UNIVERSITAT POMPEU FABRA

Multi-Sample Analysis of Moment-Structures: Asymptotic Validity of Inferences Based on Second-Order Moments

Albert Satorra

Economics working paper 16. June 1992

Multi-sample Analysis of Moment-Structures: Asymptotic Validity of Inferences Based on Second-Order Moments *

Albert Satorra
Universitat Pompeu Fabra

May 1992

Abstract

In moment structure analysis with nonnormal data, asymptotic valid infereces require the computation of a consistent (under general distributional assumptions) estimate of the matrix Γ of asymptotic variances of sample second-order moments. Such a consistent estimate involves the fourth-order sample moments of the data. In practice, the use of fourth-order moments leads to computational burden and lack of robustness against small samples. In this paper we show that, under certain assumptions, correct asymptotic inferences can be attained when Γ is replaced by a matrix Ω that involves only the second-order moments of the data. The present paper extends to the context of multi-sample analysis of second-order moment structures, results derived in the context of (single-sample) covariance structure analysis (Satorra and Bentler, 1990). The results apply to a variety of estimation methods and general type of statistics. An example involving a test of equality of means under covariance restrictions illustrates theoretical aspects of the paper.

*This paper has been presented at the "19th European Meeting of Statisticians, Bernoulli Society Meeting" in Barcelona (September 1991) and the "Seventh International Conference on Multivariate Analysis" at Penn State University (May 1992). Work supported by Spanish DGICYT grant PS89-0040.

1 Introduction

Moment-structure analysis is widely used in behavioural, medical, social and economic studies to analyze linear relations among variables some of which may be latent (see, e.g., Jöreskog and Sörbom, 1989; Bentler, 1989; Muthén, 1987; and references contained therein). Often it is of interest to undertake a multi-sample analysis where samples from several populations are analyzed simultaneously under a common model (Jöreskog, 1971; Lee and Tsui, 1982; Bentler, 1989; Muthén, 1989). Recently, multi-sample analysis has been shown to be af special relevance to deal with a single-sample with incomplete data problems (Arminger and Sobel, 1990).

A very general approach to moment structure analysis is based on fitting structured population moments to sample moments using minimum-distance (MD) methods (Chamberlain, 1982; Browne, 1984; Shapiro 1986; Fuller, 1987, Section 4.2). It has been recognized that the asymptotic variance matrix of the analyzed sample moments, which we call Γ, plays a fundamental role in designing an efficient MD analysis and in assessing the sampling variability of statistics of interest. The matrix Γ will generally involve the fourth-order moments of the observable variables. Not surprisingly, however, when the observable variables are normally distributed, i.e., under a normal theory assumption, Γ is a function of second-order moments only.

There are practical reasons for using the normal theory (NT) form of Γ instead of the distribution-free form that involves moments to the fourth order. For moderate size models the number of distinct fourth-order sample moments is large, hence they lead to computational burden and lack of robustness against small samples. On the other hand, recent developments have shown the asymptotic robustness of inferences based on an NT form of Γ in situations where the normality assumption does not hold (Anderson, 1987, 1989; Anderson and Amemiya, 1988; Browne, 1987; Browne and Shapiro, 1988; Browne, 1990; Mooijart and Bentler, 1991; Satorra and Bentler, 1990). The papers by Anderson-Amemiya, Browne-Shapiro and Bentler-Mooijart deal with NT -MD methods. Satorra-Bentler's work, however, deals with general type of MD analyses, and considers general type of statistics also, like for instance the vector of moment-residuals (difference between observed and fitted moments). All of the above mentioned work, however, is confined to the analysis of covariance structures in a single-group analysis, and consider only minimum-discrepancy type of estimators.

The present paper generalizes to multi-sample analysis of mean and covariance structures the above mentioned work on asymptotic robustness. A broad class of estimators, that go beyond minimum-discrepancy methods, will be encompassed under the present approach (instrumental variable estimators, for example, will be included in this approach). Satorra-Bentler's view of asymptotic robustness will be adopted here, in the sense that we will study the validity of inferences based on replacing the correct form of Γ by another matrix, Ω, which involves only second-order moments. The proposed matrix Ω is even computationally simpler than the correct form of Γ under normality.

In multivariate analysis a simple example of multi-sample analysis is the test of equality of population means using independent samples. This test is usually carried out under the assumption that the populations are normally distributed. In the present paper we will provide asymptotic robustness results for this test in the general case where restrictions on the variance matrices of the populations are allowed.

The plan of the paper is as follows. Section 2, presents general results for the analysis of second-order moment structures. The results of asymptotic robustness are given in Section 3. Section 4 concludes with an illustration.

With respect to notation, $\text{vec}\,A$ will denote the column-vector formed by stacking columns of A one below the other. For a symmetric matrix A, $\text{v}(A)$ will denote the column vector obtained from stringing the non-duplicated elements of A column-wise in a column vector. We will use $\text{v}(A) = D^{+}\text{vec}A$, where D^{+} is the Moore-Penrose inverse of the "Duplication" matrix D (Magnus & Neudecker, 1988). The usual notation $o_p(1)$ (idem $O_p(1)$) will be used for stochastic quantities that tend to zero (idem, are bounded) in probability. Two estimators, say $\hat{\theta}$ and $\tilde{\theta}$, will be said to be asymptotically equivalent when $\sqrt{n}\,(\hat{\theta} - \tilde{\theta}) = o_p(1)$, and this will be denoted by $\hat{\theta} \stackrel{a}{=} \tilde{\theta}$ (obviously, asymptotic equivalence implies equality of the corresponding asymptotic distributions).

