British Journal of Mathematical and Statistical Psychology (1989) 42, 215–232 Printed in Great Britain 215

Balanced versus unbalanced designs for linear structural relations in two-level data

Roderick P. McDonald†
School of Education, Macquarie University, NSW 2109, Australia

Harvey Goldstein
Institute of Education, University of London, UK

A general two-level model for multivariate data is described and illustrated by specializations to a common factor model and a path model with latent variables. The problem of estimation is treated, and in the special case of a balanced sampling design, a likelihood-based discrepancy function and a test of goodness of fit are obtained in terms of some simple sufficient statistics.

1. Introduction

There has recently been an upsurge of research activity concerned with defining and fitting suitable statistical models for multilevel data (see Goldstein, 1987). Multilevel data arise from a nested or hierarchical sampling scheme, or the sampling of a hierarchically structured population, the paradigm example of which might be the drawing of random samples of students from within random samples of classes from within random samples of schools. In this case, sampling takes place at three levels—students, classes and schools—with sampling units of each level nested within a unit of the next level.

For the analysis of a simple random sample and the simultaneous analysis of random samples from a number of distinct populations, theory providing for the definition, fitting and testing of linear models, including regression models with fixed regressors and general models for linear structural relations with latent variables (path analysis and factor analysis), has reached the stage of general practical application (see, for example, Jöreskog & Sörbom, 1979; McArdle & McDonald, 1984). The extension of such methods to multilevel data presents both theoretical problems and practical computational difficulties.

It is well known that methods for the analysis of multilevel data based on aggregation (as when we combine student measures to yield class means and perform

†Requests for reprints.

a means-on-means regression) or disaggregation (as when we assign a class-based measure to each student in the class) are inadequate and at the same time they create severe problems of interpretation (see, for example, Aitkin & Longford, 1986; Hannan, 1971; Robinson, 1950; Tate & Wongbundhit, 1983). It is therefore generally understood that statistical models are needed for such data to take account of the sampling scheme, or population structure, directly and appropriately.

A number of writers have developed multilevel models in which regression coefficients at one level are treated as random variables and regressed on variables at a higher level. Harville (1977) applies maximum likelihood and restricted maximum likelihood to normal mixed models. For a two-level model with fixed regressors, Mason, Wong & Entwisle (1984) obtain restricted maximum likelihood estimates by the EM algorithm. Longford (in press) shows how to obtain maximum likelihood estimates for a general multilevel mixed-effects model by the method of scoring. Aitkin & Longford (1986) illustrate the procedure. De Leeuw & Kreft (1986) show how to fit the Mason *et al.* model by least squares, weighted least squares and maximum likelihood, using the method of scoring. Goldstein (1986) describes a general multilevel mixed model and shows how to fit it by an iterated generalized least squares algorithm.

While each of the models in the contributions just cited can be described as the regression of a univariate dependent variable on one or more fixed regressors, Goldstein (1986) points out that an h-level model ($h \geqq 3$) of this kind can be applied to give an $(h-1)$-level model for a q-variate dependent variable, possibly with missing data, by treating the q variates measured within a unit as a level of observation. De Leeuw (1985) shows that theory for a univariate dependent variable with fixed regressors can in principle be applied with little modification to fit a multilevel recursive path model with random exogenous and endogenous variables provided that the random path coefficients of distinct variables are mutually independent.

Goldstein & McDonald (1988) develop the model in Goldstein (1986) into a more general one which includes as special cases all of the models cited, and contains models for multilevel structural relations, possibly with latent variables and with data from any level missing at random. The model covers data with a fully nested, hierarchical structure, and more general variance–covariance component models with cross-classification at any level.

The present paper gives a self-contained account, in some detail, of the properties of a two-level model for linear structural relations. The model considered is a special case of the one treated by Goldstein & McDonald (1988) and it may indeed be fitted by a three-level model as described by Goldstein (1986). The object here, however, is to see how far the well known results on estimating and testing goodness of fit for the single level case generalize to a case with more than one level of sampling. In Section 2, a general two-level model is described and illustrated by specializations to a common factor model and a path model with latent variables. Section 3 treats problems of estimation, with particular attention to the case of a balanced sampling design, with the same number of level-one units in each level-two unit. The balanced design yields a suitable discrepancy function, convenient sufficient statistics, and an overall test of fit. The more general case, which is treated by Goldstein & McDonald (1988), does not appear to yield comparable results.

2. The general two-level model

We suppose we have measures y_{kji} on $j=1,\dots,q$ variables from $i=1,\dots,n_k$ level-one units (say, individual students) randomly sampled from $k=1,\dots,m$ level-two units (say, classrooms) also randomly sampled. We may also suppose, in general, that we have measures x_{kl} on $l=1,\dots,p$ variables characterizing the level-two units (classrooms). We consider cases where there are no missing data.

We write the entire data-set in the form of a single vector

$$\mathbf{z}'=[\mathbf{z}'_1,\dots,\mathbf{z}'_m],$$

of

$$N=pm+q\sum_{k=1}^{m}n_k \text{ components,}$$

where

$$\mathbf{z}'_k=[\mathbf{x}'_k\,\mathbf{y}'_{k1}\dots\mathbf{y}'_{kn_k}],$$

with

$$\mathbf{x}'_k=[x_{k1},\dots,x_{kp}],$$

$$\mathbf{y}'_{ki}=[y_{k1i},\dots,y_{kqi}],$$

and note that $\mathbf{z}_k$ has $p+qn_k$ components.

We write

$$\boldsymbol{\mu}'_z=\mathscr{E}\{\mathbf{z}'\}=[\boldsymbol{\mu}'_1,\dots,\boldsymbol{\mu}'_m], \tag{1}$$

where

$$\boldsymbol{\mu}'_k=\mathscr{E}\{\mathbf{z}'_k\}=[\boldsymbol{\mu}'_x,\mathbf{1}'_{n_k}\otimes\boldsymbol{\mu}'_y] \tag{2}$$

and

$$\mathscr{E}\{\mathbf{x}_k\}=\boldsymbol{\mu}_x \quad \forall k, \tag{3}$$

$$\mathscr{E}\{\mathbf{y}_{ki}\}=\boldsymbol{\mu}_y \quad \forall k,i. \tag{4}$$

Writing further

$$\mathbf{y}_{ki}=\mathbf{y}_{2k}+\mathbf{y}_{1ki} \tag{5}$$

(where the subscripts 1 and 2 index the level of sampling) we assume that

$$\operatorname{cov}\{\mathbf{y}_{2k},\mathbf{y}_{1ki}\}=\mathbf{0} \quad \forall k,i \tag{6a}$$

$$\operatorname{cov}\{\mathbf{x}_k,\mathbf{y}_{1ki}\}=\mathbf{0} \quad \forall k,i. \tag{6b}$$

