boldsymbol{\alpha}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'$ is $s \times q$ Jacobian matrix. Note that because of the local
 19 estimability of $\boldsymbol{\alpha}$, we have that row vectors of \mathbf{A}_0 belong to the linear space generated
 20 by rows of \mathbf{A}_0 , and hence the expression in (5.37) does not depend on a particular choice
 21 of the generalized inverse of $\mathbf{A}'_0 \mathbf{V}_0 \mathbf{A}_0$. In particular, for correctly specified discrepancy
 22 function this expression becomes $\mathbf{A}_0(\mathbf{A}'_0 \mathbf{V}_0 \mathbf{A}_0)^{-1} \mathbf{A}'_0$.

23 We can also consider a situation when the model does not hold. Under the assump-
 24 tions of Proposition 5.4, in particular, that $\boldsymbol{\theta}^*$ is the unique optimal solution of problem
 25 (4.3), we have that $\hat{\boldsymbol{\theta}}_n$ converges w.p.1 to $\boldsymbol{\theta}^*$ and, by (5.14), that (Shapiro, 1983, Theo-
 26 rem 5.4):

$$27 n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \Rightarrow N(0, \mathbf{H}_{\theta\theta}^{-1} \mathbf{H}_{\theta x} \boldsymbol{\Gamma} \mathbf{H}'_{\theta x} \mathbf{H}_{\theta\theta}^{-1}). \quad (5.38)$$

28 5.4. The situation where the population value of the parameter vector lies on the 29 boundary of the parameter space

30
 31 In the previous sections we discussed asymptotics of MDF test statistics and estimators
 32 under the assumption that the set \mathcal{E}_0 can be approximated, at a considered point, by a
 33 linear space \mathcal{T} . In this section we consider a situation when $\boldsymbol{\theta}_0$ is a boundary point of
 34 the set Θ , and as a consequence the set \mathcal{E}_0 should be approximated at $\boldsymbol{\xi}_0 = \mathbf{g}(\boldsymbol{\theta}_0)$ by
 35 a (convex) cone rather than a linear space. This may happen if the set Θ is defined by
 36 inequality constraints and some of these inequality constraints are active at the popu-
 37 lation point. For instance, in Example 2.1 (Factor Analysis model) the diagonal entries
 38 of the matrix $\boldsymbol{\Psi}$ should be nonnegative. If the population value $\boldsymbol{\Psi}_0$ has zero diagonal
 39 entries (i.e., some residual variances are zeros), then the corresponding value of the
 40 parameter vector can be viewed as lying on the boundary of the feasible region. One
 41 can think about more sophisticated examples where, for instance, it is hypothesized that
 42 some of the residual variances (i.e., diagonal entries of $\boldsymbol{\Psi}$) are bigger than the others,
 43
 44
 45

or that elements of matrix \mathbf{A} are nonnegative. Statistical theory of parameter estimation under inequality type constraints is often referred to as order restricted statistical inference. The interested reader can be referred to the recent comprehensive monograph (Silvapulle and Sen, 2005) for a thorough treatment of that theory. We give below a few basic results which are relevant for our discussion.

Suppose that the model holds, the Jacobian matrix \mathbf{A}_0 has full column rank q and the set Θ is approximated at θ_0 by convex cone \mathcal{C} . We have then that

$$n\widehat{F} \Rightarrow \min_{\beta \in \mathcal{C}} (\mathbf{Z} - \mathbf{A}_0\beta)' \mathbf{V}_0 (\mathbf{Z} - \mathbf{A}_0\beta), \tag{5.39}$$

where $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Gamma})$. Let us look at the optimal value of the minimization problem in the right-hand side of (5.39). Suppose, for the sake of simplicity, that \mathbf{A}_0 has full column rank q . Then we can decompose this optimal value into a sum of two terms as follows

$$\begin{aligned} & \min_{\beta \in \mathcal{C}} (\mathbf{Z} - \mathbf{A}_0\beta)' \mathbf{V}_0 (\mathbf{Z} - \mathbf{A}_0\beta) \\ &= (\mathbf{Z} - \mathbf{A}_0\tilde{\beta})' \mathbf{V}_0 (\mathbf{Z} - \mathbf{A}_0\tilde{\beta}) + \min_{\beta \in \mathcal{C}} (\tilde{\beta} - \beta)' \mathbf{A}_0' \mathbf{V}_0 \mathbf{A}_0 (\tilde{\beta} - \beta), \end{aligned} \tag{5.40}$$

