

(WP)
HB
34
.U474
1991

UNIVERSITAT POMPEU FABRA 

*Asymptotic Robust Inferences in the Analysis of Mean
and Covariance Structures*

Albert Satorra

Economics working paper 3, June 1991



Asymptotic robust inferences in the analysis of mean and covariance structures

Albert Satorra *

Universitat Pompeu Fabra

revised version, June 1991

Abstract

Structural equation models are widely used in economic, social and behavioral studies to analyze linear interrelationships among variables, some of which may be unobservable or subject to measurement error. Alternative estimation methods that exploit different distributional assumptions are now available. The present paper deals with issues of asymptotic statistical inferences, such as the evaluation of standard errors of estimates and chi-square goodness-of-fit statistics, in the general context of mean and covariance structures. The emphasis is on drawing correct statistical inferences regardless of the distribution of the data and the method of estimation employed. A (distribution-free) consistent estimate of Γ , the matrix of asymptotic variances of the vector of sample second-order moments, will be used to compute robust standard errors and a robust chi-square goodness-of-fit statistic. Simple modifications of the usual estimate of Γ will also permit correct inferences in the case of multi-stage complex samples. We will also discuss the conditions under which, regardless of the distribution of the data, one can rely on the usual (non-robust) inferential statistics. Finally, a multivariate regression model with errors-in-variables will be used to illustrate, by means of simulated data, various theoretical aspects of the paper.

Key words: structural equation models, minimum distance, standard errors, chi-square goodness-of-fit, asymptotic distribution-free, stochastic independence, non-normality, complex samples.

*A preliminary version of this paper was presented at the "1990 International Conference on Measurement Errors in Surveys" (Tucson, Arizona, November 1990). The comments of Bengt Muthén and two anonymous referees on an earlier draft of this paper are greatly appreciated. Work supported by DGICYT grant PS89-0040.



1 Introduction

Structural equation models are widely used in economic, social and behavioral studies when studying the interrelationships among variables, some of which may be unobservable (latent) or subject to measurement error (see, e.g., Jöreskog 1981; Jöreskog and Sörbom, 1989; Bentler, 1983 and 1989; Muthén, 1987; Bollen, 1989, and references contained therein). Computer programs which implement various estimation methods in a general class of structural equation models are now available (e.g., LISREL of Jöreskog and Sörbom 1983; EQS of Bentler, 1989; LISCOMP of Muthén, 1987; LINCOS of Schoenberg, 1989; SAS (PROC CALIS), 1990, among others)

A general approach to inference in structural equation models consists of fitting structured population moments to sample moments, usually variances and covariances, using minimum distance (MD) methods (e.g., Chamberlain, 1982; Browne, 1974, 1984; Fuller, 1987, Section 4.2). The so called (pseudo) maximum likelihood (PML) approach, which uses normality as a working assumption, can also be seen to be asymptotically equivalent to the MD estimator associated with a normal theory (NT) weight matrix W_{NT} (Browne, 1974). For the precise definition of PML and NT-MD see Section 5 below.

Even though PML and NT-MD fitting functions are deduced from the assumption that the vector of observable variables is normally distributed, parameter estimates are consistent regardless of whether or not the normality assumption holds (i.e., consistency is a robust quality of PML and NT-MD parameter estimates). This robustness property, however, does not carry over to inferential statistics obtained under the normality assumption. That is, the usual standard errors of estimates and the so called "p-values" of test statistics associated with PML and NT-MD may be incorrect with non-normal data. Since in practice the distribution of the observable variables is often skewed with non-normal kurtosis, the lack of robustness of the, say, normal theory (NT) inferential statistics is of practical concern.

Given any type of distribution of the data, an asymptotically optimal MD (AO-MD) analysis is attained when W is the inverse (or generalized inverse) of the matrix of fourth-order sample moments of the data. In the context of covariance structure analysis, this is the so called "asymptotic distribution free" ("ADF") approach of Browne (1982, 1984), Bentler (1983) and Muthén (1989). For moderate size of models, however, fourth-order sample moments may be large in number and highly unstable in small samples, hence AO-MD methods may suffer from computational burden as well as lack of robustness against small samples (see Muthén and Kaplan, 1985 and 1990, for Monte Carlo evidence on the small sample size performance of "ADF" methods). In fact, such problems of AO-MD methods have contributed to the fact that PML or NT-MD are still very popular, even though for general types of distribution of the data they do not yield asymptotic optimality. On the other hand, the sensitivity of NT inferential statistics to non-normality suggests providing PML or NT-MD analysis with (asymptotic) robust standard errors of estimates and

a robust chi-square goodness-of-fit statistic.

The general theory for asymptotically correct standard errors of estimates in the context of MD estimation has been available for many years (Ferguson, 1958; Chiang, 1956); robust standard errors for mean and covariance structures have also been deduced in the context of PML estimation (White, 1982; Arminger and Sobel, 1990; Arminger and Schoenberg, 1989). Recently, the EQS computer program (Bentler, 1989) has made robust standard errors for NT-MD available in practice, though they are confined to covariance structures with unrestricted means; LISREL 7 (Jöreskog & Sörbom, 1989) also provide such standard errors but with an additional assumption of independence between sample means and sample variances and covariances. In fact, for PML estimation, robust standard errors for mean and covariance structures are available in the program LINCOS (Schoenberg, 1989; see also, Arminger and Schoenberg, 1989).

Traditional structural equation modeling has focused on models in which means are unrestricted, and hence only covariance structures needed to be considered. A notable exception to that, however, is Bentler (1983) and the recent work of Arminger and Schoenberg (1989) and Arminger and Sobel (1990). In fact, most theoretical developments of "asymptotic distribution free" methods have been undertaken only in the context of covariance structures (Browne, 1984). There are two distinct approaches to deal with models that restrict means and covariances. One approach considers the analysis of a moment vector that contains the means and covariances of the observable variables (e.g., Jöreskog and Sörbom, 1989; Bentler, 1989; Muthén, 1987, 1989). Another approach considers the analysis of the uncentered second-order moment matrix of the vector of observable variables augmented with a variable constant to one; that is, the analysis of an augmented moment matrix. The first approach requires modification of the usual fitting function used and hence of the conventional software for covariance structure analysis. Here we adopt the second approach which, as we will argue below, is conceptually simpler and has the major advantage that can be implemented without modifying the conventional software for covariance structure. Augmented moment matrices have been analyzed in the context of normality and maximum likelihood estimation, e.g., Jöreskog & Sörbom (1984) and the recent work of Meredith and Tisak (1990). Here the augmented moment matrix structure will be analyzed under general type of estimation methods and arbitrary distribution of the data.

The present paper addresses issues of correct asymptotic inferences in the analysis of mean and covariance structures. A simple expression for a distribution-free consistent estimate of Γ , the variance matrix of the vector of sample moments, will enable us to redefine AO methods for mean and covariance structures and to develop robust standard errors and a robust chi-square goodness-of-fit statistic associated to PML and NT-MD analyses. We propose a simple modification of the estimate of Γ to encompass the interesting case in practice of multi-stage complex samples. Further, conditions will be given under which, despite non normality, robust standard errors and a robust chi-square goodness-of-fit statistics are not required. The latter point

generalizes recent results on asymptotic robustness (Anderson, 1989; Browne and Shapiro, 1988; Satorra & Bentler, 1990, and Amemiya and Anderson 1990) which have so far been confined to the analysis of covariance structures with unrestricted means.