2 Multi-sample analysis of moment structures: asymptotic theory

Let $z_g, g = 1, \ldots, G$, be a p_g-dimensional vector of observable variables with second-order moment matrix

$$\mathcal{M}_g \equiv E z_g z_g', \tag{1}$$

where G is the number of groups and E denotes mathematical expectation. Consider the multi-sample vector of population moments

$$\mu \equiv [(v\mathcal{M}_1)', (v\mathcal{M}_2)', \ldots, (v\mathcal{M}_G)']', \tag{2}$$

and let $\mu = \mu(\theta)$ be a specific moment-structure for μ, where θ is a q-dimensional vector of parameters that varies in an open set Θ of R^q and $\mu(.)$ is continuously differentiable. Let $\mathcal{M}_g \equiv \mathcal{M}_g(\theta)$, g=1,2,..., G, be the moment-structure induced by $\mu = \mu(\theta)$ in group g.

Consider the multi-sample data

$$\{z_{gi}; i = 1, 2, \ldots, n_g, g = 1, 2, \ldots, G\} \tag{3}$$

where the z_{gi}'s are independent realizations of z_g, and the z_g's are mutually stochastically independent. Here n_g is the sample size for group g, whereas $n \equiv \sum_{i=1}^{n} n_g$ is the total sample size. Throughout the paper, $n \to \infty$ means that $n_g \to \infty$ for each g.

Consider now the sample second-order moment matrix for group g,

$$M_g \equiv \sum_{i=1}^{n_g} z_{g_i} z_{g_i}' / n_g, \tag{4}$$

the corresponding vector of sample moments $m_g \equiv v(M_g)$, and the sample vector of second-order moments of the G groups

$$m \equiv [m_1', m_2', \ldots, m_G']'. \tag{5}$$

Straightforward application of the Central Limit theorem shows that $\sqrt{n_g} m_g$, $g = 1, 2, \ldots, G$, is asymptotically normal, with asymptotic variance matrix (avar)

$$\Gamma_g \equiv \text{avar}\,(\sqrt{n_g} m_g) = E(\text{v}\, z_g z_g')(\text{v}\, z_g z_g')' - E(\text{v}\, z_g z_g')E(\text{v}\, z_g z_g')', \tag{6}$$

where "avar" denote asymptotic variance matrix. An estimate of Γ_g which is consistent regardless of the distribution of z_g (provided that z_g has finite eigth-order moments) is the following $p_g^\star \times p_g^\star$ matrix of fourth-order sample moments:

$$\hat{\Gamma}_g = \sum_{i=1}^{n_g} (d_{gi} - m_g)(d_{gi} - m_g)' / (n_g - 1), \tag{7}$$

where $d_{gi} \equiv \text{v}\,(z_{gi} z_{gi}')$. Throughout, we will use the notation $p_g^\star \equiv p_g(p_g + 1)/2, g = 1, 2, \ldots, G$.

Since the m_g's, g=1,2,...,G, are independent, the asymptotic variance matrix of $\sqrt{n}m$ is of the form

$$\Gamma \equiv avar(\sqrt{n}m) = \mathrm{diag}\,(n/n_1\Gamma_1, ..., n/n_g\Gamma_g, ..., n/n_g\Gamma_G), \tag{8}$$

where "diag" denotes a block-diagonal matrix. Consequently, a consistent estimate of Γ will be

$$\hat{\Gamma} \equiv avar(\sqrt{n}m) = \mathrm{diag}\,(n/n_1\hat{\Gamma}_1, ..., n/n_g\hat{\Gamma}_g, ..., n/n_g\hat{\Gamma}_G), \tag{9}$$

where the $\hat{\Gamma}_g$'s are given in (7). Note that the matrices Γ and $\hat{\Gamma}$ are both of order $p \times p$, where $p \equiv \sum_{g=1}^{G} p_g{}^{\star}$.

Consider now the $p \times q$ derivative matrix $\Delta \equiv (\partial/\partial\theta')\mu(\theta)$, its orthogonal complement, denoted $\Delta_\perp$ (i.e., $\Delta_\perp$ is a $p \times (p-q)$ matrix of full column rank such that $\Delta_\perp{}'\Delta = 0$, a zero matrix), and the partitioned matrices

$$\Delta = [\Delta_1{}', \ldots, \Delta_g{}', \ldots, \Delta_G{}']', \tag{10}$$

where $\Delta_g = (\partial/\partial\theta')\mathrm{v}\,\mathcal{M}_g(\theta)$, and

$$\Delta_\perp{}' = [H_1, \ldots, H_g, \ldots, H_G], \tag{11}$$

say, conformably with the partition of Δ above.

