Fusion of Data Sets in Multivariate Linear Regression with Errors-in-Variables

Albert Satorra [1]

Departament d'Economia i Empresa
Universitat Pompeu Fabra
Ramon Trias Fargas 23-25, 08005 BARCELONA, Spain

Summary: We consider the application of normal theory methods to the estimation and testing of a general type of multivariate regression models with errors-in-variables, in the case where various data sets are merged into a single analysis and the observable variables deviate possibly from normality. The various samples to be merged can differ on the set of observable variables available. We show that there is a convenient way to parametrize the model so that, despite the possible non-normality of the data, normal-theory methods yield correct inferences for the parameters of interest and for the goodness-of-fit test. The theory described encompasses both the functional and structural model cases, and can be implemented using standard software for structural equations models, such as LISREL, EQS, LISCOMP, among others. An illustration with Monte Carlo data is presented.
among others.

1 Introduction

Consider the classical regression equation

$$Y_i = \alpha + \beta x_i + v_i, \quad i = 1, \ldots, n, \tag{1}$$

where, for individual i, Y_i and x_i are the values of the response and explanatory variables respectively, v_i is the value of the disturbance term, and α and β are respectively the intercept and slope parameters. Suppose we are unable to observe x_i. Instead we observe X_i which satisfy

$$X_i = x_i + u_i, \tag{2}$$

where u_i is an error term. We assume the v_i and u_i are iid with mean 0 and variances σ_v^2 and σ_u^2 respectively. Equations (1) and (2) with the associated assumptions define the classical simple regression model with errors-in-variables. When the x_i are a set of fixed values (across hypothetical sample replications) the model is called a functional model. When the x_i are random (i.e. varying across sample replications) we have the so-called structural model. See Fuller (1987) for a comprehensive overview of measurement error models in regression analysis. For recent work on the importance of assessing measurement reliability in multivariate linear regression, see Gleser (1992).

As it is well known, when $\sigma_u^2 > 0$ and X_i is used instead of x_i in (1), the usual least-squares estimate is not consistent for β; in fact, under normality, equations (1) and (2) with the associated assumptions fail to identify the parameters of the model. To obtain a consistent estimate of β, additional information related with the measurement error variance σ_u^2 is required. The parameter β is identified, for example, when we specify the value of σ_u^2, or the ratio of variances σ_u^2/σ_v^2, amid of other possibilities. In practice, however, it may be difficult to have such exact information on the value of error variances.

An alternative to specifying the size of measurement error variances is the multiple indicator approach. Consider for example two indicators X_{i1} and X_{i2} ($i = 1, \ldots, n$) which satisfy

$$X_{i1} = x_i + u_{i1}, \quad X_{i2} = x_i + u_{i2}, \tag{3}$$

where the u_{i1} and u_{i2} are uncorrelated error terms. Equations (1) and (3) with the associated assumptions yield an identified model. Inferences for this type of models is usually carried out under normality and a single sample with complete data (see, e.g., Fuller, 1987).

In practice, however, the data may be composed of several subsamples with observable variables missing in the different subsamples. In the described regression model for example, we may have one sample with information only on Y_i and X_{i1} (X_{i2} missing), and a second sample with information only on X_{i1} and X_{i2} (Y_i missing). In the present paper we focus on the joint analysis of various samples of this type using Normal Theory (NT) methods while allowing the data to deviate from the normality assumption. NT methods are available in standard software for structural equation modeling, such as LISREL (Jöreskog & Sörbom, 1989), EQS (Bentler, 1989), LISCOMP (Muthén, 1987), the SAS Procedure CALIS (e.g., Hatcher, 1994). A clear advantage of the NT methods is that they require first- and second-order moments only.

