ECONOMICS WORKING PAPERS SERIES

Economics Working Paper 35

On the Asymptotic Optimality of Alternative Minimum-Distance Estimators in Linear Latent-Variable Models*

Albert Satorra
Universitat Pompeu Fabra
and
Heinz Neudecker
University of Amsterdam

Revised February 1993

UNIVERSITAT POMPEU FABRA

Balmes, 132
Telephone (343) 484 97 00
Fax (343) 484 97 02
08008 Barcelona
e- mail econwp@upf.es

Abstract

In the context of linear latent-variable models, and a general type of distribution of the data, the asymptotic optimality of a subvector of minimum-distance estimators whose weight matrix uses only second-order moments is investigated. The asymptotic optimality extends to the whole vector of parameter estimators, if additional restrictions on the third-order moments of the variables are imposed. Results related to the optimality of normal (pseudo) maximum-likelihood methods are also encompassed. The results derived concern a wide class of latent-variable models and estimation methods used routinely in software for the analysis of latent-variable models such as LISREL, EQS and CALIS. The general results are specialized to the context of multivariate regression and simultaneous equations with errors in variables.

1 Introduction

Consider a vector s of sample moments that converges in probability to σ, a vector of population moments, and let $\sigma = \sigma(\theta)$ be a model for σ. Here, θ is a vector of parameters and $\sigma(.)$ is a continuously differentiable vector-valued function. Consider the minimum-distance (MD) estimator of θ defined as the solution $\hat{\theta}$ of

$$Min_{\theta\in\Theta}(s - \sigma(\theta))'V_n(s - \sigma(\theta)), \tag{1}$$

where Θ is the parameter space of θ and V_n is a positive semi-definite matrix that converges in probability (when sample size $n \to \infty$) to V, a positive semi-definite matrix (with V_n of the same rank as V). Under typical regularity conditions, it can be derived (e.g., Chamberlain, 1982; Browne, 1984; Bentler and Dijkstra, 1985; Shapiro, 1986; Fuller, 1987 Sect. 4.2; Satorra, 1989) that $\hat{\theta}$ is consistent and asymptotically normal with asymptotic variance matrix (avm)

$$\text{avm }(\hat{\theta}) = (\Delta'V\Delta)^{-1}\Delta'V\Gamma V\Delta(\Delta'V\Delta)^{-1}, \tag{2}$$

where $\Delta \equiv (\partial/\partial\theta')\sigma(\theta)$ and $\Gamma \equiv \text{avm}(s)$. If $V = \Gamma^-$ and Δ is contained in the column space of Γ (alternatively, if $V = V\Gamma V$), the corresponding estimator, which we denote by $\tilde{\theta}$, is asymptotically optimal (AO) since it has minimum avm within the class of MD estimators that minimize (1); in this case, (2) reduces to

$$\text{avm}(\tilde{\theta}) = (\Delta'\Gamma^-\Delta)^{-1}, \tag{3}$$

where for matrix H, H^- denotes a g-inverse of H (i.e. , $HH^-H = H$).

In this paper we will be concerned with models where σ is the vector of non-redundant (uncentered) second-order moments of a vector of observable variables z (possibly an augmented vector), and s is the corresponding vector of sample moments. We will discuss situations where there is equality among *submatrices* on the right-hand side of (2) and (3) even when $V \neq \Gamma^-$; i.e., we are concerned with conditions for asymptotic optimality of a subvector of MD estimators. In particular, we will show that, under certain conditions to be made explicit below, the normal (pseudo) maximum-likelihood estimator of a ***subvector*** τ of θ is AO even when z is nonnormally distributed.

Basically, two types of MD estimators are considered. The AO estimator $\tilde{\theta}$ introduced above, and the MD estimator $\hat{\theta}$ associated with

$$V_n = 2^{-1}\, D'(S^{-1} \otimes S^{-1})D, \tag{4}$$

or with a weight matrix whose probability limit is

$$\Omega^{-1} = 2^{-1} D'(\Sigma^{-1} \otimes \Sigma^{-1})D, \tag{5}$$

where S and Σ are, respectively, the sample and population (uncentered) second-order moment matrices of z. Here D and D^{+} are respectively the "duplication" and "elimination" matrix so that vec $A = D\text{v}\,(A)$ for symmetric matrix A, where "vec" is the usual columnwise vectorization operator and $\text{v}\,(A)$ is obtained from vec A after eliminating the duplicated elements associated with the symmetry of A. It holds that $\text{v}\,(A) = D^{+}\text{vec}\,A$ where $D^{+} = (D'D)^{-1}D$ is the Moore-Penrose inverse of D (see Magnus and Neudecker, 1986, 1991). For reasons to be given below, $\hat{\theta}$ will be called a normal MD (NMD) estimator.

The results concerning the NMD estimator $\hat{\theta}$ are of high practical relevance since it is (asymptotically) equivalent to the usual normal (pseudo) maximum-likelihood estimator (PML) of θ. For the wide class of latent-variable models considered in this paper, PML and NMD estimators are available in conventional computer programs for structural-equation models, such as LISCOMP (Muthén, 1987), LISREL (Jöreskog and Sörbom, 1989), EQS (Bentler, 1989), LINCS (Schoenberg, 1989) and CALIS (SAS, 1990).

The present paper relates to the work of Anderson and Amemiya (1988), Browne and Shapiro (1988), Anderson (1989), Satorra and Bentler (1990), and Mooijaart and Bentler (1991) on asymptotic robustness in covariance-structure analysis. We consider a more general class of models, however, since we encompass models that restrict means in addition to covariances. This is not a trivial extension, since it implies the solution of issues associated with generalized inverse matrices. On the other hand, the present work focuses on the robustness of the property of asymptotic optimality. The issue of asymptotic optimality was also considered in Browne-Shapiro's work, who exploited Shapiro's (1987) results on this topic (Shapiro's results, however, relied heavily on Rao and Mitra's (1971, Chapter 8) conditions for optimality in least-squares estimation). In contrast to that, our paper is self-contained and does not draw on the theory of optimality in least-squares estimation.

