A SIMPLE GAUSS-NEWTON PROCEDURE FOR COVARIANCE STRUCTURE ANALYSIS WITH HIGH-LEVEL COMPUTER LANGUAGES

ROBERT CUDECK

UNIVERSITY OF MINNESOTA

KELLI J. KLEBE

UNIVERSITY OF COLORADO, COLORADO SPRINGS

SUSAN J. HENLY

COLLEGE OF NURSING
UNIVERSITY OF NORTH DAKOTA

An implementation of the Gauss-Newton algorithm for the analysis of covariance structures that is specifically adapted for high-level computer languages is reviewed. With this procedure one need only describe the structural form of the population covariance matrix, and provide a sample covariance matrix and initial values for the parameters. The gradient and approximate Hessian, which vary from model to model, are computed numerically. Using this approach, the entire method can be operationalized in a comparatively small program. A large class of models can be estimated, including many that utilize functional relationships among the parameters that are not possible in most available computer programs. Some examples are provided to illustrate how the algorithm can be used.

Key words: covariance structures, Gauss-Newton method, simplex models, second order factor analysis, dichotomized variables.

A simple method for estimating structural equation models (e.g., Bollen, 1989) using high-level computer languages is described. Following Lee and Jennrich (1984), and Browne and Du Toit (1992), the procedure is formulated so that the user specifies only the covariance structure, sample data, and initial values for the parameters. While the numerical method is a simple one, it is straightforward to use to estimate a variety of standard models such as patterned covariance structures, factor analysis models, and latent variable regression models. Furthermore, some models can be estimated with this method that cannot be handled using widely available commercial programs such as LISREL (Jöreskog & Sörbom, 1989; but others also, as reviewed by Long, 1990). Some structures that are awkward, cumbersome or time consuming to set up using complicated reparameterization in LISREL (Green & Palmquist, 1991; Rindskopf, 1983, 1984; Wothke & Browne, 1990), and which may not be feasible for large problems, may be easier with this method.

There are several motivations for considering a general algorithm of this kind. The first arises as a consequence of the diversity of structures in the class. It is surprising how a small change to a model may make it difficult or impossible to fit using most

available programs (e.g., Bentler & Lee, 1983). As in nonlinear regression (e.g., Pickle, 1991), a very general program is required so that the largest possible number of models can be studied. The algorithm outlined is applicable to a wider class of models than can be estimated with most programs.

A second motivation for a procedure of this kind is methodological. Developments in covariance structure analysis have accumulated at a steady rate in recent years, and show no indication of dropping off. Yet there is often considerable delay between the publication of a new technique and its incorporation into the major programs. Furthermore, computer programs are idiosyncratic. One software author may include a particular goodness-of-fit index or other computational idea for structural models that another author ignores. Methodological progress is most rapid when the basic calculations are understood by a wide community of researchers.

A third motivation is pedagogical. Students of quite different levels of experience have found the exercise of implementing and applying the algorithm valuable.

Several approaches based on general computer programs have already been presented for covariance structure analysis. Typically, one supplies Fortran statements to describe the model. For example, Browne and Du Toit (1992) introduced a program that handles the widest range of problems yet defined for single populations, including arbitrary models for mean structures. An important feature of this program is its use of a very efficient method for nonlinear constraints that is much more flexible than has yet been implemented. Fraser (1979) described a program based on a model of McDonald (1978) in which nonstandard structures are specified via user-supplied subroutines for both the model and its derivatives. Lee and Jennrich (1984) have shown how the subprogram PAR of the BMDP system, a program for nonlinear regression, can be applied to covariance structure analysis. PAR can produce generalized least squares and ordinary least squares estimates, but it does not seem possible to use this method for maximum likelihood estimation. (To produce maximum likelihood estimates as iteratively reweighted GLS estimates requires that the GLS weight matrix be set equal to the current value of the estimated covariance matrix at the beginning of each iteration; that is, to set $W = \Sigma(\gamma)$ in (2) below. But then the gradient and approximate Hessian must be computed *ignoring* the dependency of W on γ. PAR does not do this, but instead repeatedly updates the weight matrix during each iteration for the secant approximations to the gradient. The resulting procedure gives true generalized least squares estimates, Swain, 1975, not normal theory maximum likelihood.)

The approach taken here is similar in spirit to these methods, but is based instead on high-level languages such as Gauss, Speakeasy, or Matlab. We have used Proc IML in the SAS system (SAS/IML, 1985) because it is widely available on mainframe computers. Therefore, additional special purpose hardware or software are not required. The hallmark of these languages is that their library of intrinsic functions includes linear algebraic operators in addition to standard scalar functions. This makes it possible to write extremely compact programs that, in a few lines, duplicate the work of thousands of lines of code of a lower level language. Similar to Browne and Du Toit (1992) and to Lee and Jennrich (1984), derivatives needed for the algorithm are computed numerically. The method is illustrated by discussing several models in detail.

Overview of the Procedure

The basic problem in the analysis of covariance structures is to estimate the parameters of a model, $\gamma = (\gamma_1, \ldots, \gamma_q)'$, by minimizing the difference between the covariance matrix implied by the model, $\Sigma = \Sigma(\gamma)$, and a sample covariance matrix, S,

based on $N = n - 1$ p-variate observations, $\mathbf{X}_1, \ldots, \mathbf{X}_N$. The most frequently used discrepancy functions are normal theory maximum likelihood,

$$M(\Sigma, \mathbf{S}) = \text{tr } (\Sigma^{-1}\mathbf{S}) - \ln |\Sigma^{-1}\mathbf{S}| - p, \tag{1}$$

generalized least squares (GLS),

$$G(\Sigma, \mathbf{S}) = \frac{1}{2} \text{ tr } \{[\mathbf{W}(\mathbf{S} - \Sigma)]^2\}, \tag{2}$$

where typically $\mathbf{W} = \mathbf{S}^{-1}$, and ordinary least squares

$$L(\Sigma, \mathbf{S}) = \frac{1}{2} \text{ tr } (\mathbf{S} - \Sigma)^2.$$

In general, these functions are minimized by means of an iterative method. Several different algorithms can be used for this problem (e.g., Thisted, 1988, sec. 4), although not all are equally efficient. In an important paper, Lee and Jennrich (1979) showed that the classical Fisher scoring method for minimizing (1) is equivalent to the Gauss-Newton method that is generally recommended for weighted least squares problems such as (2). This means that essentially the same algorithm can be used to produce both kinds of estimates.

Let $F(\Sigma, \mathbf{S})$ denote one of the functions $M(\Sigma, \mathbf{S})$, $G(\Sigma, \mathbf{S})$, or $L(\Sigma, \mathbf{S})$. The basic algorithm is defined by the step

$$\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} - \mathbf{H}^{-1}\mathbf{g}, \tag{3}$$

where $\mathbf{g}' = (g_1, \ldots, g_q)$ is the gradient of F, and the symmetric matrix $\mathbf{H} = \{h_{ij}, i, j = 1, \ldots, q\}$ is an approximation to the Hessian, both evaluated at $\boldsymbol{\gamma}^{(t)}$, the value of the parameter vector at iteration t. Disregarding terms of order two or higher in the Taylor series approximation to the second derivatives, several authors (e.g., Browne, 1982, Sec. 1.9; Jöreskog, 1981, sec. 3.3; Lee & Jennrich, 1979, sec. 3) have exploited the relationship

$$\frac{\partial^2 F}{\partial \gamma_i \partial \gamma_j} \approx h_{ij} = \frac{1}{2} \text{ tr } (\mathbf{A}^{-1}\dot{\Sigma}_i \mathbf{A}^{-1}\dot{\Sigma}_j), \tag{4}$$

and

$$\frac{\partial F}{\partial \gamma_i} = g_i = \frac{1}{2} \text{ tr } (\mathbf{A}^{-1}(\Sigma - \mathbf{S})\mathbf{A}^{-1}\dot{\Sigma}_i), \tag{5}$$

under the assumption of multivariate normality if (1) or (2) are chosen. Depending on the discrepancy function, define $\mathbf{A} = \Sigma(\boldsymbol{\gamma}^{(t)})$, $\mathbf{A} = \mathbf{S}$, or $\mathbf{A} = \mathbf{I}$ for $M(\Sigma, \mathbf{S})$, $G(\Sigma, \mathbf{S})$, or $L(\Sigma, \mathbf{S})$, respectively. The matrix $\dot{\Sigma}_i$ contains the partial derivatives of the covariance matrix with respect to the i-th parameter evaluated at $\boldsymbol{\gamma}^{(t)}$; that is,

$$\dot{\Sigma}_i : \left\{ \dot{\sigma}_{jk} = \frac{\partial \sigma_{jk}(\boldsymbol{\gamma}^{(t)})}{\partial \gamma_i} \right\}.$$

With (1) or (2), if the distribution assumption of multivariate normality is tenable and if the model holds, then at the minimum, the likelihood ratio test statistic, nF, for the hypothesis H_0: $\Sigma = \Sigma(\boldsymbol{\gamma})$ is distributed approximately as chi square with

$\frac{1}{2}(p^2 + p) - q$ degrees of freedom in large samples. The square root of diagonal elements of $(nH)^{-1}$ are approximations to the standard errors for $\hat{\gamma}$.

