THE ANALYSIS OF HIERARCHICAL AND UNBALANCED COMPLEX SURVEY DATA USING MULTILEVEL MODELS

MATHILDA DU TOIT

THE ANALYSIS OF HIERARCHICAL AND UNBALANCED COMPLEX SURVEY DATA USING MULTILEVEL MODELS

by

Mathilda du Toit

Submitted in partial fulfilment of the requirements

for the degree

Doctor of Philosophy in the subject Applied Statistics

in the Faculty of Science

University of Pretoria

Pretoria
December 1995

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

It was a privilege to undertake this study under the guidance of Professor D.J. Stoker and I am indebted to him for his constructive comments and encouragement.

I would like to thank the University of Pretoria and the Human Sciences Research Council for financial assistance and access to survey data. Halbe Barnardt and Johan du Toit assisted greatly by typing parts of the manuscript. I would also like to thank Bruce Sanderson who did the final editing. Finally I would like to thank Michael Browne, Darrell Bock and Stephen du Toit for making available their FORTRAN matrix manipulation subroutines.

This study would have been impossible without the continuous support and encouragement of Stephen du Toit. I gratefully acknowledge his contribution.

NOTATION

The following notation will be used:

π	: constant, $\pi = 3.14159...$
e	: Euler's constant, $e = 2.71828...$
$\exp(x)$	: e^x; $-\infty < x < \infty$
$ln\ x$	: natural logarithm of the real number x, $x \geq 0$
δ_{ij}	: Kronecker's delta, $\delta_{ij} = 1$ if $i = j$, 0 otherwise
$\mathbf{I}_n$	: an identity matrix of order $n \times n$
$\mathbf{A} : p \times q$	: a matrix of order $p \times q$
$\mathbf{a} : p \times 1$	: a column vector of order $p \times 1$
a_{ij} or $[\mathbf{A}]_{ij}$	: the element in the i-th row and j-th column of the matrix $\mathbf{A}$
a_i or $[\mathbf{a}]_i$	: the i-th element of the vector $\mathbf{a}$
$\mathbf{A}^{-1}$	: the inverse of the matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	: the determinant of the matrix $\mathbf{A}$
tr $\mathbf{A}$	: the trace of the matrix $\mathbf{A}$
Diag($\mathbf{A}$)	: a diagonal matrix formed from the diagonal elements of $\mathbf{A}$
vec $\mathbf{A}$	: the $pq \times 1$ vector formed from the q columns of the $p \times q$ matrix $\mathbf{A}$
vecs $\mathbf{A}$	: the $\frac{1}{2}p(p+1) \times 1$ vector of nonduplicated elements of the $p \times p$ symmetric matrix $\mathbf{A}$
$\mathbf{0}$	: a null matrix, that is, $[\mathbf{0}]_{ij} = 0$ for all i and j
$\mathbf{J}_{ij}$	: a matrix with all elements equal to zero with the exception of the element in the i-th row and j-th column, which is equal to 1
$\mathbf{j}$	: a column vector with all elements equal to one
$\mathbf{A} \otimes \mathbf{B}$	: the right direct product or Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$ as defined by:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & . & a_{1q}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & . & a_{2q}\mathbf{B} \\ . & . & . & . \\ a_{p1}\mathbf{B} & a_{p2}\mathbf{B} & . & a_{pq}\mathbf{B} \end{bmatrix}$$

$E(\mathbf{y}): p \times 1$	: the expected value of the random vector $\mathbf{y}$ with typical element $E(y)_i$
$Cov(\mathbf{y},\mathbf{y}'): p \times p$	: the covariance matrix of the random vector $\mathbf{y}$ with typical element $E(y_i - E(y_i))(y_j - E(y_j))'$

SUMMARY

TITLE: THE ANALYSIS OF HIERARCHICAL AND UNBALANCED COMPLEX SURVEY DATA USING MULTILEVEL MODELS

CANDIDATE: MATHILDA DU TOIT

PROMOTER: PROFESSOR D.J. STOKER

DEPARTMENT: STATISTICS

DEGREE: PHD (APPLIED STATISTICS)

Multilevel analysis allows characteristics of different groups to be included in models of individual behaviour. Most analyses of social sciences data entail the analysis of data with built-in hierarchies, usually obtained as a consequence of complex sampling methods. The formulation of such models and estimation procedures may be seen as an effort to develop a new family of analytical tools that correspond to the classical experimental designs.

The purpose of this dissertation is to investigate the efficient analysis of level-3 models, which includes the estimation of the unknown parameters and statistical inference. Use is made of the Expected Maximization algorithm and the Iterative Generalized Least Squares algorithm. As most data sets from the social sciences are quite large, the feasibility of analysing large data sets efficiently is investigated. Attention is given to the problem of developing a computer program that is easy to use as a standard statistical package.

Theoretical results required for the estimation of the unknown parameters are extended to a general level-3 model, allowing for complex variance structures on all levels of the hierarchy. Since it often happens that there may be more than one response variable of interest, for example in a personality test with a number of items, the analysis of models with two or more continuous response variables is also considered.

Survey data in the social sciences are usually of a categorical nature. It is shown how data with a categorical response variable can be analysed within the general framework developed. The theory is extended to accommodate the simultaneous analysis of more than one categorical response variable. Suggestions for further research are given, including guidelines for the handling of non-linear multilevel models.

Most of the theory derived in this study is illustrated with examples based on real data and has been implemented in FORTRAN programs.

OPSOMMING

TITEL: DIE ONTLEDING VAN HIËRARGIESE EN ONGEBALANSEERDE KOMPLEKSE OPNAMEDATA MET BEHULP VAN MEERPEIL MODELLE

KANDIDAAT: MATHILDA DU TOIT

PROMOTOR: PROFESSOR D.J. STOKER

DEPARTEMENT: STATISTIEK

GRAAD: PHD (TOEGEPASTE STATISTIEK)

Meerpeil modelle fasiliteer die insluiting van karakteristieke van verskillende groepe in die modellering van individuele gedrag. Die analise van geesteswetenskaplike data behels dikwels die ontleding van data met 'n inherente hiërargiese struktuur, wat gewoonlik verkry word as 'n gevolg van die gebruik van komplekse steekproefnemingsmetodes. Die formulering van sodanige modelle kan gesien word as 'n poging tot die ontwikkeling van nuwe analitiese hulpmiddels wat ooreenstem met die klassieke eksperimentele ontwerpe.

Die doel van hierdie verhandeling is om wyses waarop 3-peil modelle doeltreffend ontleed kan word, te ondersoek. Hierdie ondersoek sluit die beraming van die onbekende parameters en die uitvoer van statistiese inferensie in. Daar word gebruik gemaak van die maksimering van voorwaardelike verwagtingswaardes en van iteratiewe veralgemeende kleinste kwadrate. Aangesien data afkomstig vanuit geesteswetenskaplike navorsing gewoonlik omvangryk is, word klem gelê op die doeltreffende analise van groot datastelle. Aandag word ook geskenk aan die wyse waarop 'n gebruikersvriendelike rekenaarprogram ontwikkel kan word.

Die teoretiese resultate benodig vir die beraming van die onbekende parameters word uitgebrei om 'n algemene 3-peil model daar te stel wat voorsiening maak vir die hantering van komplekse variansie strukture op elke peil van die hiërargie. Aangesien daar dikwels meer as een afhanklike veranderlike betrokke is in 'n navorsingsondersoek,

soos byvoorbeeld in 'n persoonlikheidstoets, word die analise van modelle met twee of meer kontinue afhanklike veranderlikes ook beskou.

Opname data is gewoonlik van 'n kategoriese aard en daar word aangetoon hoe data met 'n kategoriese afhanklike veranderlike binne die algemene raamwerk ontleed kan word. Die teorie word uitgebrei om voorsiening te maak vir die gelyktydige ontleding van meer as een kategoriese afhanklike veranderlike. Voorstelle vir verdere navorsing, insluitende riglyne vir die hantering van nie-lineêre meerpeil modelle, word gegee.

Meeste van die teorie wat afgelei is, word met voorbeelde gebaseer op werklike data toegelig. Rekenaarprogramme is geskryf om hierdie voorbeelde te ontleed.

CHAPTER 1
INTRODUCTION

Multilevel models deal with the analysis of data where observations are nested within groups. Social, behavioural and even economic data often have a hierarchical structure. A frequently cited example is in education, where students are grouped in classes. Classes are grouped in schools, schools in education departments and so on. We thus have variables describing individuals, but the individuals may be grouped into larger or higher order units.

Traditionally, fixed parameter linear regression models are used for the analysis of such data, and statistical inference is based on the assumptions of linearity, normality, homoscedascity and independence. Ideally, only the first of these assumptions should be used. It has been shown by Aitkin and Longford (1986), that the aggregation of variables over individual observations may lead to misleading results. Both the aggregation of individual variables to a higher level of observation and the disaggregation of higher order variables to an individual level have been somewhat discredited (Bryk & Raudenbush, 1992). It has also been pointed out by Holt, Scott and Ewings (1980), that serious inferential errors may result from the analysis of complex survey data if it is assumed that the data have been obtained under a simple random sampling scheme.

Random regression models have been developed to model continuous data (Bock, 1983), and also dichotomous repeated measurement data (Gibbons & Bock, 1987), where certain characteristics of the data preclude the use of traditional ANOVA models. In random regression models, however, there is still no possibility of including higher level variables. In order to accommodate both random coefficients and higher order variables, multilevel models should be used.

Multilevel analysis allows characteristics of different groups to be included in models of individual behaviour. Most analyses of social sciences data entail the analysis of data with built-in hierarchies, usually obtained as a consequence of complex sampling methods. Thus, the scope for application of multilevel models is very wide. The formulation of such models and estimation procedures may be seen as an effort to develop a new family of analytical tools that correspond to the classical experimental designs. These models are much more flexible in that they are capable of handling unbalanced data, the analysis of variance-covariance components and the analyis of both continuous and discrete response variables. As the characteristics of individual groups are incorporated into the multilevel model, the hierarchical structure of the data is taken into account and correct estimates of standard errors are obtained. The exploration of variation between groups, which may be of interest in its own right, is facilitated. Valid tests and confidence intervals can also be constructed and stratification variables used in the sample design can be incorporated into the model.

The use of multilevel models has been hampered in the past by the fact that closed form mathematical formulas to estimate the variance and covariance components have only been available for perfectly balanced designs. Iterative numerical procedures must be used to obtain efficient estimates for unbalanced designs. Among the procedures suggested are the full maximum likelihood (Goldstein, 1986 and Longford, 1987), and the restricted maximum likelihood as proposed by Mason *et al* (1983) and Bryk & Raudenbush (1986). Another approach is the procedure of Bayes estimation (Dempster *et al*, 1981). Some other procedures include the use of Iteratively Reweighted Generalized Least Squares (Goldstein, 1986), and a Fisher scoring algorithm (Longford, 1987). Owing to the sizes of matrices involved in the iterative procedure and the computer storage space required during estimation, the implementation of such a procedure is not straightforward.

The purpose of this dissertation is to investigate level-3 multilevel models and the estimation of the unknown parameters. Use is made of the Expected Maximization (EM) algorithm and the Iterative Generalized Least Squares (IGLS) algorithm. In particular, attention is given to the efficient computer implementation of estimation procedures for the estimation of the unknown parameters. As most data sets from the social sciences are quite large, the feasibility of analysing large data sets efficiently is also investigated. Attention is also given to the problem of developing a computer program that is easy to use as a standard statistical package in order to make multilevel model analysis readily accessible for the social science researcher who will benefit most from using it.