More recently, Browne (1990) has also considered models that involve restrictions on means as well as on covariances. We deviate from Browne's work in several respects. One difference is that we are concerned not only with asymptotic optimality of the *whole* vector of parameter estimators, but also with AO of a *subvector* of estimators (for example, the estimators of regression coeficients in a regression model). This enables us to get results on AO without restricting the third-order moments of observable variables, as required in Browne (1990). Further, we involve means in the analysis by using the augmented moment matrices (as advocated, for example, in Bentler, 1983). The use of augmented moment matrices enables us to simplify theory and to carry out mean-and-covariance structure analysis by using standard software for covariance structures. Finally, in contrast with Browne's approach, we do not use the theory of cumulants.

For an extensive discussion of minimum-distance estimation in mean-and-covariance structure analysis, and robustness results with regard to the validity of standard inference under violation of distributional assumptions, see the recent work of Satorra (1992a); his work, however, does not touch on the issue of asymptotic efficiency.

The plan of the paper is as follows. Section 2 presents the model framework to be considered and general results that give conditions for AO of NMD and PML. Specialization of the results to a general type of multivariate regression models, possibly with errors in variables, is discussed subsequently in Section 3.

The notation "var" for the variance matrix, and $r^* \equiv r(r+1)/2$ will be used. Given a square matrix H, the leading principal submatrix of H of order $t \times t$ will be denoted by $[H]_{t\times t}$. Finally, the column space generated by a matrix A, will be denoted as $\mathcal{M}(A)$.

2 Models and main results

The class of models to be considered in the present paper includes the following general latent-variable model:

$$\begin{cases} z &= \Lambda\eta + \varepsilon \\ \eta &= B\eta + \xi, \end{cases} \tag{6}$$

where z is a $p \times 1$ vector of observable variables, η is an $m \times 1$ vector of (possibly) latent variables, ε is a $p \times 1$ vector of measurement errors of zero mean, ξ is a random vector composed of disturbance terms of equations and exogenous variables and Λ and B are parameter matrices. The variable ξ is assumed to be of zero mean except (possibly) for its last component which may be identically equal to unity. It is assumed that ε and ξ are uncorrelated, i.e. $E\varepsilon\xi' = 0$, where E denotes mathematical expectation, and that the (uncentered) second-order moment matrices $\Psi \equiv E\varepsilon\varepsilon'(p \times p)$ and $\Phi \equiv E\xi\xi'(m \times m)$ are finite. This is a model that encompasses the class of so-called LISREL models (factor-analysis models, regression with errors in variables, and so forth). Note that (6) implies a specific moment structure, $\sigma = \sigma(\theta)$, for the vector $\sigma \equiv \mathrm{v}\,(Ezz')$ of (uncentered) second-order moments, where θ is a parameter vector that assembles the elements to be estimated in the matrices Λ, B, Φ and Ψ. For model (6), the first and second derivatives of $\sigma = \sigma(\theta)$ can be found in Neudecker and Satorra (1991).

Note that (6) can be writen as

$$z = \Lambda(I - B)^{-1}\xi + \varepsilon = [\Lambda(I - B)^{-1}, I][\xi', \varepsilon']', \tag{7}$$

That is, model (6) is a specific case of the following multivariate linear relation (e.g., Anderson, 1989)

$$z = \mu + \sum_{i=1}^{L} A_i \delta_i \tag{8}$$

where the δ_i's are mutually uncorrelated random (multivariate) variables of zero mean; that is

$$E\delta_i\delta_j' = 0, \text{ when } i \neq j, \tag{9}$$

and $\mu = Ez$ is the mean of z. Two cases will be considered in the present paper: the case where $z = (y', 1)'$ is an augmented vector (where y may have $Ey = 0$); and the case where where z is not an augmented vector but $Ez = 0$. This is not restrictive, since the remaining case, z is not an augmented vector and $Ez \neq 0$, can be reformulated to one of these.

The following assumption needs to be introduced.

ASSUMPTION A1:

a) $z = \mu(\tau) + \sum_{i=1}^{L} A_i(\tau)\delta_i$, as in (8), where $\mu(\tau)$ and the $A_i(\tau)$'s are continuously differentiable functions of a t-dimensional parameter τ;

b) the $q \times 1$ parameter vector

$$\theta = [\tau', \mathrm{v}'(K_{11}) \ldots, \mathrm{v}'(K_{ii}),' \ldots, \mathrm{v}'(K_{LL})]', \tag{10}$$

with $K_{ii} \equiv E\delta_i\delta_i'$, i=1,...,L, is unrestricted;

c) $\Sigma = Ezz'$ is a positive definite matrix. ■

Note that Assumption A1 b) implies that the moment matrices K_{ii} are unrestricted symmetric matrices. The Assumption A1 c) plays a crucial role in the derivations.