The present paper relates with work of Arminger and Sobel (1990), where Pseudo-Maximum Likelihood (PML) analysis is applied to different data sets arising from missing data; Arminger and Sobel, however, do not address the issue of analyzing a functional model and the issue of correct NT inferences when data deviates from the normality assumption. The present work can also be seen as generalizing to multi-sample analysis and non-normality the results of Dham and Fuller (1986) that show the validity of NT-inferences in the case of a functional model. Our paper relates also with the asymptotic robustness results of Anderson and Amemiya (1988) and Amemiya and Anderson (1990); we extend such asymptotic robustness results to the case of multiple samples. Finally, we should mention that the following results are a specialization to the multivariate regression model set-up of more general results developed in Satorra (1995).

The structure of the paper is as follows. In Section 2 we present the general model set-up under consideration. Section 3 describes the NT generalized-least-squares analysis and provides a theorem which summarizes the major results of the paper. Finally, Section 4 presents a limited Monte Carlo study illustrating in finite samples and a specific model context the practical implications of the paper.

2 Multivariate Regression and Moment Structures

Let

$$\begin{cases} Y_i &=& Bx_i + v_i \\ X_i &=& \Lambda x_i + u_i, \end{cases} \qquad i = 1, \ldots, n, \tag{4}$$

where for individual i, $Y_i\,(p \times 1)$, $X_i\,(q \times 1)$ and $x_i\,(m \times 1)$ are respectively the values of the response, indicator and true (latent) regressor variables; $v_i\,(p \times 1)$ and $u_i\,(q \times 1)$ are respectively the values of the disturbance and measurement error terms; $B\,(p \times m)$ is a matrix of (unknown) parameters, and $\Lambda\,(q \times m)$ is a matrix of known coefficients. Assume that the v_i and u_i are independent and both iid with mean 0 and variance matrices Φ_v and Ψ_u respectively; furthermore, assume

$$n^{-1} \sum_{i=1}^{n} x_i x_i{}' \to \Phi_x, \tag{5}$$

as $n \to \infty$, where Φ_x is a positive-definite matrix and the convergence is in probability in the case of a structural model. We further assume that for each i: (structural model) x_i, v_i and u_i are mutually independent; (functional model) v_i and u_i are independent with Φ_v and Ψ_u unrelated with the x_i. We consider that Λ is a 0-1 matrix of full column rank, [2] and $\Psi_u = H\psi_u$, where H is a 0-1 matrix and ψ_u is a vector of unrestricted parameters. Note that by allowing a component of X_i and x_i to be constant to 1, we can accommodate as parameters of the model the intercepts of the regression equations and the mean vector of x_i. In the model set-up of the present paper, we impose that the u_i are normally distributed.

Clearly, (4) can be written as

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix} u_i + \begin{pmatrix} B \\ \Lambda \end{pmatrix} x_i + \begin{pmatrix} I \\ 0 \end{pmatrix} v_i; \tag{6}$$

or, alternatively, as

$$\begin{cases} z_i &=& \Lambda^\star z_i^\star + u_i^\star \\ z_i^\star &=& B^\star z_i^\star + v_i^\star, \end{cases} \tag{7}$$

where

$$z_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}, \quad z_i^\star = \begin{pmatrix} Y_i \\ x_i \end{pmatrix},$$

$$\Lambda^\star = \mathrm{diag}\,(I, \Lambda), \quad B^\star = \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}$$

and

$$u_i^\star = \begin{pmatrix} 0 \\ u_i \end{pmatrix}, \quad v_i^\star = \begin{pmatrix} v_i \\ x_i \end{pmatrix}.$$

(Here 0 and I denote zero and identity matrices respectively of dimensions clearly determined by the context.). We denote by $\ell \equiv (p + q)$ the dimension of z_i. We should note that expression (7) enables the direct use of standard software for

---

[2]In the context of a multiple indicator model, Λ is of the form $1_K \otimes I_m$ where 1_K denotes a vector of ones and K is the number of "replicates".

structural equation models for the purpose of model estimation and testing.

