Biometrika (1970), **57**, 2, *p.* 239
With 1 *text-figure*
Printed in Great Britain

A general method for analysis of covariance structures

BY K. G. JÖRESKOG
Educational Testing Service, Princeton, New Jersey

SUMMARY

It is assumed that observations on a set of variables have a multivariate normal distribution with a general parametric form of the mean vector and the variance-covariance matrix. Any parameter of the model may be fixed, free or constrained to be equal to other parameters. The free and constrained parameters are estimated by maximum likelihood. A wide range of models is obtained from the general model by imposing various specifications on the parametric structure of the general model. Examples are given of areas and problems, especially in the behavioural sciences, where the method may be useful.

1. GENERAL METHODOLOGY

1·1. *The general model*

We consider a data matrix $\mathbf{X} = \{x_{ai}\}$ of N observations on p response variables and the following model. Rows of $\mathbf{X}$ are independently distributed, each having a multivariate normal distribution with the same variance-covariance matrix $\mathbf{\Sigma}$ of the form

$$\mathbf{\Sigma} = \mathbf{B}(\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}^2)\mathbf{B}' + \mathbf{\Theta}^2, \tag{1}$$

and mean vectors given by

$$E(\mathbf{X}) = \mathbf{A}\mathbf{\Xi}\mathbf{P}, \tag{2}$$

where $\mathbf{A} = \{a_{as}\}$ is an $N \times g$ matrix of rank g and $\mathbf{P} = \{p_{ti}\}$ is a $h \times p$ matrix of rank h, both being fixed matrices with $g \leqslant N$ and $h \leqslant p$; $\mathbf{\Xi} = \{\xi_{st}\}$, $\mathbf{B} = \{\beta_{ik}\}$, $\mathbf{\Lambda} = \{\lambda_{km}\}$, the symmetric matrix $\mathbf{\Phi} = \{\phi_{mn}\}$, and the diagonal matrices $\mathbf{\Psi} = \{\delta_{kl}\psi_k\}$ and $\mathbf{\Theta} = \{\delta_{ij}\theta_i\}$ are parameter matrices. Potthoff & Roy (1964) considered the model (2) without a parametric form of $\mathbf{\Sigma}$.

Thus the general model is one where means, variances and covariances are structured in terms of other sets of parameters that are to be estimated. We shall allow for any one of the parameters in $\mathbf{\Xi}$, $\mathbf{B}$, $\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\mathbf{\Psi}$ and $\mathbf{\Theta}$ to be known *a priori* and for one or more subsets of the remaining parameters to have identical but unknown values. Thus parameters are of three kinds: (i) fixed parameters that have been assigned given values; (ii) constrained parameters that are unknown but equal to one or more other parameters; and (iii) free parameters that are unknown and not constrained to be equal to any other parameter.

It should be noted that if $\mathbf{B}$ is replaced by $\mathbf{B}\mathbf{T}_1^{-1}$, $\mathbf{\Lambda}$ by $\mathbf{T}_1\mathbf{\Lambda}\mathbf{T}_2^{-1}$, $\mathbf{\Phi}$ by $\mathbf{T}_2\mathbf{\Phi}\mathbf{T}_2'$ and $\mathbf{\Psi}^2$ by $\mathbf{T}_1\mathbf{\Psi}^2\mathbf{T}_1'$ while $\mathbf{\Theta}$ is left unchanged, then $\mathbf{\Sigma}$ is unaffected. This holds for all nonsingular matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ such that $\mathbf{T}_1\mathbf{\Psi}^2\mathbf{T}_1'$ is diagonal. Thus there is a great deal of indeterminacy in the general model. Hence, in order to obtain a unique set of parameters and a corresponding unique set of estimates, some additional restrictions must be imposed. In some cases these restrictions are given in a natural way by the particular application of the model. In other cases they can be chosen in any convenient way by specifying certain parameters to be fixed to certain values; see §§2·2–2·6.

1·2. *Estimation in the general model*

The main problem considered in this paper is the estimation of the free and constrained parameters of the general model. Since in most cases simple analytical solutions do not exist, the likelihood function is maximized numerically using a modification of the method of Fletcher & Powell (1963). This is an efficient iterative method that makes use of derivatives of the likelihood function and the inverse of the information matrix.

The logarithm of the likelihood is

$$\log L = -\tfrac{1}{2}pN\log(2\pi) - \tfrac{1}{2}N\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{\alpha=1}^{N}\sum_{i=1}^{p}\sum_{j=1}^{p}(x_{\alpha i}-\mu_{\alpha i})\,\sigma^{ij}(x_{\alpha j}-\mu_{\alpha j}),$$

where $\mu_{\alpha i}$ and σ^{ij} are elements of $E(\mathbf{X}) = \mathbf{A\Xi P}$ and $\boldsymbol{\Sigma}^{-1}$, respectively. Writing

$$\mathbf{T} = \frac{1}{N}(\mathbf{X}-\mathbf{A\Xi P})'(\mathbf{X}-\mathbf{A\Xi P}), \tag{3}$$

we can readily see that maximizing $\log L$ is equivalent to minimizing

$$F = \log|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{T\Sigma}^{-1}). \tag{4}$$

The function F is regarded as a function of $\boldsymbol{\Xi}$, $\mathbf{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ by remembering that $\mathbf{T}$ is a function of $\boldsymbol{\Xi}$ by (3) and $\boldsymbol{\Sigma}$ is a function of $\mathbf{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ by (1). The derivatives of F are $\partial F/\partial\boldsymbol{\Xi} = -(2/N)\,\boldsymbol{\Lambda}'(\mathbf{X}-\mathbf{A\Xi P})\boldsymbol{\Sigma}^{-1}\mathbf{P}'$, $\partial F/\partial\mathbf{B} = 2\boldsymbol{\Omega B\Gamma}$, $\partial F/\partial\boldsymbol{\Lambda} = 2\mathbf{B}'\boldsymbol{\Omega B\Lambda\Phi}$, $\partial F/\partial\boldsymbol{\Phi} = 2\boldsymbol{\Lambda}'\mathbf{B}'\boldsymbol{\Omega B\Lambda} - \mathrm{diag}(\boldsymbol{\Lambda}'\mathbf{B}'\boldsymbol{\Omega B\Lambda})$, $\partial F/\partial\boldsymbol{\Psi} = 2\,\mathrm{diag}(\mathbf{B}'\boldsymbol{\Omega B})\boldsymbol{\Psi}$ and $\partial F/\partial\boldsymbol{\Theta} = 2\,\mathrm{diag}(\boldsymbol{\Omega})\boldsymbol{\Theta}$, where $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}-\mathbf{T})\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}^2$.