Given an estimator of θ, say $\hat{\theta}$, consider the following matrix and vector of fitted-moments for group g (g=1,2,...,G):

$$\hat{\Sigma}_g \equiv \mathcal{M}_g(\hat{\theta})$$

and

$$\hat{\sigma}_g \equiv \mathrm{v}\,(\hat{\mu}_g),$$

respectively, and the multi-sample vector of fitted moments

$$\hat{\mu} = [\hat{\mu}_1{}', \hat{\mu}_2{}', \ldots, \hat{\mu}_G{}']'$$

and vector of residuals

$$u \equiv m - \hat{\mu} = [(m_1 - \hat{\mu}_1)', (m_1 - \hat{\mu}_2)', \ldots, (m_1 - \hat{\mu}_G)']'. \tag{12}$$

Most of the estimators used in structural equation models can be shown to be asymptotically equivalent to a continuously differentiable function of sample moments. This is the case for instance of the minimum discrepancy (MD) estimators defined as the solution $\hat{\theta}$ to

$$\min_{\theta\in\Theta}(m - \mu(\theta))'V_n(m - \mu(\theta)), \tag{13}$$

where Θ is the parameter space of θ and V_n is a positive semi-definite matrix that converges in probability to a positive semi-definite matrix V. A typical expression for V_n will be

$$V_n = \operatorname{diag}(n_1/n\hat{\Omega}_1, ..., n_g/n\hat{\Omega}_g, ..., n_G/n\hat{\Omega}_G) \tag{14}$$

where

$$\hat{\Omega}_g = 2D^+(M_g \otimes M_g)D^{+\prime}, \tag{15}$$

which leads to the so called normal MD (NMD) estimates. It is also the case of the pseudo maximum likelihood (PML) estimator which minimizes

$$F(m, \mu(\theta)) = \sum_{g=1}^{G} (n_g/n) F_g(m_g, \mu_g(\theta)) \tag{16}$$

where

$$F(m, \mu(\theta)) = \ln \mid \mathcal{M}_g \mid + \operatorname{tr} M_g(\mathcal{M}_g{}^{-1}) - \ln \mid M_g \mid - p_g. \tag{17}$$

Other estimatars that are also functions of sample moments are the instrumental variable estimators (Jennrich, 1987). More specifically, assume that $\hat{\theta} \stackrel{a}{=} \theta(m)$, where $\theta(.)$ is a continuously differentiable vector valued function of m with the property that

$$\theta(\mu(\theta)) = \theta; \tag{18}$$

that is, $\hat{\theta}$ is asymptotically equal to a Fisher-consistent estimate of θ (Rao, 1973, p.345; Dijkstra, 1983, Kano, 1991, Satorra, 1989b). Clearly, since m converges in probability to μ, (18) yields

$$\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1), \tag{19}$$

i.e., $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator of θ_0. Further, the delta-method (e.g., Rao, 1973) gives

$$\operatorname{avar} \sqrt{n}\theta(m) = \operatorname{avar} \sqrt{n}\hat{\theta} = \Theta\Gamma\Theta', \tag{20}$$

where $\Theta \equiv (\partial/\partial m')\theta(\mu)$. Since $\Theta \equiv [\Theta_1, ..., \Theta_g, ..., \Theta_G]$, where $\Theta_g \equiv (\partial/\partial m_g')\theta(\mu)$, using (9), we can write (20) as

$$\operatorname{avar} \sqrt{n}\theta(m) = \sum_{g=1}^{G} \Theta_g \Gamma_g \Theta_g'. \tag{21}$$

Often the interest focuses on the asymptotic distribution of a subvector $\hat{\tau}$ of $\hat{\theta}$ (for example, the regression estimates in a regression model). The asymptotic variance matrix of $\hat{\tau}$ will be given by

$$\operatorname{avar}\sqrt{n}\hat{\theta} = \Theta_1 \Gamma \Theta_1' = \sum_{g=1}^{G} \Theta_{1g} \Gamma_g \Theta_{1g}', \tag{22}$$

where $\Theta_{1g} \equiv (\partial/\partial m_g')\tau(\mu)$ with $\tau(.)$ being the corresponding subvector of $\theta(.)$.

An immediate implication of (18) (i.e. Fisher-consistency) is

$$\Theta\Delta = I_q \tag{23}$$

where I_q denotes the $q \times q$ identity matrix; consequently, given an arbitrary partition $\theta = ({\theta_1}', {\theta_2}')'$ of θ, it holds that

$$\Theta_1\Delta_1 = I,\ \Theta_2\Delta_2 = I,\ \Theta_1\Delta_2 = 0,\ \Theta_2\Delta_1 = 0, \tag{24}$$

where $\Delta = [\Delta_1, \Delta_2]$ and $\Theta = [{\Theta_1}', {\Theta_2}']'$ are the partitions of Δ and Θ associated with the partition of θ given above. Throughout the paper, I and 0 will denote respectively the identity and zero matrices of dimensions determined by the context.

