

INFERENCE IN LINEAR MODELS

BY T. W. ANDERSON*

Stanford University

A general linear model can be written as $\mathbf{Y} = \mathbf{X}\mathbf{B}' + \mathbf{U}$, where \mathbf{Y} is an $N \times p$ matrix of observable dependent variables, \mathbf{X} is an $N \times q$ matrix of independent variables, \mathbf{B}' is a $q \times p$ matrix of parameters, and \mathbf{U} is an $N \times p$ matrix of unobservable random variables. The elements of \mathbf{X} may be observable or alternatively unobservable (that is, latent); they may be nonstochastic or stochastic. The model includes regression, linear functional and structural relations, multivariate analysis of variance, factor analysis, and some simultaneous equations models. This paper considers the relationships between various models and presents methods of estimating the parameters under various conditions. Testing hypotheses about the rank of $\mathbf{X}\mathbf{B}'$ (the dimensionality of the latent variables when \mathbf{X} is not observed) are also treated.

1. A Linear Model. In this paper we consider a general linear model in multivariate analysis that includes regression models, multivariate analysis of variance (MANOVA) models, and factor analysis models. Some of these models go by names of linear functional relationships, linear structural relationships, and canonical correlations. An attempt will be made to use a unified approach to these models.

Suppose we observe the $p \times 1$ vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$. A linear model is given by

$$\mathbf{y}_\alpha = \mathbf{B}\mathbf{x}_\alpha + \mathbf{u}_\alpha, \quad \alpha = 1, \dots, N, \quad (1.1)$$

where $\mathbf{u}_1, \dots, \mathbf{u}_N$ are unobservable random $p \times 1$ vectors; we suppose them to be independently identically distributed (iid) with

$$\mathcal{E}\mathbf{u}_\alpha = \mathbf{0}, \quad \mathcal{E}\mathbf{u}_\alpha\mathbf{u}_\alpha' = \Sigma_u. \quad (1.2)$$

The $p \times q$ matrix \mathbf{B} consists of parameters, and $\mathbf{x}_1, \dots, \mathbf{x}_N$ are $q \times 1$ vectors. We shall consider these vectors as observed or alternatively as unobserved (or latent); they may be fixed (that is, nonstochastic) or alternatively random (or stochastic). The model can also be written

$$\mathbf{Y} = \mathbf{X}\mathbf{B}' + \mathbf{U}, \quad (1.3)$$

* This paper was prepared under National Science Foundation Grant DMS 89-04851 at Stanford University.

AMS 1980 Subject Classifications: 62H15, 62H25, 62J05, 62J10.

Key words and phrases: Factor analysis, latent variables, linear functional, MANOVA, rank, regression and structural relations.

2 INFERENCE IN LINEAR MODELS

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_N \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \mathbf{u}'_1 \\ \vdots \\ \mathbf{u}'_N \end{pmatrix}. \quad (1.4)$$

This paper is primarily expository. An important objective is to compare the models obtained by assigning different properties to \mathbf{X} and to show the interrelationships among the models. In each case the maximum likelihood estimators of \mathbf{B} and Σ_u under normality will be given as examples of estimation procedures. Also, likelihood ratio criteria for testing hypotheses about \mathbf{B} are given. A particular question is the rank of \mathbf{B} , which is the dimensionality of the space of $\mathbf{B}\mathbf{x}_\alpha$. The estimators and criteria in the different models will be compared. We shall not discuss alternative estimation and testing methods although they are important; we shall not develop the distributions or asymptotic distributions of the estimators and criteria. This paper updates (and abbreviates) the author's Wald lectures to the IMS in 1982 [Anderson (1984a)]. More details and references can be found in Chapters 8 and 12 of Anderson (1984b).

2. A Regression Model: \mathbf{X} Observed and Nonstochastic.

2.1. Estimators

For convenience we assume that the rank of \mathbf{X} is q . Then the least squares estimator of \mathbf{B} in the model of Section 1 is given by

$$\hat{\mathbf{B}}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (2.1)$$

We shall use the notation

$$\text{vec } \mathbf{A} = \text{vec}(\mathbf{a}_1, \dots, \mathbf{a}_m) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}, \quad (2.2)$$

where \mathbf{A} is an $n \times m$ matrix, and for the Kronecker product

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{bmatrix}. \quad (2.3)$$

The estimator $\hat{\mathbf{B}}$ is unbiased:

$$\mathcal{E}\hat{\mathbf{B}} = \mathbf{B}. \quad (2.4)$$

The covariances of the elements are given by

$$\text{cov} \left(\text{vec } \hat{\mathbf{B}} \right) = (\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma_u. \quad (2.5)$$

If \mathbf{U} is normally distributed, then $\widehat{\mathbf{B}}$, which is the maximum likelihood estimator of \mathbf{B} , is normally distributed. Under quite general conditions

$$\sqrt{N} \operatorname{vec}(\widehat{\mathbf{B}} - \mathbf{B}) \xrightarrow{d} N \left[\mathbf{0}, \left(\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}' \mathbf{X} \right)^{-1} \otimes \Sigma_u \right]. \quad (2.6)$$

Of course, it is assumed that $(1/N) \mathbf{X}' \mathbf{X}$ converges to a nonsingular limit.

An unbiased estimator of Σ_u is

$$\mathbf{S} = \frac{1}{N-q} (\mathbf{Y} - \mathbf{X} \widehat{\mathbf{B}}')' (\mathbf{Y} - \mathbf{X} \widehat{\mathbf{B}}'). \quad (2.7)$$

If \mathbf{U} is normal, $(N-q)\mathbf{S}$ has the Wishart distribution with covariance matrix Σ_u and $N-q$ degrees of freedom [denoted by $W(\Sigma_u, N-q)$]. Under normality the maximum likelihood estimator is $\widehat{\Sigma}_u = [(N-q)/N]\mathbf{S}$. Quite generally, $\widehat{\Sigma}_u \xrightarrow{P} \Sigma_u$.

2.2. MANOVA I (fixed effects)

A special case of the regression model is the analysis of variance with fixed effects. The balanced model is

$$\mathbf{y}_{\alpha j} = \boldsymbol{\mu}_{\alpha} + \mathbf{u}_{\alpha j}, \quad j = 1, \dots, k, \quad \alpha = 1, \dots, n. \quad (2.8)$$

The number of observations is $N = kn$. (We have modified the original indexing of the observations.) Let

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}'_{11} \\ \vdots \\ \mathbf{y}'_{1k} \\ \mathbf{y}'_{21} \\ \vdots \\ \mathbf{y}'_{nk} \end{bmatrix}. \quad (2.9)$$

If $\boldsymbol{\varepsilon}_k = (1, \dots, 1)'$ with k components, then

$$\mathcal{E}\mathbf{Y} = \begin{bmatrix} \boldsymbol{\varepsilon}_k \boldsymbol{\mu}'_1 \\ \vdots \\ \boldsymbol{\varepsilon}_k \boldsymbol{\mu}'_n \end{bmatrix} = \mathbf{X}\mathbf{M}, \quad (2.10)$$

where

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{\varepsilon}_k & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varepsilon}_k & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\varepsilon}_k \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}'_1 \\ \boldsymbol{\mu}'_2 \\ \vdots \\ \boldsymbol{\mu}'_n \end{bmatrix}. \quad (2.11)$$

4 INFERENCE IN LINEAR MODELS

The matrix \mathbf{M} takes the place of \mathbf{B}' . The least squares estimator of \mathbf{M} is

$$\begin{aligned}\widehat{\mathbf{M}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= (k\mathbf{I}_n)^{-1} \begin{bmatrix} \sum_{j=1}^k \mathbf{y}'_{1j} \\ \vdots \\ \sum_{j=1}^k \mathbf{y}'_{nj} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}}'_1 \\ \vdots \\ \bar{\mathbf{y}}'_n \end{bmatrix},\end{aligned}\quad (2.12)$$

where

$$\bar{\mathbf{y}}_\alpha = \frac{1}{k} \sum_{j=1}^k \mathbf{y}_{\alpha j}. \quad (2.13)$$