The structure of the paper is as follows. Section 2 presents the family of models to be dealt with. Section 3 develops asymptotic theory for (uncentered) second-order moment structures, summarizing the basic theory for MD estimation and developing a general type of chi-square goodness-of-fit statistic. Section 4 deals with the estimation of Γ and the AO-MD analysis. Section 5 presents robust standard errors and a robust chi-square goodness-of-fit statistic associated with PML and NT-MD. Section 6 deals with asymptotic robustness of normal-theory inferential statistics. Section 7 illustrates the theoretical developments of the paper using simulated data, and Section 8 concludes.

2 Linear relation models

We will deal with the following general latent-variable model:

$$\begin{cases} z &= \Lambda\eta + \varepsilon \\ \eta &= B\eta + \xi, \end{cases} \quad (1)$$

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, and ξ is a random vector composed of disturbance-terms of simultaneous equations and (possibly) unobservable exogenous variables. The parameter matrices Λ ($p \times m$) and B ($m \times m$), and the uncentered second-order moment matrices of ε and ξ , Ψ ($p \times p$) and Φ ($m \times m$) respectively, will be structured as continuously differentiable functions of a q -dimensional parameter vector, say θ . Without loss of generality, the matrix $(I - B)$ will be assumed to be invertible and the last component of the vectors ξ , η and z in equation (1) to be a variable constant to 1, say $c1$. This model encompasses factor analysis, multivariate regression with (or without) measurement error, and structural equation models with measurement error (i.e. the family of so called "LISREL" models). The gradient vector and Hessian matrix associated with model (1), for different fitting functions, is given in Neudecker and Satorra (1990).

The presence of the variable $c1$ will allow us to impose structure on the means as well as on the covariances of z . For such an approach of encompassing mean and covariance structures see, e.g., Jöreskog and Sörbom (1984) and Bentler (1989). In the computer program EQS (Bentler, 1989), $c1$ is called the "independent variable V999".

For the developments of our paper it is important to note that (1) can be rewritten

as

$$z = \Lambda(I - B)^{-1}\xi + \varepsilon = [\Lambda(I - B)^{-1}, I][\xi', \varepsilon']' = A\delta, \quad (2)$$

say, where $A := [\Lambda(I - B)^{-1}, I]$, I is an identity matrix of appropriate dimensions and $\delta' := [\xi', \varepsilon']$. That is, model (1) is a specific case of a linear structure

$$z = \sum_{i=1}^L A_i \delta_i, \quad (3)$$

where the δ_i 's are uncorrelated random variables, and the matrices A_i , $\Phi_{ii} := E\delta_i\delta_i'$, $i = 1, \dots, L$, are restricted possibly to being functions of θ . Here the δ_i 's will be assumed to be of zero mean except for δ_L , which is taken to be the variable $c1$; hence, (3) can be written as

$$z = \sum_{i=1}^{L-1} A_i \delta_i + \mu c1, \quad (4)$$

where $\mu := Ez$, the mean of z , is also allowed to be a function of θ . The (pseudo) variance of the (pseudo) variable $c1$ will be denoted as ϕ_c .

The linear structure (3), without the possibility of constraining the mean of z has been considered recently in different papers (e.g., Anderson, 1989; Browne and Shapiro, 1988; and Satorra and Bentler, 1990; for such general type of linear structures, see also Bentler, 1983).

As an illustration of the above model, consider the following multivariate regression model with measurement error

$$\begin{cases} Y_1 = \alpha + \beta x + \zeta_1 \\ Y_2 = \alpha + \beta x + \zeta_2 \\ X = x + u, \end{cases} \quad (5)$$

where α and β are parameters, and the ζ_1 's, the Y_i 's, X , x and u are scalar random variables. The variables ζ_1 , ζ_2 , x and u are assumed to be mutually uncorrelated. Since in model (5) the intercepts (and the slopes) of first and second equations are restricted to being equal, the model impose restrictions on the means of the observable variables.

Note that (5) is a special case of (1), where $z := (Y_1, Y_2, X, c1)'$, $\eta := (Y_1, Y_2, x, c1)'$, $\varepsilon = (0, 0, u, 0)'$ and $\xi := (\zeta_1, \zeta_2, x, c1)'$, $\Lambda := I$,

$$B := \begin{pmatrix} 0 & 0 & \beta & \alpha \\ 0 & 0 & \beta & \alpha \\ 0 & 0 & 0 & \mu_x \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\Phi := \begin{pmatrix} \phi_{11} & 0 & 0 & 0 \\ 0 & \phi_{22} & 0 & 0 \\ 0 & 0 & \phi_{33} & 0 \\ 0 & 0 & 0 & \phi_c \end{pmatrix}$$

and $\Psi := \text{diag}(0, 0, \psi, 0)$.

Note that in this example the variable $c1$ plays a crucial role since its inclusion enables us to structure the intercepts. In the econometric literature, the model (5) is known as seemingly unrelated regression (SUR) model with error in the variables.

3 Moment structure analysis: asymptotic inferences

The structural equation model (1) implies a moment-structure $\Sigma = \Sigma(\theta)$, where θ is the (q -dimensional) vector of parameters, for the population matrix of second-order moments of z

$$\Sigma := Ezz'. \quad (6)$$

Since the last component of z is $c1$, Σ contains means and uncentered second-order moments. In fact, writing $z' = (y', c1)'$, Σ is the so called augmented moment matrix of y .

Given a sample z_1, z_2, \dots, z_n of n independent observations of z , consider the following matrix of sample second-order moments:

$$S = \sum_{i=1}^n (z_i z_i') / n. \quad (7)$$

The MD estimator $\hat{\theta}$ of θ is defined as the minimizer of ¹

$$F = (s - \sigma(\theta))' \hat{W} (s - \sigma(\theta)), \quad (8)$$

where $s = \text{vech}S$ and $\sigma = \text{vech}\Sigma$ are the (reduced) vectors of sample and population moments, respectively, and \hat{W} is a matrix converging in probability to W , a positive definite matrix. Here $\text{vech}S$ is the column vector formed stacking the $p^* := p(p+1)/2$ different elements of S . It holds that $\text{vec}(S) = D\text{vech}(S)$, where $\text{vec}(\cdot)$ denotes the usual column vectorization of a matrix, and D is the 0-1 duplication matrix; we can also write, $\text{vech}(S) = D^+ \text{vec}(S)$, where $D^+ := (D'D)^{-1}D$ is a g-inverse of D (see, e.g., Magnus and Neudecker, 1988). Generally, a fitting function $F = F(S, \Sigma(\theta))$ where $F = F(\cdot, \cdot)$ is non-negative, continuous in both arguments and zero when $S = \Sigma$ could be used (e.g., Browne, 1984).

To include restrictions on the means, some computer programs (e.g., LISREL, EQS and LISCOMP) recently incorporated a fitting function that is the sum of two parts: one corresponding to the fit of means and the other to the fit of variances and covariances. This approach implies to introduce basic modifications on conventional software for covariance structure analysis. In contrast to that approach, we adopt

¹ typical regularity assumptions, as e.g. θ identifiable, will be assumed in order that $\hat{\theta}$ is well defined, consistent and asymptotically normal. For a set of general regularity conditions see Satorra (1989).

the simpler framework of fitting the matrix of sample moments S defined above (see (7)) to the corresponding matrix Σ of population moments defined in (6) and which is structured as a function of θ through the model representation (1). The structure on the means will be imposed through $c1$. By adopting this approach we can analyze mean and covariance structures using, without modification, conventional software for covariance structures with S taking the role of the "covariance matrix" to be analyzed.