The minimization problem is that of minimizing the function $F(\boldsymbol{\Xi}, \mathbf{B}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Theta})$ taking into account that some parameters may be fixed and some may be constrained to be equal to some others. Such a problem may be formalized as follows.

Let $F(\mathbf{z}) = F(z_1, \ldots, z_k)$ be a function with first order derivatives $\partial F/\partial\mathbf{z}$ and approximate second order derivatives $\partial^2 F/(\partial\mathbf{z}\,\partial\mathbf{z}')$. Let some $(k-l)$ of the z's be fixed and denote the remaining l z's by $y_1, \ldots, y_l$ $(l \leqslant k)$. The function F is now considered as a function $G(\mathbf{y})$ of $y_1, \ldots, y_l$. Derivatives $\partial G/\partial\mathbf{y}$ and $\partial^2 G/(\partial\mathbf{y}\,\partial\mathbf{y}')$ are obtained from $\partial F/\partial\mathbf{z}$ and $\partial^2 F/(\partial\mathbf{z}\,\partial\mathbf{z}')$ by omitting rows and columns corresponding to the fixed z's. Among $y_1, \ldots, y_l$, let there be some m distinct arguments denoted $x_1, \ldots, x_m$ $(m \leqslant l)$, so that each y_i is equal to one and only one x_g, but possibly several y's equal the same x. Let

$$m_{ig} = \begin{cases} 1 & (y_i = x_g), \\ 0 & \text{otherwise.} \end{cases} \qquad (i = 1, \ldots, l;\ g = 1, \ldots, m).$$

The function F, or G, is now a function $H(\mathbf{x})$ of $x_1, \ldots, x_m$ and we have

$$\left.\begin{aligned} \partial H/\partial x_g &= \sum_{i=1}^{l}(\partial G/\partial y_i)\,m_{ig}, \\ \partial^2 H/(\partial x_g\,\partial x_h) &= \sum_{i=1}^{l}\sum_{j=1}^{l}\{\partial^2 G/(\partial y_i\,\partial y_j)\}\,m_{ig}\,m_{jh}. \end{aligned}\right\} \tag{5}$$

Thus, derivatives of H are simple sums of derivatives of G.

The minimization of $H(\mathbf{x})$ is now a straightforward application of the Fletcher and Powell method. This method makes use of a matrix $\mathbf{E}$, which is evaluated in each iteration. Initially, $\mathbf{E}$ is any positive definite matrix approximating the inverse of $\partial^2 H/(\partial\mathbf{x}\,\partial\mathbf{x}')$. In subsequent

iterations $\mathbf{E}$ is improved, using the information built up about the function, so that ultimately $\mathbf{E}$ converges to an approximation of the inverse of $\partial^2 H/(\partial \mathbf{x}\,\partial \mathbf{x}')$ at the minimum. If there are many parameters, the number of iterations may be excessive, but can be considerably decreased by the provision of a good starting point and a good initial estimate of $\mathbf{E}$. Such an estimate may be obtained from the information matrix

$$E\{\partial^2 F/(\partial \mathbf{z}\,\partial \mathbf{z}')\} = \tfrac{1}{2}NE\{(\partial F/\partial \mathbf{z})(\partial F/\partial \mathbf{z}')\}$$

and the use of (5). The elements of the information matrix are rather complicated and are not given here.

1·3. *Standard errors of estimates and confidence intervals*

When multiplied by $2/N$ the final matrix $\mathbf{E}$ obtained with the Fletcher and Powell procedure gives an approximation to the variance-covariance matrix of the estimated parameters. Experience with the method, however, has revealed that this approximation is not sufficiently accurate for most purposes. A better estimate of the variance-covariance matrix of the estimated parameters may be obtained by computing the information matrix at the minimum. This could in principle be used to set up large sample confidence regions for the parameters. In most cases, however, only the diagonal elements, corresponding to variances of estimates, would normally be of interest. If $\mathbf{P}$ is square and nonsingular or if $\mathbf{\Sigma}$ is known, or an independent estimate $\mathbf{\Sigma}$ is available, exact small sample confidence intervals can be set up for the parameters in $\mathbf{\Xi}$; see Potthoff & Roy (1964).

1·4. *Tests of hypotheses*

Let H_0 be any specific hypothesis concerning the parametric structure of the general model and let H_1 be an alternative hypothesis. One can then test H_0 against H_1 by means of the likelihood ratio technique. Let F_0 be the minimum of F under H_0 and let F_1 be the minimum of F under H_1. Then $F_1 \leqslant F_0$ and minus two times the logarithm of the likelihood ratio becomes

$$u = N(F_0 - F_1). \tag{6}$$

Under H_0 this is distributed, in large samples, as a χ^2 distribution with degrees of freedom equal to the difference in number of parameters estimated under H_1 and H_0.

In general this requires the computation of the solution both under H_0 and H_1. However, for most of the useful alternatives H_1, the solution is known and the value of F_1 can be computed from some simple sample statistics. Because of its importance in most practical work one such general alternative is considered here.

Suppose that under H_1, $E(\mathbf{X}) = \mathbf{A\Xi P}$, where $\mathbf{P}$ is square and nonsingular, and that $\mathbf{\Xi}$ and $\mathbf{\Sigma}$ are unconstrained. Then, under H_1, the maximum likelihood estimates of $\mathbf{\Xi}$ and $\mathbf{\Sigma}$ are

$$\hat{\mathbf{\Xi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{XP}, \tag{7}$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\mathbf{X}'\{\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\}\mathbf{X}. \tag{8}$$

The test statistic becomes

$$u = N(F_0 - \log|\hat{\mathbf{\Sigma}}| - p), \tag{9}$$

with degrees of freedom

$$d = gp + \tfrac{1}{2}p(p+1) - m,$$

where m is the number of independent parameters estimated under H_0.