The rows of $\widehat{\mathbf{M}}$ are independent, and $\bar{\mathbf{y}}_\alpha \sim N[\mu_\alpha, (1/k)\Sigma_u]$ when \mathbf{U} is normal. The estimator $\widehat{\mathbf{M}}$ is unbiased: $\mathcal{E}\widehat{\mathbf{M}} = \mathbf{M}$ and $\text{cov}(\text{vec } \widehat{\mathbf{M}}) = (1/k)\Sigma_u \otimes \mathbf{I}_n$. Consider testing the null hypothesis that the fixed effects are identical, that is, $H : \mu_1 = \cdots = \mu_n$ or

$$H : (\mu_1 - \bar{\mu}, \dots, \mu_n - \bar{\mu}) = \mathbf{0}, \quad (2.14)$$

where

$$\bar{\mu} = \frac{1}{n} \sum_{\alpha=1}^n \mu_\alpha. \quad (2.15)$$

Write the matrix of means as

$$\mathbf{M} = \begin{bmatrix} \mu'_1 - \bar{\mu}' \\ \vdots \\ \mu'_n - \bar{\mu}' \end{bmatrix} + \varepsilon_n \bar{\mu}'. \quad (2.16)$$

The hypothesis is that the rank of the first matrix on the right-hand side of (2.16) is 0. The “effect sum of squares” is

$$\mathbf{H} = k \sum_{\alpha=1}^n (\bar{\mathbf{y}}_\alpha - \bar{\mathbf{y}})(\bar{\mathbf{y}}_\alpha - \bar{\mathbf{y}})', \quad (2.17)$$

where $\bar{\mathbf{y}} = (1/n) \sum_{\alpha=1}^n \bar{\mathbf{y}}_\alpha$; the “error sum of squares” is

$$\mathbf{G} = \sum_{\alpha=1}^n \sum_{j=1}^k (\mathbf{y}_{\alpha j} - \bar{\mathbf{y}}_\alpha)(\mathbf{y}_{\alpha j} - \bar{\mathbf{y}}_\alpha)'. \quad (2.18)$$

The likelihood ratio criterion λ for testing H is defined by

$$\lambda^{2/n} = \frac{|\mathbf{G}|^k}{|\mathbf{G} + \mathbf{H}|^k}. \quad (2.19)$$

We reject H if $\lambda^{2/n} < \lambda_0^{2/n}$, where λ_0 is a constant such that

$$\Pr\{\lambda < \lambda_0\} = \alpha \quad (2.20)$$

under H and α is a specified significance level. [The probability (2.20) does not depend on Σ_u or $\bar{\mu}$ when H is true.]

2.3. Regression matrices of specified rank

The null hypothesis that the effect means are equal is that in the p -dimensional space of \boldsymbol{y}_α the mean vectors represent a point, that is, the dimension of the space of effect means is 0. We now consider the general case of

$$\text{rank } \begin{bmatrix} \boldsymbol{\mu}'_1 - \bar{\mu}' \\ \vdots \\ \boldsymbol{\mu}'_n - \bar{\mu}' \end{bmatrix} = m, \quad (2.21)$$

where m is a specified integer between 0 and $p - 1$. That $m = 1$, for example, is that the means $\boldsymbol{\mu}_\alpha$ lie on a line; in that case the classes or populations can be ordered. That $m = 2$ is that the effect means lie in a plane. We now consider estimating the matrix $(\boldsymbol{\mu}_1 - \bar{\mu}, \dots, \boldsymbol{\mu}_n - \bar{\mu})$ under the condition (2.21). Let

$$\widetilde{\mathbf{H}} = \frac{1}{n} \mathbf{H}, \quad \widetilde{\mathbf{G}} = \frac{1}{n(k-1)} \mathbf{G}, \quad (2.22)$$

and let the roots of

$$|\widetilde{\mathbf{H}} - d\widetilde{\mathbf{G}}| = 0 \quad (2.23)$$

be

$$d_1 > \dots > d_p. \quad (2.24)$$

Define the vectors \mathbf{w}_i as satisfying

$$\widetilde{\mathbf{H}} \mathbf{w}_i = d_i \widetilde{\mathbf{G}} \mathbf{w}_i, \quad i = 1, \dots, p, \quad (2.25)$$

$$\mathbf{w}'_i \widetilde{\mathbf{G}} \mathbf{w}_i = 1, \quad i = 1, \dots, p. \quad (2.26)$$

Define

$$\mathbf{W}_1 = (\mathbf{w}_1, \dots, \mathbf{w}_m), \quad \mathbf{W}_2 = (\mathbf{w}_{m+1}, \dots, \mathbf{w}_p), \quad (2.27)$$

$$\mathbf{D}_1 = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_m \end{bmatrix}, \quad \mathbf{D}_2 = \begin{bmatrix} d_{m+1} & 0 & \dots & 0 \\ 0 & d_{m+2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_p \end{bmatrix}, \quad (2.28)$$

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix}, \quad \mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2), \quad (2.29)$$

$$\mathbf{Z} = (\mathbf{W}')^{-1} = (\mathbf{Z}_1, \mathbf{Z}_2), \quad (2.30)$$

6 INFERENCE IN LINEAR MODELS

where the matrices are conformable; note that

$$\mathbf{I} = \mathbf{WZ}' = (\mathbf{W}_1\mathbf{Z}_1' + \mathbf{W}_2\mathbf{Z}_2'). \quad (2.31)$$

The maximum likelihood estimator (under normality) of $(\boldsymbol{\mu}_1 - \bar{\boldsymbol{\mu}}, \dots, \boldsymbol{\mu}_n - \bar{\boldsymbol{\mu}})'$ of rank m is

$$\begin{bmatrix} \bar{\mathbf{y}}_1' - \bar{\mathbf{y}}' \\ \vdots \\ \bar{\mathbf{y}}_n' - \bar{\mathbf{y}}' \end{bmatrix} \mathbf{W}_1 \mathbf{Z}_1'. \quad (2.32)$$

Note that the columns of \mathbf{W}_1 are the characteristic vectors of $\tilde{\mathbf{G}}^{-1} \tilde{\mathbf{H}}$ associated with the m largest characteristic roots. These maximum likelihood estimators were derived by Anderson (1951); Fisher (1938) considered a related problem in discriminant analysis.

To test the null hypothesis $H : m = m_0$, where m is given by (2.21) and m_0 is specified, the likelihood ratio criterion λ is given by

$$\lambda^{2/n} = \prod_{i=m_0+1}^p \left(\frac{k-1}{k-1+d_i} \right)^k. \quad (2.33)$$

We reject H if $\lambda < \lambda_0$, where λ_0 is a suitable constant. Equivalently

$$-2 \log \lambda = nk \sum_{i=m_0+1}^p \log \left(1 + \frac{d_i}{k-1} \right). \quad (2.34)$$

We reject H if $-2 \log \lambda > -2 \log \lambda_0$. As $k \rightarrow \infty$, under the null hypothesis

$$-2 \log \lambda \sim \chi^2_{(p-m_0)(n-m_0-1)}. \quad (2.35)$$

Note that the hypothesis is rejected if the $p - m_0$ smallest roots are too large.

If (2.21) holds, then there exists a $(p - m) \times p$ matrix Γ such that

$$\Gamma(\boldsymbol{\mu}_1 - \bar{\boldsymbol{\mu}}, \dots, \boldsymbol{\mu}_n - \bar{\boldsymbol{\mu}}) = \mathbf{0} \quad (2.36)$$

or

$$\Gamma \boldsymbol{\mu}_\alpha = \Gamma \bar{\boldsymbol{\mu}}, \quad \alpha = 1, \dots, n. \quad (2.37)$$

Note that Γ is not unique; (2.36) or (2.37) can be multiplied on the left by an arbitrary nonsingular matrix. The set of equations (2.36) or (2.37) is known as a set of *linear functional relationships*. A maximum likelihood estimator of Γ (under normality) is

$$\hat{\Gamma} = \mathbf{W}_2'. \quad (2.38)$$

Note that the rows of $\hat{\Gamma}$ (the columns of \mathbf{W}_2) are the characteristic vectors of $\tilde{\mathbf{G}}^{-1} \tilde{\mathbf{H}}$ associated with the $p - m_0$ smallest characteristic roots.