We define

$$\mathbf{V}_k = \text{cov}\{\mathbf{z}_k\}, (p+qn_k)\times(p+qn_k), \tag{7a}$$

$$\mathbf{V} = \text{cov}\{\mathbf{z}\}, N\times N, \tag{7b}$$

$$\mathbf{\Sigma}_{xx} = \text{cov}\{\mathbf{x}_k\}, p\times p, \quad \forall k \tag{7c}$$

$$\mathbf{\Sigma}_{yy} = \text{cov}\{\mathbf{y}_{ki}\}, q\times q, \quad \forall k, i \tag{7d}$$

$$\mathbf{\Sigma}_{xy} = \text{cov}\{\mathbf{x}_k, \mathbf{y}_{ki}\}, p\times q, \quad \forall k, i \tag{7e}$$

$$\mathbf{\Sigma}_2 = \text{cov}\{\mathbf{y}_{2k}\}, q\times q, \quad \forall k \tag{7f}$$

$$\mathbf{\Sigma}_1 = \text{cov}\{\mathbf{y}_{1ki}\}, q\times q, \quad \forall k, i. \tag{7g}$$

Noting that

$$\text{cov}\{\mathbf{y}_{ki}, \mathbf{y}_{ki'}\} = \text{cov}\{\mathbf{y}_{2k}\} = \mathbf{\Sigma}_2 \quad \forall k, i \neq i' \tag{8a}$$

and

$$\text{cov}\{\mathbf{x}_k, \mathbf{y}_{ki}\} = \text{cov}\{\mathbf{x}_k, \mathbf{y}_{2k}\} = \mathbf{\Sigma}_{xy} \tag{8b}$$

and hence by (6a),

$$\mathbf{\Sigma}_{yy} = \mathbf{\Sigma}_2 + \mathbf{\Sigma}_1, \tag{9}$$

we may think of $\mathbf{\Sigma}_1$ as the within-level-two-units, between-level-one-units, covariance matrix of $\mathbf{y}_{ki}$, and $\mathbf{\Sigma}_2$ as its between-level-two-units covariance matrix.

Structural models for such a data-set may be defined by restricting the elements of $\mathbf{\Sigma}_{xx}$, $\mathbf{\Sigma}_{xy}$, $\mathbf{\Sigma}_2$, $\mathbf{\Sigma}_1$ and, possibly, $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$ to be functions of some basic set of parameters $\boldsymbol{\pi}' = [\pi_1, \ldots, \pi_t]$. Thus, $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$ could be expressed in terms of a design matrix (known values of fixed regressors) with fixed regression coefficients to be estimated, and the covariance matrices could be structured as implied by a factor model or path model.

For example, we may define a general two-level common factor model by writing, say,

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_{2k} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_x \\ \mathbf{\Lambda}_2 \end{bmatrix} \mathbf{v}_2 + \begin{bmatrix} \mathbf{e}_{xk} \\ \mathbf{e}_{2k} \end{bmatrix} \tag{10a}$$

and

$$\mathbf{y}_{1ki} = \mathbf{\Lambda}_1 \mathbf{v}_1 + \mathbf{e}_{1ki} \tag{10b}$$

with

$$\mathbf{\Phi}_2 = \mathrm{cov}\{\mathbf{v}_2\} \tag{11a}$$

$$\mathbf{\Phi}_1 = \mathrm{cov}\{\mathbf{v}_1\}, \tag{11b}$$

and diagonal matrices

$$\mathbf{\Psi}_x = \mathrm{cov}\{\mathbf{e}_{xk}\} \tag{12a}$$

$$\mathbf{\Psi}_2 = \mathrm{cov}\{\mathbf{e}_{2k}\} \tag{12b}$$

$$\mathbf{\Psi}_1 = \mathrm{cov}\{\mathbf{e}_{1ki}\}, \quad \forall k, i. \tag{12c}$$

The matrices $\mathbf{\Lambda}_x$, $p \times s$, say, and $\mathbf{\Lambda}_2$, $q \times s$, are patterned factor loading matrices respectively relating the level-two variables and the level-one variables to the level-two factors $\mathbf{v}_2$, while $\mathbf{\Lambda}_1$, $q \times t$, say, is a patterned factor loading matrix relating the level-one variables to the level-one factors $\mathbf{v}_1$. The covariance matrices $\mathbf{\Phi}_2$, $s \times s$, of the level-two factors, and $\mathbf{\Phi}_1$, $t \times t$, of the level-one factors, are also possibly patterned, including possibly diagonal for an orthogonal model. The diagonal matrices $\mathbf{\Psi}_x$, $\mathbf{\Psi}_2$, $\mathbf{\Psi}_1$ are the corresponding residual variance (uniqueness) matrices.

We then have

$$\mathbf{\Sigma}_{xx} = \mathbf{\Lambda}_x \mathbf{\Phi}_2 \mathbf{\Lambda}_x' + \mathbf{\Psi}_x \tag{13a}$$

$$\mathbf{\Sigma}_{xy} = \mathbf{\Lambda}_x \mathbf{\Phi}_2 \mathbf{\Lambda}_2' \tag{13b}$$

$$\mathbf{\Sigma}_2 = \mathbf{\Lambda}_2 \mathbf{\Phi}_2 \mathbf{\Lambda}_2' + \mathbf{\Psi}_2 \tag{13c}$$

$$\mathbf{\Sigma}_1 = \mathbf{\Lambda}_1 \mathbf{\Phi}_1 \mathbf{\Lambda}_1' + \mathbf{\Psi}_1, \tag{13d}$$

hence

$$\mathbf{\Sigma}_{yy} = [\mathbf{\Lambda}_2 \vdots \mathbf{\Lambda}_1] \begin{bmatrix} \mathbf{\Phi}_2 & \\ & \mathbf{\Phi}_1 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_2' \\ \mathbf{\Lambda}_1' \end{bmatrix} + \mathbf{\Psi}_2 + \mathbf{\Psi}_1. \tag{14}$$

Identifiable, restrictive models can be obtained by appropriate choices of pattern in the three factor loading and two factor covariance matrices of the model. In most applications we would expect to set $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_1$, thus supposing that the level-one observed variables imply the same definition ('interpretation') of the factors at both levels, but we would still wish to estimate distinct factor covariance matrices ϕ_2, ϕ_1, corresponding to their between-level-two and between-level-one-within-level-two variability. Even in the case where there are no observed level-two variables (i.e. $p = 0$), so that the model reduces to (14), it would still be generally necessary to fit both a level-two uniqueness matrix ψ_2 and a level-one uniqueness matrix ψ_1, as well as level-two and level-one factor covariance matrices.