2. APPLICATIONS

2·1. *The special case, $g = 1$*

In this part of the paper several models and problems are considered that may conveniently be handled by means of the method described in § 1. To begin with we shall consider models where all observation vectors have the same mean vector. This is specified in (2) by letting $\mathbf{A}$ be a column vector with all ones. The parameter matrix $\boldsymbol{\Xi}$ is then a row vector $\boldsymbol{\xi}'$ and the common mean vector is $\boldsymbol{\xi}'\mathbf{P}$. If $\mathbf{P}$ is square and nonsingular and $\boldsymbol{\xi}$ is unconstrained, all information regarding parameters $\mathbf{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ in $\boldsymbol{\Sigma}$ is provided by the usual sample variance-covariance matrix $\mathbf{S}$. If, on the other hand, the common mean vector is constrained, either by means of $\mathbf{P}$ or by constraints on $\boldsymbol{\xi}$, this information affects the estimation of parameters in $\boldsymbol{\Sigma}$ and an overall estimation has to be done.

2·2. *Analysis of sets of congeneric measurements*

Most measurements employed in the behavioural sciences contain sizeable errors of measurement and any adequate theory or model must take this fact into account. Of particular importance is the study of congeneric measurements, i.e. those measurements that are assumed to measure the same thing.

Classical test theory (Lord & Novick, 1968) assumes that a test score x is the sum of a true score τ and an error score e, where e and τ are uncorrelated. A set of test scores $x_1, \ldots, x_p$ with true scores $\tau_1, \ldots, \tau_p$ is said to be congeneric if every pair of true scores τ_i and τ_j have unit correlation. Such a set of test scores can be represented as

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\beta}\tau + \mathbf{e},$$

where $\mathbf{x}' = (x_1, \ldots, x_p)$, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ is a vector of regression coefficients, $\mathbf{e}' = (e_1, \ldots, e_p)$ is the vector of error scores, $\boldsymbol{\mu}$ is the mean vector of $\mathbf{x}$ and τ is a true score, for convenience scaled to zero mean and unit variance. The elements of $\mathbf{x}$, $\mathbf{e}$ and τ are regarded as random variables for a population of examinees. Let $\theta_1^2, \ldots, \theta_p^2$ be the variances of $e_1, \ldots, e_p$, respectively, i.e. the error variances. The corresponding true score variances are $\beta_1^2, \ldots, \beta_p^2$. One important problem is that of estimating these quantities. The variance-covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Theta}^2, \tag{10}$$

where $\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \ldots, \theta_p)$. This is a special case of (1), obtained by specifying $\mathbf{B} = \boldsymbol{\beta}$, $\boldsymbol{\Lambda} = \boldsymbol{\Phi} = 1$ and $\boldsymbol{\Psi} = 0$.

Parallel tests and tau-equivalent tests, in the sense of Lord & Novick (1968), are special cases of congeneric tests. Parallel tests have equal true score variances and equal error variances, i.e.

$$\beta_1^2 = \ldots = \beta_p^2, \theta_1^2 = \ldots = \theta_p^2.$$

Tau-equivalent tests have equal true score variances but possibly different error variances. These two models are obtained from (1) by specification of equality of the corresponding set of parameters. Equality of means may also be assumed. This is specified, for example, by choosing $\boldsymbol{\Xi} = \mu$ and $\mathbf{P} = (1, \ldots, 1)$.

The previous model may be generalized to several sets of congeneric test scores. If there are q sets of such tests, with $m_1, \ldots, m_q$ tests respectively, we write $\mathbf{x}' = (\mathbf{x}_1', \ldots, \mathbf{x}_q')$, where $\mathbf{x}_g'(g = 1, \ldots, q)$ is the vector of observed scores for the gth set. Associated with the vector

$\mathbf{x}_g$ there is a true score τ_g and vectors $\boldsymbol{\mu}_g$ and $\boldsymbol{\beta}_g$ defined as in the previous paragraph. If the different true scores $\tau_1, \ldots, \tau_q$ are all mutually uncorrelated each set of tests can be analyzed separately. In most cases, however, these true scores are correlated with each other and an overall analysis of the entire set of tests must be made. Let $p = m_1 + \ldots + m_q$, $\boldsymbol{\mu} = E(\mathbf{x})$ and $\mathbf{e}$ be the vector of error scores. Furthermore, let $\boldsymbol{\tau}' = \{\tau_1, \ldots, \tau_q\}$ and let

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_2 & \ldots & \mathbf{0} \\ . & . & \ldots & . \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\beta}_q \end{bmatrix}. \tag{11}$$

Then

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\tau} + \mathbf{e}.$$

Let $\boldsymbol{\Gamma}$ be the variance-covariance (correlation) matrix of $\boldsymbol{\tau}$. Then the variance-covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Gamma}\mathbf{B}' + \boldsymbol{\Theta}^2. \tag{12}$$

The model (12) is obtained from (1) by specifying $\boldsymbol{\Lambda} = \mathbf{I}$, $\boldsymbol{\Phi} = \boldsymbol{\Gamma}$ with unities in the diagonal, and $\boldsymbol{\Psi} = \mathbf{0}$. The matrix $\mathbf{B}$ is specified to have zero values in certain positions as in (11). Any subset of tests may be assumed to be tau-equivalent or parallel by specifying the equality of the corresponding β and θ parameters.

2·3. *Factor analysis*

Factor analysis is a widely used technique especially among psychologists and other behavioural scientists. The basic idea is that for a given set of response variates $x_1, \ldots, x_p$ one wants to find a set of underlying or latent factors $f_1, \ldots, f_k$, fewer in number than the observed variates, that will account for the intercorrelations of the response variates, in the sense that when the factors are partialled out from the observed variates there no longer remain any correlations between these. This leads to the model

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{z}, \tag{13}$$

where $E(\mathbf{x}) = \boldsymbol{\mu}$, $E(\mathbf{f}) = \mathbf{0}$ and $E(\mathbf{z}) = \mathbf{0}$, $\mathbf{z}$ being uncorrelated with $\mathbf{f}$. Let $\boldsymbol{\Phi} = E(\mathbf{f}\mathbf{f}')$ which may be taken as a correlation matrix and $\boldsymbol{\Psi}^2 = E(\mathbf{z}\mathbf{z}')$ which is diagonal. Then the variance-covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}$ becomes

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2. \tag{14}$$

If $(p-k)^2 < p+k$, this relationship can be tested statistically, unlike (13) which involves hypothetical variates and cannot be verified directly. Equation (14) may be obtained from the general model (1) by specifying $\mathbf{B} = \mathbf{I}$ and $\boldsymbol{\Theta} = \mathbf{0}$.

