DOCUMENT RESUME

ED 051 257 TM 000 291

AUTHOR	Joreskog, Karl G.
TITLE	A General Method for Estimating a Linear Structural Equation System.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-70-54
PUB DATE	Sep 70
NOTE	43p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Algorithms, Analysis of Covariance, Analysis of Variance, *Computer Programs, Critical Path Method, Economics, Goodness of Fit, *Mathematical Models, Mathematics, Psychology, Sciences, *Standard Error of Measurement, *Statistics, Tests
IDENTIFIERS	*Measurement Errors

ABSTRACT

A general method for estimating the unknown coefficients in a set of linear structural equations is described. In its most general form the method allows for both errors in equations (residuals, disturbances) and errors in variables (errors of measurement, observational errors) and yields estimates of the residual variance-covariance matrix and the measurement error variances as well as estimates of the unknown coefficients in the structural equations, provided all these parameters are identified. Two special cases of this general method are discussed separately. One is when there are errors in equations but no errors in variables. The other is when there are errors in variables but no errors in equations. The methods are applied and illustrated using artificial, economic and psychological data. (Author)

RESEARCH BULLETIN

RB-70-54

A GENERAL METHOD FOR ESTIMATING A LINEAR STRUCTURAL EQUATION SYSTEM

Karl G. Jöreskog

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
September 1970

A General Method for Estimating a Linear Structural Equation System

Karl G. Jöreskog

Educational Testing Service

Abstract

A general method for estimating the unknown coefficients in a set of linear structural equations is described. In its most general form the method allows for both errors in equations (residuals, disturbances) and errors in variables (errors of measurement, observational errors) and yields estimates of the residual variance-covariance matrix and the measurement error variances as well as estimates of the unknown coefficients in the structural equations, provided all these parameters are identified. Two special cases of this general method are discussed separately. One is when there are errors in equations but no errors in variables. The other is when there are errors in variables but no errors in equations. The methods are applied and illustrated using artificial, economic and psychological data.

A General Method for Estimating a Linear Structural Equation System*

1. Introduction

We shall describe a general method for estimating the unknown coefficients in a set of linear structural equations. In its most general form the method will allow for both errors in equations (residuals, disturbances) and errors in variables (errors of measurement, observational errors) and will yield estimates of the residual variance-covariance matrix and the measurement error variances as well as estimates of the unknown coefficients in the structural equations, provided all these parameters are identified. After giving the results for this general case, two special cases will be considered. The first is the case when there are errors in equations but no errors in variables. This case has been studied extensively by econometricians (see e.g., Goldberger, 1964, Chapter 7). The second case is when there are errors in variables but no errors in equations. Models of this kind have been studied under the name of path analysis by biometricians (see e.g., Turner & Stevens, 1959). sociologists (see e.g., Blalock, 1964) and psychologists (Werts & Linn, 1970).

It is assumed that the observed variables have a multinormal distribution and the unknown parameters are estimated by the maximum likelihood method. The estimates are computed numerically using a modification of the Fletcher-Powell minimization algorithm (Fletcher & Powell, 1963; Gruvaeus & Jöreskog, 1970). Standard errors of the estimated parameters may be obtained by computing the inverse of the information matrix. A computer program,

*This research has been supported in part by grant NSF-GB-12959 from the National Science Foundation. The author wishes to thank Professor Arthur Goldberger for his comments on an earlier draft of the paper and Marielle van Thillo, who wrote the computer programs, checked the mathematical derivations and gave other valuable assistance throughout the work.

LISREL, in FORTRAN IV, that performs all the necessary computations has been written and tested out on the IBM 360/65; a write-up of this is under preparation (Jöreskog & van Thillo, 1970).

In the first special case referred to above, where there are no errors of measurement in the observed variables the general method to be presented is equivalent to the full information maximum likelihood (FIML) method of Koopmans, Rubin and Leipnik (1950) also called full information least generalized residual variance (FILGRV) method (Goldberger, 1964, Chapter 7), provided that no constraints are imposed on the residual variance-covariance matrix and the variance-covariance matrix of the independent variables. However, with the general method described here, it is possible to assign fixed values to some elements of these matrices and also to have equality constraints among the remaining elements.

2. The General Model

Consider random vectors $\eta' = (\eta_1, \eta_2, \ldots, \eta_m)$ and $\xi' = (\xi_1, \xi_2, \ldots, \xi_n)$ of true dependent and independent variables, respectively, and the following system of linear structural relations

$$B\eta = \Gamma\xi + \zeta \tag{1}$$

where $B(m \times m)$ and $\Gamma(m \times n)$ are coefficient matrices and $\zeta' = (\zeta_1, \zeta_2, \ldots, \zeta_m)$ is a random vector of residuals (errors in equations, random disturbance terms). Without loss of generality it may be assumed that $\mathcal{E}(\eta) = \mathcal{E}(\zeta) = 0$ and $\mathcal{E}(\xi) = 0$. It is furthermore assumed that ζ is uncorrelated with ξ and that B is nonsingular.

The vectors η and ξ are not observed but instead vectors $y' = (y_1, y_2, \ldots, y_m)$ and $x' = (x_1, x_2, \ldots, x_n)$ are observed, such that

$$y = \mu + \eta + \epsilon \tag{2}$$

$$x = \nu + \xi + \delta \tag{3}$$

where $\mu = \mathcal{E}(y)$, $\nu = \mathcal{E}(x)$ and ϵ and δ are vectors of errors of measurement in y and x , respectively. It is convenient to refer to y and x as the observed variables and η and ξ as the true variables. The errors of measurement are assumed to be uncorrelated with the true variates and among themselves.

Let $\Phi(n \times n)$ and $\Psi(m \times m)$ be the variance-covariance matrices of ξ and ζ , respectively, Θ_ϵ^2 and Θ_δ^2 the diagonal matrices of error variances for y and x , respectively. Then it follows, from the above assumptions, that the variance-covariance matrix $\Sigma[(m + n) \times (m + n)]$ of $z = (y', x')'$ is

$$\Sigma = \begin{pmatrix} B^{-1}\Gamma\Phi\Gamma'B'^{-1} + B^{-1}\Psi B'^{-1} + \Theta_\epsilon^2 & B^{-1}\Gamma\Phi \\ \Phi\Gamma'B'^{-1} & \Phi + \Theta_\delta^2 \end{pmatrix} . \tag{4}$$

The elements of Σ are functions of the elements of B , Γ , Φ , Ψ , Θ_δ and Θ_ϵ . In applications some of these elements are fixed and equal to assigned values. In particular this is so for elements in B and Γ , but we shall allow for fixed values even in the other matrices. For the remaining nonfixed elements of the six parameter matrices one or more subsets may have identical but unknown values. Thus parameters in B , Γ , Φ , Ψ , Θ_δ and Θ_ϵ 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.

Before an attempt is made to estimate a model of this kind, the identification problem must be examined. The identification problem depends on the specification of fixed, constrained and free parameters. Under a given specification, a given structure $\mathbf{B}$, Γ, Φ, Ψ, Θ_δ, Θ_ϵ generates one and only one Σ but there may be several structures generating the same Σ. If two or more structures generate the same Σ, the structures are said to be equivalent. If a parameter has the same value in all equivalent structures, the parameter is said to be identified. If all parameters of the model are identified, the whole model is said to be identified. When a model is identified one can usually find consistent estimates of all its parameters. Some rules for investigating the identification problem when there are no errors in variables are given by Goldberger (1964, pp. 306-318).

3. Estimation of the General Model

Let $z_1, z_2, \cdots, z_N$ be N observations of $z = (y', x')'$. Since no constraints are imposed on the mean vector $(\mu', \nu')'$ the maximum likelihood estimate of this is the usual sample mean vector $\bar{z} = (\bar{y}', \bar{x}')'$. Let

$$S = \frac{1}{N} \sum_{\alpha=1}^{N} (z_\alpha - \bar{z})(z_\alpha - \bar{z})' \tag{5}$$

be the usual sample variance-covariance matrix, partitioned as

$$S[(m+n) \times (m+n)] = \begin{bmatrix} S_{yy}(m \times m) & S_{yx}(m \times n) \\ S_{xy}(n \times m) & S_{xx}(n \times n) \end{bmatrix} . \tag{6}$$

The logarithm of the likelihood function, omitting a function of the observations, is given by

$$\log L = -\frac{1}{2} N[\log |\Sigma| + \mathrm{tr}(S\Sigma^{-1})] \quad . \tag{7}$$

This is regarded as a function of the independent distinct parameters in B , Γ , Φ , Ψ , Θ_δ and Θ_ϵ and is to be maximized with respect to these, taking into account that some elements may be fixed and some may be constrained to be equal to some others. Maximizing $\log L$ is equivalent to minimizing

$$F = (N/2)[\log |\Sigma| + \mathrm{tr}(S\Sigma^{-1})] \quad . \tag{8}$$

Such a minimization problem may be formalized as follows.