In a similar way, we may develop a two-level version of any model for linear structural relations—for example, the LISREL model of Jöreskog & Sörbom (1979), the COSAN model of McDonald (1978, 1980) or McArdle's RAM model (McArdle & McDonald, 1984)—by writing the model as required for $\mathbf{x}_k$, $\mathbf{y}_{2k}$ and $\mathbf{y}_{1ki}$. The RAM model will serve as an illustration. For single-level data it can be defined by

$$\mathbf{v} = \mathbf{F}(\mathbf{A}\mathbf{v}^* + \mathbf{u}) \tag{15}$$

where $\mathbf{v}$ is a $t \times 1$ vector of observed variables, $\mathbf{v}^*$ is a $s \times 1$ vector containing the components of $\mathbf{v}$ and all additional latent variables in the model, $\mathbf{F}$ is a matrix of unities and zeros such that

$$\mathbf{v} = \mathbf{F}\mathbf{v}^*, \tag{16a}$$

and

$$\mathbf{v}^* = \mathbf{A}\mathbf{v}^* + \mathbf{u}, \tag{16b}$$

where $\mathbf{A}$, $s \times s$, is patterned such that a_{lm} is the path coefficient from v_m to v_l, and, with

$$\mathbf{S} = \mathrm{cov}\{\mathbf{u}\}, \tag{16c}$$

s_{lm} is the residual covariance corresponding to a non-directed path between v_l and v_m. Then the covariance structure of $\mathbf{v}$ is given by

$$\mathrm{cov}\{\mathbf{v}\} = \mathbf{F}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{S}(\mathbf{I} - \mathbf{A}')^{-1}\mathbf{F}' \tag{17}$$

(see McArdle & McDonald, 1984).

The two-level counterpart can easily be recognized to be given by

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_{2k} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_x & \\ & \mathbf{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_k^* \\ \mathbf{y}_{2k}^* \end{bmatrix} = \begin{bmatrix} \mathbf{F}_x & \\ & \mathbf{F}_2 \end{bmatrix} \left\{ \mathbf{A} \begin{bmatrix} \mathbf{x}_k^* \\ \mathbf{y}_{2k}^* \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{xk} \\ \mathbf{u}_{2k} \end{bmatrix} \right\}, \tag{18}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{xx} & \mathbf{A}_{x2} \\ \mathbf{A}_{2x} & \mathbf{A}_{22} \end{bmatrix} \tag{19}$$

together with

$$\mathbf{y}_{1ki} = \mathbf{F}_y \mathbf{y}_{1ki}^* = \mathbf{F}_y[\mathbf{A}_1 \mathbf{y}_{1ki}^* + \mathbf{u}_{1ki}], \tag{20}$$

where $\mathbf{F}_x$, $\mathbf{F}_y$, and $\mathbf{A}$, $\mathbf{A}_1$ are obvious counterparts of $\mathbf{F}$, $\mathbf{A}$ in the single-level model. The four submatrices of $\mathbf{A}$, namely $\mathbf{A}_{xx}$, $\mathbf{A}_{x2}$, $\mathbf{A}_{2x}$, $\mathbf{A}_{22}$ all represent level-two paths. These are, respectively, paths between level-two variables, paths from level-one to level-two, paths from level-two to level-one, and paths between level-one variables. In addition, we have level-one paths between level-one variables represented in $\mathbf{A}_1$. If the model is recursive at each level, there will exist a permutation of $[\mathbf{x}_k', \mathbf{y}_{ki}']$ such that $\mathbf{A}$ is lower triangular, and a permutation of $\mathbf{y}_{ki}$ such that $\mathbf{A}_1$ is lower triangular. It need not be that these two permutations are the same in respect of $\mathbf{y}_{ki}$. We can easily entertain the possibility that a causal path reverses its sign, or, indeed, its direction, between level one and level two. For example, we might conjecture that self-esteem partly determines and increases academic achievement in terms of between-individual-within-class variability while academic achievement partly determines and decreases self-esteem in terms of between-class variability by a frame-of-reference effect (see Marsh, 1984, for evidence of relations of this type.)

The structures for the matrices of the model given by (18) and (20) are, then,

$$\boldsymbol{\Sigma}_{xx} = [\mathbf{F}_x\ \mathbf{0}](\mathbf{I}-\mathbf{A})^{-1}\mathbf{S}_2(\mathbf{I}-\mathbf{A}')^{-1}\begin{bmatrix}\mathbf{F}'_x\\ \mathbf{0}'\end{bmatrix} \tag{21a}$$

$$\boldsymbol{\Sigma}_{xy} = [\mathbf{F}_x\ \mathbf{0}](\mathbf{I}-\mathbf{A})^{-1}\mathbf{S}_2(\mathbf{I}-\mathbf{A}')^{-1}\begin{bmatrix}\mathbf{0}'\\ \mathbf{F}'_2\end{bmatrix} \tag{21b}$$

$$\boldsymbol{\Sigma}_2 = [\mathbf{0}\ \mathbf{F}_y](\mathbf{I}-\mathbf{A})^{-1}\mathbf{S}_2(\mathbf{I}-\mathbf{A}')^{-1}\begin{bmatrix}\mathbf{0}'\\ \mathbf{F}'_2\end{bmatrix} \tag{21c}$$

$$\boldsymbol{\Sigma}_1 = \mathbf{F}_y(\mathbf{I}-\mathbf{A}_1)^{-1}\mathbf{S}_1(\mathbf{I}-\mathbf{A}'_1)\mathbf{F}'_2, \tag{21d}$$

where

$$\mathbf{S}_2 = \operatorname{cov}\left\{\begin{bmatrix}\mathbf{u}_{xk}\\ \mathbf{u}_{2k}\end{bmatrix}\right\} \tag{22a}$$

$$\mathbf{S}_1 = \operatorname{cov}\{\mathbf{u}_{1ki}\}. \tag{22b}$$

For the remainder of this paper we consider, as above, the vectors $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$ and matrices $\boldsymbol{\Sigma}_{xx}$, $\boldsymbol{\Sigma}_{xy}$, $\boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}_1$ to be functions of a vector of parameters $\boldsymbol{\pi}$, where, it will be understood, $\boldsymbol{\pi}$ typically contains the quantities to be estimated in the matrices of factor loadings, path coefficients, and the like, considered in these two examples.

