British Journal of Mathematical and Statistical Psychology (1989) **42**, 55–62 *Printed in Great Britain*

© 1989 The British Psychological Society

Multiple-group structural modelling with non-normal continuous variables

Bengt Muthén

Graduate School of Education, University of California, Los Angeles, CA 90024, USA

An estimator is considered for mean and covariance structure analysis of non-normal continuous variables. This estimator extends the ADF estimator of Browne (1982, 1984) to include structured means. The behaviour of the estimator is compared to that of normal-theory generalized least-squares in simulated data.

1. Introduction

Simultaneous factor analysis in several groups has been treated by Joreskog (1971) and Sorbom (1974). These authors studied maximum likelihood estimation under multivariate normality, testing various forms of invariance across groups by means of chi-square. In this paper we are concerned with non-normal, continuous variables which are common in social science data. In recent years there has been an increasing interest in the robustness of multivariate hypothesis testing to deviations from normality; see, for instance, Ito (1969) and Mardia (1970, 1971, 1974). Often tests on means are found to be rather robust, whereas tests involving covariances are less robust (see, for example, Mardia, 1971). More recently, Browne (1982, 1984) has suggested generalized least squares covariance structure estimators for non-normal data. Here we study one such type of estimator as it applies to the multiple-group structural modelling situation. It is shown that the normal theory approach gives unduly large chi-square values as compared to the present approach, when testing invariance hypotheses.

2. Preliminaries

While Joreskog and Sorbom studied maximum-likelihood (ML) estimation, we will briefly describe their approach as it applies to generalized least squares (GLS) estimation under multivariate normality. In the large sample situations we will be concerned with, the ML vs. GLS difference is negligible, since these estimators are asymptotically equivalent (Browne, 1974).

Consider a random vector of variables y $(p \times 1)$, observed in group (population) g,

$$y^{(g)} = \nu^{(g)} + \Lambda^{(g)}\eta^{(g)} + \varepsilon^{(g)}, \tag{1}$$

\# 26

where $\nu^{(g)}(p \times 1)$ is a vector of location parameters, $\Lambda^{(g)}(p \times m)$ is a matrix of regression (loading) parameters, $\eta^{(g)}(m \times 1)$ is a random vector of latent variables (factors), and $\varepsilon^{(g)}(p \times 1)$ is a vector of residuals with covariance matrix $\mathbf{\Theta}^{(g)}$.

Consider the structural equations

$$\eta^{(g)} = \alpha^{(g)} + \mathbf{B}^{(g)}\eta^{(g)} + \zeta^{(g)}, \tag{2}$$

where $\alpha^{(g)}(m \times 1)$ is a vector of intercepts, $\mathbf{B}^{(g)}(m \times m)$ is a matrix of slopes for which $I - \mathbf{B}^{(g)}$ is non-singular, and $\zeta^{(g)}$ is a random vector of residuals with covariance matrix $\mathbf{\Psi}^{(g)}$. Ordinary assumptions give (cf. Muthén, 1984)

$$E(y^{(g)}) = \nu^{(g)} + \Lambda^{(g)}(I - \mathbf{B}^{(g)})^{-1}\alpha^{(g)} \tag{3}$$

$$V(y^{(g)}) = \Lambda^{(g)}(I - \mathbf{B}^{(g)})^{-1}\mathbf{\Psi}^{(g)}(I - \mathbf{B}^{(g)})'^{-1}\Lambda^{(g)\prime} + \Theta^{(g)}. \tag{4}$$

Furthermore, for each group g, let $\sigma_1^{(g)}(p \times 1)$ denote the vector valued function of the mean vector parameters given in (3), and $\sigma_2^{(g)}(p(p+1)/2 \times 1)$ the vector valued function corresponding to the distinct covariance matrix elements of (4). Let $\sigma^{(g)\prime} = (\sigma_1^{(g)\prime}\sigma_2^{(g)\prime})$. Assuming multivariate normality and independent random sampling from G groups, Joreskog and Sorbom used the mean vectors and covariance matrices in the groups in order to estimate the model parameters. For each group, let $s_1^{(g)}(p \times 1)$ and $s_2^{(g)}(p(p+1)/2 \times 1)$ denote these mean and covariance elements, respectively. Let $s^{(g)\prime} = (s_1^{(g)\prime}s_2^{(g)\prime})$.