In Chapter 2 the concept of multilevel modelling is introduced. The standard fixed parameter linear regression model and the disadvantages of such an approach to the analysis of hierarchically structured data are discussed. This is followed by the definition and subsequent discussion of a level-2 hierarchical model. The advantages of a multilevel modelling approach is considered. A brief overview of estimation procedures that may be used for the analysis of unbalanced hierarchical data is given and the use of the IGLS algorithm is considered for a level-2 model. Computer programs for the analysis of multilevel data are briefly discussed and the difficulties concerning the implementation of algorithms noted.

Chapter 3 contains the extension of the level-2 model discussed in Chapter 2 to allow for the model coefficients to be random across a third level of the hierarchy. Theoretical results required for the estimation of the unknown parameters in the level-3 model are given, based on the EM optimization algorithm. The estimation procedure is illustrated with a practical application.

The EM algorithm is a fast, robust method for obtaining maximum likelihood estimates of the unknown parameters. In certain cases, however, a large number of iterations may be required for the procedure to converge. A more serious disadvantage of the EM algorithm is that standard errors of the estimators cannot be obtained. The EM algorithm also does not facilitate statistical inference such as hypothesis and contrast testing.

In Chapter 4 the estimation of the unknown parameters in a level-3 model using an IGLS algorithm is discussed. Although the mathematical equations on which this procedure is based appear to be straightforward, simplification of these equations is necessary to ensure that the optimization algorithm is computationally efficient. Some of the shortcomings of the EM algorithm used in Chapter 3 are addressed. To illustrate the theoretical principles involved, two practical applications are given. The level-3 model discussed in this chapter is, however, still limited as complex level-1 structures cannot be accommodated. It is also not possible to use different sets of predictors in the fixed and stochastic parts of the model.

A general level-3 model, allowing for complex variance structures on all levels of the hierarchy, is introduced in Chapter 5. Cases where no coefficients are random at a specific level are considered. Three examples are given. The examples range from a case which is similar in structure to the cases discussed in the previous chapters, to an example of a level-3 model with complex variance structures on all levels of the hierarchy. Note that the estimation procedure discussed in this chapter is computationally less efficient than the procedures discussed in previous chapters. Its chief advantage lies in its ability to handle a wide variety of models, as provision is also made for the situation where there are no random coefficients on a particular level of the model.

It often happens that there may be more than one response variable of interest, for example in a personality test with a number of items. The analysis of models with two or more continuous response variables is considered in Chapter 6 by the introduction of a multivariate multilevel model for the analysis of continuous response variables. The mathematical implications of missing data are discussed. Two practical examples are given.

Survey data in the social sciences are usually of a categorical nature. Most of the questions answered by respondents concern a respondent's attitude to a particular matter. Possible outcomes are normally either of the 'yes/no/don't know' type or are based on a scale, where '1' may indicate strong agreement and '5' strong disagreement with a statement. Data sets are frequently of a hierarchical nature, for instance single respondents are nested within households which may be nested within suburbs, which may in turn be nested in metropolitan and non-metropolitan areas. In Chapter 7 the analysis of data with categorical response variables is considered. The theory is also extended to accommodate the simultaneous analysis of more than one categorical response variable. The theory is implemented in a number of practical examples.

Finally, in Chapter 8 suggestions for further research are given, including guidelines for the handling of non-linear multilevel models in the theoretical framework of the preceding chapters.

All the data sets used in the practical applications given in Chapters 3 to 7 are from large sample surveys. Estimated parameters are, therefore, considered to be approximately distributed as multivariate normal variates. This large-sample property of the estimators is used in the calculation of standard errors and p-values.

Most of the theory discussed in this study has been implemented in FORTRAN programs. For illustrative purposes the program MULTVAR, used for the practical applications in Chapter 6, is included on the accompanying diskette. The Appendix contains installation information.

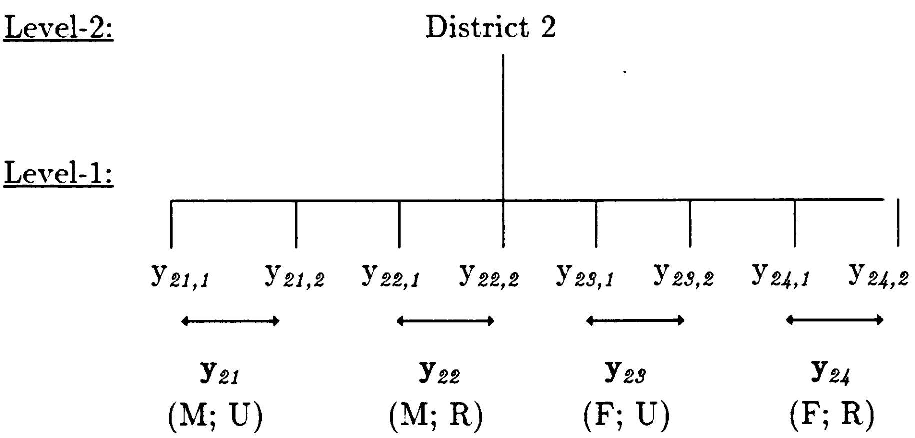
and

$$\mathbf{y}_{ij} = \begin{bmatrix} y_{ij,1} \\ y_{ij,2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{ij,c-1} \end{bmatrix} = \begin{bmatrix} ln \, \dfrac{f_{ij,1}}{f_{ij.}} \\ ln \, \dfrac{f_{ij,2}}{f_{ij.}} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ ln \, \dfrac{f_{ij,c-1}}{f_{ij.}} \end{bmatrix} . \tag{7.3.2}$$

Example 7.3.1

Consider a data set in which the response variable has $c = 3$ categories and the predictors are GENDER (MALE, FEMALE) and AREA (URBAN, SEMI-URBAN and RURAL). The level-2 units are 30 districts. There is no semi-urban area in District 2 and hence there are $s_2 = 4$ subpopulations, these being (MALE, URBAN), (MALE, RURAL), (FEMALE, URBAN) and (FEMALE, RURAL).

A schematic representation of the data for District 2 is as follows:



CHAPTER 2
INTRODUCTION TO MULTILEVEL MODELS

2.1 INTRODUCTION

Social and behavioural data usually have a hierarchical structure. A frequently cited example is in education, where students are grouped in classes. Classes are grouped in schools, schools in education departments and so on. We thus have variables describing individuals, but the individuals may be grouped into larger or higher order units. In economics there is the problem of relating the micro and macro levels.

Traditionally, the analysis of such data was done by using fixed parameter linear regression models with either aggregation or disaggregation of variables available for respondents or higher order units. Both these procedures, however, have been somewhat discredited (Prosser, Rasbash & Goldstein, 1991). In Section 2.2 a fixed parameter linear regression model will be discussed and illustrated with a small example. Attention will be given to the disadvantages associated with such an approach.

In Section 2.3 the concept of multilevel modelling will be introduced and a definition of a multilevel model will be given. A level-2 multilevel model will be considered. The data used in Section 2.2 will be re-analysed as a simple level-2 model. Advantages of a multilevel approach to the analysis of complex survey data will be given.

Estimation procedures and various algorithms which may be used for estimation will be discussed in Section 2.4. Problems with the implementation of algorithms for the estimation of the unknown parameters in a multilevel model will be considered. In Section 2.5 conclusions will be given.

2.2 FIXED PARAMETER LINEAR REGRESSION MODELS

In this section a fixed parameter linear regression model with the aggregation of individual variables to a higher level is illustrated.

Example 2.2.1

Consider the dental measurement data set first analyzed by Potthoff and Roy (1964). The data set contains the dental measurements of 11 girls and 16 boys at each of the ages 8, 10, 12 and 14 years. Each measurement is the distance in millimetres between the center of the pituitary and the pterygomaxillary fissure.

Suppose we wish to investigate the relationship between the measurements y_{ij}, $j = 1, 2, 3, 4$ for child i and the ages at which the measurement were taken. Denote these ages for individual i by x_{i1} , x_{i2} , x_{i3} and x_{i4} .

Traditionally a single linear equation may be estimated by pooling all 27 cases. The measurement may then be expressed as a linear function of the ages at which measurements are taken and could be written as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij} \ , j = 1, 2, 3, 4. \tag{2.2.1}$$

Let

$$\mathbf{x}'_{ij} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix}$$

and

$$\boldsymbol{\beta}' = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}$$

From (2.2.1), the measurement for individual i can be rewritten as

$$y_{ij} = \mathbf{x}'_{ij}\,\boldsymbol{\beta} + \mathbf{e}_{ij} \ , \quad i = 1, 2, \ldots, N \, ; j = 1, 2, 3, 4.$$

where $N = 27$ denotes the total number of children for which measurements were available.

Using matrix notation, the set of regression equations given above may be written as

$$\mathbf{y}_i = \mathbf{X}_i\, \boldsymbol{\beta} + \mathbf{e}_i\,, \quad i = 1, 2, \ldots, N \tag{2.2.2}$$

which can be rewritten as

$$\mathbf{y} = \mathbf{X}\, \boldsymbol{\beta} + \mathbf{e},$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \text{ and } \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_N \end{bmatrix}.$$

It is assumed that e_{11}, e_{12}, . . . are uncorrelated with mean zero and constant variance σ^2. Thus,

$$\mathrm{E}(\mathbf{e}) = \mathbf{0} \tag{2.2.3}$$

and

$$\mathrm{Cov}(\mathbf{e}, \mathbf{e}') = \sigma^2\, \mathbf{I}\,. \tag{2.2.4}$$

Under the assumptions given by (2.2.3) and (2.2.4), the ordinary least squares estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained as

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{y}$$

where

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

and

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}') = \sigma^2\,(\mathbf{X}'\,\mathbf{X})^{-1}.$$

If, however, a separate linear function is fitted for each child, 27 regression models of the form given in (2.2.2) will be obtained. Figures 2.2.1 and 2.2.2 are graphical representations of these fitted regression lines. In Figure 2.2.1 the regression lines for girls are given and the regression line for all the children pooled together. Figure 2.2.2 is a similar representation for the male group.