By standard asymptotic arguments, the vector of residuals is asymptotically normal with

$$\operatorname{avar}\sqrt{n}u = [I - \Delta\Theta]\Gamma[I - \Delta\Theta]' = C({\Delta_\perp}'\Gamma\Delta_\perp)C', \tag{25}$$

where $C \equiv [I - \Delta\Theta]\Delta_\perp({\Delta_\perp}'\Delta_\perp)^{-1}$. This follows from (18) and (19) which imply that

$$\hat{\mu} = \mu + \Delta(\hat{\theta} - \theta_0) = \mu + \Delta\Theta(m - \mu) + o(1/\sqrt{n}), \tag{26}$$

Note also that, from (9) and (10), we can write

$${\Delta_\perp}'\Gamma\Delta_\perp = \sum_{g=1}^{G} H_g \Gamma_g H_g'. \tag{27}$$

The following (residual-based) goodness-of-fit statistic will be of interest

$$T = nu'\hat{A}^{-1}u, \tag{28}$$

where $\hat{A}$ is a consistent estimate of [1]

$$A \equiv \Delta_\perp(\Delta_\perp{}'\Gamma\Delta_\perp)^{-1}\Delta_\perp{}'. \tag{29}$$

An obvious expression for $\hat{A}$ will be obtained from (29) by evaluating the derivative matrix Δ at $\hat{\theta}$ and substituting $\hat{\Gamma}$ of (9) for Γ. By standard arguments on the distribution of quadratic forms of normal variables, when the model holds T will be asymptotically chi-squared distributed with r degrees of freedom, where $r \equiv \text{rank}(\Delta_\perp{}'\Gamma\Delta_\perp)$ (cf., Moore, 1977). In the case of single-group analysis of covariance structures, T is Browne's (1984, Proposition 4) quadratic form chi-square goodness-of-fit statistic. The statistic T can also be shown to be asymptotically equal to other more conventional goodness-of-fit statistics (see, e.g., Satorra, 1989a).

3 Asymptotic robustness

Assume $z_g, g = 1, 2, \ldots, G$, has the following linear latent-variable structure:

$$z_g = \sum_{i=1}^{h(g)} B_{ig}\delta_{ig}, \tag{30}$$

where, B_{ig} is a $p_g \times r_{ig}$ matrix, and the δ_{ig}'s are mutually independent r_{ig}-dimensional random vectors (whose components may be observable or latent).

Clearly, under (30), we can write

$$\text{vec}\, z_g z_g{}' = \sum_{i=1}^{h(g)}(B_{ig} \otimes B_{ig})\text{vec}(\delta_{ig}\delta_{ig}{}') + \sum_{i \neq j}(B_{jg} \otimes B_{ig})\text{vec}\, \delta_{ig}\delta_{jg}{}'; \tag{31}$$

thus, premultiplying both sides of the equality by D^+, and taking expectations, we obtain

$$\mu_g = D^+ \text{vec}\, E(z_g z_g{}') = \sum_{i=1}^{h(g)} D^+(B_{ig} \otimes B_{ig})D\text{v}\Phi_{iig} + \sum_{i \neq j} D^+(B_{jg} \otimes B_{ig})\text{vec}\, \Phi_{ijg}, \tag{32}$$

where $\Phi_{iig} \equiv E(\delta_{ig}\delta_{ig}{}')$ and $\Phi_{ijg} \equiv E(\delta_{ig}\delta_{ig}{}')$. The following assumption will be needed.

ASSUMPTION A In addition to (30), assume that, for $g = 1, 2, \ldots, G$,

[1] note that the right hand side of (29) is a generalized inverse of the right hand side of (25)

a) $E\delta_{ig} = 0, i = 1, \ldots, h(g) - 1$, and $\delta_{h(g)g}$ is constant equal to 1.

b) θ is unrestricted and partitioned as

$$\begin{aligned} \theta &= [\tau', (\mathrm{v}\,\Phi_{221})', \ldots, (\mathrm{v}\,\Phi_{ii1})', \ldots, (\mathrm{v}\,\Phi_{h(1)h(1)1})', \ldots \\ &\quad \ldots, (\mathrm{v}\,\Phi_{22g})', \ldots (\mathrm{v}\,\Phi_{iig})', \ldots, (\mathrm{v}\,\Phi_{h(g)h(g)g})', \ldots, \\ &\quad \ldots, (\mathrm{v}\,\Phi_{22G})', \ldots, (\mathrm{v}\,\Phi_{iiG})', \ldots, (\mathrm{v}\,\Phi_{h(G)h(G)G})']' \\ &\quad = [\tau', \psi'], \end{aligned} \tag{33}$$

c) $B_{ig} = B_{ig}(\tau)$ and $\Phi_{11g} = \Phi_{11g}(\tau)$ are continuously differentiable functions

d) the δ_{1g}'s ($g = 1, 2, \ldots, G$) have the third- and fourth-order moments as under normality

e) the δ_{1g}'s ($i = 2, \ldots, n_g, g = 1, 2, \ldots, G$) are mutually independent.