Let $\lambda' = (\lambda_1, \lambda_2, \cdots, \lambda_p)$ be a vector of all the elements of B , Γ , Φ , Ψ , Θ_δ and Θ_ϵ arranged in a prescribed order. Then F may be regarded as a function $F(\lambda)$ of $\lambda_1, \lambda_2, \cdots, \lambda_p$, which is continuous and has continuous derivatives $\partial F/\partial\lambda_s$ and $\partial^2 F/\partial\lambda_s\partial\lambda_t$ of first and second order, except where Σ is singular. The totality of these derivatives is represented by a gradient vector $\partial F/\partial\lambda$ and a symmetric matrix $\partial^2 F/\partial\lambda\partial\lambda'$. Now let some $p - q$ of the λ's be fixed and denote the remaining λ's by π_1, $\pi_2, \cdots, \pi_q$, $q \leq p$. The function F is now considered as a function $G(\pi)$ of $\pi_1, \pi_2, \cdots, \pi_q$. Derivatives $\partial G/\partial\pi$ and $\partial^2 G/\partial\pi\partial\pi'$ are obtained from $\partial F/\partial\lambda$ and $\partial^2 F/\partial\lambda\partial\lambda'$ by omitting rows and columns corresponding to the fixed λ's. Among $\pi_1, \pi_2, \cdots, \pi_q$, let there be some r distinct parameters denoted $\kappa_1, \kappa_2, \cdots, \kappa_r$, $r \leq q$, so that each π_i is equal to one and only one κ_j , but possibly several π's equal the same κ. Let $\underset{\sim}{K} = (k_{ij})$ be a matrix of order $q \times r$ with elements $k_{ij} = 1$ if $\pi_i = \kappa_j$ and $k_{ij} = 0$ otherwise. The function F (or G) is now a function $H(\underset{\sim}{\kappa})$ of $\kappa_1, \kappa_2, \ldots, \kappa_r$ and we have

$$\partial H/\partial \underset{\sim}{\kappa} = \underset{\sim}{K}'(\partial G/\partial \underset{\sim}{\pi}) \tag{9}$$

$$\partial^2 H/\partial \underset{\sim}{\kappa}\partial \underset{\sim}{\kappa}' = \underset{\sim}{K}'(\partial G/\partial \underset{\sim}{\pi}\partial \underset{\sim}{\pi}')\underset{\sim}{K} \quad . \tag{10}$$

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

The minimization of $H(\underset{\sim}{\kappa})$ is now a straightforward application of the Fletcher-Powell method for which a computer program is available (Gruvaeus & Jöreskog, 1970). This method makes use of a matrix $\underset{\sim}{E}$, which is evaluated in each iteration. Initially $\underset{\sim}{E}$ is any positive definite matrix approximating the inverse of $\partial^2 H/\partial \underset{\sim}{\kappa}\partial \underset{\sim}{\kappa}'$. In subsequent iterations $\underset{\sim}{E}$ is improved, using the information built up about the function so that ultimately $\underset{\sim}{E}$ converges to an approximation of the inverse of $\partial^2 H/\partial \underset{\sim}{\kappa}\partial \underset{\sim}{\kappa}'$ 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 initial estimate of $\underset{\sim}{E}$. Such an estimate may be obtained by inverting the information matrix

$$\mathcal{E}(\partial^2 H/\partial \underset{\sim}{\kappa}\partial \underset{\sim}{\kappa}') = \underset{\sim}{K}'\mathcal{E}(\partial^2 G/\partial \underset{\sim}{\pi}\partial \underset{\sim}{\pi}')\underset{\sim}{K} \quad , \tag{11}$$

where $\mathcal{E}(\partial^2 G/\partial \underset{\sim}{\pi}\partial \underset{\sim}{\pi}')$ is obtained from

$$\mathcal{E}(\partial^2 F/\partial \underset{\sim}{\lambda}\partial \underset{\sim}{\lambda}') \approx \mathcal{E}(\partial F/\partial \underset{\sim}{\lambda}\, \partial F/\partial \underset{\sim}{\lambda}') \tag{12}$$

as described above. When the minimum of H has been found, the inverse of the information matrix may be computed again to obtain standard errors of all the parameters in $\underset{\sim}{\kappa}$. A general method for obtaining the elements of $\mathcal{E}(\partial F/\partial \underset{\sim}{\lambda} \partial F/\partial \underset{\sim}{\lambda}')$ is given in Appendix A2.

The application of the Fletcher-Powell method requires formulas for the derivatives of F with respect to the elements of $\underset{\sim}{B}$, $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Phi}$, $\underset{\sim}{\Psi}$, $\underset{\sim}{\Theta}_\delta$ and $\underset{\sim}{\Theta}_\epsilon$. These may be obtained by matrix differentiation as shown in Appendix A1. Writing $\underset{\sim}{A} = \underset{\sim}{B}^{-1}$, $\underset{\sim}{D} = \underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}$ and

$$\underset{\sim}{\Omega} = \begin{pmatrix} \underset{\sim}{\Omega}_{yy} & \underset{\sim}{\Omega}_{yx} \\ \underset{\sim}{\Omega}_{xy} & \underset{\sim}{\Omega}_{xx} \end{pmatrix} = \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\Sigma} - \underset{\sim}{S})\underset{\sim}{\Sigma}^{-1} , \qquad (13)$$

the derivatives are

$$\partial F/\partial \underset{\sim}{B} = -N(\underset{\sim}{A}'\underset{\sim}{\Omega}_{yy}\underset{\sim}{D}\underset{\sim}{\Phi}\underset{\sim}{D}' + \underset{\sim}{A}'\underset{\sim}{\Omega}_{yy}\underset{\sim}{A}\underset{\sim}{\Psi}\underset{\sim}{A}' + \underset{\sim}{A}'\underset{\sim}{\Omega}_{yx}\underset{\sim}{\Phi}\underset{\sim}{D}') \qquad (14)$$

$$\partial F/\partial \underset{\sim}{\Gamma} = N(\underset{\sim}{A}'\underset{\sim}{\Omega}_{yy}\underset{\sim}{D}\underset{\sim}{\Phi} + \underset{\sim}{A}'\underset{\sim}{\Omega}_{yx}\underset{\sim}{\Phi}) \qquad (15)$$

$$\partial F/\partial \underset{\sim}{\Phi} = N(\underset{\sim}{D}'\underset{\sim}{\Omega}_{yy}\underset{\sim}{D} + \underset{\sim}{D}'\underset{\sim}{\Omega}_{yx} + \underset{\sim}{\Omega}_{xy}\underset{\sim}{D} + \underset{\sim}{\Omega}_{xx}) \qquad (16)$$

$$\partial F/\partial \underset{\sim}{\Psi} = N\underset{\sim}{A}'\underset{\sim}{\Omega}_{yy}\underset{\sim}{A} \qquad (17)$$

$$\partial F/\partial \underset{\sim}{\Theta}_\delta = N\underset{\sim}{\Omega}_{xx}\underset{\sim}{\Theta}_\delta \qquad (18)$$

$$\partial F/\partial \underset{\sim}{\Theta}_\epsilon = N\underset{\sim}{\Omega}_{yy}\underset{\sim}{\Theta}_\epsilon \qquad (19)$$

In these expressions we have not taken into account that $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are symmetric and that $\underset{\sim}{\Theta}_\delta$ and $\underset{\sim}{\Theta}_\epsilon$ are diagonal matrices. The off-diagonal zero elements of Θ_δ and Θ_ϵ are treated as fixed parameters and the off-diagonal elements of Φ and Ψ as constrained parameters.

When the maximum likelihood estimates of the parameters have been obtained, the goodness of fit of the model may be tested, in large samples, by the likelihood ratio technique. Let H_0 be the null hypothesis of the model under the given specifications of fixed, constrained and free parameters. The alternative hypothesis H_1 may be that Σ is any positive definite matrix.

Under H_1, the maximum of $\log L$ is (see e.g., Anderson, 1958, Chapter 3),

$$\log L_1 = -\frac{1}{2} N(\log |S| + m + n) \quad .$$

Under H_0, the maximum of $\log L$ is equal to minus the minimum value F_0 of F. Thus minus 2 times the logarithm of the likelihood ratio becomes

$$U = 2F_0 - N \log|S| - N(m + n) \quad . \tag{20}$$

If the model holds, U is distributed, in large samples, as χ^2 with

$$d = \frac{1}{2}(m + n)(m + n + 1) - r \tag{21}$$

degrees of freedom, where, as before, r is the total number of independent parameters estimated under H_0.

4. The Special Case of No Errors of Measurement

If there are no errors of measurement in $\underset{\sim}{y}$ and $\underset{\sim}{x}$, the model (1) may be written

$$\underset{\sim}{B}\underset{\sim}{y} = \Gamma\underset{\sim}{x} + \underset{\sim}{u} \tag{22}$$

where we have written $\underset{\sim}{u}$ instead of $\underset{\sim}{\zeta}$. In (22) we have altered the model slightly, compared to (1), (2) and (3), in that the mean vectors have been eliminated. This is no limitation, however, since constant terms in the equations can be handled by using an x-variable that has the value 1 for every observation. In this case, of course, $\underset{\sim}{S}$ should be the raw moment matrix instead of the dispersion matrix.

This type of model has been studied for many years by econometricians under the names of causal chains and interdependent systems (e.g., Wold & Jureen, 1953). The variables $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are economic variables and in the econometric terminology, the variables are classified as exogenous and endogenous variables, the idea beign that the exogenous variables are given from the outside and the endogenous variables are accounted for by the model. From a statistical point of view the distinction is rather between the independent or predetermined variables $\underset{\sim}{x}$ and the dependent variables $\underset{\sim}{y}$. The residual $\underset{\sim}{u}$ represents a random disturbance term assumed to be uncorrelated with the predetermined variables. Observations $\underset{\sim}{y}_\alpha$ and $\underset{\sim}{x}_\alpha$ on $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are usually in the form of a time series.