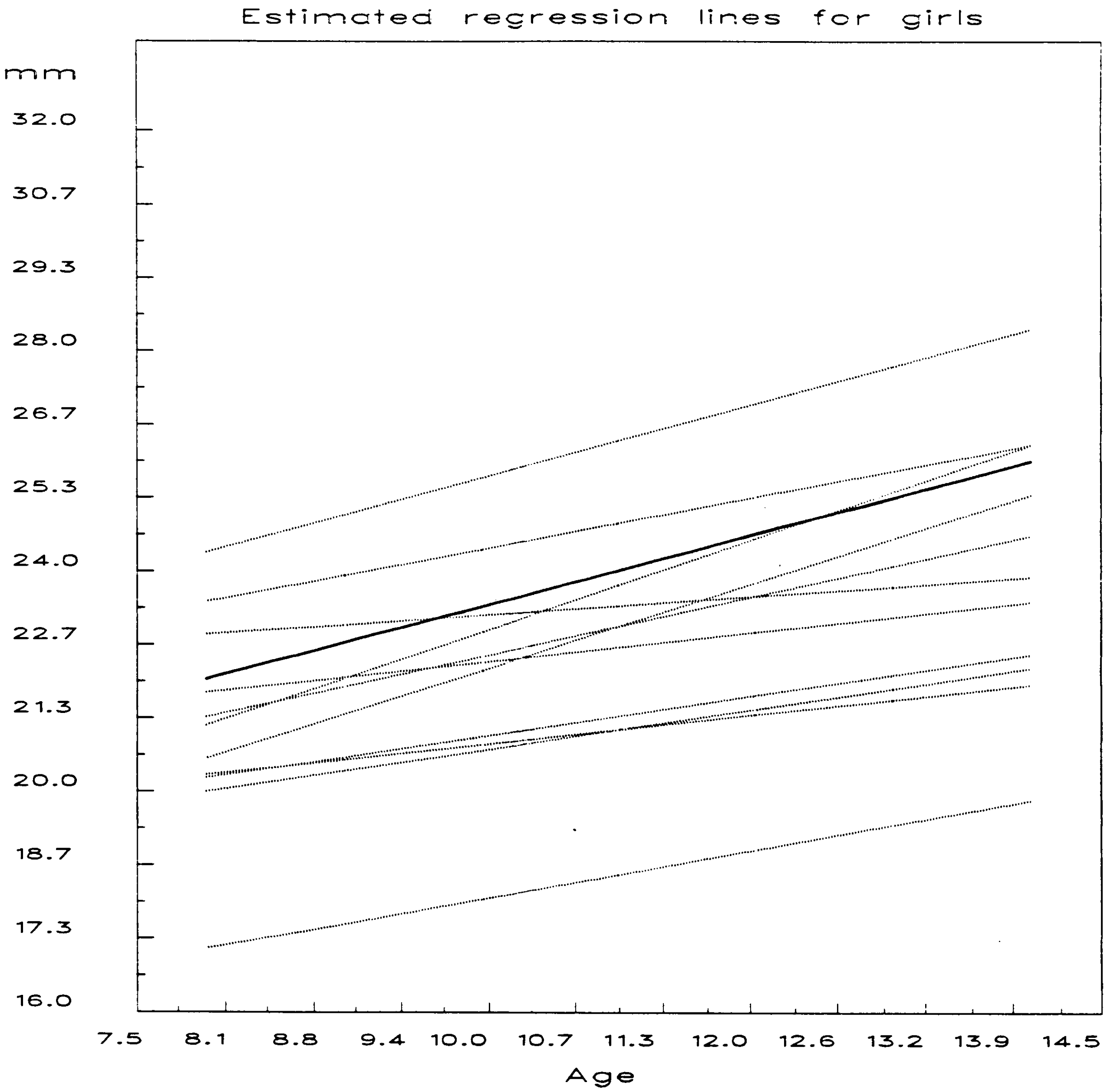


Figure 2.2.1: OLS regression lines of measurements on age for girls

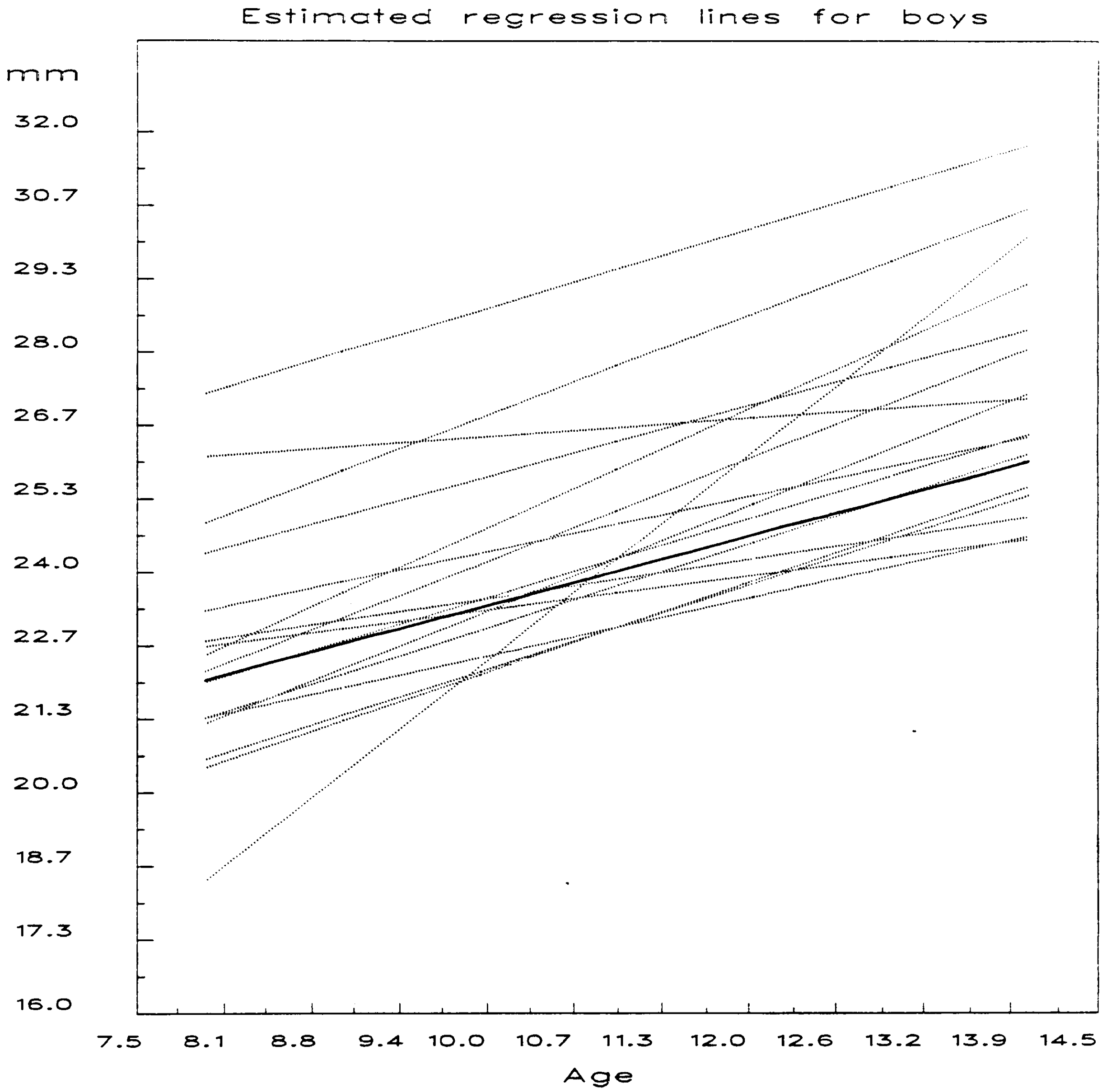


Figure 2.2.2: OLS regression lines of measurements on age for boys

From Figure 2.2.1 it can be seen that the intercepts and slopes for individual girls vary considerably. Most of the OLS regression lines for girls are below the OLS regression line for the combined group. In the case of the boys, the intercepts and slopes also vary considerably. Most of the OLS regression lines for the boys are above the regression line for the total group.

The large differences between lines cannot be ascribed to measurement error only, as measurement equipment for this type of measurement is usually fairly accurate. The error variance for the total group is 6.5609 mm^2. It would thus seem likely that each individual may have his or her own intercept and slope. □

This example illustrates some of the dangers inherent in ignoring the hierarchical structure of data. By aggregating higher order variables to an individual level, for instance assigning teacher, class and school characteristics to an individual, we know that students in the same class will have the same value on each of the class variables. The assumption of independence of observation that is basic for the classical statistical techniques cannot be used in such a situation (Bryk & Raudenbush, 1992).

On the other hand, individual level variables may be aggregated to a higher level and an analysis done on a higher level. This will in effect discard all the within-group information, which may account for a considerable part of the total variation. Consequently, relations between aggregated variables may be stronger and certainly very different from the relation between non-aggregated variables. Aitkin and Longford (1986), in describing the analysis of educational data, with students nested within schools, show that it is misleading to aggregate student variables to form school means for a "means on means" analysis. The same conclusion is reached when the hierarchical structure of the data is ignored. The effect can be seen in high collinearity among predictors and as large standard errors for estimates.

In addition, standard fixed parameter regression models do not allow for the exploration of variation between groups, which may be of interest in its own right. Examples include the variation between schools or, in the case of the previous example, between individuals. Correlations induced by the hierarchy will also be ignored, thereby biasing the standard errors estimated (Rasbash, 1993).

In the situation where different groups or individuals each has a different regression model, one can assume that the intercepts and slopes are a random sample from a population or group of individual intercepts and slopes. This will give a random coefficient regression model.

It is still not possible, however, to include variables from a higher level of the hierarchy. To facilitate this, multilevel models can be used.

2.3 MULTILEVEL MODELS

The term *multilevel* refers to a hierarchical or nested relationship among units in a system. Multilevel analysis allows characteristics of different groups to be incorporated into models of individual behaviour. Returning to the field of education, students may for instance be regarded as level-1 units in a hierarchy, nested within schools (level-2 units) in turn nested within educational departments (level-3). Each of the levels in this structure is formally represented by its own submodel. These submodels express relationships among variables within a given level. They also specify how variables at one level may influence relations occurring at another, as reported by Bryk and Raudenbush (1992).

Example 2.3.1

Consider the data set consisting of dental measurements at different ages as discussed in Section 2.2. In this case the individual children are the level-2 units and the dental measurements at different ages the level-1 units. There are four level-1 units nested within each level-2 unit and there are 27 level-2 units.

From du Toit (1993), it follows that the assumption that the regression coefficients $\boldsymbol{\beta}$ vary from one individual to another, may be accommodated by regarding the unknown regression parameters as random variables with mean $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Phi}$.

The model for the 27 level-2 units can then be defined as

$$\mathbf{y}_i = \mathbf{X}\,\mathbf{b}_i + \mathbf{e}_i\ , \qquad i = 1, 2, \ldots, 27 \tag{2.3.1}$$

where the 4×2 matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}$$

with the first column denoting the intercept term and the second column giving the ages at which measurements were made.

It is assumed that $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{27}$ are a random sample from a multivariate normal distribution with

$$\mathrm{E}(\mathbf{b}_i) = \boldsymbol{\beta}$$

and

$$\mathrm{Cov}(\mathbf{b}_i, \mathbf{b}_i') = \boldsymbol{\Phi} \, .$$

The vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{27}$ are assumed to be independently and identically distributed as $\mathrm{N}(\mathbf{0}, \sigma^2\, \mathbf{I})$ independent of $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{27}$. Under these assumptions it follows that

$$\mathrm{E}(\mathbf{y}_i) = \mathbf{X}\, \boldsymbol{\beta} \qquad (2.3.2)$$

and

$$\mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i') = \mathbf{X}\, \boldsymbol{\Phi}\, \mathbf{X}' + \sigma^2\, \mathbf{I} \, . \qquad (2.3.3)$$

If, however, the gender of an individual is to be taken into account, (2.3.1) can be rewritten as

$$\mathbf{y}_i = \mathbf{X}\, \mathbf{b}_i + \mathbf{e}_i \, ,$$

where

$$\mathbf{b}_i = \begin{bmatrix} \beta_0 + u_{i1} \\ \beta_1 + u_{i2} \\ \beta_2 + 0 \end{bmatrix} \tag{2.3.4}$$

so that

$$\mathbf{y}_i = \mathbf{X}_{(f)}\,\boldsymbol{\beta} + \mathbf{X}_{(2)}\,\mathbf{u}_i + \mathbf{e}_i\,,\; i = 1, 2, \ldots, 27\,. \tag{2.3.5}$$

The matrix $\mathbf{X}_{(f)}$ is the design matrix for the fixed part of the model. If gender is coded '1' for boys and '-1' for girls, the matrix $\mathbf{X}_{(f)}$, in the case of a female, is given by

$$\mathbf{X}_{(f)} = \begin{bmatrix} 1 & 8 & -1 \\ 1 & 10 & -1 \\ 1 & 12 & -1 \\ 1 & 14 & -1 \end{bmatrix}.$$

The vector $\mathbf{b}_i$, as given by (2.3.4), defines a model with intercept and slope coefficients which are allowed to vary over the units. A fixed gender effect is also included. The vector $\boldsymbol{\beta}$ as given in (2.3.5) contains coefficients for the fixed part of the model, while the vector $\mathbf{u}_i$ contains those coefficients allowed to vary over level-2 units.