Clearly, under Assumption A, for $g = 1, 2, \ldots, G$, it holds that

$$\mu_g = \sum_{i=1}^{h(g)} (B_{ig} \otimes B_{ig}) D \mathrm{v}\,\Phi_{iig}; \tag{34}$$

and, consequently, the derivative matrix $\Delta_g \equiv (\partial/\partial\theta')\mu_g(\theta)$ can be partitioned as

$$\begin{aligned} \Delta_g &= [\Delta_{g1}, 0, D^+(B_{2g} \otimes B_{2g})D, \ldots, D^+(B_{ig} \otimes B_{ig})D, \ldots \\ &\quad \ldots\ldots, D^+(B_{h(g)g} \otimes B_{h(g)g})D, 0], \end{aligned} \tag{35}$$

where $\Delta_{g1} \equiv (\partial/\partial\tau')\mu_g(\theta)$. It should be noted that in the partitioned matrix $\Delta = [\Delta_1', \ldots, \Delta_g', \ldots, \Delta_G']'$, the elements above and below the submatrices $D^+(B_{ig} \otimes B_{ig})D$ are zero; consequently, $H_g \Delta_g = 0$ $(g = 1, \ldots, G)$, since $\Delta_\perp{}'\Delta = 0$. Hence, using (11) and (35), we get

$$H_g D^+(B_{ig} \otimes B_{ig})D = 0, \;\; i = 2, 3, \ldots, h(g); \tag{36}$$

further, using (24), we get

$$\Theta_{1g} D^+(B_{ig} \otimes B_{ig})D = 0, \;\; i = 2, \ldots, h(g). \tag{37}$$

Recall that $\Theta_{1g} \equiv (\partial/\partial\mu_g{}')\tau(\mu)$. The following Lemma will now be needed.

LEMMA 2. (cf., Satorra (1991)) Let $z = \sum_{i=1}^{L} B_i \delta_i$ with the δ_i's being mutually independent and of zero mean, with the exception of δ_L, where the mean vector $E\delta_L$ may differ from zero. Then

$$\begin{aligned}\mathrm{var}(D^+ \mathrm{vec}\, zz') &= 2D^+(Ezz' \otimes Ezz')D^{+\prime} \qquad (38)\\ &+ \sum_{i=1}^{L-1} \{2D^+(B_i \otimes B_L)[E\delta_i(v\delta_i\delta_i')']D'(B_i \otimes B_i)'D^{+\prime}\\ &+2D^+(B_i \otimes B_i)D[E(v\delta_i\delta_i')\delta_i'](B_i \otimes B_L)'D^{+\prime}\\ &+D^+(B_i \otimes B_i)D[\mathrm{var}(v\delta_i\delta_i')\\ &-2D^+E(\delta_i\delta_i') \otimes E(\delta_i\delta_i')D^{+\prime}]D'(B_i \otimes B_i)'D^{+\prime}\}\\ &-2D^+(Ez \otimes Ez)(Ez \otimes Ez)'D^+\end{aligned}$$

PROOF. See Satorra (1991).

Note that if δ_j, say, is normally distributed then

$$\mathrm{var}\, v(\delta_j\delta_j') = 2D^+ E(\delta_j\delta_j') \otimes E(\delta_j\delta_j')D^{+\prime} \qquad (39)$$

$$E(\mathrm{v}\, \delta_j\delta_j')\delta_j' = 0; \qquad (40)$$

consequently, the terms on the right-hand side of (38) which correspond to third- and fourth-order moments of δ_j vanish when δ_j is normally distributed. In particular, when all the δ_j' s are normally distributed, then

$$\mathrm{var}(D^+ \mathrm{vec}\, zz') =$$

$$2D^+[(Ezz' \otimes Ezz') - (EzEz' \otimes EzEz')]D^{+\prime}.$$

We will now define

$$\Omega \equiv \mathrm{diag}[n/n_1\Omega_1, \ldots, n/n_g\Omega_g, \ldots, n/n_g\Omega_G], \qquad (41)$$

where

$$\Omega_g \equiv 2D^+(Ez_g z_g' \otimes Ez_g z_g')D^{+\prime}. \qquad (42)$$

We will also define $\Omega^\star$ as Ω in (41) but replacing Ω_g by

$$\Omega_g^\star \equiv 2D^+[(Ez_g z_g' \otimes Ez_g z_g') - (Ez_g Ez_g' \otimes Ez_g Ez_g')]D^{+\prime}. \qquad (43)$$

Note that a (consistent) estimate of Ω, is obtained replacing in (42) M_g of (4) for $Ez_g z_g'$. To obtain a consistent estimate of $\Omega_g^\star$ we will replace, in addition, the sample mean of z_g for Ez_g. The following theorem synthesizes the basic results of asymptotic robustness.

THEOREM 1 Let Assumption A hold. Then

$$\Theta_1 \Gamma \Theta_1' = \Theta_1 \Omega \Theta_1' = \Theta_1 \Omega^\star \Theta_1' \tag{44}$$

and

$$\Delta_\perp' \Gamma \Delta_\perp = \Delta_\perp' \Omega \Delta_\perp = \Delta_\perp' \Omega^\star \Delta_\perp \tag{45}$$

PROOF. It follows directly from the orthogonality results (36) and (37) and Lemma 1. ■

Note that the matrix $\Omega^\star$ is the asymptotic variance matrix of the multi-sample vector m of sample moments, under a normality assumption. Hence, the above theorem says that, under Assumption A, the vector of residuals u, the test statistic T and the estimator $\hat{\tau}$ have an asymptotic distribution which is insensitive to the deviation of Γ from its normal-theory form $\Omega^\star$. Consequently, under Assumption A, in order to evaluate the asymptotic variance matrix of the estimator $\hat{\tau}$ and vector of residuals u (formulae (22) and (25) respectively), as well as to compute T (see (28) and (29)), a consistent estimate of $\Omega^\star$, or of the simpler expression Ω, can be used instead of a consistent estimate of Γ. The possibility of using an estimate of Ω instead of an estimate of Γ will be of interest in practice since it will enable us to compute consistent estimates of standard errors of parameter estimates and residuals, and asymptotically chi-square test statistics, using only sample second-order moments. The next section will provide an illustration of how these results can be used.