Equation (22) is usually referred to as the structural form of the model. When (22) is premultiplied by $\underset{\sim}{B}^{-1}$ one obtains the reduced form

$$y = \Pi x + u^* \quad , \tag{23}$$

where $\Pi = B^{-1}\Gamma$ and $u^* = B^{-1}u$. u^* is the vector of residuals in the reduced form.

In this case, Θ_δ and Θ_ϵ in (4) are zero and therefore $|\Sigma|$ and Σ^{-1} in (7) can be written explicitly. It is readily verified that

$$|\Sigma| = |B|^{-2}|\Phi||\Psi|$$

and

$$\Sigma^{-1} = \begin{pmatrix} B'\Psi^{-1}B & -B'\Psi^{-1}\Gamma \\ -\Gamma'\Psi^{-1}B & \Gamma'\Psi^{-1}\Gamma + \Phi^{-1} \end{pmatrix} \quad .$$

Using these results, $\log L$ becomes

$$\log L = -\tfrac{1}{2} N[\log |\Phi| + \operatorname{tr}(S_{xx}\Phi^{-1})] - \tfrac{1}{2} N\{\log |\Psi| - \log |B|^2$$
$$+ \operatorname{tr}[(BS_{yy}B' - BS_{yx}\Gamma' - \Gamma S_{xy}B' + \Gamma S_{xx}\Gamma')\Psi^{-1}]\} \quad .$$

If Φ is unconstrained, maximizing $\log L$ with respect to Φ gives $\hat{\Phi} = S_{xx}$, which is to be expected, since Φ in this case is the variance-covariance matrix of x. After the likelihood has been maximized with respect to Φ, the reduced likelihood is equal to a constant plus

$$\log L^* = -\tfrac{1}{2} N\{\log |\Psi| - \log |B|^2$$
$$+ \operatorname{tr}[(BS_{yy}B' - BS_{yx}\Gamma' - \Gamma S_{xy}B' + \Gamma S_{xx}\Gamma')\Psi^{-1}]\} \quad . \tag{24}$$

If also ψ is unconstrained, further simplification can be obtained, for then (24) is maximized with respect to ψ, for given B and Γ, when ψ is equal to

$$\psi = BS_{yy}B' - BS_{yx}\Gamma' - \Gamma S_{xy}B' + \Gamma S_{xx}\Gamma' \quad , \tag{25}$$

so that the function to be maximized with respect to B and Γ becomes a constant plus

$$\begin{aligned} \log L^{**} &= -\tfrac{1}{2} N[\log |\psi| - \log |B|^2] \\ &= -\tfrac{1}{2} N \log(|\psi|/|B|^2) \\ &= -\tfrac{1}{2} N \log |B^{-1}\psi B'^{-1}| \\ &= -\tfrac{1}{2} N \log |\psi^*| \quad , \end{aligned} \tag{26}$$

where

$$\psi^* = S_{yy} - S_{yx}\Pi' - \Pi S_{xy} + \Pi S_{xx}\Pi' \quad . \tag{27}$$

In deriving (26), we started from the likelihood function (7) based on the assumption of multinormality of y and x. Such an assumption may be very unrealistic in most economic applications. Koopmans, Rubin and Leipnik (1950) derived (24) and (26) from the assumption of multinormal residuals u, which is probably a better assumption. However, the criterion (26) has intuitive appeal regardless of distributional assumptions and connections with the maximum likelihood method. The matrix ψ in (25) is the variance-covariance matrix of the residuals u in the structural form (22) and the matrix ψ^* in (27) is the variance-covariance matrix of the residuals u^* in the reduced form (23). Maximizing (26) is equivalent to minimizing $|\psi^*|$. Since $|\psi^*|$ is a generalized variance, this method has been called the full information least generalized residual variance (FILGRV) method (see, e.g., Goldberger, 1964, Chapter 7). Several other estimation criteria based on ψ^* have been proposed. Brown (1960) suggested the minimization of $\mathrm{tr}(\psi^*)$ and Zellner (1962) proposed the minimization of $\mathrm{tr}(W^{-1}\psi^*)$ where W is proportional to $S_{yy\cdot x} = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$. Malinvaud (1966, Chapter 9) considered the family of estimation criteria $\mathrm{tr}(A\psi^*)$ with arbitrary positive definite weighting matrices A .

Since the original article by Koopmans, Rubin and Leipnik (1950) several authors have contributed to the development of the FILGRV method (Chernoff & Divinsky, 1953; Klein, 1953, 1969; Brown, 1959; Eisenpress, 1962; Eisenpress & Greenstadt, 1964; Chow, 1968; Wegge, 1969). This paper will add another computational algorithm to those already existing.

Minimizing $|\psi^*|$ is equivalent to minimizing

$$F = \log |\psi| - \log |B|^2 \quad . \tag{28}$$

Matrix derivatives of F with respect to B and Γ may be obtained by matrix differentiation as shown in Appendix A3. The results are

$$\partial F/\partial B = 2\psi^{-1}(BS_{yy} - \Gamma S_{xy}) - B'^{-1} \tag{29}$$

$$\partial F/\partial \Gamma = 2\psi^{-1}(\Gamma S_{xx} - BS_{yx}) \quad . \tag{30}$$

The function F is to be minimized with respect to the elements of B and Γ taking into account that some elements are fixed and others are constrained in some way. As will be demonstrated in sections 5 and 6, allowing for equalities among the elements of $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$, is not sufficient to handle some economic applications. Instead, more general constraints may be involved. Usually these constraints are linear but even models with nonlinear constraints have been studied (see, e.g., Klein, 1969). Such constraints can be handled as follows.

Let $\underset{\sim}{\pi}' = (\pi_1, \pi_2, \ldots, \pi_q)$ be the vector of all nonfixed elements in $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$. Each of these elements may be a known linear or nonlinear function of $\underset{\sim}{\kappa}' = (\kappa_1, \kappa_2, \ldots, \kappa_r)$, the parameters to be estimated, i.e.,

$$\pi_i = f_i(\underset{\sim}{\kappa}) \quad , \qquad i = 1,2,\ldots,q \; . \tag{31}$$

Then F is regarded as a function $H(\kappa)$ of $\kappa_1, \kappa_2, \ldots, \kappa_r$. The derivatives of H of first and second order are again given by (9) and (10), but now K is the matrix of order $q \times r$ whose ijth element is $\partial f_i / \partial \kappa_j$. The function $H(\underset{\sim}{\kappa})$ may be minimized by the Fletcher-Powell method as before. The advantage of this method compared to the more general one of the preceding section is that the function now contains many fewer parameters and the minimization is therefore faster. The Fletcher-Powell algorithm is relatively easy to apply even in the nonlinear case and the iterations converge quadratically from an arbitrary starting point to a minimum of the function, although there is no guarantee that this is the absolute minimum if several local minima exist.

5. Analysis of Artificial Data

The following hypothetical economic model is taken from Brown (1959),

$$C = a_0 + a_1 W + a_2 \Pi + u_1 \tag{32a}$$

$$W = b_0 + b_1 Y + b_2 Y_{-1} + u_2 \tag{32b}$$

$$W + \Pi + T_g = Y \tag{32c}$$

$$C + E = Y \tag{32d}$$

where the dependent variables are

C = consumer expenditures

W = wage-salary bill

Π = nonwage income

Y = total income, production and expenditure

and the predetermined variables are

T_g = government net revenue

E = all nonconsumer spending on newly produced final goods

Y_{-1} = value of Y lagged one time period

and where u_1 and u_2 are random disturbance terms assumed to be uncorrelated with the predetermined variables. This hypothetical model will be used to illustrate some of the ideas and methods of the previous sections.

To begin with we shall assume that the variables involved in this model are not directly observed. Instead they are assumed to represent true variables that can only be measured with errors. Such an assumption may not be unreasonable, as pointed out by Johnston (1963):

> To be realistic we must recognize that most economic statistics contain errors of measurement, so that they are only approximations to the underlying "true" values. Such errors may arise because totals are estimated on a sample basis or, even if a complete enumeration is attempted, errors and inaccuracies may creep in. Often, too, the published statistics may represent an attempt to measure concepts which are different from those postulated in the theory (p. 148).

Converting the variables to deviations from mean values and writing $\underset{\sim}{\eta}' = (C,W,\Pi,Y)$, $\underset{\sim}{\xi}' = (T_g,E,Y_{-1})$ and $\underset{\sim}{\zeta}' = (u_1,u_2,0,0)$, model (32) may be written in the form of (1) as

$$\begin{pmatrix} 1 & -a_1 & -a_2 & 0 \\ 0 & 1 & 0 & -b_1 \\ 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \underset{\sim}{\eta} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & b_2 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \underset{\sim}{\xi} + \underset{\sim}{\zeta} \quad . \tag{33}$$

There are 19 independent parameters in this model, namely 4 in $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$, 6 in

$$\underset{\sim}{\Phi} = \begin{pmatrix} \sigma^2_{T_g} & & \\ \sigma_{T_gE} & \sigma^2_E & \\ \sigma_{T_gY_{-1}} & \sigma_{EY_{-1}} & \sigma^2_{Y_{-1}} \end{pmatrix} \quad , \tag{34}$$

3 in

$$\underset{\sim}{\Psi} = \begin{bmatrix} \sigma^2_{u_1} & & & \\ \sigma_{u_1u_2} & \sigma^2_{u_2} & & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad , \tag{35}$$

and 6 in $\underset{\sim}{\Theta}_\delta = \mathrm{diag}(\theta_{T_g},\theta_E,\theta_{Y_{-1}})$ and $\underset{\sim}{\Theta}_\epsilon = \mathrm{diag}(\theta_C,\theta_W,\theta_\Pi,\theta_Y)$. Note that since (32c) and (32d) are error-free equations, $\underset{\sim}{\Psi}$ has the form (35) with zero variances and covariances for u_3 and u_4 . Also since Y_{-1} is Y lagged, we have assumed that the error variances in Y and Y_{-1} are the same. Therefore, $\underset{\sim}{\Theta}_\delta$ and $\underset{\sim}{\Theta}_\epsilon$ have only 6 independent elements.