For later comparison we shall want a canonical form for MANOVA I. For an integer l let

$$\mathbf{Q}_l = \begin{bmatrix} 1/\sqrt{l} & 1/\sqrt{l} & \cdots & 1/\sqrt{l} \\ q_{21} & q_{22} & \cdots & q_{2l} \\ \vdots & \vdots & & \vdots \\ q_{l1} & q_{l2} & \cdots & q_{ll} \end{bmatrix} \quad (2.39)$$

be a $l \times l$ orthogonal matrix. Define

$$\begin{bmatrix} \mathbf{y}_{\alpha 1}^{*''} \\ \mathbf{y}_{\alpha 2}^{*''} \\ \vdots \\ \mathbf{y}_{\alpha k}^{*''} \end{bmatrix} = \mathbf{Q}_k \begin{bmatrix} \mathbf{y}'_{\alpha 1} \\ \mathbf{y}'_{\alpha 2} \\ \vdots \\ \mathbf{y}'_{\alpha k} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{u}_{\alpha 1}^{*''} \\ \mathbf{u}_{\alpha 2}^{*''} \\ \vdots \\ \mathbf{u}_{\alpha k}^{*''} \end{bmatrix} = \mathbf{Q}_k \begin{bmatrix} \mathbf{u}'_{\alpha 1} \\ \mathbf{u}'_{\alpha 2} \\ \vdots \\ \mathbf{u}'_{\alpha k} \end{bmatrix}. \quad (2.40)$$

Thus

$$\mathbf{y}_{\alpha 1}^* = \sqrt{k} \bar{\mathbf{y}}_{\alpha} = \sqrt{k} \boldsymbol{\mu}_{\alpha} + \sqrt{k} \bar{\mathbf{u}}_{\alpha} \sim N(\sqrt{k} \boldsymbol{\mu}_{\alpha}, \boldsymbol{\Sigma}_u), \quad (2.41)$$

where $\bar{\mathbf{u}}_{\alpha} = (1/k) \sum_{j=1}^k \mathbf{u}_{\alpha j}$, and

$$\mathbf{y}_{\alpha j}^* = \mathbf{u}_{\alpha j}^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u), \quad j = 2, \dots, k. \quad (2.42)$$

The vectors $\mathbf{y}_{\alpha j}^*$, $\alpha = 1, \dots, n$, $j = 1, \dots, k$, are independent. Further let

$$\begin{bmatrix} \mathbf{y}_{11}^{**'} \\ \vdots \\ \mathbf{y}_{n1}^{**'} \end{bmatrix} = \mathbf{Q}_n \begin{bmatrix} \mathbf{y}'_{11} \\ \vdots \\ \mathbf{y}'_{n1} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{u}_{11}^{**'} \\ \vdots \\ \mathbf{u}_{n1}^{**'} \end{bmatrix} = \mathbf{Q}_n \begin{bmatrix} \mathbf{u}'_{11} \\ \vdots \\ \mathbf{u}'_{n1} \end{bmatrix}. \quad (2.43)$$

Then

$$\mathbf{y}_{11}^{**} = \sqrt{n} \bar{\mathbf{y}}_1^* = \sqrt{nk} \bar{\mathbf{y}} \sim N(\sqrt{nk} \bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_u), \quad (2.44)$$

where $\bar{\mathbf{y}}_1^* = (1/n) \sum_{\alpha=1}^n \mathbf{y}_{\alpha 1}^*$, and

$$\mathbf{y}_{\alpha 1}^{**} = \mathbf{u}_{\alpha 1}^{**} \sim N(\boldsymbol{\gamma}_{\alpha}, \boldsymbol{\Sigma}_u), \quad \alpha = 2, \dots, n. \quad (2.45)$$

where $\boldsymbol{\gamma}_{\alpha}$ depends on α , \mathbf{Q}_n , and $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n$. The vectors $\mathbf{y}_{\alpha 1}^{**}$, $\alpha = 1, \dots, n$, are independent. Then

$$\mathbf{H} = \sum_{\alpha=1}^n (\mathbf{y}_{\alpha 1}^* - \bar{\mathbf{y}}_1^*)(\mathbf{y}_{\alpha 1}^* - \bar{\mathbf{y}}_1^*)' = \sum_{\alpha=2}^n \mathbf{y}_{\alpha 1}^{**} \mathbf{y}_{\alpha 1}^{**'}, \quad (2.46)$$

$$\mathbf{G} = \sum_{\alpha=1}^n \sum_{j=2}^k \mathbf{y}_{\alpha j}^* \mathbf{y}_{\alpha j}^{*'}. \quad (2.47)$$

Here it is clear that \mathbf{H} and \mathbf{G} have independent Wishart distributions with $n-1$ and $n(k-1)$ degrees of freedom, respectively; the distribution of \mathbf{H} may be noncentral.

8 INFERENCE IN LINEAR MODELS

Now we consider estimation and hypothesis testing of \mathbf{B} for the general regression model. Note that

$$\Pr \{ \text{rank } \hat{\mathbf{B}} = \min(p, q) \} = 1. \quad (2.48)$$

We may want to estimate \mathbf{B} of which a submatrix has lower rank. Let

$$\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2), \quad (2.49)$$

where \mathbf{B}_1 and \mathbf{B}_2 have q_1 and q_2 columns, respectively; we want to estimate \mathbf{B}_2 of rank m [$0 \leq m \leq \min(p, q_2)$], where m is specified. Let \mathbf{X} be partitioned into submatrices of q_1 and q_2 columns

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2). \quad (2.50)$$

The model is

$$\begin{aligned} \mathbf{Y} &= (\mathbf{X}_1, \mathbf{X}_2) \begin{pmatrix} \mathbf{B}'_1 \\ \mathbf{B}'_2 \end{pmatrix} + \mathbf{U} \\ &= \mathbf{X}_1 \mathbf{B}'_1 + \mathbf{X}_2 \mathbf{B}'_2 + \mathbf{U}. \end{aligned} \quad (2.51)$$

Let

$$\mathbf{A} = \mathbf{X}' \mathbf{X} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad (2.52)$$

$$\begin{aligned} \mathbf{A}_{22.1} &= \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \\ &= (\mathbf{X}_2 - \mathbf{X}_1 \mathbf{A}_{11}^{-1} \mathbf{A}_{12})' (\mathbf{X}_2 - \mathbf{X}_1 \mathbf{A}_{11}^{-1} \mathbf{A}_{12}) \\ &= \mathbf{X}'_{2.1} \mathbf{X}_{2.1}, \end{aligned} \quad (2.53)$$

where $\mathbf{X}_{2.1} = \mathbf{X}_2 - \mathbf{X}_1 \mathbf{A}_{11}^{-1} \mathbf{A}_{12}$. Let

$$\begin{aligned} q\widetilde{\mathbf{H}} &= \mathbf{H} = (\mathbf{X}_{2.1} \widehat{\mathbf{B}}'_2)' (\mathbf{X}_{2.1} \widehat{\mathbf{B}}'_2) \\ &= \widehat{\mathbf{B}}_2 \mathbf{A}_{22.1} \widehat{\mathbf{B}}'_2. \end{aligned} \quad (2.54)$$

$$\begin{aligned} (N - q)\widetilde{\mathbf{G}} &= \mathbf{G} = (N - q)\widehat{\Sigma}_u \\ &= (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}')' (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}'). \end{aligned} \quad (2.55)$$

The maximum likelihood estimator of \mathbf{B}'_2 of rank m (under normality) is

$$\widehat{\mathbf{B}}'_2(m) = \widehat{\mathbf{B}}'_2 \mathbf{W}_1 \mathbf{Z}'_1. \quad (2.56)$$

Since

$$\mathbf{W}'_2 \hat{\mathbf{B}}_2(m) = \mathbf{0}, \quad (2.57)$$

an estimator for Γ satisfying $\Gamma \mathbf{B}_2 = \mathbf{0}$ is \mathbf{W}'_2 .

2.4. An example from econometrics

Simultaneous equation models are used extensively for macro-economics. Estimation of the coefficients of the relationships involves — at least implicitly — the estimation of regression matrices of specified rank. We illustrate by the simple example of a pair of demand and supply functions.