Once a structure $\Sigma(\gamma)$ has been specified, the only difficulty in implementing (3) to estimate a model is in computing the q matrices of partial derivatives $\dot{\Sigma}_i$ used in (4) and (5). Instead of setting these up for each different model, they can be approximated numerically to quite a reasonable degree by using the finite forward difference method of differentiation extended to matrix derivatives (e.g., Burden & Faires, 1989, sec. 4.1):

$$\dot{\Sigma}_i = \frac{\partial \Sigma}{\partial \gamma_i}(\gamma^{(t)}) \approx \frac{\Sigma(\gamma^{(t)} + e_i h) - \Sigma(\gamma^{(t)})}{h}, \quad i = 1, \ldots, q.$$

Here h is a small positive constant (e.g., $h = 10^{-7}$), and e_i is a vector of length q whose only nonzero entry is unity at the i-th position.

This sketch completes the essential features of the procedure; however, two common problems encountered during estimation merit further comment. First, when maximum likelihood is chosen as the discrepancy function, a poor initial estimate of γ can generate $A = \Sigma(\gamma^{(t)})$ in (4) and (5) that is numerically singular. In this situation, Browne and Du Toit (1992) have used the effective method of switching to GLS until $\Sigma(\gamma^{(t)})$ is again positive definite, at which point a switch back to maximum likelihood is again made. This practice was followed here as well. Second, if the approximate Hessian is singular, because, for example, the model is unidentified, then the update in (3) cannot be computed directly. Singularity of H may also occur because of a particular value of $\gamma^{(t)}$, or may occur because one or more parameters are superfluous, or may result from a condition of the data. One way to handle this situation corresponds roughly to a suggestion made by Jennrich and Sampson (1966) to solve the system of linear equations implied by (3), $Hd = -g$, for d by successive sweep operations, but only if the relevant pivot element of H exceeds a specified tolerance. If a particular row and column of H are not swept, then the associated element in d is set to zero, and the corresponding element in $\gamma^{(t)}$ is not updated. This modification is invaluable because it usually allows the iterative process to continue. Further, the unswept elements can serve as a guide to modifying the model appropriately.

The appendix contains a listing of this algorithm written for Proc IML (SAS/IML *User's Guide*, 1985) to estimate the well-known Stability of Alienation model (Wheaton, Muthén, Alwin, & Summers, 1977). Other high-level matrix languages such as Speakeasy, Gauss, or APL would provide an equally appropriate environment for implementing this method. IML is convenient, however, since it is widely available on mainframe computers, and allows one to take advantage of the SAS system for other operations. To adapt this code for a particular problem, one need only change the listing given here by (i) describing the structure, $\Sigma(\gamma)$; (ii) supplying the data, most often in the form of a sample covariance matrix; and (iii) providing an initial approximation to the parameter vector, $\gamma^{(0)}$.

Examples

These ideas will be illustrated with two factor analysis models, and two structures for the quasi-simplex pattern. It is possible to estimate versions of the simplex models using commercially available programs, but both require complicated reparameterizations because the structures are not special cases of these programs. While reparameterizations are effective, they obscure the sense of the model and often require secondary calculations to obtain the desired parameter estimates and their standard errors. The first two models are discussed in some detail because the data sets are realistic, and

because this approach to estimating them highlights certain features of each that have not been pointed out before.

A Power Function for True Score Correlations

The first example is based on a power function for true score correlations originally described by Anderson (1960). Morrison (1990, sec. 9.9) subsequently reviewed this structure and briefly compared its properties to related models. Jöreskog (1970) gave a comprehensive overview of simplex structures and formulated a version of this model that allows it to be estimated within the LISREL framework. However, we know of no instance where the power model in its original form has been applied directly to data.

In a simplex it is assumed that a particular pattern applies to the observed score correlation matrix. With a quasi-simplex the observed scores are assumed to be measured with error, so that the associated true score correlations follow the simplex pattern directly. Since the latter assumption is more realistic in the behavioral sciences, quasi-simplex structures are more justifiable. Let $X_j, j = 1, \ldots, p$, have $\mathscr{E}(X_j) = 0$, and let

$$X_j = \xi(t_j) + \varepsilon_j.$$

The true scores, $\xi(t_j)$, are a function of unknown scale points t_j, ordered such that $0 \le t_1 \le \cdots \le t_p$. It is assumed that the components have expectations $\mathscr{E}(\xi(t_j)) = \mathscr{E}(\varepsilon_j) = 0$, and variances $\delta_j^2 = \text{var}(\xi(t_j))$ and $\theta_j^2 = \text{var}(\varepsilon_j)$, while $\text{cov}(\xi(t_j), \varepsilon_k) = 0$ for all $j, k = 1, \ldots, p$. A quasi-simplex is said to follow a Markov process if the correlation function between $\xi(t_j)$ and $\xi(t_k)$ has the form

$$\rho_{jk} = \rho^{|t_j - t_k|}, \tag{6}$$

for a parameter $0 < \rho < 1$. The symmetric matrix $\mathbf{P}(p \times p)$, with diag $(\mathbf{P}) = \mathbf{I}$, is composed of these elements: $\mathbf{P} = \{\rho_{jk}\}$. Since $t_j \ge t_k$ for $j > k$, $\mathbf{P}$ displays the characteristic simplex pattern in which correlations decrease in magnitude as one moves away from the diagonal. Let $\mathbf{D}_\delta = \text{diag}(\delta_1, \ldots, \delta_p)$ and $\mathbf{D}_\theta = \text{diag}(\theta_1, \ldots, \theta_p)$. Then the covariance matrix for $\mathbf{X}$ is

$$\mathbf{\Sigma} = \mathbf{D}_\delta \mathbf{P} \mathbf{D}_\delta + \mathbf{D}_\theta^2. \tag{7}$$

There are several indeterminacies in this structure that must be eliminated before it can be estimated. The scale and origin of $\mathbf{t}$ are arbitrary, because in (6) only the differences $t_j - t_k$ enter into the exponent. The coefficient of the power function is also arbitrary, since for a different pair of scale values, $\ddot{t}_j \ne t_j$ and $\ddot{t}_k \ne t_k$, there exists $\ddot{\rho} \ne \rho$, such that

$$\ddot{\rho}^{\,\ddot{t}_j - \ddot{t}_k} = \rho^{\,t_j - t_k}, \tag{8a}$$

when

$$\ln \ddot{\rho} = \frac{(t_j - t_k) \ln \rho}{\ddot{t}_j - \ddot{t}_k}. \tag{8b}$$

There are several ways to identify the model. As an example, four of the scale points, such as t_1, t_2, t_{p-1}, and t_p, can be constrained in various ways. This approach allows ρ to be estimated. Alternatively, one can fix ρ plus three of the scale points. Since estimates of t_j are the most interesting components of the model, it seems preferable to take the latter approach so that an additional value of t_j can be estimated. One set of three constraints is $t_1 = 0$, $t_2 = 1$, $t_p = 1 + t_{p-1}$.