When $k > 1$ there is an indeterminacy in (14) arising from the fact that a nonsingular linear transformation of $\mathbf{f}$ changes $\boldsymbol{\Lambda}$ and in general also $\boldsymbol{\Phi}$ but leaves $\boldsymbol{\Sigma}$ unchanged. The usual way to eliminate this indeterminacy in exploratory factor analysis (see, for example, Lawley & Maxwell, 1963; Jöreskog, 1967; Jöreskog & Lawley, 1968) is to choose $\boldsymbol{\Phi} = \mathbf{I}$ and $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ to be diagonal and to estimate the parameters in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ subject to these conditions. This leads to an arbitrary set of factors which may then be subjected to a rotation or a linear transformation to another set of factors which can be given a more meaningful interpretation.

In terms of the general model (1), the indeterminacy in (14) may be eliminated by assigning zero values, or any other values, to k^2 elements in $\mathbf{\Lambda}$ and/or $\mathbf{\Phi}$, in such a way that these assigned values will be destroyed by all nonsingular transformations of the factors except the identity transformation. There may be an advantage in eliminating the indeterminacy this way, in that, if the fixed parameters are chosen in a reasonable way, the resulting solution will be directly interpretable and the subsequent rotation of factors may be avoided.

Specification of parameters *a priori* may also be used in a confirmatory factor analysis, where the experimenter has already obtained a certain amount of knowledge about the variates measured and is in a position to formulate a hypothesis that specifies the factors on which the variates depend. Such an hypothesis may be specified by assigning values to some parameters in $\mathbf{\Lambda}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$; see, for example, Jöreskog & Lawley (1968) and Jöreskog (1969). If the number of fixed parameters in $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ exceeds k^2, the hypothesis represents a restriction of the common factor space and a solution obtained under such an hypothesis cannot be obtained by a rotation of an arbitrary solution such as obtained in an exploratory analysis. The method of analysis described in § 1 may be used to estimate the free parameters and to test the hypothesis.

Model (10) is formally equivalent to a factor analytic model with one common factor and model (12) is equivalent to a factor analytic model with q correlated nonoverlapping factors. In the latter case the factors are the true scores $\boldsymbol{\tau}' = (\tau_1, \ldots, \tau_q)$ of the tests. These true scores may themselves satisfy a factor analytic model, i.e.

$$\boldsymbol{\tau} = \mathbf{\Lambda}\mathbf{f} + \mathbf{s},$$

where $\mathbf{f}$ is a vector of order k of common true score factors, $\mathbf{s}$ is a vector of order q of specific true score factors and $\mathbf{\Lambda}$ is a matrix of order $q \times k$ of factor loadings. Let $\mathbf{\Phi}$ be the variance-covariance matrix of $\mathbf{f}$ and let $\mathbf{\Psi}^2$ be a diagonal matrix whose diagonal elements are the variances of the specific true score factors $\mathbf{s}$. Then $\mathbf{\Gamma}$, the variance-covariance matrix of $\boldsymbol{\tau}$, becomes

$$\mathbf{\Gamma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}^2. \tag{15}$$

Substituting (15) into (12) gives $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma} = \mathbf{B}(\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}^2)\,\mathbf{B}' + \mathbf{\Theta}^2. \tag{16}$$

Model (16) is a special case of (1) by specifying zero values in $\mathbf{B}$ as before. To define $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ uniquely it is necessary to impose k^2 independent conditions on these to eliminate the indeterminacy due to rotation. Model (16) is a special case of the second order factor analytic model.

2·4. *Estimation of variance components*

We give an example of an experimental design that has one random way of classification and a possibly unbalanced fixed classification. It is possible to deal with all models of this kind considered by Bock & Bargmann (1966) and many other models. Suppose the data x_{vij} are classified according to one random mode $v = 1, \ldots, N$, one fixed mode $i = 1, 2, 3$ and another fixed mode $j = 1, 2, 3$ ($i = 1, 2$) and $j = 1, 2$ ($i = 3$). One model that may be considered is

$$x_{vij} = a_v + b_{vi} + c_{vj} + e_{vij}, \tag{17}$$

where a_v, b_{vi}, c_{vj} and e_{vij} are independent random variates with means α, β_i, γ_j and 0, and

variances $\sigma_a^2, \sigma_{b_i}^2, \sigma_{c_j}^2$ and $\sigma_{e_{ij}}^2$, respectively. Writing $\mathbf{x}_\nu' = (x_{\nu 11}, x_{\nu 12}, x_{\nu 13}, x_{\nu 21}, x_{\nu 22}, x_{\nu 23}, x_{\nu 31}, x_{\nu 32})$, $\mathbf{u}_\nu' = (a_\nu, b_{\nu 1}, b_{\nu 2}, b_{\nu 3}, c_{\nu 1}, c_{\nu 2}, c_{\nu 3})$ and

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}, \tag{18}$$

we may write (17) as

$$\mathbf{x}_\nu = \mathbf{H}\mathbf{u}_\nu + \mathbf{e}_\nu, \tag{19}$$

where $\mathbf{e}_\nu$ is a random error vector of the same form as $\mathbf{x}_\nu$. The mean vector of $\mathbf{x}_\nu$ is $\mathbf{H}E(\mathbf{u}_\nu)$ and the variance-covariance matrix is

$$\mathbf{\Sigma} = \mathbf{H}\mathbf{\Phi}\mathbf{H}' + \mathbf{\Psi}^2, \tag{20}$$

where $\mathbf{\Phi}$ is a diagonal matrix whose diagonal elements are σ_a^2, $\sigma_{b_1}^2$, $\sigma_{b_2}^2$, $\sigma_{b_3}^2$, $\sigma_{c_1}^2$, $\sigma_{c_2}^2$ and $\sigma_{c_3}^2$ and $\mathbf{\Psi}^2$ is a diagonal matrix whose elements are the $\sigma_{e_{ij}}^2$. In terms of the general model (1) and (2), this model may be represented by choosing $\mathbf{\Xi} = E(\mathbf{u}_\nu')$, $\mathbf{P} = \mathbf{H}'$, $\mathbf{B} = \mathbf{I}$, $\mathbf{\Lambda} = \mathbf{H}$ and $\mathbf{\Theta} = \mathbf{0}$. Matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}^2$ are as defined in (20) and the matrix $\mathbf{A}$ in (2) is a column vector with all ones. The general method of analysis yields maximum likelihood estimates of the fixed effects in $E(\mathbf{u}_\nu)$ and of the variance components σ_a^2, $\sigma_{b_i}^2$, $\sigma_{c_j}^2$ and $\sigma_{e_{ij}}^2$.