4 An Illustration

Consider the example "Effects of Head Start Program" reported in the LISREL 7 manual (Example 10.2, p. 280 of Jöreskog and Sörbom, 1989; see also "Example: Efects of Head Start", in the EQS manual of Bentler, 1989, p. 186). In Jöreskog and Sörbom (1989) we read "Sörbom used data on 303 white children from Head Start summer program, consisting of a Head Start sample (N=148) and a matched Control sample (N=155) [...]. The variables used in Sörbom's reanalysis were x1= Mother's education, x2= Father's education, x3= Father's occupation, x4= Family income, y1=Score on the Metropolitan Readiness Test, y2= Score on the Illinois Test of Psycholinguistic Abilities" (p. 253). The variables x1 to x4 are regarded as indicators of socio-economic status (SES) while y1 and y2 are indicators of ability.

The main interest of the study is to see whether there is a significant difference in the effects of the program among the two mentioned groups of children. The model postulates the following structural relation:

$$\eta_g = \alpha_g + \gamma_g \xi_g + \zeta_g, \tag{46}$$

where η_g, ξ_g and $\zeta_g$are Ability, SES and the disturbance term in the regression equation, respectively, of group g, $g = 1, 2$. The coefficients α_g and γ_g are respectively the intercept and slope parameters of group g. In fact, significant difference among the two groups implies that the α_g's differ. Measurement equations will be added to (46).

The model can be represented by a measurement equation (assumed to be invariant for both groups)

$$\begin{pmatrix} Y_{g1} \\ Y_{g2} \\ X_{g1} \\ X_{g2} \\ X_{g3} \\ X_{g4} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \iota_{y1} \\ \lambda_y & 0 & \iota_{y2} \\ 0 & 1 & \iota_{x1} \\ 0 & \lambda_{x2} & \iota_{x2} \\ 0 & \lambda_{x3} & \iota_{x3} \\ 0 & \lambda_{x4} & \iota_{x4} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_g \\ \xi_g \\ 1 \end{pmatrix} + \begin{pmatrix} \epsilon_{g1} \\ \epsilon_{g2} \\ \epsilon_{g3} \\ \epsilon_{g4} \\ \epsilon_{g5} \\ \epsilon_{g6} \\ 0 \end{pmatrix}, g = 1, 2; \tag{47}$$

the structural equation for Head Start Group (HSG):

$$\begin{pmatrix} \eta_1 \\ \xi_1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & \gamma_1 & \alpha \\ 0 & 0 & \kappa \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \xi_1 \\ 1 \end{pmatrix} + \begin{pmatrix} \zeta_{11} \\ \zeta_{12} \\ 1 \end{pmatrix}; \tag{48}$$

and the structural equation for the Control Group (CG):

$$\begin{pmatrix} \eta_1 \\ \xi_1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & \gamma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_2 \\ \xi_2 \\ 1 \end{pmatrix} + \begin{pmatrix} \zeta_{21} \\ \zeta_{22} \\ 1 \end{pmatrix}. \tag{49}$$

The diagonal elements of the covariance matrix of measurement error terms [2] in (47) are free parameters, except for the last element of the diagonal which is fixed to zero. The correlation between ϵ_{g3} and ϵ_{g4} is also set a free

[2]The so called matrix "TE" in LISREL terminology

parameter. The diagonal elements of the covariance matrix of the last vector on the right hand side of (49) [3] are also free parameters.

The above model has the form of a linear structure (30), since we can write $z_g \equiv (Y_{g1}, Y_{g2}, X_{g1}, X_{g2}, X_{g3}, X_{g4}, 1)'$ and

$$\begin{array}{rr} \delta_{1g} \equiv \zeta_{g1}, & \delta_{2g} \equiv \zeta_{g2}, \\ \delta_{3g} \equiv \epsilon_{g1}, & \delta_{4g} \equiv \epsilon_{g2}, \\ \delta_{5g} \equiv (\epsilon_{g3}, \epsilon_{g4})' & \delta_{6g} \equiv \epsilon_{g5}, \\ \delta_{7g} \equiv \epsilon_{g6}, & \epsilon_{8g} \equiv 1 \end{array} \tag{50}$$

Note that the B_{ig}'s of (30), for $i = 1, 2, 8$, are, respectively, the 1st, 2nd and 3rd columns of the following matrix product:

$$\begin{pmatrix} 1 & 0 & \tau_{y1} \\ \lambda_y & 0 & \tau_{y2} \\ 0 & 1 & \tau_{x1} \\ 0 & \lambda_{x2} & \tau_{x2} \\ 0 & \lambda_{x3} & \tau_{x3} \\ 0 & \lambda_{x4} & \tau_{x4} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \gamma_1 & \alpha \\ 0 & 0 & \kappa \\ 0 & 0 & 0 \end{pmatrix}^{-1} ; \tag{51}$$

the B_{ig}'s, for $i = 3, 4, 6, 7$, are respectively the 1st, 2nd, 5th and 6th columns of an identity matrix of dimension 7×7; finally, B_{5g} is the matrix formed by the 3th and 4th columns of a 7×7 identity matrix. Moreover, α_g and κ_g are set to zero when $g = 2$. Note that the variable δ_{8g}, noted by 1, is the constant 1.

