PSYCHOMETRIKA—VOL. 43, NO. 3.
SEPTEMBER, 1978

STATISTICAL ASPECTS OF A THREE-MODE FACTOR ANALYSIS MODEL

P. M. BENTLER

UNIVERSITY OF CALIFORNIA, LOS ANGELES

SIK-YUM LEE

THE CHINESE UNIVERSITY OF HONG KONG

A special case of Bloxom's version of Tucker's three-mode model is developed statistically. A distinction is made between modes in terms of whether they are fixed or random. Parameter matrices are associated with the fixed modes, while no parameters are associated with the mode representing random observation vectors. The identification problem is discussed, and unknown parameters of the model are estimated by a weighted least squares method based upon a Gauss-Newton algorithm. A goodness-of-fit statistic is presented. An example based upon self-report and peer-report measures of personality shows that the model is applicable to real data. The model represents a generalization of Thurstonian factor analysis; weighted least squares estimators and maximum likelihood estimators of the factor model can be obtained using the proposed theory.

Key words: factor analysis, three-mode models.

Two generalizations of principal components analysis have been proposed for the analysis of data that can be cross-classified in terms of three modes of measurement. The pioneering and most comprehensive model is the one developed by Tucker [1966]. It postulates that an observed three-way data observation can be decomposed into components attributable to each of the three modes as well as an internal core matrix. The components of each of the modes are of dimensionality equal to or smaller than the number of variables in that mode, and the three dimensions can be unequal. The core matrix, like the original data, represents a three-way data observation, but of order given by the dimensionality of the components rather than the variables. In contrast, the model of Carroll and Chang [1970] is a restricted version of the Tucker model: a three-way observation is similarly decomposed into components associated with each of the modes of classification, but the component dimensionality of each mode is required to be equal, and the core matrix is "a kind of 3-way analogue of an identity matrix" [Carroll & Chang, 1970, p. 310]. These models are most properly called component models rather than factor analytic models because there is no specific provision for the concept of uniqueness. Bloxom [1968], however, has rewritten the Tucker model in a covariance structure form that includes the concept of uniqueness. We shall limit ourselves to the factor analytic case of three-mode models, and thus ignore the multidimensional scaling specialization.

In this paper we develop a three-mode model that can be considered to be a special case of Bloxom's [1968] version of Tucker's [1966] model. As a consequence, the proposed model is quite restricted in parametric structure. This restriction creates a clear drawback

This investigation was supported in part by a Research Scientist Development Award (K02-DA00017) and a research grant (DA01070) from the U. S. Public Health Service. The very helpful comments of several anonymous reviewers are gratefully acknowledged.

Requests for reprints should be sent to P. M. Bentler, Department of Psychology, University of California, Los Angeles, California 90024.

to the general applicability of the model to a wide variety of data, but it has the advantage that when the model fits three-mode data, it is extremely economical in terms of the number of parameters as compared either to Bloxom's model or to more general models such as those of Jöreskog [1973] or Bentler [1976]. (In the multidimensional scaling area, the Carroll-Chang model is sometimes also preferred to the Tucker model for similar reasons of economy of representation.) Nonetheless, the model represents more than a mathematical exercise, since it can be shown to be applicable to real data in the three-mode context as well as being, conceptually, a generalization of two-mode factor analysis. The statistical aspects of fitting the model to the data via the covariance matrix are emphasized, particularly, in obtaining consistent, asymptotically efficient parameter and standard error estimates, and a goodness of fit test statistic. A computer program is developed to implement the basic theory.

The Model

Consider the $(pm \times 1)$ observed random vector x with structure

$$x = \mu + (A \otimes B)(\alpha \otimes \beta) + Z\zeta . \tag{1}$$

The vector μ is a vector of population means. The notation $(A \otimes B)$ refers to the right Kronecker product of matrices $(A \otimes B) = [a_{ij}B]$. The matrices A, B, and Z are parameter matrices of order $(p \times k)$, $(m \times r)$, and $(pm \times pm)$, respectively; the matrix Z is taken to be diagonal. The parameter matrix A corresponds to one variable classification mode, and the matrix B to the other mode. The data for the third mode is not represented by parameters, only by random vectors. The vectors α, β, and ζ are random vectors of order $(k \times 1)$, $(r \times 1)$, and $(pm \times 1)$, respectively. We assume that the expectations $\mathcal{E}(\alpha\alpha')$, $\mathcal{E}(\beta\beta')$, and $\mathcal{E}(\zeta\zeta')$ are given by ϕ_α, ϕ_β, and I_ζ (the latter is an identity matrix of appropriate order), and that the random vectors α, β, and ζ are statistically independent of each other. Consequently, we obtain as $\mathcal{E}(x-\mu)(x-\mu)'$ the covariance structure