The asymptotic variance matrix of estimates and test statistics will now be obtained. Under fairly general conditions, it holds that

$$\sqrt{n}(s - \sigma) \rightarrow_L N(0, \Gamma), \quad (9)$$

where \rightarrow_L denotes convergence in distribution, and $N(0, \Gamma)$ denotes a normal distribution of zero mean and variance matrix Γ , a finite $p^* \times p^*$ matrix (recall that $p^* = p(p+1)/2$). From (9), and typical regularity conditions, it follows that the expression for the matrix of asymptotic variances of $\hat{\theta}$ will be (e.g., Satorra, 1989):

$$avar(\hat{\theta}) = n^{-1}(\Delta'W\Delta)^{-1}\Delta'W\Gamma W\Delta(\Delta'W\Delta)^{-1}, \quad (10)$$

where $\Delta := (\partial/\partial\theta')\sigma(\theta)$, a $p^* \times q$ matrix.

When W and Γ are such that $\Delta'W\Gamma W\Delta = \Delta'W\Delta$ then, obviously, (10) reduces to

$$avar(\hat{\theta}) = n^{-1}(\Delta'W\Delta)^{-1}, \quad (11)$$

and, in that case, the corresponding fitting function is said to be asymptotically optimal for the given model and distribution of the data (Satorra, 1989).

A chi-square goodness-of-fit statistic based on the residuals of the fit of S to $\Sigma = \Sigma(\theta)$ can also be developed. Let $\hat{\sigma} = vech\Sigma(\hat{\theta})$ and Δ_{\perp} be an orthogonal complement of Δ (i.e. Δ_{\perp} is a $p^* \times (p^* - q)$ matrix of full column rank such that $\Delta'_{\perp}\Delta = 0$). It can easily be shown that the "residual" vector $\sqrt{n}(s - \hat{\sigma})$ has an asymptotic normal distribution with asymptotic variance matrix

$$avar(\sqrt{n}(s - \hat{\sigma})) = (I - (\Delta'W\Delta)^{-1}\Delta'W)\Gamma(I - (\Delta'W\Delta)^{-1}\Delta'W)'; \quad (12)$$

hence, the following goodness-of-fit statistic

$$T = n(s - \hat{\sigma})'\hat{A}(s - \hat{\sigma}), \quad (13)$$

where \hat{A} is a consistent estimate of

$$\Delta_{\perp}(\Delta'_{\perp}\Gamma\Delta_{\perp})^{-}\Delta'_{\perp}, \quad (14)$$

and " $-$ " denotes g-inverse, is asymptotically chi-square distributed with degrees of freedom equal to $r := rank(\Delta_{\perp}(\Delta'_{\perp}\Gamma\Delta_{\perp})^{-}\Delta'_{\perp})$. This is the generalization to the analysis of moment structures, with Γ non-singular, of the goodness-of-fit statistic developed by Browne (1984) in the context of covariance structure analysis.

A more typical version of the goodness-of-fit statistic is n times the fitting function at its minimum, i.e. $nF(\hat{\theta})$. It can be shown (Satorra, 1989) that when the fitting function is asymptotically optimal, then $nF(\hat{\theta})$ is asymptotically equal to T of (13); however, when asymptotic optimality does not hold, then $nF(\hat{\theta})$ will in general not be asymptotically chi-square. Likelihood ratio, score and Wald type test statistics for testing a specific set of restrictions can also be developed in line with the arguments of Satorra (1989).

A scaled (adjusted for mean) goodness-of-fit statistic

$$ST = nF(\hat{\theta})/\kappa, \quad (15)$$

where

$$\kappa := \text{tr}[(W - W\Delta(\Delta'W\Delta)^{-1}\Delta'W)\Gamma]/r = (\text{tr}[(\Delta_{\perp}'W^{-1}\Delta_{\perp})^{-1}\Delta_{\perp}'\Gamma\Delta_{\perp}])/r \quad (16)$$

has also been proposed in covariance structure analysis (cf. Satorra and Bentler, 1988) on the basis that it would improve the chi-square approximation under a general type of distribution of z . This is the so-called "Satorra-Bentler chi-square statistic" implemented in EQS (Bentler, 1990). An adjusted (adjusted for mean and variance) chi-square goodness of fit statistic could also be considered (see Satorra and Bentler, 1988).

4 Consistent estimation of Γ and AO-MD analysis

The asymptotic variance matrix Γ of the vector of sample moments plays a fundamental role in assessing the sampling variability of statistics of interest (i.e. in drawing correct statistical inferences) and also in defining the optimal MD analysis. For general type of distributions of the data, Γ involves the fourth-order moments of the observable variables. When the vector of observable variables is normally distributed then Γ expresses as a function of only second-order moments.

In order to estimate the asymptotic variance matrix of estimates as well as to compute the goodness-of-fit statistic, an estimate of the matrix Γ is required. As show below, for general type of distributions of the observable variables, and also in the case of complex samples, such a consistent estimator of Γ is readily available by standard theory.

Define $d_i := \text{vech}(z_i z_i'), i = 1, 2, \dots, n$; hence, $s = \sum_{i=1}^n d_i/n$. Since the d_i 's are uncorrelated with each other, an unbiased estimate of the variance matrix of $\sqrt{n}s = \sum_{i=1}^n d_i/\sqrt{n}$ will be the following ($p^* \times p^*$) matrix of fourth-order sample moments (see Result 1 of Appendix A):

$$\hat{\Gamma} = \sum_{i=1}^n (d_i - s)(d_i - s)'/(n - 1). \quad (17)$$

In the case of complex samples, a similar type of estimate of Γ can also be developed. Consider a population divided into H strata ($h = 1, 2, \dots, H$) within each of which I_h primary sample units (PSU) are randomly chosen (with replacement). Within each PSU further stages of sampling may be undertaken (consider, e.g., two further levels of sampling). Define

$$d_{hi} := \sum_{tc} \text{vech}(z_{hict} z'_{hict}), \quad (18)$$

where z_{hict} is the vector value associated with the t -th third-stage unit of c -th second-stage unit of i -th PSU of stratum h , with the summation going over all the units within the i -th PSU (of course, further levels of sampling could be considered by adding more subscripts, besides t and c).

Since within stratum h the d_{hi} 's ($i = 1, \dots, l_h$) are uncorrelated, by standard results (see Result 1 of Appendix A) a consistent estimate of matrix Γ will be (cf., Skinner, Holt and Smith, 1989, p.48)

$$\hat{\Gamma} = \sum_{h=1}^H \left[\sum_{i=1}^{l_h} (I_h / (I_h - 1)) (d_{hi} - s_h)(d_{hi} - s_h)' \right] / n, \quad (19)$$

where $s_h = \sum_{i=1}^{l_h} d_{hi} / I_h$, and n is the total sample size (total number of last-stage sample units over all strata). Note that when $H = 1$ and $l_h = n$, that is when there is only one strata and each PSU is a final sample unit, then (19) reduces to (17).

Note that since the last column of z is variable $c1$, $\hat{\Gamma}$ will be a singular matrix and will partition as

$$\begin{pmatrix} \hat{\Gamma}^* & 0 \\ 0 & 0 \end{pmatrix},$$

where $\hat{\Gamma}^*$ is a matrix of dimensions $(p^* - 1) \times (p^* - 1)$ and 0 denotes a zero matrix of appropriate dimensions.