Although the value of ρ is somewhat arbitrary, care must be taken in deciding on a particular value. This is because

$$\delta_1 \rho^{t_2 - t_1} = \frac{\sigma_{21}}{\delta_2}, \text{ and } \delta_p \rho^{t_p - t_{p-1}} = \frac{\sigma_{p1}\delta_{p-1}}{\sigma_{p-1,1}}, \tag{9}$$

(Anderson, 1960, p. 212) where σ_{jk} is an element of Σ in (7). (Here we take the general case in which the differences $t_2 - t_1$ and $t_p - t_{p-1}$ are not necessarily unity.) Errors and true scores are uncorrelated, so that $\sigma_j^2 = \delta_j^2 + \theta_j^2$. This equality in conjunction with (9) imposes a lower bound on ρ that is typically greater than the nominal restriction $\rho > 0$. To ensure that both $\theta_1^2 > 0$ and $\theta_p^2 > 0$, the value specified must satisfy

$$\rho > \max (b_1, b_p), \tag{10}$$

where

$$b_1 = \left(\frac{\sigma_{21}}{\sigma_{11}\delta_2}\right)^{(t_2 - t_1)^{-1}}, \text{ and } b_p = \left(\frac{\sigma_{p1}\delta_{p-1}}{\sigma_{pp}\sigma_{p-1,1}}\right)^{(t_p - t_{p-1})^{-1}}.$$

Maximum likelihood estimates of the standard deviations are equal to their sample counterparts, so $\hat{\sigma}_{11} = s_{11}$ and $\hat{\sigma}_{pp} = s_{pp}$, but the other components of b_1 and b_p are unknown. Thus, the above relationships cannot be used to determine ρ before the model is applied to data. In the problems tried to date we have arbitrarily taken $\rho = 0.95$, which has worked satisfactorily. The true score correlations $\rho_{j1}, j \geq 2$, and ρ_{kp}, $k \leq p - 1$, vary as a function of ρ, but all other correlations, $\rho_{jk}, p - 1 \geq j > k \geq 2$, are invariant with respect to ρ. This is because the scale values, $t_3, \ldots, t_{p-1}$, are free to vary according to (8) as a function of ρ. With ρ fixed, there are a total of $3p - 3$ parameters in the model, namely $(\delta_1, \ldots, \delta_p, \theta_1, \ldots, \theta_p, t_3, \ldots, t_{p-1})$.

In some cases, it may be reasonable to impose other restrictions on D_δ or D_θ. Depending on how this is done, additional values of t can often be estimated. For example, if the residual variances are homogeneous, then $D_\theta = I\theta$ for the single parameter $\theta \geq 0$. Only one element of t must be fixed in this case, most reasonably $t_1 = 0$. The number of parameters would be only $2p$, $(\delta_1, \ldots, \delta_p, \theta, t_2, \ldots, t_p)$.

It is sometimes useful to write the model in the form

$$\Sigma = D_\delta(P + D_\theta^2)D_\delta, \tag{11}$$

so that the residual variances are proportional to true score standard deviations. If both D_δ and D_θ are unconstrained, then (11) is equivalent to (7) and D_θ^2 in the former is simply a rescaling of D_θ^2 in the latter. If either D_δ or D_θ are restricted in some manner, however, then the two structures are not in general equivalent, and one form may be preferable because the fit is better, even though the degrees of freedom are the same.

Negative residual variance estimates can sometimes occur, unless these estimates are constrained to be nonnegative. Consequently, it is prudent to estimate the standard deviations of the residuals, $(\theta_1, \ldots, \theta_p)$, and then add their squares to the diagonal elements of Σ as in (7) and (11). This well-known parameterization restricts the variance estimates to the admissible region $\theta_j^2 \geq 0$.

Both forms of the model are scale invariant as long as D_δ is unconstrained. If the covariance matrix is rescaled by an arbitrary diagonal matrix D_α in $D_\alpha\Sigma D_\alpha$, then t and the correlation matrix P will be unchanged, while the rescaled version of D_δ and D_θ will be $D_\delta D_\alpha$ and $D_\theta D_\alpha^2$, respectively. The test statistic is the same in either rescaling. Finally, it should be noted that because of the power function (6), this model is appropriate only for sample data in which the covariances among variables are nonnegative.

TABLE 1

Summary Data for Eleven Trial Blocks of Reaction Time.

Trial	1	2	3	4	5	6	7	8	9	10	11
1	—										
2	.696	—									
3	.551	.808	—								
4	.483	.710	.893	—							
5	.372	.606	.784	.856	—						
6	.373	.590	.750	.811	.899	—					
7	.330	.541	.712	.762	.844	.887	—				
8	.297	.503	.663	.739	.820	.881	.903	—			
9	.288	.470	.625	.699	.768	.827	.849	.905	—		
10	.246	.420	.559	.643	.724	.759	.801	.836	.891	—	
11	.250	.399	.547	.627	.688	.730	.758	.817	.867	.897	—
SD	92.6	63.1	50.3	40.9	37.9	33.2	30.5	28.8	27.6	27.6	25.8
Mean	242	193	169	149	142	133	127	123	119	116	112

Note: Data are described in Woltz (1988). $N = 401$.

To illustrate this model, some data originally gathered by Woltz (1988) were analyzed. In this study, Air Force recruits were trained in a complex procedural learning task during a single experimental session. Both response errors and latencies were recorded, but only the reaction time measures were analyzed here. The data were summarized into eleven blocks of sixty-four trials each. Four hundred twelve subjects had complete data at the end of the experiment. Of this number, eleven cases were set aside due to unusually slow or fast response times. The resulting summary data for $N = 401$ subjects are shown in Table 1.

The model in (11) was fit by maximum likelihood. The value of the likelihood ratio test statistic for the null hypothesis that the model holds compared to the unstructured alternative hypothesis was $\chi^2 = 45.8$, on df = 36, with an upper tail probability of 0.13. Parameter estimates are in Table 2. Since $b_1 = 0.70$ and $b_p = 0.94$, the fixed coefficient of the power function $\rho = 0.95$ satisfies (10). Examining $\hat{t}_j$ showed that progressing from the first to the fifth trial blocks corresponded to the greatest improvements in true reaction time. The improvement between later blocks was steady but less dramatic. After trial five, improvement between trial blocks corresponded to an increment of about one unit in the t-scale for the remainder of the experiment. Two estimated residual variances were at their boundary. It is interesting that the estimate $\hat{\theta}_1^2 = 0.863$ is roughly ten times greater than the residual for the variable with next largest value. Thus there is appreciable "noise" in the reaction time measurements for trial block one. The effect of this error is best observed by comparing the estimated true score correlations between Trial 1 and Trials 2–11, $\hat{\rho}_{j1}$, $j \geq 2$, in the lower section of

TABLE 2

Maximum Likelihood Estimates of Quasi–simplex Model with Power Function.

	1	2	3	4	5	6	7	8	9	10	11
$\hat{t}_j$:	0.00*	1.00*	5.16	7.10	9.35	10.4	11.4	12.2	13.4	14.4	15.4*
	—	—	.419	.541	.704	.775	.830	.868	.931	.989	—
$\hat{\delta}_j$:	67.8	63.1	50.3	40.3	36.9	32.4	29.2	28.3	27.1	26.5	25.4
	4.24	2.23	1.78	1.49	1.40	1.21	1.13	1.05	1.00	1.02	1.03
$\hat{\theta}_j^2$:	.863	0.00‡	0.00‡	.028	.055	.048	.093	.039	.039	.085	.034
	.138	—	—	.015	.015	.012	.015	.011	.011	.015	.018

$\hat{\rho}_{jk} = 0.950^{|\hat{t}_j - \hat{t}_k|}$.   (Estimated True Score Correlations)

```
1.0
 .95  1.0
 .77   .81  1.0
 .70   .73   .91  1.0
 .62   .65   .81   .89  1.0
 .59   .62   .76   .84   .95  1.0
 .56   .59   .73   .80   .90   .95  1.0
 .54   .56   .70   .77   .87   .92   .96  1.0
 .50   .53   .66   .73   .81   .86   .91   .94  1.0
 .48   .50   .62   .69   .77   .82   .86   .89   .95  1.0
 .45   .48   .59   .65   .73   .77   .81   .85   .90   .95  1.0
```

Note: Standard errors are below estimates.   *Fixed parameter.   ‡Parameter on boundary.