$$\Sigma = (A\phi_\alpha A' \otimes B\phi_\beta B') + Z^2 . \tag{2}$$

Suppose that A is of dimension (1×1), the unit scalar. It follows immediately that

$$\Sigma = B\phi_\beta B' + Z^2 , \tag{3}$$

which is the oblique generalization of Thurstone's [1947] multiple factor model. Principal components are a special case with $\phi_\beta = I$ and $Z^2 = 0$. In this paper we concentrate on the simple three-mode case of (2) given by

$$\Sigma = (AA' \otimes BB') + Z^2 , \tag{4}$$

where the factors are considered to be orthogonal.

The model proposed by Tucker [1966], as modified by Bloxom [1968], on the other hand, can be represented in the present context as

$$x = (A \otimes B)G\gamma + Z\zeta \tag{5}$$

where G is a $(kr \times t)$ matrix representation of the core box, and γ is a $(t \times 1)$ random vector for the third mode. The cross product structure (possibly, covariance structure) is given by

$$\Sigma = (A \otimes B)G\phi G'(A \otimes B)' + Z^2 \tag{6}$$

where ϕ is a symmetric nonnegative definite matrix of covariances among the γ. In the current model (2) we take $G\phi G' = (\phi_\alpha \otimes \phi_\beta)$, and in (4) $G\phi G' = I$. Although these assumptions are restrictive, in some contexts they are not unreasonable, and they make the statistical estimation of the parameters a tractable problem.

Jöreskog [1973] has introduced the second-order factor model

$$\Sigma = B(\Lambda\phi\Lambda' + \Psi^2)B' + \theta^2 \,. \tag{7}$$

If one equates the model matrices of equations (4) and (7) as $B = (A \otimes I)$, $\Lambda = (I \otimes B)$, $\phi = I$, $\Psi^2 = 0$, and $\theta^2 = Z^2$, it follows that (4) is a restricted version of (7); for example, $(A \otimes I)$ is a $(pm \times km)$ factor pattern matrix in which a number of elements are constrained to be equal and others are fixed as zero. Consequently, the model (4) can be estimated via programs that implement (7), specifically via ACOVS [Jöreskog, Gruvaeus & van Thillo, 1970]. In this paper we develop a class of estimation methods that take into account the special nature of the Kronecker product and can accept singular estimated and data covariance matrices.

Parameter Identification

It is obvious that the parameter matrices A and B in (1) can be replaced by AT_1, and BT_2, where T_1 and T_2 are orthonormal transformation matrices. The consequent covariance structure (4) remains identical. In order to fix A and B uniquely, we set the elements a_{ij} of A equal to zero if $j > i$, and similarly $b_{ij} = 0$ if $j > i$. This procedure follows Anderson and Rubin [1956], and it is appropriate when estimating the parameters in an exploratory situation. The resulting matrices can be individually transformed later, if desired, or placed into the well-known oblique forms $A\phi_\alpha A'$ and $B\phi_\beta B'$. When (2) is treated as a confirmatory model, the fixed elements may be nonzero and may appear anywhere, provided the parameters remain identified.

There exists another indeterminacy in (2). Any Kronecker product $(X \otimes Y)$ can be equivalently written as $(kX \otimes k^{-1}Y)$, where k is a nonzero constant. Consequently, the scale of the parameters for each mode are only defined relative to each other, and the scale of one mode can be fixed arbitrarily. This is done in the current work by setting $a_{11} = 1.0$.

Estimation Method

Let $\Sigma_o = \Sigma(\theta_o)$ be the covariance structure defined by (4) where θ_o is the q by 1 vector of the unknown parameters in A, B, and Z^2. In general, we regard θ as a vector of mathematical variables and $\Sigma = \Sigma(\theta)$ as a function of θ. Let the random vectors x_i, $i = 1, \cdots, N$ of (1) be independently and multivariate normally distributed with mean vector μ and covariance matrix $\Sigma(\theta)$. Let $n = N - 1$, S be the sample covariance matrix, and $\tilde{S}$ be the vector formed from the elements on the diagonal and below, taken row by row. Thus, $\tilde{S} = (S_{11}, S_{21}, S_{22}, \cdots, S_{pp})$.

