PSYCHOMETRIKA—VOL. 54, NO. 1, 131–151
MARCH 1989

ALTERNATIVE TEST CRITERIA IN COVARIANCE STRUCTURE ANALYSIS: A UNIFIED APPROACH

ALBERT SATORRA

UNIVERSITY OF BARCELONA

In the context of covariance structure analysis, a unified approach to the asymptotic theory of alternative test criteria for testing parametric restrictions is provided. The discussion develops within a general framework that distinguishes whether or not the fitting function is asymptotically optimal, and allows the null and alternative hypothesis to be only approximations of the true model. Also, the equivalent of the information matrix, and the asymptotic covariance matrix of the vector of summary statistics, are allowed to be singular. When the fitting function is not asymptotically optimal, test statistics which have asymptotically a chi-square distribution are developed as a natural generalization of more classical ones. Issues relevant for power analysis, and the asymptotic theory of a testing related statistic, are also investigated.

Key words: covariance structure analysis, test statistics, asymptotic equivalence, noncentral chi-square, power of the test.

1. Introduction

In covariance structure analysis we often need to assess the validity of the restrictions that a model specification H_0, say, imposes on a more general (less restricted) model, say H (see, e.g., Bentler, 1983*b*; Browne, 1982; Jöreskog, 1981; Saris & Stronkhorst, 1984, and references contained therein). To deal with this problem, hypothesis testing is a standard approach in which the specifications H_0 and H are taken respectively as the null and the alternative hypothesis of a test. The minimization of a real-valued function $F = F(\mathbf{s}, \boldsymbol{\sigma})$, measuring the discrepancy between a vector $\mathbf{s}$ of summary statistics and a vector $\boldsymbol{\sigma}$ of parameters satisfying a specific model, generates the corresponding test statistics.

Three types of alternative test criteria have already been proposed in covariance structure analysis. The first test is the difference (D) type of test statistic, which is based on the difference between the minima of the fitting function F obtained when analyzing respectively H_0 and H (e.g., the "chi-square difference" test statistic, used by Jöreskog, 1977, p. 273, or Lee & Bentler, 1980, p. 132). This test is by far the most frequently applied statistic used to compare nested models. The second test is the score (S) type of test statistic, which is based on a vector of derivatives ("scores") and requires only that one minimizes the fitting function F under H_0 (e.g., the "Modification Index" of LISREL, Jöreskog & Sörbom, 1984, p. 1.42; Sörbom, 1986; and the "Lagrange multiplier" of EQS, Bentler, 1986). The third test is the Wald (W) type test statistic, which only needs to minimize F under H (Bentler, 1986; Lee, 1985). This statistic was the latest to be introduced to structural modeling, although, as will be

This research has been supported by the U.S.-Spanish Joint Committee for Cultural and Educational Cooperation, grant number V-B.854020. The author wishes to express his gratitude to P. M. Bentler who provided very helpful suggestions and research facilities—with an stimulating working environment—at the University of California, Los Angeles, where this work was undertaken. Thanks are also due to W. E. Saris who provided very valuable comments to earlier versions of this paper. Finally, it has also to be acknowledged the editor's and reviewers suggestions which led to substantial improvements of this article.

Request for reprints should be sent to Albert Satorra, Department of Statistics and Econometrics, Faculty of Economics, University of Barcelona, Avinguda Diagonal 690, Barcelona-08034, SPAIN.

0033-3123/89/$00.75/0

noted, it is a classical one. When F corresponds to a maximum likelihood estimation, the D, S and W type of test statistics are, respectively, the (log) likelihood ratio (Wilks, 1938), score (Rao, 1947) or Lagrange multiplier (Silvey, 1959), and Wald (1943) test statistics.

In the context of maximum likelihood estimation, the asymptotic equuivalence between the three types of test statistics is well known from Silvey (1959); for a very interesting review of this asymptotic equivalence, see Buse (1982). A unified development of these three principles within the context of maximum likelihood and in relation with econometrics is given in Engle (1984). In the context of covariance structure analysis, the equivalence between D and W is proven by Lee (1985). Also, the multivariate distribution of a set of D statistics arising from a sequence of nested structural models has been studied in Steiger, Shapiro, and Browne (1985). Bentler and Chou (1986) propose the use of S and W statistics for model modification in covariance structure analysis. The asymptotic equivalence of these alternative test criteria in the context of generalized least squares estimation is also noted in Bentler and Dijkstra (1985, p. 18).

A standard assumption of F being *asymptotically optimal* (in the sense of leading to efficient estimators and asymptotic chi-square statistics) is used in all the references above. This assumption restricts the possible discrepancy functions to be used in a specific application, because a Hessian matrix of F needs to relate in a specific form with the asymptotic covariance matrix of the random vector s of summary statistics (see Assumption 6, below). For instance, a Wishart maximum likelihood discrepancy function (i.e., normal theory), say F_{ML}, would not be asymptotically optimal when the distribution of the data is not normal. Very often in applications F will not satisfy this optimality condition. Additionally, the above references assume that the asymptotic covariance matrix of s, and the equivalent of the information matrix of the model, are nonsingular. The asymptotic theory of estimators and chi-square goodness of fit statistics in those cases where the asymptotic covariance matrix of s may be singular and/or the model is overparameterized is investigated in Shapiro (1986). However, Shapiro gives no specific attention to the D, S, and W types of test statistics. (In recent work, Shapiro, 1987, investigates also conditions for the asymptotic robustness of goodness of fit and D type of chi-square statistics.)

In the practice of covariance structure analysis, power considerations are needed for interpreting the values of test statistics (see Saris, Satorra & Sörbom, 1987; Saris & Stronkhorst, 1984, chap. 11; Satorra & Saris, 1985). The (nonnull) distribution of the above mentioned test statistics when the model H_0 (or H) does not hold exactly, and the expressions for approximating the corresponding noncentrality parameter (ncp), require also specific consideration.

The purpose of the present paper is two-fold. First, to provide a self contained review of asymptotic theory regarding the above referenced test criteria. The review develops in a general framework where the models H and H_0 may only be approximately true, the information matrix and (or) the asymptotic covariance matrix of s may be singular, and the standard assumption of F being asymptotically optimal is relaxed somewhat. Although part of this review overlaps with existing theory, the general framework taken may provide novelty. Also, this review sets the groundwork for developing the second purpose of the article, which is to investigate more general statistics to deal with the case where the discrepancy function is not asymptotically optimal. Asymptotic chi-square statistics, which may be adequate for instance when F is normal based and the data are not normal, are derived as natural generalizations of standard test statistics. Additionally, asymptotic theory is provided for a testing related statistic proposed recently in Saris et al. (1987) for assessing the substantive significance of the hypothesized restrictions.

Section 2 starts with the background and notation while section 3 presents the assumptions and needed preliminary results. The main results are given in sections 4 and 5, with section 4 developing the more standard results corresponding to F being asymptotically optimal, and section 5 giving new statistics adequate in case of more general discrepancy functions. Finally, section 6 develops the corresponding discussion.

2. Background and Notation

Let s denote a p-vector of sample statistics that converges in probability (when sample size $n \to \infty$) to the "true" p-vector, say $\boldsymbol{\sigma}$, of population parameters. For example, in covariance structure analysis, typically, s and $\boldsymbol{\sigma}$ contain the nonredundant elements of the sample and population covariance matrices, respectively. Also consider the two nested models for $\boldsymbol{\sigma}$, H: $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, and H_0: $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta_0$, where $\boldsymbol{\theta}$ is a q-dimensional parameter vector, and Θ and Θ_0 are subsets of R^q such that $\Theta_0 \subset \Theta$. We assume that $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is a twice continuously differentiable vector-valued function of $\boldsymbol{\theta}$, and that Θ_0 is the kernel of a continuously differentiable r-dimensional vector-valued function of $\boldsymbol{\theta}$, say $\mathbf{a} = \mathbf{a}(\boldsymbol{\theta})$ (i.e., $\Theta_0 = \{\boldsymbol{\theta} \in \Theta / \mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}\}$). For instance, when using LISREL (Jöreskog & Sörbom, 1984), the equality $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ may "fix" some of the elements of $\boldsymbol{\theta}$ at specific values, and (or) may impose an identity among two or more elements of $\boldsymbol{\theta}$. Also, a real-valued function $F = F(\mathrm{s}, \boldsymbol{\sigma})$ of s and $\boldsymbol{\sigma}$ is assumed to be specified for fitting alternative models. That is, the estimate $\tilde{\boldsymbol{\theta}}$ (or $\hat{\boldsymbol{\theta}}$) of $\boldsymbol{\theta}$ under the model H (or H_0) is defined as the value of $\boldsymbol{\theta}$ for which $F(\boldsymbol{\theta}) = F(\mathrm{s}, \boldsymbol{\sigma}(\boldsymbol{\theta}))$ attains its minimum over Θ (or Θ_0). $F = F(\mathrm{s}, \boldsymbol{\sigma})$ will be called a *discrepancy function* (Browne, 1982, 1984; Shapiro 1983, 1986).

We shall make use of the p-vectors $\tilde{\boldsymbol{\sigma}} \equiv \boldsymbol{\sigma}(\tilde{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\sigma}} \equiv \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})$; the values $\tilde{F} \equiv F(\mathrm{s}, \tilde{\boldsymbol{\sigma}})$ and $\hat{F} \equiv F(\mathrm{s}, \hat{\boldsymbol{\sigma}})$; the (1/2 gradient) q-dimensional vector-valued function $\mathbf{d}(\mathrm{s}, \boldsymbol{\theta}) \equiv (1/2)\partial F(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ of s and $\boldsymbol{\theta}$; the $q \times q$ matrix-valued function $J(\mathrm{s}, \boldsymbol{\theta}) \equiv (1/2)\partial^2 F(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ of s and $\boldsymbol{\theta}$ and the $p \times q$ and $r \times q$, respectively, matrix-valued functions $\Delta(\boldsymbol{\theta}) \equiv \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $A(\boldsymbol{\theta}) \equiv \partial\mathbf{a}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ of $\boldsymbol{\theta}$.