Typically, the analysis is based on a sample $z_1, \ldots, z_n$ of n independent observations of z and the data are summarized by the ($p^\star \times 1$) vector of sample moments $s \equiv \mathrm{v}\ (S)$, where

$$S \equiv \sum_{i=1}^{n} z_i z_i'/n \tag{11}$$

is the usual sample (uncentered) second-order moment matrix. Straightforward application of the central limit theorem shows that the vector of sample moments s is asymptotically normal with mean $\sigma \equiv \mathrm{v}(\mathrm{var}\ z)$ and asymptotic variance matrix

$$\Gamma \equiv \mathrm{avm}(\sqrt{n}\, s) = \mathrm{var}\ (\mathrm{v}\,(zz')), \tag{12}$$

which we assume to be finite. A consistent estimate of Γ can be obtained easily as follows. Let $d_i \equiv \mathrm{v}\,(z_i z_i'), i = 1, 2, \ldots, n$; thus, $s = \sum_{i=1}^{n} d_i/n$. Clearly, when z has finite eighth-order moments, the $(p^\star \times p^\star)$ matrix of sample fourth-order moments

$$\hat{\Gamma} = \sum_{i=1}^{n} (d_i - s)(d_i - s)'/(n-1) \tag{13}$$

is an unbiased and consistent estimator of Γ. Consequently, the MD analysis with Δ contained in the column space of Γ and weight matrix V_n equal to a generalized inverse of the matrix $\hat{\Gamma}$ will be AO . Further, substitution of $\hat{\Gamma}$ of (13) for Γ in (2) produces a consistent estimator of the avm $(\hat{\theta})$. The standard errors of parameter estimators obtained in this way will be called asymptotic robust standard errors, since they are valid (asymptotically) regardless of the distribution of the data. In the context of regression analysis, such asymptotic robust standard errors can be seen to be identical to the (heteroskedasticity-) robust standard errors proposed in White (1982).

In order to arrive at the main results of this paper, we will now obtain some implications of (8) and (9). Clearly, (8) implies

$$\mathrm{v}\,(zz') = D_z^+(\mu \otimes \mu) + \sum_{i=1}^{L} D_z^+(A_i \otimes A_i) D_{\delta_i} \mathrm{v}\,(\delta_i \delta_i') +$$

$$\sum_{\substack{1 \le i,j \le L \\ i \ne j}} D_z^+(A_j \otimes A_i) \mathrm{vec}\ \delta_i \delta_j' +$$

$$\sum_{i=1}^{L} D_z^+(\mu \otimes A_i)\delta_i + \sum_{i=1}^{L} D_z^+(A_i \otimes \mu)\delta_i;$$

hence,

$$\sigma \equiv E\text{v}\,(zz') = D_z^+(\mu \otimes \mu) + \sum_{i=1}^{L} D_z^+(A_i \otimes A_i)D_{\delta_i}\text{v}\,(K_{ii}), \tag{14}$$

where (9) was used. Combining the above equation (14) with Assumption A1 b), we obtain that the derivative matrix $\Delta \equiv (\partial/\partial\theta')\sigma(\theta)$ will be partitioned as

$$\begin{aligned}\Delta &= [\Delta_1, D^+(A_1 \otimes A_1)D, ..., D^+(A_i \otimes A_i)D, ..., D^+(A_L \otimes A_L)D] \\ &= [\Delta_1, \Delta_2],\end{aligned} \tag{15}$$

say, where $\Delta_1 \equiv (\partial/\partial\tau')\sigma(\theta)$ is a $p^\star \times t$ matrix.

In the sequel $\hat{\theta}$ and $\tilde{\theta}$ denote the MD estimators when V equals Ω^{-1} or Γ^-, respectively, where Ω^{-1} is given in (5), under the assumption that Δ is contained in the column space of Γ. Recall that any MD estimator with weight matrix V_n converging in probability to Ω^{-1} (alternatively, to Γ^-) will have the same asymptotic properties as $\hat{\theta}$ (alternatively, $\tilde{\theta}$). The corresponding estimator of the subvector τ of θ will be denoted by $\hat{\tau}$ (alternatively, $\tilde{\tau}$).

Before going into the Theorems of the paper we need to consider some preliminary results.

LEMMA 1. (cf. Satorra, 1992b) Let $z = \mu + \sum_{i=1}^{L} A_i\delta_i$ where the δ_i's are mutually *independent* and of zero mean. Then

$$\begin{aligned}\text{var}\,(\text{v}\,(zz')) = \bar{\Omega} + \sum_{i=1}^{L}\Big\{&2D^+(A_i \otimes \mu)\,[E\,\delta_i(\text{v}\;\delta_i\delta_i')']\;D'(A_i \otimes A_i)'D^{+\prime} + \\ &+2\,D^+(A_i \otimes A_i)\,D\;[E(\text{v}\;\delta_i\delta_i')\delta_i']\;(A_i \otimes \mu)'\,D^{+\prime} + \\ &+D^+(A_i \otimes A_i)\,D\,[\text{var}\,(\text{v}\,\delta_i\delta_i') - 2\,D^+E(\delta_i\delta_i') \otimes E(\delta_i\delta_i')D^{+\prime}]\;D'(A_i \otimes A_i)'D^{+\prime}\Big\},\end{aligned} \tag{16}$$

where

$$\bar{\Omega} = \Omega - 2D^+(\mu\mu' \otimes \mu\mu')D^{+\prime}, \tag{17}$$

$$\Omega = 2D^+(\Sigma \otimes \Sigma)D^{+\prime}, \tag{18}$$

$\mu = Ez$ and $\Sigma = Ezz'$.