The matrix $\mathbf{X}_{(2)}$ is the random parameter matrix on level-2 of the model and is given by

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}.$$

Let

$$E(\mathbf{u}_i) = \mathbf{0}$$

and

$$\mathrm{Cov}(\mathbf{u}_i, \mathbf{u}_i') = \mathbf{\Phi}_{(2)}$$

while $\mathbf{e}_1$, $\mathbf{e}_2$, . . . , $\mathbf{e}_{27}$ are assumed to be identically and independently $\mathrm{N}(\mathbf{0}, \sigma^2\, \mathbf{I})$. Then (cf. (2.3.2) and (2.3.3))

$$\mathrm{E}(\mathbf{y}_i) = \mathbf{X}_{(f)}\, \boldsymbol{\beta} \tag{2.3.6}$$

and

$$\mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i') = \mathbf{X}_{(2)}\, \mathbf{\Phi}_{(2)}\, \mathbf{X}_{(2)}' + \sigma^2\, \mathbf{I}\,. \tag{2.3.7}$$

Fitting of this model using the program ML3E (Prosser, Rasbash & Goldstein, 1991) gives the following computer output:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\boldsymbol{\beta}}$	**STD. ERR**	**Z-VALUE**	**PR>|Z|**
CONS	16.560	0.8151	20.3165	0.0000
SLOPE	0.660	0.0699	9.4421	0.0000
GENDER	1.073	0.3644	2.9446	0.0032

(ii) **Random part of the model:**

Level-2:

PARAMETER	$\hat{\mathbf{u}}$	**STD. ERR**	**Z-VALUE**	**PR>|Z|**
CONS/CONS	7.0030	5.2860	1.3248	0.1852
SLOPE/CONS	-0.4325	0.4366	-0.9906	0.3219
SLOPE/SLOPE	0.0462	0.0395	1.1696	0.2422

Level-1:

CONS/CONS	1.7160	0.3303	5.1953	0.0000

From the fixed part of the output it can be seen that all the estimates of the fixed parameters are highly significant at a 5 % level of significance. From the random part of the output it follows that the estimates on level-2 of the model are not significant at a 5 % level of significance. It can however, not be concluded that there is no stochastic component at this level, as the number of level-2 units is only 27. The estimated error variance (level-1) is 1.7160. This estimate of error variance is considerably lower than the estimated error variance of 6.5609 obtained in Section 2.2 through use of a fixed parameter linear model. □

According to Goldstein (1991), "multilevel models not only provide more efficient estimates than traditional approaches, they also allow the exploration of variation between clusters which may be of interest in its own right". The output obtained for this example seems to be in concurrence with this statement.

It has also been pointed out by Holt, Scott and Ewings (1980), that, if complex survey data are analysed under the assumption that the data were obtained from a simple random sampling scheme, serious inferential errors may be made. Stratification variables from a complex sample design can be incorporated into a multilevel model as fixed explanatory variables. Thus their additive and interactive effects may be studied (Goldstein, 1991).

When the dependence among individual responses within the same unit is disregarded, incorrectly estimated standard errors may occur. This problem is resolved by including a unique random effect for each unit in the statistical model, which is the case with a multilevel model. The variability in these random effects is taken into account in estimating standard errors. According to Bryk and Raudenbush (1992), "these standard error estimates adjust for the intraclass correlation (or related to it, the design effect) that occurs as a result of cluster sampling".

In the User's Guide to ML3E (Prosser, Rasbash & Goldstein, 1991) the following advantages of multilevel modelling are given. Firstly, coefficients at one level of the hierarchy can be viewed as variables that are functions of characteristics of units at another level. Secondly, the coefficients of within-unit relations among variables may be estimated more accurately than in the case of the same coefficients in a single-level analysis for each group. Thirdly, the more appropriate model specification helps to resolve the problem of incorrectly estimated precision which occurs in the single level analysis of data with a hierarchical structure. Finally, the variances and covariances of coefficients may be of interest in their own right.

In conclusion it is noted that multilevel models may obviate the necessity of creating large numbers of dummy variables, a procedure which may make the interpretation of results rather difficult. Multilevel models have the capability to handle unbalanced data and missing responses. They allow for covariance components on different levels of the hierarchy and both discrete and continuous outcomes may be studied successfully. Given that discrete or continuous covariates may also be included at the different levels of the analysis, it is clear that multilevel modelling provides a flexible alternative to the classical analysis of variance.

2.4 PARAMETER ESTIMATION FOR MULTILEVEL MODELS

In a multilevel model it is possible to estimate fixed effects, random coefficients on the different levels of the hierarchy and variance-covariance components. Closed form mathematical formulas to estimate the variance and covariance components are only available for perfectly balanced designs. Numerical procedures must be used to obtain efficient estimates for unbalanced designs.

A number of approaches to the problem of obtaining estimates for unbalanced data have been developed in recent years (Bryk & Raudenbush, 1992). Among these are the full maximum likelihood approach (Goldstein, 1986 and Longford, 1987), restricted maximum likelihood approach (Mason *et al*, 1983 and Raudenbush & Bryk, 1986), and the EM algorithm developed by Dempster, Laird and Rubin (1977). Other numerical approaches include the use of IGLS algorithm (Goldstein, 1986), and the use of a Fisher scoring algorithm (Longford, 1987). A number of these algorithms have been implemented in computer programs. Some of these programs will be discussed at the end of this section.

The use of the EM algorithm for the estimation of unknown coefficients will be discussed in detail in Chapter 3. The Iterative Generalized Least Squares procedure, of which a description is given below, will be used in Chapters 4 to 7.

The general level-2 model can be rewritten (du Toit, 1993), as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{e}_i^* . \tag{2.4.1}$$

The vector $\boldsymbol{\beta}$ is a vector of fixed coefficients, while $\mathbf{e}_i^*$ is a vector of variables allowed to vary over level-1 and level-2 of the hierarchy. The matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ denote the fixed and random parameter design matrices respectively.

Let $\mathbf{X}_{(f)i} = \mathbf{X}_i$ and let

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{X}_{(2)i} & \mathbf{X}_{(1)i} \end{bmatrix}, \tag{2.4.2}$$

where $\mathbf{X}_{(2)i}$ and $\mathbf{X}_{(1)i}$ are the level-2 and level-1 random parameter design matrices. Also let

$$\mathbf{e}_i^* = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{e}_i \end{bmatrix},$$

where $\mathbf{u}_i$ and $\mathbf{e}_i$ denote the vectors of random parameters on level-2 and level-1 of the model. The model (cf. (2.3.5))

$$\mathbf{y}_i = \mathbf{X}_{(f)i} \boldsymbol{\beta} + \mathbf{X}_{(2)i} \mathbf{u}_i + \mathbf{X}_{(1)i} \mathbf{e}_i$$

is therefore equivalent to (2.4.1).

It is assumed that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$ are a random sample from a $\mathrm{N}(\mathbf{0}, \boldsymbol{\Phi}_{(2)})$ random variable and that $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$ are an independent random sample of a $\mathrm{N}(\mathbf{0}, \boldsymbol{\Phi}_{(1)})$ random variable.

Under the distributional assumptions given above, it follows that $\text{Cov}(\mathbf{y}_i, \mathbf{y}_i')$, hereafter denoted by Σ_i, is given by

$$\Sigma_i = \text{Cov}(\mathbf{y}_i, \mathbf{y}_i') = \mathbf{Z}_i \begin{bmatrix} \mathbf{\Phi}_{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{(1)} \end{bmatrix} \mathbf{Z}_i' . \tag{2.4.3}$$

Let $\mathbf{V}_i$ be a consistent estimator of Σ_i and consider the quadratic function

$$Q_{\beta} = \sum_{i=1}^{N} [\mathbf{y}_i - \mathbf{X}_{(f)i}\,\boldsymbol{\beta}]'\, \mathbf{V}_i^{-1}\, [\mathbf{y}_i - \mathbf{X}_{(f)i}\,\boldsymbol{\beta}] . \tag{2.4.4}$$

If it is assumed that $\mathbf{V}_i$ is known, the generalized least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained as the minimum of (2.4.4) and is given by

$$\hat{\boldsymbol{\beta}} = [\sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{X}_{(f)i}]^{-1} [\sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{y}_i] \tag{2.4.5}$$

with

$$\text{Cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}') = [\sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{X}_{(f)i}]^{-1} .$$

Suppose further that $\boldsymbol{\beta}$ is known and let

$$\mathbf{Y}_i^* = (\mathbf{y}_i - \mathbf{X}_{(f)i}\,\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_{(f)i}\,\boldsymbol{\beta})'.$$

Then

$$\text{E}(\mathbf{Y}_i^*) = \Sigma_i \tag{2.4.6}$$

and from (2.4.3) and (2.4.2)

$$\mathbf{V}_i = \mathbf{X}_{(2)i}\, \hat{\mathbf{\Phi}}_{(2)}\, \mathbf{X}'_{(2)i} + \mathbf{X}_{(1)i}\, \hat{\mathbf{\Phi}}_{(1)}\, \mathbf{X}'_{(1)i} . \tag{2.4.7}$$

Using the result (see for example Browne, 1974)

$$\text{vec}(\mathbf{C}\ \mathbf{A}\ \mathbf{C}') = \mathbf{C} \otimes \mathbf{C}\ \text{vec}\ \mathbf{A} ,$$

it follows that the vector of elements of $\mathbf{V}_i$ is given by (cf. (2.4.7))

$$\text{vec } \mathbf{V}_i = (\mathbf{X}_{(2)i} \otimes \mathbf{X}_{(2)i}) \text{ vec } \boldsymbol{\Phi}_{(2)} + (\mathbf{X}_{(1)i} \otimes \mathbf{X}_{(1)i}) \text{ vec } \boldsymbol{\Phi}_{(1)} \,. \tag{2.4.8}$$

There is an unique matrix (see for example Browne, 1974) $\mathbf{G}_m : m^2 \times \frac{1}{2}\, m(m+1)$ such that

$$\text{vec } \mathbf{A} = \mathbf{G}_m \text{ vecs } \mathbf{A} \tag{2.4.9}$$

with $\mathbf{A}$ a symmetric $m \times m$ matrix.