Data were generated from this model by assigning the following values to each of the 19 parameters

$$\left.\begin{array}{l}
a_1 = 0.8 \quad a_2 = 0.4 \quad b_1 = 0.3 \quad b_2 = 0.2 \\
\sigma^2_{T_g} = 1.0 \quad \sigma^2_E = 2.0 \quad \sigma^2_{Y_{-1}} = 3.0 \\
\sigma_{T_gE} = 0.1 \quad \sigma_{T_gY_{-1}} = 0.2 \quad \sigma_{EY_{-1}} = 0.1 \\
\sigma^2_{u_1} = 0.2 \quad \sigma^2_{u_2} = 0.3 \quad \sigma_{u_1u_2} = 0.1 \\
\theta_{T_g} = 0.4 \quad \theta_E = 0.6 \quad \theta_{Y_{-1}} = 0.5 \\
\theta_C = 0.5 \quad \theta_W = 0.6 \quad \theta_\Pi = 0.9 \quad \theta_Y = 0.5
\end{array}\right\} \tag{36}$$

The resulting $\underset{\sim}{\Sigma}$, obtained from (4) and rounded to 3 decimals, is

	C	W	Π	Y	T_g	E	Y_{-1}
C	4.599						
W	2.481	2.069					
Π	4.659	2.159	7.514				
Y	6.449	3.731	7.409	10.799			
T_g	-0.692	-0.138	-1.454	-0.592	1.160		
E	2.100	1.250	2.750	4.100	0.100	2.360	
Y_{-1}	0.442	0.763	-0.421	0.542	0.200	0.100	3.250

(37)

For the purpose of illustrating the estimation method of section 3, the above matrix is regarded as a sample dispersion matrix $\underset{\sim}{S}$ to be analyzed. The order of the vector $\underset{\sim}{\lambda}$ is 78, since there are 78 elements in $\underset{\sim}{B}$, $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Phi}$, $\underset{\sim}{\Psi}$, $\underset{\sim}{\Theta}_\delta$ and $\underset{\sim}{\Theta}_\epsilon$ all together. Of these, 54 are fixed and 24 are nonfixed, so that $\underset{\sim}{\pi}$ is of order 24. Because of the symmetry of $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ and the imposed equality of θ_Y and $\theta_{Y_{-1}}$, there are 19 independent parameters, so that the order of $\underset{\sim}{\kappa}$ is 19.

The minimization of $H(\underset{\sim}{\kappa})$ started at the point

$$a_1 = 0.6 \; , \quad a_2 = 0.3 \; , \quad b_1 = 0.4 \; , \quad b_2 = 0.1$$

$$\sigma^2_{T_g} = 2.0 \; , \quad \sigma^2_E = 2.0 \; , \quad \sigma^2_{Y_{-1}} = 2.0$$

$$\sigma_{T_gE} = \sigma_{T_gY_{-1}} = \sigma_{EY_{-1}} = 0.0$$

$$\sigma^2_{u_1} = 0.3 \; , \quad \sigma^2_{u_2} = 0.3 \; , \quad \sigma_{u_1u_2} = 0.0$$

$$\theta_{T_g} = 0.4 \; , \quad \theta_E = 0.6 \; , \quad \theta_{Y_{-1}} = 0.5$$

$$\theta_C = 0.5 \; , \quad \theta_W = 0.6 \; , \quad \theta_\Pi = 0.9 \; , \quad \theta_Y = 0.5$$

From this point seven steepest descent iterations were performed. Thereafter Fletcher-Powell iterations were used and it took 23 such iterations to reach a point where all derivatives were less than 0.00005 in absolute value. At this point, the solution was correct to four decimals and the $\underset{\sim}{\Sigma}$ in (37) was reproduced exactly. Twenty-three Fletcher-Powell iterations required for convergence is not considered excessive since no information about second-order derivatives was used and it takes at least 19 Fletcher-Powell iterations to build up an estimate of the matrix of second order derivatives.

We now consider model (32a-d) in the case when the variables are observed without errors of measurement. Then the method of section 3 cannot be applied directly since the two identities (32c) and (32d) imply that $\underset{\sim}{\Sigma}$ is singular. Therefore, two of the endogenous variables must be eliminated from the system. It seems most convenient to eliminate C and Y. When these variables have been eliminated, the structural equations become

$$\begin{pmatrix} 1 - a_1 & 1 - a_2 \\ 1 - b_1 & - b_1 \end{pmatrix} \begin{pmatrix} W \\ \Pi \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ b_1 & 0 & b_2 \end{pmatrix} \begin{pmatrix} T_g \\ E \\ Y_{-1} \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad . \tag{38}$$

This system may be estimated by the method of section 4.

To illustrate the application of the estimation procedure we use a dispersion matrix $\underset{\sim}{S}$ obtained from $\underset{\sim}{\Sigma}$ in (37) by subtracting the error variances from the diagonal elements and deleting rows and columns corresponding to C and Y. There are 6 nonfixed elements in $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$, namely β_{11}, β_{12}, β_{21}, β_{22}, γ_{21} and γ_{23}. These are the elements of the vector $\underset{\sim}{\pi}$. These elements are functions of a_1, a_2, b_1 and b_2 defined by [compare equation (31)]

$$\begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ \gamma_{21} \\ \gamma_{23} \end{pmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad . \tag{39}$$

Thus the function F is a function of 4 independent parameters.

The function F was minimized using only Fletcher-Powell iterations starting from the point

$$a_1 = 0.6 \quad a_2 = 0.3 \quad b_1 = 0.4 \quad b_2 = 0.1 \quad .$$

The solution point, found after 8 iterations, was, as expected, $a_1 = 0.8$, $a_2 = 0.4$, $b_1 = 0.3$, $b_2 = 0.2$ with

$$\hat{\underset{\sim}{\Psi}}^* = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{pmatrix} \quad , \quad |\hat{\underset{\sim}{\Psi}}^*| = 0.05 \quad .$$

6. An Economic Application

In this section we apply methods SFLGRV and RFLGRV to a small economic model taken from the literature. The model is Klein's model of United States economy presented in Klein (1950, pp. 58-66):

Consumption: $C = a_0 + a_1P + a_2P_{-1} + a_3W + u_1$ (40a)

Investment: $I = b_0 + b_1P + b_2P_{-1} + b_3K_{-1} + u_2$ (40b)

Private wages: $W^* = c_0 + c_1E + c_2E_{-1} + c_3A + u_3$ (40c)

Product: $Y + T = C + I + G$ (40d)

Income: $Y = P + W$ (40e)

Capital: $K = K_{-1} + I$ (40f)

Wages: $W = W^* + W^{**}$ (40g)

Private product: $E = Y + T - W^{**}$, (40h)

where the endogenous variables are

C = consumption

I = investment

W^* = private wage bill

P = profits

Y = national income

K = end-of-year capital stock

W = total wage bill

E = private product

and the predetermined variables are the lagged endogenous variables P_{-1}, K_{-1} and E_{-1} and the exogenous variables

1 = unity

W^{**} = government wage bill

T = indirect taxes

G = government expenditures

A = time in years from 1931.

All variables except 1 and A are in billions of 1934 dollars.

This model contains eight dependent variables and eight predetermined variables. There are three equations involving residual terms. The other five equations are identities. Using the five identities (40d) - (40h), P, Y, K, W and E may be solved for and substituted into (40a) - (40c). This gives a model with the following structural form

$$\begin{pmatrix} 1 - a_1 & -a_1 & a_1 - a_3 \\ -b_1 & 1 - b_1 & b_1 \\ -c_1 & -c_1 & 1 \end{pmatrix} \begin{pmatrix} C \\ I \\ W^{*} \end{pmatrix}$$

$$= \begin{pmatrix} a_0 & a_3 - a_1 & -a_1 & a_1 & C & a_2 & 0 & 0 \\ b_0 & -b_1 & -b_1 & b_1 & 0 & b_2 & b_3 & 0 \\ c_0 & -c_1 & 0 & c_1 & c_3 & 0 & 0 & c_2 \end{pmatrix} \begin{pmatrix} 1 \\ W^{**} \\ T \\ G \\ A \\ P_{-1} \\ K_{-1} \\ E_{-1} \end{pmatrix} . \qquad (41)$$

There are 24 nonfixed elements in $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$. These are all linear functions of the 12 unknown coefficients in (40a-c) as follows

$$
\begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \beta_{31} \\ \beta_{32} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{14} \\ \gamma_{16} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \\ \gamma_{24} \\ \gamma_{26} \\ \gamma_{27} \\ \gamma_{31} \\ \gamma_{32} \\ \gamma_{34} \\ \gamma_{35} \\ \gamma_{38} \end{bmatrix}
=
\begin{bmatrix}
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}
+
\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
. \qquad (42)
$$