Now suppose that we have G different selection matrices T_g, $g = 1, \ldots, G$, where T_g is of dimension $\ell_g \times \ell$ with $\ell_g \leq \ell$. Suppose that instead of z_i we observe only

$$\tilde{z}_{ig} = T_g z_i,$$

where each i $(i = 1, \ldots, n)$ is associated to one of the matrices T_g. Let the cases i associated with T_g form the group (or subsample) g, and denote by n_g the size of the gth group $(g = 1, \ldots, G)$. For the validity of the asymptotic theory to be developed below, we require that G is "small" compared to n, so that $n_g/n \to c_g > 0$ as $n \to \infty$.

Define now the $(\ell_g \times \ell_g)$ uncentered sample cross-product moment matrix

$$S_g \equiv n_g^{-1} \sum_i \tilde{z}_{ig} \tilde{z}'_{ig},$$

where here $\sum_i$ denotes sum over the n_g cases in group g. Clearly, under the current model set-up,

$$S_g \xrightarrow{P} \Sigma_g,$$

where $\xrightarrow{P}$ denotes convergence in probability and Σ_g is a finite $\ell_g \times \ell_g$ matrix.

From (6), and the definition of $\tilde{z}_{ig}$, we have

$$\tilde{z}_{ig} = \Pi_{g0} u_i^{(g)} + \Pi_{g1} x_i^{(g)} + \Pi_{g2} v_i^{(g)}, \tag{8}$$

where

$$\Pi_{g0} \equiv T_g \begin{pmatrix} 0 \\ I \end{pmatrix}, \quad \Pi_{g1} \equiv T_g \begin{pmatrix} B \\ \Lambda \end{pmatrix}, \quad \Pi_{g2} \equiv T_g \begin{pmatrix} I \\ 0 \end{pmatrix}, \tag{9}$$

and where $u_i^{(g)}$ and $v_i^{(g)}$ are subvectors of u_i and v_i respectively so that Π_{g0} and Π_{g2} are of full column rank. Note that the vector $x_i^{(g)}$ is just the whole vector x_i.

The following moment structure for Σ_g can be derived from (8) and the current assumptions:

$$\Sigma_g = \Pi_{g0} \Phi_{g0} \Pi_{g0}' + \Pi_{g1} \Phi_{g1} \Pi_{g1}' + \Pi_{g2} \Phi_{g2} \Pi_{g2}', \tag{10}$$

where $\Phi_{g0} \equiv \Phi_{u^{(g)}}$, $\Phi_{g1} \equiv \Phi_{x^{(g)}}$ and $\Phi_{g2} \equiv \Phi_{v^{(g)}}$ are respectively the "population" moment matrices associated to $u_i^{(g)}$, $x_i^{(g)}$ and $v_i^{(g)}$. Consider now the following (unconstrained) vector of parameters

$$\vartheta_1 \equiv (\mathrm{v}'(\Phi_{v^{(1)}}), \ldots, \mathrm{v}'(\Phi_{v^{(G)}}), \mathrm{v}'(\Phi_{x^{(1)}}), \ldots, \mathrm{v}'(\Phi_{x^{(G)}}))', \tag{11}$$

where the $\mathrm{v}(\Phi_{v^{(g)}})$ may be of varying dimensions (determined by T_g) while the $\mathrm{v}'(\Phi_{x^{(g)}})$ are all of dimension $m(m+1)/2$ (For a symmetric matrix A, $\mathrm{v}(A)$ denotes the vectorized matrix A with the redundant elements due to symmetry suppressed.). Furthermore, define

$$\tau \equiv \begin{pmatrix} \mathrm{vec}(B) \\ \psi_u \end{pmatrix} \quad \text{and} \quad \vartheta \equiv \begin{pmatrix} \tau \\ \vartheta_1 \end{pmatrix}. \tag{12}$$

4

Let denote by $t^\star$ and t the dimensions of τ and ϑ respectively.