In the above example the design matrix $\mathbf{H}$ of the experiment was of full column rank. However, in many cases this is not so and each of the components of $\mathbf{u}_\nu$ cannot be separately estimated. In general, if $\mathbf{H}$ is of order $p \times r$ and of rank k, one may choose k independent linear functions, each one linearly dependent on the rows of $\mathbf{H}$ and estimate the mean vector and variance-covariance matrix of these functions; see, for example, Graybill (1961, p. 228). It is customary to choose linear combinations that are mutually uncorrelated but this is not necessary in the analysis by our method. Let $\mathbf{L}$ be the matrix of coefficients of the chosen linear functions and let $\mathbf{K}$ be any matrix such that $\mathbf{H} = \mathbf{K}\mathbf{L}$. For example, $\mathbf{K}$ may be obtained from

$$\mathbf{K} = \mathbf{H}\mathbf{L}'(\mathbf{L}\mathbf{L}')^{-1}. \tag{21}$$

The model may then be reparameterized to full rank by defining $\mathbf{u}^* = \mathbf{L}\mathbf{u}$. We then have $\mathbf{x} = \mathbf{H}\mathbf{u} + \mathbf{e} = \mathbf{K}\mathbf{L}\mathbf{u} + \mathbf{e} = \mathbf{K}\mathbf{u}^* + \mathbf{e}$. The mean vector of $\mathbf{x}$ is $\mathbf{K}E(\mathbf{u}^*)$ and the variance-covariance matrix of $\mathbf{x}$ is represented as

$$\mathbf{\Sigma} = \mathbf{K}\mathbf{\Phi}\mathbf{K}' + \mathbf{\Psi}^2, \tag{22}$$

where $\mathbf{\Phi}$ is the variance-covariance matrix of $\mathbf{u}^*$ and $\mathbf{\Psi}^2$ is as before. The general method of analysis yields estimates of $E(\mathbf{u}^*)$, $\mathbf{\Psi}^2$ and $\mathbf{\Phi}$. The last matrix may be taken to be diagonal, if desired.

2·5. *Analysis of sets of ordered responses*

Suppose observations are available on a set of ordered response variates $x_1, \ldots, x_p$. These may be repeated measurements on the same sampling unit or any other set of ordered but correlated responses. Such data are often employed in growth studies. Perhaps, the most

comprehensive treatment of the estimation, testing and confidence interval problems in this context, is that of Potthoff & Roy (1964). Potthoff and Roy use the general model (2) assuming that $\boldsymbol{\Sigma}$ is essentially known, although their hypothesis testing procedure is valid conditional on any given $\boldsymbol{\Sigma}$. In this section we shall consider models where a structure is imposed on $\boldsymbol{\Sigma}$ and possibly also on the means and where the problem is to estimate parameters of the means as well as parameters of the variances and covariances. The emphasis is on estimation rather than hypothesis testing.

Consider the Wiener stochastic process

$$x_t = \mu_t + u_1 + \ldots + u_t \quad (t = 1, \ldots, p), \tag{23}$$

where $\mu_t = E(x_t)$ and $u_1, \ldots, u_p$ are independent increments. This may be written in matrix form as

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{M}\mathbf{u}, \tag{24}$$

where $\mathbf{M}$ is a lower triangular matrix whose nonzero elements are all unity. The variance-covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \mathbf{M}\boldsymbol{\Phi}\mathbf{M}', \tag{25}$$

where $\boldsymbol{\Phi}$ is the diagonal matrix whose diagonal elements are the variances of the independent increments $u_1, \ldots, u_p$. While the estimation of $\boldsymbol{\Phi}$ is a simple matter in this case, this is no longer true if the data are fallible so that the variates contain errors of measurement. With such errors included, (24) becomes

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{M}\mathbf{u} + \mathbf{e}, \tag{26}$$

where $\mathbf{e}$ is a vector of errors of measurement. The variance-covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \mathbf{M}\boldsymbol{\Phi}\mathbf{M}' + \boldsymbol{\Psi}^2, \tag{27}$$

where $\boldsymbol{\Psi}^2$ is the diagonal matrix of error variances, assuming $e_1, \ldots, e_p$ to be mutually uncorrelated. Mukherjee (1966) developed a method for finding the maximum likelihood estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}^2$ both for homogeneous and heterogeneous error variances. These estimates may also be obtained by our method by specifying $\mathbf{B} = \mathbf{I}$, $\boldsymbol{\Lambda} = \mathbf{M}$ and $\boldsymbol{\Phi}$ to be diagonal, and $\boldsymbol{\Theta} = \mathbf{0}$. If error variances are assumed to be homogeneous one specifies equality of all diagonal elements of $\boldsymbol{\Psi}$ in addition, in order that $\boldsymbol{\Psi} = \psi\mathbf{I}$.

Next consider a Markov model, namely the first order autoregressive series,

$$x_t = \mu_t + \kappa_t(x_{t-1} - \mu_{t-1}) + u_t \quad (t = 2, \ldots, p),$$

or in matrix form, with u_1 defined as $x_1 - \mu_1$,

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{K}\mathbf{u}, \tag{28}$$

with variance-covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{K}\boldsymbol{\Phi}\mathbf{K}', \tag{29}$$

where

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \kappa_2 & 1 & \ldots & 0 \\ \kappa_2\kappa_3 & \kappa_3 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ \prod_{i=2}^{p} \kappa_i & \prod_{i=3}^{p} \kappa_i & \ldots & 1 \end{bmatrix}. \tag{30}$$