The following statistics are a main focus of this study:

$$D \equiv n(\hat{F} - \tilde{F}), \tag{1}$$

$$S \equiv n\hat{\mathbf{d}}'(\hat{J})^{-1}\hat{\mathbf{d}}, \tag{2}$$

$$W \equiv n\tilde{\mathbf{a}}'[\tilde{A}(\tilde{J})^{-1}\tilde{A}']^{-1}\tilde{\mathbf{a}}, \tag{3}$$

and

$$\hat{\mathbf{z}} \equiv -\hat{A}(\hat{J})^{-1}\hat{\mathbf{d}}, \tag{4}$$

where $\hat{\mathbf{d}} \equiv \mathbf{d}(\mathrm{s}, \hat{\boldsymbol{\theta}})$, $\hat{J} \equiv J(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\theta}})$, $\tilde{J} \equiv J(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\theta}})$, $\tilde{\mathbf{a}} \equiv \mathbf{a}(\tilde{\boldsymbol{\theta}})$ and $\tilde{A} \equiv A(\tilde{\boldsymbol{\theta}})$. Often, T will be used to denote any one of the test statistics (1), (2) and (3).

It shall be noted that expression (2) can be replaced by an equivalent one in which a vector of Lagrange multipliers $\hat{\mathbf{l}}$, instead of the "score" vector $\hat{\mathbf{d}}$, is involved. Consider the Lagrangian function $L \equiv (1/2)F - \mathbf{l}'\mathbf{a}$, where $\mathbf{l}$ is a r-vector of Lagrange multipliers. The first order conditions for the constrained estimator $\hat{\boldsymbol{\theta}}$ yield $\hat{\mathbf{d}} = \hat{A}'\hat{\mathbf{l}}$, where $\hat{\mathbf{l}}$ is the Lagrange multiplier vector associated with $\hat{\boldsymbol{\theta}}$. Substituting $\hat{\mathbf{d}}$ in (2), produces

$$S = n\hat{\mathbf{l}}'\hat{A}(\hat{J})^{-1}\hat{A}'\hat{\mathbf{l}}, \tag{5}$$

which is the alternative form of S, introduced in Silvey (1959), that inspires the alternative notation for S as a "Lagrange multiplier" statistic.

Note that when F is $(-2/n)$ times the log of the likelihood function, then D, S and W are, respectively, the typical (log) likelihood ratio, score and Wald test statistics (see, e.g., Cox & Hinkley, 1974, chap. 9). In covariance structure analysis, D is known as the "chi-square difference" test statistic (Jöreskog, 1970, 1981). Note that when $\mathbf{a}(\boldsymbol{\theta}) \equiv \theta_j$, a component of the parameter vector $\boldsymbol{\theta}$, then the statistic W in (3) can easily recognized to be the square of the parameter estimate divided by the standard error, that is, the square of the typical "t-value" used to test "whether the true parameter is zero" (Jöreskog & Sörbom, 1984, p. III.12); thus, the typical "t-test" is equivalent (i.e., gives exactly the same answer) to a Wald test.

In the context of statistics used for assessing the validity of the restrictions $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$, the statistic $\hat{\mathbf{z}}$ is also worth considering (see Saris et al., 1987, for the motivation and theory of such statistic in its univariate version). The statistics "estimated change" and "parameter change" implemented in the newest versions of the computer programs LISREL (Jöreskog & Sörbom, 1984) and EQS (Bentler, 1986), respectively, are examples of the statistic $\hat{\mathbf{z}}$. It will be shown below that $\hat{\mathbf{z}}$ is asymptotically equal to $\mathbf{a}(\tilde{\boldsymbol{\theta}})$, being then the "change" in $\mathbf{a}(\boldsymbol{\theta})$ as a result of dropping the restriction $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$. For instance, in the case where $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ is just restricting an element of $\boldsymbol{\theta}$ at the value zero, $\hat{\mathbf{z}}$ equals (asymptotically) the estimated value of such an element when it is treated as a free parameter. Notice that for obtaining $\hat{\mathbf{z}}$, only the restricted model H_0 needs to be analyzed.

When H specifies that $\boldsymbol{\sigma}$ is unrestricted, the corresponding statistic D is the typical chi-square goodness of fit statistic used for assessing the validity of the model H_0. The statistics $n\tilde{F}$ and $n\hat{F}$ are also chi-square goodness of fit statistics for testing the specifications H and H_0, respectively (e.g., Browne, 1982, p. 97).

This paper will review test statistics that have asymptotically a central (noncentral) chi-square distribution when the null hypothesis holds exactly (approximately). Typically, the rejection region of the (nominal)α-level test is $[T \geqslant c_\alpha]$, where T denotes the "chi-square" test statistic and c_α is a ("critical") value such that $\Pr\{\chi^2_r \geqslant c_\alpha\} = \alpha$, where r is the associated degrees of freedom (df) of the test statistics and χ^2_r denotes a central chi-square distribution with r df. The noncentral chi-square distribution with noncentrality parameter (ncp) λ, and df r, will be denoted as $\chi^2_r(\lambda)$. The power of the test associated to a specific true vector $\boldsymbol{\sigma}$ will be approximated by $\Pr\{\chi^2_r(\lambda) \geqslant c_\alpha\}$, where the ncp λ will depend on $\boldsymbol{\sigma}$.

3. Assumptions and Preliminary Results

The following assumptions will be referred to throughout.

Assumption 1. F is such that

(i) $F(\mathbf{s}, \boldsymbol{\sigma}) \geqslant 0$ for all $\mathbf{s}$ and $\boldsymbol{\sigma}$;
(ii) $F(\mathbf{s}, \boldsymbol{\sigma}) = 0$ if $\mathbf{s} = \boldsymbol{\sigma}$;
(iii) F is twice continuously differentiable in $\mathbf{s}$ and $\boldsymbol{\sigma}$.

This is a typical condition satisfied for all the discrepancy functions currently in use (Browne, 1982, p. 81; Shapiro, 1986). An example is the following generalized least squares (GLS) discrepancy function

$$F(\mathbf{s}, \boldsymbol{\sigma}) = (\mathbf{s} - \boldsymbol{\sigma})' Y (\mathbf{s} - \boldsymbol{\sigma}), \tag{6}$$

where Y is a $p \times p$ nonnegative weight matrix (Browne, 1974). Note that unlike Browne (1982), the condition that $F(\mathbf{s}, \boldsymbol{\sigma}) = 0$ only if $\mathbf{s} = \boldsymbol{\sigma}$ is not used.

Remark 3.1. Often the weight matrix Y is stochastic, being also a function of the data and possibly depending on other statistics than s (e.g., when using the GLS asymptotically distribution free methods, Y will be a function also of the fourth order moments of the data). When that occurs, we need to strengthen Assumption 1 with the condition that, as sample size $n \to \infty$, Y is of constant rank and converges in probability to a nonnegative definite matrix. (This point will be pursued in Remark 3.2.)

Assumption 2. There exists a vector $\boldsymbol{\sigma}_0$ such that $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(\boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is an inner point of Θ satisfying $\mathbf{a}(\boldsymbol{\theta}_0) = \mathbf{0}$ and $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0) \xrightarrow{L} N(\boldsymbol{\mu}, \Gamma)$, where "$\xrightarrow{L}$" indicates convergence in distribution, $\boldsymbol{\mu}$ is a (finite) p-vector and Γ is a nonnegative definite $p \times p$ matrix.

Usually a central limit theorem can be invoked to ensure the asymptotic normality of s under very mild regularity conditions of the data (see, e.g., van Praag, Dijkstra & Van Velzen, 1985). The condition of $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$ to have asymptotic mean equal to $\boldsymbol{\mu}$, possibly a nonnull vector, amounts to saying that the true $\boldsymbol{\sigma}$, say $\boldsymbol{\sigma}^0$, changes with n in the following form:

$$\boldsymbol{\sigma}^0 \equiv \boldsymbol{\sigma}_n^0 = \boldsymbol{\sigma}_0 + n^{-1/2}\boldsymbol{\mu};$$

that is, the deviation $n^{-1/2}\boldsymbol{\mu}$, between the true $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}_0$, decreases with n at the rate $n^{-1/2}$. This assumption of a sequence of "local alternatives" is a technical device that prevents the (nonnull) distributions of the test statistics to be degenerate at the limit, while it allows for certain degree of ("structural") *misspecification* of the model. Intuitively, one has to view $\boldsymbol{\mu} \neq 0$ as allowing for H_0 (or H) to be only "approximately" true. However, it must be stressed that when $\boldsymbol{\sigma}^0$ highly deviates from H_0, then this assumption may be unrealistic and the terms that are ignored in the asymptotic expansions used below may become nonnegligible; that is, the asymptotic theory developed below does not apply when H_0 is grossly misspecified. For a similar assumption, see Browne (1984), Shapiro (1983), Satorra and Saris (1985), Steiger et al. (1985), and Bentler and Dijkstra (1985).

An interesting case to consider is when H is true, but H_0 is not. This can be formalized by saying that $\boldsymbol{\sigma}^0 = \boldsymbol{\sigma}(\boldsymbol{\theta}_0 + \boldsymbol{\delta} n^{-1/2})$, where $\boldsymbol{\delta}$ is a nonnull q-vector. Obviously, this implies that:

$$\boldsymbol{\mu} = \Delta(\boldsymbol{\theta}_0)\boldsymbol{\delta}. \tag{7}$$

The case where the form (7) for $\boldsymbol{\mu}$ holds will be frequently referred to in the present paper.

The following is an identification condition for the model H_0.

Assumption 3. Θ_0 is a compact subset of R^q and $\boldsymbol{\theta}_0$ is the unique minimizer of $F(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}(\cdot))$ in Θ_0.

In fact, the compactness assumption of Θ_0 could be substituted by the less stringent "level condition" of Shapiro (1983, Definition 2.2).

Clearly a nonrestrictive assumption, ensuring that the constraints are not linearly dependent, is the following.

Assumption 4. The $r \times q$ matrix $A \equiv \partial \mathbf{a}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ has full row rank.

Note that by Assumption 1 the following Hessian matrix

$$\left.\frac{\partial^2 F(\mathbf{s}, \boldsymbol{\sigma})}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'}\right|_{(\mathbf{s}, \boldsymbol{\sigma})=(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_0)} = 2V, \tag{8}$$

say, is nonnegative definite; thus, the matrix

$$2J = \left.\frac{\partial^2 F(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 2\Delta' V \Delta$$

where $J \equiv J(\boldsymbol{\sigma}_0, \boldsymbol{\theta}_0)$ and $\Delta \equiv \Delta(\boldsymbol{\theta}_0)$, is also a nonnegative definite matrix. (Note that Assumptions 1 and 2 imply that $\partial F(\boldsymbol{\sigma}_0, \boldsymbol{\sigma})/\partial \boldsymbol{\sigma}|_{\boldsymbol{\sigma}=\boldsymbol{\sigma}_0}$ is zero, which yields the above expression $2\Delta' V \Delta$ for the matrix of second derivatives of F.)