Jöreskog and Sörbom (1989) say that "There seems to be no significant effect for the Head Start program when controlling for social status [...]" (p. 257), on the bases that they observe a non-significant z-value (estimate divided by its standard error) for α. Jöreskog and Sörbom's analysis was based on the assumption that the distribution of the observable variables was normal; in fact they carried out a PML analysis. Our concern here is whether the above conclusion can still be defended without the normality assumption, that is, whether the NT standard error of the estimate of α is asymptotically correct even for non-normal data.

Since in this example the variances of the corresponding δ_{ig}'s of (30) are unrestricted, i.e. they are parameters of the model, if the δ_{ig}'s are assumed to be independent (not necesarely normally distributed), Assumption A holds and hence the conclusions of the Theorem. Consequently, under the above

[3] The matrix "PSI", in LISREL terminology

assumption we can use Ω instead of Γ to compute standard errors, chi-square goodness-of-fit statistics and the variance matrix of the vector of residuals. In fact this is the type of analysis undertaken when the "multi-sample" option of LISREL (Jöreskog and Sörbom, 1984) is used with "ML" as estimation method and "CM" as a "matrix to be analyzed". The matrices of each group to be analyzed are the moment matrices shown in Table 1 (which was deduced from Table 10.4 of Jöreskog and Sörbom, 1988). Note that $M_g, g = 1, 2$, can be partitioned as

$$M = \begin{pmatrix} M_{yy} & M_{yx} & m_y \\ M_{xy} & M_{xx} & m_x \\ m_y & m_x & 1 \end{pmatrix}$$

where the M's are (uncentered) sample second-order moments and the m's are sample mean vectors. The results of a PML analysis are shown in Table 2. It can be seen that figures on Table 2 match[4] with the figures shown on Table 10.6 of the LISREL.

In conclusion, Theorem 1 applied to this example implies that even in the case of non-normal distribution of latent factors and errors, if the δ_{ig}'s are mutually stochastically independent, then correctness of some of the NT inferential statistics can be claimed. In particular, the standard errors that in Table 2 appear with an asterisc ($\star$) are asymptotically correct even under non-normality. The robust inferential statistics are the standard errors of the estimates of the parameters λ's, ι's, γ's and α and κ which correspond to the subvector τ of θ of Section 3. Moreover, Theorem 1 ensures also that the P-value of .238 associated to the chi-square goodness-of-fit test of the model will also be valid with non-normal data under the assumption of independence. Note, however, that the standard errors corresponding to estimates of variances of the ϵ's ξ's, and ζ's will not necessarily be correct when such variables are non-normally distributed.

[4] up to rounding errors introduced when obtaining Table 1 from correlations and standard deviations reported in the LISREL manual

References

Amemiya, Y. & Anderson, T.W. 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." Pp. 9-36 in *Proceedings of the Second International Conference in Statistics,*edited by T. Pukkila & S.Puntanen. Tampere, Finland: University of Tampere.

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, M.E., 1990. "Pseudo-maximum likelihood estimation of mean and covariance structures with missing data" *Journal of the American Statistical Association, 85:*195-203.

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

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

Browne, M.W. 1987 . "Robustness in statistical inference in factor analysis and related models" *Biometrika, 74,* 375-384.

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

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

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

Jennrich, R.I. 1987. "Tableau algorithms for factor analysis by instrumental variable methods" *Psychometrika, 52.* 469-476.

Jöreskog, K. 1971. "Simultaneous factor analysis in several populations" *Psychometrika, 57:* 409-426.

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

Kano, Y. 1991. "Asymptotic Properties of Statistical Inference Based on Fisher-Consistent Estimators in the Analysis of Covariance Structures" *Unpublished Manuscript.*

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

Magnus J. & Neudecker, H. 1988. *Matrix differential calculus* . New York: Wiley.

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

Moore, D.S. 1977. "Generalized Inverses, Wald's Method, and the Construction of Chi-Squared Tests of Fit." *Journal of the American Statistical Association, 72:* 131-137.

Muthén, B. 1987. *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model* (User's Guide). Mooresville, IN: Scientific Software.

Muthén, B. 1989. "Multiple group structural modeling with non-normal continuous variables." *British Journal of Mathematical and Statistical Psychology, 42:* 55-62.

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

Satorra, A. 1989b. "Robustness of inferences based on the normality assumption: the instrumental variables case" *Paper presented at the European Meeting of the Psychometric Society,* Louvain, August 1989.

Satorra, A 1991. "On the asymptotic variance matrix of sample second order moments in multivariate linear relations." *Economics Working Paper 8.* Universitat Pompeu Fabra, Barcelona.

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. 1986. "Asymptotic theory of overparameterized models." *Journal of the American Statistical Association, 81:* 142-149.