where $\tilde{\beta} = (\mathbf{A}_0' \mathbf{V}_0 \mathbf{A}_0)^{-1} \mathbf{A}_0' \mathbf{V}_0 \mathbf{Z}$ is the corresponding unconstrained minimizer (compare with (5.11)). The term in the left-hand side of (5.40) can be viewed as a squared distance from \mathbf{Z} to the cone $\{\mathbf{A}_0\beta: \beta \in \mathcal{C}\}$, where the distance is defined with respect to the weight matrix \mathbf{V}_0 . The first term in the right-hand side of (5.40) is the corresponding unconstrained minimum over $\beta \in \mathbb{R}^q$, and can be viewed as the squared distance from \mathbf{Z} to the linear space generated by \mathbf{A}_0 , and the second term can be considered as the squared distance from $\tilde{\beta}$ to \mathcal{C} . In a sense the above decomposition (5.40) is just the Pythagoras Theorem. Suppose, further, that the discrepancy function is correctly specified. Recall that it is assumed that the asymptotic covariance matrix $\mathbf{\Gamma}$ has maximal rank. By reducing the saturated space, if necessary, we can assume here that $\mathbf{\Gamma}$ is nonsingular, and hence “correctly specified” means that $\mathbf{V}_0 = \mathbf{\Gamma}^{-1}$. It follows then that $\tilde{\beta} \sim N(\mathbf{0}, (\mathbf{A}_0' \mathbf{V}_0 \mathbf{A}_0)^{-1})$.

Assuming that the model holds and $\mathbf{V}_0 = \mathbf{\Gamma}^{-1}$, we have that the first term in the right-hand side of (5.40) has chi-square distribution, with $\nu = m - q$ degrees of freedom, and is distributed independently of the second term. The second term in the right-hand side of (5.40) follows a mixture of chi-square distributions (such distributions are called *chi-bar-squared* distributions). With various degrees of generality this result was derived in (Bartholomew, 1961; Kudô, 1963; Nüesch, 1966), it was shown in (Shapiro, 1985d) that this holds for any convex cone \mathcal{C} . Denote by $n\tilde{F}$ the corresponding unconstrained MDF test statistic, i.e.,

$$\tilde{F} := \min_{\theta \in \mathbb{R}^q} F(\hat{\xi}_n, \mathbf{g}(\theta)).$$

We have that $n\tilde{F}$ is asymptotically equivalent to the first term in the right-hand side of (5.40). Under the above assumptions we obtain the following results:

- (i) The unrestricted MDF test statistic $n\tilde{F}$ converges in distribution to chi-square with $\nu = m - q$ degrees of freedom and is asymptotically independent of the difference statistic $n\widehat{F} - n\tilde{F}$.

(ii) The difference statistic $n\hat{F} - n\tilde{F}$ is asymptotically equivalent to the second term in the right hand side of (5.40) and converges in distribution to a mixture of chi-square distributions, that is,

$$\lim_{n \rightarrow \infty} \text{Prob}\{n\hat{F} - n\tilde{F} \geq c\} = \sum_{i=0}^q w_i \text{Prob}\{\chi_i^2 \geq c\}, \quad (5.41)$$

where χ_i^2 is a chi-square random variable with i degrees of freedom, $\chi_0^2 \equiv 0$ and w_i are nonnegative weights such that $w_0 + \dots + w_q = 1$.

Of course, the asymptotic distribution, given by the right-hand side of (5.41), depends on the weights w_i , which in turn depend on the covariance matrix of $\hat{\beta}$ and cone \mathcal{C} (recall that, for correctly specified discrepancy function, the covariance matrix of $\hat{\beta}$ is $(\mathbf{D}'_0 \mathbf{V}_0 \mathbf{D}_0)^{-1}$). A general property of these weights is that $\sum_{i=0}^q (-1)^i w_i = 0$ if at least two of these weights are nonzeros. If θ_0 is an interior point of Θ , and hence $\mathcal{C} = \mathbb{R}^q$, then $w_0 = 1$ and all other weights are zeros. In that case we have the same asymptotics of the MDF statistic $n\hat{F}$ as given in Theorem 5.1. Often the set $\Theta \subset \mathbb{R}^q$ is defined by inequality constraints. Then, under mild regularity conditions (called *constraint qualifications*), the approximating cone \mathcal{C} is obtained by linearizing the active at θ_0 inequality constraints. In particular, if only one inequality constraint is active at θ_0 , then \mathcal{C} is defined by one linear inequality constraint and hence is a half space of \mathbb{R}^q . In that case $w_0 = w_1 = 1/2$ and all other weights are zeros. If two inequality constraints are active at θ_0 , then only weights w_0, w_1 and w_2 can be nonzeros, with $w_1 = 1/2$, etc. For a general discussion of how to calculate these weights we can refer, e.g., to (Shapiro, 1988; Silvapulle and Sen, 2005).