In relation with (9), (10), (11) and (12), consider the (multiple-group) moment structure

$$\Sigma_g = \Sigma_g(\vartheta), \quad g = 1, \ldots, G, \tag{13}$$

where $\Sigma_g(.)$ is the matrix-valued function of ϑ associated with (10), with $\Pi_{g\ell} = \Pi_{g\ell}(\tau)$ and $\Phi_{g0} = H\psi_u$ ($\ell = 1, 2, 3; g = 1, \ldots, G$) as specified above. Note that in this model specification the non-redundant elements of Φ_x and of Φ_v are unconstrained parameters (even across-groups) of the model (this is in fact a key assumption for the theorem to be presented in Section 3). [3]

We now consider the estimation and testing of the multiple-group moment structure (13). We will do so by fitting simultaneously the S_g to the $\Sigma_g(\vartheta)$. We will use a fitting function that is optimal under the following NT assumption: the z_i are iid normally distributed. In fact, the major import of the present paper is a theorem that identifies conditions under which the NT-inferences are correct even though the assumption NT does not hold. In structural equation modeling, the validity of NT inferences when the NT does not hold has been called asymptotic robustness (Anderson, 1987).

To facilitate notation, we define

$$s \equiv \text{vec}\{s_g; g = 1, \ldots, G\} \quad \text{and} \quad \sigma \equiv \text{vec}\{\sigma_g; g = 1, \ldots, G\},$$

where $s_g \equiv \text{v}(S_g)$ and $\sigma_g \equiv \text{v}(\Sigma_g)$, and we write (13) as

$$\sigma = \sigma(\vartheta), \tag{14}$$

where $\sigma(.)$ is a continuously differentiable vector-valued function of ϑ. Now we will fit s to $\sigma(\vartheta)$.

3   NT Generalized-Least-Squares

The NT Generalized-Least-Squares (NT-GLS) estimate of ϑ is defined as

$$\hat{\vartheta} \equiv \text{arg min}[s - \sigma(\vartheta)]'\hat{V}^\star[s - \sigma(\vartheta)], \tag{15}$$

where $\hat{V}^\star$ converges in probability to the following block-diagonal matrix:

$$V^\star \equiv \text{diag}\left\{c_g V_g^\star; g = 1, \ldots, G\right\},$$

where $\frac{n_g}{n} \to c_g$ when $n \to \infty$,

$$V_g^\star \equiv \frac{1}{2}D'(\Sigma_g{}^{-1} \otimes \Sigma_g{}^{-1})D, \tag{16}$$

---

[3]We should note that the results to be described below apply also when B and Λ are continuously differentiable functions of the subvector of parameters τ.

and D is the duplication matrix for symmetry of Magnus and Neudecker (1988). It is well known that under the NT assumption this NT-GLS estimate is asymptotically optimal. For seminal work on GLS estimation of covariance structures in single-sample analysis, see Browne (1974); NT-GLS estimation in multi-sample analysis is treated in Lee and Tsui (1982).

We need to impose the following identification assumption: $\sigma(\vartheta^\star) = \sigma(\vartheta)$ implies (locally) that $\vartheta^\star = \vartheta$, and the Jacobian matrix $\Delta = \frac{\partial \sigma(\vartheta)}{\partial \vartheta'}$ is of full column rank in a neighborhood of ϑ. This assumption is needed to ensure uniqueness and consistency of the estimate $\hat{\vartheta}$.

An alternative to the NT-GLS estimate is the PML approach. Under the NT assumption,

$$F_{ML} \equiv \sum_{g=1}^{G} \frac{n_g}{n} [\log \mid \Sigma_g(\vartheta) \mid + \text{trace}\{S_g \Sigma_g(\vartheta)^{-1}\} - \log \mid S_g \mid -\ell_g] \qquad (17)$$

is an affine transformation of the log-likelihood function. Thus, under general distribution assumption for the z_i, maximization of $F_{ML} = F_{ML}(\vartheta)$ yields the so-called PML estimate of ϑ. This PML estimate is in fact asymptotically equivalent to the NT-GLS estimate described above. For a comprehensive overview of PML estimation in structural equation modeling, see Arminger and Sobel (1990). Multi-sample analysis of structural equation models using the maximum likelihood method was first proposed by Jöreskog (1971) in the context of factor analysis. Now we will review the asymptotic variance matrix of the NT-GLS and PML estimates.