Table 1: Moment matrices of first- and second-order moments

Group 1 (Head Start Group)						
14.165						
11.598	11.134					
7.665	6.713	5.515				
19.932	17.125	12.261	35.720			
70.619	61.889	41.917	106.240	401.155		
34.620	29.897	20.688	52.637	194.794	98.598	
3.520	3.081	2.088	5.358	19.672	9.562	1
Group 1 (Control Group)						
16,588						
13.417	12.252					
10.345	9.042	8.183				
25.885	22.003	18.226	51.900			
79.593	68.167	53.991	132.823	431.982		
39.639	34.095	26.941	66.227	212.313	108.798	
3.839	3.290	2.600	6.435	20.415	10.070	1

Table 2: Parameter estimates with standard errors between brackets below (Results obtained using LISREL 7 under the "ML" option and the moment matrices shown in Table 1). The chi-square goodness-of-fit statistic was equal to 27.44 (23 degrees of freedom, P = .238). With an " $\star$ " we indicate that the corresponding NT standard error is asymptotically valid even under non-normality

$$\hat{\Lambda}_{HSG\&CG} = \begin{bmatrix} 1.00 & 0.00 & 20.35 \\ & & (.28)* \\ .85 & 0 & 10.08 \\ (.14) & & (.21)* \\ 0 & 1 & 3.86 \\ & & (.09)* \\ 0 & .85 & 3.33 \\ & (.14)* & (.08)* \\ 0 & 1.20 & 2.57 \\ & (.22)* & (.09)* \\ 0 & 2.75 & 6.42 \\ & (.51)* & (.22)* \\ 0 & 0 & 1.0 \end{bmatrix}$$

$$\hat{B}_{HSG} = \begin{bmatrix} 0 & 2.13 & 0.18 \\ & (.55)* & (.37)* \\ 0 & 0 & -.38 \\ & & (.10)* \\ 0 & 0 & .0 \end{bmatrix} \qquad \hat{B}_{CG} = \begin{bmatrix} 0 & 2.13 & 0 \\ & (.55)* & \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\hat{\Psi}_{HSG} = \begin{bmatrix} 6.34 & & \\ (1.47) & & \\ 0 & .31 & \\ & (.10) & \\ 0 & 0 & 1 \end{bmatrix} \qquad \hat{\Psi}_{CG} = \begin{bmatrix} 6.18 & & \\ (1.47) & & \\ 0 & .40 & \\ & (.13) & \\ 0 & 0 & 1 \end{bmatrix}$$

$$\hat{\Theta}_{HSG} = \begin{bmatrix} \begin{matrix}6.31\\(1.53)\end{matrix} & & & & & & \\ 0 & \begin{matrix}1.46\\(0.98)\end{matrix} & & & & & \\ 0 & 0 & \begin{matrix}1.41\\(0.18)\end{matrix} & & & & \\ 0 & 0 & \begin{matrix}.48\\(.14)\end{matrix} & \begin{matrix}1.44\\(.18)\end{matrix} & & & \\ 0 & 0 & 0 & 0 & \begin{matrix}.71\\(.12)\end{matrix} & & \\ 0 & 0 & 0 & 0 & 0 & \begin{matrix}4.61\\(0.73)\end{matrix} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{\Theta}_{\epsilon\ CG} = \begin{bmatrix} \begin{matrix}7.32\\(1.59)\end{matrix} & & & & & & \\ 0 & \begin{matrix}1.64\\(1.00)\end{matrix} & & & & & \\ 0 & 0 & \begin{matrix}1.46\\(0.19)\end{matrix} & & & & \\ 0 & 0 & \begin{matrix}.42\\(.13)\end{matrix} & \begin{matrix}1.08\\(.14)\end{matrix} & & & \\ 0 & 0 & 0 & 0 & \begin{matrix}.83\\(.14)\end{matrix} & & \\ 0 & 0 & 0 & 0 & 0 & \begin{matrix}7.44\\(1.06)\end{matrix} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

RECENT WORKING PAPERS

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).

8. Albert Satorra, *The Variance Matrix of Sample Second-order Moments in Multivariate Linear Relations* (January 1992).

9. Teresa Garcia-Milà and Therese J. McGuire, *Industrial Mix as a Factor in the Growth and Variability of States' Economies* (January 1992).

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).

12. Daniel Serra and Charles ReVelle, *The PQ-Median Problem: Location and Districting of Hierarchical Facilities.* Part I (April 1992).

13. Daniel Serra and Charles ReVelle, *The PQ-Median Problem: Location and Districting of Hierarchical Facilities.* Part II: Heuristic Solution Methods (April 1992).

14. Juan Pablo Nicolini, *Ruling out Speculative Hyperinflations: a Game Theoretic Approach* (April 1992).

15. Albert Marcet and Thomas J. Sargent, *Speed of Convergence of Recursive Least Squares Learning with ARMA Perceptions* (May 1992).

16. Albert Satorra, *Multi-Sample Analysis of Moment-Structures: Asymptotic Validity of Inferences Based on Second-Order Moments* (June 1992).

UNIVERSITAT POMPEU FABRA

Balmes, 132

Telephone (343) 484 97 00

Fax (343) 484 97 02

08008 Barcelona