Table 2, with the corresponding observed score correlations shown in Table 1. The differences $(\hat{\rho}_{j1} - r_{j1})$ are rather large, generally on the order of 0.20. On the other hand, the other estimated true score correlations seldom differ by more than 0.05 from their observed score counterparts. Overall, the quantitative summary seems quite reasonable for an experiment of this kind. This completes the first illustration of the method.

Other structures for the quasi-simplex can also be fit directly by this method. For example, a quasi-Wiener simplex (e.g., Jöreskog, 1970, sec. 3) is generated by the correlation function

$$\rho_{kj} = \left(\frac{w_j}{w_k}\right)^{1/2},$$

for the scale points $w_k > w_j$. Taking $\mathbf{D}_w = \text{diag}(\sqrt{w_1}, \ldots, \sqrt{w_p})$ and $\mathbf{D}_\theta = \text{diag}(\theta_1^2, \ldots, \theta_p^2)$, the covariance structure is

$$\mathbf{\Sigma} = \mathbf{D}_w \mathbf{P} \mathbf{D}_w + \mathbf{D}_\theta,$$
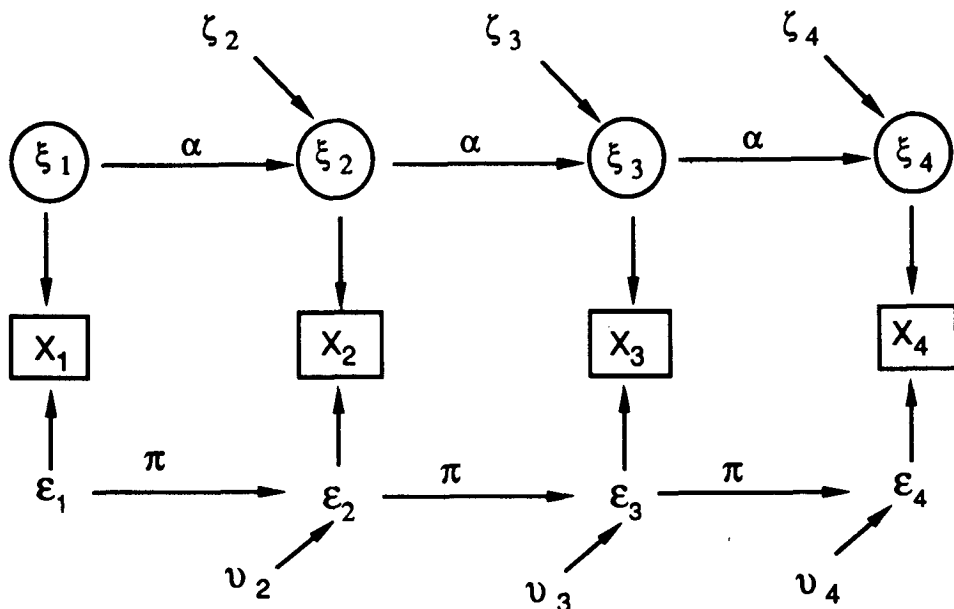
FIGURE 1.
Model for quasi-simplex with autoregressive structure between true scores and between error scores.

or more simply,

$$\sigma_{jk} = \begin{cases} w_k + \theta_k^2, & k = j \\ w_k, & k < j \end{cases}.$$

This structure is identified if one restriction, for example $\theta_p^2 = 0$, is imposed, so a total of $2p - 1$ parameters are required.

A Double Autoregressive Model

Wiley and Wiley (1974) presented a rather different idea for a quasi-simplex in a longitudinal design. In this model, there is one autoregressive structure between the true scores and a second autoregressive relationship between the error scores. A representation of the model is in Figure 1. Analogous to the development above, consider the decomposition $X_j = \xi_j + \varepsilon_j$, for $j = 1, \ldots, p$. The true scores are not independent but rather are related by the linear function

$$\xi_j = \alpha \xi_{j-1} + \zeta_j,$$

for $j \geq 2$. Similarly, the error scores have the relationship,

$$\varepsilon_j = \pi \varepsilon_{j-1} + \nu_j,$$

for $j \geq 2$. The common coefficient of each true score regression is α, while π is the coefficient common to each regression between error scores at adjacent time points. The variables ζ_j and ν_j are the residuals of the true score and error score regressions, respectively. It is assumed that all other pairs of variables are uncorrelated; thus, cov $(\xi_j, \varepsilon_k) = $ cov $(\xi_j, \nu_k) = $ cov $(\zeta_j, \varepsilon_k) = $ cov $(\zeta_j, \nu_k) = 0$.

A distinguishing feature of this model is that the error variance is taken as constant for all variables:

$$\text{var} (\varepsilon_1) = \ldots = \text{var} (\varepsilon_p) = \gamma^2. \tag{12}$$

This implies that the variance of the regression residuals for error scores is also constant,

$$\text{var} (\nu_j) = (1 - \pi^2) \text{ var} (\varepsilon_1), \qquad j \geq 2.$$

As a consequence of (12) and the autoregressive structure among residual scores, the covariance matrix among ε_j is a power function of only two terms

$$\varphi_{jk} = \text{cov} (\varepsilon_j, \varepsilon_k) = \gamma^2 \pi^{|j - k|}.$$

For example, when $p = 4$,

$$\Phi = \{\varphi_{jk}\} = \gamma^2 \begin{bmatrix} 1 & & & \\ \pi & 1 & & \\ \pi^2 & \pi & 1 & \\ \pi^3 & \pi^2 & \pi & 1 \end{bmatrix}.$$

Define the variance of the true scores for the first variable as $\psi_1 = \text{var} (\xi_1)$, and let the variances of the regression residuals be $\psi_j = \text{var} (\zeta_j)$, $j \geq 2$, with $D_\psi = \text{diag} (\psi_1, \ldots, \psi_p)$. Because true scores and error scores are uncorrelated, the structure is

$$\Sigma = \Omega + \Phi, \tag{13}$$

where

$$\Omega = (I - A)^{-1} D_\psi (I - A')^{-1}$$

$$= \begin{bmatrix} \psi_1 & & & \\ \alpha\psi_1 & \alpha^2\psi_1 + \psi_2 & & \\ \alpha^2\psi_1 & \alpha^3\psi_1 + \alpha\psi_2 & \alpha^4\psi_1 + \alpha^2\psi_2 + \psi_3 & \\ \alpha^3\psi_1 & \alpha^4\psi_1 + \alpha^2\psi_2 & \alpha^5\psi_1 + \alpha^3\psi_2 + \alpha\psi_3 & \alpha^6\psi_1 + \alpha^4\psi_2 + \alpha^2\psi_3 + \psi_4 \end{bmatrix}.$$

Here $A(p \times p)$ is a matrix with all elements on the first subdiagonal equal to α:

$$A = \begin{bmatrix} 0 & & & & & \\ \alpha & 0 & & & & \\ 0 & \alpha & 0 & & & \\ \cdot & & & \ddots & & \\ \cdot & & & & 0 & \\ \cdot & & & & & \\ 0 & 0 & 0 & \cdots & \alpha & 0 \end{bmatrix}$$

The model has an attractive simplicity. There are only $p + 3$ parameters, namely (α, π, γ^2, $\psi_1, \ldots, \psi_p$). Obviously the model is not invariant with respect to arbitrary rescalings of the variables because restrictions are imposed directly on the residual variances according to (12). Consequently, the sample covariance matrix must be analyzed.

Although this model was proposed some time ago, it has seldom been applied, perhaps because the structure cannot be easily specified in most computer programs.

TABLE 3

Correlations and Standard Deviations for Eight Semesters of GPA.

Semester	1	2	3	4	5	6	7	8
1	—							
2	.556	—						
3	.456	.490	—					
4	.439	.445	.562	—				
5	.399	.418	.496	.512	—			
6	.415	.383	.456	.469	.551	—		
7	.387	.364	.445	.442	.500	.544	—	
8	.342	.339	.354	.416	.453	.482	.541	—
SD	.57	.57	.57	.54	.59	.58	.59	.59

Note: From Humphreys (1968, Table 2); N is approximately 1,600.