An AO-MD analysis will be attained by the use of the following weight matrix

$$\hat{W}_{AO} = \begin{pmatrix} \hat{\Gamma}^{*-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad (20)$$

since then the probability limit of \hat{W}_{AO} , say W_{AO} , will obviously satisfy the asymptotic optimality condition of $W_{AO} \Gamma W_{AO} = W_{AO}$. Furthermore, the use of (20) will yield a goodness-of-fit statistic T of (13) numerically equal to $nF(\hat{\theta})$; under the null hypothesis that the model holds, T (or $nF(\hat{\theta})$) will be both asymptotically chi-square with degrees of freedom given by² (Satorra, 1989, Theorem 4.2)

$$\text{rank}(W) - q = p(p+1)/2 - q - 1. \quad (21)$$

²To ensure a unique minimizer of (8) the "pseudo" variance ϕ_c should be "fixed" to 1, i.e. ϕ_c should not be a free parameter of the model.

Note that the above analysis is asymptotically efficient within the class of MD fitting functions (8), but it involves the inversion of a matrix of fourth-order sample moments. Such inversion may turn out to be computationally expensive, or inaccurate, or it may even not exist due to a small sample size (or a small number of PSU per strata). Specifically, $\hat{\Gamma}^*$ will not be invertible when n , or the number of PSU's, is lower than $(p^* - 1)p^*/2$. One analysis that may not be asymptotically efficient, but which is computationally much more feasible than the AO method described above, consists of using PML or NT-MD together with robust standard errors and a robust goodness-of-fit test statistic.

5 PML and NT-MD analysis

Let $z' = (y', c1)'$, then under the assumption that y is normally distributed it can be shown that the log likelihood function is an affine transformation of (see Appendix B)

$$F_{ML} = \ln |\Sigma| + \text{tr} S \Sigma^{-1} - \ln |S| - p, \quad (22)$$

such that the minimization of $F_{ML} = F(S, \Sigma(\theta))$ gives maximum likelihood estimation. The use of F_{ML} when the normality assumption does not necessarily holds will be called pseudo maximum likelihood (PML) analysis.

It can be seen that the Hessian matrix $\partial^2 F_{ML} / \partial \sigma \partial \sigma'$ evaluated at (Σ, Σ) is equal to (see, e.g., Neudecker and Satorra, 1991)

$$W = W_{NT} := (1/2) D' (\Sigma^{-1} \otimes \Sigma^{-1}) D. \quad (23)$$

Since the asymptotic properties of statistics of interest are characterized by this Hessian matrix (Shapiro, 1985; Satorra, 1989; Newey, 1988), PML will be asymptotically equivalent to MD analysis with weight matrix W as given by (23). By NT-MD analysis it will be understood the use of a MD fitting function (8) with the asymptotic limit of \hat{W} equal to W of (23). Obviously a specific choice for \hat{W} is the matrix of (23) with S substituting for Σ . In the case of covariance structure analysis, this equivalence between PML and NT-MD was first proven by Browne (1974).

It should be noted that PML and NT-MD analyses as defined above are available in conventional software for covariance structure analysis. For example, in the computer program LISREL (Jöreskog and Sörbom, 1984, 1989) the minimization of F_{ML} of (22) will be invoked when one specifies "ML" as the estimation method and S as the "covariance matrix" to be analyzed.

The asymptotic efficiency of maximum likelihood analysis (e.g., Cox and Hinkley, 1974) and the fact that when y is normally distributed then PML is maximum likelihood, guarantees the asymptotic optimality of PML in case of normality. The above mentioned equivalence between PML and NT-MD imply that such asymptotic efficiency extends also to NT-MD analysis. In fact when y is normally distributed,

this asymptotic optimality of NT-MD follows also directly from result iv) of Lemma 2 of Appendix B.

When y is normally distributed, it follows from Lemma 1 of Appendix B that the asymptotic variance matrix of s is given by

$$\Gamma = \Gamma_{NT} := \Omega - 2D^+(\mu\mu' \otimes \mu\mu')D^{+'}, \quad (24)$$

where

$$\Omega := 2D^+(\Sigma \otimes \Sigma)D^{+'}(= W_{NT}^{-1}). \quad (25)$$

Obviously a consistent estimate of Ω will be obtained replacing S for Σ in (25). In this case, since it holds that (see Lemma 1 of Appendix B)

$$\Delta'W_{NT}\Gamma_{NT}W_{NT}\Delta = \Delta'W_{NT}\Delta, \quad (26)$$

the general expression (10) of $\text{avar}(\hat{\theta})$ will simplify to (11). Further, when y is normally distributed and Δ belongs to the space generated by Γ_{NT} , then it can easily be seen that

$$\text{avar} \sqrt{n}(s - \hat{\sigma}) = (\Gamma_{NT} - \Delta(\Delta'W_{NT}\Delta)^{-1}\Delta'), \quad (27)$$

hence we can write the chi-square goodness-of-fit statistic of (13) as (this follows from i) and iii) of Lemma 1 of Appendix B)

$$\begin{aligned} T = n(s - \hat{\sigma})'[W_{NT} - W_{NT}\Delta(\Delta'W_{NT}\Delta)^{-1}\Delta'W_{NT}](s - \hat{\sigma}) = \\ n(s - \hat{\sigma})'\Delta_{\perp}(\Delta'_{\perp}\Omega\Delta_{\perp})^{-1}\Delta'_{\perp}(s - \hat{\sigma}), \end{aligned} \quad (28)$$

where, in a specific analysis, obvious consistent estimates would replace population values. Since T above is asymptotically equivalent to $nF(\hat{\theta})$ (e.g., Satorra, 1989), the latter will also be asymptotically chi-square when the model holds. That is, the conventional standard errors and goodness-of-fit statistics obtained by a computer program for the analysis of covariance structures when S is analyzed as a "covariance matrix" will be correct when y is normally distributed.

For general type of distributions of z , i.e. when y is not normally distributed, it follows from the general theory of Sections 3 and 4 above that asymptotic robust (i.e., correct for any distribution of y) standard errors and an asymptotic robust chi-square statistic are obtained substituting $\hat{\Gamma}$ of (17) or (19) for Γ in (10) and (13) above³. This approach produces what we will call (asymptotic) robust standard errors of estimates and robust chi-square goodness-of-fit statistic.

With regard to computational aspects of such robust standard errors and the chi-square goodness-of-fit statistic, substituting $\hat{\Gamma}$ of (17) for Γ in (10), the following

³It can be noted that by specifying the pseudo variance ϕ_c as "free" parameter of the model, the use of a generalized inverse can be avoided in (14)

variance matrix of estimates is obtained:

$$avar(\hat{\theta}) = n^{-1}(\Delta' \hat{W} \Delta)^{-1} \left[\sum_{\alpha=1}^n (t_{\alpha})(t_{\alpha})' / (n-1) \right] (\Delta' \hat{W} \Delta)^{-1}, \quad (29)$$

where

$$t_{\alpha} := \Delta' \hat{W}(d_{\alpha} - s), \quad (30)$$

and the derivative matrix Δ is evaluated at $\hat{\theta}$. Note that this step of computing standard errors will require a second pass through the data in order to compute the q -dimensional t_{α} 's vectors. Note further that the j th element of t_{α} , $j = 1, 2, \dots, q$, can be expressed as

$$(t_{\alpha})_j = (\Delta' \hat{W}(d_{\alpha} - s))_j = 2^{-1} tr(\hat{\Sigma}^{-1}(z_{\alpha} z_{\alpha}' - S) \hat{\Sigma}^{-1}(\partial \Sigma(\hat{\theta}) / \partial \theta_j)). \quad (31)$$

That is, the consistent asymptotic variance matrix estimate given by the expression (29) above parallels Arminger and Schoenberg's (1989) computation of robust standard errors in line with what they say of "[...] the asymptotic covariance matrix can be estimated consistently without computing the empirical fourth order moment matrix of the data" (p. 410) (See also formulae (24) on p.414 of the mentioned paper). It is not clear to us, however, that such an approach is computationally faster than just using the expression (9) of Section 3 with consistent estimates replacing population matrices, and the sample matrices (17) or (19) being used as a consistent estimates of Γ .