There is a non-unique matrix $\mathbf{H}_p : \frac{1}{2}\, p(p+1) \times p^2$ such that

$$\text{vecs } \mathbf{A} = \mathbf{H}_p \text{ vec } \mathbf{A} \,. \tag{2.4.10}$$

By using (2.4.9) and (2.4.10), (2.4.8) can be rewritten as

$$\begin{aligned} \text{vecs } \mathbf{V}_i &= \mathbf{H}_{n_i^*}[\left(\mathbf{X}_{(2)i} \otimes \mathbf{X}_{(2)i}\right) \mathbf{G}_q \text{ vecs } \boldsymbol{\Phi}_{(2)} + \left(\mathbf{X}_{(1)i} \otimes \mathbf{X}_{(1)i}\right) \mathbf{G}_m \text{ vecs } \boldsymbol{\Phi}_{(1)}] \\ &= \mathbf{H}_{n_i^*}\left[\; \left(\mathbf{X}_{(2)i} \otimes \mathbf{X}_{(2)i}\right) \mathbf{G}_q \quad \left(\mathbf{X}_{(1)i} \otimes \mathbf{X}_{(1)i}\right) \mathbf{G}_m \;\right] \begin{bmatrix} \text{vecs } \boldsymbol{\Phi}_{(2)} \\ \text{vecs } \boldsymbol{\Phi}_{(1)} \end{bmatrix} \end{aligned} \tag{2.4.11}$$

where $n_i^* = \sum_{j=1}^{n_i} n_{ij}$. Let

$$\mathbf{X}_i^* = \mathbf{H}_{n_i^*}\left[\; \left(\mathbf{X}_{(2)i} \otimes \mathbf{X}_{(2)i}\right) \mathbf{G}_q \quad \left(\mathbf{X}_{(1)i} \otimes \mathbf{X}_{(1)i}\right) \mathbf{G}_m \;\right]$$

and

$$\boldsymbol{\tau} = \begin{bmatrix} \text{vecs } \boldsymbol{\Phi}_{(2)} \\ \text{vecs } \boldsymbol{\Phi}_{(1)} \end{bmatrix},$$

then (2.4.11) can be written as

$$\text{vecs } \mathbf{V}_i = \mathbf{X}_i^* \, \boldsymbol{\tau} \,. \tag{2.4.12}$$

Let

$$\mathbf{y}^* = \text{vecs } \mathbf{Y}^* \tag{2.4.13}$$

so that (cf. (2.4.6))

$$\text{E}(\mathbf{y}^*) = \text{vecs } \Sigma \tag{2.4.14}$$

where Σ denotes the true population covariance matrix.

Furthermore, let $\mathbf{W}$ denote the covariance matrix of vecs $\hat{\Sigma}$. It can be shown (du Toit, 1993) that

$$\mathbf{W}_i^{-1} = \frac{N}{2}\, \mathbf{G}'_{n_i}\, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\, \mathbf{G}_{n_i} \tag{2.4.15}$$

is a consistent estimator of $\text{Cov}(\mathbf{y}_i^*, \mathbf{y}_i^{*\prime})$.

Now consider the quadratic form

$$\mathbf{Q}_T = \sum_{i=1}^{N} \left\{ (\mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\tau})' \, \mathbf{W}_i^{-1} (\mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\tau}) \right\} . \tag{2.4.16}$$

Minimization of $\mathbf{Q}_\tau$ with respect to $\boldsymbol{\tau}$ yields

$$\hat{\boldsymbol{\tau}} = [\, \sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{X}_i^* \,]^{-1} \,[\, \sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^* \,] \tag{2.4.17}$$

where $\mathbf{y}^* = \text{vecs } \mathbf{Y}^*$ (cf. (2.4.13)).

The estimate of $\boldsymbol{\beta}$ (cf. (2.4.5)) is based on the assumption that $\mathbf{V}$ is known. Note that the estimation of $\boldsymbol{\tau}$ (cf. (2.4.17)) is based on the assumption that $\mathbf{W}$ is a consistent estimator of the covariance of vecs $(\mathbf{Y}^*)$ and that $\boldsymbol{\beta}$ is known, since $\boldsymbol{\beta}$ is required to evaluate $\mathbf{y}^*$.

To obtain the IGLS estimates of the unknown parameters, the following iterative procedure is followed:

(i) Set $\widehat{\mathbf{V}}_i = \mathbf{I}$ in (2.4.5).

(ii) Calculate an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

(iii) Calculate $\mathbf{W}$ as given by (2.4.15).

(iv) Calculate $\widehat{\boldsymbol{\tau}}$ using (2.4.17).

(v) Obtain a revised estimate of $\mathbf{V}$ using (2.4.11).

Steps (ii) to (v) are repeated until convergence is obtained, for example until $|\widehat{\boldsymbol{\tau}}_{k+1} - \widehat{\boldsymbol{\tau}}_k| < \epsilon$; $|\widehat{\boldsymbol{\beta}}_{k+1} - \widehat{\boldsymbol{\beta}}_k| < \boldsymbol{\epsilon}$, with $\epsilon_k = 10^{-6}$, where ϵ_k is a typical element of $\boldsymbol{\epsilon}$ (du Toit, 1993). Approximate standard errors of the elements of $\widehat{\boldsymbol{\beta}}$ are obtained as the square roots of the diagonal elements of the matrix $[\sum_{i=1}^{N} \mathbf{X}'_{(f)i} \mathbf{V}_i^{-1} \mathbf{X}'_{(f)i}]^{-1}$. Likewise, standard errors of the elements of $\widehat{\boldsymbol{\tau}}$ are obtained as the square roots of the diagonal elements of the matrix $[\sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^{*}]^{-1}$.

This algorithm is known as IGLS (see for example Goldstein, 1986). Under the assumption of multivariate normality, the IGLS estimates are equivalent to the Maximum Likelihood estimates of the corresponding unknown parameters $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\tau}}$ (see for example Browne and du Toit, 1992).

As mentioned earlier, a number of the available algorithms for the estimation of unknown parameters in hierarchical linear models have been implemented in computer programs. The names of a selection of the available programs and the algorithms they used for estimation are given in a table below, extracted from a review of multilevel programs by Kreft, de Leeuw and van der Leeden (1994).

Authors of program	Program name	Algorithms used
Jennrich and Schluster	BMDP-5V	Newton-Raphson Fisher scoring Generalized EM
Hermalin and Anderson	GENMOD	Restricted ML / EM
Bryk, Raudenbush and Congdon	HLM	EM / ML
Prosser, Rasbash and Goldstein	ML3E	IGLS
Longford	VARCL	Fisher scoring / ML

Most of the programs listed are capable of analysing a level-2 hierarchical model. The program ML3E is capable of handling three-level hierarchical data. The size of data sets which can be analysed with the program ML3E, however, is somewhat limited (Kreft, de Leeuw and van der Leeden, 1994). The authors nevertheless recommended ML3E as the most appropriate program for people doing research in multilevel analysis. Since this article was published, the SAS Institute have also included a multilevel program called MIXED in their statistical analysis package. From this information the conclusion can be drawn that there is a need for a computationally efficient stand-alone procedure for level-3 models which can handle large data sets, as commonly occurs in the social sciences.

In order to implement the IGLS algorithm in a program, the size of matrices to be used in the procedure must be taken into account, as this affects both the storage space and computational time required. From (2.4.5) we see that the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is a function of both the fixed parameter design matrix $\mathbf{X}_{(f)i}$ and the inverse of the covariance matrix $\mathbf{V}_i$. The covariance matrix is of the order $n_i \times n_i$ where n_i is the number of units or respondents within the i-th level-2 unit. If $n_i = 500$, the estimation of $\boldsymbol{\beta}$ will entail the inversion of a 500×500 matrix which requires 500^3 operations. The storage space required to store this inverse matrix in memory is also large.

From (2.4.11) it follows that, in order to obtain an estimator $\hat{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$, the Kronecker products $\mathbf{X}_{(2)i} \otimes \mathbf{X}_{(2)i}$ and $\mathbf{X}_{(1)i} \otimes \mathbf{X}_{(1)i}$ must be calculated. The order of the level-2 random parameter design matrix $\mathbf{X}_{(2)i}$ is $n_i \times m$, where m indicates the number of random coefficients on level-2. The Kronecker product is thus of the order $n_i^2 \times m^2$, which, for $n_i = 500$ and $m = 5$, gives a matrix with $500^2 \times 5^2$ elements.

It is clear that, in order to implement the IGLS algorithm in a computer program, expressions containing matrices or submatrices of lower order are needed if the program is to be computationally efficient. Ways must be found to minimize the storage requirements of such a program. Ideally, a program should also be capable of handling large data sets and facilitating complex variation on all levels of the model.

In Chapters 3 to 7 the simplification of expressions needed for the successful implementation of both the EM algorithm and the IGLS algorithm are proposed.

Results will be incorporated in computer programs. Applications will be given in order to illustrate the estimation and data size capabilities of the programs.

2.5 SUMMARY

In this chapter the concept of multilevel modelling was introduced. In Section 2.2 the standard fixed parameter linear regression model and the disadvantages of such an approach to the analysis of hierarchically structured data were discussed. This was followed in Section 2.3 with a definition and discussion of a level-2 hierarchical model. The. advantages of a multilevel modelling approach were considered.

Section 2.4 provided a brief overview of estimation procedures that may be used for the analysis of unbalanced hierarchical data. The IGLS algorithm was considered for a level-2 model. Computer programs for the analysis of multilevel data were briefly discussed and the difficulties concerning the implementation of algorithms noted.

In the next chapter a level-3 model will be defined and the EM algorithm, used for the estimation of the unknown parameters, will be derived.

CHAPTER 3
THE EM ALGORITHM FOR PARAMETER ESTIMATION OF LEVEL-3 MODELS

3.1 INTRODUCTION

In many fields of application, such as psychology, sociology, medicine and economics, data commonly have a nested structure which may contain more than two levels. Examples of this type of data are

(i) comparison of the wages of factory employees in nine provinces, where provinces are the level-3 units, factories the level-2 units and employees the level-1 units; and

(ii) the matriculation results of pupils in private and public schools falling under the various education departments. In this case the education departments are the level-3 units, the schools the level-2 units and the pupils the level-1 units.

In Section 2.3 of Chapter 2 properties of level-2 models were reviewed. The more complex structures of a level-3 model are considered in detail in this chapter. In Section 3.2 the level-3 model is defined and distributional assumptions given.

Section 3.3 contains relevant concepts of the Bayesian approach to statistical inference. The development of the EM algorithm by Dempster, Laird and Rubin (1977), provided a feasible approach to the estimation of variance and covariance components. The applicability of this approach to hierarchical structures was demonstrated by Dempster *et al* (1981). The Expected Maximization algorithm used to obtain estimates of the unknown parameters is discussed in Section 3.4. A practical application of the EM algorithm for the level-3 model is given in Section 3.5.

3.2 EXTENSION OF A LEVEL-2 TO A LEVEL-3 MODEL

Consider data collected from N schools, each of which contains n_i students. Suppose that the relationship between an individual's achievement (y) and his or her intelligence quotient (x) is to be investigated. A linear relationship between these variables for the i-th school is given by (see for example du Toit, 1993),

$$y_{ij} = b_{0i} + b_{1i}\, x_{ij} + e_{ij}\,, \quad i = 1, 2, \ldots, N; \;\; j = 1, 2, \ldots, n_i \;. \qquad (3.2.1)$$

In (3.2.1) y_{ij} and x_{ij} denote the achievement and IQ respectively for student j in school i. (Note: the students are regarded as level-1 units and the schools as level-2 units).

The term e_{ij} is a random variable assumed to have a mean of zero, while the terms b_{0i} and b_{1i} are treated as random variables at level-2 since, in general, they can vary across schools.