REMARK 1. When y in $z = (y', 1)'$ is normally distributed, (16) above implies

$$\operatorname{var}(\mathrm{v}(zz')) = \bar{\Omega}. \tag{19}$$

LEMMA 2. When $z = (y', 1)'$, for general distribution of y, it holds
i)

$$\Gamma = \begin{pmatrix} \Gamma^\star & 0 \\ 0 & 0 \end{pmatrix} \tag{20}$$

and ii)

$$\bar{\Omega} = \begin{pmatrix} \bar{\Omega}^\star & 0 \\ 0 & 0 \end{pmatrix}, \tag{21}$$

where $\Gamma^\star$ and $\bar{\Omega}^\star$ are $(p^\star - 1) \times (p^\star - 1)$ positive semi-definite matrices;

iii) under Assumption A1 c), $\Gamma^\star$ and $\bar{\Omega}^\star$ are positive definite;

iv) under Assumption A1 b),

$$\Delta = \begin{bmatrix} \Delta^\star \\ 0 \end{bmatrix}, \tag{22}$$

where $\Delta^\star$ is a $(p^\star - 1) \times q$ matrix;

v) Under Assumption A1 c),

$$\mathcal{M}(\Delta) \subset \mathcal{M}(\Gamma), \tag{23}$$

and

$$\mathcal{M}(\Gamma) = \mathcal{M}(\bar{\Omega}), \tag{24}$$

and, consequently,

$$\mathcal{M}(\Delta) \subset \mathcal{M}(\bar{\Omega}); \tag{25}$$

vi) under Assumption A1 c) and Δ contained in the column space of Γ and $\bar{\Omega}$,

$$\Delta'\Gamma^{-}\Delta = \Delta^{\star\prime}\Gamma^{\star-1}\Delta^\star \tag{26}$$

and

$$\Delta'\bar{\Omega}^{-}\Delta = \Delta^{\star\prime}\bar{\Omega}^{\star-1}\Delta^\star \tag{27}$$

PROOF. The equality (20) follows trivially from the definition of Γ in (12) and the fact that when $z = (y', 1)'$ then the last component of the vector $\mathrm{v}(zz')$ is constant to unity. We will prove (21) by showing that $\bar{\Omega}$ is also a variance matrix of a vector whose last component is constant to unity. Effectively, defining $\tilde{z} \equiv (\tilde{y}', 1)'$, where $\tilde{y}$ is the normally distributed vector that has first- and second-order moments equal to those of y, and using Remark 1, we obtain

$$\text{var}\,(\text{v}\,(\tilde{z}\tilde{z}')) = \bar{\Omega}, \tag{28}$$

since $\mu = Ez = E\tilde{z}$ and $\Sigma = Ezz' = E\tilde{z}\tilde{z}'$; hence (21) is proved. Under Assumption A1 c), the above random vectors y and $\tilde{y}$ have a non-singular covariance matrix, hence result iii) follows. Result iv) follows trivially by differentiation of a constant (unity). Result v) is a trivial consquence of i) to iv). Finally, we prove vi) by noting that under (23) and (25) the left-hand side expressions of (26) and (27) are invariant under choice of g-inverses and

$$\begin{pmatrix} \Gamma^{\star-1} & 0 \\ 0 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} \bar{\Omega}^{\star-1} & 0 \\ 0 & 0 \end{pmatrix}$$

are g-inverses of Γ and $\bar{\Omega}$ respectively ■

When z is not an augmented vector, then results (23), (24) and (25) trivially hold.

REMARK 2. Note that (15) and (16) can be combined to write Γ in the form

$$\Gamma \equiv \bar{\Omega} + \Delta_2 C \Delta_2' + \Delta_2 B + B' \Delta_2', \tag{29}$$

where Δ_2 is the submatrix of Δ defined in (15), $C = \text{diag}(C_{11}, \ldots, C_{ii}, \ldots, C_{LL})$ and $B = (B_1', \ldots, B_i', \ldots, B_L')'$ with

$$C_{ii} = \text{var}\,(\text{v}\,\delta_i\delta_i') - 2\,D^+ E(\delta_i\delta_i') \otimes E(\delta_i\delta_i') D^{+\prime}$$

and

$$B_i = 2\,[E(\text{v}\;\delta_i\delta_i')\delta_i']\;(A_i \otimes \mu)'\,D^{+\prime};$$

further, (29) can be rewritten as

$$\Gamma \equiv \bar{\Omega} + \Delta J C J' \Delta' + \Delta J B + B' J' \Delta', \tag{30}$$

where $J \equiv [0 : I_{(q-t)}]'$ with $I_{(q-t)}$ an identity matrix of dimension $(q-t)$.

When $z = (y', 1)'$, then pre- and post- multiplying the equality (30) by the matrices $J^{\star\prime}$ and $J^\star$ respectively, where $J^\star \equiv [I_{(p^\star-1)} : 0]'$, we obtain

$$\Gamma^\star \equiv \bar{\Omega}^\star + \Delta^\star J C J' \Delta^{\star\prime} + \Delta^\star J B^\star + B^{\star\prime} J' \Delta^{\star\prime}, \tag{31}$$

where $B^\star \equiv BJ^\star$.

The matrices C and B will obviously vary with the type of non-normality of z and will simply vanish when y of $z = (y', 1)'$ is normally distributed. Note also that when $Ez = 0$, then B equals zero and $\bar{\Omega} = \Omega$.

LEMMA 3. Under Assumption A1, it holds that i) the MD estimator associated with $V = \bar{\Omega}^-$ has the same asymptotic variance matrix as the one associated with $V = \Omega^{-1}$; ii)

$$\Delta'\bar{\Omega}^-\Delta = \Delta'\Omega^{-1}\Delta; \tag{32}$$

and iii)

$$[(\Delta'\Gamma^-\Delta)^{-1}]_{t\times t} = [(\Delta'\bar{\Omega}^-\Delta)^{-1}]_{t\times t}. \tag{33}$$