Also an alternative way to compute the test statistic T of (13) is as follows. Let $\hat{\Delta}'_{\perp}$ denote the orthogonal complement of Δ evaluated at $\hat{\theta}$, then in (10a) one could use

$$\hat{A} = \hat{\Delta}'_{\perp} \left[\sum_{\alpha=1}^n (b_{\alpha})(b_{\alpha})' / (n-1) \right]^{-1} \hat{\Delta}'_{\perp}, \quad (32)$$

where $b_{\alpha} := \hat{\Delta}'_{\perp}(d_{\alpha} - s)$. Note that the b_{α} 's are also of reduced dimension (r -dimensional vectors), hence only the inversion of a matrix of reduced dimension ($r \times r$) is required.

Finally, a consistent estimate of the scaling correction κ of (15) can be easily seen to be

$$\hat{\kappa} = \left[\sum_{\alpha=1}^n b'_{\alpha} (\hat{\Delta}'_{\perp} \hat{W}^{-1} \hat{\Delta}'_{\perp})^{-1} b_{\alpha} \right] / nr. \quad (33)$$

In fact, $u_{\alpha} = b'_{\alpha} (\hat{\Delta}'_{\perp} \hat{W}^{-1} \hat{\Delta}'_{\perp})^{-1} b_{\alpha}$ could be interpreted as the "influence" of case α on the departure of $nF(\hat{\theta})$ from its chi-square distribution (the influence being nil when $u_{\alpha} = r$).

The next section will show that there are situations where despite the fact that y is non-normally distributed, usual inferences based on the assumption that y is normally distributed can still be trusted.

6 Asymptotic robustness of inferences based on second-order moments

The results of this section are summarized by the following theorem.

Theorem. 1 Assume z decomposes as in (3) and additionally i) the δ_i 's are mutually independent (not only uncorrelated), ii) the unconstrained parameter vector θ partitions as $\theta = [\tau', \omega']'$, where ω is the vector formed with the distinct elements of the moment matrices Φ_{ii} 's associated with non-normally distributed δ_i 's, and iii) the matrices $A_i, i = 1, 2, \dots, L$, and Φ_{ii} 's associated with normally distributed δ_i 's, are continuously differentiable functions of τ . Then, for any choice of weight matrix W , it holds that

(a) The goodness-of-fit statistic T of (13) with Ω of (25) substituting for Γ is asymptotically chi-square when the model holds.

(b) The expression of $\text{avar}(\hat{\theta})$ of (10) with Ω substituting for Γ yields correct variances and covariances of estimates of the subvector of parameters τ .

Proof. See Appendix C.

In Section 5 it was shown that when $z = (y', c1)'$ and y is normally distributed, the use of S as the "covariance matrix" in conventional software for covariance structure analysis produced correct asymptotic inferences. The theorem above shows that such correctness of normal theory inferences extends also to the case of y non-normally distributed, provided certain model conditions hold and attention restricts to the estimate of the subvector of parameters τ . The theorem shows also the validity of the usual normal theory chi-square goodness-of-fit when the normality assumption is violated. A fundamental assumption turns out to be the independence, and not only uncorrelation, between the basic random constituents of the model. Restricted to covariance structure analysis and PML (or NT-MD) estimation, results (a) and (b) were derived by Anderson (1987, 1989) and Anderson and Amemiya (1988), Browne and Shapiro (1988) and Mooijart and Bentler (1990). For general type of discrepancy functions (as PML and MD with any choice of W), but confined also to the context of covariance structure analysis, (a) and (b) were derived by Satorra and Bentler (1990).

Note that Theorem above applies to MD analysis with any choice of weight matrix W . For example, W could in fact be the identity matrix, as in the so called "unweighted least squares" (ULS) analysis. In this example of ULS, theorem above says that the variance matrix of the estimate of τ , as well as the chi-square goodness-of-fit statistic, obtained under the "normality" assumption, i.e. using an estimate of Ω instead of the (distribution free) consistent estimate of Γ , will be valid for general type of distributions of y when the conditions of the theorem hold⁴. In fact the current version of LISREL provides such normal theory standard errors for ULS

⁴Note that here for the conditions of the theorem to hold, the pseudo parameter ϕ_c should be a free parameter of the model

estimates, which will then be correct even for non-normal data when the conditions of the theorem hold.

The next section will illustrate some of the theoretical aspects of the paper using a specific model context and simulated data.

7 Illustration

To illustrate the performance in finite samples of the above asymptotic results a specific model with simulated data was considered. Let model (5) with the parameter values shown in the first column of Table 1. Recall that model (5) assumes that variables ζ_1, ζ_2, x and u are uncorrelated. Note that since intercepts are restricted to be equal, model (5) will imply restrictions on the means of the observable variables.

Two situations dealing with non-normal data will be considered. First, Case I where the data are non-normal and the assumption of independence between ζ_1, ζ_2, x and u cannot be assumed. In this case the AO methods of Section 4 above will be required for an efficient statistical inference, or when using the generally non-optimal PML or NT-MD analyses, robust standard errors and a robust goodness-of-fit statistic will be required for correct asymptotic inferences. Second, Case II, where the data are non-normal but the conditions for asymptotic robustness, as the independence between ζ_1, ζ_2, x and u , hold and consequently usual PML (or NT-MD) analysis gives a correct chi-square goodness-of-fit statistic and correct standard errors for some of the parameters of interest. Finally, we consider Case III of a complex sample situation where the observations are clustered in groups with high intraclass correlation (i.e. the cases are not independent).

The Monte Carlo study consisted on replicating a number of times the generation of a sample of size n from model (5), with population values of the parameters as shown in the first column of Table 1. Summary statistics of the Monte Carlo results are presented in the corresponding tables. The NT-MD analysis used W as in (23), with S substituting for Σ . The robust standard errors and the robust chi-square statistic were computed as explained above in Section 3.

Table 1 shows a summary of the Monte Carlo results corresponding to Case I, where non-normal data was generated with uncorrelation, but lack of independence, between ζ_1, ζ_2, x and u . In fact, we have generated ζ_1, ζ_2 and u to be heteroskedastic with variance changing with x . The results show clearly that with this type of data robust standard errors and a robust chi-square goodness-of-fit statistic are required for NT-MD analysis. The performance of the scaled goodness-of-fit statistic (sbchi2, in the table) is also acceptable.

Table 1 about here

Table 2 shows the results corresponding to Case II, where variables x , the ζ_i 's and u were chosen to be mutually independent centered chi-square distributions with one degree of freedom (rescaled to have variance 1). Since in this case the basic random constituents of the model are independent, Theorem 1 above guarantees that the usual normal theory chi-square can be trusted and that the usual standard errors of estimates of β, α and μ will be asymptotically correct. However, the usual normal theory standard errors of estimates of ϕ 's and ψ 's are non necessarily correct. This is reflected by the numbers shown in the table.

Table 2 and 3 about here

A situation where the cases are not independent is considered in Case III. The data were generated such that cases were clustered in 20 groups of size 30 with high intraclass correlation (.8). This is a case of complex sample where the specific estimate (19) of Γ , which takes into account the clustered structure of the cases, is required. Note that here the number of groups is small, hence the AO analysis of Section 4 above would have to face a singular matrix $\hat{\Gamma}^*$. The results of the simulation are described in Table 3. Clearly, robust standard errors as well as the robust chi-square statistics are required in this case.