Let Γ denote the asymptotic variance matrix of the scaled vector of sample moments $\sqrt{n}s$. Under the NT assumption, the asymptotic variance matrix of $\sqrt{n}s$ is (e.g., Satorra, 1993)

$$\Gamma^\star = \text{diag}\left\{c_g^{-1} \Gamma_g^\star; \, g = 1, \ldots, G\right\},$$

where

$$\Gamma_g^\star \equiv \Omega_g - \Upsilon_g,$$
$$\Omega_g \equiv 2D^+(\Sigma_g \otimes \Sigma_g)D^{+\prime}, \qquad (18)$$
$$\Upsilon_g \equiv 2D^+(\mu_g \mu_g' \otimes \mu_g \mu_g')D^{+\prime}$$

and

$$\mu_g \equiv E(z_g),$$

where E denotes the expectation operator under the NT assumption. Furthermore, let

$$\Omega \equiv \text{diag}\left\{c_g^{-1} \Omega_g; \, g = 1, \ldots, G\right\}, \qquad (19)$$

and let $\hat{\Omega}$ denote the estimate of Ω obtained by substituting S_g for Σ_g and c_g^{-1} for $\frac{n}{n_g}$ in (18) and (19) respectively. Note that $V^\star = \Omega^{-1}$ (and that $\hat{V}^\star = \hat{\Omega}^{-1}$).

6

By using the standard delta-method, under general distribution assumption for the z_i, the asymptotic variance matrix of the NT-GLS and PML estimates is known to be (e.g., Satorra and Neudecker, 1994)

$$\text{avar}(\hat{\vartheta}) = (\Delta'V^\star\Delta)^{-1}\Delta'V^\star\Gamma V^\star\Delta(\Delta'V^\star\Delta)^{-1}. \tag{20}$$

Under the NT assumption we have

$$\text{avar}_{NT}(\hat{\vartheta}) = (\Delta'V^\star\Delta)^{-1}. \tag{21}$$

In addition to parameter estimation, we are interested in testing the goodness-of-fit of the model. The following goodness-of-fit test statistic will be used:

$$T^\star = n(s - \hat{\sigma})'(\hat{V}^\star - \hat{V}^\star\hat{\Delta}(\hat{\Delta}'\hat{V}^\star\hat{\Delta})^{-1}\hat{\Delta}'\hat{V}^\star)(s - \hat{\sigma}); \tag{22}$$

under NT, and when the model holds, $T^\star$ is asymptotically chi-square distributed with

$$r = \text{rank}(P\Gamma^\star P') \tag{23}$$

degrees of freedom (df), where $P = I - \Delta(\Delta'V^\star\Delta)^{-1}V^\star$ (Satorra, 1993).

Direct specialization to the present model set-up of results of Satorra (1995) yield the following theorem.

THEOREM 1 (cf., Satorra, 1995, Theorem 1)   *Under (6) and (13) and the current assumptions (NT is not included), the NT-GLS and PML estimates verify*

1. *$\hat{\vartheta}$ is a consistent estimate of ϑ.*

2. *The $t^\star \times t^\star$ leading principal submatrices of $\text{avar}(\hat{\vartheta})$ and $\text{avar}_{NT}(\hat{\vartheta})$ (recall (20) and (21)) coincide (that is, the asymptotic variance matrix $\text{avar}(\hat{\tau})$ is the same as under NT).*

3. *$\hat{\tau}$ is an efficient estimate within the class of GLS estimates of (15).*

4. *The asymptotic distribution of the goodness-of-fit test statistic $T_V^\star$ of (22) is chi-square with degrees of freedom r given by (23).*

Note that this theorem guarantees correct NT-inferences for the parameter vector τ, including the asymptotic efficiency, and also asymptotic chi-squaredness for the goodness-of-fit test, without the requirement for the z_i to be iid normally distributed. It is required, however, that for each i: (structural model) the x_i, v_i, u_i are mutually independent, not only uncorrelated; (functional model) the v_i, u_i are independent with the limit of the second-order moments of the x_i to be finite. With regard to the model specification, it is also required that the variances of the possibly non-normal constituents of the model, such as the x_i and v_i, are not constrained even across groups. See Satorra (1995) for full details of the proof of this theorem (in this reference, the theorem is formulated in a more general model context).