6. Asymptotic robustness of the MDF statistical inference

An important condition in the analysis of the previous section was the assumption of correct specification of the discrepancy function. In particular, the discrepancy functions F_{ML} and F_{GLS} , defined in (3.5) and (3.9), respectively, are motivated by the assumption that the underlying population has a normal distribution and are correctly specified in that case. Nevertheless, these discrepancy functions are often applied in situations where the normality assumption has no justification or even can be clearly wrong. It turns out, however, that the asymptotic chi-square distribution of MDF test statistics, discussed in Theorems 5.1 and 5.2, can hold under considerably more general conditions than correct specification of the discrepancy function. This was discovered in Amemiya and Anderson (1990) and Anderson and Amemiya (1988) for a class of factor analysis models, and in Browne (1987), Browne and Shapiro (1988), and Shapiro (1987) for general linear models, by using approaches based on different techniques. In this section we are going to discuss this theory following the Browne–Shapiro approach, which in a sense is more general although uses a slightly stronger assumption of existence of fourth-order moments. As in the previous section we assume that $n^{1/2}(\hat{\xi}_n - \xi_0) \Rightarrow N(\mathbf{0}, \mathbf{\Gamma})$.

Let us start with the following algebraic result (Shapiro, 1987, Theorem 3.1). It is based on a verification that the corresponding quadratic form has a chi-square distribution. (For the sake of notational convenience we drop here the superscript of the Jacobian matrix $\mathbf{\Delta}^*$.)

PROPOSITION 6.1. *Suppose that the assumption of a sequence of local alternatives holds and the set Ξ_0 is approximated at the point $\xi^* = \mathbf{g}(\theta^*)$ by a linear space \mathcal{T} generated by the columns of the matrix $\mathbf{\Delta} = \mathbf{\Delta}(\theta^*)$ (e.g., the point θ^* is regular). Suppose, further, that the discrepancy function is correctly specified with respect to an $m \times m$ positive semidefinite matrix $\mathbf{\Gamma}_0$ of maximal rank. Then $n\hat{F}$ is asymptotically chi-squared, with degrees of freedom ν and the noncentrality parameter δ given in (5.18) and (5.21), respectively, if and only if $\mathbf{\Gamma}$ is representable in the form*

$$\mathbf{\Gamma} = \mathbf{\Gamma}_0 + \mathbf{\Delta}\mathbf{C}' + \mathbf{C}\mathbf{\Delta}', \tag{6.1}$$

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

In particular, we have that for the normal-theory discrepancy functions F_{ML} and F_{GLS} (in the analysis of covariance structures) the MDF test statistics are asymptotically chi-squared if and only if the corresponding $p^2 \times p^2$ asymptotic covariance matrix $\mathbf{\Gamma}$ can be represented in the form

$$\mathbf{\Gamma} = \mathbf{\Gamma}_N + \mathbf{\Delta}\mathbf{C}' + \mathbf{C}\mathbf{\Delta}', \tag{6.2}$$

where matrix $\mathbf{\Gamma}_N$ is defined in (5.1).

The representation (6.1) is slightly more general than the following representation

$$\mathbf{\Gamma} = \mathbf{\Gamma}_0 + \mathbf{\Delta}\mathbf{D}\mathbf{\Delta}', \tag{6.3}$$

where \mathbf{D} is an arbitrary $q \times q$ symmetric matrix. Clearly, (6.3) is a particular form of the representation (6.1) with $\mathbf{C} := \frac{1}{2}\mathbf{\Delta}\mathbf{D}$. It turns out that under various structural assumptions, in the analysis of covariance structures, it is possible to show that the corresponding asymptotic covariance matrix $\mathbf{\Gamma}$ is of the form

$$\mathbf{\Gamma} = \mathbf{\Gamma}_N + \mathbf{\Delta}\mathbf{D}\mathbf{\Delta}', \tag{6.4}$$

and hence to verify that the normal-theory MDF test statistics have asymptotically chi-square distributions. We also have the following result about asymptotic robustness of MDF estimators (Shapiro, 1987, Corollary 5.4).

PROPOSITION 6.2. *Suppose that the model holds, the set Ξ_0 is approximated at the point ξ_0 by a linear space \mathcal{T} generated by the columns of the matrix $\mathbf{\Delta} = \mathbf{\Delta}(\theta_0)$, the Jacobian matrix $\mathbf{\Delta}$ has full column rank q , the MDF estimator $\hat{\theta}_n$ is a consistent estimator of θ_0 , the discrepancy function is correctly specified with respect to an $m \times m$ positive semidefinite matrix $\mathbf{\Gamma}_0$ of maximal rank, and the representation (6.3) holds. Then $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to normal $N(\mathbf{0}, \mathbf{\Pi})$, the MDF estimator $\hat{\theta}_n$ is asymptotically efficient within the class of MDF estimators, and*

$$\mathbf{\Pi} = \mathbf{\Pi}_0 + \mathbf{D}, \tag{6.5}$$

where $\mathbf{\Pi}_0 := (\mathbf{\Delta}'\mathbf{V}_0\mathbf{\Delta})^{-1}$.