3. Estimation and testing fit

From the definitions and assumptions (1) through (9) given above, it follows that

$$\mathbf{V}_k = \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \vdots & \mathbf{1}'_{n_k}\otimes\boldsymbol{\Sigma}_{xy}\\ \mathbf{1}_{n_k}\otimes\boldsymbol{\Sigma}_{yx} & \vdots & \mathbf{I}_{n_k}\otimes\boldsymbol{\Sigma}_1 + \mathbf{1}_{n_k}\mathbf{1}'_{n_k}\otimes\boldsymbol{\Sigma}_2\end{bmatrix} \tag{23}$$

and

$$\mathbf{V} = \bigoplus_{k=1}^{m} \mathbf{V}_k, \tag{24}$$

where $\otimes$ represents the direct (Kronecker) product and $\oplus$ the direct sum of matrices. It is convenient here and in the sequel to write $\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}'_{xy}$ and the like.

By well-known theory, if $\mathbf{z}_k$ has a normal density function, the maximum likelihood estimate of $\boldsymbol{\pi}$ is a solution of

$$\frac{\partial l}{\partial \boldsymbol{\pi}} = \mathbf{0}, \tag{25}$$

where

$$l = \log|\mathbf{V}| + (\mathbf{z} - \boldsymbol{\mu}_z)'\mathbf{V}^{-1}(\mathbf{z} - \boldsymbol{\mu}_z) \tag{26}$$

differs by a constant from twice the negative of the log-likelihood. Further,

$$l = \sum_{k=1}^{m} l_k, \tag{27}$$

where

$$l_k = \log|\mathbf{V}_k| + (\mathbf{z}_k - \boldsymbol{\mu}_k)'\mathbf{V}_k^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k). \tag{28}$$

Since $\mathbf{V}_k$ is of order $p + qn_k$, which may be extremely large in applications, the problem is to reduce the terms in l_k to a practical form for computation.

We define

$$\boldsymbol{\Sigma}_{2\cdot x} = \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}, \tag{29}$$

$$\boldsymbol{\Sigma}_k = n_k\boldsymbol{\Sigma}_{2\cdot x} + \boldsymbol{\Sigma}_1, \tag{30}$$

$$\bar{\mathbf{y}}_k = n_k^{-1} \sum_{i=1}^{n_k} \mathbf{y}_{ki}, \tag{31a}$$

$$\bar{\bar{\mathbf{y}}} = m^{-1} \sum_{k=1}^{m} \bar{\mathbf{y}}_k, \tag{31b}$$

$$\bar{\mathbf{x}} = m^{-1} \sum_{k=1}^{m} \mathbf{x}_k. \tag{31c}$$

By Appendix A, the kth term of the function of likelihood (28) may be written as

$$\begin{aligned} l_k = {} & \log|\boldsymbol{\Sigma}_{xx}| + (n_k - 1)\log|\boldsymbol{\Sigma}_1| + \log|\boldsymbol{\Sigma}_k| \\ & + \mathrm{Tr}\,\{[\boldsymbol{\Sigma}_{xx}^{-1} + n_k\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}](\mathbf{x}_k - \boldsymbol{\mu}_x)(\mathbf{x}_k - \boldsymbol{\mu}_x)'\} \\ & - 2n_k\,\mathrm{Tr}\,\{\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}(\bar{\mathbf{y}}_k - \boldsymbol{\mu}_y)(\mathbf{x} - \boldsymbol{\mu}_x)'\} \\ & + \mathrm{Tr}\left\{\boldsymbol{\Sigma}_1^{-1} \sum_{i=1}^{n_k} (\mathbf{y}_{ki} - \boldsymbol{\mu}_y)(\mathbf{y}_{ki} - \boldsymbol{\mu}_y)'\right\} \end{aligned}$$

$$-n_k \operatorname{Tr}\{[\Sigma_1^{-1}-\Sigma_k^{-1}](\bar{\mathbf{y}}_k-\mu_y)(\bar{\mathbf{y}}_k-\mu_y)'\}. \tag{32}$$

In this form, the order of the largest matrix requiring inversion or computation of its determinant is the greater of p and q. In a structural model in which μ_x, μ_y, Σ_{xx}, Σ_{xy}, Σ_1, Σ_2 are prescribed functions of a vector of parameters π, it is therefore computationally feasible, even when n_k is very large for some or all k, to obtain an ML estimate of the parameters by minimizing $\sum_{k=1}^{m} l_k$ with respect to π, using one of a number of available algorithms for minimizing a function.

In the special case of a balanced sampling design, with $n_k=n\,\forall k$ it is possible to express l as a function in a convenient set of sufficient statistics. For, in such cases, writing

$$\Sigma_k=\Sigma=n\Sigma_{2\cdot x}+\Sigma_1 \tag{33}$$

yields, from (32),

$$\begin{aligned} l=&\,m\{\log|\Sigma_{xx}|+(n-1)\log|\Sigma_1|+\log|\Sigma|\}\\ &+m\operatorname{Tr}\{[\Sigma_{xx}^{-1}+n\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}][\mathbf{S}_{xx}+(\bar{\mathbf{x}}-\mu_x)(\bar{\mathbf{x}}-\mu_x)']\}\\ &-2mn\operatorname{Tr}\{\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma^{-1}[\mathbf{S}_{yx}+(\bar{\bar{\mathbf{y}}}-\mu_y)(\bar{\mathbf{x}}-\mu_x)']\}\\ &+mn\operatorname{Tr}\{\Sigma_1^{-1}[\mathbf{S}_t+(\bar{\bar{\mathbf{y}}}-\mu_y)(\bar{\bar{\mathbf{y}}}-\mu_y)']\}\\ &-mn\operatorname{Tr}\{[\Sigma_1^{-1}-\Sigma^{-1}][\mathbf{S}_b+(\bar{\bar{\mathbf{y}}}-\mu_y)(\bar{\bar{\mathbf{y}}}-\mu_y)']\}, \end{aligned} \tag{34}$$

where

$$\mathbf{S}_{xx}=m^{-1}\sum_{k=1}^{m}(\mathbf{x}_k-\bar{\mathbf{x}})(\mathbf{x}_k-\bar{\mathbf{x}})' \tag{35a}$$

$$\mathbf{S}_{yx}=m^{-1}\sum_{k=1}^{m}(\bar{\mathbf{y}}_k-\bar{\bar{\mathbf{y}}})(\bar{\mathbf{x}}_k-\bar{\mathbf{x}})' \tag{35b}$$

$$\mathbf{S}_t=(mn)^{-1}\sum_{k=1}^{m}\sum_{i=1}^{n}(\mathbf{y}_{ki}-\bar{\bar{\mathbf{y}}})(\mathbf{y}_{ki}-\bar{\bar{\mathbf{y}}})' \tag{35c}$$

$$\mathbf{S}_b=m^{-1}\sum_{k=1}^{m}(\bar{\mathbf{y}}_k-\bar{\bar{\mathbf{y}}})(\bar{\mathbf{y}}_k-\bar{\bar{\mathbf{y}}})' \tag{35d}$$

$$\mathbf{S}_w=(mn)^{-1}\sum_{k=1}^{m}\sum_{i=1}^{n}(\mathbf{y}_{ki}-\bar{\mathbf{y}}_k)(\mathbf{y}_{ki}-\bar{\mathbf{y}}_k)' \tag{35e}$$

whence

$$\mathbf{S}_t=\mathbf{S}_b+S_w. \tag{36}$$

Thus clearly in the balanced case the statistics $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$, $\mathbf{S}_{xx}$, $\mathbf{S}_{xy}$, $\mathbf{S}_b$, $\mathbf{S}_w$ are minimal sufficient.