In the notation of Goldstein and MacDonald (1988), b_{0i} and b_{1i} can be expressed in terms of fixed and random components as

$$b_{0i} = \beta_0 + u_{0i} \qquad (3.2.2)$$

and

$$b_{1i} = \beta_1 + u_{1i} \;. \qquad (3.2.3)$$

From (3.2.2) and (3.2.3), expression (3.2.1) can be written in matrix notation as follows:

$$\mathbf{y}_i = \mathbf{X}_i\, \mathbf{b}_i + \mathbf{e}_i \qquad (3.2.4)$$

where

$$\mathbf{X}_i = \begin{bmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{in_i} \end{bmatrix}, \qquad (3.2.5)$$

$$\mathbf{b}_i = \begin{bmatrix} b_{0j} \\ b_{1j} \end{bmatrix}, \qquad (3.2.6)$$

and where the $n_i \times 1$ vector of responses $\mathbf{y}_i$ has typical element y_{ij} and the vector $\mathbf{e}_i$ has typical element e_{ij} .

The following distributional assumptions are usually made:

1. $\mathbf{X}_i$ is a matrix of fixed inputs while $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ is a non-random parameter vector with elements that have to be estimated

2. $\mathbf{e}_i \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$

3. $\mathbf{u}_i = \begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Phi})$

4. $\mathrm{Cov}(\mathbf{e}_i, \mathbf{u}_i') = \mathbf{0}$

Under these assumptions it follows that

$$\mathbf{y}_i \sim \mathrm{N}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{X}_i \boldsymbol{\Phi} \mathbf{X}_i' + \sigma^2 \mathbf{I}_{n_i}) .$$

If students at schools belonging to different departments of education are to be considered, the level-2 model defined by (3.2.4) to (3.2.6) may be extended. In this case the departments of education will be referred to as level-3 units, the schools as level-2 units and the students as level-1 units.

To accommodate variation of the random coefficients over level-2 and level-3 of the hierarchy, the following model is considered:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij} \mathbf{b}_{ij} + \mathbf{e}_{ij}, \quad i = 1, 2, \ldots, N; \; j = 1, 2, \ldots, n_i \tag{3.2.7}$$

where

$\mathbf{y}_{ij} : n_{ij} \times 1$ is a vector of responses with typical element y_{ijk},
$\mathbf{X}_{ij} : n_{ij} \times m$ is a design matrix for the (i,j)-th unit and
$\mathbf{b}_{ij} : m \times 1$ is a vector of random coefficients.

To allow for variation across level-3 units, let

$$\mathbf{b}_{ij} = \mathbf{S}_i \mathbf{c}_i + \mathbf{u}_{ij} , \tag{3.2.8}$$

where $\mathbf{S}_i : m \times q$ is a design matrix for the i-th unit at the third level and $\mathbf{c}_i : q \times 1$ is a vector of random coefficients varying across level-3. It is assumed that

$$\mathbf{c}_i = \boldsymbol{\beta} + \mathbf{v}_i \tag{3.2.9}$$

where $\boldsymbol{\beta}$ is a vector of fixed coefficients. It is assumed that $\mathbf{e}_{11}, \mathbf{e}_{12}, \ldots, \mathbf{e}_{Nn_N}$ are independently distributed; $\mathbf{u}_{11}, \mathbf{u}_{12}, \ldots, \mathbf{u}_{Nn_N}$ are a random sample of the random variate $\mathbf{u}$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are a random sample of the random variate $\mathbf{v}$.

In addition, it is assumed that

$$\begin{aligned} \mathbf{e}_i &\sim \mathrm{N}(\mathbf{0}, \Phi_{(1)} \mathbf{I}_{n_i}) \\ \mathbf{u} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Phi}_{(2)}) \\ \mathbf{v} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Phi}_{(3)}) \end{aligned} \tag{3.2.10}$$

and that

$$\begin{aligned} \mathrm{Cov}(\mathbf{e}_i, \mathbf{u}') &= \mathbf{0} \\ \mathrm{Cov}(\mathbf{e}_i, \mathbf{v}') &= \mathbf{0} \\ \mathrm{Cov}(\mathbf{u}, \mathbf{v}') &= \mathbf{0}\,. \end{aligned}$$

From (3.2.8) and (3.2.9), it follows that (3.2.7) can be expressed as

$$\mathbf{y}_{ij} = \mathbf{X}^*_{ij} \boldsymbol{\beta} + \mathbf{Z}^*_{ij} \mathbf{e}^*_{ij} \tag{3.2.11}$$

where

$$\mathbf{X}^*_{ij} = \mathbf{X}_{ij} \mathbf{S}_i\,, \tag{3.2.12}$$

$$\mathbf{Z}^*_{ij} = \begin{bmatrix} \mathbf{X}_{(3)ij} & \mathbf{X}_{(2)ij} & \mathbf{X}_{(1)ij} \end{bmatrix}, \tag{3.2.13}$$

with

$$\mathbf{X}_{(3)ij} = \mathbf{X}_{ij}\,\mathbf{S}_i\,,$$

$$\mathbf{X}_{(2)ij} = \mathbf{X}_{ij}\,,$$

$$\mathbf{X}_{(1)ij} = \mathbf{I}_{n_i}\,,$$

and

$$\mathbf{e}^*_{ij} = \begin{bmatrix} \mathbf{v}_i \\ \mathbf{u}_{ij} \\ \mathbf{e}_{ij} \end{bmatrix}. \qquad (3.2.14)$$

Using the distributional assumptions given in (3.2.10), it follows that

$$\begin{aligned} \mathrm{E}(\mathbf{y}_{ij}) &= \mathbf{X}^*_{ij}\,\boldsymbol{\beta} \\ &= \mathbf{X}_{(3)ij}\,\boldsymbol{\beta} \end{aligned} \qquad (3.2.15)$$

and

$$\Sigma_{ij} = \mathrm{Cov}(\,\mathbf{y}_{ij},\,\mathbf{y}'_{ij})$$

$$= \mathbf{X}_{(3)ij}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)ij} + \mathbf{X}_{(2)ij}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{X}'_{(2)ij} + \mathbf{X}_{(1)ij}\,\boldsymbol{\Phi}_{(1)}\,\mathbf{X}'_{(1)ij}\,. \qquad (3.2.16)$$

Example 3.2.1

Suppose that five pupils are selected randomly from school 1 falling under education department 1. The response variable (y) is the pupil's knowledge with regard to the transmission of AIDS through proven means. As a possible linear predictor (x) the age of the pupils is used. In addition, it is also known that 0.75 hours is spent weekly on sex education at schools falling under this department. In general, let s_{ij} denote the time spent weekly on sex education at school j falling under education department i.

From (3.2.7) it follows that the vector of responses can be written as

$$\begin{bmatrix} y_{11,1} \\ y_{11,2} \\ y_{11,3} \\ y_{11,4} \\ y_{11,5} \end{bmatrix} = \begin{bmatrix} 1 & x_{11,1} \\ 1 & x_{11,2} \\ 1 & x_{11,3} \\ 1 & x_{11,4} \\ 1 & x_{11,5} \end{bmatrix} \begin{bmatrix} b_{11,0} \\ b_{11,1} \end{bmatrix} + \begin{bmatrix} e_{11,1} \\ e_{11,2} \\ e_{11,3} \\ e_{11,4} \\ e_{11,5} \end{bmatrix}.$$

The vector $\mathbf{b}_{ij}$ (cf. (3.2.8)) can be expressed as

$$\begin{bmatrix} b_{11,0} \\ b_{11,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & s_{11} & 0 \\ 0 & 1 & 0 & s_{11} \end{bmatrix} \begin{bmatrix} c_{1,0} \\ c_{1,1} \\ c_{1,2} \\ c_{1,3} \end{bmatrix} + \begin{bmatrix} u_{11,0} \\ u_{11,1} \end{bmatrix},$$

while $\mathbf{c}_i$ (cf. (3.2.9)) can be expressed as

$$\begin{bmatrix} c_{1,0} \\ c_{1,1} \\ c_{1,2} \\ c_{1,3} \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \\ \alpha_1 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} v_{1,0} \\ v_{1,1} \\ v_{1,2} \\ v_{1,3} \end{bmatrix}.$$

The parameters α_0 and β_0 represent the average value of y given $x = 0$ and the time s spent on sex education. The parameters α_1 and β_1 represent the increase or decrease in y for a unit increase in x.

If, for example, the vector of response $\mathbf{y}_{11}$ and the design matrix $\mathbf{X}_{11}$ are given by

$$\mathbf{y}_{11} = \begin{bmatrix} 74 \\ 71 \\ 65 \\ 81 \\ 91 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_{11} = \begin{bmatrix} 1 & 16 \\ 1 & 13 \\ 1 & 13 \\ 1 & 15 \\ 1 & 17 \end{bmatrix},$$

the matrix $\mathbf{X}^*_{11}$ is calculated as (cf. (3.2.12))

$$\mathbf{X}^*_{11} = \begin{bmatrix} 1 & 16 \\ 1 & 13 \\ 1 & 13 \\ 1 & 15 \\ 1 & 17 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0.75 & 0 \\ 0 & 1 & 0 & 0.75 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 16 & 0.75 & 12.00 \\ 1 & 13 & 0.75 & 9.75 \\ 1 & 13 & 0.75 & 9.75 \\ 1 & 15 & 0.75 & 11.25 \\ 1 & 17 & 0.75 & 12.75 \end{bmatrix}$$

and (cf. (3.2.13))

$$\mathbf{Z}^*_{11} = \left[\ \mathbf{X}^*_{11} \quad \mathbf{X}_{11} \quad \mathbf{I} \ \right] .$$

3.3 PRINCIPLES OF BAYESIAN ANALYSIS

A short review of standard results normally employed in Bayesian analysis (see for example Degroot, 1970), is given in this section. These results are required for the derivation of the EM-algorithm discussed in the next section.

Let $f(\mathbf{y}_i, \mathbf{c}_i)$ denote a joint probability density function consisting of an $n_i \times 1$ vector $\mathbf{y}_i$ of random observations and of a $q \times 1$ vector $\mathbf{c}_i$ of random parameters. The joint probability density function can be written as

$$f(\mathbf{y}_i, \mathbf{c}_i) = f(\mathbf{y}_i | \mathbf{c}_i)\ g_i(\mathbf{c}) \qquad (3.3.1)$$

$$= p(\mathbf{c}_i | \mathbf{y}_i)\ h(\mathbf{y}_i) \qquad (3.3.2)$$

where $h(\mathbf{y}_i)$ is the marginal distribution of $\mathbf{y}_i$, $g_i(\mathbf{c})$ is the probability density function of $\mathbf{c}_i$.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

From equations (3.3.1) and (3.3.2), it follows that

$$p\,(\mathbf{c}_i|\,\mathbf{y}_i) = \frac{f\,(\mathbf{y}_i|\,\mathbf{c}_i)\;g_i\,(\mathbf{c})}{h\,(\mathbf{y}_i)}\;. \tag{3.3.3}$$

Using standard results on conditional distributions, the marginal distribution of $\mathbf{y}_i$ can be expressed as

$$h\,(\mathbf{y}_i) \;=\; \int_{\mathbf{c}} f\,(\mathbf{y}_i|\,\mathbf{c}_i)\;g_i\,(\mathbf{c})\;\mathrm{d}\mathbf{c}\;, \tag{3.3.4}$$

where $\int_{\mathbf{c}}$ denotes $\int\limits_{c_1=-\infty}^{\infty}\int\limits_{c_2=-\infty}^{\infty}\cdots\int\limits_{c_1=-\infty}^{\infty}$.