As applied to GLS, the Joreskog–Sorbom approach results in minimizing the fitting function (cf. Muthén, 1983, and references therein)

$$F = \sum_{g=1}^{G} (s^{(g)} - \sigma^{(g)})' W^{(g)-1}(s^{(g)} - \sigma^{(g)}), \tag{5}$$

with respect to the model parameters. Here, $W^{(g)}$ is chosen as a consistent estimator of the asymptotic covariance matrix of $s^{(g)}$. Let $W^{(g)}$ be partitioned according to $s^{(g)}$,

$$W^{(g)} = \begin{bmatrix} W_{11}^{(g)} & \text{symm.} \\ W_{21}^{(g)} & W_{22}^{(g)} \end{bmatrix}. \tag{6}$$

Under multivariate normality, $W_{11}^{(g)} = N^{(g)-1}\mathbf{S}^{(g)}$, $W_{21}^{(g)} = 0$, and $W_{22}^{(g)} = \mathbf{K}(S^{(g)} \times S^{(g)})\mathbf{K}'$ where $N^{(g)}$ is the sample size in group g, $\mathbf{S}^{(g)}$ is the sample covariance matrix, and $\mathbf{K}$ is a constant matrix selecting elements (see, for example, Browne, 1974; Kendall & Stuart, 1977).

Let the vector of all parameters be denoted $\theta(s \times 1)$. A particular hypothesis considers restrictions on θ so that there are only r free and distinct parameters to be estimated. Of particular interest in multiple-group situations are hypotheses regarding equality of certain θ-parameters from different groups, especially

$$\nu^{(1)} = \nu^{(2)} = \cdots = \nu^{(G)} = \nu \tag{7}$$

$$\Lambda^{(1)} = \Lambda^{(2)} = \cdots = \Lambda^{(G)} = \Lambda, \tag{8}$$

implying measurement parameter invariance (see, for example, Sorbom, 1974). Let t denote the total number of elements in the $\sigma^{(g)'}s$, $t = G(p + p(p+1)/2)$. It follows that any such hypothesis can be tested by the value F_0 of (5), calculated at the minimum, since, asymptotically, F_0 follows a chi-square distribution with $t - r$ degrees of freedom (see, for example, Browne, 1974, 1982).

In a single-group ($G = 1$) covariance structure context, Browne (1982, 1984) considered the GLS type fitting function

$$F = (s_2 - \sigma_2)' W^{-1} (s_2 - \sigma_2), \tag{9}$$

where in this case the elements of W represent a consistent estimator of the asymptotic covariance matrix of s_2 using

$$(N-1)\operatorname{cov}(s_{ij}, s_{kl}) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} + \frac{N-1}{N}\kappa_{ijkl}. \tag{10}$$

Under normality of y, the fourth-order cumulants, κ_{ijkl} of (10) would all be zero and $W_{22}^{(g)}$ of (6) would obtain the simple Kronecker structure given beneath (6). In Browne's approach, the normality assumption is avoided. We will use Browne's name for this estimator: ADF ('asymptotically distribution free'). For details on the ADF estimation including its asymptotic behaviour, see Browne (1982, 1984).

3. Multiple-group ADF

The ADF approach would seem to be directly applicable to the multiple-group problem outlined above. Since $W_{11}^{(g)}$ of (6) does not change when the normality assumption is relaxed, it only remains to find $W_{21}^{(g)}$, i.e. a consistent estimator of the asymptotic covariances between s_2 and s_1, which are no longer zero.

Let μ_i denote the mean variable i, and

$$\mu_{ijk} = E(y_i - \mu_i)(y_j - \mu_j)(y_k - \mu_\kappa), \tag{11}$$

a multivariate third-order moment about the mean. We note that

$$\mu_{ijk} = \kappa_{ijk}, \tag{12}$$

where κ_{ijk} is a third-order cumulant (cf. Kendall & Stuart, 1977; Kaplan, 1952). Asymptotically,

$$N\operatorname{cov}[(s_2)_{ij}, (s_1)_k] = \kappa_{ijk}, \tag{13}$$

where the i, j subscript on the s_2 vector refers to one of its $p(p+1)/2$ distinct covariance matrix elements and the k subscripts on the s_1 vector refers to one of its p elements.

A consistent estimator of the s_2, s_1 covariance matrix is formed as follows (cf. Mooijart, 1985). Consider the p-dimensional vector y^* for observation i,

$$y_i^{*\prime} = (y_{11} - \bar{y}_1 \ldots y_{pi} - \bar{y}_p), \tag{14}$$

where $\bar{y}_i$ is the sample mean of y_i and creates the $p(p+1)/2$-dimensional vector a_i,

$$a_i' = *\,(y_{1i}^* y_{1i} {}^* y_{2i}^* y_{1i} {}^* y_{2i}^* y_2 {}^* y_{3i}^* y_{1i}$$

$$\ldots y_{pi}^* y_{p-1,i} y_{pi}^* y_{pi}^*). \tag{15}$$

For each group g, deleting the group index, we may then create W_{21} as

$$W_{21} = N^{-2} \sum_{i=l}^{N} a_i y_i^{*\prime}, \tag{16}$$

where N is the sample size for this particular group.