Remark 3.2. It has to be noted that the matrix $\partial^2 F(\mathbf{s}, \boldsymbol{\sigma})/\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'|_{(\mathbf{s}, \boldsymbol{\sigma})=(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_0)}$ may be a stochastic matrix, as for instance when F is a GLS discrepancy function with a stochastic weight matrix Y (see (6)); when this is the case, V will be taken to be the corresponding probability limit of Y, which is assumed to be a non-negative matrix (recall Remark 3.1).

An assumption that is needed to ensure that $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ implies, at least, a set of "just-identifying" restrictions on $\boldsymbol{\theta}$ (usually H_0 will overidentify $\boldsymbol{\theta}$) is the following.

Assumption 5. There exists a partition $[A_1', A_2']$ of A' such that: A_1 is a $t \times q$ matrix, $t = q$–rank (J) and $[J, A_1']$ has rank q.

Note that in case of maximum likelihood estimation, the matrix J is $(1/n)$ times the ("Hessian" form of the) Fisher information matrix (see, e.g., Satorra & Saris 1985, p. 85), and the condition that J be nonsingular is a standard one. However, exactly as it happens in restricted maximum likelihood estimation (Silvey 1959), an adjustment is needed when J is singular but Assumption 5 holds: one substitutes J^{-1} by $(J + A_1'A_1)^{-1}$, and rank (A) by rank (A_2), whenever the inverse of J is required and rank (A) is used. This adjustment will be assumed to hold throughout the present article, even when no explicit discussion of this point is made.

The following assumption relates the (1/2) Hessian matrix V, defined in (8), with the asymptotic covariance matrix Γ of s.

Assumption 6. $V\Gamma V = V$.

Note that when V is nonsingular, Assumption 6 implies that $\Gamma = V^{-1}$, a condition which has been highly exploited to derive optimal asymptotic properties of estimators and test statistics (Browne, 1984; Shapiro, 1983; Bentler & Dijkstra, 1985).

It will be seen below that the use of a discrepancy function F for which Assumption 6 is satisfied guarantees the asymptotic optimality of the statistics. We will follow the convention of calling a F for which Assumption 6 is verified an *asymptotically optimal* (AO) discrepancy function. In contrast with Assumptions 1 to 5, which are not very restrictive in practice. Assumption 6 is likely to be violated in applications. Results of this paper address discrepancy functions which are not AO.

It will be seen below that the standard theory for the statistics D, W and S will hold under the following somewhat less restrictive version of Assumption 6.

*Assumption 6**. $\Delta' V \Gamma V \Delta = \Delta' V \Delta$.

The following lemma provides some symptotic equalities that will be needed when proving the theorems of the sections below (for the proof of this lemma, see Appendix A).

Lemma 1: Assumptions 1 to 5 guarantee that:

(i) $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{=} (J \cdot A)\Delta' V n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$,

(ii) $n^{1/2}\tilde{\mathbf{a}} \overset{a}{=} A(J \cdot A_1)\Delta' V n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$,

(iii) $n^{1/2}\tilde{\mathbf{d}} \overset{a}{=} -JJ^{-1}A_2'(A_2(J \cdot A_1)A_2')^{-1}A_2(J \cdot A_1)\Delta' V n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$,

(iv) $n\hat{F} \overset{a}{=} n(\mathbf{s} - \boldsymbol{\sigma}_0)'(V - V\Delta(J \cdot A)\Delta' V)(\mathbf{s} - \boldsymbol{\sigma}_0)$,

(v) $n^{1/2}\hat{\mathbf{z}} \overset{a}{=} A(J \cdot A_1)\Delta' V n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$,

(vi) $n^{1/2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0) \overset{a}{=} \Delta(J \cdot A)\Delta' V n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$

where $u_n \overset{a}{=} v_n$ means that $u_n - v_n \to 0$ as $n \to \infty$ (the convergence is in probability when one or both sides of the equality are stochastic quantities) and

$$(J \cdot A) \equiv J^{-1} - J^{-1}A'(AJ^{-1}A')^{-1}AJ^{-1}, \tag{9}$$

with $(J \cdot A_1)$ of the form (9) with A replaced by A_1. Note that, as pointed in the discussion of Assumption 5, whenever J is singular, J^{-1} is defined as $(J + A_1'A_1)^{-1}$. (Note that in the right side of (iii) JJ^{-1} is not simplified due to the possibility of J being singular.) Obviously, when J is nonsingular, $A_2 \equiv A$ and $(J \cdot A_1) \equiv J^{-1}$.

Remark 3.3. When J is singular the model H is said to be overparameterized (Shapiro, 1986) and $\tilde{\boldsymbol{\theta}}$ is not uniquely defined. In the Lemma 1 above, $\tilde{\boldsymbol{\theta}}$ is chosen to be the estimator (uniquely) determined when imposing to H the ("just identifying") restrictions $\mathbf{a}_1(\boldsymbol{\theta}) = \mathbf{0}$. Hence, the Results (i) and (iv) of Lemma 1 will hold also when changing $\hat{\boldsymbol{\theta}}$, $\hat{F}$ and A to $\tilde{\boldsymbol{\theta}}$, $\tilde{F}$ and A_1 respectively.

The next section gives results corresponding to the most typical case of F satisfying Assumptions 1 to 6.

4. Results for Asymptotically Optimal Discrepancy Functions

The following theorem states the asymptotic equivalence of D, S and W and gives their common asymptotic distribution. Also, whenever a true covariance $\boldsymbol{\sigma}^0$ is specified, explicit expressions for approximating the corresponding ncp are provided. For similar and closely related results, see Silvey (1959), Satorra and Saris (1985), Lee (1985), Steiger et al. (1985) and Bentler and Dijkstra (1985).

Theorem 4.1. Let Assumptions 1 to 5 hold, then

(a) $D \overset{a}{=} S \overset{a}{=} W$.

(b) When additionally Assumption 6* holds,

$$T \xrightarrow{L} \chi_r^2(\lambda),$$

where T is any one of the statistics D, S and W, r is rank (A_2) and

$$\lambda = \mu' V\Delta(J \cdot A_1)A_2'(A_2(J \cdot A_1)A_2')^{-1}A_2(J \cdot A_1)\Delta' V\mu. \tag{10}$$

(c) $\lambda \overset{a}{=} T^0$, where T^0 is any of the (nonstochastic) values of D, S or W obtained when σ^0 substitutes s.

Proof. By subtracting the asymptotic expressions of $n\hat{F}$ and $n\tilde{F}$ obtained from (iv) of Lemma 1 (see Remark 3.3), and using Lemma 2 of the appendix, it is obtained that

$$D \overset{a}{=} n(\mathbf{s} - \sigma_0)' V\Delta(J \cdot A_1)A_2'(A_2(J \cdot A_1)A_2')^{-1}A_2(J \cdot A_1)\Delta' V(\mathbf{s} - \sigma_0). \tag{11}$$

Then, substituting in (2) and (3) the terms $n^{1/2}\tilde{\mathbf{a}}$ and $n^{1/2}\hat{\mathbf{d}}$ by their equivalent expressions given in (ii) and (iii) of Lemma 1, it follows that S and W equal the right hand side of (11); thus, proving (a) of the theorem. To prove (b), note first that Assumption 2, combined with standard asymptotic arguments, implies that the quadratic form of (11) is asymptotically distributed as $\mathbf{z}'U\mathbf{z}$, where $\mathbf{z} \sim N(\mu, \Gamma)$ and

$$U = V\Delta(J \cdot A_1)A_2'(A_2(J \cdot A_1)A_2')^{-1}A_2(J \cdot A_1)\Delta' V.$$

Now, straightforward algebra shows that the sufficient condition $U\Gamma U = U$ for $\mathbf{z}'U\mathbf{z}$ to be (noncentral) chi-square distributed, with df and ncp equal to trace $(U\Gamma)$ and $\mu' U \mu$ respectively (see, e.g., Theorem 9.2.1 of Rao & Mitra, 1971), holds when $\Delta' V\Gamma V\Delta = \Delta' V\Delta$ (i.e., Assumption 6*). Finally, (c) follows by noting that the arguments used in (a) apply also when σ^0 substitutes s, with $\overset{a}{=}$ denoting now that the difference between the left and right side of the equality converges to 0 as $n \to \infty$ (here convergence in probability is not needed as one deals with nonstochastic quantities). □

Remark 4.1. Assumption 6* was needed only when proving that $U\Gamma U = U$; without that assumption, and using also standard results of distribution of quadratic forms in normal variables, it can still be guaranteed that T is asymptotically distributed as a weighted sum of independent chi-square distributions with 1 df, with the weights being determined by the eigenvalues of the matrix $U\Gamma$ (see Satorra & Bentler, 1988b, where this point is exploited and a specific scaling correction based on the trace of $U\Gamma$ is proposed). Note that (a) of Theorem 4.1 guarantees that this asymptotic distribution is the same for the three statistics D, S and W.

When H holds then $\mu = \Delta\delta$ (see (7)), and if in addition J is nonsingular, then the ncp λ of (10) reduces to

$$\lambda = \delta' A'(AJ^{-1}A')^{-1}A\delta. \tag{12}$$

Using this expression, the power of the test against different δ characterizing different "directions" of misspecification of H_0 within H can be investigated.

For the completeness of this paper, the asymptotic distribution of $n\hat{F}$ will be given (Bentler & Dijkstra, 1985; Browne, 1982, 1984; Satorra & Bentler, 1988b; Satorra & Saris, 1985; Shapiro, 1983, 1986). This asymptotic distribution and the relations between D, S, W and the goodness of fit statistics $n\tilde{F}$ and $n\hat{F}$ are reviewed in the next theorem (see also Steiger et al., 1985).

Theorem 4.2. Let Assumptions 1 to 6 hold, then

(a) $n\hat{F} \overset{L}{\to} \chi^2_{r^*}(\lambda^*)$, where $r^* = \text{rank}(V) - q + r$ and

$$\lambda^* = \mu'(V - V\Delta(J \cdot A)^{-1}\Delta' V)\mu. \tag{13}$$

(b) $nF^0 \overset{a}{=} \lambda^*$, where $F^0 = \text{Min}_{\theta \in \Theta_0} F(\sigma^0, \sigma(\theta))$.