PROOF. Result i) follows from the fact that Ω^{-1} is a g-inverse of $\bar{\Omega}$ (see Satorra & Neudecker, 1993), (24) and (25); ii) results from noting that when $\mathcal{M}(\Delta) \subset \mathcal{M}(\Omega)$ then $\Delta'\bar{\Omega}^-\Delta$ is invariant under the choice of g-inverse. We will now prove the result iii). When z is not an augmented vector, then $\bar{\Omega} = \Omega$ and $\bar{\Omega}$ and Γ are positive definite, hence

$$(\Delta'\Gamma^{-1}\Delta)^{-1} - (\Delta'\Omega^{-1}\Delta)^{-1} = (\Delta'\Gamma^{-1}\Delta)^{-1}(\Delta'\Omega^{-1}\Delta - \Delta'\Gamma^{-1}\Delta)(\Delta'\Omega^{-1}\Delta)^{-1}$$

$$= (\Delta'\Gamma^{-1}\Delta)^{-1}\Delta'\Gamma^{-1}(\Gamma - \Omega)\Omega^{-1}\Delta(\Delta'\Omega^{-1}\Delta)^{-1}$$

$$= (\Delta'\Gamma^{-1}\Delta)^{-1}\Delta'\Gamma^{-1}(\Delta JCJ'\Delta' + \Delta JB + B'J'\Delta')\Omega^{-1}\Delta(\Delta'\Omega^{-1}\Delta)^{-1}$$

$$= JCJ' + JB\Omega^{-1}\Delta(\Delta'\Omega^{-1}\Delta)^{-1} + (\Delta'\Gamma^{-1}\Delta)^{-1}\Delta'\Gamma^{-1}B'J'$$

where (30) was used; consequently,

$$[(\Delta'\Gamma^{-1}\Delta)^{-1}]_{t\times t} - [(\Delta'\Omega^{-1}\Delta)^{-1}]_{t\times t} = [(\Delta'\Gamma^{-1}\Delta)^{-1} - (\Delta'\Omega^{-1}\Delta)^{-1}]_{t\times t} = 0,$$

When z is an augmented vector, then we can use (26) and (27) and proceed similarly with (31) replacing (30). ∎

We are now in a position to state the main result of the paper.

THEOREM 1 Assume $z = \mu + \sum_{i=1}^{L} A_i\delta_i$ as in (8) and (9) with the δ_i's being *mutually independent*, and let Assumption A1 hold. Then the asymptotic variance matrices of $\hat{\tau}$ and $\tilde{\tau}$ are both equal to $[(\Delta'\Omega^{-1}\Delta)^{-1}]_{t\times t}$.

PROOF By Lemma 3 we know that the MD estimator associated with $V = \bar{\Omega}^-$ has the same asymptotic variance matrix as the one associated with $V = \Omega^{-1}$; hence, we can proceed evaluating the asymptotic variance matrix of the MD estimator associated with $V = \bar{\Omega}^{-}$. Using (24), (25) and (30), we obtain

$$[(\Delta'\bar{\Omega}^{-}\Delta)^{-1}\Delta'\bar{\Omega}^{-}\Gamma\bar{\Omega}^{-}\Delta(\Delta'\bar{\Omega}^{-}\Delta)^{-1}]_{t\times t} = [(\Delta'\bar{\Omega}^{-}\Delta)^{-1}\Delta'\bar{\Omega}^{-}\bar{\Omega}\bar{\Omega}^{-}\Delta(\Delta'\bar{\Omega}^{-}\Delta)^{-1}]_{t\times t}. \tag{34}$$

Using (25), and the definition of g-inverse, we get

$$(\Delta'\bar{\Omega}^{-}\Delta)^{-1}\Delta'\bar{\Omega}^{-}\bar{\Omega}\bar{\Omega}^{-}\Delta(\Delta'\bar{\Omega}^{-}\Delta)^{-1} = (\Delta'\bar{\Omega}^{-}\Delta)^{-1}. \tag{35}$$

The proof of Theorem 1 is concluded by combining (32), (33), (34) and (35). ■

Theorem 1 says that under certain conditions the NMD estimator $\hat{\tau}$ is asymptotically optimal within the class of MD estimators defined in (1).

REMARK 3 When some of the moment matrices K_{ii}, $i = 1, 2, \ldots, L$, of (10) are also restricted to be continuously differentiable functions of τ, then the conclusions of the theorem also hold if, in addition to the assumptions of the theorem, each δ_i with restricted moment matrix K_{ii} satisfies

$$\text{var}\,[\text{v}\,(\delta_i\delta_i')] = 2\,D^{+}[E(\delta_i\delta_i') \otimes E(\delta_i\delta_i')]D^{+\prime} \tag{36}$$

and

$$E\delta_i(\text{v}\,\delta_i\delta_i')' = 0. \tag{37}$$

The above remark results from noting that there will be a one-to-one correspondence between the submatrices of Δ_2 in (15) that drop out due to the restrictions on K_{ii}'s, and the terms on the rhs of (16) that vanish due to (36) and (37). Note that properties (36) and (37) will hold, of course, when δ_i is normally distributed.

There are situations where such asymptotic optimality of $\hat{\tau}$ carries over to the whole vector of estimators $\hat{\theta}$.

THEOREM 2 Assume the conditions of Theorem 1 hold and, additionally, B in the representation (29) of Γ equals zero. Then, the (*whole*) vectors of estimators $\hat{\theta}$ and $\bar{\theta}$ have the same asymptotic variance matrix. Further, the avm of $\hat{\tau}$ and $\bar{\tau}$ is equal to $[(\Delta'\Omega^{-1}\Delta)^{-1}]_{t\times t}$.