In the next section we present a limited Monte Carlo study to illustrate the performance in finite samples of the asymptotic results of the paper.

4 Monte Carlo illustration

Consider the regression with errors-in-variables model set-up of equations (1) and (3) of Section 1, from which we simulate two-sample data of sizes $n_1 = 800$ and $n_2 = 400$, with X_{i2} missing in the first subsample and Y_i missing in the second subsample. The values of x_i and v_i are simulated as independent chi-square distributed of 1 degree of freedom (i.e., a highly skewed non-normal distribution) conveniently scaled; the u_{i1} and u_{i2} are simulated as uncorrelated iid normally distributed. Each Monte Carlo run consisted on replicating 1000 times the steps of (a) generating two-sample data, and (b) estimating (for each two-sample data) a specified model using the NT-GLS approach of Section 2.

The three models differ with respect to the restrictions imposed on equality across groups of means and variances of x_i. Two of the models considered are structural, one is a functional model.

Unrestricted structural model (USM): The mean of x_i is a parameter restricted to be equal across groups; the variance of x_i is a parameter different for each group.

Restricted structural model (RSM): The variance of x_i is a parameter restricted to be equal across groups; the mean of x_i is a parameter different for each group.

Unrestricted functional model (UFM): the (pseudo) mean and variance of x_i are (pseudo) parameters different for each group.

In all the three models, the variance of v_i in the first group and the variance of u_{i2} in the second group are unrestricted parameters of the model; the parameters α and β and the variance of u_{i1} are restricted to be equal across groups. With regard to the degrees of freedom of $T^\star$, the difference between the number of distinct moments (10 moments) and the number of free parameters in the model, yields 1 df in the case of models RSM and USM, and 2 df in the case of UFM. Tables 1-3 present the Monte Carlo results for models USM, UFM and RSM respectively.

Note that USM, UFM and RSM are in fact three alternative correctly specified models, since the restrictions of equality of parameters across groups imposed by the models are in fact true in the population. Only USM and UFM, however, satisfy the conditions for invoking Theorem 1; RSM does not satisfy such conditions, since it restricts across groups the variance of x_i, a non-normal constituent of the model.

In all the three models, we expect consistency of the parameter estimates. This is in fact corroborated by comparing the second and third columns of Tables 1-3. The second column corresponds to the population values of the parameters and the third column corresponds to the mean (across 1000 replications) of NT-GLS parameter estimates.

In the case of models USM and UFM, i.e. Tables 1 and 2 respectively, Theorem 1 guarantees asymptotic correctness of NT-inferences for the subvector of parameters τ: the parameters α, β, $\sigma_{u_1}^2$, $\sigma_{u_2}^2$ $\mu(x_i^{(1)})$ and $\mu(x_i^{(2)})$ in USM, and α, β, $\sigma_{u_1}^2$

and $\sigma^2_{u_2}$ in UFM. Thus, the rows of Tables 1-2 corresponding to the components of τ should show: (a) the Monte Carlo standard deviation of parameter estimates (column $sd(\hat{\vartheta})$) close to the Monte Carlo mean of NT standard errors (column $E(se)$); (b) the Monte Carlo variance of $d \equiv \frac{\hat{\vartheta} - \vartheta}{se(\hat{\vartheta})}$ (column $var(d)$) close to unity; and, (c) the Monte Carlo two-sided tails of d close to nominal values (columns 5%, 10% and 20%). Theorem 1 ensures also the asymptotic chi-squaredness of the goodness-of-fit test $T^\star$; thus, in Tables 1-2, the Monte Carlo mean (and variance) of $T^\star$ should be close to the degrees of freedom (twice the degrees of freedom), and the tails of the Monte Carlo distribution of $T^\star$ should be close to the nominal values. These expectations are clearly corroborated by the results shown in Tables 1 and 2. This correct performance of the NT-inferences is remarkable despite clear non-normality of the z_i.