(c) $n\hat{F}$ is asymptotically independent of T, where T denotes any of the statistics D, S or W.

(d) The asymptotic covariance between $n\hat{F}$ and T is equal to $2 \times \text{rank}(A_2) + 4\lambda$, where λ is given in (10).

Proof. The theorem follows easily when combining (iv) of Lemma 1, and (11), with standard theory of distribution of quadratic forms in normal variables. Here Assumption 6 is needed. (We recall that r is the number of components of $\mathbf{a}(\boldsymbol{\theta})$.) □

Consider next the case that H holds (i.e. $\boldsymbol{\mu} = \Delta\boldsymbol{\delta}$) and J is nonsingular. In that case, it can easily be seen that the expressions of λ and λ^* of (10) and (13), respectively, are identical; thus the following result applies.

Corollary 4.1. Let Assumptions 1 to 6 and the model H hold, with J being nonsingular, then

$$T \xrightarrow{L} \chi^2_{r^*}(\lambda),$$

where λ is given in (10), and r^* equals rank (A) whenever T is D, S or W, and equals $\text{rank}(V) - q + r$ when T is $n\hat{F}$.

An important observation is the following. When H holds the use of D, S or W instead of $n\hat{F}$, for testing H_0, leads to a decrease in the df of the corresponding chi-square distribution, while the ncp remains the same. Thus, whenever the model H is known a priori to be true, there is a clear increase of power of the test by using D, S or W instead of $n\hat{F}$.

Test statistics which are asymptotically equivalent to $T\,(D,\ W$ or $S)$ will be obtained when in (2) and (3) the matrices $\tilde{J}$ and $\hat{J}$, and the vectors $n^{1/2}\tilde{\mathbf{a}}$ and $n^{1/2}\hat{\mathbf{d}}$, are replaced by asymptotically equivalent statistics. For instance, a root-n consistent estimator $\hat{\boldsymbol{\theta}}_+$ of $\boldsymbol{\theta}_0$ (i.e., $n^{1/2}(\hat{\boldsymbol{\theta}}_+ - \boldsymbol{\theta}_0)$ is bounded in probability as $n \to \infty$) that satisfies H_0 produces the "linearized" estimator $\hat{\boldsymbol{\theta}}_L = \hat{\boldsymbol{\theta}}_+ + (J \cdot A)_+ \Delta'_+ V_+(\mathbf{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_+))$ (Bentler & Dijkstra, 1985), where the subindex "$_+$" indicates evaluation at $\hat{\boldsymbol{\theta}}_+$. It can be verified that $n^{1/2}\hat{\boldsymbol{\theta}}_L \overset{a}{=} n^{1/2}\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the minimum discrepancy function estimator as defined in section 2. Thus, if in (2) $\mathbf{d}_L = \mathbf{d}(\mathbf{s}, \hat{\boldsymbol{\theta}}_L)$ and J_+ substitute $\hat{\mathbf{d}}$ and $\hat{J}$ respectively, a "linearized" score test statistic, say S_L, that is asymptotically equivalent to S is obtained. In the same way, if $\tilde{\boldsymbol{\theta}}_+$ is a root-n consistent estimator of $\boldsymbol{\theta}$ in H, a "linearized" Wald statistic, say W_L, that has the same asymptotic distribution as W, is obtained by changing in (3) $\tilde{\mathbf{a}}$ and $\tilde{J}$ to $\mathbf{a}_L = \mathbf{a}(\tilde{\boldsymbol{\theta}}_L)$ and J_+, respectively, with $\tilde{\boldsymbol{\theta}}_L = \tilde{\boldsymbol{\theta}}_+ + (J \cdot A)_+^{-1}\Delta'_+ V_+(\mathbf{s} - \boldsymbol{\sigma}(\tilde{\boldsymbol{\theta}}_+))$ (now, the subindex "$_+$" indicates evaluation at $\tilde{\boldsymbol{\theta}}_+$). Linearized score and Wald statistics are available in EOS (Bentler, 1986).

Often in practice the researcher cast doubts just on the validity of a subset of the restrictions implied by $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$, having certainty on the validity of the others. Consider the case of a partition $\mathbf{a}' \equiv [\mathbf{b}_1', \mathbf{b}_2']$ of $\mathbf{a} \equiv \mathbf{a}(\boldsymbol{\theta})$ such that the matrix A of derivatives partitions accordingly as $A' \equiv [B_1', B_2']$ (when J is singular, also assume that A_1 of Assumption 5 is contained in B_1). Consider the model H_1, say, defined by H and the additional restrictions of $\mathbf{b}_1(\boldsymbol{\theta}) = \mathbf{0}$. Then, one may just be interested in the test of $\mathbf{b}_2(\boldsymbol{\theta}) = \mathbf{0}$ when the *adopted* model is H_1 (i.e., the model H with the restriction $\mathbf{b}_1(\boldsymbol{\theta}) = \mathbf{0}$); that is, the interest centers on the *restricted test* of H_0 against H_1 (see Aitchison, 1962, for the theory and motivation of the restricted test in the case of maximum likelihood). Denote by T_0 any one of the statistics D, S and W corresponding to the restricted test of H_0 against H_1; and denote by T_1 the test statistics (D_1, S_1 and W_1) when testing H_1 against H. As in all above, T still corresponds to the (unrestricted) test of H_0

against H. The following Corollary 4.2 shows that there are a variety of asymptotically equivalent test statistics for the test of the subset of restrictions $\mathbf{b}_2(\boldsymbol{\theta}) = \mathbf{0}$ (i.e., the test of H_0 against H_1).

Corollary 4.2. Let Assumptions 1 to 5 hold, then

(a) $T_0 \overset{a}{=} T - T_1$, where $T - T_1$ can be any of the nine statistics arising from the different choices of T and T_1 (T and T_1 are chosen among D, S and W, and D_1, S_1 and W_1 respectively).

(b) When additionally Assumption 6* holds, then

$$R \overset{L}{\to} \chi^2_{r^{**}}(\lambda^{**}),$$

where R denotes either T_0 or $T - T_1$, r^{**} is the rank of B_2 and

$$\begin{aligned}\lambda^{**} &= \underline{\boldsymbol{\mu}' V\Delta}(J \cdot B_1)B_2'(B_2(J \cdot B_1)B_2')^{-1}B_2\Delta' V\boldsymbol{\mu} \\ &= \boldsymbol{\mu}' V\Delta J^{-1}[A'(AJ^{-1}A')^{-1}A - B_1'(B_1J^{-1}B_1')^{-1}\underline{B_1]J^{-1}}\Delta' V\boldsymbol{\mu}. \end{aligned} \tag{14}$$

(Here $(J \cdot B_1)$ is defined as in (9) with A changed to B_1.)

(c) $R^0 \overset{a}{=} \lambda^{**}$, where R^0 is the nonstochastic value of R when $\boldsymbol{\sigma}^0$ substitutes $\mathbf{s}$.

Proof. Using (a) of Theorem 4.1, we get

$$T \overset{a}{=} D = n\hat{F} - n\tilde{F} = (n\hat{F} - n\hat{F}_1) + (n\hat{F}_1 - n\tilde{F}) = D_1 + \underline{D_0} \overset{a}{=} T_1 + T_0$$

(here $n\hat{F}_1$ is used to denote the chi-square goodness of fit statistic associated with H_1), which implies (a) and (c) of Corollary 4.2. The result (b) follows directly from (b) of Theorem 4.1, after an obvious (local) reparameterization is applied to H_1, and the matrix equality stated in Lemma 2 of the appendix (note that here the matrix equality of Lemma 2 is applied with A_1 and A_2 changed to B_1 and B_2, respectively). □

Note that attending to the specific kind of departure from H_0 (or H) we can simplify the expression of the ncp of (14). Effectively, if H holds then $\boldsymbol{\mu} = \Delta\boldsymbol{\delta}$, and λ^{**} reduces to

$$\lambda^{**} = \boldsymbol{\delta}'[A'(AJ^{-1}A')^{-1}A - B_1'(B_1J^{-1}B_1')^{-1}B_1]\boldsymbol{\delta}; \tag{15}$$

if in addition H_1 holds, then $B_1\boldsymbol{\delta} = \mathbf{0}$ and we get

$$\lambda^{**} \overset{a}{=} \boldsymbol{\delta}'[A'(AJ^{-1}A')^{-1}A]\boldsymbol{\delta},$$

which gives the same ncp as in (12), where the (nonrestricted) test of H_0 against H was considered. Therefore, when H_1 holds there is clearly an increase in the power of the test as a result of using a restricted test of H_0 against H_1, due to the fact that the ncp remains the same while the df's are reduced. Note, however, that when H_1 does not hold, and the test of $\mathbf{b}_2(\boldsymbol{\theta}) = \mathbf{0}$ is carried out by a restricted test of H_0 within H_1, the ncp λ^{**} may still be different from zero even though $\mathbf{b}_2(\boldsymbol{\theta}) = \mathbf{0}$ (as $\mathbf{b}_1(\boldsymbol{\theta})$ may not be zero and hence the right hand side of (15) does not vanish). See Satorra and Saris (1983) for some Monte Carlo evidence on this point.

When assessing model modification, the researcher may be faced with the choice between two different sets of restrictions, say $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$. Let T_a and T_c be any of the D, S or W test statistics corresponding to the restrictions $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$, respectively. (The function $\mathbf{c} = \mathbf{c}(\boldsymbol{\theta})$ is assumed to be continuously differentiable with the matrix $C \equiv \partial \mathbf{c}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ of full row rank; when J is singular, we also assume that $C' \equiv [A_1', C_2']$, where A_1 is the same matrix as in Assumption 5.) A typical example appears, for instance, when using LISREL (Jöreskog & Sörbom, 1984) and the modification indexes of two different fixed parameters are considered. One can also consider the statistic T_{ac}, say, that corresponds to the simultaneous test of both sets of restrictions $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$. By just applying standard results of independence of quadratic forms in normal variables (see e.g., Rao & Mitra, 1971) to the asymptotic expressions of T_a and T_c implied by (11), and using simple algebra, we get the following result.