From annual observations, United States, 1921-1941 the following raw moment matrices are obtained:

$$
S_{yy} = \begin{array}{c} \\ C \\ I \\ W^* \end{array}
\begin{array}{c} \begin{array}{ccc} C & I & W^* \end{array} \\ \begin{pmatrix} 62166.63 & & \\ 1679.01 & 286.02 & \\ 42076.78 & 1217.92 & 28560.86 \end{pmatrix} \end{array} ,
$$

$$
\underset{\sim}{S}_{xy} = \begin{array}{c} \\ 1 \\ W^{**} \\ T \\ G \\ A \\ P_{-1} \\ K_{-1} \\ E_{-1} \end{array}
\begin{array}{c} \begin{array}{ccc} C & I & W^{*} \end{array} \\
\begin{pmatrix}
1133.90 & 26.60 & 763.60 \\
5977.33 & 103.80 & 4044.07 \\
7858.86 & 160.40 & 5315.62 \\
11633.68 & 243.19 & 7922.46 \\
577.70 & -105.60 & 460.90 \\
18929.37 & 655.33 & 12871.73 \\
227767.38 & 5073.25 & 153470.56 \\
66815.25 & 1831.13 & 45288.51
\end{pmatrix} \end{array} ,
$$

$$
\underset{\sim}{S}_{xx} = \begin{array}{c} \\ 1 \\ W^{**} \\ T \\ G \\ A \\ P_{-1} \\ K_{-1} \\ E_{-1} \end{array}
\begin{array}{c} \begin{array}{cccccccc} 1 & W^{**} & T & G & A & P_{-1} & K_{-1} & E_{-1} \end{array} \\
\begin{pmatrix}
21.00 & & & & & & & \\
107.50 & 626.87 & & & & & & \\
142.90 & 789.27 & 1054.95 & & & & & \\
208.20 & 1200.19 & 1546.11 & 2369.94 & & & & \\
0.00 & 238.00 & 176.00 & 421.70 & 770.00 & & & \\
343.90 & 1746.22 & 2348.48 & 3451.86 & -11.90 & 5956.29 & & \\
4210.40 & 21683.18 & 28766.23 & 42026.14 & 590.60 & 69073.54 & 846132.70 & \\
1217.70 & 6364.43 & 8436.53 & 12473.50 & 495.60 & 20542.22 & 244984.77 & 72200.03
\end{pmatrix} \end{array} .
$$

The following estimated model was obtained

$$
\left.\begin{aligned}
C &= 18.318 - 0.229P + 0.384P_{-1} + 0.802W + u_1 \\
I &= 27.278 - 0.797P + 1.051P_{-1} - 0.148K_{-1} + u_2 \\
W^{*} &= 5.766 + 0.235E + 0.284E_{-1} + 0.234A + u_3
\end{aligned}\right\} \qquad (43)
$$

with

$$
\hat{\underset{\sim}{\Sigma}}{}^{*} = \begin{pmatrix} 43.775 & & \\ 80.456 & 265.856 & \\ 9.834 & 80.247 & 37.540 \end{pmatrix} . \qquad (44)
$$

The standard errors of the estimated parameters may be obtained from a formula for the asymptotic variance-covariance matrix developed by Rothenberg and Leenders (1964).

7. The Special Case of No Residuals

When there are no residuals in (1), the relations between η and ξ are exact. The joint distribution of η and ξ is singular and of rank n. In the equation (4) for Σ, the second term in Σ_{yy} vanishes. In general, when there are fixed and constrained elements in B and Γ or in Φ, Θ_δ and Θ_ϵ, this model has to be estimated by the method of section 3. This may be done by choosing $\Psi = 0$ and specifying the fixed elements and the constraints as described in that section.

The matrix Σ can also be written

$$\Sigma = \Lambda\Phi\Lambda' + \Theta^2 , \tag{45}$$

where

$$\Lambda = \begin{pmatrix} B^{-1}\Gamma \\ I \end{pmatrix} \quad \text{and} \quad \Theta = \begin{pmatrix} \Theta_\epsilon & 0 \\ 0 & \Theta_\delta \end{pmatrix} , \tag{46}$$

from which it is seen that the model is identical to a certain restricted factor analysis model. Several special cases will now be considered.

If $B = I$ and Γ is unconstrained, i.e., all elements of Γ are regarded as free parameters, model (45) is formally equivalent to an unrestricted factor model (Jöreskog, 1969). The matrix Λ in (46) may be obtained from any Λ^* of order $(m + n) \times n$ satisfying

$$\Sigma = \Lambda^* \Lambda^{*\prime} + \Theta^2 \tag{47}$$

by a transformation of Λ^* to a reference variables solution where the x's are used as reference variables. Maximum likelihood estimates of Λ^* and Θ may be obtained by the method of Jöreskog (1967a,b) which also yields a large sample χ^2 test of goodness of fit. Let the estimate of Λ^* be partitioned as

$$\hat{\Lambda}^* = \begin{bmatrix} \hat{\Lambda}_1^* \\ \hat{\Lambda}_2^* \end{bmatrix} , \tag{48}$$

where $\hat{\Lambda}_1^*$ is of order $m \times n$ and $\hat{\Lambda}_2^*$ of order $n \times n$. Then the maximum likelihood estimates of Γ and Φ are

$$\hat{\Gamma} = \hat{\Lambda}_1^* \hat{\Lambda}_2^{*-1} , \tag{49}$$

$$\hat{\Phi} = \hat{\Lambda}_2^* \hat{\Lambda}_2^{*\prime} . \tag{50}$$

If $B = I$ and Γ is constrained to have some fixed elements while the remaining elements in Γ are free parameters, model (45) is formally equivalent to a restricted factor model in the sense of Jöreskog (1969). This model may be estimated by the procedure described in the same paper and, in large samples, standard errors of the estimates and a goodness of fit test can also be obtained. A computer program for this procedure is available (Jöreskog & Gruvaeus, 1967).

A more general case is when B is lower triangular. The structural equation system for the true variates is then a causal chain. In general such a causal chain may be estimated by the method described in section 3 of the paper, though there may be simpler methods. One example occurs when the system is normalized by fixing one element in each row of $\underset{\sim}{\Gamma}$ to unity and $\underset{\sim}{B}$ has the form

$$\underset{\sim}{B} = \begin{bmatrix} \beta_{11} & 0 & & 0 \\ \beta_{21} & \beta_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \beta_{m1} & \beta_{m2} & & \beta_{mm} \end{bmatrix}$$

where all the β's are free parameters. Then there is a one-to-one transformation between the free parameters of $\underset{\sim}{B}$ and the free elements of $\underset{\sim}{A} = \underset{\sim}{B}^{-1}$. One may therefore estimate $\underset{\sim}{\Lambda}$ instead of $\underset{\sim}{B}$. In this case, the variance-covariance matrix $\underset{\sim}{\Sigma}$ is of the form

$$\underset{\sim}{\Sigma} = \underset{\sim}{B}^{*}\underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{B}^{*\prime} + \underset{\sim}{\Theta}^{2} \tag{51}$$

where

$$\underset{\sim}{B}^{*} = \begin{pmatrix} \underset{\sim}{A} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{I} \end{pmatrix} \quad , \quad \underset{\sim}{\Lambda} = \begin{pmatrix} \underset{\sim}{\Gamma} \\ \underset{\sim}{I} \end{pmatrix} \quad , \quad \underset{\sim}{\Theta} = \begin{pmatrix} \underset{\sim}{\Theta}_{\epsilon} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Theta}_{\delta} \end{pmatrix} \quad . \tag{52}$$

Model (51) is a special case of a general model for covariance structures developed by Jöreskog (1970) and may be estimated using the computer program ACOVS (Jöreskog, Gruvaeus & van Thillo, 1970). In this model $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Phi}$, $\underset{\sim}{\Theta}_{\delta}$ and $\underset{\sim}{\Theta}_{\epsilon}$ may contain fixed parameters and even parameters constrained to be equal in groups. The computer program gives maximum likelihood estimates of the free parameters in $\underset{\sim}{A}$, $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Phi}$, $\underset{\sim}{\Theta}_{\delta}$ and $\underset{\sim}{\Theta}_{\epsilon}$ and, in large samples, standard errors of these estimates and a test of overall goodness of fit of the model can also be obtained.

More generally, the above mentioned method may be used whenever $\underset{\sim}{\Sigma}$ can be written in the form (51) such that there is a one-to-one correspondence between the free parameters in $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$ and the distinct free elements in $\underset{\sim}{B}^*$ and $\underset{\sim}{\Lambda}$. For a less trivial example, see Jöreskog (1970, section 2.6).

8. A Psychological Application

In this section we consider a simplified model for the prediction of achievements in mathematics (M) and science (S) at different grade levels. To estimate the model we make use of longitudinal data from a growth study conducted at Educational Testing Service (Anderson & Maier, 1963; Hilton, 1969). In this study a nationwide sample of fifth graders was tested in 1961 and then again in 1963, 1965 and 1967 as seventh, ninth and eleventh graders, respectively. The test scores employed in this model are the verbal (V) and quantitative (Q) parts of SCAT (Scholastic Aptitude Test) obtained in 1961 and the achievement tests in mathematics (M_5, M_7, M_9, M_{11}) and science (S_5, S_7, S_9, S_{11}) obtained in 1961, 1963, 1965, and 1967, respectively. The achievement tests have been scaled so that the unit of measurement is approximately the same at all grade levels.