Browne [1974] has shown that the residual quadratic form

$$(\tilde{S} - \tilde{\Sigma})' \operatorname{cov}(\tilde{S})^{-1} (\tilde{S} - \tilde{\Sigma}) \tag{8}$$

is asymptotically equal to

$$\frac{n\operatorname{tr}[(S - \Sigma)\Sigma_o^{-1}]^2}{2} \,. \tag{9}$$

Since Σ_o^{-1} is unknown, it is replaced by some weight matrix V and (9) becomes, except for a constant, the weighted least squares function

$$Q(V, \theta) = \frac{\operatorname{tr}[(S - \Sigma)V]^2}{2} \,. \tag{10}$$

When $V = S^{-1}$ exists, (10) is of special interest and becomes

$$Q(\theta) = \frac{\operatorname{tr}[(S - \Sigma)S^{-1}]^2}{2} \,. \tag{11}$$

By definition, the weighted least squares estimator $\hat{\theta}$ is the vector θ such that $Q(\theta)$ is minimized, provided such a vector exists. It can be shown that this estimator is scale-free, so that the correlation matrix can be analyzed in place of the covariance matrix. Under mild regularity conditions, Browne [1974] and Lee [1977] proved that $\hat{\theta}$ possesses the following asymptotic properties:

(i) The weighted least squares estimator $\hat{\theta}$ is consistent.
(ii) The weighted least squares estimator $\hat{\theta}$ and the maximum likelihood estimator are asymptotically equivalent.
(iii) The weighted least squares estimator $\hat{\theta}$ is a "best weighted least squares" estimator in the sense that for any other weighted least squares estimator $\hat{\theta}^*$, $\mathrm{cov}(\hat{\theta}^*) - \mathrm{cov}(\hat{\theta})$ is positive semidefinite.
(iv) The asymptotic distribution of $\hat{\theta}$ is multivariate normal with mean vector θ_o and covariance matrix $(2n^{-1})[\partial\Sigma_o(\Sigma_o^{-1} \otimes \Sigma_o^{-1})\partial\Sigma_o']^{-1}$ where $\partial\Sigma_o = \partial\Sigma_o/\partial\theta = \partial\Sigma_o/\partial\theta|_{\theta=\theta_o}$.
(v) The asymptotic distribution of $nQ(\hat{\theta}) = 2^{-1} n\mathrm{tr}\,[(S - \Sigma)S^{-1}]^2$ is χ^2, with degrees of freedom equal to $[pm(pm + 1)/2] - q$.

The loss function $Q(\theta)$ measures how well the model fits the data. Property (v) enables us to test the hypothesis that Σ_o has the structure (4) against the alternative that Σ_o is any symmetric positive definite matrix.

Minimization Procedure

The weighted least squares estimates are determined as the solution of $\partial Q/\partial\theta = 0$. These equations cannot be solved algebraically, so that some iterative procedure has to be used. The most popular nonlinear algorithms are the methods of steepest descent, Fletcher-Powell, Fletcher-Reeves, Gauss-Newton, and Newton-Raphson. The steepest descent method tends to be very slow. It is shown by Lee [1977] that the Gauss-Newton algorithm is at least comparable in speed to the Fletcher-Powell and Fletcher-Reeves algorithms and, moreover, the Fletcher-Powell and Fletcher-Reeves algorithms may produce questionable estimates of the standard errors of the estimators. The Newton-Raphson algorithm is not robust to bad starting values, and it requires the second order partial derivatives of the function. For complicated covariance structures, such as (4), these second order partial derivatives are difficult to derive and take a long time to compute. Hence, we prefer to use the Gauss-Newton algorithm to obtain the estimates. The algorithm is conceptually simple, usually quite robust to bad starting values, and requires only the first-order partial derivatives of the function.