If random error terms are added to (28), $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{\Sigma} = \mathbf{K}\boldsymbol{\Phi}\mathbf{K}' + \boldsymbol{\Theta}^2, \tag{31}$$

where $\boldsymbol{\Theta}^2$ is the diagonal matrix of error variances. This is of the same form as (27) but in this case $\mathbf{K}$ is neither a fixed matrix nor a matrix whose nonfixed elements are free parameters. A reparameterization is therefore necessary to fit this model into the form (1). Let $\alpha_i = \kappa_2 \ldots \kappa_i$ $(i = 2, \ldots, p)$ and $\mathbf{D}_\alpha$ be a diagonal matrix with the first diagonal element equal to unity and the ith diagonal element equal to α_i $(i = 2, \ldots, p)$. Then

$$\mathbf{K} = \mathbf{D}_\alpha \mathbf{M} \mathbf{D}_\alpha^{-1},$$

where $\mathbf{M}$, as before, is a lower triangular matrix whose nonzero elements are all unity. Let $\boldsymbol{\Phi}$, as before, be the diagonal dispersion matrix of $\mathbf{u}$ and $\boldsymbol{\Phi}^* = \mathbf{D}_\alpha^{-1} \boldsymbol{\Phi} \mathbf{D}_\alpha^{-1}$. Then

$$\boldsymbol{\Sigma} = \mathbf{D}_\alpha \mathbf{M} \boldsymbol{\Phi}^* \mathbf{M}' \mathbf{D}_\alpha + \boldsymbol{\Theta}^2, \tag{32}$$

which may be represented by (1) by choosing $\mathbf{B}$ as the diagonal matrix $\mathbf{D}_\alpha$, $\boldsymbol{\Lambda}$ as $\mathbf{M}$, $\boldsymbol{\Phi}$ as the diagonal matrix $\boldsymbol{\Phi}^*$, $\boldsymbol{\Psi}$ as $\mathbf{0}$ and $\boldsymbol{\Theta}$ as $\boldsymbol{\Theta}$. It should be noted that there is a one to one transformation between $(\kappa_2, \ldots, \kappa_p, \phi_1, \ldots, \phi_p)$ and $(\alpha_2, \ldots, \alpha_p, \phi_1^*, \ldots, \phi_p^*)$.

One disadvantage with the model (32) is that not all parameters are uniquely determined by $\boldsymbol{\Sigma}$. An investigation of the identification problem reveals that constants may be added to or subtracted from the first and the last diagonal elements in $\boldsymbol{\Theta}$, the effect of which may be counterbalanced by a certain change in $\mathbf{D}_\alpha$ and $\boldsymbol{\Phi}^*$. To eliminate this indeterminacy, two independent conditions have to be imposed but there does not seem to be any natural way to do so. However, if it is reasonable to assume that error variances are equal for all variates, which may be the case when all variates are measured in the same units, the model (32) may be replaced by the more restrictive model

$$\boldsymbol{\Sigma} = \mathbf{D}_\alpha \mathbf{M} \boldsymbol{\Phi}^* \mathbf{M}' \mathbf{D}_\alpha + \theta^2 \mathbf{I}, \tag{33}$$

which does not contain any indeterminacies.

Autoregressive series of higher orders may also be considered, but except in certain special cases the resulting $\boldsymbol{\Sigma}$ is not of the form (1). Models of this kind for the structure of the variance-covariance matrix may be used in connexion with various structures on the mean values. If one variate is measured on p occasions $t_1, \ldots, t_p$ and there are g independent groups of observations with n_s observations in the sth group, $n_1 + \ldots + n_g = N$, one may wish to consider polynomial growth curves like

$$E(x_t) = \mu_t = \xi_{s0} + \xi_{s1} t + \ldots + \xi_{sh} t^h, \tag{34}$$

for the sth group $(s = 1, \ldots, g)$. Such a model may be represented in the form of (2) by letting $\mathbf{A}$, $\boldsymbol{\Xi}$ and $\mathbf{P}$ be as follows. The matrix $\mathbf{A}$ is of order $N \times g$ and has n_1 rows $(1, 0, \ldots, 0)$, n_2 rows $(0, 1, 0, \ldots, 0)$, ... and n_g rows $(0, \ldots, 0, 1)$. Further

$$\boldsymbol{\Xi} = \begin{bmatrix} \xi_{10} & \xi_{11} & \cdots & \xi_{1h} \\ \xi_{20} & \xi_{21} & \cdots & \xi_{2h} \\ \cdot & \cdot & \cdots & \cdot \\ \xi_{g0} & \xi_{g1} & \cdots & \xi_{gh} \end{bmatrix}, \tag{35}$$

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ t_1 & t_2 & \ldots & t_p \\ t_1^2 & t_2^2 & \ldots & t_p^2 \\ \cdot & \cdot & \cdots & \cdot \\ t_1^h & t_2^h & \ldots & t_p^h \end{bmatrix}. \tag{36}$$

If this model is used together with (27), for example, one can estimate the parameters in $\boldsymbol{\Xi}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ simultaneously. If desired, one can also estimate these parameters under the condition that some of the coefficients in the polynomial growth curves have assigned values or that one or more of them are the same for certain groups.

The above models may be extended to the case when several variates are measured at different occasions. To consider the simplest case of only two variates, let

$$E(x_{it}) = \xi_{s0}^{(i)} + \xi_{s1}^{(i)} t + \ldots + \xi_{sh}^{(i)} t^h \quad (s = 1, \ldots, g;\ i = 1, 2). \tag{37}$$

If the variates x_{it} are ordered so that

$$(x_{1t_1}, \ldots, x_{1t_p}, x_{2t_1}, \ldots, x_{2t_p}) \tag{38}$$

corresponds to a row of the data matrix $\mathbf{X}$, then $\mathbf{A}$ is as before,

$$\boldsymbol{\Xi} = \begin{bmatrix} \xi_{10}^{(1)} & \xi_{11}^{(1)} & \cdots & \xi_{1h}^{(1)} & \xi_{10}^{(2)} & \xi_{11}^{(2)} & \cdots & \xi_{1h}^{(2)} \\ \xi_{20}^{(1)} & \xi_{21}^{(1)} & \cdots & \xi_{2h}^{(1)} & \xi_{20}^{(2)} & \xi_{21}^{(2)} & \cdots & \xi_{2h}^{(2)} \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \xi_{g0}^{(1)} & \xi_{g1}^{(1)} & \cdots & \xi_{gh}^{(1)} & \xi_{g0}^{(2)} & \xi_{g1}^{(2)} & \cdots & \xi_{gh}^{(2)} \end{bmatrix}, \tag{39}$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^* \end{bmatrix},$$

where $\mathbf{P}^*$ is the same as $\mathbf{P}$ in (36) and $\mathbf{0}$ is a zero matrix. The variance-covariance matrix $\boldsymbol{\Sigma}$ of the random vector in (38) may be assumed, for example, to be of the form