The model is depicted in Figure 1, where V, Q, M_5, M_7, M_9, M_{11}, S_5, S_7, S_9 and S_{11} denote the true scores of the tests and $\zeta_1, \zeta_2, \ldots, \zeta_8$ the corresponding residuals. The model for the true scores is

$$M_5 = a_1 V + a_2 Q + \zeta_1 \tag{53a}$$

$$S_5 = b_1 V + b_2 Q + \zeta_2 \tag{53b}$$

$$M_7 = c_1 M_5 + \zeta_3 \tag{53c}$$

$$S_7 = d_1 S_5 + d_2 M_7 + \zeta_4 \tag{53d}$$

$$M_9 = e_1 M_7 + \zeta_5 \tag{53e}$$

$$S_9 = f_1 S_7 + f_2 M_9 + \zeta_6 \tag{53f}$$

$$M_{11} = g_1 M_9 + \zeta_7 \tag{53g}$$

$$S_{11} = h_1 S_9 + h_2 M_{11} + \zeta_8 \tag{53h}$$

This model postulates the major influences of a student's achievement in mathematics and science at various grade levels. At grade 5 the main determinants of a student's achievements are his verbal and quantitative abilities at that stage. At higher grade levels, however, the achievements are mainly determined by his achievements in the earlier grades. Thus, achievements in mathematics in grade i is determined mainly by the achievements in mathematics in grade $i - 2$, whereas achievements in science in grade i is determined mainly by the achievements in science in grade $i - 2$ and in mathematics in grade i, $i = 7,9,11$.

The structural form of this model is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -c_1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -d_1 & -d_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -e_1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -f_1 & -f_2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -g_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -h_1 & -h_2 & 1 \end{pmatrix} \begin{pmatrix} M_5 \\ S_5 \\ M_7 \\ S_7 \\ M_9 \\ S_9 \\ M_{11} \\ S_{11} \end{pmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} V \\ Q \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \\ \zeta_8 \end{pmatrix} . \tag{54}$$

It is seen that this model is a causal chain. The model can be estimated by the method described in section 3, provided some assumption is made about the intercorrelations of residuals $\zeta_1, \zeta_2, \cdots, \zeta_8$. Without such an assumption the model is not identified. We have chosen to make the assumption that all residuals are uncorrelated except ζ_1 and ζ_2. This assumption does not seem to be too unrealistic.

The data that we use consist of a random sample of 730 boys taken from all the boys that took all tests at all occasions. The variance-covariance matrices are

$$
\underset{\sim}{S}_{yy} = \begin{array}{c} \\ M_5 \\ S_5 \\ M_7 \\ S_7 \\ M_9 \\ S_9 \\ M_{11} \\ S_{11} \end{array}
\begin{array}{c}
\begin{array}{cccccccc} M_5 & S_5 & M_7 & S_7 & M_9 & S_9 & M_{11} & S_{11} \end{array} \\
\begin{pmatrix}
130.690 & & & & & & & \\
115.645 & 179.617 & & & & & & \\
116.162 & 123.838 & 193.537 & & & & & \\
90.709 & 114.364 & 120.426 & 148.[illegible]48 & & & & \\
119.564 & 125.22[illegible] & 155.883 & 120.492 & 215.894 & & & \\
104.430 & 135.074 & 137.827 & 133.231 & 159.783 & 218.067 & & \\
119.712 & 126.470 & 149.930 & 112.218 & 175.497 & 149.045 & 264.071 & \\
90.916 & 116.950 & 117.439 & 109.187 & 133.839 & 147.115 & 143.218 & 190.765
\end{pmatrix}
\end{array} ,
$$

$$
\underset{\sim}{S}_{xy} = \begin{array}{c} \\ V \\ Q \end{array}
\begin{array}{c}
\begin{array}{cccccccc} M_5 & S_5 & M_7 & S_7 & M_9 & S_9 & M_{11} & S_{11} \end{array} \\
\begin{pmatrix}
97.544 & 122.919 & 106.837 & 96.252 & 108.748 & 107.750 & 107.042 & 94.613 \\
78.527 & 82.389 & 87.859 & 65.703 & 91.502 & 72.534 & 89.617 & 64.453
\end{pmatrix}
\end{array} ,
$$

$$
\underset{\sim}{S}_{xx} = \begin{array}{c} \\ V \\ Q \end{array}
\begin{array}{c}
\begin{array}{cc} V & Q \end{array} \\
\begin{pmatrix}
138.014 & \\
73.518 & 80.751
\end{pmatrix}
\end{array} .
$$

The estimated model is

$$M_5 = 0.640V + 0.415Q + \hat{\zeta}_1 \qquad (55a)$$

$$S_5 = 1.296V - 0.175Q + \hat{\zeta}_2 \qquad (55b)$$

$$M_7 = 1.097M_5 + \hat{\zeta}_3 \qquad (55c)$$

$$S_7 = 0.325S_5 + 0.493M_7 + \hat{\zeta}_4 \tag{55d}$$

$$M_9 = 1.027M_7 + \hat{\zeta}_5 \tag{55e}$$

$$S_9 = 0.703S_7 + 0.383M_9 + \hat{\zeta}_6 \tag{55f}$$

$$M_{11} = 0.951M_9 + \hat{\zeta}_7 \tag{55g}$$

$$S_{11} = 0.658S_9 + 0.184M_{11} + \hat{\zeta}_8 \tag{55h}$$

The estimated variance-covariance matrix of the true scores V and Q is

$$\hat{\underset{\sim}{\Phi}} = \begin{array}{c c} & \begin{array}{cc} V & Q \end{array} \\ \begin{array}{c} V \\ Q \end{array} & \begin{pmatrix} 105.48 & \\ 73.95 & 76.68 \end{pmatrix} \end{array} .$$

Estimated residual variances and error variances for each measure are given below

Measure	Residual Variance	Error Variance
V	--	33.1
Q	--	4.4
M_5	10.0	25.4
S_5	22.5	11.8
M_7	26.4	40.3
S_7	29.5	24.3
M_9	25.2	29.3
S_9	28.5	36.1
M_{11}	75.7	18.8
S_{11}	20.0	47.7

The estimated correlation between ζ_1 and ζ_2 is 0.17.

The estimated reduced form for the true scores is

$$M_5 = 0.640V + 0.415Q + \hat{\zeta}_1^* \tag{56a}$$

$$S_5 = 1.296V - 0.175Q + \hat{\zeta}_2^* \tag{56b}$$

$$M_7 = 0.702V + 0.455Q + \hat{\zeta}_3^* \tag{56c}$$

$$S_7 = 0.767V + 0.167Q + \hat{\zeta}_4^* \tag{56d}$$

$$M_9 = 0.721V + 0.467Q + \hat{\zeta}_5^* \tag{56e}$$

$$S_9 = 0.815V + 0.296Q + \hat{\zeta}_6^* \tag{56f}$$

$$M_{11} = 0.686V + 0.444Q + \hat{\zeta}_7^* \tag{56g}$$

$$S_{11} = 0.663V + 0.277Q + \hat{\zeta}_8^* \tag{56h}$$

The relative variance contributions of V and Q, the residual ζ^* and the error, to each test's total variance are shown below:

Measure	V and Q	Residual	Error
M_5	0.73	0.03	0.19
S_5	0.78	0.15	0.07
M_7	0.59	0.20	0.21
S_7	0.56	0.28	0.16
M_9	0.56	0.30	0.14
S_9	0.52	0.32	0.16
M_{11}	0.42	0.51	0.07
S_{11}	0.42	0.33	0.25

It is not easy to give a clear-cut interpretation of these results. Inspecting first the equations (55c), (55e) and (55g), it is seen that a unit increase in M_{i-2} tends to have a smaller effect on M_i the larger i is. This agrees with the fact that the growth curves in mathematics "flattens" out at the higher grade levels. One would expect that the coefficient g_1 in (55g), like c_1 in (55c) and e_1 in (55e), would be greater than one, since, in general, for these data, the correlation of status, M_{i-2}, and gain, $M_i - M_{i-2}$, are positive although usually very small. However, the large residual variance $\hat{\xi}_7$ suggests that M_9 alone is not sufficient to account for M_{11}. This is probably due to the fact that mathematics courses at the higher grades change character from being mainly "arithemetic computation" to involving more "algebraic reasoning."

Inspecting next the equations (55d), (55f) and (55h) describing science achievements, it is seen that the influence of mathematics on science tends to decrease at the higher grades. This is natural since science courses in the lower grades are based mainly on "logical reasoning" whereas in the higher grades they are based on "memorizing of facts." The effect of science achievements on science two years later first increases and then decreases. This is probably because the science courses specialize into different courses (Biology, Physics, etc.) at grade 11 whereas the science test at the lower grades measures some kind of overall "science knowledge."

Whatever may be the best interpretations of these results, the example serves to illustrate that it is possible to have both errors in equations and errors in variables and still have an estimable model.

References

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Anderson, S. B., & Maier, M. H. 34,000 pupils and how they grew. Journal of Teacher Education, 1963, 14, 212-216.

Blalock, H. M. Causal inferences in nonexperimental research. Chapel Hill, N. C.: University of North Carolina Press, 1964.

Brown, T. M. Simplified full maximum likelihood and comparative structural estimates. Econometrica, 1959, 27, 638-653.

Brown, T. M. Simultaneous least squares: a distribution free method of equation system structure estimation. International Economic Rev 1960, 1, 173-191.

Chernoff, H., & Divinsky, N. The computation of maximum-likelihood esti of linear structural equations. In W. C. Hood & T. C. Koopmans (Studies in econometric method, Cowles Commission Monograph 14. New Wiley, 1953. Pp. 236-269.

Chow, G. C. Two methods of computing full-information maximum likeli estimates in simultaneous stochastic equations. International Ec Review, 1968, 9, 100-112.