Let y_{1t} be the quantity of a good consumed in period t , y_{2t} be the price of the good produced, x_{1t} be the aggregate income of consumers in the market, x_{2t} be the cost of one raw material for producing the good, and x_{3t} be the cost of a second raw material. The demand relation is

$$y_{1t} = \gamma_0 + \alpha y_{2t} + \gamma_1 x_{1t} + v_{1t}, \quad (2.58)$$

where v_{1t} is an unobserved random term. The quantity of the good desired by the consumers depends on the price at which the good is offered and on the income of the consumers. It is a stochastic relationship. The supply relation may be

$$y_{1t} = \phi_0 + \delta y_{2t} + \phi_2 x_{2t} + \phi_3 x_{3t} + v_{2t}, \quad (2.59)$$

where v_{2t} is an unobservable random term. Here y_{1t} represents the quantity that producers would supply at the price y_{2t} . In the market the price is adjusted so that the quantity desired by the consumers equals the quantity the producers are willing to supply. This price and quantity form the solution to (2.58) and (2.59) for y_{1t} and y_{2t} . We write the solution as the regression

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \pi_{10} & \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{20} & \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} \begin{bmatrix} 1 \\ x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \quad (2.60)$$

where u_{1t} and u_{2t} are linear combinations of v_{1t} and v_{2t} . It will be convenient to re-write (2.58) as

$$\alpha_1 y_{1t} + \alpha_2 y_{2t} = \gamma_0 1 + \gamma_1 x_{1t} + v_{1t}. \quad (2.61)$$

If we write $\alpha_1 y_{1t} + \alpha_2 y_{2t}$ in terms of the regression form, we obtain

$$(\alpha_1, \alpha_2) \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = (\alpha_1, \alpha_2) \begin{bmatrix} \pi_{10} & \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{20} & \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} \begin{bmatrix} 1 \\ x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} + (\alpha_1, \alpha_2) \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \quad (2.62)$$

10 INFERENCE IN LINEAR MODELS

This is

$$\alpha_1 y_{1t} + \alpha_2 y_{2t} = (\alpha_1 \pi_{10} + \alpha_2 \pi_{20})1 + (\alpha_1 \pi_{11} + \alpha_2 \pi_{21})x_{1t} + (\alpha_1 u_{1t} + \alpha_2 u_{2t}) \quad (2.63)$$

if

$$(\alpha_1, \alpha_2) \begin{bmatrix} \pi_{12} & \pi_{13} \\ \pi_{22} & \pi_{23} \end{bmatrix} = (0, 0). \quad (2.64)$$

Then (2.63) agrees with the demand function (2.61) if $\gamma_0 = \alpha_1 \pi_{10} + \alpha_2 \pi_{20}$ and $\gamma_1 = \alpha_1 \pi_{11} + \alpha_2 \pi_{21}$. For (2.64) to hold, one needs

$$\text{rank} \begin{bmatrix} \pi_{12} & \pi_{13} \\ \pi_{22} & \pi_{23} \end{bmatrix} = 1. \quad (2.65)$$

To estimate the coefficients of (2.61) the matrix of regression coefficients is estimated under the rank condition (2.65). Then an estimator of (α_1, α_2) is a solution to (2.64) with the π 's replaced by their estimators; the nonuniqueness in the solution may be eliminated by requiring α_1 to be 1. This estimation method is known as the Limited Information Maximum Likelihood procedure [Anderson and Rubin (1949)].

3. A Regression Model: X Observed and Stochastic. We shall now use our basic model (1.1) to develop the model that is usually associated with canonical correlations and canonical variates. Suppose \mathbf{x}_α has the form

$$\mathbf{x}_\alpha = \begin{pmatrix} 1 \\ \mathbf{x}_\alpha^{(2)} \end{pmatrix}. \quad (3.1)$$

Let $\mathbf{x}_1^{(2)}, \dots, \mathbf{x}_N^{(2)}$ be independent random vectors with $q_2 = q - 1$ components and

$$\mathcal{E}\mathbf{x}_\alpha^{(2)} = \boldsymbol{\mu}_x, \quad \mathcal{E}(\mathbf{x}_\alpha^{(2)} - \boldsymbol{\mu}_x)(\mathbf{x}_\alpha^{(2)} - \boldsymbol{\mu}_x)' = \boldsymbol{\Sigma}_x, \quad (3.2)$$

and independent of $\mathbf{u}_1, \dots, \mathbf{u}_N$. If we write $\mathbf{B}_1 = \boldsymbol{\mu}_y - \mathbf{B}_2 \boldsymbol{\mu}_x$, the model (1.1) is

$$\mathbf{y}_\alpha = \boldsymbol{\mu}_y + \mathbf{B}_2(\mathbf{x}_\alpha^{(2)} - \boldsymbol{\mu}_x) + \mathbf{u}_\alpha, \quad \alpha = 1, \dots, N. \quad (3.3)$$

Here $\mathcal{E}\mathbf{y}_\alpha = \boldsymbol{\mu}_y$ and, since $\mathbf{x}_\alpha^{(2)}$ and \mathbf{u}_α are uncorrelated, the covariance matrix of $(\mathbf{y}_\alpha', \mathbf{x}_\alpha^{(2)'})'$ is

$$\text{cov} \begin{bmatrix} \mathbf{y}_\alpha \\ \mathbf{x}_\alpha^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_u + \mathbf{B}_2 \boldsymbol{\Sigma}_x \mathbf{B}_2' & \mathbf{B}_2 \boldsymbol{\Sigma}_x \\ \boldsymbol{\Sigma}_x \mathbf{B}_2' & \boldsymbol{\Sigma}_x \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x \end{bmatrix}, \quad (3.4)$$

say.

The canonical correlations between \mathbf{y}_α and $\mathbf{x}_\alpha^{(2)}$ are defined by

$$0 = \begin{vmatrix} -\rho \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & -\rho \Sigma_x \end{vmatrix} = \begin{vmatrix} -\rho(\Sigma_u + B_2 \Sigma_x B_2') & B_2 \Sigma_x \\ \Sigma_x B_2' & -\rho \Sigma_x \end{vmatrix}. \quad (3.5)$$

If we multiply (3.5) on the right by a determinant of value 1 (for $\rho \neq 0$), we obtain

$$\begin{aligned} 0 &= \begin{vmatrix} -(\Sigma_u + B_2 \Sigma_x B_2')\rho & B_2 \Sigma_x \\ \Sigma_x B_2' & -\Sigma_x \rho \end{vmatrix} \cdot \begin{vmatrix} I & \mathbf{0} \\ \frac{1}{\rho} B_2' & I \end{vmatrix} \\ &= \begin{vmatrix} -(\Sigma_u + B_2 \Sigma_x B_2')\rho + \frac{1}{\rho} B_2 \Sigma_x B_2' & B_2 \Sigma_x \\ \mathbf{0} & -\rho \Sigma_x \end{vmatrix} \\ &= \rho^{q_2-p} \begin{vmatrix} B_2 \Sigma_x B_2' - (\Sigma_u + B_2 \Sigma_x B_2')\rho^2 & B_2 \Sigma_x \\ \mathbf{0} & -\Sigma_x \end{vmatrix} \\ &= \rho^{q_2-p} |-\Sigma_x| \cdot |B_2 \Sigma_x B_2' - (\Sigma_u + B_2 \Sigma_x B_2')\rho^2| \\ &= \rho^{q_2-p} |-\Sigma_x| \cdot |B_2 \Sigma_x B_2'(1 - \rho^2) - \Sigma_u \rho^2|. \end{aligned} \quad (3.6)$$

If we replace $\rho^2/(1 - \rho^2)$ by λ , the equation (3.6) is equivalent to

$$0 = |B_2 \Sigma_x B_2' - \lambda \Sigma_u|. \quad (3.7)$$

Note that the number of canonical correlations different from 0 is the rank of B_2 .