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_3 \\ \boldsymbol{\Phi}_3 & \boldsymbol{\Phi}_2 \end{bmatrix} \begin{bmatrix} \mathbf{M}' & \mathbf{0} \\ \mathbf{0} & \mathbf{M}' \end{bmatrix} + \boldsymbol{\Psi}^2, \tag{40}$$

where $\boldsymbol{\Phi}_1$, $\boldsymbol{\Phi}_2$ and $\boldsymbol{\Phi}_3$ are diagonal matrices of order p. Such a covariance structure results from

$$z_{it} = z_{i,t-1} + u_{it} \tag{41}$$

if the increments u_{it} are uncorrelated between occasions but correlated within occasions, z_{it} being the deviation of x_{it} from the mean. The diagonal elements of $\boldsymbol{\Phi}_i$ are the variances of $u_{i1}, \ldots, u_{ip}$ $(i = 1, 2)$ and the diagonal elements of $\boldsymbol{\Phi}_3$ are the covariances between u_{1t} and u_{2t} $(t = 1, \ldots, p)$. In this case one can estimate the elements of $\boldsymbol{\Xi}$, $\boldsymbol{\Phi}_1$, $\boldsymbol{\Phi}_2$, $\boldsymbol{\Phi}_3$ and $\boldsymbol{\Psi}$ simultaneously. If desired, one can test the hypothesis that growth curves are the same for several groups and for the two variates measured.

2·6. *Path analysis and linear structural relationships*

Path analysis, due to Wright (1918), is a technique sometimes used to assess the direct causal contribution of one variable to another in a nonexperimental situation. The problem, in general, is that of estimating the parameters of a set of linear structural equations, representing the cause and effect relationships hypothesized by the investigator. Such a system of equations involves variates of two kinds: independent or cause variables $\xi_1, \ldots, \xi_m$ and dependent or effect variables $\eta_1, \ldots, \eta_n$. The technique consists of solving the structural equations for the dependent variables in terms of the independent to obtain the reduced form of the equations and then estimating the regression of the dependent variables on the independent from this reduced form. Models of this kind and a variety of estimation techniques have been extensively studied by econometricians, for example, Wold & Jureen (1953) and

Wold (1964); biometricians, for example, Turner & Stevens (1959) and references therein; and sociologists, for example, Blalock (1964). There seem to be three types of difficulties associated with these techniques, namely (i) if there are errors of measurement in the independent variates, these errors will give rise to inconsistent regression estimates; (ii) if the parameters in the structural equations and the parameters in the reduced form are not in a one-to-one correspondence, one or more of the parameters of structural equations may be overidentified or underidentified; and (iii) since the regression technique is applied to each equation separately, one does not get an overall test of the entire causal structure. The purpose of this section is to show how some of these models may be analyzed by means of a covariance-structure approach and how this to a certain extent eliminates the above difficulties. We shall first give an example and then consider a more general case.

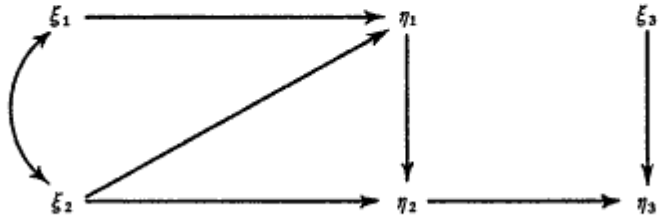

Fig. 1. Path diagram for fictitious problem.

Consider the path diagram of Fig. 1, involving true variates ξ_1, ξ_2, ξ_3, η_1, η_2 and η_3. The structural equations are

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2, \quad \eta_2 = \alpha_{21}\eta_1 + \gamma_{22}\xi_2, \quad \eta_3 = \alpha_{32}\eta_2 + \gamma_{33}\xi_3. \tag{42}$$

There are six parameters γ_{11}, γ_{12}, α_{21}, γ_{22}, α_{32} and γ_{33} in these equations to be estimated. The reduced form equations are

$$\left.\begin{aligned} \eta_1 &= \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + 0\xi_3, \\ \eta_2 &= \alpha_{21}\gamma_{11}\xi_1 + (\alpha_{21}\gamma_{12} + \gamma_{22})\xi_2 + 0\xi_3, \\ \eta_3 &= \alpha_{32}\alpha_{21}\gamma_{11}\xi_1 + \alpha_{32}(\alpha_{21}\gamma_{12} + \gamma_{22})\xi_2 + \gamma_{33}\xi_3. \end{aligned}\right\} \tag{43}$$

There are nine regression coefficients in these equations, two of which are constrained to be zero, the remaining seven being functions of the six parameters of the structural equations. It is readily seen that α_{32} is overdetermined. If we estimate the regression coefficients in the reduced form equations by the regression method, assuming ξ_1, ξ_2 and ξ_3 without error, there is no guarantee that the estimates $\hat{\beta}_{31}/\hat{\beta}_{21}$ and $\hat{\beta}_{32}/\hat{\beta}_{22}$ of α_{32} are the same.

Defining $\mu = \alpha_{21}$, $\nu = \alpha_{32}\alpha_{21}$, $\delta = \gamma_{11}$, $\epsilon = \gamma_{12}$, $\phi = \gamma_{12} + (\gamma_{22}/\alpha_{21})$ and $\kappa = \gamma_{33}/\alpha_{32}\alpha_{21}$, we may easily verify that there is a one-to-one transformation between $(\alpha_{21}, \alpha_{32}, \gamma_{11}, \gamma_{12}, \gamma_{22}, \gamma_{33})$ and $(\mu, \nu, \delta, \epsilon, \phi, \kappa)$ and that $\alpha_{21} = \mu$, $\alpha_{32} = \nu/\mu$, $\gamma_{11} = \delta$, $\gamma_{12} = \epsilon$, $\gamma_{22} = \mu(\phi - \epsilon)$ and $\gamma_{33} = \nu\kappa$. Let

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{\Lambda} = \begin{bmatrix} \delta & \epsilon & 0 \\ \delta & \phi & 0 \\ \delta & \phi & \kappa \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{44}$$