Recently Green and Palmquist (1991) presented a clever, but complicated, parameterization of this model that can be implemented with available software. However, if these programs are used to fit the model, the number of latent variables required for the nonlinear constraints in (12) is very large: $10p - 4$. So while estimation of the model with standard programs is possible, it may be more convenient and practical to use a direct method instead.

To illustrate how this model can be estimated with the current procedure, some data originally reported by Humphreys (1968) were analyzed. They are reproduced in Table 3. The variables are $p = 8$ semesters of GPA for a sample of approximately $N = 1,600$ students. These data were previously analyzed by Werts, Linn, and Jöreskog (1978) using an equivalent form of the power function model in (7) with (6) above.

The value of the maximum likelihood test statistic for the hypothesis that the model in (13) holds against the general alternative was $\chi^2 = 38.4$ on df = 25, with an upper tail probability of 0.04. Parameter estimates and associated standard errors were

$$\hat{\alpha} = .940 \ (.010), \quad \hat{\pi} = .056 \ (.021), \quad \hat{\gamma}^2 = .149 \ (.005),$$

$$\hat{\psi}_j \quad \begin{matrix} .176 & .006 & .044 & .003 & .041 & .019 & .027 & .035 \\ (.010) & (.006) & (.007) & (.006) & (.006) & (.006) & (.007) & (.009) \end{matrix}$$

In interpreting these estimates, it should be recalled that $\psi_1 = \text{var} (\xi_1)$, while $\psi_j = \text{var} (\zeta_j)$, for $j \geq 2$. The autoregression effect among true scores was therefore relatively strong. The largest of the true score regression residual variances was only .044. Comparing values of $\hat{\psi}_j$ to $\text{se}(\hat{\psi}_j)$ suggested that ψ_2 and ψ_4 may be zero. The modest value of $\hat{\pi} = .056$ and the relatively large value of $\hat{\gamma}^2 = .149$ means that the corresponding

autoregression effect among the error scores was quite small. The largest of the estimated error score correlations were those between adjacent variables, for which the correlation between ε_j and ε_{j-1} is $\hat{\pi} = 0.056$. This suggests that a modified model without the auto-regressive structure for error scores might be tried, although we did not pursue the possibility here.

Standardized Second-Order Factors

The next example illustrates that a seemingly minor change to a model that is otherwise easily estimated by all programs may make the modified model impossible to handle without a more general approach. Using a set of twelve ability variables, Rindskopf and Rose (1988) studied several second-order factor analysis models,

$$\Sigma = \Lambda(\mathbf{B}\Phi\mathbf{B}' + \mathbf{D}_\nu)\Lambda' + \mathbf{D}_\psi. \tag{14}$$

In one instance (Table 4, p. 63), they specified that $\mathbf{D}_\psi = \{\psi_j\}$ and $\mathbf{D}_\nu = \{\nu_k\}$ be diagonal matrices of order twelve and five, respectively, Φ was a 2×2 correlation matrix with diag (Φ) = $\mathbf{I}$, and the first-order and second-order factor patterns had the structure

$$\Lambda' = \begin{bmatrix} \lambda_1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 & \lambda_6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_7 & \lambda_8 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & \lambda_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_9 & \lambda_{10} & \lambda_{11} & \lambda_{12} \end{bmatrix},$$

and

$$\mathbf{B}' = \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} & 0 & 0 \\ 0 & 0 & 0 & \beta_{42} & \beta_{52} \end{bmatrix},$$

respectively. To identify the variances of the first-order factors, they used the restrictions $\lambda_1 = \lambda_3 = \lambda_5 = \lambda_7 = \lambda_9 = 1$.

Alternatively, one could estimate all twelve coefficients in Λ, and impose the restriction

$$\text{diag } (\mathbf{B}\Phi\mathbf{B}' + \mathbf{D}_\nu) = \mathbf{I}. \tag{15}$$

Then the second-order factors are standardized with unit variance. Elements of $\mathbf{D}_\nu$ are not free parameters but instead satisfy the identity

$$\mathbf{D}_\nu = \mathbf{I} - \text{diag } (\mathbf{B}\Phi\mathbf{B}').$$

This method of identification is particularly convenient if the subsets of variables associated with each first-order factor are parallel or tau-equivalent (see Jöreskog, 1981, sec. 2.2). For if the variances of the second-order factors are constrained as in (15), one can avoid the arbitrary step of fixing one of the factor loadings to unity to identify the factor variances.

Rinsdkopf (1984) described a way to impose the needed restriction with current programs, but the parameterization works only for models in which each first-order factor loads on one second-order factor. If the first-order factors are complex, the structure cannot be estimated using this method. For example, one might wish to estimate β_{12} under model (14) maintaining restriction (15). While this appears not to be a major change to the structure, the addition of β_{12} takes the resulting model out of the

TABLE 4

Maximum Likelihood Estimates of a Second–Order Factor Analysis Model.

λ_1	.922	ψ_1	.150	β_{11}	.459	ν_1	.560
λ_2	.933	ψ_2	.130	β_{21}	.806	ν_2	.351
λ_3	.914	ψ_3	.165	β_{31}	.633	ν_3	.599
λ_4	.864	ψ_4	.253	β_{42}	.564	ν_4	.682
λ_5	.891	ψ_5	.207	β_{52}	.870	ν_5	.244
λ_6	.865	ψ_6	.253	β_{12}	.252		
λ_7	.901	ψ_7	.189				
λ_8	.866	ψ_8	.250	ϕ_{21}	.715		
λ_9	.605	ψ_9	.634				
λ_{10}	.580	ψ_{10}	.664				
λ_{11}	.725	ψ_{11}	.475				
λ_{12}	.487	ψ_{12}	.762				

$$\hat{\mathbf{B}}\hat{\boldsymbol{\Phi}}\hat{\mathbf{B}}' + \hat{\mathbf{D}}_\nu = \begin{bmatrix} 1.0 & & & & \\ .53 & 1.0 & & & \\ .43 & .47 & 1.0 & & \\ .30 & .33 & .27 & 1.0 & \\ .47 & .52 & .42 & .49 & 1.0 \end{bmatrix}$$

Note: The unique variances of first–order factors, ν_k, are not free parameters, but rather are constrained so that $\beta_{k1}^2 + \beta_{k2}^2 + 2\beta_{k1}\beta_{k2}\phi_{21} + \nu_k = 1$

class of structures that can be handled by standard programs. On the other hand, the model is simple to estimate with a more general procedure.

The estimates in Table 4 were obtained by fitting this structure to the data presented by Rindskopf and Rose (1988, Table 1). Because elements of $\mathbf{D}_\nu$ are not free parameters, there are a total of $q = 31$ parameters, with df = 47. The value of the test statistic is $\chi^2 = 50.49$, with an upper tail probability of 0.34. The correlation matrix among first-order factors is in the lower section of the table, where as required, diag $(\hat{\mathbf{B}}\hat{\boldsymbol{\Phi}}\hat{\mathbf{B}}' + \hat{\mathbf{D}}_\nu) = \mathbf{I}$.

It is likely that the constraints needed to estimate this particular model will be incorporated into commercial programs before long. That present versions of these

programs cannot estimate this structure is a secondary point. The more important issue is that general programs frequently are required to estimate models that are outside the domain of standard programs. This can be important even when seemingly minor changes are needed.

A Model for Dichotomized Variables

Christoffersson (1975) considered a factor analysis model for dichotomized variables that utilizes the marginal distribution of single items and the joint distributions of item pairs. Since the model was developed for binary variables, and includes transformations that relate dichotomized observed variables to continuous latent variables, it is completely outside the domain of popular programs for structural equation models. Nonetheless, it is a simple matter to employ the general method used here to produce least squares estimates of this model.