From the derivatives of l_k in (32) with respect to $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_{xx}$, $\boldsymbol{\Sigma}_{xy}$, $\boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}_1$, we may obtain the derivatives with respect to $\boldsymbol{\pi}$ of which these are functions in any specified restrictive model by use of the matrix chain rule and rearrangement rules (McDonald & Swaminathan, 1973). The former derivatives are given in Appendix B. In the general case the resulting likelihood equations (B1)–(B2) and (B5)–(B13) cannot be solved in closed form to estimate these matrices as unrestricted sets of parameters.

On the other hand, in the special case of a balanced sampling design, the resulting likelihood equations [(B3)–(B4) and (B14)–(B17)], are, by inspection, jointly satisfied by $\boldsymbol{\mu}_x = \bar{\mathbf{x}}$, $\boldsymbol{\mu}_y = \bar{\mathbf{y}}$ and

$$\boldsymbol{\Sigma}_{xx} = \mathbf{S}_{xx} \tag{37a}$$

$$\boldsymbol{\Sigma}_{xy} = \mathbf{S}_{xy} \tag{37b}$$

$$\boldsymbol{\Sigma}_1 = n(n-1)^{-1}\mathbf{S}_w \tag{37c}$$

$$\boldsymbol{\Sigma} = n(\mathbf{S}_b - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}) \tag{37d}$$

whence

$$\boldsymbol{\Sigma}_{2\cdot x} = \mathbf{S}_b - (n-1)^{-1}\mathbf{S}_w - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} \tag{38}$$

whence

$$\boldsymbol{\Sigma}_2 = \mathbf{S}_b - (n-1)^{-1}\mathbf{S}_w. \tag{39}$$

That is, in the balanced case, with unrestricted parameter matrices, the maximum likelihood estimates are obtainable in closed form as simple functions of the sufficient statistics $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$, $\mathbf{S}_{xx}$, $\mathbf{S}_{xy}$, $\mathbf{S}_b$, $\mathbf{S}_w$.

Substituting the unrestricted estimates in the function of likelihood yields

$$\begin{aligned} l = m\{&\log|\mathbf{S}_{xx}| + (n-1)\log|n(n-1)^{-1}\mathbf{S}_w| \\ &+ \log|\mathbf{S}_b - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}| + p + nq\} = l_\Omega \text{ say.} \end{aligned} \tag{40}$$

If $\boldsymbol{\Sigma}_{xx}$, $\boldsymbol{\Sigma}_{xy}$, $\boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}_1$ are constrained to be functions of $\boldsymbol{\pi}$, $t \times 1$, and we write

$$l_\omega = \min_{\boldsymbol{\pi}} l(\boldsymbol{\pi}), \tag{41}$$

then the quantity

$$U = l_\omega - l_\Omega \tag{42}$$

is a suitable discrepancy function for measuring the misfit of the model, since it is bounded below by zero and the bound is attained if and only if the model fits the sample exactly. Further, by well-known theory, U has an asymptotic chi-square distribution that yields an overall test of the restrictive hypothesis, with, in an identified model, degrees of freedom given by

$$d=(1/2)p(p+1)+q(q+1)+pq-t.$$

Alternatively, we might measure the misfit of the model by a suitable function of U and d. (Measures of goodness-of-fit of generally false restrictive models that nevertheless approximate the data seem currently open to a number of methodological and statistical questions, so it seems best not to make specific recommendations here.)

We note that we may alternatively estimate the parameters in the balanced case with unrestricted mean vectors by a generalized least squares procedure in which, defining

$$\mathbf{W}=\begin{bmatrix} \mathbf{S}_{xx} & \cdot & \mathbf{1}'_n \otimes \mathbf{S}_{xy} \\ \mathbf{1}_n \otimes \mathbf{S}_{yx} & \cdot & \mathbf{I}_n \otimes \mathbf{S}_w + \mathbf{1}_n \mathbf{1}'_n \otimes \mathbf{S}_b \end{bmatrix} \tag{43}$$

we minimize

$$s(\boldsymbol{\pi})=[\text{vec}(\mathbf{V}-\mathbf{W})]'[\mathbf{W}^{-1} \otimes \mathbf{W}^{-1}][\text{vec}(\mathbf{V}-\mathbf{W})], \tag{44}$$

that is

$$s(\boldsymbol{\pi})=\text{Tr}\{[\mathbf{W}^{-1}(\mathbf{V}-\mathbf{W})]^2\} \tag{45}$$

since $\mathbf{W}$ so defined is a consistent estimator of $\mathbf{V}$ (see Browne, 1982).

Since $\mathbf{W}^{-1}$ may be written as the obvious analogue of $\mathbf{V}^{-1}$ as given in (A8)–(A10), it is easily verified that

$$\mathbf{W}^{-1}(\mathbf{V}-\mathbf{W})=\begin{bmatrix} \mathbf{A} & \cdot & \mathbf{1}'_n \otimes \mathbf{B} \\ \mathbf{1}_n \otimes \mathbf{C} & \cdot & \mathbf{I}_n \otimes \mathbf{D} + \mathbf{1}_n \mathbf{1}'_n \otimes \mathbf{E} \end{bmatrix}, \tag{46}$$

where

$$\mathbf{A}=\mathbf{S}_{xx}^{-1}\boldsymbol{\Delta}_{xx}+n\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}^{-1}[\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\boldsymbol{\Delta}_{xx}-\boldsymbol{\Delta}_{yx}] \tag{47a}$$

$$\mathbf{B}=[\mathbf{S}_{xx}^{-1}+n\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}]\boldsymbol{\Delta}_{xy}-\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}^{-1}[\boldsymbol{\Delta}_1+n\boldsymbol{\Delta}_2] \tag{47b}$$

$$\mathbf{C}=\mathbf{S}^{-1}[\boldsymbol{\Delta}_{yx}-\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\boldsymbol{\Delta}_{xx}] \tag{47c}$$

$$\mathbf{D}=\mathbf{S}_1^{*-1}\boldsymbol{\Delta}_1 \tag{47d}$$

$$\mathbf{E}=\mathbf{S}^{-1}\boldsymbol{\Delta}_2-n^{-1}(\mathbf{S}_1^{*-1}-\mathbf{S}^{-1})\boldsymbol{\Delta}_1-\mathbf{S}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\boldsymbol{\Delta}_{xy}, \tag{47e}$$