Now suppose further that η_1, η_2, η_3, ξ_1, ξ_2 and ξ_3 cannot be observed directly, but that instead

$$y_i = \nu_i + \eta_i + \epsilon_i, \quad x_i = \mu_i + \xi_i + \delta_i \quad (i = 1, 2, 3) \tag{45}$$

are observed, where ν_i and μ_i are the means of y_i and x_i respectively and ϵ_i and δ_i are errors of measurement. Assuming that errors of measurement are uncorrelated with the true variates and among themselves, we find from (44) and (45) that the variance-covariance matrix of y_1, y_2, y_3, x_1, x_2 and x_3 is

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\mathbf{B}' + \boldsymbol{\Theta}^2, \tag{46}$$

where $\boldsymbol{\Phi}$ is the variance-covariance matrix of ξ_1, ξ_2 and ξ_3, and $\boldsymbol{\Theta}^2$ is a diagonal matrix containing the six error variances. This is of the form (1) with $\boldsymbol{\Psi} = \mathbf{0}$ and with certain elements of $\boldsymbol{\Lambda}$ constrained to be equal as in (44). The sample variance-covariance matrix $\mathbf{S}$ of y_1, y_2, y_3, x_1, x_2 and x_3 provides information for estimating the free parameters in $\mathbf{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$. In large samples the χ^2 statistic (6) may be used to test the overall goodness of fit of the causal structure. In this case there are 6 parameters in $\mathbf{B}$ and $\boldsymbol{\Lambda}$, 6 in $\boldsymbol{\Phi}$ and 6 in $\boldsymbol{\Theta}$ to be estimated, so that the degrees of freedom are $21 - 18 = 3$.

Now consider a general case. Let $\boldsymbol{\xi}$ be a random $m \times 1$ vector of true independent variables, $\boldsymbol{\eta}$ be a random $n \times 1$ vector of dependent variables and

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\delta} \end{bmatrix}, \tag{47}$$

where, as before, $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are mean vectors and $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are vectors of errors of measurement being uncorrelated among themselves and with the true variates. Let the structural equations be written as

$$\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi}. \tag{48}$$

The reduced form becomes, by assuming $\mathbf{I} - \mathbf{A}$ to be nonsingular,

$$\boldsymbol{\eta} = (\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi}, \tag{49}$$

which may also be written as

$$\begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} = \begin{bmatrix} (\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\Gamma} \\ \mathbf{I} \end{bmatrix} \boldsymbol{\xi}, \tag{50}$$

where the matrix on the right-hand side is of order $(n + m) \times m$. In general, there will be some fixed, usually zero, elements in $\mathbf{A}$ and $\boldsymbol{\Gamma}$, the remaining elements being parameters to be estimated. Suppose there exist two matrices $\mathbf{B}$ and $\boldsymbol{\Lambda}$ containing fixed, constrained and free elements to conform with the specification for model (1), such that

$$\mathbf{B}\boldsymbol{\Lambda} = \begin{bmatrix} (\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\Gamma} \\ \mathbf{I} \end{bmatrix}, \tag{51}$$

and such that there is a one-to-one transformation between the parameters in $\mathbf{A}$ and $\boldsymbol{\Gamma}$ and the distinct free elements in $\mathbf{B}$ and $\boldsymbol{\Lambda}$. Then (50) can be written

$$\begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} = \mathbf{B}\boldsymbol{\Lambda}\boldsymbol{\xi}. \tag{52}$$

Substituting (52) into (47), we find that the observed variates are given by

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\mu} \end{bmatrix} + \mathbf{B}\boldsymbol{\Lambda}\boldsymbol{\xi} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\delta} \end{bmatrix}, \tag{53}$$

with variance-covariance matrix of the form (46), where $\boldsymbol{\Phi}$ is the variance-covariance

matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}^2$ is the diagonal matrix of error variances of $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$. The free parameters in $\mathbf{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ may be estimated by the method described in § 1·1, provided (53) represents a nontrivial hypothesis, i.e. provided the degrees of freedom are positive. The number of degrees of freedom is equal to $\frac{1}{2}(m+n)(m+n+1)$, the number of variances and covariances in $\boldsymbol{\Sigma}$, minus $\frac{1}{2}m(m+1)$, the number of variances and covariances in $\boldsymbol{\Phi}$, minus $m+n$, the number of error variances in $\boldsymbol{\Theta}^2$, minus the number of free distinct parameters in $\mathbf{B}$ and $\boldsymbol{\Lambda}$, or $\mathbf{A}$ and $\boldsymbol{\Gamma}$.

The above approach is limited to such cases where (51), with its specifications, holds. In those cases this approach clearly eliminates the three types of difficulties listed in the beginning of this section.

This work was supported by the National Science Foundation. A computer program implementing the method described in this paper has been written by G. Gruvaeus and M. van Thillo for the IBM 360/65. Copies of the programe are available from the author.

References

Blalock, H. M. (1964). *Causal Inferences in Nonexperimental Research*. University of North Carolina Press.

Bock, R. D. & Bargmann, R. E. (1966). Analysis of covariance structures. *Psychometrika* **31**, 507–34.

Fletcher, R. & Powell, M. J. D. (1963). A rapidly convergent descent method for minimization. *Comput. J.* **6**, 163–8.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, vol. 1. New York: McGraw-Hill.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32**, 443–82.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34**, 183–202.

Jöreskog, K. G. & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. *Brit. J. Math. Statist. Psychology* **21**, 85–96.

Lawley, D. N. & Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method*. London: Butterworth.

Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores* (with contributions by A. Birnbaum). Reading, Mass.: Addison-Wesley.

Mukherjee, B. N. (1966). Derivation of likelihood-ratio tests for Guttman quasi-simplex covariance structures. *Psychometrika* **31**, 97–123.

Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–16.

Turner, M. E. & Stevens, C. D. (1959). The regression analysis of causal paths. *Biometrics*, **15**, 236–58.

Wold, H. (Editor). (1964). *Econometric Model Building. Essays on the Causal Chain Approach*. Amsterdam: North-Holland.

Wold, H. & Jureen, L. (1953). *Demand Analysis*. New York: Wiley.

Wright, S. (1918). On the nature of size factors. *Genetics* **3**, 367–74.

[*Received November* 1969. *Revised January* 1970]