In (2.50) let $\mathbf{X}_1 = \boldsymbol{\epsilon}_N$ and $\mathbf{X}_2 = (\mathbf{x}_1^{(2)}, \dots, \mathbf{x}_N^{(2)})'$. Then

$$\mathbf{A}_{22.1} = \sum_{\alpha=1}^N \mathbf{x}_\alpha^{(2)} \mathbf{x}_\alpha^{(2)'} - N \bar{\mathbf{x}}^{(2)} \bar{\mathbf{x}}^{(2)'} = \sum_{\alpha=1}^N (\mathbf{x}_\alpha^{(2)} - \bar{\mathbf{x}}^{(2)})(\mathbf{x}_\alpha^{(2)} - \bar{\mathbf{x}}^{(2)})', \quad (3.8)$$

$$\hat{\mathbf{B}}_2 = \sum_{\alpha=1}^N (\mathbf{y}_\alpha - \bar{\mathbf{y}})(\mathbf{x}_\alpha^{(2)} - \bar{\mathbf{x}}^{(2)})' \mathbf{A}_{22.1}^{-1}, \quad (3.9)$$

where $\bar{\mathbf{y}} = (1/N) \sum_{\alpha=1}^N \mathbf{y}_\alpha$ and $\bar{\mathbf{x}}^{(2)} = (1/N) \sum_{\alpha=1}^N \mathbf{x}_\alpha^{(2)}$. Then in (2.23) we use $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$ defined by

$$q \tilde{\mathbf{H}} = \mathbf{H} = \hat{\mathbf{B}}_2 \mathbf{A}_{22.1} \hat{\mathbf{B}}_2', \quad (3.10)$$

$$(N - q) \tilde{\mathbf{G}} = \mathbf{G} = \sum_{\alpha=1}^N (\mathbf{y}_\alpha - \bar{\mathbf{y}})(\mathbf{y}_\alpha - \bar{\mathbf{y}})' - \hat{\mathbf{B}}_2 \mathbf{A}_{22.1} \hat{\mathbf{B}}_2', \quad (3.11)$$

which is a form of (2.55). If $\mathbf{x}_\alpha^{(2)}$ is normally distributed, then $(\mathbf{y}_\alpha', \mathbf{x}_\alpha^{(2)'})'$ is normally distributed, the maximum likelihood estimator of Σ_x is

$$\hat{\Sigma}_x = \frac{1}{N} \mathbf{A}_{22.1}, \quad (3.12)$$

12 INFERENCE IN LINEAR MODELS

and the maximum likelihood estimator of Σ_y is

$$\widehat{\Sigma}_y = \widehat{\Sigma}_u + \widehat{\mathbf{B}}_2 \widehat{\Sigma}_x \widehat{\mathbf{B}}_2' = \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{y}_\alpha - \bar{\mathbf{y}})(\mathbf{y}_\alpha - \bar{\mathbf{y}})'. \quad (3.13)$$

The maximum likelihood estimators of the canonical correlations are the roots of (3.6) with Σ_u , Σ_x and \mathbf{B}_2 replaced by their maximum likelihood estimators.

Under normality the likelihood function of $\mu_y, \mu_x, \Sigma_u, \mathbf{B}_2$ and Σ_x (or alternatively $\mathbf{B}_1, \mu_x, \Sigma_u, \mathbf{B}_2$ and Σ_x) given the observations factors into the likelihood function of $\mathbf{B}_1, \mathbf{B}_2$, and Σ_u given the observations times the likelihood function of μ_x and Σ_x given $\mathbf{x}_1^{(2)}, \dots, \mathbf{x}_N^{(2)}$. Hence, likelihood ratio criteria concerning the rank of \mathbf{B}_2 (or alternatively Σ_{yx}) are the criteria of Section 2, and the maximum likelihood estimator of \mathbf{B}'_2 is given by (2.56) with \mathbf{H} and \mathbf{G} defined by (3.10) and (3.11). (Note that \mathbf{W}_1 and \mathbf{Z}_1 can be defined in terms of \mathbf{H} and \mathbf{G} .)

4. A Latent Variable Model: \mathbf{X} Unobserved and Nonstochastic. When \mathbf{X} is unobserved, $\mathcal{E}\mathbf{Y} = \mathbf{X}\mathbf{B}'$ is an $N \times p$ matrix of rank at most q (since \mathbf{X} has q columns and \mathbf{B}' has q rows); $\mathcal{E}\mathbf{Y}$ is a matrix of parameters. If $p < q$, there are more entries in \mathbf{X} than in \mathbf{Y} and hence \mathbf{X} and \mathbf{B} are unidentified.

We shall restrict attention to the case of $q \leq p$ and $\text{rank } \mathbf{X} = q$. Then the columns of \mathbf{X} define a q -dimensional subspace in the N -dimensional space of the columns of \mathbf{Y} and the columns of $\mathcal{E}\mathbf{Y} = \mathbf{X}\mathbf{B}'$ lie in that subspace.

Indeterminacy. We see that for any nonsingular matrix \mathbf{A}

$$\mathbf{X}\mathbf{B}' = \mathbf{X}\mathbf{A}\mathbf{A}^{-1}\mathbf{B}' = (\mathbf{X}\mathbf{A})(\mathbf{A}^{-1}\mathbf{B}'). \quad (4.1)$$

Thus we can make the transformation

$$\mathbf{X} \rightarrow \mathbf{X}\mathbf{A}, \quad \mathbf{B}' \rightarrow \mathbf{A}^{-1}\mathbf{B}' \quad (4.2)$$

without affecting $\mathcal{E}\mathbf{Y} = \mathbf{X}\mathbf{B}'$. We can eliminate this indeterminacy by requiring

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}^* \\ \mathbf{I}_q \end{pmatrix}. \quad (4.3)$$

If we also partition \mathbf{y}_α and \mathbf{u}_α as $(\mathbf{y}_\alpha^{(1)'}, \mathbf{y}_\alpha^{(2)'})'$ and $(\mathbf{u}_\alpha^{(1)'}, \mathbf{u}_\alpha^{(2)'})'$, we can write (1.1) as

$$\mathbf{y}_\alpha^{(1)} = \mathbf{B}^*\mathbf{x}_\alpha + \mathbf{u}_\alpha^{(1)}, \quad (4.4)$$

$$\mathbf{y}_\alpha^{(2)} = \mathbf{x}_\alpha + \mathbf{u}_\alpha^{(2)}.$$

This is sometimes known as the “errors in variables” model. The unknown parameters are \mathbf{B}^* , $\mathbf{x}_1, \dots, \mathbf{x}_n$, and Σ_u . Another way of eliminating the indeterminacy is to require

$$\frac{1}{N} \mathbf{X}' \mathbf{X} = \mathbf{I} \quad (4.5)$$

and some conditions on \mathbf{B} .

Even if the indeterminacy of transformations (4.2) is eliminated, all of the parameters may not be identified; restrictions on Σ_u may also be necessary. In Section 6 the case of Σ_u diagonal (factor analysis) is discussed.

Linear functional relationships. Let Γ be a $(p - q) \times p$ matrix such that

$$\Gamma \mathbf{B} = \mathbf{0}, \quad (4.6)$$

where $\mathbf{0}$ is $(p - q) \times q$. Then

$$(\mathcal{E}\mathbf{Y}) \Gamma' = \mathbf{X} \mathbf{B}' \Gamma' = \mathbf{0}. \quad (4.7)$$

The components of (4.7) are called *linear functional relationships*.

There is an indeterminacy in (4.6) or in (4.7) because (4.6), for instance, can be multiplied on the left by an arbitrary nonsingular $(p - q) \times (p - q)$ matrix \mathbf{C} ; then

$$\mathbf{C} \Gamma \mathbf{B} = \mathbf{0}. \quad (4.8)$$

Thus we can make the transformation

$$\Gamma \rightarrow \mathbf{C} \Gamma \quad (4.9)$$

without affecting (4.7). To eliminate this indeterminacy we can require

$$\Gamma = (\mathbf{I}_{p-q}, \Gamma^*). \quad (4.10)$$

Then if (4.3) holds

$$\mathbf{0} = \Gamma \mathbf{B} = (\mathbf{I}_{p-q}, \Gamma^*) \begin{pmatrix} \mathbf{B}^* \\ \mathbf{I}_q \end{pmatrix} = \mathbf{B}^* + \Gamma^*; \quad (4.11)$$

that is,

$$\Gamma^* = -\mathbf{B}^*, \quad (4.12)$$

where each matrix is $(p - q) \times q$. Thus the two ways of writing the model are equivalent, and an estimator of \mathbf{B}^* is an estimator of $-\Gamma^*$ as well.

The elements of \mathbf{X} are called latent variables or factor scores or incidental parameters; the elements of \mathbf{B} are structural parameters; and the elements of \mathbf{U} are errors. Increasing the sample size (increasing N) increases the number of values of the latent variables (the number of rows of \mathbf{X}), but does not increase the number of structural parameters. Usually it is desirable to have

14 INFERENCE IN LINEAR MODELS

a small value of q (number of factors). A model with such a small value is called parsimonious.

Even if q is small, however, the normal likelihood does not have a maximum [Anderson and Rubin (1956)]. Hence, a maximum likelihood estimator of $\mathbf{X}\mathbf{B}$ (and/or \mathbf{X} and \mathbf{B}) and Σ_u does not exist. It is possible, however, to estimate the parameters by other means if they are identified. One method is to use the maximum likelihood estimators of \mathbf{B} (or \mathbf{B}^*) and Σ_u derived under the assumption that \mathbf{X} is random; see Section 5.