The posterior density function (pdf), which is used in the Bayesian approach to make inferences about parameters, is given by $p\,(\mathbf{c}_i|\,\mathbf{y}_i)$. The pdf $g_i\,(\mathbf{c})$ is known as the prior pdf of $\mathbf{c}_i$. The posterior pdf $p\,(\mathbf{c}_i|\,\mathbf{y}_i)$ contains the sample information in the likelihood function $f\,(\mathbf{y}_i|\,\mathbf{c}_i)$ and the prior information in the pdf $g_i\,(\mathbf{c})$.

Consider the multilevel model (cf. Section 3.2)

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{i1} \\ \cdot \\ \mathbf{y}_{ij} \\ \cdot \\ \cdot \\ \mathbf{y}_{in_i} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{(2)i1}\,\mathbf{b}_{i1} \\ \cdot \\ \mathbf{X}_{(2)ij}\,\mathbf{b}_{ij} \\ \cdot \\ \cdot \\ \mathbf{X}_{(2)in_i}\,\mathbf{b}_{in_i} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{i1} \\ \cdot \\ \mathbf{e}_{ij} \\ \cdot \\ \cdot \\ \mathbf{e}_{in_i} \end{bmatrix} \tag{3.3.5}$$

where $\mathbf{y}_{ij}$ is an $n_{ij}\times 1$ vector with typical element y_{ijk} and $\mathbf{X}_{ij}$ is an $n_{ij}\times m$ design matrix for the (i,j)-th unit.

From Section 3.2 (cf. (3.2.8))

$$\mathbf{b}_{ij} \;=\; \mathbf{S}_i\,\mathbf{c}_i \;+\; \mathbf{u}_{ij}\;,$$

where (cf. (3.2.9))

$$\mathbf{c}_i = \boldsymbol{\beta} + \mathbf{v}_i \,.$$

The elements of the $m \times 1$ vector $\mathbf{b}_{ij}$ are considered to be random, varying across the level-2 and level-3 units.

Under the assumptions given in Section 3.2, it follows that

$$\mathbf{y}_i \sim \mathrm{N}(\mathbf{X}_{(3)i}\,\boldsymbol{\beta} \,,\, \Sigma_i\,)$$

where

$$\Sigma_i \;=\mathbf{X}_{(3)i}\;\boldsymbol{\Phi}_{(3)i}\,\mathbf{X}'_{(3)i}\;+\;\boldsymbol{\Lambda}_i \qquad (3.3.6)$$

and $\boldsymbol{\Lambda}_i$ is a diagonal matrix with typical element

$$\boldsymbol{\Lambda}_{ij} \;=\; \Phi_{(1)}\,\mathbf{I}_{n_{ij}} \;+\; \mathbf{X}_{(2)ij}\;\boldsymbol{\Phi}_{(2)}\,\mathbf{X}'_{(2)ij}\;.$$

Since (cf. (3.2.9), Section 3.2)

$$\mathbf{c}_i \;\sim\; \mathrm{N}(\,\boldsymbol{\beta}\,,\,\boldsymbol{\Phi}_{(3)})\;,$$

the joint distribution of $\mathbf{y}_i$ and $\mathbf{c}_i$ is given by

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{c}_i \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{X}_{(3)i}\,\boldsymbol{\beta} \\ \boldsymbol{\beta} \end{pmatrix},\begin{pmatrix} \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i}+\boldsymbol{\Lambda}_i & \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)} \\ \boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} & \boldsymbol{\Phi}_{(3)} \end{pmatrix}\right).$$

Using standard results on conditional distributions, the covariance of $\mathbf{c}_i$, given $\mathbf{y}_i$, is given by

$$\mathrm{Cov}(\mathbf{c}_i\,|\,\mathbf{y}_i) = \Sigma_{cc} \;-\Sigma_{cy}\;\Sigma_{yy}^{-1}\;\Sigma_{yc}$$

$$= \Phi_{(3)} - \Phi_{(3)} X'_{(3)i} (X_{(3)i} \Phi_{(3)} X'_{(3)i} + \Lambda_i)^{-1} X_{(3)i} \Phi_{(3)} .$$

After applying the result (see, for example, du Toit, 1993)

$$(\Phi^{-1} + X' \Lambda^{-1} X)^{-1} = \Phi^{-1} - \Phi^{-1} X'(\Lambda + X \Phi X')^{-1} X \Phi \quad (3.3.7)$$

$\mathrm{Cov}(c_i \mid y_i)$ can be written as

$$\mathrm{Cov}(c_i|y_i) = (\Phi_{(3)}^{-1} + X'_{(3)i} \Lambda_i^{-1} X_{(3)i})^{-1}$$

$$= (\Phi_{(3)}^{-1} + S'_i \sum_{j=1}^{n_i} A_{ij} S_i)^{-1} \quad (3.3.8)$$

where

$$A_{ij} = X'_{(2)ij} \Lambda_{ij}^{-1} X_{(2)ij} \quad (3.3.9)$$

(see Appendix 3.1 for further simplification of this term).

Using the result (du Toit, 1993),

$$\Phi X' (\Lambda + X \Phi X')^{-1} = (X' \Lambda^{-1} X + \Phi^{-1})^{-1} X' \Lambda^{-1} , \quad (3.3.10)$$

the expected value of c_i, given y_i, may be expressed as

$$E(c_i|y_i) = \mathrm{Cov}(c_i|y_i) X'_{(3)i} \Lambda_i^{-1} (y_i - X_{(3)i} \beta) + \beta$$

$$= \mathrm{Cov}(c_i|y_i) \{ S'_i \sum_{j=1}^{n_i} b_{ij} - S'_i \sum_{j=1}^{n_i} (A_{ij}) S_i \beta \} + \beta \quad (3.3.11)$$

where

$$b_{ij} = X'_{(2)ij} \Lambda_{ij}^{-1} y_{ij} \quad (3.3.12)$$

(see Appendix 3.1 for further simplification of this term)

The vector $\mathbf{y}_i$ can be written as

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{X}_{(2)i1} (\mathbf{S}_i \mathbf{c}_i + \mathbf{u}_{i1}) + \mathbf{e}_{i1} \\ \vdots \\ \mathbf{X}_{(2)ij} (\mathbf{S}_i \mathbf{c}_i + \mathbf{u}_{ij}) + \mathbf{e}_{ij} \\ \vdots \\ \mathbf{X}_{(2)in_i} (\mathbf{S}_i \mathbf{c}_i + \mathbf{u}_{in_i}) + \mathbf{e}_{in_i} \end{bmatrix}$$

so that

$$\mathbf{y}_i = \mathbf{X}_{(2)i} \mathbf{S}_i \mathbf{c}_i + \begin{bmatrix} \mathbf{X}_{(2)i1} \mathbf{u}_{i1} \\ \vdots \\ \mathbf{X}_{(2)ij} \mathbf{u}_{ij} \\ \vdots \\ \mathbf{X}_{(2)in_i} \mathbf{u}_{in_i} \end{bmatrix} + \mathbf{e}_i \, .$$

It further follows that the covariance matrix of $\mathbf{y}_i$ and $\mathbf{b}'_{ij}$ (cf. (3.2.8)) can be written as

$$\mathrm{Cov}(\mathbf{y}_i \, , \mathbf{b}'_{ij}) = \mathrm{Cov} \left\{ \mathbf{X}_{(2)i} \mathbf{S}_i \mathbf{c}_i + \begin{bmatrix} \mathbf{X}_{(2)i1} \mathbf{u}_{i1} \\ \vdots \\ \mathbf{X}_{(2)ij} \mathbf{u}_{ij} \\ \vdots \\ \mathbf{X}_{(2)in_i} \mathbf{u}_{in_i} \end{bmatrix} , \; [\, \mathbf{c}_i{}' \mathbf{S}_i{}' + \mathbf{u}'_{ij} \,] \right\}$$

$$= \mathbf{X}_{(2)i} \, \mathbf{\Phi}_{(3)} \, \mathbf{S}_i{}' + \mathbf{C}_{ij} \, ,$$

where

$$\mathbf{C}'_{ij} = [\, \mathbf{0}' \; \ldots \; \mathbf{0}' \quad \mathbf{\Phi}_{(2)} \mathbf{X}'_{(2)ij} \quad \mathbf{0}' \; \ldots \; \mathbf{0}' \,] \; .$$

Under the distributional assumptions given in the previous section, the joint distribution of $\mathbf{y}_i$ and $\mathbf{b}_{ij}$, where

$$\mathbf{y}_i \sim \mathrm{N}(\mathbf{X}_{(3)i}\boldsymbol{\beta} , \mathbf{X}_{(3)i} \ \boldsymbol{\Phi}_{(3)} \mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i)$$

and

$$\mathbf{b}_{ij} \sim \mathrm{N}(\mathbf{S}_i \boldsymbol{\beta} , \ \mathbf{S}_i \boldsymbol{\Phi}_{(3)} \mathbf{S}'_i + \boldsymbol{\Phi}_{(2)}) ,$$

is given by (see for example Morrison (1991))

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_{ij} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{X}_{(3)i}\boldsymbol{\beta} \\ \mathbf{S}_i \boldsymbol{\beta} \end{pmatrix} ; \begin{matrix} \mathbf{X}_{(3)i}\boldsymbol{\Phi}_{(3)}\mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i & \boldsymbol{\Sigma}_{yb} \\ \boldsymbol{\Sigma}_{by} & \mathbf{S}_i \boldsymbol{\Phi}_{(3)} \mathbf{X}'_{(3)i} + \boldsymbol{\Phi}_{(2)} \end{matrix} \right)$$

with

$$\boldsymbol{\Sigma}_{by} = \mathbf{S}_i \boldsymbol{\Phi}_{(3)} \mathbf{X}'_{(3)i} + \mathbf{C}'_{ij} .$$

The covariance of $\mathbf{b}_{ij}$, given $\mathbf{y}_i$, is

$$\mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) = \mathbf{S}_i \boldsymbol{\Phi}_{(3)} \mathbf{S}'_i + \boldsymbol{\Phi}_{(2)} - \boldsymbol{\Sigma}_{by} (\mathbf{X}_{(3)i} \ \boldsymbol{\Phi}_{(3)} \mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i)^{-1} \boldsymbol{\Sigma}_{yb} .$$

Using results (3.3.7) and (3.3.10) the covariance of $\mathbf{b}_{ij}$, given $\mathbf{y}_i$, can be expressed as

$$\begin{aligned} \mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) = {} & \mathbf{S}_i \,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\, \mathbf{S}'_i + \boldsymbol{\Phi}_{(2)} - [\mathbf{S}_i \boldsymbol{\Sigma}_{cy} + \mathbf{C}'_{ij}] \times \\ & (\mathbf{X}_{(3)i} \boldsymbol{\Phi}_{(3)} \mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i)^{-1} [\boldsymbol{\Sigma}_{yc} \mathbf{S}'_i + \mathbf{C}_{ij}] \\ = {} & \mathbf{S}_i \ \boldsymbol{\Phi}_{(3)} \mathbf{S}'_i + \boldsymbol{\Phi}_{(2)} - \mathbf{S}_i \,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\, \mathbf{S}'_i \mathbf{X}'_{(2)i} \ \boldsymbol{\Lambda}_i^{-1} \mathbf{C}_{ij} - \\ & \mathbf{C}_{ij} \boldsymbol{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)i} \mathbf{S}_i \ \mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\, \mathbf{S}'_i + \mathbf{S}_i \ \mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\, \mathbf{S}'_i - \mathbf{C}_{ij} \ \boldsymbol{\Lambda}_{ij}^{-1} \ \mathbf{C}_{ij}. \end{aligned} \tag{3.3.13}$$