Theorem 4.3. Let Assumptions 1 to 5 and Assumption 6* hold, and

$$A(J \cdot A_1)C' = 0, \tag{16}$$

then T_a and T_c are asymptotically independent and $T_{ac} \overset{a}{=} T_a + T_c$.

Remark 4.3. Note that under the assumptions of the theorem, T_a and T_c are also asymptotically chi-square distributed (see (b) of Theorem 4.1). It can easily be seen that just under Assumptions 1 to 5, the equality

$$A(J \cdot A_1)\Delta' V \Gamma V \Delta (J \cdot A_1) C' = 0 \tag{17}$$

guarantees the asymptotic independence of T_a and T_b whether they have a chi-square distribution or not; now, however, the asymptotic decomposition of T_{ac} as a sum of T_a and T_c will not hold in general.

In practice it may be interesting to have available test statistics for which an asymptotic chi-square distribution can be guaranteed without the discrepancy function F being asymptotically optimal. The next section presents such statistics.

5. Generalized Score and Wald Statistics

Generalized score and Wald test statistics, say GS and GW, which will be proven to be asymptotically chi-square statistics even when the discrepancy function F is not AO, are defined as follows:

$$GS \equiv n\hat{\mathbf{d}}'\hat{J}^{-1}\hat{A}'[(\hat{A}\hat{J}^{-1}\hat{\Delta}'V)\bar{\Gamma}(V\hat{\Delta}\hat{J}^{-1}\hat{A}')]^{-1}\hat{A}\hat{J}^{-1}\hat{\mathbf{d}}, \tag{18}$$

where $\bar{\Gamma}$ denotes a consistent estimate of Γ; and

$$GW = n\tilde{\mathbf{a}}'[(\tilde{A}\tilde{J}^{-1}\tilde{\Delta}'V)\bar{\Gamma}(V\tilde{\Delta}\tilde{J}^{-1}\tilde{A}')]^{-1}\tilde{\mathbf{a}}. \tag{19}$$

For the sake of simplicity, in (18) and (19) it is assumed that J, V and Γ are nonsingular matrices. (If this condition were not met, $AJ^{-1}\Delta' V$ would be changed to $(A(J \cdot A_1)\Delta' V)$ and the condition of $[A_2(J \cdot A_1)\Delta' V \Gamma V \Delta (J \cdot A_1) A_2']$ to be nonsingular would be added.)

The following theorem gives the asymptotic distribution of such test statistics.

Theorem 5.1. Let Assumptions 1 to 5 hold, with J, V and Γ being nonsingular matrices, then

(a) $GS \overset{a}{=} GW$,
(b) $GT \overset{L}{\to} \chi_r^2(\lambda^{***})$ with $r = \text{rank}(A)$,

$$\lambda^{***} = \boldsymbol{\mu}' V \Delta J^{-1} A' [AJ^{-1}\Delta' V \Gamma V \Delta J^{-1} A']^{-1} AJ^{-1}\Delta' V \boldsymbol{\mu}, \tag{20}$$

and GT denoting either GS or GW.

(c) $GT^0 \overset{a}{=} \lambda^{***}$,
where GT^0 is the corresponding value of GT when $\boldsymbol{\sigma}^0$ substitutes s.

Proof. Using (ii) and (iii) of Lemma 1, it is easily obtained that

$$GT \overset{a}{=} n(\mathrm{s} - \boldsymbol{\sigma}_0)'V\Delta J^{-1}A'[AJ^{-1}\Delta' V\Gamma V\Delta J^{-1}A']^{-1}AJ^{-1}\Delta' V(\mathrm{s} - \boldsymbol{\sigma}_0), \tag{21}$$

which proves (a). The results (b) and (c) follow easily from (21), and the stated assumptions, after applying the above used typical results on quadratic forms in normal variables. □

By comparing the quadratic forms on the right hand side of (21) and (11), the following corrollary is obvious.

Corollary 5.1: With the conditions of Theorem 5.1, if the equality

$$AJ^{-1}\Delta' V\Gamma V\Delta J^{-1}A' = A(\Delta' V\Delta)^{-1}A' \tag{22}$$

holds, then GT (GS or GW) is asymptotically equal to T (D, S or W). (Thus, GT and T have the same asymptotic chi-square distribution.)

Remark 5.1. This corollary implies that the condition (22) (together with Assumptions 1 to 5) is a sufficient condition for T to be asymptotically chi-square distributed. Note that condition (22) depends on the model and the restrictions being tested, and that is implied by Assumption 6, or just by Assumption 6*.

A sufficient condition for the goodness of fit statistic $n\tilde{F}$ to be asymptotically chi-square distributed is implied, also, by the corollary above. Effectively, $n\tilde{F}$ can be represented as a statistic T for an appropriate set of restrictions imposed on the elements of $\boldsymbol{\sigma}$, which may be viewed as p functionally independent parameters (see, e.g., Satorra & Saris, 1985, p. 87). It can easily be seen that, for such representation of $n\tilde{F}$ as a statistic T the equality (22) reduces to (recall that here V is non singular)

$$P\Gamma P' = PV^{-1}P', \tag{23}$$

where P is a full row rank $(p - q) \times p$ matrix such that $P\Delta = 0$ (P' is an orthogonal complement of the Jacobian matrix Δ; basically, the argument is that the restrictions $\mathbf{a}(\boldsymbol{\sigma}) = \mathbf{0}$ imposed on $\boldsymbol{\sigma}$ need to be such that $\mathbf{a}(\boldsymbol{\sigma}(\boldsymbol{\theta})) = \mathbf{0}$, in a neighborhood of $\boldsymbol{\theta}_0$; thus, the rule for a derivative of a composite function gives the stated orthogonality between P and Δ). This is in agreement with Shapiro (1986), where the equality (23) is stated as a sufficient (and when Γ is nonsingular, also a necessary) condition for $n\tilde{F}$ to be asymptotically chi-square distributed (provided that conditions similar to Assumptions 1 to 5 hold).

Furthermore, with respect to the above reparameterization of H, the right hand side of (21) becomes

$$n(\mathrm{s} - \boldsymbol{\sigma}_0)'P'[P\Gamma P']^{-1}P(\mathrm{s} - \boldsymbol{\sigma}_0) \overset{a}{=} n(\mathrm{s} - \tilde{\boldsymbol{\sigma}})'P'[P\Gamma P']^{-1}P(\mathrm{s} - \tilde{\boldsymbol{\sigma}}) =$$

$$n(\mathrm{s} - \tilde{\boldsymbol{\sigma}})'[\Gamma^{-1} - \Gamma^{-1}\Delta(\Delta'\Gamma^{-1}\Delta)^{-1}\Delta'\Gamma^{-1}](\mathrm{s} - \tilde{\boldsymbol{\sigma}}) \overset{a}{=}$$

$$n(\mathrm{s} - \tilde{\boldsymbol{\sigma}})'[\tilde{\Gamma}^{-1} - \tilde{\Gamma}^{-1}\tilde{\Delta}(\tilde{\Delta}'\tilde{\Gamma}^{-1}\tilde{\Delta})^{-1}\tilde{\Delta}'\tilde{\Gamma}^{-1}](\mathrm{s} - \tilde{\boldsymbol{\sigma}}) \equiv GF,$$

say, (the first asymptotic equality used (vi) of Lemma 1 combined with the fact that $P\Delta = 0$; the second equality uses a typical result on matrix algebra stated, for example, in Rao (1965, p. 77). The statistic GF is proposed in Browne (1982, p. 99; 1984, p. 69) as a (asymptotically chi-square) goodness of fit statistic for the model H, to be used in conjunction with GLS estimators in general. Clearly, under the conditions of Theorem 5.1, the generalized goodness of fit statistic GF will be asymptotically chi-square distributed with df = rank (P) = $(p - q)$ (Browne 1982, 1984).

Note that when H holds, then λ^{***} in (20) simplifies to

$$\lambda^{***} = \boldsymbol{\delta}' A' \underline{(AJ^{-1}}\Delta' V\Gamma V \Delta J^{-1} A')^{-1} A \boldsymbol{\delta}, \tag{24}$$

which value may change among different non AO discrepancy functions (as λ^{***} depends on F, through V); thus, the (asymptotic) power of the associated test may change also.

The above corollary says that, in case F is AO (or, less restrictively, (22) holds), T and GT can not be distinguished asymptotically; however, when F is not AO, T is not necessarily a chi-square statistic. If F_{AO} denotes an alternative AO discrepancy function, the statistic GT competes asymptotically with the statistic T_{AO} (D, S and W), say, associated with F_{AO}. When F is not AO, the point of comparison of the power of the test associated with the use of GT instead of T_{AO} arises. If V_{AO} and J_{AO} denotes, respectively, the matrices V and J associated with F_{AO}, then it has to be compared the ncp given by (10),

$$\boldsymbol{\mu}'(V_{\text{AO}} \Delta J_{\text{AO}}^{-1} A' (A J_{\text{AO}}^{-1} A')^{-1} A J_{\text{AO}}^{-1} \Delta' V_{\text{AO}}) \boldsymbol{\mu} = \lambda_{\text{AO}}, \text{ say,}$$

with the ncp given in (20). Consider the simplest case, where V is nonsingular and H holds. Then, as the following matrix

$$J^{-1} \Delta' V\Gamma V \Delta J^{-1} - J_{\text{AO}}^{-1} \tag{25}$$

is a positive semidefinite matrix (see Bentler & Dijkstra, 1985, p. 17), the difference $(\lambda_{\text{AO}} - \lambda^{***})$ is a nonnegative number. When J and J_{AO} are nonsingular matrices, (25) is equal to zero if and only if the estimator $\tilde{\boldsymbol{\theta}}$, based on F, is asymptotically efficient (see Shapiro, 1986). Thus, when $\tilde{\boldsymbol{\theta}}$ does not have minimum asymptotic variance, GT is not an asymptotically optimal test statistic; that is, the test induced by T_{AO} may have, asymptotically, greater power than the one induced by GT.

For the sake of completeness of this paper, we give next the asymptotic distribution of the (restricted) estimator $\hat{\boldsymbol{\theta}}$ in the general case where F is not necessarily AO. This is a result which can be traced back to Ferguson (1958, Theorem 1). (See also Bentler & Dijkstra, 1985; Browne 1982, 1984; Satorra & Bentler, 1988b; and Shapiro 1983, 1986.)