We have that if the representation (6.3) holds, then the MDF test statistics designed for the asymptotic covariance matrix $\mathbf{\Gamma}_0$ still have asymptotically a chi-square distribution (under a sequence of local alternatives) and the asymptotic covariance matrix of the corresponding MDF estimators needs a simple correction given by formula (6.5). In the remainder of this section we discuss situations in the analysis of covariance structures which lead to the representation (6.4).

We assume below, in the remainder of this section, the setting of the analysis of covariance structures, with structural model $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ and with $\mathbf{\Gamma}$ being the $p^2 \times p^2$ asymptotic covariance matrix of $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$, where $\mathbf{s} := \text{vec}(\mathbf{S})$ and $\boldsymbol{\sigma}_0 := \text{vec}(\mathbf{\Sigma}_0)$. We assume that the underlying population has finite fourth-order moments, and hence the asymptotic covariance matrix $\mathbf{\Gamma}$ is well defined. As before, we denote by $\mathbf{\Gamma}_N$ and $\mathbf{\Pi}_N$ the normal-theory asymptotic covariance matrices given in (5.1) and (5.35), respectively. We also assume that the employed discrepancy function is *correctly specified with respect to a normal distribution* of the data, i.e., V_0 is a generalized inverse of $\mathbf{\Gamma}_N$. Recall that the normal-theory discrepancy functions F_{ML} and F_{GLS} satisfy this property.

6.1. Elliptical distributions

In this section we assume that the underlying population has an *elliptical* distribution. We may refer to (Muirhead, 1982) for a thorough discussion of elliptical distributions. In the case of elliptical distributions the asymptotic covariance matrix $\mathbf{\Gamma}$ has the following structure:

$$\mathbf{\Gamma} = \alpha \mathbf{\Gamma}_N + \beta \boldsymbol{\sigma}_0 \boldsymbol{\sigma}'_0, \quad (6.6)$$

where $\alpha = 1 + \kappa$, $\beta = \kappa$, and κ is the kurtosis parameter of a considered elliptical distribution. This basic asymptotic result was employed in the studies of Muirhead and Waternaux (1980), Tyler (1982, 1983) and Browne (1982, 1984).

It can be seen that the corrected covariance matrix $\alpha^{-1} \mathbf{\Gamma}$ has the structure specified in Eq. (6.4), provided that $\boldsymbol{\sigma}_0$ can be represented as a linear combination of columns of the Jacobian matrix $\mathbf{\Delta} = \mathbf{\Delta}(\boldsymbol{\theta}_0)$. If the point $\boldsymbol{\theta}_0$ is regular, and hence \mathcal{E}_0 is a smooth manifold near $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(\boldsymbol{\theta}_0)$ and the tangent space \mathcal{T} , to \mathcal{E}_0 at $\boldsymbol{\sigma}_0$, is the linear space generated by the columns of $\mathbf{\Delta}$ (i.e., can be written in the form (5.4)), then this condition is equivalent to the condition that $\boldsymbol{\sigma}_0 \in \mathcal{T}$. This, in turn, holds if the set \mathcal{E}_0 is *positively homogeneous*, i.e., it satisfies the property that if $\boldsymbol{\sigma} \in \mathcal{E}_0$ and $t > 0$, then $t\boldsymbol{\sigma} \in \mathcal{E}_0$. For structural models, positive homogeneity of \mathcal{E}_0 can be formulated in the following form (this condition was introduced in Browne (1982) where models satisfying this condition were called *invariant under a constant scaling factor*):

(C) For every $t > 0$ and $\boldsymbol{\theta} \in \Theta$ there exists $\boldsymbol{\theta}^* \in \Theta$ such that $t \mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Sigma}(\boldsymbol{\theta}^*)$.

The above condition (C) is easy to verify and it holds for many models used in applications. For example, it holds for the factor analysis model (2.2). By the above discussion we have the following results, which in somewhat different forms were obtained in Tyler (1983) and Browne (1982, 1984), and in the present form in Shapiro and Browne (1987).

1 THEOREM 6.1. Suppose that the assumption of a sequence of local alternatives holds, 1
 2 the set Ξ_0 is approximated at the point $\sigma^* = \sigma(\theta^*)$ by a linear space \mathcal{T} , the represen- 2
 3 tation (6.6) holds with $\alpha > 0$ and that $\sigma_0 \in \mathcal{T}$. Let $\hat{\alpha}$ be a consistent estimator of the 3
 4 parameter α . Then $\hat{\alpha}^{-1}n\hat{F}$ has asymptotically a chi-squared distribution with ν degrees 4
 5 of freedom and the noncentrality parameter $\alpha^{-1}\delta$, where ν and δ are defined in (5.18) 5
 6 and (5.21), respectively. 6

7
 8 Recall that the condition “ $\sigma_0 \in \mathcal{T}$ ”, used in the above theorem, holds if the set Ξ_0 is 8
 9 positively homogeneous, which in turn is implied by condition (C) (invariance under a 9
 10 constant scaling factor). We also have the following result about asymptotic robustness 10
 11 of the MDF estimators (Shapiro and Browne, 1987). 11