and

$$\mathbf{S}=n(\mathbf{S}_b-\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy})+\mathbf{S}_w, \tag{48}$$

and we write

$$\mathbf{\Delta}_{xx} = \mathbf{\Sigma}_{xx} - \mathbf{S}_{xx} \tag{49a}$$

$$\mathbf{\Delta}_{xy} = \mathbf{\Sigma}_{xy} - \mathbf{S}_{xy} \tag{49b}$$

$$\mathbf{\Delta}_1 = \mathbf{\Sigma}_1 - \mathbf{S}_w \tag{49c}$$

$$\mathbf{\Delta}_2 = \mathbf{\Sigma}_2 - \mathbf{S}_b. \tag{49d}$$

It follows that

$$s(\pi) = m \operatorname{Tr} \{\mathbf{A}^2 + 2n\mathbf{BC} + n\mathbf{D}^2 + n^2\mathbf{E}^2 + 2n\mathbf{DE}\}, \tag{50}$$

a form which allows a computationally practical treatment of this alternative.

Investigators have shown a great deal of ingenuity in developing algebraic devices enabling the application of computer programs such as the LISREL series of Jöreskog & Sörbom (1979) or the COSAN program for McDonald's (1978, 1980) model to problems to which they may at first sight appear inapplicable. It is natural to ask whether there may exist an algebraic formulation of a multilevel model such as those described here, and an arrangement of the data, that would allow efficient estimates to be obtained by LISREL or COSAN. (It seems likely that crude estimates of some kind could be obtained in the balanced case using the sample covariance matrices defined above as submatrices of a suitably partitioned matrix of acceptably small order, but such estimates may not have good properties.)

In the balanced case with unrestricted mean vectors, we may indeed obtain GLS or ML estimates of the parameters in the models discussed in Section 2, using COSAN (and in special cases by LISREL) if n is small enough to allow the available computer configuration to accept an input covariance matrix of order $p+nq$. In such a case, we may obtain GLS estimates with COSAN by fitting the model

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{I}_p & & \\ & \mathbf{1}_n \otimes \mathbf{I}_q & \mathbf{I}_n \otimes \mathbf{I}_q \end{bmatrix} \cdot \begin{bmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} & \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_2 & \\ & & \mathbf{I}_n \otimes \mathbf{\Sigma}_1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_p & \\ & \mathbf{1}'_n \otimes \mathbf{I}_q \\ & \mathbf{I}_n \otimes \mathbf{I}_q \end{bmatrix} \tag{51}$$

(using the algebraic identities given in McDonald, 1978, 1980 and McArdle & McDonald, 1984, to express $\mathbf{\Sigma}_{xx}$, $\mathbf{\Sigma}_{xy}$, $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ as matrix functions of the parameters), to the $(p+nq) \times (p+nq)$ sample matrix $\mathbf{W}$ defined in (43). It is easily verified from (34) or directly from (26) that in the balanced case with unrestricted means, ML estimates may similarly be obtained by minimizing the discrepancy function

$$l = \operatorname{Tr} \{\mathbf{W\Sigma}^{-1}\} - \log |\mathbf{W\Sigma}^{-1}| - (p+nq), \tag{52}$$

where, again, $\mathbf{W}$ is given by (43) and $\mathbf{\Sigma}$ by (51). It does not seem possible to obtain these estimates by defining $\mathbf{\Sigma}$ and $\mathbf{W}$ of smaller order. (Because of the algebraic structure of LISREL, it seems that its application to this class of model would best be investigated case by case.) In practice, n will commonly be too large to allow this device to be employed. Further, it does not seem possible to carry this alternative treatment over to the unbalanced case.

4. Summary and discussion

A general model for linear structural relations in two-level data has been described and illustrated with a two-level common factor model and a two-level extension of McArdle's RAM model. The derivatives of the likelihood with respect to the parameter matrices of the model have been obtained, as a basis for numerical algorithms for estimating the parameters in any fully specified case.

In the balanced sampling design, with the same number of level-one units in each level-two unit, the sample analogue mean vectors and covariance matrices have been shown to be minimal sufficient statistics. It was further shown that these or simple functions of them are maximum likelihood estimates of the unrestricted parameter matrices. Hence we obtain a discrepancy function based on the ratio of likelihoods, respectively under a restrictive hypothesis and without restriction, which will also yield an asymptotic chi-square test for any identified model. In the balanced case these results further provide the basis for a generalized least squares estimation procedure that is non-iterative in the sense that the weight matrix is not updated.

The results show that it will commonly be impractical to rearrange a two-level model for linear structural relations so as to fit it by an existing computer program such as LISREL or COSAN in the balanced case, and *a fortiori* the unbalanced case is unlikely to admit such treatment. They also suggest that there may be strategic advantages in developing special computer programs for balanced designs in addition to programs for what will almost certainly be the more common case of the unbalanced design.

In unbalanced designs, there may be little to be gained in the way of special properties for models less general than that given by Goldstein & McDonald (1988) which, *inter alia*, takes account in a natural way of data missing at random at any level of sampling. In developing the necessary computer programs we can expect to find it desirable to test the efficiency of a variety of numerical algorithms, including quasi-Newton methods and Fisher's method of scoring, possibly in combination with the iterated generalized least squares algorithm given by Goldstein (1986).

Acknowledgement

We thank George Cooney for his comments on the manuscript.

References

Aitkin, M. & Longford, N. (1986). Statistical modelling in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society*, **A, 1949,** 1–43.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*. Cambridge: Cambridge University Press.

De Leeuw, J. (1985). *Path Models with Random Coefficients*. Leiden, The Netherlands: Department of Data Theory FSW/RUL.

De Leeuw, J. & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, **11**, 57–85.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, **73**, 43–56.

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. London: Griffin; New York: Oxford University Press.

Goldstein, H. & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, **53**, 435–467.

Hannan, M. T. (1971). *Aggregation and Disaggregation in Sociology*. Lexington, MA: Heath-Lexington.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320–340.

Jöreskog, K. G. & Sörbom, D. (1979). *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books.

Longford, N. (*in press*). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Journal of the American Statistical Association*.

Marsh, H. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, **47**, 213–231.

Mason, W. M., Wong, G. Y. & Entwisle, B. (1984). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological Methodology 1983*. San Francisco: Jossey-Bass.

McArdle, J. J. & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 234–251.

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

McDonald, R. P. (1980). A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *British Journal of Mathematical and Statistical Psychology*, **33**, 161–183.