Theorem 5.2. Suppose Assumptions 1–5 hold, then

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N((J \cdot A)^{-1} \Delta' V \boldsymbol{\mu},\ (J \cdot A) \Delta' V\Gamma V \Delta (J \cdot A))$$

Proof. It is obvious when considering (i) of Lemma 1 and Assumption 2. □

Thus, the covariance matrix of $\hat{\boldsymbol{\theta}}$ will be estimated by

$$n^{-1}(J \cdot A) \Delta' V\Gamma V \Delta (J \cdot A), \tag{26}$$

substituting consistent estimates for population matrices (i.e., evaluating $(J \cdot A)^{-1} \Delta' V$ at $\hat{\boldsymbol{\theta}}$ and changing Γ to $\bar{\Gamma}$). Note that Theorem 5.2 allows one to obtain asymptotically correct standard errors of the estimators under fairly general conditions; for instance, one may obtain asymptotically correct standard errors for the (maximum likelihood) normal theory estimators even when the distribution of the data is not normal.

Note also that when Assumption 6* is verified, the above asymptotic covariance matrix reduces to $n^{-1}(J \cdot A)$, which is the typical form of the covariance matrix of the restricted estimator $\hat{\boldsymbol{\theta}}$ under standard assumptions (Browne, 1982; Shapiro, 1983; Bentler & Dijkstra, 1985). Furthermore, when V is nonsingular,

$$(J \cdot A) - (J \cdot A)\Delta' V \Gamma V \Delta (J \cdot A) = (J \cdot A)\Delta' V (V^{-1} - \Gamma) V \Delta (J \cdot A);$$

thus, when Assumption 6* is not verified, the typical standard errors associated with $(J \cdot A)$ may be asymptotically biased. This bias does not need to be systematically positive, or negative, as $(V^{-1} - \Gamma)$ is not in general a positive (negative) definite matrix.

Obviously, when considering the asymptotic distribution of the unrestricted estimator $\tilde{\boldsymbol{\theta}}$, $(J \cdot A)$ need to change to $(J \cdot A_1)$ (i.e., to J^{-1}, when J is nonsingular). Note that when $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ amounts to fix a single component of $\boldsymbol{\theta}$ to zero, the statistic GW in (19) can easily be recognized as the square of the ratio between an estimate and its standard error (i.e., a square of a "t-value"), where now the standard error is the asymptotically correct one extracted from (26). (Recall that the t-value associated with W uses the standard error extracted from J^{-1}; that is, the correct standard error when Assumption 6 hold.)

Finally, we will consider the asymptotic distribution of the ("parameter change") statistic $\hat{\mathbf{z}}$ defined in (4). This statistic is proposed in Saris et al. (1987) as a tool for assessing the substantive significance of dropping the restrictions $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$. The next theorem states, first, the asymptotic equality of $n^{1/2}\hat{\mathbf{z}}$ and $n^{1/2}\mathbf{a}(\tilde{\boldsymbol{\theta}})$; then gives the asymptotic distribution of $n^{1/2}\hat{\mathbf{z}}$; finally, provides an approximation to the asymptotic mean of $n^{1/2}\hat{\mathbf{z}}$.

Theorem 5.3. Let Assumptions 1 through 5 hold, then

(a) $n^{1/2}\hat{\mathbf{z}} \overset{a}{=} n^{1/2}\mathbf{a}(\tilde{\boldsymbol{\theta}})$.
(b) $n^{1/2}\hat{\mathbf{z}} \xrightarrow{L} N(A(J \cdot A_1)\Delta' V \boldsymbol{\mu},\ A(J \cdot A_1)\Delta' V \Gamma V \Delta (J \cdot A_1) A')$.
(c) $n^{1/2}\mathbf{z}^0 \overset{a}{=} A(J \cdot A_1)\Delta' V \boldsymbol{\mu}$,

where $\mathbf{z}^0$ be the value of $\hat{\mathbf{z}}$ obtained when $\boldsymbol{\sigma}^0$ substitutes $\mathbf{s}$.

Proof. (a) is a trivial consequence of (i) and (v) of Lemma 1; while (b) and (c) follows immediately from (v) of Lemma 1 and Assumption 2. □

6. Discussion

The sections above provide insights of practical relevance. Result (a) of Theorem 4.1 suggests that, as $n \to \infty$, the values of the alternative test statistics D, S and W get closer to each other and have the same asymptotic distribution, even if that asymptotic distribution is not chi-square (which may be the case when Assumption 6* does not hold; see also Remark 4.1). This asymptotic equality between D, S and W justifies the practice of interpreting the statistic S (or W) as the approximate decrease (increase) on the value of the goodness of fit statistic $n\hat{F}$ ($n\tilde{F}$), of the fitted model H_0 (or H), when dropping (adding) the restriction $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$. Note that this equivalence is ensured just by Assumptions 1 to 5; that is, it will hold regardless of F being AO or not (e.g., it will hold even when in a LISREL or EQS analysis one uses the "unweighted least squares" discrepancy function). The (asymptotic) equality, hence equivalence with respect to asymptotic performance (e.g., they will lead to tests with the same asymptotic size and power), extends also to the statistics GS and GW whenever Assumption 6*, or the less restrictive condition (22), holds. However, the (asymptotic) equality between GT (GS and GW) and D may break down when (22) does not hold; which means, for instance, that when D is not asymptotically a chi-square statistic, the (asymptotically chi-square) statistic GS (GW) can be a "poor predictor" of the change on $n\hat{F}$ ($n\tilde{F}$) as a result of dropping (adding) the restriction $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$.

When F is AO (or just when equality (22) holds), the common asymptotic distribution of the statistics D, S and W is chi-square. Under "nonrestrictive" conditions (just Assumptions 1 to 5), GS and GW are also asymptotically chi-square distributed. As mentioned in section 2, to each of the above asymptotic chi-square statistics corresponds an α-level test of H_0 against H, with rejection region $[T \geq c_\alpha]$. The (asymptotic) noncentral chi-square behavior of the test statistic when the model is not exactly true suggest the approximation $\Pr\{\chi^2_{df}(\lambda) \geq c_\alpha\}$ for the power of the test associated with a specific alternative $\boldsymbol{\sigma}^0$, with λ being determined by $\boldsymbol{\sigma}^0$. The theorems of sections 4 and 5 provide alternative asymptotically equivalent expressions for the ncp λ to be used in practice. For instance, result (c) of Theorem 4.1 shows that if $\boldsymbol{\sigma}^0$ is chosen as the alternative value to which the power value is referred, the noncentrality parameter λ can be approximated by using any one of the (nonstochastic) values of D, S and W obtained when $\boldsymbol{\sigma}^0$ is "analysed" instead of s (e.g., in a maximum likelihood LISREL analysis of $\boldsymbol{\sigma}^0$ under the specification H_0, the modification indexes, the square of the t-values and also the chi-square statistic can be viewed as asymptotic approximations of ncp's).

The relative merits of the variety of tests defined above, and alternative approximations of λ, with respect to small sample size performance (in having the predicted size and power) is a topic which remains to be investigated (by using, e.g., Monte Carlo methods, or higher order asymptotics). A Monte Carlo study of the accuracy of D^0 as a noncentrality parameter for fitting the finite sample size distribution of D, in case of normality, maximum likelihood and a specific model context, is reported in Satorra and Saris (1983). Satorra and Saris (1985) advocate the use of $nF^0 = n\{\mathrm{Min}_\Theta F(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}(\boldsymbol{\theta}))\}$ when approximating the power of the chi-square goodness of fit test associated with $n\hat{F}$ (in case of maximum likelihood and normal data). Satorra et al. (1987) compare in a small sample size study, and different "degrees" of the incorrectness of the model, alternative approximations for the ncp λ. Recently, the small sample size behavior, and comparative performance in model modification, of some of the above statistics has been investigated, using Monte Carlo methods, by Luijben, Boomsma and Molenaar (1987), and Chou and Bentler (1987).

Although Assumptions 1 to 5 are quite plausible in applications, the assumption of F being AO will often be violated in practice. The Corollary 5.1 can be seen as providing sufficient conditions for the statistics D, S and W to be asymptotically chi-square distributed when F is not AO; see, especially, Remark 5.1. Clearly, when F is not AO, or the equality (22) does not hold, the statistics D, S and W can be misleading, as they are not necessarily asymptotic chi-square statistics. An alternative is, of course, to use a more appropriate discrepancy function F in order to reach this AO condition. However, the use of an AO discrepancy function F may imply extensive computations. A sensible alternative is to use the GS and GW statistics; after all, when F is AO, the generalized statistic GT has the same asymptotic distribution as the statistic T (Corollary 5.1), and will lead to test of the same characteristics with respect to (asymptotic) size and power. One may argue, however, that test statistics which are asymptotic chi-square variates, and do not require heavy computations, can be obtained via linearized estimators (see Bentler & Dijkstra 1985, and the statistics S_L and W_L discussed in section 4). It has also been shown (section 5) that a statistic T associated with an AO discrepancy function may even have greater power than a statistic GT associated with a non AO discrepancy function. Clearly additional research that studies the robustness against small sample size, and performance with respect to power, of these competing test statistics is required.

Notice that Assumptions 1 to 5 guarantee (see (a) of Theorem 4.1)

$$n\hat{F} = n\tilde{F} + T, \tag{27}$$

asymptotically, which implies (using e.g., Theorem 9.3.6 of Rao & Mitra, 1971) that whenever the asymptotic distribution of $n\hat{F}$ and $n\tilde{F}$ (or $\eta^{-1}n\hat{F}$ and $\eta^{-1}n\tilde{F}$, for some positive real-value η) is chi-square, so will be the asymptotic distribution of the statistic T (or $\eta^{-1}T$). Also, when a scaling correction factor applies to the chi-square goodness of fit statistics $n\hat{F}$ and $n\tilde{F}$, then the same scaling correction factor applies to the statistics S and W (i.e., the elliptical corrections of Browne, 1982, 1984, and Bentler, 1983*b*, extend also to the statistics D, S and W). Nonstandard conditions under which $n\hat{F}$ ($n\tilde{F}$) is asymptotically chi-square distributed, and conditions under which a specific scaling correction factor apply (in an asymptotically exact form, or approximately) are investigated in Satorra and Bentler (1988a, 1988b). Satorra and Bentler (1988b) propose a scaling correction to the D, S and W statistics that may induce an approximate chi-square behavior of the resulting statistics under general conditions. The correction consists of dividing the statistics D, S and W by (1/df)trace($U\Gamma$), with consistent estimates replacing U and Γ (recall that U was defined in section 4). For further non standard conditions guaranteeing the asymptotic chi-squaredness of test statistics, see the recent papers of Shapiro (1987) and Browne (1987).