PROOF When B in (29) equals zero, we can write

$$\Gamma = \bar{\Omega} + \Delta_2 C\Delta_2' = \bar{\Omega} + \Delta JCJ'\Delta' \tag{38}$$

where $J = [0 : I_{(q-t)}]'$ with $I_{(q-t)}$ an identity matrix of dimension $q - t$; consequently, the avm of $\hat{\theta}$ of (2) will be transformed to

$$(\Delta'\bar{\Omega}^{-}\Delta)^{-1}\Delta'\bar{\Omega}^{-}\Gamma\bar{\Omega}^{-}\Delta(\Delta'\bar{\Omega}^{-}\Delta)^{-1} = (\Delta'\bar{\Omega}^{-}\Delta)^{-1} + JCJ'. \qquad (39)$$

The proof concludes by noting that, under the stated conditions,

$$(\Delta'\Gamma^{-}\Delta)^{-1} = (\Delta'\bar{\Omega}^{-}\Delta)^{-1} + JCJ' \qquad (40)$$

(cf., Neudecker and Satorra, 1991c) ∎

Theorem 2 above says that when B of (29) is zero then the NMD vector of estimators $\tilde{\theta}$ is AO within the class of MD estimators defined in (1). Note, however, that the standard errors provided by $(\Delta'\bar{\Omega}^{-}\Delta)^{-1}$ will be correct only for the subvector of estimates $\tilde{\tau}$. Correct standard errors of estimators for the other parameters can, of course, be estimated from (2) by replacing Γ by $\hat{\Gamma}$ of (13) (these latter estimates are the asymptotic robust standard errors mentioned above).

As can be seen from the inspection of (16), there are two situations where the matrix B of (29) is zero (and hence AO of the whole vector $\tilde{\theta}$ is attained); these are listed in Assumption A2.

ASSUMPTION A2 One of the following two conditions holds

a) the third-order moments of the δ_i's (i.e. the terms $E\,\delta_i(\mathrm{v}\ \delta_i\delta_i')'$) are zero;

b) $Ez = 0$ ∎

A typical case where Assumption A2 b) always holds is covariance-structure analysis, where the vector s in (1) is the vector of non-redundant variances and covariances of a vector of observable variables (equivalently, the (uncentered) second-order moments of a variable z for which $Ez = 0$). Hence, in covariance-structure analysis, if robustness results on AO apply, they will concern the *whole* NMD vector of parameter estimators, and not only a *subvector* τ of θ. In this case, however, AO will relate to the class of MD estimators defined by (1), with s being a vector of centered second-order moments only. When restricted to covariance-structure analysis, Theorems 1 and 2 above are in agreement with the work on asymptotic robustness referenced in Section 1. When $z = (y', 1)'$, however, then s in (1) is composed of first- and second-order moments and AO for the whole vector of estimators may not apply. Assumption A2 a) guarantees the AO of the whole vector of NMD estimators $\hat{\theta}$ within the class of MD estimators defined by (1) with s containing first- and second-order moments.

It has been shown elsewhere (e.g., Satorra, 1992) that when the vector of observable variables z is normally distributed the loglikelihood function is an affine transformation of $F(\theta) = F_{ML}(s, \sigma(\theta))$ where

$$F_{ML}(s, \sigma) \equiv \ln \mid \Sigma \mid + \mathrm{tr}\, S\Sigma^{-1} - \ln \mid S \mid - p; \tag{41}$$

hence, for a general type of distribution of z, the minimizer $\hat{\theta}$ of $F(\theta)$ is a (pseudo) maximum-likelihood (PML) estimator. In fact, it can be shown that an MD estimator with V_n as in (4) is (asymptotically) equivalent to the PML estimator[1]. Consequently, Theorem 1 and 2 above give also conditions for the AO of PML.

The typical case in practice is to use PML or MD estimation with V_n defined in (4), with the standard errors of the estimators being computed by using (3) with V_n substituted for Γ^-. The possibility that NMD and PML analysis retains the property of AO with regard to a subvector of estimates of τ, and that usual standard errors are correct, is of high practical relevance. Effectively, NMD and PML are the most typical analyses when we use standard software such as Jöreskog and Sörbom's LISREL (1989), Bentler's EQS (1989) or CALIS (a procedure of SAS). To carry out NMD or PML analyses in the mentioned computer programs, the augmented moment matrix S of (11) replaces the usual covariance matrix and the "ML" option of such programs is used. Theorem 1 implies that, under the stated conditions, for parameters different from variances and covariances of non-normal constituents of the model, the usual "ML" estimators are AO and the usual "ML" standard errors are correct. When the additional conditions of Theorem 2 hold, namely conditions a) or b) of Assumption A2, the property of asymptotic optimality carries over to the whole vector of parameter estimators.

The above general results will now be specialized to the context of multivariate regression type of models.

3 Multivariate regression models

Consider the following model of seemingly unrelated regressions (SURE) with stochastic regressors,

$$y = \alpha + Bx + \zeta, \tag{42}$$

[1]This follows from the fact that $2^{-1}\partial^2 F_{ML}/\partial s \partial s' \mid_{s=\sigma} = \Omega^{-1}$ (e.g, Neudecker and Satorra, 1991a); consequently, there is asymptotic equality between the minimizer of (41) and the MD estimator with $V = 2^{-1}\partial^2 F_{ML}/\partial s \partial s' \mid_{s=\sigma}$ (Shapiro, 1985; Newey, 1988).

where $y(p_1 \times 1)$ is a vector of responses, $B(p_1 \times p_2)$ is a matrix of regression coefficients, $x(p_2 \times 1)$ is a vector of regressors of mean μ_x and $\zeta(p_1 \times 1)$ is a vector of disturbance terms of mean zero. Consider the moment matrices $Exx' \equiv K_{xx}(p_2 \times p_2)$ and $E\zeta\zeta' \equiv \Psi(p_1 \times p_1)$. Denoting $z \equiv (y', x', 1)'$, we can write (42) as

$$z = \begin{pmatrix} y \\ x \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & B & \alpha \\ 0 & 0 & \mu_x \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ x \\ 1 \end{pmatrix} + \begin{pmatrix} \zeta \\ x - \mu_x \\ 1 \end{pmatrix} = A\delta, \tag{43}$$

where

$$\delta = (\zeta', x' - \mu_x', 1)', \tag{44}$$

$A = (I - \Upsilon)^{-1}$ and

$$\Upsilon = \begin{pmatrix} 0 & B & \alpha \\ 0 & 0 & \mu_x \\ 0 & 0 & 0 \end{pmatrix}, \tag{45}$$

with I and 0 denoting identity and zero matrices respectively of appropriate dimensions.