For the weighted least squares function $Q(\theta)$, Lee [1977] showed that a step of the Gauss-Newton algorithm consists of

$$\Delta\theta_i = -U_i^{-1} \left.\frac{\partial Q}{\partial\theta}\right|_{\theta=\theta}, \tag{12}$$

where $U_i = \partial\Sigma(S^{-1}\otimes S^{-1})\partial\Sigma' \mid_{\theta=\theta_i}$ and $\partial Q/\partial\theta = \partial\Sigma(S^{-1}\otimes S^{-1})(\tilde{S} - \tilde{\Sigma})'$. ($\tilde{S}$ denotes a row vector which takes all elements of S row by row.) After $\Delta\theta_i$ has been evaluated, a new θ_{i+1} is taken to be

$$\theta_{i+1} = \theta_i + \Delta\theta_i\,. \tag{13}$$

The process is repeated until the root mean square of $\Delta\theta_i$ or the root mean square of the gradient vector, $\partial Q/\partial\theta$, is sufficiently small.

Although in most cases U_i is at least positive semidefinite, the algorithm frequently fails when it is not modified. A useful modification consists of "step-halving", in which $\Delta\theta$, $\Delta\theta/2$, $\cdots$, are chosen until a step is found which reduces $Q(\theta)$. If U_i is singular or nearly singular, the techniques of Marquardt [1963] or Greenstadt [1967] can be used to replace U_t by a positive definite matrix. Alternatively, the stepwise regression method of Jennrich and Sampson [1968] can be used.

Under mild regularity conditions, Lee [1977] showed that in the Gauss-Newton algorithm, as θ converges to $\hat{\theta}$, $U_t(\theta)$ converges to $U(\hat{\theta})$ and $2U(\hat{\theta})^{-1}/n$ is asymptotically equivalent to the inverse of the information matrix based on the distribution of S. Hence, the standard error estimates are naturally taken to be the square roots of the diagonal elements of $2U(\hat{\theta})^{-1}/n$. As a result, the standard error estimates can be obtained easily.

The derivatives required to implement this algorithm in the current context are obtained in the Appendix.

Example

To illustrate the model and the methods of analysis that have been developed, we obtained the intercorrelations among four variables measured by each of two methods. The four variables are: ambition, attractiveness, leadership, and extraversion. The two methods of measurement are: self-report and peer-report. Seventy-two subjects were asked to describe themselves using bipolar adjectives and short phrases; each variable represents a composite score computed across twenty items. Subjects were asked to bring friends, and the friend was asked to describe the subject using the same adjectives and phrases. The intercorrelation matrix among these variables is presented in Table 1, which has been organized so that the first four variables represent the self-report data, and the second four variables represent the peer-report data.

TABLE 1

Correlations Among Four Variables Assessed by Two Methods of Measurement

	1	2	3	4	5	6	7	8
1	1.000							
2	.223	1.000						
3	.337	.418	1.000					
4	.223	.290	.693	1.000				
5	.402	.070	.226	.210	1.000			
6	.035	.442	.251	.219	.233	1.000		
7	.160	.196	.603	.639	.379	.314	1.000	
8	.093	.180	.451	.645	.269	.283	.582	1.000

TABLE 2

Parameter Estimates and Standard Errors (In Parentheses)

	A		B	
	I		I	II
1	1.00*	1	.74 (.11)	0*
2	.85 (.10)	2	.30 (.13)	.36 (.12)
		3	.41 (.13)	.71 (.10)
		4	.27 (.13)	.80 (.09)

Z

1	.63 (.11)	2	.71 (.08)	3	.52 (.07)	4	.47 (.07)
5	.69 (.09)	6	.74 (.08)	7	.58 (.06)	8	.63 (.07)

*Fixed value

The analysis was designed to uncover a common factor structure within the self-report and peer-report data, as well as a possibly quantitative cross-method relation. The matrix B was taken to represent the content structure of the variables, which was assumed to be of dimensionality two. An initial factor loading matrix for B was obtained by a principal factor analysis of the self-report data. The matrix A was assumed to be single-dimensional, with the two methods of measurement initialized as having an equal weight. An initial estimate of Z^2 was obtained from the uniquenesses of the principal factor analysis, with uniquenesses for the two methods taken as equal. The program took 23 iterations to converge upon the parameter and standard error estimates reported in Table 2. The criterion for convergence was $\text{RMS}(\Delta\theta) \leq .0001$ or $\text{RMS}(\partial Q/\partial\theta) \leq .0001$. The solution yielded $\chi^2 = 29.86$. With the sixteen unknown parameters of Table 2, there are 20 degrees of freedom. From the chi-square table, it is evident that the model (4) cannot be rejected at $\alpha = .05$ level.