In the case of model RSM, however, Table 3 shows that NT-inferences for parameters deviate dramatically from the correct performance (only NT-inferences for the means $\mu(x_i^1)$ and $\mu(x_i^2)$ seem to be appropriate), and the empirical distribution of $T^\star$ also deviates highly from the desired chi-square distribution. Note that this lack of robustness of NT-inferences resulted from the restriction of equality across groups of the variance of the x_i, a restriction which is in fact true in the population generating the data.

References:

AMEMIYA, Y. and 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. In T. Pukkila & S. Puntanen (edits.) *Proceedings of the Second International Conference in Statistics, Pp. 9-36.* Tampere, Finland: University of Tampere.

ANDERSON, T.W. and AMEMIYA, Y. (1988): The asymptotic normal distribution of estimates in factor analysis under general conditions." *The Annals of Statistics, 16, 759-771.*

ARMINGER, G. and 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. (1974):. Generalized least squares estimates in the analysis of covariance structures. *South African Statistical Journal, 8, 1-24.*

DHAM, P.F. and FULLER, W.A. (1986): Generalized Least Squares Estimation of the Functional Multivariate Linear Errors-in-variables Model. *Journal of Multivariate Analysis, 19, 132-141.*

FULLER, W.A. (1987): *Measurement Error Models.* New York: Wiley.

GLESER, L.J. (1992): The Importance of Assessing Measurement Reliability in Multivariate Regression. *Journal of the American Statistical Association, 87, 696-707.*

HATCHER, L. (1994): A step-by-step approach to using the SAS System for factor analysis and structural equation modeling, Cary, NC SAS Institute.

JÖRESKOG, K (1971): Simultaneous factor analysis in several populations *Psychometrika, 57, 409-426.*

JÖRESKOG, K. and SÖRBOM, D. (1989): LISREL 7, A Guide to the Program and Applications (2nd ed.). Chicago: SPSS Inc.

LEE, S.-Y., and 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.

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

SATORRA, A. (1993): Asymptotic robust inferences in multiple-group analysis of augmented-moment structures", in pp. 211-229 *Multivariate Analysis: Future Directions 2,* (edits. by C.M. Cuadras and C.R. Rao). Elsevier: Amsterdam.

SATORRA, A. (1995): Asymptotic robust inferences in Multi-Sample Analysis of Multivariate Linear Relations, Economics Working Paper 126, Universitat Pompeu Fabra, Barcelona

SATORRA, A. and NEUDECKER, H. (1994): On the asymptotic optimality of alternative minimum-distance estimates in linear latent-variable models. *Econometric Theory, 10, 867-883.*

Table 1: Monte Carlo results for USM model, where $\sigma^2(x_i^{(1)})$ and $\sigma^2(x_i^{(2)})$ are unrestricted parameters and $\mu(x_i^{(1)}) = \mu(x_i^{(2)})$.

para.	true val.	distribution of estimates						
		$E(\hat{\vartheta})^a$	$sd(\hat{\vartheta})^b$	$E(se)^c$	$var(d)^d$	5% ^{e}	10%	20%
$\sigma^2(v_i^{(1)})$	1.00	1.00	.25	.23	1.24	7.10	13.20	23.50
$\sigma^2(x_i^{(1)})$	1.00	1.00	.15	.08	3.43	29.40	37.10	49.50
$\sigma^2(x_i^{(2)})$	1.00	1.00	.20	.08	5.62	39.80	49.90	59.40
$\sigma^2(u_{2i}^{(j)})$	.30	.30	.05	.05	1.01	5.40	10.40	19.90
$\sigma^2(u_{2i}^{(2)})$	.40	.40	.05	.05	.89	3.70	8.20	16.60
β	2.00	2.00	.12	.12	1.04	4.90	9.20	20.20
α	1.00	1.00	.37	.37	1.03	4.40	9.00	20.50
$\mu(x^{(j)})$	3.00	3.00	.03	.03	1.05	4.50	10.10	20.90

distribution of $T^\star$ (df=2)

mean	var	5%	10%	20%
2.12	4.44	6.40	11.40	21.40

^{a} Monte Carlo mean of estimates

^{b} Monte Carlo standard deviation of estimates

^{c} Monte Carlo mean of NT standard errors

^{d} Monte Carlo variance of $d \equiv \frac{\hat{\vartheta} - \vartheta}{se(\hat{\vartheta})}$