Let B, α and μ_x be continuously differentiable functions of a t-dimensional vector τ (a specific model, for example, is the one in which τ collects the distinct elements in B, α and μ_x). Assume that the moment matrices of x and ζ, K_{xx} and Ψ, respectively, are unrestricted. Define, finally, $\theta \equiv [\tau', (\mathrm{v}\, K_{xx})', (\mathrm{v}\, \Psi)']'$. Under this set-up, (8) , (9) and Assumption A1 hold with the role of the δ_i's being taken by $x - \mu_x$ and ζ. When ζ and x are stochastically independent, then Theorem 1 guarantees that PML or NMD estimators of τ are AO within the class of MD estimators defined by (1) with $s \equiv \mathrm{v}\,(S)$ and S defined in (11); further, they have a variance matrix which can be evaluated by (3) using a consistent estimator of Ω instead of Γ. Note that a consistent estimator of Ω is readily available from (18) by just substituting S for Σ.

Consider now a more general case, where ζ is partitioned as $\zeta = (\zeta_1', \zeta_2', ..., \zeta_T')'$, with $\Psi = \mathrm{diag}(\Psi_{11}, \Psi_{22}, \ldots, \Psi_{TT})$ conformable with the partition of ζ. Let $\delta = (\zeta', x' - \mu_x', 1)' = (\zeta_1', \zeta_2', \ldots, \zeta_T', x' - \mu_x', 1)'$. Suppose the model does not restrict the matrices Ψ_{ii}, nor the moment matrix K_{xx} of x. Under the assumption that $x, \zeta_1, \ldots, \zeta_T$ are mutually independent, the conditions of Theorem 1 are satisfied and hence PML and NMD are AO for estimating τ, within the class of MD estimates defined by (1). Note that now we allow for a multivariate regression where the disturbance term has a block diagonal covariance matrix. In the case that one or more covariance matrices Ψ_{ii} are restricted to be (continuously differentiable) functions of τ, then the results of optimality of PML and NMD hold also if each ζ_i with restricted variance matrix satisfies (36) and (37) (a condition which is obviously satisfied when the corresponding ζ_i is normally distributed).

Note that the classical simultaneous econometric model is encompassed by (43) if appropriate zero matrices in Υ of (45) are substituted for parameters to be estimated (taking care, of course, that the model is identified). Thus, the same conclusions of asymptotic robustness would apply in that case, namely the asymptotic efficiency of PML (which in this case is known as the classical "full-information maximum-likelihood" estimator) even under non-normality, when there is independence among x and ζ (or if there is mutual independence among $x, \zeta_1, \ldots, \zeta_T$, in the case of zero restrictions on the covariances of disturbance terms of equations).

The above arguments extend also to a SURE model with errors in variables. Consider (43) with $z = (y', x', 1)'$ replaced by the unobservable vector $z^\star \equiv (y^{\star\prime}, x^{\star\prime}, 1)'$ (where $z^\star$ may be of different dimension than z). Consider, additionally, the following measurement equation for z

$$z = \Lambda z^\star + \epsilon, \tag{46}$$

where Λ is a matrix of parameters restricted to be a function of τ (Λ could be for instance the identity matrix) and ϵ is the vector of measurement errors. In this set-up the model can be written as

$$z = [\Lambda(I - \Upsilon)^{-1}, I]((\zeta', x^{\star\prime} - \mu_{x^\star}', 1, \epsilon')', \tag{47}$$

which is of the form (8) with $\delta = (\zeta', x^{\star\prime} - \mu_{x^\star}', 1, \epsilon')'$. The above subvectors can, of course, be further partitioned. From Theorem 1, when the mixed fourth-order moments among the elements of the partition of δ are the same as with independence, and the variance matrix of non-normal elements of the partition is unrestricted, then NMD and PML analysis are optimal within the class of MD estimators defined by (1). Furthermore, the avm of $\bar{\tau}$ (corresponding to parameters distinct from the variances of non-normal elements of the partition of δ) are given by the same formula as when maximum likelihood is not misspecified. In all the cases discussed in this section, if the third-order moments of the δ_i's are zero then Theorem 2 applies and the asymptotic optimality extends to the whole vector of estimators.

Recall that the general model (6) can be written as (8), where $A \equiv [\Lambda(I - B)^{-1}, I]$ can be a function of τ and $\delta' \equiv [\xi', \epsilon']'$ (which can, of course, be partitioned further). This general model set-up encompasses the SURE models considered above as well as the simultaneous equation models with errors in variables. The results of Section 2 show that when there is *stochastic independence* among the (assumedly) uncorrelated elements of the partition of δ, and there are no restrictions on the moment matrices of each element of the partition, then PML and NMD give AO estimates of τ; further, the avm of estimators of τ can be computed with a consistent estimator of Ω replacing Γ.

a comprehensive measurement model (User's Guide). Mooresville, IN: Scientific Software.

Neudecker, H. & Satorra, A. 1991a. "Linear structural relations: gradient and hessian of the fitting function." *Statistics & Probability Letters* , *11:* 57-61.

Neudecker, H. , & Satorra, A. 1991c "Simple proof of a general matrix equality" *South African Statist. J.*, *25* 79-82

Newey, W.K. 1988. "Asymptotic Equivalence of Closest Moments and GMM Estimators." *Econometric Theory, 4:* 336-340.