Let the correlation matrix among continuous, unobserved variables, $Y_1, \ldots, Y_p$, be

$$\mathbf{P} = \lambda\lambda' + \mathbf{D}_\delta, \tag{16}$$

where diag $(\mathbf{P}) = \mathbf{I}$, and where it is assumed for simplicity that a single factor holds. As in (15), elements of $\mathbf{D}_\delta$ are not free parameters, but rather $\mathbf{D}_\delta = \mathbf{I} - \text{diag} (\lambda\lambda')$. Associated with each variate is a coefficient, h_j, the threshold for the observed binary variable, Y_j^*, with scores

$$Y_j^* = \begin{cases} 1 & \text{if } Y_j \geq h_j, \\ 0 & \text{otherwise.} \end{cases}$$

In this context, the structure $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ does not represent the covariance matrix among observed dichotomous scores. Instead, it is the matrix of proportions of endorsed responses

$$\boldsymbol{\Sigma} = \{\sigma_{jk}\} = \begin{cases} \text{Prob } (Y_j^* = 1) & j = k \\ \text{Prob } (Y_j^* = 1, \ Y_k^* = 1) & j \neq k \end{cases}. \tag{17}$$

Denote the density of the standard normal distribution by $\varphi_1(z)$, and of the standard bivariate normal density, with coefficients $\mathbf{z} = (z_1, z_2)'$, by

$$\varphi_2(\mathbf{z}) = \frac{1}{2\pi(1 - \rho_{jk}^2)^{1/2}} \exp\left(-\frac{1}{2}\mathbf{z}'[\mathbf{P}]_{jk}^{-1}\mathbf{z}\right),$$

where $[\mathbf{P}]_{jk} = \begin{pmatrix} 1 & \rho^{jk} \\ \rho_{jk} & 1 \end{pmatrix}$ is a 2×2 submatrix for items j and k from $\mathbf{P}$ in (16). The distribution functions are, respectively,

$$\Phi_1(h) = \int_{-\infty}^{h} \varphi_1(z) \, dz,$$

and

$$\Phi_2(h_j, h_k; [\mathbf{P}]_{jk}) = \int_{-\infty}^{h_j} \int_{-\infty}^{h_k} \varphi_2(\mathbf{z}) \, dz_k \, dz_j.$$

Rewriting (17), the components of $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ are

$$\sigma_{jk}(\gamma) = \begin{cases} 1 - \Phi_1(h_j) & j = k, \qquad \text{(18a)} \\ 1 - \Phi_1(h_j) - \Phi_1(h_k) + \Phi_2(h_j, h_k; [\mathbf{P}]_{jk}) & j \neq k, \qquad \text{(18b)} \end{cases}$$

where $\gamma = (h_1, \ldots, h_p, \lambda_1, \ldots, \lambda_p)'$.

Although the model is not a covariance structure for continuous variables, the marginal and joint proportions of the dichotomous items can be computed as the elements of a symmetric, matrix-valued function of γ. To compute GLS estimates or maximum likelihood estimates would entail a modification of this procedure because these estimators utilize additional information on the distribution of Y^*. Ordinary least squares, however, can be applied directly. Compared to the more routine applications of the method, all that is needed are approximations to the distribution functions $\Phi_1(h)$ and $\Phi_2(h_j, h_k; [\mathbf{P}]_{jk})$. The joint probability (18b) can be efficiently written as a single integral (Froemel, 1971), so that Gaussian quadrature, which is straightforward to implement in high-level computer languages, works well here.

Christofferson presented GLS estimates of this model for two sets of five items originally studied by Bock and Lieberman (1970), and that have been subsequently analyzed by many writers. The joint and marginal frequencies for the $p = 5$ items of the LSAT Section 6 based on $N = 1,000$ are shown in the upper section of Table 5. Least squares estimates of h_j, λ_j and $\delta_j = 1 - \lambda_j^2$ are shown in the lower section of the table. The estimated thresholds are similar to the GLS estimates reported by Christofferson, but there is greater discrepancy for some $\hat{\lambda}_j$ between the OLS and GLS solutions. The loadings in Table 5 also differ somewhat from other sets of least squares estimates reported by Mooijaart (1983) and Muthén (1978) for these variables. The differences are due to their decision to fix h_j at the sample values $\hat{h}_j = \Phi_1^{-1}(s_{jj})$, and solve only for λ_j, while we have obtained all eight estimates simultaneously.

This structure is equivalent to the two parameter normal ogive item response theory model (Takane & de Leeuw, 1987). The correspondence between the two is described especially clearly in McDonald (1985). He also related both the factor analysis and IRT models to the class of nonlinear latent variable structures developed by McDonald (1967). (We are grateful to an associate editor for bringing this matter to our attention.) The equivalence of these three models is quite pleasing, as is the fact that several specializations of them can be estimated with essentially the same computing scheme. There are practical benefits as well. For example, with this data set there is a six-fold improvement in computing efficiency for the factor analysis model estimated as a nonlinear latent variable structure than when estimated directly.

Discussion

For many problems it is obviously more efficient to estimate a model with available computer programs than to bother with a special algorithm. This is especially true for popular models such as restricted factor analysis or latent variable regressions. On the other hand, many interesting structures exist that cannot be estimated with existing software, or that require a complicated parameterization to set up. A more general method is preferable for them. This method may also be valuable when investigating other issues related to structural equation modeling such as fit indices that require elements of the Hessian matrix or other derivatives (e.g., Sörbom, 1975).

Examples have been presented in some detail to illustrate that the procedure is appropriate for realistic problems. Indirect versions of the first two models can be estimated with popular computer programs, but it is often preferable to fit the models directly, for then the actual model parameters, not functions of them, can be estimated, and the appropriate standard errors are computed.

TABLE 5

Data and Least Squares Estimates of Model (18).

Joint and Marginal Frequencies

1.	924				
2.	664	709			
3.	524	418	553		
4.	710	553	445	763	
5.	806	630	490	678	870

Least Squares Estimates

$\hat{h}_j$	−1.408	−0.560	−0.128	−0.724	−1.147
$\hat{\lambda}_j$	0.455	0.310	0.605	0.304	0.213
$\hat{\delta}_j$	0.793	0.904	0.634	0.908	0.955

Note: Data from Bock and Lieberman (1970), LSAT Section 6. $N = 1,000$

This procedure represents a basic framework for structural equation models. Its simplicity is an asset. It can be generalized in several ways depending on requirements. Including a facility for multiple populations is one obvious extension. Models with a mean structure can be accommodated by augmenting the covariance matrix, as discussed, for example, by Jöreskog (1981, sec. 5.1), or via a more explicit representation. A larger class of distributions can be considered (Browne, 1984) by modifying the discrepancy function that is at the heart of the procedure. A valuable enhancement is to replace the basic step (3) by a complete implementation of Jennrich and Sampson's (1966) stepwise least squares solution for the Gauss-Newton update vector. None of these generalizations is difficult to incorporate into the basic scheme.

A referee requested some comparative information on the relative speed of this procedure versus LISREL 7. Processing times for a few models are shown in Table 6. Generally, LISREL is two to seven times faster than this procedure for models that LISREL can estimate directly. For very large problems, such as complicated path models or restricted factor analysis models that involve many variables, the performance difference would be even greater. Considering the many features implemented in LISREL, the relative advantage in computing time is all the more impressive. If speed is the major consideration, one would certainly not use a more general method to estimate one of LISREL's sub-models. On the other hand, LISREL is about ten times slower for the double autoregressive model (13).

More to the point, LISREL 7 cannot estimate model (7) except in an indirect way that one may not be interested in obtaining, and it cannot handle constraints such as

TABLE 6
Computing Time in Seconds for Selected Models

Model	p	q	LISREL 7	GN	Model
1	8	11	63	6	Double Autoregressive Model
2	6	16	1	2	Stability of Alienation
3	9	22	2	12	Restricted Factor Analysis
4	12	30	7	37	2nd Order Factor Analysis

Note: GN denotes the general Gauss–Newton procedure. Computing time is based on a 33MHz microcomputer. Model 1 corresponds to structure (13);

Models 2, 3, and 4 are described in Jöreskog and Sörbom (1989, p. 204),

Jöreskog and Sörbom (1989, p. 136), and Rindskopf and Rose (1988,

p. 63), respectively.

(15) in factor analysis models, or any version of model (18). Browne and Du Toit (1992) discuss others that are outside the LISREL framework. The intention has not been to minimize the value of popular programs, but to emphasize the advantages of more general procedures for problems that cannot be estimated with standard software at all.