8 Conclusions

We have discussed general approaches to inference for structural equation models that impose restrictions on the means and covariances of the vector z of observable variables. The emphasis has been on drawing correct statistical inferences regardless of the distribution of z and the estimation method used. Asymptotic robust standard errors and a (asymptotic) robust chi-square goodness-of-fit statistic has been derived for PML and MD analyses, encompassing the case of multi-stage complex samples. The asymptotically optimal MD analysis has been reviewed, and we argued that, even for moderate size models, such AO analysis may be computationally costly and lack robustness against small sample size.

Recent results on asymptotic robustness of normal theory methods (e.g. Satorra and Bentler, 1990) have been extended to the context of mean and covariance structure analysis. Here the correct expression for Γ is replaced by a matrix Ω that is a function of second-order moments and does not equal Γ even under normality. When the usual assumption of uncorrelation among random constituents of the model is replaced by the stronger assumption of independence (and no restrictions are imposed on the variances and covariances of non-normally distributed constituents of the model), then robust inferential statistics are not required.

In practice, the deviation between the usual (non-robust) and robust standard errors,

and between the robust and non-robust chi-square goodness-of-fit statistic, should be taken as an indication that the assumption of independence does not hold and, in such a case, the robust inferential statistics will be the ones to be trusted. Such a discrepancy between robust and non-robust inferential statistics is to be expected, for example, when the variance of the error terms varies across cases (i.e., under heteroskedasticity), a situation in which robust inferential statistics are certainly required. In fact, a general recommendation would be to use PML or NT-MD together with robust inferential statistics. To follow this recommendation would require minimal modifications of the software at present in use.

Finally, a SUR model with error-in-variables has served to illustrate the type of models being considered and the performance in finite samples of asymptotic results of the paper.

References

- AMEMIYA, Y. & ANDERSON (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, 3, 1453-1463.
- ANDERSON, T. W. (1987). Multivariate linear relations. In T. Pukkila & S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics*. p. 9-36. Tampere, Finland.
- 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.
- ARMINGER, G. , & SOBEL, E.M. (1990). Pseudo-maximum likelihood estimation of Mean and covariance structures with missing data, *Journal of the American Statistical Association*, 85, 195-203.
- ARMINGER, G., & SHOENBERG, R.J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika*, 54, 409-425.
- 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. BMDP Statistical Software, Inc. : Los Angeles.
- 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.
- BOLLEN, K.A. (1989). *Structural equations with latent variables*. Wiley & Sons: New York.
- 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. , & 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. *J Annals of Mathematical Statistics*, 27, 336-351.
- COX, D. AND HINKLEY, D. V. (1974). *Theoretical statistics*. London: Chapman & Hall.
- FERGUSON, T. (1959). A method of generating best asymptotically normal es-

- timates with application to the estimation of bacterial densities, *Annals of Mathematical Statistics* 29: 1046-1062.
- JÖRESKOG, K., & SÖRBOM, D. (1984). LISREL IV: A Guide to the Program and Applications ITS . : Chicago.
- JÖRESKOG, K., & SÖRBOM, D. (1989). LISREL 7 A Guide to the Program and Applications (2nd ed.). SPSS Inc. : Chicago.
- MAGNUS , J., & NEUDECKER, H. (1988). Matrix differential calculus . New York: Wiley.
- MEREDITH, W. & TISAK, J. (1990). Latent Curve Analysis *Psychometrika*, 53, 107-122.
- MUTHÉN, B. (1987). LISCOMP: Analysis of linear structural equations with a comprehensive measurement model (User's Guide). Mooresville, IN: Scientific Software.
- MOOIJART, A., & BENTLER, P. M. (1987). Robustness of normal theory statistics in structural equation models. Technical Report 87-11, Leiden University.
- NEWBY, W.K. (1988). Asymptotic Equivalence of Closest Moments and GMM Estimators. *Econometric Theory*, 4, 336-340.
- NEUDECKER, H., SATORRA, A. (1990). Linear structural relations: gradient and hessian of the fitting function. *Statistics & Probability Letters* , 11, 57-61.
- SAS (1990). Technical Report P-200. SAS/STAT Software: CALIS and LOGISTIC Procedures. SAS Institute Inc: Cary, NC.
- SATORRA, A. (1989) Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika*, 54, 131-151.
- SATORRA, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. *Quality & Quantity* 24, 367-386.
- SATORRA, A (1991). On the asymptotic variance matrix of sample second order moments in multivariate linear relations. (submitted).
- SATORRA, A., & BENTLER, P. M. (1986A). Some robustness issues of goodness of fit statistics in covariance structure analysis. ASA 1986 Proceedings of the Business and Economic Statistics Section, 549-554.
- SATORRA, A., & BENTLER, P.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. ASA 1988 Proceedings of the Business and Economic Statistics Section, 308-313.
- SATORRA, A. & BENTLER, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis*, 10, 235-249.
- 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.
- SCHOENBERG, R.J. (1989). *LINCS: Linear covariance structure analysis. User's Guide*. Kent, WA: RJS Software.
- SKINNER, C.J., HOLDT, D., & SMITH, T.M.F. (1989). *Analysis of Complex Surveys*. New York: Wiley.
- WOLTER, K.M. (1985). *Introduction to variance estimation*. New York: Springer Verlag.
- WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1-25.

APPENDIX

Consistent estimation of Γ .

Result 1

Let u_1, u_2, \dots, u_n be uncorrelated random vectors with common mean μ and finite covariance matrices $\Omega_i, i = 1, 2, \dots, n$. Define $u. := \sum_{i=1}^n u_i/n$. Then

$$E \left(\frac{n}{n-1} \right) \sum_{i=1}^n (u_i - u.)(u_i - u.)' = \text{var} \left(\sum_{i=1}^n u_i \right),$$

where "var" denotes variance matrix.

Note. In fact, the variance matrix $(n/(n-1)) \sum_{i=1}^n (u_i - u.)(u_i - u.)'$ is the usual "random group" estimator of variance of, e.g., Wolter (1985, p. 21) and Skinner et. al. (1989).

Proof. It follows from

$$\begin{aligned} E \sum_{i=1}^n (u_i - u.)(u_i - u.)' &= \\ \sum_{i=1}^n [E(u_i - \mu)(u_i - \mu)' - 2E(u_i - \mu)(u. - \mu)' + E(u. - \mu)(u. - \mu)'] &= \\ = \sum_{i=1}^n [\Omega_i - 2/n \Omega_i + (1/n^2) \sum_{j=1}^n \Omega_j] = ((n-1)/n) \sum_{i=1}^n \Omega_i &= \\ ((n-1)/n) \text{var} \left(\sum_{i=1}^n u_i \right). \end{aligned}$$

Note that only the assumption of uncorrelation among the u_i 's has been used.