12
 13 THEOREM 6.2. Suppose that the model holds, the set Ξ_0 is approximated at the point 13
 14 $\sigma_0 \in \Xi_0$ by a linear space \mathcal{T} generated by the columns of the matrix $\Delta = \Delta(\theta_0)$, the 14
 15 MDF estimator $\hat{\theta}_n$ is a consistent estimator of θ_0 , the representation (6.6) holds with 15
 16 $\alpha > 0$ and that $\sigma_0 = \Delta\zeta$ for some $\zeta \in \mathbb{R}^q$. Then $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution 16
 17 to normal $N(\mathbf{0}, \Pi)$, the MDF estimator $\hat{\theta}_n$ is asymptotically efficient within the class of 17
 18 MDF estimators, and 18
 19

$$20 \quad \Pi = \alpha\Pi_N + \beta\zeta\zeta'. \quad (6.7) \quad 20$$

21
 22 The vector ζ can be obtained by solving the system of linear equations $\Delta\zeta = \sigma_0$, 22
 23 which is consistent if $\sigma_0 \in \mathcal{T}$. If the model is invariant under a constant scaling factor, 23
 24 i.e., the above condition (C) holds, then we can view $\theta^* = \theta^*(\theta, t)$ as a function of θ 24
 25 and t . If, moreover, $\theta^*(\theta, t)$ is differentiable in t , then by differentiating both sides of 25
 26 the equation $\sigma(\theta^*(\theta, t)) = t\sigma(\theta)$ we obtain vector ζ in the form 26
 27

$$28 \quad \zeta = \left. \frac{\partial\theta^*(\theta_0, t)}{\partial t} \right|_{t=1}. \quad (6.8) \quad 28$$

29
 30 For example, if the model is linear in θ , then $\theta^* = t\theta$ and hence $\zeta = \theta_0$. Of course, in 31
 32 practice the (unknown) population value θ_0 should be replaced by its estimator $\hat{\theta}$. 32

33
 34 **6.2. Linear latent variate models** 34

35
 36 Presentation of this section is based on Browne and Shapiro (1988). We assume here 36
 37 that the observed $p \times 1$ vector variate X can be written in the form 37

$$38 \quad X = \mu + \sum_{i=1}^s A_i z_i, \quad (6.9) \quad 38$$

39
 40 where μ is a $p \times 1$ mean vector, z_i is an (unobserved) $m_i \times 1$ vector variate and A_i is 42
 43 a (deterministic) $p \times m_i$ matrix of regression weights of X onto z_i , $i = 1, \dots, s$. We 43
 44 assume that random vectors z_i and z_j are independently distributed for all $i \neq j$ and 44
 45 have finite fourth-order moments. The above model implies the following structure of 45

1 the covariance matrix Σ of X : 1

$$2 \quad \Sigma = \sum_{i=1}^s A_i \Phi_i A_i', \quad (6.10) \quad 2$$

3 where Φ_i is the $m_i \times m_i$ covariance matrix of z_i , $i = 1, \dots, s$. 3

4 For example, consider the factor analysis model: 4

$$5 \quad X = \mu + \Lambda f + u, \quad (6.11) \quad 5$$

6 where Λ is a $p \times k$ matrix of factor loadings, $f = (f_1, \dots, f_k)'$ is a $k \times 1$ common vector 6
 7 variate, and $u = (u_1, \dots, u_p)'$ is a $p \times 1$ unique factor vector variate. It is assumed 7
 8 that random variables $f_1, \dots, f_k, u_1, \dots, u_p$, are mutually independently distributed, 8
 9 and hence random vectors f and u are independent. This implies that the covariance 9
 10 matrices Φ and Ψ , of f and u , respectively, are diagonal. Of course, we can write 10
 11 model (6.11) in the form 11
 12 12
 13 13
 14 14
 15 15

$$16 \quad X = \mu + \sum_{i=1}^k \Lambda_i f_i + \sum_{j=1}^p E_j u_j, \quad (6.12) \quad 16$$

17 where Λ_i is the i th column vector of Λ and E_j is the j th coordinate vector. This shows 17
 18 that this model is a particular case of model (6.9) with $s = k + p$ and $m_i = 1$, $i =$ 18
 19 $1, \dots, s$. Now model (6.11) (or, equivalently, model (6.12)) generates the following 19
 20 covariance structures model: 20
 21 21
 22 22
 23 23

$$24 \quad \Sigma = \Lambda \Phi \Lambda' + \Psi. \quad (6.13) \quad 24$$

25 The only difference between the above model (6.13) and the model (2.2) is that in (2.2) 25
 26 the covariance matrix Φ is assumed to be the identity matrix. 26