Eisenpress, H. Note on the computation of full-information maximum-lik estimates of coefficients of a simultaneous system. Econometrica, 30, 343-348.

Eisenpress, H., & Greenstadt, J. The estimation of non-linear econom systems. Econometrica, 1966, 34, 851-861.

Fletcher, R., & Powell, M. J. D. A rapidly convergent descent method for minimization. The Computer Journal, 1963, 6, 163-168.

Goldberger, A. S. Econometric theory. New York: Wiley, 1964.

Gruvaeus, G., & Jöreskog, K. G. A computer program for minimizing a function of several variables. Research Bulletin 70-14. Princeton, N. J.: Educational Testing Service, 1970.

Hilton, T. L. Growth study annotated bibliography. Progress Report 69-11. Princeton, N. J.: Educational Testing Service, 1969.

Johnston, J. Econometric methods. New York: McGraw-Hill, 1963.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1967, 32, 443-482. (a)

Jöreskog, K. G. UMLFA--A computer program for unrestricted maximum likelihood factor analysis. Research Memorandum 66-20. Princeton, N. J.: Educational Testing Service, revised edition, 1967. (b)

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Jöreskog, K. G., & Gruvaeus, G. RMLFA--A computer program for restricted maximum likelihood factor analysis. Research Memorandum 67-21. Princeton: N. J.: Educational Testing Service. 1967.

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

Jöreskog, K. G., & van Thillo, M. LISREL--A general computer program for estimating linear structural relationships. Research Bulletin 70-00. Princeton, N. J.: Educational Testing Service, in preparation.

Klein, L. R. Economic fluctuations in the United States, 1921-1941, Cowles Commission Monograph 11. New York: Wiley, 1950.

Klein, L. R. A textbook of econometrics. Evanston: Row, Peterson, 1953.

Klein, L. R. Estimation of interdependent systems in macroeconometrics. Econometrica, 1969, 37, 171-192.

Koopmans, T. C., Rubin, H., & Leipnik, R. B. Measuring the equation systems of dynamic economics. In T. C. Koopmans (Ed.), Statistical inference in dynamic economic models, Cowles Commission Monograph 10. New York: Wiley, 1950. Pp. 53-237.

Malinvaud, E. Statistical methods of econometrics. Chicago: Rand-McNally, 1966.

Rothenberg, T. G., & Leenders, C. T. Efficient estimation of simultaneous equation systems. Econometrica, 1964, 32, 57-76.

Turner, M. E., & Stevens, C. D. The regression analysis of causal paths. Biometrics, 1959, 15, 236-258.

Wegge, L. L. A family of functional iterations and the solution of maximum likelihood estimation equations. Econometrica, 1969, 37, 122-130.

Werts, C. E., & Linn, R. L. Path analysis: Psychological examples. Psychological Bulletin, 1970, 74 (3), 193-212.

Wold, H., & Jureen, L. Demand analysis. New York: Wiley, 1953.

Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 1962, 57, 348-368.

A. Appendices of Mathematical Derivations

A1. Matrix Derivatives of Function F in Section 3

The function is

$$F = \log |\Sigma| + \mathrm{tr}(S\Sigma^{-1}) \tag{A1}$$

which is regarded as a function of B, Γ, Φ, Ψ, Θ_δ, Θ_ϵ defined by (4). To derive the matrix derivatives we shall make use of matrix differentials. In general, $dX = (dx_{ij})$ will denote a matrix of differentials and if F is a function of X and $dF = \mathrm{tr}(C dX')$ then $\partial F/\partial X = C$.

Writing $A = B^{-1}$ and $D = B^{-1}\Gamma = A\Gamma$ we have

$$dA = -B^{-1}dBB^{-1} = -AdBA \tag{A2}$$

$$\begin{aligned} dD &= B^{-1}d\Gamma + dA\Gamma \\ &= Ad\Gamma - AdBA\Gamma \\ &= Ad\Gamma - AdBD \quad . \end{aligned} \tag{A3}$$

Furthermore, since in general,

$$d\log |X| = \mathrm{tr}(X^{-1}dX)$$

and

$$\begin{aligned} d\,\mathrm{tr}(AX^{-1}) &= \mathrm{tr}(AdX^{-1}) \\ &= -\mathrm{tr}(AX^{-1}dXX^{-1}) \\ &= -\mathrm{tr}(X^{-1}AX^{-1}dX) \end{aligned}$$

we obtain from (A1),

$$dF = d\log|\Sigma| + d\,\mathrm{tr}(S\Sigma^{-1})$$

$$= \mathrm{tr}(\Sigma^{-1}d\Sigma) - \mathrm{tr}(\Sigma^{-1}S\Sigma^{-1}d\Sigma)$$

$$= \mathrm{tr}[(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})d\Sigma]$$

$$= \mathrm{tr}(\Omega d\Sigma)$$

$$= \mathrm{tr}(\Omega_{yy}d\Sigma_{yy} + \Omega_{yx}d\Sigma_{xy} + \Omega_{xy}d\Sigma_{yx} + \Omega_{xx}d\Sigma_{xx}) \quad , \tag{A4}$$

where Ω is defined by (12) and $d\Sigma$ is partitioned the same way as Ω in (12).

From (4) and the definitions of A and D we have

$$\Sigma_{yy} = D\Phi D' + A\Psi A' + \Theta_\epsilon^2 \tag{A5}$$

$$\Sigma_{xy} = \Sigma'_{yx} = \Phi D' \tag{A6}$$

$$\Sigma_{xx} = \Phi + \Theta_\delta^2 \tag{A7}$$

from which we obtain

$$d\Sigma_{yy} = D\Phi dD' + Dd\Phi D' + dD\Phi D'$$

$$+ A\Psi dA' + Ad\Psi A' + dA\Psi A'$$

$$+ 2\Theta_\epsilon d\Theta_\epsilon \tag{A8}$$

$$d\Sigma_{xy} = \Phi dD' + d\Phi D' \tag{A9}$$

$$d\Sigma_{xx} = d\Phi + 2\Theta_\delta d\Theta_\delta \quad . \tag{A10}$$

Substitution of dA and dD from (A2) and (A3) into (A8) and (A9) gives

$$d\Sigma_{yy} = D\Phi d\Gamma' A' - D\Phi D' dB' A'$$

$$+ Ad\Gamma\Phi D' - AdBD\Phi D'$$

$$- A\Psi A' dB' A' - AdBA\Psi A'$$

$$+ Dd\Phi D' + Ad\Psi A' + 2\Theta_\epsilon d\Theta_\epsilon \quad , \qquad \text{(A11)}$$

$$d\Sigma_{xy} = \Phi d\Gamma' A' - \Phi D' dB' A' + d\Phi D' \qquad \text{(A12)}$$

Substitution of (A11), (A12) and (A10) into (A4), noting that $tr(C'dX) = tr(dX'C) = tr(CdX')$ and collecting terms, shows that the matrices multiplying dB', $d\Gamma'$, $d\Phi$, $d\Psi$, $d\Theta_\delta$ and $d\Theta_\epsilon$ are the matrices on the right sides of equations (14), (15), (17), (18) and (19) respectively. These are therefore the corresponding matrix derivatives.

A2. Information Matrix for the General Model of Section 3

In this section we shall prove a general theorem concerning the expected second-order derivatives of any function of the type (8) and show how this theorem can be applied to compute all the elements of the information matrix (12).

We first prove the following

Lemma: Let $S = (1/N) \sum_{\alpha=1}^{N} (z_\alpha - \bar{z})(z_\alpha - \bar{z})'$, where $z_1, z_2, \ldots, z_N$ are independently distributed according to $N(\mu, \Sigma)$. Then the asymptotic distribution of the elements of $\Omega = \Sigma^{-1}(\Sigma - S)\Sigma^{-1}$ is multivariate normal with means zero and variances and covariances given by

$$N\mathcal{E}(\omega_{\alpha\beta}\omega_{\mu\nu}) = \sigma^{\alpha\mu}\sigma^{\beta\nu} + \sigma^{\alpha\nu}\sigma^{\beta\mu} \quad . \tag{A13}$$

Proof: The proof follows immediately by multiplying $\omega_{\alpha\beta} = \sum_g \sum_h \sigma^{\alpha g}(\sigma_{gh} - s_{gh})\sigma^{h\beta}$ and $\omega_{\mu\nu} = \sum_i \sum_j \sigma^{\mu i}(\sigma_{ij} - s_{ij})\sigma^{j\nu}$ and using the fact that the asymptotic variances and covariances of S are given by

$$N\mathcal{E}[(\sigma_{gh} - s_{gh})(\sigma_{ij} - s_{ij})] = \sigma_{gi}\sigma_{hj} + \sigma_{gj}\sigma_{hi}$$

(see e.g., Anderson, Theorem 4.2.4).

We can now prove the following general theorem.