B. Log-likelihood function and related results for PML and NT-MD

Let $y_1, y_2, \dots, y_i, \dots, y_n$ be n i.i.d. observations of a $(p-1)$ -dimensional normally distributed vector y of mean μ^* and covariance matrix Σ^* , and let S and Σ be the sample and population moments as defined in (6) and (7) of Section 3 above, with $z' = (y', c1)$. The log-likelihood function $l(\theta)$ is easily seen to be

$$\begin{aligned} l(\theta) &= -2^{-1}n(p-1)\ln 2\pi - 2^{-1}n \ln |\Sigma^*| - 2^{-1} \text{tr} \Sigma^{*-1} \sum_{i=1}^n (y_i - \mu^*)(y_i - \mu^*)' = \\ & -2^{-1}n(p-1) \ln 2\pi - 2^{-1}n \ln |\Sigma^*| - 2^{-1} \text{tr} \Sigma^{*-1} \sum_{i=1}^n y_i y_i' + n \mu^{*\prime} \Sigma^{*-1} y.' - 2^{-1}n \mu^{*\prime} \Sigma^{*-1} \mu^*, \end{aligned}$$

where

$$y_{\cdot} = \sum_{i=1}^n y_i/n.$$

Hence, noting that $|\Sigma| = |\Sigma^*|$ and that

$$tr \Sigma^{-1} S = n^{-1} tr (\Sigma^{*-1} \sum_{i=1}^n y_i y_i') - 2 tr (\Sigma^{*-1} \mu^* y_{\cdot}') + 1 + \mu^{*'} \Sigma^{*-1} \mu^*,$$

we can write

$$l(\theta) = -2^{-1} n [(p-1) \ln 2\pi + \ln |\Sigma| + tr \Sigma^{-1} S];$$

consequently, $l(\theta) = -2^{-1} n F_{ML} + b$, say, where b does not depend on θ . That is, the estimators that minimize F_{ML} of (22) are ML estimators.

The following results will also be used.

Lemma 1.

Given W_{NT} and Γ_{NT} as defined in (23) and (24) respectively, it holds that

i)

$$\Gamma_{NT} W_{NT} \Gamma_{NT} = \Gamma_{NT},$$

When Δ is in the column space of Γ_{NT} , then

(ii)

$$\Delta' W_{NT} \Gamma_{NT} W_{NT} \Delta = \Delta' W_{NT} \Delta$$

and

(iii)

$$(\Gamma_{NT} - \Delta(\Delta' W_{NT} \Delta)^{-1} \Delta') Q (\Gamma_{NT} - \Delta(\Delta' W_{NT} \Delta)^{-1} \Delta') = (\Gamma_{NT} - \Delta(\Delta' W_{NT} \Delta)^{-1} \Delta'),$$

where $Q := (W_{NT} - W_{NT} \Delta(\Delta' W_{NT} \Delta)^{-1} \Delta' W_{NT})$.

(iv) When Δ and $(s - \sigma(\theta))$ are in the column space of Γ_{NT} , then the MD estimator associated with $W = W_{NT}$ is the same as the one associated with $W = \Gamma_{NT}^-$, for any choice of g-inverse.

Proof Since, given the definitions of μ and Γ of Section 2 above, it holds that $\mu' \Sigma^{-1} \mu = 1$, we get that

$$2D^+(\mu\mu' \otimes \mu\mu') D^{+'} (1/2) D' (\Sigma^{-1} \otimes \Sigma^{-1}) D 2D^+(\mu\mu' \otimes \mu\mu') D^{+'} = 2D^+(\mu\mu' \otimes \mu\mu') D^{+'},$$

consequently,

$$\Gamma_{NT} W_{NT} \Gamma_{NT} = (W_{NT}^{-1} - \Upsilon) W_{NT} (W_{NT}^{-1} - \Upsilon) = W_{NT}^{-1} - 2\Upsilon + \Upsilon W_{NT} \Upsilon =$$

$$(W_{NT}^{-1} - \Upsilon) = \Gamma_{NT},$$

where $\Upsilon := 2D^+(\mu\mu' \otimes \mu\mu')D^{+'}$, which proves i) of the Lemma.

Results ii) and iii) follow trivially from i) and the stated condition of Δ to be in the column space of Γ_{NT} .

To prove iv), note first that the MD estimator is the root of $\Delta'W(s - \sigma(\theta)) = 0$. Since by the assumption stated in iv), $\Delta = \Gamma_{NT}R$ and $(s - \sigma(\theta)) = \Gamma_{NT}R^*$, say, where R and R^* are conformable matrix and vector, respectively, it holds that

$$\Delta'W(s - \sigma(\theta)) = R'\Gamma_{NT}W\Gamma_{NT}R^* = R'\Gamma_{NT}R^*,$$

hence the root of $\Delta'W_{NT}(s - \sigma(\theta)) = 0$ does not depend on the choice of g-inverse. From i) note that a choice of g-inverse is in fact W_{NT} .

Since the condition on Δ is obviously verified by the type of models defined in Section 2 (when the pseudo variance ϕ_c is a fixed parameter), equality ii) of Lemma 1 is the Condition 6* of Satorra (1989, p.137), which guaranty the correctness of the simple formulae (11) of the variance matrix of estimates (see also Shapiro, 1987).

C. Proof of Theorem 1.

Consider first the simple case where each moment matrix $\Phi_{ii} := E\delta_i\delta_i'$, $i = 1, 2, \dots, L$, is a (symmetric) free matrix, thus the parameter vector θ partitions as

$$\theta = [\tau', \text{vech}'(\Phi_{11}), \dots, \text{vech}'(\Phi_{ii}), \dots, \text{vech}'(\Phi_{LL})]'$$

Since (3) imply

$$\sigma := E \text{vech}(zz') = \sum_{i=1}^L D^+(A_i \otimes A_i)Dv(\Phi_{ii}),$$

the derivative matrix $\Delta := (\partial/\partial\theta')\sigma(\theta)$ will partition as:

(C3)

$$\Delta = [\Delta_1, D^+(A_1 \otimes A_1)D, \dots, D^+(A_i \otimes A_i)D, \dots, D^+(A_L \otimes A_L)D] = [\Delta_1, \Delta_2],$$

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

The following result will be needed.

Lemma 2 (cf. , Satorra, 1991). Let $z = \sum_{i=1}^L A_i\delta_i$, as in (3), with the δ_i 's being mutually independent and of zero mean, with the exception of δ_L , a scalar-variate constant to 1. Then

(C4)

$$\Gamma = \Omega +$$

$$\begin{aligned} & \sum_{i=1}^{L-1} [2D^+(A_i \otimes A_L)[E \delta_i(\text{vech } \delta_i \delta_i')]D'(A_i \otimes A_i)'D^{+'} + \\ & 2D^+(A_i \otimes A_i)D[E(\text{vech } \delta_i \delta_i')\delta_i'](A_i \otimes A_L)'D^{+'} + \\ & D^+(A_i \otimes A_i)D[\text{var}(\text{vech } \delta_i \delta_i') - 2D^+E(\delta_i \delta_i') \otimes E(\delta_i \delta_i')D^{+'}]D'(A_i \otimes A_i)'D^{+'} - \\ & 2D^+(\mu \otimes \mu)(\mu \otimes \mu)'D^{+'}, \end{aligned}$$

where

$$\Omega = 2D^+(\Sigma \otimes \Sigma)D^{+'}.$$

Note. We could of course write $D^+(\mu \otimes \mu)(\mu \otimes \mu)'D^{+'} = D^+(\mu\mu' \otimes \mu\mu')D^{+'}$.

Proof. See Satorra (1991).

Since $\Delta'_\perp \Delta = 0$, we have that

(C5)

$$\Delta'_\perp D^+(A_i \otimes A_i)D = 0$$

for $i = 1, 2, \dots, L$; Consequently given the form of Γ in (C4) it is verified that

(C6)

$$\Delta'_\perp (\Delta'_\perp \Gamma \Delta_\perp)^- \Delta'_\perp = \Delta'_\perp (\Delta'_\perp \Gamma_{NT} \Delta_\perp)^- \Delta'_\perp.$$

Since the right hand side of (C6) above is equal to $\Delta'_\perp (\Delta'_\perp \Omega \Delta_\perp)^- \Delta'_\perp$, result a) of the theorem is proved.