^{e} Monte Carlo nominal 5%, 10% and 20% two-sided tails for d

Table 2: Monte Carlo results for UFM model, where $\sigma^2(x_i^{(1)})$, $\sigma^2(x_i^{(2)})$, $\mu(x_i^{(1)})$ and $\mu(x_i^{(2)})$ are unrestricted (pseudo) parameters.

		distribution of estimates						
para.	true val.	$E(\hat{\vartheta})^a$	$sd(\hat{\vartheta})^b$	$E(se)^c$	$\mathrm{var}(d)^d$	5% [e]	10%	20%
$\sigma^2(v_i^{(1)})$	1.00	1.00	.20	.16	1.41	8.70	16.80	26.60
$\sigma^2(u_{1i}^{(j)})$	.30	.30	.03	.03	1.08	6.20	10.90	21.80
$\sigma^2(u_{2i}^{(2)})$	.40	.40	.03	.04	.99	5.10	10.50	19.70
β	2.00	2.00	.08	.08	1.04	6.10	10.40	21.60
α	1.00	1.00	.17	.17	1.00	5.10	9.50	18.40

distribution of $T^\star$ (df=1)				
mean	var	5%	10%	20%
1.05	2.27	5.40	10.00	19.90

[a] Monte Carlo mean of estimates

[b] Monte Carlo standard deviation of estimates

[c] Monte Carlo mean of NT standard errors

[d] Monte Carlo variance of $d \equiv \frac{\hat{\vartheta}-\vartheta}{se(\hat{\vartheta})}$

[e] Monte Carlo nominal 5%, 10% and 20% two-sided tails for d

Table 3: Monte Carlo results for RSM model, where $\mu(x_i^{(1)})$ and $\mu(x_i^{(2)})$ are unrestricted parameters and $\sigma^2(x_i^{(1)}) = \sigma^2(x_i^{(2)})$.

para.	true val.	distribution of estimates						
		$E(\hat{\vartheta})^a$	$sd(\hat{\vartheta})^b$	$E(se)^c$	$var(d)^d$	5% [e]	10%	20%
$\sigma^2(v_i^{(1)})$	1.00	.99	.29	.22	1.72	13.50	20.80	32.80
$\sigma^2(x_i^{(j)})$	1.00	.98	.12	.06	4.37	34.90	43.60	55.30
$\sigma^2(u_{1i}^{(j)})$	.30	.30	.06	.05	1.49	10.80	17.30	29.50
$\sigma^2(u_{2i}^{(2)})$	.40	.40	.05	.05	1.12	5.50	10.90	22.70
β	2.00	2.01	.15	.11	1.85	14.80	23.20	35.00
α	1.00	.98	.45	.33	1.84	15.10	22.70	34.80
$\mu(x^{(1)})$	3.00	3.00	.04	.04	1.06	5.50	12.20	20.90
$\mu(x^{(2)})$	3.00	3.00	.05	.05	1.07	6.40	10.70	21.10

distribution of $T^\star$ (df=2)

mean	var	5%	10%	20%
5.68	40.81	33.20	41.30	53.80

[a] Monte Carlo mean of estimates

[b] Monte Carlo standard deviation of estimates

[c] Monte Carlo mean of NT standard errors

[d] Monte Carlo variance of $d \equiv \frac{\hat{\vartheta} - \vartheta}{se(\hat{\vartheta})}$

[e] Monte Carlo nominal 5%, 10% and 20% two-sided tails for d

13