Theorem: Under the conditions of the above lemma let the elements of Σ be functions of two parameter matrices $M = (\mu_{gh})$ and $N = (\nu_{ij})$ and let $F(M,N) = \frac{1}{2} N[\log|\Sigma| + \mathrm{tr}(S\Sigma^{-1})]$ with $\partial F/\partial M = NA\Omega B$ and $\partial F/\partial N = NC\Omega D$. Then we have asymptotically

$$(1/N)\mathcal{E}(\partial^2 F/\partial\mu_{gh}\partial\nu_{ij}) = (A\Sigma^{-1}C')_{gi}(B'\Sigma^{-1}D)_{hj} + (A\Sigma^{-1}D)_{gj}(B'\Sigma^{-1}C')_{hi} \quad . \tag{A14}$$

Proof: Writing $\partial F/\partial\mu_{gh} = Na_{g\alpha}\omega_{\alpha\beta}b_{\beta h}$ and $\partial F/\partial\nu_{ij} = Nc_{i\mu}\omega_{\mu\nu}d_{\nu j}$, where it is assumed that every repeated subscript is to be summed over, we have

$$\begin{aligned}(1/N)\mathcal{E}(\partial^2 F/\partial\mu_{gh}\partial\nu_{ij}) &= (1/N)\mathcal{E}(\partial F/\partial\mu_{gh}\partial F/\partial\nu_{ij}) \\ &= N\,\mathcal{E}(a_{g\alpha}\omega_{\alpha\beta}b_{\beta h}c_{i\mu}\omega_{\mu\nu}d_{\nu j}) \\ &= N\,a_{g\alpha}b_{\beta h}c_{i\mu}d_{\nu j}\mathcal{E}(\omega_{\alpha\beta}\omega_{\mu\nu}) \\ &= a_{g\alpha}b_{\beta h}c_{i\mu}d_{\nu j}(\sigma^{\alpha\mu}\sigma^{\beta\nu} + \sigma^{\alpha\nu}\sigma^{\beta\mu})\end{aligned}$$

$$= (a_{g\alpha}\sigma^{\alpha\mu}c_{i\mu})(b_{\beta h}\sigma^{\beta\nu}d_{\nu j}) + (a_{g\alpha}\sigma^{\alpha\nu}d_{\nu j})(b_{\beta h}\sigma^{\beta\mu}c_{i\mu})$$

$$= (A\Sigma^{-1}C')_{gi}(B'\Sigma^{-1}D)_{hj} + (A\Sigma^{-1}D)_{gj}(B'\Sigma^{-1}C')_{hi} \quad .$$

It should be noted that the theorem is quite general in that both M and N may be row or column vectors or scalars and M and N may be identical in which case, of course, $A \equiv C$ and $B \equiv D$.

We now show how the above theorem can be applied repeatedly to compute all the elements of the information matrix (12). To do so we write the derivatives (14) - (19) in the form required by the theorem.

Let $A = B^{-1}$ and $D = B^{-1}\Gamma$, as before, and

$$T[m \times (m + n)] = [A' \quad 0] \tag{A15}$$

$$P[(m + n) \times m] = \begin{bmatrix} D\Phi D' + A\Psi A' \\ \Phi D' \end{bmatrix} \tag{A16}$$

$$Q[(m + n) \times n] = \begin{pmatrix} D\Phi \\ \Phi \end{pmatrix} \tag{A17}$$

$$R[(m + n) \times n] = \begin{pmatrix} D \\ I \end{pmatrix} \tag{A18}$$

Then it is readily verified that

$$\partial F/\partial B = -NT\Omega P \tag{A19}$$

$$\partial F/\partial \Gamma = NT\Omega Q \tag{A20}$$

$$\partial F/\partial \Phi = NR'\Omega R \tag{A21}$$

$$\partial F/\partial \Psi = NT\Omega T' \tag{A22}$$

$$\partial F/\partial \underset{\sim}{\Theta} = N\underset{\sim}{\Omega}\underset{\sim}{\Theta} \tag{A23}$$

In the last equation we have combined (18) and (19) using $\underset{\sim}{\Theta} = \begin{pmatrix} \underset{\sim}{\Theta}_\delta & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Theta}_\epsilon \end{pmatrix}$.

A3. Matrix Derivatives of Function F in Section 4

The function is defined by

$$F = \log |\underset{\sim}{\Psi}| - \log |\underset{\sim}{B}|^2 \quad , \tag{A24}$$

where

$$\underset{\sim}{\Psi} = \underset{\sim}{B}\underset{\sim}{S}_{yy}\underset{\sim}{B}' - \underset{\sim}{B}\underset{\sim}{S}_{yx}\underset{\sim}{\Gamma}' - \underset{\sim}{\Gamma}\underset{\sim}{S}_{xy}\underset{\sim}{B}' + \underset{\sim}{\Gamma}\underset{\sim}{S}_{xx}\underset{\sim}{\Gamma}' \quad . \tag{A25}$$

One finds immediately that

$$\begin{aligned} dF &= tr(\underset{\sim}{\Psi}^{-1}d\underset{\sim}{\Psi}) - 2tr(\underset{\sim}{B}^{-1}d\underset{\sim}{B}) \\ &= tr[\underset{\sim}{\Psi}^{-1}(d\underset{\sim}{B}\underset{\sim}{S}_{yy}\underset{\sim}{B}' + \underset{\sim}{B}\underset{\sim}{S}_{yy}d\underset{\sim}{B}' - d\underset{\sim}{B}\underset{\sim}{S}_{yx}\underset{\sim}{\Gamma}' - \underset{\sim}{\Gamma}\underset{\sim}{S}_{xy}d\underset{\sim}{B}')] \\ &\quad - 2tr(\underset{\sim}{B}^{-1}d\underset{\sim}{B}) \\ &\quad + tr[\underset{\sim}{\Psi}^{-1}(-\underset{\sim}{B}\underset{\sim}{S}_{yx}d\underset{\sim}{\Gamma}' - d\underset{\sim}{\Gamma}\underset{\sim}{S}_{xy}\underset{\sim}{B}' + d\underset{\sim}{\Gamma}\underset{\sim}{S}_{xx}\underset{\sim}{\Gamma}' + \underset{\sim}{\Gamma}\underset{\sim}{S}_{xx}d\underset{\sim}{\Gamma}')] \\ &= 2tr\{[\underset{\sim}{\Psi}^{-1}(\underset{\sim}{B}\underset{\sim}{S}_{yy} - \underset{\sim}{\Gamma}\underset{\sim}{S}_{xy}) - \underset{\sim}{B}'^{-1}]d\underset{\sim}{B}'\} \\ &\quad + 2tr[\underset{\sim}{\Psi}^{-1}(\underset{\sim}{\Gamma}\underset{\sim}{S}_{xx} - \underset{\sim}{B}\underset{\sim}{S}_{yx})d\underset{\sim}{\Gamma}'] \quad , \end{aligned}$$

so that the derivatives $\partial F/\partial \underset{\sim}{B}$ and $\partial F/\partial \underset{\sim}{\Gamma}$ are those given by (29) and (30).

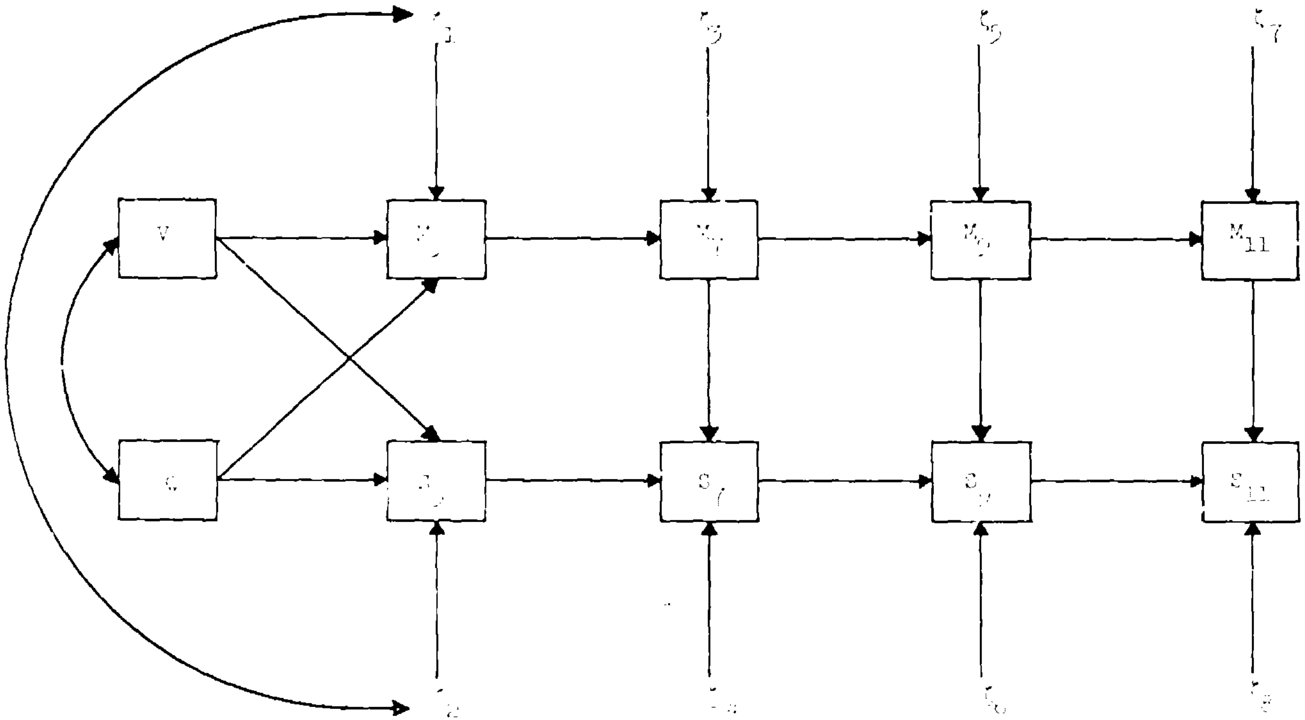

Fig. 1. Model for Achievements in Mathematics and Science