Above we have defined W for each group g in the expression for the fitting function F in (5). With this choice of W, we may term this GLS type estimation 'multiple-group ADF'. The asymptotic properties of this estimator are analogous to those of the single-group ADF, described by Browne (1982, 1984). Note that for each group the sample size must be larger than $p + p(p+1)/2$ to ensure a non-singular $W^{(g)}$. Fortunately there is a programmed algorithm available for finding the minimum of F given a specific model structure of the kind considered here. This is the LISCOMP program (Muthén, 1987; see also Muthén, 1984). A fitting function such as in (5) is considered, where a suitable weight matrix can be read by the program. In addition to parameter estimates, an asymptotic chi-square measure of model fit and asymptotic standard errors of the estimates are produced.

4. Applying multiple-group ADF

Simulated, non-normal data will now be considered in order to study the behaviour of the multiple-group ADF as compared to the normal theory GLS for multiple groups. For each of two groups, the random data generation was as follows. A factor model for nine observed variables and three correlated factors was chosen with three indicators per factor and a simple measurement structure. Parameter values are given in Table 2. To reflect commonly observed data in the social and behavioural sciences, each observed variable was chosen to be discrete with ten scale steps scored 0 to 9, where for simplicity the variables were taken to have equal univariate distributions. In the population the category percentages for each variable were chosen to give a sizeable skew and kurtosis: 2%, 2%, 2%, 2%, 2%, 3%, 5%, 14%, 51%, 17%. This choice resulted in a univariate skew and kurtosis of 2.11 and 4.10 respectively. The population means were all 7.25, the variances 3.91, and the correlations were 0.39 and 0.18 for variables within and between factors, respectively. Multivariate normal variables were used to generate the observed variables by categorization, yielding the

Table 1. Chi-square model: ν and λ invariant. All variables have equal skew. Degrees of freedom = 60

	Case 1	Case 2
N_1	1050	700
N_2	1050	1400
ADF		
Mean	64.07	64.31
Variance	125.71	140.83
Reject freq.	9	11
GLS		
Mean	84.91	85.16
Variance	225.09	261.30
Reject freq.	65	62

Mean and variance are calculated from 100 replications. Reject freq. gives the number of times the chi-square statistic exceeded the 5 per cent critical value (79.1).

Table 2. Parameter estimates[a]. Case 1: $N_1 = 1050$, $N_2 = 1050$. Model: ν and λ invariant. All variables have equal skew

Parameter		True value	ADF Group 1	ADF Group 2	GLS Group 1	GLS Group 2
Intercept	ν	7.254	7.304 (0.7)[b]	inv.	7.256 (0.0)	inv.
Loading	λ	1.000	1.003	inv.	1.003	inv.
Error variance	θ	2.380	2.236 (−6.2)	2.231 (−6.3)	2.329 (−2.1)	2.325 (−2.3)
Factor variance	ψ_{ii}	1.530	1.446 (−5.5)	1.449 (−5.29)	1.521 (−0.59)	1.524 (−0.4)
Factor covariance	ψ_{ij}	0.712	0.633 (−11.1)	0.632 (−11.2)	0.697 (−2.11)	0.698 (−2.0)
Factor mean	α	0.000	fixed	0.004	fixed	0.003

[a] Parameter estimates are the average of the free parameters within parameter types.
[b] Per cent over- or underestimation relative to the true value.

above features. The two groups were taken to have the same parameter values. This means that (3), (4), (7), and (8) hold for these data.

Two cases were studied. In Case 1, 1050 random observations were taken in both groups, while for Case 2, 700 and 1400 observations were chosen, respectively. The use of unequal sample sizes in the two groups would seem to reflect real data situations. It is also known that tests on covariance matrices are sensitive to unequal group sizes (see, for example, Mardia, 1971).

The random sampling procedure was repeated 100 times within the LISCOMP program. For each replication, both the normal theory GLS chi-square and the ADF chi-square was used to test (7) and (8). The number of degrees of freedom for this hypothesis is 60. The results on chi-square are given in Table 1. It is seen that the ADF chi-square results are closer to expectation than the GLS chi-square results.