The interpretive meaning of these results appears to be as follows. A common two-dimensional factor structure (B) accounts for the correlations among the variables, when taking into account the relative weights that must be applied to the factor structure for each of the methods of measurement (1.0 for self-report, .85 for peer-report). According to (1), $(A \otimes B)$ provides the Thurstonian factor loadings for all eight variables. Thus, the complete 8×2 loading matrix would be given by $[B', .85B']'$. Unlike ordinary factor analysis, however, the complete loading matrix is not needed to reproduce the common portions of the data. Obviously, B is taken as the estimator that reproduces the correlations among self-report variables, via BB'. The correlations among peer-report variables, estimated at $.85^2BB'$, are systematically somewhat smaller than among self-report variables, but only marginally so in view of the standard error of .10 associated with the multiplier .85. The reproduced cross-method correlations, given by $.85BB'$, are estimated at an intermediate value. Finally, turning to the psychological content of the variables under consideration, it appears that Variable 1 (ambition) assesses a different dimension from Variables 3 and 4 (leadership and extraversion). Variable 2 (attractiveness) is not a particularly good indicator of either dimension. In this instance, there appears to be no particular reason to transform B to a more meaningful factorial structure, although an oblique transformation might produce a cleaner simple structure.

Discussion

The statistical approach to three-mode models rather naturally considers these models as covariance structure models of a particular sort, and, when this approach is taken, the statistical ideas associated with such models become relevant [e.g., Jöreskog, 1973]. The simplest conceptual notion is that of variables and random observations on the variables. Some of the modes can quite obviously be taken to represent measurable variables in the traditional sense, though the factor structure that generates the variables is a very specialized one. On the other hand, one of the modes must be taken to represent no more than a series of observations on the variables, having no associated parameters. Such a distinction between variables and observations on variables is quite a natural one for the area of psychological testing where subjects describe themselves, their friends, or their knowledge with respect to the specific variables under investigation. No ambiguity about this distinction would seem possible for the data used to illustrate the proposed model and estimation method. In other circumstances, however, an ambiguity might arise.

Our approach to estimating model (4) has involved the statistical theory associated with weighted least-squares estimation in the context of multi-normal variables. Since the ordinary factor model is a special case of the model developed in the current paper, it is apparent that the procedures we have developed also include as a special case the generalized least-squares approach to the Thurstonian model as developed by Jöreskog and Goldberger [1972]. Furthermore, Lee [1977] has proven that when the weight matrix V in (10) is replaced at every iteration by the matrix $[\Sigma_i(\theta)^{-1}]$, if the iterative procedure converges, it will converge to the maximum-likelihood estimator. Consequently, (4) can be utilized with our equations to provide maximum-likelihood solutions to the exploratory factor analysis model, as described by Jöreskog [1967]. In spite of this generality, it must be acknowledged that the current model remains too specialized to be able to fit the wide variety of data envisioned by Tucker [1966]. In the particular example being investigated, the correlation matrices within and between each of the two methods of measurement were assumed to be generated by a simple proportional process, except for the uniquenesses. Such a process was found to be useful because of its economy of representation; alternative models, such as those based on separate additive common factors for content and method components of the variables, would require a greater number of parameters. Nonetheless, it is clear that an important task for the future must include the statistical development of Tucker's more general model, as modified, perhaps, by Bloxom [1968]. A statistical approach to the more general model would, of course, also subsume the developments of the current paper. However, we believe that our results demonstrate that a statistical approach to multimode models can be pursued successfully.

If a researcher does not wish to adopt a factor analytic multimode model, perhaps because of such concerns as factor indeterminacy, or small sample size, the principal component equivalent of (4) with $Z^2 = 0$ could be adopted. In such an instance, ordinary least squares estimation might be desirable. This can be accomplished with the current equations by setting $V = I$ in (9). However, it should be recognized that in the absence of relevant statistical theory, the inferential methods associated with the current approach cannot be applied.

Appendix: Partial Derivatives

In order to implement the Gauss-Newton method, the partial derivatives $\partial\Sigma$ are required. We will obtain these derivatives with the help of the following additional notation.

For any diagonal matrix Z, let D_Z denote the column vector of all diagonal elements of Z.