27 The linear model (6.9) implies the following structure of the asymptotic covariance 27
 28 matrix Γ (Browne and Shapiro, 1988, Theorem 2.1): 28
 29 29

$$30 \quad \Gamma = \Gamma_N + \sum_{i=1}^s (A_i \otimes A_i) C_i (A_i' \otimes A_i'), \quad (6.14) \quad 30$$

31 where C_i is the $m_i^2 \times m_i^2$ fourth-order cumulant matrix of z_i , $i = 1, \dots, s$. 31
 32 32
 33 33

34 Suppose now that the weight matrices have parametric structures $A_i = A_i(\mathbf{v})$, $i =$ 34
 35 $1, \dots, s$, where $\mathbf{v} \in \mathcal{Y}$ is a parameter vector varying in space $\mathcal{Y} \subset \mathbb{R}^\ell$. Then (6.10) 35
 36 becomes the following covariance structures model 36
 37 37

$$38 \quad \Sigma(\theta) = \sum_{i=1}^s A_i(\mathbf{v}) \Phi_i A_i(\mathbf{v})', \quad (6.15) \quad 38$$

39 with the parameter vector $\theta := (\mathbf{v}', \varphi_1', \dots, \varphi_s')'$, where $\varphi_i := \text{vec}(\Phi_i)$, $i = 1, \dots, s$. 39
 40 Note that the only restriction on the $m_i^2 \times 1$ parameter vectors φ_i imposed here is that 40
 41 the corresponding covariance matrix Φ_i should be positive semidefinite. Note also that 41
 42 if $m_i > 1$ for at least one i , then such choice of the parameter vector results in over- 42
 43 parameterization of model (6.15) since φ_i will have duplicated elements. It is possible 43
 44 44
 45 45

1 to include in the parameter vector θ only nonduplicated elements of matrices Φ_i . This, 1
 2 however, is not essential at the moment. 2

3 By applying the ‘vec’ operator to both sides of Eq. (6.15), we can write this model 3
 4 in the form 4

$$5 \quad \sigma(\theta) = \sum_{i=1}^s (A_i(\mathbf{v}) \otimes A_i(\mathbf{v}))\varphi_i. \quad (6.16) \quad 6$$

7
 8 It can be seen that the model is linear in parameters $\varphi_i, i = 1, \dots, s$, and 8
 9

$$10 \quad \frac{\partial \sigma(\theta)}{\partial \varphi_i} = A_i(\mathbf{v}) \otimes A_i(\mathbf{v}). \quad (6.17) \quad 10$$

11
 12 That is, the corresponding Jacobian matrix can be written as 12
 13

$$14 \quad \Delta(\theta) = [\Delta(\mathbf{v}), A_1(\mathbf{v}) \otimes A_1(\mathbf{v}), \dots, A_s(\mathbf{v}) \otimes A_s(\mathbf{v})]. \quad (6.18) \quad 14$$

15 Together with (6.14) this implies that Eq. (6.4) holds with matrix 15
 16

$$17 \quad D = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & C_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & C_2 & \dots & \mathbf{0} \\ & \dots & \dots & \dots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & C_s \end{bmatrix}. \quad (6.19) \quad 17$$

18
 19
 20
 21 We obtain the following result (Browne and Shapiro, 1988, Proposition 3.3). 22
 23

24 **THEOREM 6.3.** Consider the linear latent variate model (6.9) and the corresponding 24
 25 covariance structures model (6.15). Suppose that random vectors $\mathbf{z}_i, i = 1, \dots, s$, are 25
 26 mutually independently distributed, the assumption of a sequence of local alternatives 26
 27 (for the covariance structures model) holds, and Ξ_0 is a smooth manifold near the 27
 28 point σ^* . Then the MDF test statistic $n\hat{F}$ has asymptotically noncentral chi-squared 28
 29 distribution with $\nu = p(p + 1)/2 - \text{rank}(\Delta_0)$ degrees of freedom and the noncentrality 29
 30 parameter δ . 30
 31

32 In particular, the above theorem can be applied to the factor analysis model (6.11). 32
 33 Note that the MDF test statistics for the model (6.13), with the covariance term Φ , 33
 34 and model (2.2), without this term, are the same since Φ can be absorbed into Λ and 34
 35 hence the corresponding set Ξ_0 is the same. Note also that in order to derive the 35
 36 asymptotic chi-squaredness of the MDF test statistics we only used the corresponding 36
 37 independence condition, no other assumptions about distributions of \mathbf{f} and \mathbf{u} were made 37
 38 (except existence of fourth-order moments). For the factor analysis model this result was 38
 39 first obtained by Amemiya and Anderson (1990) by employing different techniques and 39
 40 without the assumption of finite fourth-order moments. 40