Tables 2 and 3 give the parameter estimates for Cases 1 and 2. ADF shows a

Table 3. Parameter estimates[a]. Case 2: $N_1=700$, $N_2=1400$. Model: ν and λ invariant. All variables have equal skew

Parameter		True value	ADF Group 1	ADF Group 2	GLS Group 1	GLS Group 2
Intercept	ν	7.254	7.332 (1.1)[b]	inv.	7.259 (0.1)	inv.
Loading	λ	1.000	1.001 (0.1)	inv.	1.002 (0.2)	inv.
Error variance	θ	2.380	2.168 (−8.9)	2.269 (−4.7)	2.308 (−3.0)	2.336 (−1.8)
Factor variance	ψ_{ii}	1.530	1.393 (−9.0)	1.477 (−3.5)	1.508 (−1.4)	1.531 (−0.1)
Factor covariance	ψ_{ij}	0.712	0.595 (−16.4)	0.652 (−8.4)	0.690 (−3.1)	0.700 (−0.2)
Factor mean	α	0.000	fixed	−0.039	fixed	−0.001

[a]Parameter estimates are the average of the free parameters within parameter types.
[b]Per cent over- or underestimation relative to the true value.

Table 4. Sampling variability[a]. Case 1: $N_1=1050$, $N_2=1050$. Model: ν and λ invariant. All variables have equal skew

Parameter		ADF Group 1	ADF Group 2	GLS Group 1	GLS Group 2
Intercept	ν	0.052 0.057	inv. inv.	0.054 0.054	inv. inv.
Loading	λ	0.074 0.080	inv. inv.	0.056 0.074	inv. inv.
Error	θ	0.193 0.203	0.193 0.201	0.140 0.216	0.140 0.220
Factor variance	ψ_{ii}	0.190 0.198	0.191 0.216	0.142 0.193	0.143 0.213
Factor covariance	ψ_{ij}	0.096 0.107	0.095 0.096	0.081 0.105	0.081 0.100
Factor mean	α	fixed	0.064 0.070	fixed	0.066 0.066

[a]The two entries are:
mean of estimated standard errors;
empirical standard deviations of estimates.

somewhat larger bias than GLS and the bias is larger with unequal sample sizes in the groups.

Tables 4 and 5 give the sampling variability for Cases 1 and 2. The empirical variability is somewhat larger overall for ADF than for GLS. At the same time, the estimated ADF standard errors are closer to the empirical variability than are the estimated GLS standard errors.

In conclusion, we find in these examples that normal theory GLS chi-square

Table 5. Sampling variability[a]. Case 2: $N_1=700$, $N_2=1400$. Model: ν and λ invariant. All variables have equal skew

Parameter		ADF Group 1	ADF Group 2	GLS Group 1	GLS Group 2
Intercept	ν	0.060	inv.	0.064	inv.
		0.067	inv.	0.062	inv.
Loading	λ	0.074	inv.	0.056	inv.
		0.079	inv.	0.074	inv.
Error	θ	0.220	0.177	0.167	0.124
		0.243	0.184	0.261	0.194
Factor variance	ψ_{ii}	0.208	0.180	0.159	0.133
		0.226	0.197	0.217	0.195
Factor covariance	ψ_{ij}	0.108	0.086	0.096	0.072
		0.128	0.091	0.122	0.089
Factor mean	α	fixed	0.067	fixed	0.070
			0.074		0.069

[a]The two entries are:
mean of estimated standard errors;
empirical standard deviations of estimates.

testing leads to too frequent rejections and that normal theory GLS standard errors underestimate actual variability in the estimates. The multiple-group ADF estimator, however, performs reasonably well.

Acknowledgement

An earlier version of this paper was presented at the American Statistical Association meetings, 13–16 August, 1984, in Philadelphia, Pennsylvania. This research was supported by grant no. SES-8312583 from the National Science Foundation.

References

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1–24. Reprinted in D. J. Aigner & A. S. Goldberger, (Eds), *Latent Variables in Socio-economic Models*. Amsterdam: North-Holland, 1977.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*. Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62–82.

Ito, K. (1969). On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. In P. R. Krishnaiah (Ed), *Proceedings of the International Symposium on Multivariate Analysis*, vol. 2. New York: Academic Press.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, **36**, 409–426.

Kaplan, E. L. (1952). Tensor notation and the sampling cumulants of *k*-statistics. *Biometrika*, **39**, 319–323.

Kendall, M. & Stuart, A. (1977). *The Advanced Theory of Statistics*, vol. 1, 4th ed. New York: Macmillan.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

Mardia, K. V. (1971). The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika*, **58**, 105–121.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankya, Series B*, **36**, 115–128.

Mardia, K. V. (1983). Assessment of multinormality and the robustness of Hotelling's T^2 test. *Applied Statistics*, **24**, 163–171.

Mooijart, A. (1985). A note on computationally efficiency in asymptotically distribution-free correlational models. *British Journal of Mathematical and Statistical Psychology*, **38**, 112–115.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22, 43–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, **49**, 115–132.

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

Sorbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, **27**, 229–239.

Received 29 September 1987; revised version received 22 August 1988