Let $K_{(ij)}^{\{p\}}$ denote a $(p \times p)$ matrix with the unit scalar in the $(ij)^{\text{th}}$ position, and with zeros elsewhere.

Let $X_{(ij)}^{\{p\}}$ denote the matrix $(K_{(ij)}^{\{p\}} \otimes X)$.

For any $(p \times p)$ matrix $Y = [y_{ij}]$, let $Y_{(u\,v)}^{\{pm\}}$ denote the partitioned matrix whose $(i, j)^{\text{th}}$ block is given by $y_{ij(u\,v)}^{\{m\}}$. Note that $y_{ij(u\,v)}^{\{m\}}$ is a $(m \times m)$ matrix, whose $(u, v)^{\text{th}}$ element is equal to y_{ij}.

Let $A^* = AA' = [a_{rs}^*]$ and $B^* = BB' = [b_{rs}^*]$.

We are now ready to proceed with the differentiation. We shall require two components, and then we combine the results to yield the required derivative. First we note that for the element a_{ij} of $A = [a_{ij}]$,

$$\frac{\partial a_{rs}^*}{\partial a_{ij}} = \sum_{t=1}^{k} \left[a_{rt} \frac{\partial a_{st}}{\partial a_{ij}} + a_{st} \frac{\partial a_{rt}}{\partial a_{ij}} \right]$$

$$\text{(A1)} \qquad = \begin{cases} 0 & r \neq i, \quad s \neq i \\ a_{rj} & s = i \\ a_{sj} & r = i \\ 2a_{ij} & r = s = i \end{cases} .$$

From the definition, we obviously have

$$\text{(A2)} \qquad \frac{\partial \Sigma}{\partial a_{rs}^*} = \frac{\partial (AA' \otimes BB')}{\partial a_{rs}^*} = (BB')_{(rs)}^{\{p\}} .$$

By elementary calculus, we obtain

$$\frac{\partial \Sigma}{\partial a_{ij}} = \sum_{s=1}^{p} \sum_{r=1}^{p} \frac{\partial a_{rs}^*}{\partial a_{ij}} \frac{\partial \Sigma}{\partial a_{rs}^*}$$

$$\text{(A3)} \qquad = \sum_{s=1}^{p} \sum_{r=1}^{p} \frac{\partial a_{rs}^*}{\partial a_{ij}} (BB')_{(rs)}^{\{p\}} .$$

We now expand (A3) and simplify the results using (A1) repeatedly.

$$\frac{\partial \Sigma}{\partial a_{ij}} = \sum_{s=1}^{p} \left[\frac{\partial a_{1s}^*}{\partial a_{ij}} (BB')_{(1s)}^{\{p\}}) + \cdots + \sum_{s=1}^{p} \left(\frac{\partial a_{ps}^*}{\partial a_{ij}} (BB')_{(ps)}^{\{p\}} \right]$$

$$= \frac{\partial a_{1i}^*}{\partial a_{ij}} (BB')_{(1i)}^{\{p\}} + \cdots + \sum_{s=1}^{p} \frac{\partial a_{is}^*}{\partial a_{ij}} (BB')_{(is)}^{\{p\}} + \cdots + \frac{\partial a_{pi}^*}{\partial a_{ij}} (BB')_{(pi)}^{\{p\}}$$

$$= \sum_{r \neq i}^{p} \frac{\partial a_{ri}^*}{\partial a_{ij}} (BB')_{(ri)}^{\{p\}} + \sum_{s \neq i}^{p} \frac{\partial a_{is}^*}{\partial a_{ij}} (BB')_{(is)}^{\{p\}} + \frac{\partial a_{ii}^*}{\partial a_{ij}} (BB')_{(ii)}^{\{p\}}$$

$$= \sum_{r \neq i}^{p} a_{rj} (BB')_{(ri)}^{\{p\}} + \sum_{s \neq i}^{p} a_{sj} (BB')_{(is)}^{\{p\}} + 2a_{ij} (BB')_{(ii)}^{\{p\}} .$$

This can be written in the final form as

$$\text{(A4)} \qquad \frac{\partial \Sigma}{\partial a_{ij}} = \sum_{r=1}^{p} a_{rj}(BB')^{\{p\}}_{(r\;i)} + \sum_{s=1}^{p} a_{sj}(BB')^{\{p\}}_{(i\;s)} .$$

We obtain the complete result by analogous procedures to yield the matrix

$$\text{(A5)} \qquad \frac{\partial \Sigma}{\partial A} = \begin{bmatrix} \overline{\partial \Sigma / \partial_{a11}} \\ \cdots \\ \overline{\partial \Sigma / \partial a_{1k}} \\ \cdots \\ \overline{\partial \Sigma / \partial a_{pk}} \end{bmatrix} .$$

(Recall from (12) that the notation $\overline{S}$ denotes a row vector which takes all elements of S row by row.)