We turn to the case where there is an independent estimator of Σ_u . Suppose $\mathbf{x}'_\alpha = (1, \mathbf{x}_\alpha^{(2)'}')$ and $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$, where $\mathbf{x}_\alpha^{(2)}$ has $q_2 = q - 1$ components and \mathbf{B}_2 has q_2 columns; then we can write

$$\mathbf{y}_\alpha = \boldsymbol{\phi} + \mathbf{B}_2 \left(\mathbf{x}_\alpha^{(2)} - \bar{\mathbf{x}}^{(2)} \right) + \mathbf{u}_\alpha, \quad (4.13)$$

where $\bar{\mathbf{x}}^{(2)} = (1/N) \sum_{\alpha=1}^N \mathbf{x}_\alpha^{(2)}$ and $\boldsymbol{\phi} = \mathbf{B}_1 + \mathbf{B}_2 \bar{\mathbf{x}}^{(2)}$ is a vector of parameters. Display (4.13) indicates that

$$\mathcal{E}\mathbf{Y} = \boldsymbol{\varepsilon}_N \boldsymbol{\phi}' + \left(\mathbf{X}_2 - \boldsymbol{\varepsilon}_N \bar{\mathbf{x}}^{(2)'} \right) \mathbf{B}_2'. \quad (4.14)$$

The first matrix on the right-hand side is of rank 1; the second matrix is of rank q_2 and its columns are orthogonal to $\boldsymbol{\varepsilon}_N$. We shall write this $N \times p$ matrix of rank q_2 as

$$\mathbf{N} = \begin{bmatrix} \boldsymbol{\nu}'_1 \\ \vdots \\ \boldsymbol{\nu}'_N \end{bmatrix} \quad (4.15)$$

with $\sum_{\alpha=1}^N \boldsymbol{\nu}_\alpha = \mathbf{0}$.

Let \mathbf{G} be a matrix having the distribution $W(\Sigma_u, M)$; then $\tilde{\mathbf{G}} = (1/M)\mathbf{G}$ is an estimator of Σ_u . Define

$$\mathbf{H} = \sum_{\alpha=1}^N (\mathbf{y}_\alpha - \bar{\mathbf{y}})(\mathbf{y}_\alpha - \bar{\mathbf{y}})', \quad \tilde{\mathbf{H}} = \frac{1}{N} \mathbf{H}. \quad (4.16)$$

Then $\bar{\mathbf{y}}$ is an estimator of $\boldsymbol{\phi}$, and \mathbf{H} and \mathbf{G} are to be used to estimate (4.15). Define $d_1 > \dots > d_p$ as the roots of (2.23), $\mathbf{w}_1, \dots, \mathbf{w}_p$ satisfying (2.25) and (2.26), and $\mathbf{W}_1, \mathbf{W}_2, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}, \mathbf{W}, \mathbf{Z}, \mathbf{Z}_1$, and \mathbf{Z}_2 by (2.27), (2.28), (2.29) and (2.30), with m replaced by q_2 . Then (4.15) of rank q_2 is estimated by

$$\widehat{\mathbf{N}} = \begin{bmatrix} \widehat{\boldsymbol{\nu}}'_1 \\ \vdots \\ \widehat{\boldsymbol{\nu}}'_N \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_1 - \bar{\mathbf{y}}' \\ \vdots \\ \mathbf{y}'_N - \bar{\mathbf{y}}' \end{bmatrix} \mathbf{W}_1 \mathbf{Z}'_1. \quad (4.17)$$

Note that \mathbf{W}_1 has q_2 columns and \mathbf{Z}_1 has q_2 columns corresponding to the q_2 largest characteristic roots of $\tilde{\mathbf{G}}^{-1} \tilde{\mathbf{H}}$. The estimator has the same form as (2.32).

If the indeterminacy in $(\mathbf{X}_2 - \varepsilon_N \bar{\mathbf{x}}^{(2)})\mathbf{B}'_2$ is resolved by requiring $\mathbf{B}'_2 = (\mathbf{B}'_2^*, \mathbf{I}_{q_2})$, then the estimators of the two factors are obtained by solving

$$(\widehat{\mathbf{N}}_1, \widehat{\mathbf{N}}_2) = \left[(\widehat{\mathbf{X}}_2 - \varepsilon_N \widehat{\mathbf{x}}^{(2)'}) \widehat{\mathbf{B}}_2^*, (\widehat{\mathbf{X}}_2 - \varepsilon_N \widehat{\mathbf{x}}^{(2)'}) \right], \quad (4.18)$$

where $\widehat{\mathbf{N}}_2$ has q_2 columns. Then

$$\widehat{\mathbf{B}}_2^* = (\widehat{\mathbf{N}}_2' \widehat{\mathbf{N}}_2)^{-1} \widehat{\mathbf{N}}_2' \widehat{\mathbf{N}}_1. \quad (4.19)$$

5. A Latent Variable Model : \mathbf{X} Unobserved and Stochastic.

Now we suppose the form and properties of the model in Section 3 hold except that \mathbf{X} is not observed, only \mathbf{Y} . Then the vectors $\mathbf{y}, \dots, \mathbf{y}_N$ are independent with mean $\mathcal{E}\mathbf{y}_\alpha = \boldsymbol{\mu}_y$ and covariance matrix

$$\text{cov}(\mathbf{y}_\alpha) = \mathbf{B}_2 \boldsymbol{\Sigma}_x \mathbf{B}'_2 + \boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_y, \quad (5.1)$$

where \mathbf{B}_2 is $p \times q_2 = p \times (q-1)$. We assume $\text{rank } \boldsymbol{\Sigma}_u = p$ and $\text{rank } \boldsymbol{\Sigma}_x = q_2 \leq p$.

As in the case of \mathbf{X} nonstochastic the model has the indeterminacy of multiplication of $\mathbf{x}_\alpha^{(2)}$ by an arbitrary nonsingular matrix \mathbf{A} (and hence $\boldsymbol{\Sigma}_x$ by \mathbf{A} on the left and \mathbf{A}' on the right) and \mathbf{B}_2 by \mathbf{A}^{-1} on the right; this indeterminacy can be eliminated by restricting \mathbf{B}_2 , for example, as in (4.3). There is a further indeterminacy due to the fact that $\text{cov}(\mathbf{y}_\alpha)$ is the sum of a positive definite matrix $\boldsymbol{\Sigma}_u$ and a positive semi-definite matrix of rank q_2 . If \mathbf{x} and \mathbf{y} are normal, $\mathbf{B}_2 \boldsymbol{\Sigma}_x \mathbf{B}'_2$ and $\boldsymbol{\Sigma}_u$ are unidentified. For identification we can (a) restrict $\boldsymbol{\Sigma}_u$ and q_2 or in some situations (b) obtain an independent estimator of $\boldsymbol{\Sigma}_u$. An example of (a) is to require $\boldsymbol{\Sigma}_u$ to be diagonal (components of \mathbf{u}_α to be uncorrelated) and q_2 small relative to p ; this model is used for factor analysis, which is discussed in Section 6.

The model can alternatively be characterized by linear restrictions on $\mathbf{B}_2(\mathbf{x}_\alpha^{(2)} - \boldsymbol{\mu}_x)$. If $\boldsymbol{\Gamma}$ is a $(p-q_2) \times p$ matrix such that $\boldsymbol{\Gamma} \mathbf{B}_2 = \mathbf{0}$, then $\boldsymbol{\Gamma} \mathbf{B}_2 \mathbf{x}_\alpha^{(2)} = \mathbf{0}, \alpha = 1, \dots, N$. These relations are called *linear structural relationships*. The distinction between functional and structural was made by Kendall and Stuart (1979).