41 It is also possible to give corrections for the asymptotic covariance matrix Π of 41
 42 $n^{1/2}(\hat{\theta}_n - \theta_0)$ (compare with formula (6.5) of Proposition 6.2). In order to do that we 42
 43 need to verify identifiability of the parameter vector θ . Let us replace now φ_i with 43
 44 parameter vector $\varphi_i^* := \text{vecs}(\Phi_i), i = 1, \dots, s$, i.e., φ_i^* is $m_i(m_i + 1)/2 \times 1$ vector 44
 45 formed from the nonduplicated elements of Φ_i . Eq. (6.4) still holds with the matrix D 45

reduced to a smaller matrix \mathbf{D}^* by replacing each $m_i^2 \times m_i^2$ matrix \mathbf{C}_i by the corresponding $m_i(m_i + 1)/2 \times m_i(m_i + 1)/2$ matrix \mathbf{C}_i^* formed by the nonduplicated rows and columns of \mathbf{C}_i . We have then that, under the above assumptions,

$$\mathbf{\Pi} = \mathbf{\Pi}_N + \mathbf{D}^*. \quad (6.20)$$

It follows that the asymptotic covariance matrix of the MDF estimator $\hat{\mathbf{v}}$ is independent of the particular distribution of the z_i , $i = 1, \dots, s$, while the asymptotic covariance matrix of the MDF estimator $\hat{\boldsymbol{\varphi}}_i^*$ needs the correction term \mathbf{C}_i^* as compared with the normal case. Asymptotic covariances between $\hat{\mathbf{v}}$ and $\hat{\boldsymbol{\varphi}}_i^*$, and between $\hat{\boldsymbol{\varphi}}_i^*$ and $\hat{\boldsymbol{\varphi}}_j^*$, for $i \neq j$, are the same as in the normal case.

REMARK 5. Suppose, furthermore, that the population value \mathbf{v}_0 , of the parameter vector \mathbf{v} , lies on the *boundary* of the parameter space \mathcal{Y} , and that \mathcal{Y} is approximated at \mathbf{v}_0 by convex cone \mathcal{C} (recall that the parameter vectors $\hat{\boldsymbol{\varphi}}_i^*$, $i = 1, \dots, s$, are assumed to be unconstrained). Let $\tilde{\mathbf{v}}$ be the *unconstrained* MDF estimator of \mathbf{v}_0 (compare with the derivations of Section 5.4). Then, under the above assumptions, $n^{1/2}(\tilde{\mathbf{v}} - \mathbf{v}_0) \Rightarrow N(\mathbf{0}, \mathbf{U})$, where the asymptotic covariance matrix \mathbf{U} is independent of the particular distribution of the z_i , $i = 1, \dots, s$. We also have then that the MDF test statistic $n\hat{F}$ converges in distribution to the sum of two stochastically independent terms (compare with Eqs. (5.39) and (5.40)), one term having the usual chi-square distribution and the other term given by $\min_{\mathbf{v} \in \mathcal{C}} (\tilde{\mathbf{v}} - \mathbf{v})' \mathbf{U}^{-1} (\tilde{\mathbf{v}} - \mathbf{v})$. It follows that the asymptotic distribution of the MDF test statistic $n\hat{F}$ is chi-bar-squared and is independent of the particular distribution of the z_i , $i = 1, \dots, s$. That is, under these assumptions, distribution of the MDF test statistic is again asymptotically robust.

Acknowledgements

Author is supported in part by the National Science Foundation award DMS-0510324. The author is indebted to Michael Browne for careful reading of the manuscript and many helpful comments and suggestions.

References

- Amemiya, Y., Anderson, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics* **18**, 1453–1463.
- Anderson, T.W., Amemiya, Y. (1988). The Asymptotic normal distribution of estimators in Factor Analysis under general conditions. *Annals of Statistics* **16**, 759–771.
- Anderson, T.W., Rubin, H. (1956). Statistical inference in factor analysis. In: Neyman, J. (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5. University of California Press, Berkeley, CA, pp. 111–150.
- Bartholomew, D.J. (1961). Ordered tests in the analysis of variance. *Biometrika* **48**, 325–332.
- Bekker, P.A., ten Berge, J.M.F. (1997). Generic global identification in factor analysis. *Linear Algebra and its Applications* **264**, 255–263.
- Bonnans, J.F., Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York.