Now we determine the derivatives for the right-hand matrices in the Kronecker product. From the definition, we have

$$\text{(A6)} \qquad \frac{\partial \Sigma}{\partial b_{ij}{}^{*}} = (AA')^{\{pm\}}_{(i\;j)} .$$

By methods similar to those used above we obtain

$$\text{(A7)} \qquad \begin{aligned} \frac{\partial \Sigma}{\partial b_{ij}} &= \sum_{s=1}^{m} \sum_{r=1}^{m} \frac{\partial b_{rs}{}^{*}}{\partial b_{ij}} \frac{\partial \Sigma}{\partial b_{rs}{}^{*}} \\ &= \sum_{s=1}^{m} \sum_{r=1}^{m} \frac{\partial b_{rs}{}^{*}}{\partial b_{ij}} (AA')^{\{pm\}}_{(r\;s)} \\ &= \sum_{r=1}^{m} b_{rj}(AA')^{\{pm\}}_{(r\;i)} + \sum_{s=1}^{m} b_{sj}(AA')^{\{pm\}}_{(i\;s)} . \end{aligned}$$

Again we form the final matrix

$$\text{(A8)} \qquad \frac{\partial \Sigma}{\partial B} = \begin{bmatrix} \overline{\partial \Sigma / \partial b_{11}} \\ \cdots \\ \overline{\partial \Sigma / \partial b_{1r}} \\ \cdots \\ \overline{\partial \Sigma / \partial b_{mr}} \end{bmatrix} .$$

Finally, it can be easily shown that

$$\text{(A9)} \qquad \left(\frac{\partial \Sigma}{\partial D_z}\right)_{ij} = \begin{cases} 2Z_{ii} \text{ if } j = pm(i-1) + i \\ 0 \text{ otherwise} \end{cases}$$

where $0 < i \leq pm$.

Equations (A5), (A8) and (A9) provide the required derivatives for $\partial \Sigma$. Since the fixed parameters are no longer unknown mathematical variables, the corresponding rows from $\partial \Sigma$ are eliminated.

REFERENCES

Anderson, T. W., & Rubin, H. Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1956, *5*, 111–150.

Bentler, P. M. Multistructure statistical model applied to factor analysis. *Multivariate Behavioral Research*, 1976, *11*, 3–25.

Bloxom, B. A note on invariance in three-mode factor analysis. *Psychometrika*, 1968, *33*, 347–350.

Browne, M. W. Generalized least square estimators in the analysis of covariance structures. *South African Statistical Journal*, 1974, *8*, 1–24.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, *35*, 283–319.

Greenstadt, J. On the relative efficiencies of gradient methods. *Mathematical Computation*, 1967, *21*, 360–367.

Jennrich, R. I., & Sampson, P. F. Application of stepwise regression to nonlinear estimation. *Technometrics*, 1968, *40*, 63–72.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 1967, *32*, 443–482.

Jöreskog, K. G. Analysis of covariance structures. In P. R. Krishnaiah (Ed.), *Multivariate analysis III*. New York: Academic Press, 1973, pp. 263–285.

Jöreskog, K. G., & Goldberger, A. S. Factor analysis by generalized least squares. *Psychometrika*, 1972, *37*, 243–260.

Jöreskog, K. G., Gruvaeus, G. T., & van Thillo, M. *RB-70-15: ACOVS, a general computer program for the analysis of covariance structures*. Princeton, N.J.: Educational Testing Service, 1970.

Lee, S. Y. *Some algorithms for covariance structure analysis*. Unpublished doctoral dissertation, University of California, Los Angeles, 1977.

Marquardt, D. W. An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal*, 1963, *11*, 431–441.

Thurstone, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966, *31*, 279–311.

Manuscript received 7/13/76
First revision received 8/16/77
Final version received 2/13/78