Appendix

IML Implementation of Stability of Alienation Model

```
TITLE1 'GAUSS-NEWTON METHOD FOR COVARIANCE STRUCTURE ANALYSIS';
TITLE2 'STABILITY OF ALIENATION MODEL';

PROC IML;

START GN;

*---- INITIALIZATION -------------------------------------------------*;
   NPAR  =      16;                    * NUMBER OF PARAMETERS;
   N     =     932;                    * SAMPLE SIZE;
   MAXIT =      50;                    * MAXIMUM ITERATIONS;
   TOL   = 0.00001;                    * CONVERGENCE CRITERION;
                                       * STARTING VALUES;
   GA1   = (1, 1, 5, -.6, -.6, -.2, 7, 4, 4, 3, 267, 5, 5, 5, 5, 2);

*--- LOWER TRIANGULAR PART OF SAMPLE CORRELATION MATRIX --------------*;
   S = ( 11.834     0       0       0       0       0,
          6.947   9.364     0       0       0       0,
          6.819   5.091  12.532     0       0       0,
          4.783   5.028   7.495   9.986     0       0,
         -3.839  -3.889  -3.841  -3.625   9.610     0,
        -21.899 -18.831 -21.748 -18.775  35.522 450.288 ) ;
   S = S + S' - DIAG(S);               * FILL IN UPPER SECTION OF MATRIX;

*------- LABLES FOR VARIABLES AND PARAMETERS -------------------------*;
   LABELS = ( " AN67"  "POW67"  " AN71"  "POW67"  "  ED"  " SEI");
   PARLAB = ("LY 2,1" "LY 4,2" "LX 2,1" " BETA" " GAM 1" " GAM 2"
             "  PHI" " PS1" " PS2" " TD 1" " TD 2" "TE 1,1"
             "TE 2,2" "TE 3,3" "TE 4,4" "TE 3,1");

*--- ESTIMATION METHOD:   0 FOR ML,  1 FOR GLS ----------------------*;
   ESTMET = 0;

*-------------------------------------------------------------------*
*   MODEL: COMPUTE COVARIANCE MATRIX "SIG" AS A FUNC OF "PAR"       *
*   STABILITY OF ALIENATION MODEL                                   *
*-------------------------------------------------------------------*;
   START MODEL (SIG,PAR);
      LY  = (1 // PAR(|1|) // 0 // 0) || (0 // 0 // 1 // PAR(|2|));
      LX  = 1 // PAR(|3|);
      BI  = INV ( (1 || 0) // (PAR(|4|) || 1) );
      GA  = PAR(|5:6|);
      PH  = PAR(|7|);
      PS  = I(2) # PAR(|8:9|);
      TD  = I(2) # PAR(|10:11|);
      TE  = I(4) # PAR(|12:15|);
            TE(|3,1|) = PAR(|16|); TE(|1,3|) = PAR(|16|);

      Q1  = LY * BI;
      SYY = Q1 * (GA * PH * GA' + PS) * Q1' + TE;
      SYX = Q1 * GA * PH * LX';
      SXX = LX * PH * LX' + TD;

      SIG = (SYY || SYX) // (SYX' || SXX);        * SIGMA;
   FINISH;       * END OF MODULE "MODEL";
```

```
    RESET NONAME;
    MNAME= ('   ML' ' GLS'); NVAR = NROW(S); UV = J(NVAR,1,1);

*--- SAMPLE COVARIANCE MATRIX MUST BE POSITIVE DEFINITE --------------*;
    RUN INVCHK (SING,B,S,UV,ST);
    IF SING = 0 THEN
          INVS = B(|1:NVAR,1:NVAR|);
    ELSE DO;
          PRINT "SAMPLE COVARIANCE MATRIX IS NOT POSITIVE DEFINITE";
          GOTO STOP;
        END;
    TESTMET = ESTMET;

*--- TEST WHETHER ESTIMATED COVARIANCE MATRIX IS SINGULAR ------------*;
    RUN MODEL (SIGMA,GA1);
       IF(TESTMET = 0) THEN RUN INVCHK (SING,B,SIGMA,UV,ST);
       IF SING = 0 THEN INVSIG = B(|1:NVAR,1:NVAR|);
       IF (ESTMET|SING) = 0 THEN TESTMET = 0;
          ELSE TESTMET = 1;
       IF TESTMET = 1 THEN
           F1 = 0.5 * TRACE((I(NVAR) - INVS * SIGMA)**2);
       ELSE
           F1 = LOG(DET(SIGMA)) + TRACE(S * INVSIG) - LOG(DET(S)) - NVAR;

*--- PRINT SAMPLE COVARIANCE MATRIX AND INITIAL PARAMETER VALUES ------*;
    PRINT "SAMPLE COVARIANCE MATRIX",,
            S(|ROWNAME=LABELS COLNAME=LABELS FORMAT=8.3|);
    DO J1 = 1 TO NPAR BY 10;
       J2 = J1 + 9; IF J2 > NPAR THEN J2 = NPAR;
       A   = GA1(|J1:J2|)';
       PN = PARLAB(|J1:J2|);
       IF J1 = 1 THEN PRINT "INITIAL PARAMETER ESTIMATES",,
                            A(|COLNAME=PN FORMAT=8.3|);
                ELSE PRINT A(|COLNAME=PN FORMAT=8.3|);
    END;


*=== MAJOR ITERATION LOOP BEGINS HERE ================================*;
    FUNVALS = F1;  MAXGR = 0;  FUNDIF = 0;  ITS = 0; NSNG = 0;
    ESTVEC  = MNAME(|1,TESTMET+1|);
    DO K = 1 TO MAXIT;
       IF TESTMET = 1 THEN
          RUN APDER(GRAD,HESS,S,SIGMA,INVS,GA1,NPAR);
       ELSE
          RUN APDER(GRAD,HESS,S,SIGMA,INVSIG,GA1,NPAR);

*--- UPDATE ESTIMATES ------------------------------------------------*;
       RUN INVCHK (SING0,B,HESS,GRAD,STAT);
       INVHESS = B(|1:NPAR,1:NPAR|);
       GA2 = GA1 - B(|1:NPAR,NPAR+1|);
       RUN MODEL (SIGMA,GA2);

*--- CHECK THAT SIGMA IS POSITIVE DEFINITE ---------------------------*;
       IF(TESTMET = 0) THEN RUN INVCHK (SING,B,SIGMA,UV,ST);
       IF SING = 0 THEN INVSIG = B(|1:NVAR,1:NVAR|);
       IF (ESTMET|SING) = 0 THEN TESTMET = 0;
          ELSE TESTMET = 1;
       IF TESTMET = 1 THEN
           F2 = 0.5 * TRACE((I(NVAR) - INVS * SIGMA)**2);
       ELSE
           F2 = LOG(DET(SIGMA)) + TRACE(S * INVSIG) - LOG(DET(S)) - NVAR;
```

```
*--- RECORD ITERATION HISTORY -----------------------------------------*;
    ITS      = ITS     // K;                     * ITERATION NUMBERS;
    ESTVEC  = ESTVEC // MNAME(|1,TESTMET+1|);    * ESTIMATION METHOD;
    FUNVALS = FUNVALS // F2;                      * FUNCTION VALUE;
    FUNDIF  = FUNDIF  // (F1 - F2);              * FUNCTION DIFFERENCES;
    GMAX = MAX(ABS(GRAD));                        * MAX GRADIENT;
    MAXGR    = MAXGR   // GMAX;
    NSNG     = NSNG    // SINGO;                  * SING. PARAMETERS;


*--- CHECK FOR CONVERGENCE, UPDATE PARAMETERS AND GO AGAIN ------------*;
    IF(GMAX < TOL) THEN GOTO RESULTS;
        ELSE DO;
            F1  = F2;
            GA1 = GA2;
            END;
    END;
*=== END OF ITERATION LOOP ===========================================*;


*--- ALL DONE.  PRINT RESULTS -----------------------------------------*;
    IF K > MAXIT THEN PRINT "MAXIMUM NUMBER OF ITERATIONS WAS REACHED";

    RESULTS:
        RESID   = SIGMA - S;                      * RESIDUAL MATRIX;
        CHISQ   = (N - 1) * F2;                   * TEST STATISTIC;
        DF      = NVAR * (NVAR + 1) * 0.5 - NPAR; * DEGREES OF FREEDOM;
        P       = 1 - PROBCHI(CHISQ,DF);          * PROBABILITY;
        INFOMAT = (2 / (N - 1)) * INVHESS;        * INFORMATION MATRIX;
        SE      = SQRT (VECDIAG (INFOMAT));       * STANDARD ERRORS;

        PRINT / "ITERATION HISTORY",,
            ITS     (|COLNAME='ITER'       FORMAT=4.0|)
            ESTVEC  (|COLNAME='METHOD'     FORMAT=$CHAR6.|)
            FUNVALS(|COLNAME='FUNCTION'    FORMAT=10.6|)
            FUNDIF  (|COLNAME='DELTA(F)'   FORMAT=12.8|)
            MAXGR   (|COLNAME='MAX(GRDNT)' FORMAT=12.7|)
            NSNG    (|COLNAME='N SING PAR' FORMAT=12.0|);

        S1 = {"        MODEL MAY NO BE IDENTIFIED",
              "     (STAT FLAG = 0 INDICATES PARAMETER IS SINGULAR)"};
        G = (GA2 || SE || (STAT(|1:NPAR|)=0))';
        CN = {"ESTIMATES" "STD ERR'S" "STATUS"};

        DO J1 = 1 TO NPAR BY 10;
            J2 = J1 + 9; IF J2 > NPAR THEN J2 = NPAR;
            A  = G(|1:3,J1:J2|);
            PN = PARLAB(|J1:J2|);
            IF J1 = 1 THEN
                    PRINT "FINAL PARAMETER ESTIMATES AND STANDARD ERRORS",,
                            A(|COLNAME=PN ROWNAME=CN FORMAT=8.3|);
                ELSE PRINT A(|COLNAME=PN ROWNAME=CN FORMAT=8.3|);
        END;
                IF SUM(STAT(|1:NPAR|)) > 0 THEN PRINT S1;
        PRINT    "ESTIMATED COVARIANCE MATRIX",,
            SIGMA (|ROWNAME=LABELS COLNAME=LABELS FORMAT=10.3|);
        PRINT / "RESIDUAL COVARIANCE MATRIX",,
            RESID (|ROWNAME=LABELS COLNAME=LABELS FORMAT=10.3|);
        PRINT    " LIKELIHOOD RATIO TEST STATISTIC",,
            CHISQ (|COLNAME="CHI-SQARE" FORMAT=10.3|)
            DF    (|COLNAME="DF"        FORMAT=5.0|)
            P     (|COLNAME="PROB"      FORMAT=8.3|);
```

```
*------------------------------------------------------------------*
*  APDER:  GRADIENT AND APPROXIMATE HESSIAN                        *
*------------------------------------------------------------------*;
     START APDER (G,H,SAMCOV,SIG,Q1,PAR,NUMPAR);
        SMALL = 0.0000001;           V2 = PAR;
        G     = J(NUMPAR,1,0);        H  = J(NUMPAR,NUMPAR,0);
        Q2    = Q1 * (SIG - SAMCOV);

*--- ELEMENTS OF GRADIENT ----------------------------------------*;
        DO I = 1 TO NUMPAR;
           V2(|I,1|) = PAR(|I,1|) + SMALL;
           RUN MODEL (SH,V2);
           W        = (SH - SIG) / SMALL;
           Q3       = Q1 * W;
           G(|I,1|) = TRACE (Q2 * Q3);

*--- ELEMENTS OF HESSIAN -----------------------------------------*;
           V2(|I|) = PAR(|I|);
           DO J = 1 TO I;
              V2(|J|)   = PAR(|J|) + SMALL;
              RUN MODEL (SH,V2);
              W         = (SH - SIG) / SMALL;
              Q4        = Q1 * W;
              H(|J,I|)  = TRACE (Q3 * Q4);
              V2(|J|)   = PAR(|J|);
           END;
        END;
        H = H + H` - DIAG (H);
     FINISH;    * END OF APDER MODULE;


*------------------------------------------------------------------*
*  INVCHK: DETERMINE IF A MATRIX IS POSITIVE DEFINITE.             *
*          IF YES, RETURN THE INVERSE                              *
*------------------------------------------------------------------*;
     START INVCHK (SING,C,A1,A2,STAT);
        A    = (A1 || A2) // (A2` || 1);
        C    = SWEEP (A,1:NROW(A1));
        STAT = (C(|,+|) = 0);
        SING = STAT(|+,|);
     FINISH; * END OF INVCHK PROGRAM;

STOP: FINISH; * END OF PROGRAM;


RUN GN;
```