Now suppose we have an independent estimator of $\boldsymbol{\Sigma}_u$. Specifically let \mathbf{G} have the distribution $W(\boldsymbol{\Sigma}_u, M)$ independent of \mathbf{H} defined in (4.16). Define $d_1 > \dots > d_p$ as the roots of (2.23) and $\mathbf{w}_1, \dots, \mathbf{w}_p$ as the vectors satisfying (2.25) and (2.26). Since \mathbf{H} estimates $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_u + \mathbf{B}_2 \boldsymbol{\Sigma}_x \mathbf{B}'_2$ and $\widetilde{\mathbf{G}} [= (1/M)\mathbf{G}]$ estimates $\boldsymbol{\Sigma}_u$, the difference $\widetilde{\mathbf{H}} - \widetilde{\mathbf{G}}$ estimates $\mathbf{B}_2 \boldsymbol{\Sigma}_x \mathbf{B}'_2$. However, this estimator is unsatisfactory because $\mathbf{B}_2 \boldsymbol{\Sigma}_x \mathbf{B}'_2$ is positive semidefinite, but $\widetilde{\mathbf{H}} - \widetilde{\mathbf{G}}$ is not necessarily. In fact, $\widetilde{\mathbf{H}} = \mathbf{Z} \mathbf{D} \mathbf{Z}'$, $\widetilde{\mathbf{G}} = \mathbf{Z} \mathbf{Z}'$, and

$$\widetilde{\mathbf{H}} - \widetilde{\mathbf{G}} = \mathbf{Z} \mathbf{D} \mathbf{Z}' - \mathbf{Z} \mathbf{Z}' = \mathbf{Z} (\mathbf{D} - \mathbf{I}_p) \mathbf{Z}', \quad (5.2)$$

16 INFERENCE IN LINEAR MODELS

where

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}, \quad \mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_p), \quad (5.3)$$

and $\mathbf{Z} = (\mathbf{W}')^{-1}$. The diagonal matrix $\mathbf{D} - \mathbf{I}_p$ is positive semidefinite only if $d_i \geq 1, i = 1, \dots, p$. With positive probability this is not the case.

To obtain a maximum likelihood estimator of $\mathbf{B}_2 \Sigma_x \mathbf{B}_2'$ that is positive semidefinite we define p^* as the number of $d_i > 1$; that is

$$d_p < \cdots < d_{p^*+1} \leq 1 < d_{p^*} < \cdots < d_1. \quad (5.4)$$

Let $q^* = \min(q_2, p^*)$; q_2 is the rank of $\mathbf{B}_2 \Sigma_x \mathbf{B}_2'$. Let

$$\mathbf{D}_1^* = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_{q^*} \end{bmatrix}, \quad \mathbf{D}_2^* = \begin{bmatrix} d_{q^*+1} & 0 & \cdots & 0 \\ 0 & d_{q^*+2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}, \quad (5.5)$$

$$\mathbf{W} = (\mathbf{W}_1^*, \mathbf{W}_2^*), \quad (5.6)$$

$$(\mathbf{W}')^{-1} = \mathbf{Z} = (\mathbf{Z}_1^*, \mathbf{Z}_2^*), \quad (5.7)$$

where \mathbf{W}_1^* and \mathbf{Z}_1^* have q^* columns. The maximum likelihood estimator of $\mathbf{B}_2 \Sigma_x \mathbf{B}_2'$ is

$$\hat{\mathbf{B}}_2 \hat{\Sigma}_x \hat{\mathbf{B}}_2' = \mathbf{Z}_1^* (\mathbf{D}_1^* - \mathbf{I}_{q^*}) \mathbf{Z}_1^{*\prime}. \quad (5.8)$$

The rank of this estimator is

$$\text{rank } \hat{\mathbf{B}}_2 \hat{\Sigma}_x \hat{\mathbf{B}}_2' = q^* = \min(q_2, p^*). \quad (5.9)$$

Note that this rank is random; it depends on \mathbf{H} and \mathbf{G} . If $\mathbf{B}' = (\mathbf{B}_2^{*\prime}, \mathbf{I}_{q_2})$, then $\hat{\Sigma}_x$ and $\hat{\mathbf{B}}_2^*$ can be found from $\hat{\mathbf{B}}_2 \hat{\Sigma}_x \hat{\mathbf{B}}_2'$.

Linear structural relationships. When the rank of $\mathbf{B}_2 \Sigma_x \mathbf{B}_2'$ is q_2 , there exists a $(p - q_2) \times p$ matrix Γ such that

$$\Gamma \mathbf{B}_2 \Sigma_x \mathbf{B}_2' = \mathbf{0}, \quad (5.10)$$

or

$$\Gamma \mathbf{B}_2 \Sigma_x \mathbf{B}_2' \Gamma' = \mathbf{0}. \quad (5.11)$$

Equivalently

$$\Gamma \mathbf{B}_2 \mathbf{x}^{(2)} = \Gamma \boldsymbol{\mu}_x \quad \text{with probability 1.} \quad (5.12)$$

An estimator of Γ is

$$\hat{\Gamma} = \mathbf{W}_2^{*\prime}, \quad (5.13)$$

which is $(p - q^*) \times p$. The number of estimated linear relations is random.

To test the null hypothesis

$$H : \text{rank } \mathbf{B}_2 \Sigma_x \mathbf{B}'_2 = q_0, \quad (5.14)$$

where q_0 is specified, the likelihood ratio criterion λ is given by

$$\lambda = \prod_{i=q_0+1}^{p^*} \frac{(M+N)^{(M+N)/2} d_i^{N/2}}{(M+N d_i)^{(M+N)/2}} \quad (5.15)$$

if that is, $p^* > q_0$, and $\lambda = 1$ if $p^* \leq q_0$. The hypothesis is rejected if $\lambda < \lambda_0$, where λ_0 is a suitable constant. The criterion leads to rejection if the $p - q_0$ smallest roots are large enough; in particular, some of them must be larger than 1.

An example of a situation having an independent estimator of Σ_u is the balanced one-way components of variance (MANOVA II). Suppose

$$\mathbf{y}_{\alpha j} = \boldsymbol{\mu}_y + \mathbf{v}_\alpha + \mathbf{u}_{\alpha j}, \quad j = 1, \dots, k, \quad \alpha = 1, \dots, n, \quad (5.16)$$

where $\mathbf{v}_\alpha \sim N(\mathbf{0}, \Sigma_v)$ independent of $\mathbf{u}_{\alpha j}$ and $\text{rank } \Sigma_v = q \leq p$. In (2.8) the nonstochastic $\boldsymbol{\mu}_\alpha$ has been replaced by the sum of $\boldsymbol{\mu}_y$ and the random factor \mathbf{v}_α . Then $\mathbf{y}_{\alpha i} \sim N(\boldsymbol{\mu}_y, \Sigma_v + \Sigma_u)$ and $\text{cov}(\mathbf{y}_{\alpha i}, \mathbf{y}_{\alpha j}) = \Sigma_v, i \neq j$. When we make the transformation (2.40), then

$$\mathbf{y}_{\alpha 1}^* = \sqrt{k} \bar{\mathbf{y}}_\alpha = \sqrt{k} \boldsymbol{\mu}_y + \sqrt{k} \mathbf{v}_\alpha + \sqrt{k} \bar{\mathbf{u}}_\alpha, \quad \alpha = 1, \dots, n, \quad (5.17)$$

has the distribution $N(\sqrt{k} \boldsymbol{\mu}_y, \Sigma_u + k \Sigma_v)$. Further $\mathbf{y}_{\alpha j}^*, j = 2, \dots, n$, has the distribution $N(\mathbf{0}, \Sigma_u)$. To summarize: $\sqrt{k} \bar{\mathbf{y}}_\alpha, \alpha = 1, \dots, n$, here has the distribution of $\mathbf{y}_\alpha, \alpha = 1, \dots, N$, in the first paragraph of this section with $\boldsymbol{\mu}$ replaced by $\sqrt{k} \boldsymbol{\mu}_y$ and $\mathbf{B}_2 \Sigma_x \mathbf{B}'_2$ replaced by $k \Sigma_v$. (Note $\mathbf{B}_2 \Sigma_x \mathbf{B}'_2$ is a way of writing a $p \times p$ positive semidefinite matrix of rank q_2 .)

Now \mathbf{H} and \mathbf{G} defined by (2.46) and (2.47) have Wishart distributions $W(\Sigma_u + k \Sigma_v, n - 1)$ and $W[\Sigma_u, n(k - 1)]$, respectively. The matrix $\widetilde{\mathbf{H}} = (1/n) \mathbf{H}$ estimates $\Sigma_u + k \Sigma_v$ and $\widetilde{\mathbf{G}} = \{1/[n(k - 1)]\} \mathbf{G}$ estimates Σ_u . Define \mathbf{D} , \mathbf{W} , and \mathbf{Z} by (2.23) to (2.29) and \mathbf{D}_1^* and \mathbf{Z}_1^* by (5.5) and (5.7) with $q^* = \min(n - 1, p^*)$. Then

$$k \widehat{\Sigma} = \mathbf{Z}_1^* (\mathbf{D}_1^* - \mathbf{I}_{q^*}) \mathbf{Z}_1^{*\prime}. \quad (5.18)$$

Klotz and Putter (1969) derived this estimator when $q_2 = p$. Anderson, Anderson, and Olkin (1986) obtained the general result. See, also, Schott and Saw (1984) and Amemiya (1985).