McDonald, R. P. & Swaminathan, H. (1973). A simple matrix calculus with applications to multivariate analysis. *General Systems*, **18**, 37–54.

Robinson, W. S. (1950). Ecological correlation and the behavior of individuals. *American Sociological Review*, **15**, 351–357.

Tate, R. L. & Wongbundhit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics*, **8**, 103–120.

Received 3 December 1987; revised version received 21 June 1988

Appendix A

Derivation of (32)

We write

$$\mathbf{V}_k = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \tag{A.1}$$

and

$$\mathbf{V}_k^{-1}=\begin{bmatrix}\mathbf{V}^{11} & \mathbf{V}^{12}\\ \mathbf{V}^{21} & \mathbf{V}^{22}\end{bmatrix} \tag{A.2}$$

where the submatrices correspond to those in (23).

Then by the well-known identity for the inverse of a partitioned matrix, with a little algebra we obtain

$$\mathbf{V}^{11}=\boldsymbol{\Sigma}_{xx}^{-1}+n_k\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}(n_k\boldsymbol{\Sigma}_{2\cdot x}+\boldsymbol{\Sigma}_1)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} \tag{A.3}$$

$$\mathbf{V}^{12}=-\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{1}'_{n_k}\otimes\boldsymbol{\Sigma}_{xy}(n_k\boldsymbol{\Sigma}_{2\cdot x}+\boldsymbol{\Sigma}_1)^{-1}] \tag{A.4}$$

$$\mathbf{V}^{22}=[\mathbf{I}_{n_k}\otimes\boldsymbol{\Sigma}_1+\mathbf{1}_{n_k}\mathbf{1}'_{n_k}\otimes\boldsymbol{\Sigma}_{2\cdot x}]^{-1}, \tag{A.5}$$

where $\boldsymbol{\Sigma}_{2\cdot x}$ is given by (29).

By the easily verified identity, for symmetric **A** and **B**,

$$(\mathbf{I}\otimes\mathbf{A}+\mathbf{1}\mathbf{1}'\otimes\mathbf{B})^{-1}=\mathbf{I}\otimes\mathbf{A}^{-1}-\mathbf{1}\mathbf{1}'\otimes(\mathbf{1}'\mathbf{1}\mathbf{A}+\mathbf{A}\mathbf{B}^{-1}\mathbf{A})^{-1}, \tag{A.6}$$

(A.5) reduces further to

$$\mathbf{V}^{22}=\mathbf{I}_{n_k}\otimes\boldsymbol{\Sigma}_1^{-1}-\mathbf{1}_{n_k}(\mathbf{1}'_{n_k}\mathbf{1}_{n_k})^{-1}\mathbf{1}_{n_k}\otimes[\boldsymbol{\Sigma}_1^{-1}-(\mathbf{n}_k\boldsymbol{\Sigma}_{2\cdot x}+\boldsymbol{\Sigma}_1)^{-1}]. \tag{A.7}$$

Using (30) we write these expressions as

$$\mathbf{V}^{11}=\boldsymbol{\Sigma}_{xx}^{-1}+n_k\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} \tag{A.8}$$

$$\mathbf{V}^{12}=-\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{1}'_{n_k}\otimes\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}] \tag{A.9}$$

$$\mathbf{V}^{22}=\mathbf{I}_{n_k}\otimes\boldsymbol{\Sigma}_1^{-1}-\mathbf{1}_{n_k}(\mathbf{1}'_{n_k}\mathbf{1}_{n_k})^{-1}\mathbf{1}_{n_k}\otimes[\boldsymbol{\Sigma}_1^{-1}-\boldsymbol{\Sigma}_k^{-1}]. \tag{A.10}$$

Next we require a computationally convenient form for the determinant of $\mathbf{V}_k$. By a well-known identity, we have

$$|\mathbf{V}_k|=|\boldsymbol{\Sigma}_{xx}|\,|\mathbf{I}_{n_k}\otimes\boldsymbol{\Sigma}_1+\mathbf{1}_{n_k}\mathbf{1}'_{n_k}\otimes\boldsymbol{\Sigma}_{2\cdot x}|. \tag{A.11}$$

Define

$$\boldsymbol{\Sigma}_i=\mathbf{I}_i\otimes\boldsymbol{\Sigma}_1+\mathbf{1}_i\mathbf{1}'_i\otimes\boldsymbol{\Sigma}_{2\cdot x}, \tag{A.12}$$

$$\Delta_i=|\boldsymbol{\Sigma}_i|. \tag{A.13}$$

Then

$$\Delta_{i+1}=\Delta_i|\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_{2\cdot x}-(\mathbf{1}'_i\otimes\boldsymbol{\Sigma}_{2\cdot x})\boldsymbol{\Sigma}_i^{-1}(\mathbf{1}_i\otimes\boldsymbol{\Sigma}_{2\cdot x})|, \tag{A.14}$$

which may be written as

$$\Delta_{i+1} = \Delta_i |\Sigma_i| |(i+1)\Sigma_{2\cdot x} + \Sigma_1| / |i\Sigma_{2\cdot x} + \Sigma_1| \tag{A.15}$$

whence

$$\Delta_{n_k} = |\Sigma_1|^{n_k - 1} |n_k \Sigma_{2\cdot x} + \Sigma_1|, \tag{A.16}$$

whence, further, we have

$$|\mathbf{V}_k| = |\Sigma_{xx}| |\Sigma_1|^{n_k - 1} |\Sigma_k|. \tag{A.17}$$

Thence we have (32).

Appendix B

Derivatives

From (32) we have

$$\frac{\partial l_k}{\partial \boldsymbol{\mu}_x} = 2\{[\Sigma_{xx}^{-1} + n_k \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_k^{-1} \Sigma_{yx} \Sigma_{xx}](\mathbf{x}_k - \boldsymbol{\mu}_x) - n_k \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_k^{-1} (\mathbf{y}_k - \boldsymbol{\mu}_y)\} \tag{B.1}$$

$$\frac{\partial l_k}{\partial \boldsymbol{\mu}_y} = 2\mathbf{n}_k \{\Sigma_k^{-1} (\bar{\mathbf{y}}_k - \boldsymbol{\mu}_y) - n_k \Sigma_k^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} (\mathbf{x}_k - \boldsymbol{\mu}_x)\} \tag{B.2}$$

and the corresponding derivatives of l are, of course, the sums of these over k. In the general case, the resulting likelihood equations cannot be solved in closed form for unrestricted $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$. In the special case of a balanced sampling design, by (27) and (33), we have

$$\frac{\partial l}{\partial \boldsymbol{\mu}_x} = 2m\{[\Sigma_{xx}^{-1} + n\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}](\bar{\mathbf{x}} - \boldsymbol{\mu}_x) - n\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_y)\} = \mathbf{0} \tag{B.3}$$

and

$$\frac{\partial l}{\partial \boldsymbol{\mu}_y} = 2mn\{\Sigma^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_y) - \Sigma^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_x)\} = \mathbf{0}, \tag{B.4}$$

which are satisfied, as we might have expected, by

$$\boldsymbol{\mu}_x = \bar{\mathbf{x}};\ \boldsymbol{\mu}_y = \bar{\mathbf{y}}. \tag{B.5}$$