References

Anderson, T. W. (1960). Some stochastic process models for intelligence test scores. In K. J. Arrow, S. Karlin, & P. Suppes (Ed.), *Mathematical methods in the social sciences* (pp. 205–220). Stanford, CA: Stanford University Press.

Bentler, P. M., & Lee, S.-L. (1983). Covariance structures under polynomial constraints: Applications to correlation and alpha-type structural models. *Journal of Educational Statistics, 8*, 207–222.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179–197.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). 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*, 62–83.

Browne, M. W., & du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research, 27*, 269–300.

Burden, R. L., & Faires, J. D. (1989). *Numerical analysis* (4th ed.). Boston: Prindle, Weber & Schmidt.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Fraser, C. (1979). *COSAN: User's guide.* Toronto: Ontario Institute for Studies in Education, Department of Measurement, Evaluation and Computer Applications.

Froemel, E. A. (1971). A comparison of computer routines for the calculation of the tetrachoric correlation coefficient. *Psychometrika, 36*, 165–173.

Green, D. P., & Palmquist, B. L. (1991). More "tricks of the trade": Reparameterizing LISREL models using negative variances. *Psychometrika, 56*, 137–145.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology, 59*, 375–380.

Jennrich, R. I., & Sampson, P. F. (1966). Application of stepwise regression to nonlinear estimation. *Technometrics, 10*, 63–72.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*, 121–145.

Jöreskog, K. G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics, 8*, 65–92.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 user's guide*. Mooresville, IN: Scientific Software.

Lee, S.-Y., & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika, 44*, 99–113.

Lee, S.-Y., & Jennrich, R. I. (1984). The analysis of structural equation models by means of derivative free nonlinear least squares. *Psychometrika, 49*, 521–528.

Long, J. S. (1990). [Review of software for structural equation modeling: EQS, EQS-EM, EzPATH, LIS-COMP, LISREL 7]. *Journal of Marketing Research, 37*, 372–378.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monograph Number 15, 32* (4, Pt. 2).

McDonald, R. P. (1978). A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 31*, 59–72.

McDonald, R. P. (1985). Unidimensional and multidimensional models for item response theory. In D. J. Weiss (Ed.), *Proceedings of the 1982 item response theory and computerized adaptive testing conference* (pp. 127–148). Minneapolis, MN: University of Minnesota, Department of Psychology.

Mooijaart, A. (1983). Two kinds of factor analysis for ordered categorical variables. *Multivariate Behavioral Research, 18*, 423–441.

Morrison, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Pickle, L. W. (1991). Maximum likelihood estimation in the new computing environment. *Statistical Computing and Statistical Graphical Newsletter, 3*, 6–15.

Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika, 48*, 73–83.

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*, 37–47.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*, 51–67.

SAS Institute. (1985). *SAS/IML user's guide, version 5*. Cary, NC: Author.

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology, 28*, 138–151.

Swain, A. J. (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika, 40*, 315–335.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393–408.

Thisted, R. A. (1988). *Elements of statistical computing*. New York: Chapman and Hall.

Werts, C., Linn, R. L., & Jöreskog, K. G. (1978). Reliability of college grades from longitudinal data. *Educational and Psychological Measurement, 38*, 89–95.

Wheaton, B., Muthén, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D. R. Heise, (Ed.), *Sociological methodology* (pp. 84–136). San Francisco: Jossey-Bass.

Wiley, J. A., & Wiley, M. G. (1974). A note on correlated errors in repeated measurements. *Sociological Methods and Research, 3*, 172–188.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117*, 319–331.

Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika, 55*, 255–262.