Note the asymptotic independence between the goodness of fit statistic $n\tilde{F}$ and T, which contrasts with the nonzero (asymptotic) covariance between T and $n\hat{F}$ (see (c) and (d) of Theorem 4.2 in Steiger et al., 1985). In practice this implies that once the goodness of fit of a model has been assessed by the chi-square goodness of fit statistic, there are still available (asymptotically) independent statistics for testing if some additional restrictions hold; however, when testing if some restrictions should be dropped (typically using a score statistic), stochastic dependence with the chi-square goodness of fit test arises (i.e., the role of the chi-square goodness of fit statistic in a process of modifying the model by dropping restrictions, may be very different than when the process of model modification consists in adding restrictions).

When a sequence of models is considered, Corollary 4.2 expands the number of asymptotically equivalent statistics in a useful way. For instance, point (a) of the corollary implies that when testing the restrictions that distinguish H_0 from H_1 (where H_1 can be any model nested "between" H_0 and H), it is not needed in the computations to change the adopted model from H to H_1, as one can just use the difference between the statistics T and T_1, which both use H as the adopted model. (The HAPRIORI procedure of EQS (Bentler, 1986) implements the testing of the restrictions associated with a nested sequence of hypotheses.)

When F is AO (or just Assumption 6* holds), (16) of Theorem 4.3 ensures the asymptotic independence between the tests statistics T_a and T_b and defines the hypothesis (restrictions) $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{b}(\boldsymbol{\theta}) = \mathbf{0}$ to be *separable* (Aitchison, 1962, p. 238); in that case, the test statistic T_{ac}, of the combined hypothesis $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{b}(\boldsymbol{\theta}) = \mathbf{0}$, decomposes (asymptotically) as the sum of T_a and T_b. When F is not AO, the asymptotic independence of T_a and T_b is ensured by (17) even when they are not chi-square distributed. It can easily be shown that (17) assures also the separability of $\mathbf{a}(\boldsymbol{\theta}) = 0$ and $\mathbf{b}(\boldsymbol{\theta}) = \mathbf{0}$ with respect to the GT statistics; that is, (17) implies the independence of the GT statistics and the corresponding decomposition $GT_{ac} \overset{a}{=} GT_a + GT_c$ (obvious notation is used).

The expressions for calculating S and GS (W and GW) involve vectors and matrices evaluated at the restricted (unrestricted) estimate $\hat{\boldsymbol{\theta}}$ (or $\tilde{\boldsymbol{\theta}}$); thus, S and GS (W and GW) have the computational advantage compared with D (D requires the fit of the restricted and unrestricted models) in requiring just the fit of one model. Obviously, to compute GS and GW a consistent estimate of Γ needs to be available also. Although for large models Γ may be a matrix of high dimension, the fact that it does not get involved in any iterative process (and does not even need to be inverted), implies that the computations of GW and GS will not be of substantively higher "cost" than the computations of S or W. Clearly, the above statistics are obtained just by matrix algebra when one has available a consistent estimate of Γ and the vector of derivatives $\mathbf{d}$, and matrices Δ and V, evaluated at specified values ($\hat{\boldsymbol{\theta}}$ or $\tilde{\boldsymbol{\theta}}$). Note that for both statistics, S and W (GS and GW), the expressions for $\mathbf{d}$ and Δ correspond with the same adopted model H. (The above "intermediate" statistics Γ, $\mathbf{d}$, Δ and V, will be available as technical output in the latest version of EQS; Bentler, 1985.)

Note that the misspecification on the model allowed by Assumption 2 (i.e., it allows the true $\boldsymbol{\sigma}^0$ not to satisfy the model H_0 (H), and a nonnull value for $\boldsymbol{\mu}$), produces the nonzero values of the ncp's. It can also be seen to induce to $\hat{\boldsymbol{\theta}}$ a "bias", with respect to $\boldsymbol{\theta}_0$, that is of order of magnitude $n^{-1/2}$. (Substituting in (i) of Lemma 1, $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$ for $\boldsymbol{\mu}$, it is obtained that this "bias term" equals $(J \cdot A)\Delta' V(\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}_0)$, which is zero when $\boldsymbol{\mu} = \mathbf{0}$, that is, when $\boldsymbol{\sigma}^0 \equiv \boldsymbol{\sigma}_0$. Of course, in a strict sense, even when $\boldsymbol{\mu} \neq \mathbf{0}$, $\hat{\boldsymbol{\theta}}$ remains consistent due to the fact that the assumption of a sequence of local alternatives guarantees that $(\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}_0) \to \mathbf{0}$ as $n \to \infty$.). However, this type of misspecification has no effect on the asymptotic distribution of the statistics, nor does it bias the estimates of the standard errors (see section 4). This contrast sharply with the effects of F not being AO, which may distort the typical asymptotic distribution of the statistics and biases (asymptotically) the typical estimates of the standard errors.

In practice it is very likely that the model is slightly misspecified (if not grossly) and that F is not AO. For small misspecification of the model, the asymptotic theory given above can provide sensible approximations to the actual distribution of the statistics of interest. For grossly misspecified models the above theory does not hold. Unfortunately, here the term "small" ("grossly") has a similar loose meaning as the "large" ("small") of the jargon of asymptotic theory (when referring to sample size): given a specific context, "small" ("grossly") will be a matter of empirical evaluation, typically using Monte Carlo methods. On theoretical grounds, it can be said that, for given data, a conflict among alternative statistics that are asymptotically equivalent under certain assumptions should be interpreted as that, either the sample size is "still" small, or some assumption is violated. Thus, a conflict among the alternative statistics T (D, S and W) may indicate that H_0 (or H) can be grossly misspecified, while a conflict among T and GT may indicate that F is not AO. However, the above interpretation need to be mediated by considerations of how much negligible are, given the actual finite sample size and possible values of the parameters, the quantities being ignored by the first order asymptotic approximations of, for example, Lemma 1. Different parameterizations may also lead to substantial changes on the quality of these approximations (and note that in general W and GW are not invariant under such reparameterizations). (In the context of logistic regression and maximum likelihood, the inadequacy of the quadratic approximation to the likelihood contour, causing finite sample size conflict among asymptotic equivalent statistics, has been investigated, for exsample, by Lustbader, Moolgavkar, and Venzon, 1984; and Jennings, 1986.)

It has to be noted that GS and GW can be related with test statistics introduced in White (1982), and Burguete, Gallant and Souza (1982), in the context of quasi-maximum likelihood estimation and nonlinear econometric models, respectively. In these references a distinction is being made between the "outer" and "hessian" form of the information matrix, which are assumed not to be equal as a result of some type of "distribution misspecification" (that is, the typical information matrix equality is not assumed to hold); then, score and Wald statistics are defined accordingly. The similarity of White's and Burguete et. al's score and Wald statistics with the statistics *GS* and *GW* introduced in section 5 becomes apparent when one regards $\Delta' V\Gamma V\Delta$ and $\Delta' V\Delta$ as the "outer" and "hessian" form, respectively, of the "information matrix" (note that, in general, neither of these matrices is strictly the information matrix, as no full specification of the likelihood function of the data is involved). Then, Assumption 6* would play the role of the "information matrix equality" which is a sufficient condition for T (D, S or W) to be asymptotically a chi-square statistic, and $n^{-1} J^{-1}$ to be the "correct" matrix of (asymptotic) variances and covariances of $\tilde{\boldsymbol{\theta}}$. Note, however, that a less restrictive condition than Assumption 6*, namely (22), is a sufficient condition for T to be asymptotically chi-square distributed. Therefore, in order to assess the adequacy of T as a chi-square statistic, an "information matrix test" of the type introduced in White (1982) could be based on (22), involving matrices of much lower dimension than when based on the equality stated in Assumption 6*.

We will also mention that the results of this paper are valid in a more general context than covariance structure analysis. The vector $\mathbf{s}$, and the vector of parameters $\boldsymbol{\sigma}$ that is modeled, may contain many types of moments: means, product-moment, frequencies (proportions), and so forth. Thus, such an approach includes a great variety of techniques as factor analysis, simultaneous equations for continuous variables, log linear or multinomial parametric models, and so forth. Of course, in order for the above results to apply, caution must be exerted with the assumptions fully stated in section 2. In practice, the observance of those assumptions will boil down to just taking care of using an appropriate consistent estimate of Γ, and distinguishing whether or not Assumption 6 is verified. (In van Praag et. al., 1985, it is shown that also a specific type of the "incomplete observations problem" can be "modeled" by a specific form of Γ; also, for the structure of Γ in the case of "controlled" or "repeated" sampling experiments, see van Praag, de Leeuw and Kloek, 1986.)

The case of a "multisample analysis" (Jöreskog & Sörbom, 1984; see also Lee & Tsui, 1982; and section 4 of Lee, 1985) is also encompassed by the theory above. In such a case, the vectors $\mathbf{s}$ and $\boldsymbol{\sigma}$ are partitioned in (stochastically independent) subvectors, $\mathbf{s}_g$ and $\boldsymbol{\sigma}_g$ ($g = 1, 2, \cdots, G$); thus, the matrix Γ is block diagonal, with diagonal blocks, Γ_g, conformable with the $\mathbf{s}_g$'s. Typically V will also be block diagonal (with the V_g's in the diagonal) and thus the assumption of F being AO (Assumption 6) will be ensured when $V_g\Gamma_gV_g = V_g$ holds for each group g.

In conclusion, a variety of competing test statistics, some of which are classical ones and others are new statistics, have been reviewed in a unified framework using asymptotics. Under a restrictive assumption (which restricts the fitting function to satisfy an optimality condition), all the above statistics can not be distinguished asymptotically, as far as first order asymptotics are used. Under very general assumptions (Assumptions 1 to 5), some of the statistics may be inadequate as not being asymptotically chi-square distributed, some may lead to tests with greater power than others, but some are still (asymptotically) chi-square statistics. The review identifies basic assumptions for their asymptotic behavior and sets a theoretical ground for their comparative performance (as far as asymptotic theory can be a guide). It has to be expected, however, that computational convenience, and small sample size performance, will be the key factors when assessing the comparative usefulness of the statistics reviewed above, when used in a practical context. Clearly, those are topics for further research that go beyond the purpose of the present paper.