We now require the derivatives of l_k, $k = 1, \ldots, m$ (hence immediately of l) with respect to Σ_{xx}, Σ_{xy}, Σ_1, Σ_2. It is convenient to write

$$\mathbf{S}_{xxk} = (\mathbf{x}_k - \boldsymbol{\mu}_x)(\mathbf{x}_k - \boldsymbol{\mu}_x)' \tag{B.6}$$

$$\mathbf{S}_{xyk} = (\mathbf{x}_k - \boldsymbol{\mu}_x)(\bar{\mathbf{y}}_k - \boldsymbol{\mu}_y)' \tag{B.7}$$

$$\mathbf{S}_{yyk} = \sum_{i=1}^{n_k} (\mathbf{y}_{ki} - \boldsymbol{\mu}_y)(\mathbf{y}_{ki} - \boldsymbol{\mu}_y)' \tag{B.8}$$

$$\mathbf{S}_{\bar{y}\bar{y}k} = (\bar{\mathbf{y}}_k - \boldsymbol{\mu}_y)(\bar{\mathbf{y}}_k - \boldsymbol{\mu}_y)'. \tag{B.9}$$

A reasonably mechanical procedure for obtaining the required matrices of derivatives is given by McDonald & Swaminathan (1973). These may be arranged in the form

$$\begin{aligned}\frac{\partial l_k}{\partial \boldsymbol{\Sigma}_{xx}} = {} & \boldsymbol{\Sigma}_{xx}^{-1} - \boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xxk}\boldsymbol{\Sigma}_{xx}^{-1} + n_k\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}[2n_k\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xyk} \\ & - n_k\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xk}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} + \boldsymbol{\Sigma}_k - n_k\mathbf{S}_{\bar{y}\bar{y}k}]\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} \\ & + 2n_k\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xyk}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_{yx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xxk}]\boldsymbol{\Sigma}_{xx}^{-1},\end{aligned} \tag{B.10}$$

$$\begin{aligned}\frac{\partial l_k}{\partial \boldsymbol{\Sigma}_{xy}} = {} & 2n_k[\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xxk}\boldsymbol{\Sigma}_{xx}^{-1} - \boldsymbol{\Sigma}_{xx}^{-1}]\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1} \\ & - 2n_k^2\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xyk} - \mathbf{S}_{xxk}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1} \\ & - 2n_k\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xyk} - n_k\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_k^{-1}(\mathbf{S}_{\bar{y}\bar{y}k} - \mathbf{S}_{yxk}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})]\boldsymbol{\Sigma}_k^{-1},\end{aligned} \tag{B.11}$$

$$\begin{aligned}\frac{\partial l_k}{\partial \boldsymbol{\Sigma}_1} = {} & (n_k - 1)\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_1^{-1}(\mathbf{S}_{yyk} - n_k\mathbf{S}_{\bar{y}\bar{y}k})\boldsymbol{\Sigma}_1^{-1} \\ & + \boldsymbol{\Sigma}_k^{-1}[\boldsymbol{\Sigma}_k - n_k\{\mathbf{S}_{\bar{y}\bar{y}k} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(2\mathbf{S}_{xxk}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} - \mathbf{S}_{xyk})\}]\boldsymbol{\Sigma}_k^{-1},\end{aligned} \tag{B.12}$$

$$\begin{aligned}\frac{\partial l_k}{\partial \boldsymbol{\Sigma}_2} = {} & n_k\boldsymbol{\Sigma}_k^{-1}[\boldsymbol{\Sigma}_k - n_k\mathbf{S}_{\bar{y}\bar{y}k}]\boldsymbol{\Sigma}_k^{-1} \\ & + n_k^2\boldsymbol{\Sigma}_k^{-1}[\mathbf{S}_{\bar{y}\bar{y}k} - 2\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xyk} \\ & + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xxk}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}]\boldsymbol{\Sigma}_k^{-1}.\end{aligned} \tag{B.13}$$

The corresponding derivatives of l are, again, the sums of these over k.

In the special case of a balanced sampling design these yield the likelihood equations

$$\begin{aligned}m^{-1}\frac{\partial l}{\partial \boldsymbol{\Sigma}_{xx}} = {} & \boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{\Sigma}_{xx} - \mathbf{S}_{xx})\boldsymbol{\Sigma}_{xx}^{-1} \\ & + n\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^{-1}[\boldsymbol{\Sigma} - n\mathbf{S}_b + n\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(2\mathbf{S}_{xy} - \mathbf{S}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})]\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\end{aligned}$$

$$+2n\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xy}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{yx}-\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xx}]\boldsymbol{\Sigma}_{xx}^{-1}=\mathbf{0}, \quad \text{(B.14)}$$

$$m^{-1}\frac{\partial l}{\partial \boldsymbol{\Sigma}_{xy}}=2n\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xx}-\boldsymbol{\Sigma}_{xx}]\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^{-1}$$

$$-2n^2\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xy}-\mathbf{S}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}]\boldsymbol{\Sigma}^{-1}$$

$$-2n\boldsymbol{\Sigma}_{xx}^{-1}[\mathbf{S}_{xy}-\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^{-1}n(\mathbf{S}_b-\mathbf{S}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})]\boldsymbol{\Sigma}^{-1}=\mathbf{0}, \quad \text{(B.15)}$$

$$m^{-1}\frac{\partial l}{\partial \boldsymbol{\Sigma}_1}=\boldsymbol{\Sigma}_1^{-1}[(n-1)\boldsymbol{\Sigma}_1-n\mathbf{S}_w]\boldsymbol{\Sigma}_1^{-1}$$

$$+\boldsymbol{\Sigma}^{-1}[\boldsymbol{\Sigma}-n\{\mathbf{S}_b-\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(2\mathbf{S}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}-\mathbf{S}_{xy})\}]\boldsymbol{\Sigma}^{-1}=0, \quad \text{(B.16)}$$

$$m^{-1}\frac{\partial l}{\partial \boldsymbol{\Sigma}_2}=n\boldsymbol{\Sigma}^{-1}-n^2\boldsymbol{\Sigma}^{-1}[\mathbf{S}_b-(2\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xy}-\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})]\boldsymbol{\Sigma}^{-1}=\mathbf{0}. \quad \text{(B.17)}$$