- 1 **Browne, M.W.** (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal* **8**, 1–24. 1
- 2 **Browne, M.W.** (1982). Covariance structures. In: Hawkins, D.M. (Ed.), *Topics in Applied Multivariate Analysis*. Cambridge University Press. 2
- 3 **Browne, M.W.** (1984). Asymptotically distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* **37**, 62–83. 3
- 4 **Browne, M.W.** (1987). Robustness in statistical inference in factor analysis and related models. *Biometrika* **74**, 375–384. 4
- 5 **Browne, M.W., Shapiro, A.** (1988). Robustness of normal theory methods in the analysis of linear latent 5
- 6 variate models. *British Journal of Mathematical and Statistical Psychology* **41**, 193–208. 6
- 7 **Chernoff, H.** (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573–578. 7
- 8 **Danskin, J.M.** (1967). *The Theory of Max-Min and its Applications to Weapon Allocation Problems. Econometrics and Operations Research*, vol. 5. Springer-Verlag, Berlin and New York. 8
- 9 **Fisher, F.M.** (1966). *The Identification Problem in Econometrics*. McGraw-Hill, New York. 9
- 10 **Jennrich, R.I.** (1987). Rotational equivalence of factor loading matrices with specified values. *Psychometrika* **43**, 421–426. 10
- 11 **Jöreskog, K.G.** (1977). Structural equation models in the social sciences: specification, estimation and testing. 11
- 12 In: Krishnaiah, P.R. (Ed.), *Applications of Statistics*. North-Holland, Amsterdam, pp. 265–287. 12
- 13 **Jöreskog, K.G.** (1981). Analysis of covariance structures (with discussion). *Scandinavian Journal of Statistics* **8**, 65–92. 13
- 14 **Kano, Y.** (1986). Conditions on consistency of estimators in covariance structure model. *Journal of the Japan* 14
- 15 *Statistical Society* **16**, 75–80. 15
- 16 **Kudô, A.** (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–418. 16
- 17 **Ledermann, W.** (1937). On the rank of reduced correlation matrices in multiple factor analysis. *Psychometrika* **2**, 85–93. 17
- 18 **Muirhead, R.J.** (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. 18
- 19 **Muirhead, R.J., Waternaux, C.M.** (1980). Asymptotic distribution in canonical correlation analysis and other 19
- 20 multivariate procedures for nonnormal populations. *Biometrika* **67**, 31–43. 20
- 21 **Nüesch, P.E.** (1966). On the problem of testing location in multivariate populations for restricted alternatives. 21
- 22 *Ann. Math. Stat.* **37**, 113–119. 22
- 23 **Rothenberg, T.J.** (1971). Identification in parametric models. *Econometrica* **39**, 577–591. 23
- 24 **Seber, G.A.F.** (1977). *Linear Regression Analysis*. Wiley, New York. 24
- 25 **Shapiro, A.** (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal* **17**, 33–81. 25
- 26 **Shapiro, A.** (1984). A note on consistency of estimators in the analysis of moment structures. *British Journal* 26
- 27 *of Mathematical and Statistical Psychology* **37**, 84–88. 27
- 28 **Shapiro, A.** (1985a). Identifiability of factor analysis: some results and open problems. *Linear Algebra and* 28
- 29 *its Applications* **70**, 1–7. 29
- 30 **Shapiro, A.** (1985b). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. 30
- 31 *South African Statistical Journal* **19**, 73–81. 31
- 32 **Shapiro, A.** (1985c). Second-order sensitivity analysis and asymptotic theory of parametrized nonlinear programs. *Mathematical Programming* **33**, 280–299. 32
- 33 **Shapiro, A.** (1985d). Asymptotic distribution of test statistics in the analysis of moment structures under 33
- 34 inequality constraints. *Biometrika* **72**, 133–144. 34
- 35 **Shapiro, A.** (1986). Asymptotic theory of overparametrized structural models. *Journal of the American* 35
- 36 *Statistical Association* **81**, 142–149. 36
- 37 **Shapiro, A.** (1987). Robustness properties of the MDF analysis of moment structures. *South African Statistical* 37
- 38 *Journal* **21**, 39–62. 38
- 39 **Shapiro, A.** (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Inter-* 39
- 40 *national Statistical Review* **56**, 49–62. 40
- 41 **Shapiro, A., Browne, M.W.** (1987). Analysis of covariance structures under elliptical distributions. *Journal of* 41
- 42 *the American Statistical Association* **82**, 1092–1097. 42
- 43 **Silvapulle, M.J., Sen, P.K.** (2005). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley, New York. 43
- 44 44
- 45 45

1	Steiger, J.H., Shapiro, A., Browne, M.W. (1985). On the multivariate asymptotic distribution of sequential	1
2	chi-square statistics. <i>Psychometrika</i> 50 , 253–264.	2
3	Tyler, E.D. (1982). Radial estimates and the test for sphericity. <i>Biometrika</i> 69 , 429–436.	3
4	Tyler, E.D. (1983). Robustness and efficiency properties of scatter matrices. <i>Biometrika</i> 70 , 411–420.	4
5	Wald, A. (1950). Note on the identification of economic relations. In: Koopmans, T.C. (Ed.), <i>Statistical Infer-</i>	5
6	<i>ence in Dynamic Economic Models</i> . John Wiley, New York.	6
7		7
8		8
9		9
10		10
11		11
12		12
13		13
14		14
15		15
16		16
17		17
18		18
19		19
20		20
21		21
22		22
23		23
24		24
25		25
26		26
27		27
28		28
29		29
30		30
31		31
32		32
33		33
34		34
35		35
36		36
37		37
38		38
39		39
40		40
41		41
42		42
43		43
44		44
45		45