Appendix

Proof of Lemma 1. The result (i) follows from the Assumptions 1 through 5 and the application of the implicit function theorem to $\partial L/\partial \boldsymbol{\tau} = \mathbf{0}$, where

$$L = F(\mathbf{s}, \boldsymbol{\sigma}(\boldsymbol{\theta})) - \mathbf{l}'\mathbf{a}(\boldsymbol{\theta}),$$

$\mathbf{l}$ is a r-vector of Lagrangian multipliers, and $\boldsymbol{\tau} = (\boldsymbol{\theta}', \mathbf{l}')'$ (see Lemma 2 of Dijkstra, 1983, p. 70; also Shapiro, 1983, 1985b, 1986). (Here the result $\partial^2 F/\partial\boldsymbol{\sigma}\partial\mathbf{s}' = -\partial^2 F/\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'$, see e.g. Shapiro, 1985a, was needed.) Note that when J is singular the (unrestricted) estimator $\tilde{\boldsymbol{\theta}}$ has been defined to be the estimator associated with the restriction $\mathbf{a}_1(\boldsymbol{\theta}) = \mathbf{0}$.

The results (ii) and (iii) follows from a first order Taylor expansion of $\mathbf{a}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, and $\partial F/\partial\boldsymbol{\theta}$ around $\tilde{\boldsymbol{\theta}}$, respectively.

To prove (iv), consider the function of s

$$f(\mathbf{s}) = min_{\boldsymbol{\tau}} L(\mathbf{s}, \boldsymbol{\tau}),$$

for which $f(\mathbf{s}) = \hat{F}$, and which under the Assumptions 1 to 5 is twice continuously differentiable (see, e.g., Theorem 4.2 of Shapiro, 1983). Then, consider the following Taylor expansion:

$$f(\mathbf{s}) = f(\boldsymbol{\sigma}_0) + \left(\frac{\partial f(\boldsymbol{\sigma}_0)}{\partial \mathbf{s}'}\right)(\mathbf{s}-\boldsymbol{\sigma}_0) + \tfrac{1}{2}(\mathbf{s} - \boldsymbol{\sigma}_0)' \left(\frac{\partial^2 f(\boldsymbol{\sigma}_0^-)}{\partial\mathbf{s}\partial\mathbf{s}'}\right)(\mathbf{s} - \boldsymbol{\sigma}_0),$$

where $\boldsymbol{\sigma}_0^-$ is between s and $\boldsymbol{\sigma}_0$. As the first and second term of the right hand side of the above equality are zero, and $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$ is bounded in probability, we get:

$$nf(\mathbf{s}) \stackrel{a}{=} n(\mathbf{s} - \boldsymbol{\sigma}_0)'(V - V\Delta(J\cdot A)^{-1}\Delta' V)(\mathbf{s} - \boldsymbol{\sigma}_0),$$

where $\partial^2 f(\boldsymbol{\sigma}_0)/\partial\mathbf{s}\partial\mathbf{s}'$ has been appropriately substituted (see, e.g., (4.4) of Shapiro, 1983, p. 48). See also Shapiro (1985b, 1986).

The result (v) follows directly from the definition of $\hat{\mathbf{z}}$ (see (4)) and (iii).

Finally, (vi) results from the following Taylor series expansion of the function $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$:

$$n^{1/2}\hat{\boldsymbol{\sigma}} \stackrel{a}{=} n^{1/2}\boldsymbol{\sigma}_0 + \Delta n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

and then applying (i) of the Lemma. In all these derivations the obvious equalities $\hat{A} \stackrel{a}{=} A$, $\tilde{A} \stackrel{a}{=} A$, $\hat{J} \stackrel{a}{=} J$ and $\tilde{J} \stackrel{a}{=} J$ have been used. □

The next lemma follows using simple matrix algebra. The proof will not be reproduced here and is available from the author when requested.

Lemma 2. Let J be a nonnegative definite matrix, and $[J, A_1']$ and $[A_1', A_2']'$ well defined partitioned matrices of full row rank, and $(J\cdot A)$ and $(J\cdot A_1)$ as in (9), then

$$(J\cdot A) = (J\cdot A_1) - (J\cdot A_1)A_2'(A_2(J\cdot A_1)A_2')^{-1}A_2(J\cdot A_1). \tag{A1}$$

Remark A1. Using (9), (A1) is easily seen to be equivalent to:

$$J^{-1}A'(AJ^{-1}A')^{-1}AJ^{-1} = J^{-1}A_1'(A_1J^{-1}A_1')^{-1}A_1J^{-1} + (J\cdot A_1)A_2'(A_2(J\cdot A_1)A_2')^{-1}A_2(J\cdot A_1).$$

References

Aitchison, J. (1962). Large-sample restricted parametric tests. *Journal of the Royal Statistical Society, Series B, 24*, 234–250.

Bentler, P. M. (1983a). Simultaneous equations as moment structure models: With an introduction to latent variable models. *Journal of Econometrics, 22*, 13–42.

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

Bentler, P. M. (1985). *Theory and implementation of EQS, a structural equations program.* Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1986). *Lagrange multiplier and Wald tests for EQS and EQS/PC* (Technical Report). Los Angeles: BMDP Statistical Software.

Bentler, P. M., & Chou, C. P. (1986). *Statistics for parameter expansion and contraction in structural models.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Kirshnaiah (Ed.), *Multivariate analysis VI*(pp. 9–42), Amsterdam: North-Holland.

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

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika, 74*, 375–84.

Burguete, J. F., Gallant, A. R., & Souza, G. (1982). On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews, 1*, 151–190.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician, 36*, 153–157.

Chou, C. P., & Bentler, P. M. (1987). *Model modification in covariance structure modeling: A comparison among likelihood ratio, lagrange multiplier and Wald tests.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D. C.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics.* London: Chapman & Hall.

Dijkstra, T. (1983). Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics, 22*, 67–90.

Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier test in econometrics. In Z. Griliches & M. Intriligator (Eds.), *Handbook of Econometrics* (pp. 775–825). Amsterdam: North-Holland.

Ferguson, T. S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics, 29*, 1046–1062.

Jennings, D. E. (1986), Judging inference adequacy in logistic regression. *Journal of the American Statistical Association, 81*, 471–476.

Jöreskog, K. G. (1970). A general method for the analysis of covariance structures. *Biometrika, 57*, 239–251.

Jöreskog, K. G. (1977). *Structural equation models in the social sciences: Specification, estimation and testing.* In P. R. Krishnaiah (Ed.), Application of statistics (p. 265–287). Amsterdam: North-Holland.

Jöreskog, K. G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics, 8*, 65–92.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods.* Mooresville: Scientific Software.

Lee, S.-Y. (1985). On testing functional constraints in structural equation models. *Biometrika, 72*, 125–131.

Lee, S.-Y., & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal, 14*, 121–136.

Lee, S.-Y., & Tsui, K. L. (1982). Covariance structure analysis in several populations. *Psychometrika, 47*, 297–308.

Luijben, T., Boomsma, A., & Molenaar, I. W. (1987). *Modification of factor analysis models in covariance structure analysis: A monte study. Groningen: University of Groningen, Heymans Bulletins Psychologische Instituten.*

Lustbader, E. D., Moolgavkar, S. H., & Venzon, D. J. (1984). Tests of the null hypothesis in case-control studies. *Biometrics, 40*, 1017–1024.

Rao, C. R. (1947). Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proceedings Cambridge Philosophy Society, 44*, 50–57.

Rao, C. R. (1965). *Linear statistical inference and its applications.* New York: John Wiley.

Rao, C. R., & Mitra, S. K. (1971). *Generalized inverse of matrices and its applications.* New York: John Wiley.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105–129). Washington, D. C.: American Sociological Association.

Saris, W. E., & Stronkhorst, L. H. (1984). *Causal modeling in nonexperimental research: An introduction to the LISREL approach.* Amsterdam: Sociometric Research Foundation.

Satorra, A., & Bentler, P. M. (1988a). *Scaling corrections for statistics on covariance structure analysis* (UCLA Statistics Series #2). Los Angeles: University of California.

Satorra, A., & Bentler, P. M. (1988b). *Model conditions for asymptotic robustness in the analysis of linear relations* (UCLA Statistic Series #4). Los Angeles: Univeristy of California.

Satorra, A. & Saris, W. E. (1983). The accuracy of a procedure for calculating the power of the likelihood ratio test as used within the LISREL framework. In C. P. Middendorp, B. Niemöller & W. E. Saris (Eds.). *Sociometric Research 1982* (pp. 127–190). Amsterdam: Sociometric Research Foundation.

Satorra, A. & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50,* 83–90.

Satorra, A., Saris, W. E., & de Pijper, W. M. (1987). A comparison of several approximations to the power of the likelihood ratio test in covariance structure analysis. *American Statistical Association 1987 Proceedings, Business and Economic Statistics Section* (pp. 393–398).

Serfling, R. J. (1980). *Approximation theorems in mathematical statistics.* New York: Wiley.

Searle, S. R. (1982). *Matrix algebra useful for statistics.* New York: Wiley.

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

Shapiro, A. (1984). A note on the consistency of estimates in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 84–88.

Shapiro, A. (1985a). Asymptotic equivalence of minimum discrepancy function estimators to G.L.S. estimators. *South African Statistical Journal, 19,* 73–81.

Shapiro, A (1985b). Second-order derivatives of extremal-value functions and optimality conditions for semi-infinite programs. *Mathematics of Operations Research, 10,* 207–219.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association, 81,* 142–149.

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

Silvey, S. D. (1959). The Lagrangian multiplier test. *Annals of Mathematical Statistics, 30,* 389–407.

Sörbom, D. (1986). *Model modification* (Research Report 86-3). Uppsala: University of Uppsala.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50,* 253–263.

van Praag R. M. S., Dijkstra, T. K., & Van Velzen J. (1985). Least-squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika, 50,* 25–36.

van Praag R.M.S., de Leeuw, J., & Kloek T. (1986). The population-sample decomposition approach to multivariate estimation methods. *Applied Stochastic Models and Data Analysis, 2,* 99–119.

Wald, A (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426–482.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50,* 1–25.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9,* 60–62.

Manuscript received 1/22/86
Final version received 11/27/87