Clearly, when

(C7)

$$\Gamma := (\Omega + \Delta_2 C \Delta'_2 + \Delta_2 B + B' \Delta'_2),$$

say, which is the form of Γ given by (C4), for any matrix W it holds that

(C8)

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

where t is the dimension of the subvector τ of θ , is free of the matrices B and C . Consequently, setting B and C equal to zero, it holds that

(C9)

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

$$[(\Delta' W \Delta)^{-1} \Delta' W \Omega W \Delta (\Delta' W \Delta)^{-1}]_{t \times t}$$

which, taking into account the expression of the asymptotic variance matrix of estimates of (10) proves b) of the theorem.

It has to be noted that the form (C7) of Γ corresponds to the case where the pseudo parameter ϕ_c is set a free parameter of the model. When ϕ_c is a parameter fixed to 1, then (C7) changes to

$$(C7^*) \quad \Gamma := (\Gamma_{NT} + \Delta_2 C \Delta_2' + \Delta_2 B + B' \Delta_2');$$

in which case we get

$$(C9) \quad \begin{aligned} & [(\Delta' W \Delta)^{-1} \Delta' W \Gamma \Delta (\Delta' W \Delta)^{-1}]_{t \times t} = \\ & [(\Delta' W \Delta)^{-1} \Delta' W \Gamma_{NT} W \Delta (\Delta' W \Delta)^{-1}]_{t \times t}. \end{aligned}$$

Consequently, when ϕ_c is a fix parameter the correctness of the NT standard errors of estimates of τ , i.e., the ones provided by the matrix

$$[(\Delta' W \Delta)^{-1} \Delta' W \Omega W \Delta (\Delta' W \Delta)^{-1}]_{t \times t}$$

will not be guarantied for W arbitrary. It will of course be guarantied when $W = W_{NT}$, since in that case it is verified that $\Delta' W \Gamma_{NT} W \Delta = \Delta' W \Omega_{NT} W \Delta$. For general type of weight matrix W the validity of NT standard errors requires ϕ_c to be "declared" a free parameter (which will be estimated at the value 1). For example in ULS analysis, where W is the identity matrix, the usual NT estimates of standard errors of τ will be correct when the analysis is performed specifying ϕ_c to be free (this motivates a note on Section 6). It should be noted that setting ϕ_c free or fixed parameter has implications only for the computation of standard errors of estimates (parameter estimates and the chi-square goodness-of-fit statistic will of course not vary).

Suppose now the case where some matrices Φ_{ii} 's of normally distributed δ_i 's are also restricted to be functions of τ . Since when δ_i is normally distributed then $[E \delta_i (\text{vech} \delta_i \delta_i')']$ and $[(\text{var } \text{vech} \delta_i \delta_i') - 2D^+ E(\delta_i \delta_i') \otimes E(\delta_i \delta_i') D^{+'}]$ are null matrices, there will be a correspondence between the elements of the partition (C3) of Δ that now drop out (due to the restriction of some Φ_{ii} 's to be function of τ) and the terms of (C4) that vanish due to the normality assumption of the corresponding δ_i 's . Consequently, the same results a) and b) of the theorem apply when the covariance matrices of normally distributed δ_i 's are restricted to be functions of τ .

Table 1

Empirical distribution of parameter estimates, estimates of standard errors and test statistics. Method of estimation NT-MD. Non-normality produced by heteroskedasticity on ζ 's and ε . Sample size $n = 500$. (number of replications 600)

	(1)	(2)	(3)	(4)	(5)
ϕ_{11}	1.00	0.96	0.77	0.23*	0.69
ϕ_{22}	1.00	0.99	0.78	0.23*	0.70
ϕ_{33}	12.00	11.73	1.01	0.80*	0.94
ψ	0.80	0.79	0.21	0.05*	0.18
β	4.00	4.01	0.10	0.05*	0.10
α	8.00	7.98	0.19	0.17*	0.19
μ	1.00	1.00	0.16	0.16*	0.16

chi-square statistics

	mean	variance	reject. freq. 5 (1) % (out of 600.)
Expected ($\chi^2, df = 2$)	2.00	4.00	30 (6)
chi2	6.03*	51.25*	200 (121)*
rchi2	2.08	4.44	28 (6)
sbchi2	2.00	5.37	36 (11)
(alpha)	(3.02)		

- (1) true values
- (2) sample mean of parameter estimates
- (3) standard deviation of parameter estimates
- (4) sample mean of NT standard errors
- (5) sample mean of robust-standard errors

* not necessarily asymptotically correct

Table 2

Empirical distribution of parameter estimates, estimates of standard errors and test statistics. Method of estimation NT-MD. Data non-normal with non-normality produced by independent chi-square distributions of one df. Sample size $n=500$. (number of replications 600)

	(1)	(2)	(3)	(4)	(5)
ϕ_{11}	1.00	1.01	0.28	0.23*	0.27
ϕ_{22}	1.00	0.98	0.28	0.23*	0.27
ϕ_{33}	8.00	7.95	1.35	0.55*	1.28
ψ	0.80	0.79	0.13	0.05*	0.12
β	4.00	4.00	0.06	0.06	0.06
α	8.00	8.01	0.18	0.17	0.17
μ	1.00	0.99	0.13	0.13	0.13

chi-square statistics			
	mean	variance	reject. freq. 5 % (out of 600.)
Expected ($\chi^2, df = 2$)	2.00	4.00	30
chi2	1.98	4.18	32
rchi2	2.02	4.11	24

- (1) true values
- (2) sample mean of parameter estimates
- (3) standard deviation of parameter estimates
- (4) sample mean of NT standard errors
- (5) sample mean of robust-standard errors

* not necessarily asymptotically correct

Table 3

Empirical distribution of parameter estimates, estimates of standard errors and test statistics. Method of estimation NT-MD. Data normal (cases clustered in 20 groups of size 30, $r=0.8$). Sample size $n=600$. (number of replications 500)

	(1)	(2)	(3)	(4)	(5)
ϕ_{11}	1.00	0.88	0.40	0.21*	0.38
ϕ_{22}	1.00	0.86	0.40	0.21*	0.38
ϕ_{33}	8.00	7.32	1.40	0.48*	1.32
ψ	0.80	0.80	0.03	0.05*	0.03
β	4.00	4.02	0.08	0.06*	0.07
α	8.00	7.98	0.24	0.18*	0.23
μ	1.00	1.01	0.57	0.12*	0.59

chi-square statistics

	mean	variance	reject. freq. 5 (1) % (out of 500)
Expected ($\chi^2, df = 2$)	2.00	4.00	25 (5)
chi2	36.90*	931.24*	438 (418)*
rchi2	2.37	5.99	35 (11)
sbchi2 (alpha)	2.00 (19.87)	3.56	21 (4)

- (1) true values
- (2) sample mean of parameter estimates
- (3) standard deviation of parameter estimates
- (4) sample mean of NT standard errors
- (5) sample mean of robust-standard errors

* not necessarily asymptotically correct

RECENT WORKING PAPERS

1. ALBERT MARCET and RAMON MARIMON, *Communication, Commitment and Growth* (June 1991).
2. ANTONI BOSCH, *Economies of Scale, Location, Age, and Sex Discrimination in Household Demand* (June 1991).
3. ALBERT SATORRA, *Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures* (June 1991).
4. JAVIER ANDRES and JAUME GARCIA, *Wage Determination in the Spanish Industry* (June 1991).

UNIVERSITAT POMPEU FABRA

Balmes, 132

Telephone (93) 484 97 00

Fax (93) 484 97 02

Barcelona 08008