6. Factor Analysis. The customary factor analysis model is that of Section 4 or 5 with Σ_u diagonal. We write the model in traditional notation as

$$\mathbf{y}_\alpha = \Lambda \mathbf{f}_\alpha + \boldsymbol{\mu} + \mathbf{u}_\alpha, \quad \alpha = 1, \dots, N, \quad (6.1)$$

where Λ is $p \times q$ ($q < p$). The matrix Λ consists of factor loadings, and the vector \mathbf{f}_α consists of factor scores [Thurstone (1947), for example]. We have

$$\mathcal{E}\mathbf{u}_\alpha = \mathbf{0}, \quad \mathcal{E}\mathbf{u}_\alpha \mathbf{u}'_\alpha = \Psi \text{ diagonal.} \quad (6.2)$$

When the factor scores are random, we assume

$$\mathcal{E}\mathbf{f}_\alpha = \mathbf{0}, \quad \mathcal{E}\mathbf{f}_\alpha \mathbf{f}'_\alpha = \Phi. \quad (6.3)$$

Then

$$\text{cov}(\mathbf{y}_\alpha) = \Sigma_y = \Lambda \Phi \Lambda' + \Psi. \quad (6.4)$$

The rank of $\Lambda \Phi \Lambda'$ is q . Note that the off-diagonal elements of Σ_y are functions of Λ and Φ ; they do not involve Ψ .

For identification we may require

$$\Phi = \mathbf{I}. \quad (6.5)$$

This model is known as the orthogonal factor analysis model. Then $\Lambda \Phi \Lambda' = \Lambda \Lambda'$, which leaves the indeterminacy of multiplication of Λ on the right by an arbitrary orthogonal matrix. To eliminate this indeterminacy we may require

$$\Lambda' \Psi^{-1} \Lambda = \Delta \text{ diagonal.} \quad (6.6)$$

Here the diagonal entries δ_i are the roots of $|\Lambda \Lambda' - \delta \Psi| = 0$.

If

$$\frac{1}{2}[(p-q)^2 - (p+q)] \quad (6.7)$$

is positive and $\delta_1 > \dots > \delta_p$, the parameters will usually be uniquely determined by $\boldsymbol{\mu}$ and Σ_y ; that is, the model is identified. A sufficient condition for identification up to multiplication of Λ on the right by an orthogonal matrix is that if any row of Λ is deleted there remain two disjoint submatrices of rank q . [See Anderson and Rubin (1956).]

Let

$$\mathbf{C} = \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{y}_\alpha - \bar{\mathbf{y}})(\mathbf{y}_\alpha - \bar{\mathbf{y}})', \quad (6.8)$$

The maximum likelihood estimators of Λ , Δ , and Ψ are a solution to

$$\text{diagonal } \widehat{\Psi} = \text{diagonal } (\mathbf{C} - \widehat{\Lambda} \widehat{\Lambda}'), \quad (6.9)$$

$$(\mathbf{C} - \widehat{\Psi}) \widehat{\Psi}^{-1} \widehat{\Lambda} = \widehat{\Lambda} \widehat{\Delta}, \quad (6.10)$$

$$\widehat{\Lambda}' \widehat{\Psi}^{-1} \widehat{\Lambda} = \widehat{\Delta} \text{ diagonal.} \quad (6.11)$$

The diagonal elements of $\widehat{\Delta}$ are q largest roots of

$$|(\mathbf{C} - \widehat{\Psi})\widehat{\Psi}^{-1} - \widehat{\delta}\mathbf{I}| = 0 \quad (6.12)$$

or

$$|\mathbf{C} - \widehat{\Psi} - \widehat{\delta}\widehat{\Psi}| = 0. \quad (6.13)$$

The columns of $\widehat{\Lambda}$ are the characteristic vectors of $(\mathbf{C} - \widehat{\Psi})\widehat{\Psi}^{-1}$, and the i -th column of $\widehat{\Psi}^{-1}\widehat{\Lambda}$ satisfies

$$(\mathbf{C} - \widehat{\Psi})\mathbf{x} = \widehat{\delta}_i \widehat{\Psi} \mathbf{x}; \quad (6.14)$$

the normalization is

$$\mathbf{x}' \widehat{\Psi} \mathbf{x} = \widehat{\delta}_i \quad (6.15)$$

[in keeping with (6.11)]. Note that $\mathbf{C} - \widehat{\Psi}$, $\widehat{\Psi}$, and $\widehat{\Psi}^{-1}\widehat{\Lambda}$ replace \mathbf{H} , \mathbf{G} and \mathbf{W}_1 in the former analysis, but here $\widehat{\Psi}$ is not observed; it is an estimator satisfying (6.9), (6.10), and (6.11). These likelihood equations can be solved iteratively. [If Λ is identified by the requirement that $\Lambda' = (\Lambda^{*'}, \mathbf{I}_q)$, and Φ is unrestricted, the maximum likelihood estimators of Λ^* and Σ_f are $\widehat{\Lambda}_2^{-1}\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2\widehat{\Lambda}_2'$, where $\widehat{\Lambda}' = (\widehat{\Lambda}_1', \widehat{\Lambda}_2')$ and $\widehat{\Lambda}_2$ has q rows.] Lawley (1940), (1941) obtained these estimators.

An investigator may ask about how many factors there are, that is, the dimensionality of f_α or the rank of $\Lambda\Phi\Lambda$. Consider testing the hypothesis

$$H : \text{rank } \Lambda\Phi\Lambda' = q_0 \quad (6.16)$$

given (6.7) against the alternative that Σ_y is positive definite (otherwise arbitrary). The likelihood ratio criterion λ is given by

$$\lambda^{2/N} = \prod_{i=q_0+1}^p (1 + \widehat{\delta}_i), \quad (6.17)$$

where $\widehat{\delta}_{q_0+1}, \dots, \widehat{\delta}_p$ are the $p - q_0$ smallest roots of (6.13). The hypothesis is rejected if $\lambda < \lambda_0$ for λ_0 a suitable constant.

REFERENCES

- AMEMIYA, YASUO. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite? *Amer. Statist.* **39**, 112–117.
- ANDERSON, BLAIR M., T. W. ANDERSON, and INGRAM OLKIN. (1986). Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance, *Ann. Statist.* **14**, 405–417.

20 INFERENCE IN LINEAR MODELS

- ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Ann. Math. Statist.* **22**, 327–351. (Correction, *Ann. Statist.* **8** (1980), 923–924.)
- ANDERSON, T. W. (1984a). Estimating linear statistical relationships, *Ann. Statist.* **12**, 1–45.
- ANDERSON, T. W. (1984b). *An Introduction to Multivariate Statistical Analysis*, Second Edition. John Wiley & Sons, Inc., New York.
- ANDERSON, T. W. and RUBIN, HERMAN. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations, *Ann. Math. Statist.* **20**, 46–63.
- ANDERSON, T. W., and RUBIN, HERMAN. (1956). Statistical inference in factor analysis, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5 (Jerzy Neyman, ed.). University of California Press, 111–150.
- FISHER, R. A. (1938). The statistical utilization of multiple measurements, *Ann. Eugen.* **8**, 376–386.
- KENDALL, MAURICE G. and STUART, ALAN. (1979). *The Advanced Theory of Statistics*, Vol. 2, Fourth Edition. Griffin, London.
- KLOTZ, JEROME and PUTTER, JOSEPH. (1969). Maximum likelihood estimation of the multivariate covariance components for the balanced one-way layout, *Ann. Math. Statist.* **40**, 1100–1105.
- LAWLEY, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood, *Proc. Roy. Soc. Edinburgh, Sect. A* **60**, 64–82.
- LAWLEY, D. N. (1941). Further investigations in factor estimation, *Proc. Roy. Soc. Edinburgh, Sect. A* **61**, 176–185.
- SCHOTT, J. R., and J. G. SAW. (1984). A multivariate one-way classification model with random effects, *J. Multivariate Anal.* **15**, 1–12.
- THURSTONE, L. L. (1947). *Multiple-Factor Analysis*. University of Chicago Press.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CA 94305
USA