Rao, C.R. & Mitra, S.K. 1971. *"Generalized Inverse of Matrices and its Applications."*. Wiley: New York.

SAS 1990. *Technical Report P-200. SAS/STAT Software: CALIS and LOGISTIC Procedures.* Cary, NC.: SAS Institute Inc.

Satorra, A. 1989. "Alternative test criteria in covariance structure analysis: a unified approach." *Psychometrika, 54:* 131-151.

Satorra, A 1992a. "Asymptotic robust inferences in the analysis of mean and covariance structures " in : *Sociological Methodology 1992,* pp. 249-278, P.V. Marsden (Edit.). Basil Blackwell: Oxford & Cambridge, MA

Satorra, A 1992b. "The variance matrix of sample second-order moments in multivariate linear relations ". *Statistics & Probability Letters, 15,* 63-69.

Satorra, A. & Bentler, P.M. 1990. "Model conditions for asymptotic robustness in the analysis of linear relations." *Computational Statistics & Data Analysis, 10:* 235-249.

Satorra, A. & Neudecker, H. 1993. "A Matrix Equality Applicable in the Analysis of Mean-and-Covariance Structures" Problem Section, *Econometric Theory, 8,* p. 581.

Schoenberg, R.J. 1989. *LINCS: Linear covariance structure analysis. User' s Guide.* Kent, WA: RJS Software.

Shapiro, A. 1985. "Asymptotic equivalence of minimum discrepancy function estimators to G.L.S. estimators." *South African Statistical Journal, 19:* 73-81.

Shapiro, A. 1986. "Asymptotic theory of overparameterized 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.

White, H. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica, 50:* 1-25.

References

Anderson, T.W. 1989. "Linear latent variable models and covariance structures." *Journal of Econometrics 41* 91-119.

Anderson, T.W. & Amemiya, Y. 1988. "The asymptotic normal distribution of estimators in factor analysis under general conditions." *The Annals of Statistics, 16:*759-771.

Bentler, P.M. 1983. "Simultaneous equation systems as moment structure models." *Journal of Econometrics, 22:* 13-42.

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

Bentler, P.M. & Dijkstra, T. 1985. "Efficient estimation via linearization in structural models" in Krishnaiah, P.R. (ed.) *Multivariate Analysis - VI* ,pp.9-42. Amsterdam: North Holland

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. 1990. "Asymptotic Robustness of Normal Theory Methods for the Analysis of Latent Curves." Pp. 211-225 in *Statistical Analysis of Measurement Errors and Applications,* edited by P. J. Brown and W.A. Fuller. Rhode Island: American Mathematical Society.

Browne, M. W. & Shapiro, A. 1988. "Robustness of normal theory methods in the analysis of linear latent variable models." *British Journal of Mathematical and Statistical Psychology, 41:* 193-208.

Chamberlain, G. 1982. "Multivariate regression models for panel data." *Journal of Econometrics 18:* 5-46

Chiang, C.L. 1956. "On regular best asymptotically normal estimates." *Annals of Mathematical Statistics, 27:* 336-351.

Fuller, W.A. 1987. *Measurement Error Models.* New York: John Wiley & Sons.

Jöreskog, K. & Sörbom, D. 1989. *LISREL 7 A Guide to the Program and Applications* (2nd ed.). Chicago: SPSS Inc

Magnus, J. R. & Neudecker, H. 1986. "Symmetry, 0-1 Matrices and Jacobians". *Econometric Theory, 2:* 157-190.

Magnus, J. R. & Neudecker, H. 1991. *Matrix differential calculus (2nd ed.).* Chichester: Wiley.

Mooijaart, A. & Bentler, P. M. 1991. "Robustness of normal theory statistics in structural equation models." *Statistica Neerlandica, 45:* 159-171.

Muthén, B. 1987. *LISCOMP: Analysis of linear structural equations with*

RECENT WORKING PAPERS

1. Albert Marcet and Ramon Marimon
Communication, Commitment and Growth. (June 1991)
[Published in *Journal of Economic Theory* (December 1992)]

2. Antoni Bosch
Economies of Scale, Location, Age and Sex Discrimination in Household Demand. (June 1991)
[Published in *European Economic Review* 35, (1991) 1589-1595]

3. Albert Satorra
Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures. (June 1991)
[Published in *Sociological Methodology* (1992), pp. 249-278, P.V. Marsden Edt. Basil Blackwell: Oxford & Cambridge, MA]

4. Javier Andrés and Jaume Garcia
Wage Determination in the Spanish Industry. (June 1991)

5. Albert Marcet
Solving Non-Linear Stochastic Models by Parameterizing Expectations: An Application to Asset Pricing with Production. (July 1991)

6. Albert Marcet
Simulation Analysis of Dynamic Stochastic Models: Applications to Theory and Estimation. (November 1991)

7. Xavier Calsamiglia and Alan Kirman
A Unique Informationally Efficient and Decentralized Mechanism with Fair Outcomes. (November 1991)
[Forthcoming in *Econometrica*]

8. Albert Satorra
The Variance Matrix of Sample Second-order Moments in Multivariate Linear Relations. (January 1992)
[Published in *Statistics & Probability Letters* Vol. 15, no. 1, (1992), pp. 63-69]

9. Teresa Garcia-Milà and Therese J. McGuire
Industrial Mix as a Factor in the Growth and Variability of States'Economies. (January 1992)
[Forthcoming in *Regional Science and Urban Economics*]

10. Walter Garcia-Fontes and Hugo Hopenhayn
Entry Restrictions and the Determination of Quality. (February 1992)

11. Guillem López and Adam Robert Wagstaff
Indicadores de Eficiencia en el Sector Hospitalario. (March 1992)
[Published in *Moneda y Crédito* Vol. 196]