From (3.3.9),

$$\mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\mathbf{X}_{(2)ij} = \mathbf{A}_{ij}$$

so that

$$\begin{aligned}\mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) = {} & \mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i' + \boldsymbol{\Phi}_{(2)} - \\ & \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij}\,\mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i'\,\mathbf{A}_{ij}\,\boldsymbol{\Phi}_{(2)} - \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij}\,\boldsymbol{\Phi}_{(2)} - \\ & \mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i'\,\mathbf{A}_{ij}\,\boldsymbol{\Phi}_{(2)} - \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij}\,\mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i' \\ = {} & (\mathbf{I} - \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij})\,\mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i'\,(\mathbf{I} - \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij})' + \\ & \boldsymbol{\Phi}_{(2)}\,(\mathbf{I} - \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij}\,)\,. \end{aligned} \tag{3.3.14}$$

The expected value of $\mathbf{b}_{ij}$, given $\mathbf{y}_i$, is given by

$$\begin{aligned}\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i) &= \mathbf{S}_i\,\boldsymbol{\beta} + \Sigma_{by}\,(\,\mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \Lambda_i^{-1}\,)^{-1}\,(\,\mathbf{y}_i - \mathbf{X}_{(3)i}\,\boldsymbol{\beta}\,) \\ &= \mathbf{S}_i\,\boldsymbol{\beta} + [\mathbf{S}_i\,\Sigma_{cy} + \mathbf{C}'_{ij}\,](\,\mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \Lambda_i^{-1}\,)^{-1}\,(\,\mathbf{y}_i - \mathbf{X}_{(3)i}\,\boldsymbol{\beta}\,) \\ &= \mathbf{S}_i\,\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) + \mathbf{C}'_{ij}\,(\mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \Lambda_i^{-1}\,)^{-1}\,(\,\mathbf{y}_i - \mathbf{X}_{(3)i}\,\boldsymbol{\beta}\,)\,.\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i) = {} & \mathbf{S}_i\,\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) + \boldsymbol{\Phi}_{(2)}\,(\,\mathbf{b}_{ij} - \mathbf{A}_{ij}\,\mathbf{S}_i\,\boldsymbol{\beta}) + \\ & \boldsymbol{\Phi}_{(2)}\,\mathbf{A}_{ij}\,\mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i'\left\{\sum_{j=1}^{n_i}\mathbf{A}_{ij}\,\mathbf{S}_i\,\boldsymbol{\beta} - \sum_{j=1}^{n_i}\mathbf{b}_{ij}\right\}.\end{aligned} \tag{3.3.15}$$

3.4 PARAMETER ESTIMATION USING EXPECTED MAXIMIZATION

In this section a number of key results will be given in terms of five propositions. It will then be shown that these results may be used in an iterative procedure to obtain Maximum Likelihood (ML) estimates of the unknown parameters in the level-3 model (cf. (3.2.11) and (3.3.5)). This optimization algorithm is known as the EM algorithm.

PROPOSITION 3.4.1

Consider a row of independent random vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ and a random sample $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N$ of the random vector $\mathbf{d}$ which are independently distributed from the $\mathbf{y}_i$. Let γ denote the unknown parameters in the pdf of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ and τ the unknown parameters in the pdf of $\mathbf{d}$. Then

$$\frac{\partial \ln L}{\partial \tau_K} = \sum_{i=1}^{N} \mathrm{E}_{\mathbf{d}|\mathbf{y}_i}\left(\frac{\partial \ln g_i(\mathbf{d})}{\partial \tau_K}\right) \tag{3.4.1}$$

and

$$\frac{\partial \ln L}{\partial \gamma_K} = \sum_{i=1}^{N} \mathrm{E}_{\mathbf{d}|\mathbf{y}_i}\left(\frac{\partial \ln f(\mathbf{y}_i|\mathbf{d})}{\partial \gamma_K}\right) \tag{3.4.2}$$

where L denotes the likelihood function of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$.

Proof

From (3.3.4) it follows that

$$L = h(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N) = \prod_{i=1}^{N} \int_{\mathbf{d}} f(\mathbf{y}_i|\mathbf{d})\, g_i(\mathbf{d})\, d\mathbf{d} .$$

Hence,

$$\ln L = \sum_{i=1}^{N} \ln \int_{\mathbf{d}} f(\mathbf{y}_i|\mathbf{d})\, g_i(\mathbf{d})\, d\mathbf{d}$$

and

$$\frac{\partial\, ln\, L}{\partial \tau_K} = \sum_{i=1}^{N} \frac{\int_{\mathbf{d}} f(\mathbf{y}_i|\mathbf{d}) \left(\frac{\partial\, g_i(\mathbf{d})}{\partial \tau_K}\right) d\mathbf{d}}{\int_{\mathbf{d}} f(\mathbf{y}_i,\mathbf{d})\ g_i(\mathbf{d})\, d\mathbf{d}} .$$

Since

$$\frac{\partial\, g_i(\mathbf{d})}{\partial \tau_K} = g_i(\mathbf{d}) \frac{\partial\, ln\, g_i(\mathbf{d})}{\partial \tau_K} ,$$

it follows that (cf. (3.3.3))

$$\frac{\partial\, ln\, L}{\partial \tau_K} = \sum_{i=1}^{N} \int_{\mathbf{d}} \left\{ \frac{\partial\, ln\, g_i(\mathbf{d})}{\partial \tau_K} \right\} p\,(\mathbf{d}|\mathbf{y}_i)\, d\mathbf{d} ,$$

and, therefore,

$$\frac{\partial\, ln\, L}{\partial \tau_K} = \sum_{i=1}^{N} \mathrm{E}_{\,\mathbf{d}\,|\,\mathbf{y}_i}\left(\frac{\partial\, ln\, g_i(\,\mathbf{y}_i\,)}{\partial \tau_K}\right) .$$

Similarly, it can be shown that

$$\frac{\partial\, ln\, L}{\partial \gamma_K} = \sum_{i=1}^{N} \mathrm{E}_{\,\mathbf{d}\,|\,\mathbf{y}_i}\left(\frac{\partial\, ln\, f(\mathbf{y}_i|\mathbf{d})}{\partial \gamma_K}\right) .$$

□

PROPOSITION 3.4.2

The maximum likelihood estimator $\hat{\beta}$ of β can be expressed as

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}(\mathbf{c}_i\,|\,\mathbf{y}_i) \qquad (3.4.3)$$

with $\mathrm{E}(\mathbf{c}_i\,|\,\mathbf{y}_i)$ defined by (3.3.11).

Proof

Denote the prior distribution of $\mathbf{c}_i$ by $g_i\,(\mathbf{c})$ where

$$g_i\,(\mathbf{c}) = (2\pi)^{\,-\,q/2}\,|\,\mathbf{\Phi}_{(3)}|^{\,-\,1/2}\,\exp -\tfrac{1}{2}\,(\mathbf{c}_i\,-\,\boldsymbol{\beta})'\,\mathbf{\Phi}_{(3)}{}^{-\,1}\,(\mathbf{c}_i\,-\,\boldsymbol{\beta}\,)\,. \qquad (3.4.4)$$

The maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained as the solution of

$$\frac{\partial\; ln\; L}{\partial \boldsymbol{\beta}} = \mathbf{0}. \qquad (3.4.5)$$

Using Proposition 3.4.1 it follows that

$$\frac{\partial\; ln\; L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \mathrm{E}_{\;\mathbf{c}_i\,|\,\mathbf{y}_i}\,(\frac{\partial}{\partial \boldsymbol{\beta}}\; ln\; g_i\,(\mathbf{c})) \qquad (3.4.6)$$

where

$$ln\;\; g_i\,(\mathbf{c}) = -\,\frac{q}{2}\;ln\;(2\pi)\;-\;\frac{1}{2}\;ln\;|\,\mathbf{\Phi}_{(3)}|\;-$$

$$\frac{1}{2}\;(\mathbf{c}_i\,-\,\boldsymbol{\beta}\,)'\;\;\mathbf{\Phi}_{(3)}^{-\,1}\,(\mathbf{c}_i\,-\,\boldsymbol{\beta}\,)\,. \qquad (3.4.7)$$

Thus,

$$\frac{\partial}{\partial \boldsymbol{\beta}}\;\;ln\;g_i\,(\mathbf{c}) = \;\mathbf{\Phi}_{(3)}^{-\,1}\,(\mathbf{c}_i\,-\,\boldsymbol{\beta}\,)\,.$$

Therefore (cf. (3.4.5) and (3.4.6))

$$\sum_{i=1}^{N} \mathrm{E}_{\;\mathbf{c}_i\,|\,\mathbf{y}_i}\,\mathbf{\Phi}_{(3)}^{-\,1}\,(\mathbf{c}_i\,-\,\widehat{\boldsymbol{\beta}}\,)\;=\;\mathbf{0}\;\;,$$

so that

$$\widehat{\boldsymbol{\beta}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_{\;\mathbf{c}_i\,|\,\mathbf{y}_i}\,\mathbf{c}_i$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}(\mathbf{c}_i\,|\,\mathbf{y}_i)\;\;. \qquad \square$$

PROPOSITION 3.4.3

The maximum likelihood estimator $\hat{\mathbf{\Phi}}_{(3)}$ of $\mathbf{\Phi}_{(3)}$ can be expressed as

$$\hat{\mathbf{\Phi}}_{(3)} = \frac{1}{N} \sum_{i=1}^{N} \Big\{ \mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i) + \big(\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\boldsymbol{\beta}}\big) \times \big(\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\boldsymbol{\beta}}\big)' \Big\} \tag{3.4.8}$$

with $\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)$ defined in (3.3.8) and $\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i)$ defined in (3.3.11).

Proof

From Proposition 3.4.1 it follows that

$$\frac{\partial \ ln \ L}{\partial \Phi_{(3)rs}} = \sum_{i=1}^{N} \mathrm{E}_{\ \mathbf{c}_i|\mathbf{y}_i} \left\{ \frac{\partial \ ln \ g_i\ (\mathbf{c})}{\partial \Phi_{(3)rs}} \right\}.$$

Equation (3.4.7) may be written as

$$ln \ g_i\ (\mathbf{c}) = -\ \frac{q}{2} \ ln\ (2\pi) - \frac{1}{2}\, ln\, |\mathbf{\Phi}_{(3)}| - \frac{1}{2} \,\mathrm{tr}\ \mathbf{\Phi}_{(3)}^{-1}\ \mathbf{G}_c \tag{3.4.9}$$

where

$$\mathbf{G}_c = (\mathbf{c}_i - \boldsymbol{\beta})(\mathbf{c}_i - \boldsymbol{\beta})' . \tag{3.4.10}$$

Using the result (Browne and du Toit, 1992)

$$\frac{\partial \ ln \ g_i\ (\mathbf{c})}{\partial \Phi_{(3)rs}} = \frac{1}{2}\ \mathrm{tr}\ \mathbf{P}_i\ \frac{\partial \mathbf{\Phi}_{(3)}}{\partial \Phi_{(3)rs}} + \mathrm{tr}\ \mathbf{R}_i\ \frac{\partial \boldsymbol{\beta}}{\partial \Phi_{(3)rs}}\ ,$$

where

$$\mathbf{P}_i = \mathbf{\Phi}_{(3)}^{-1}\ (\mathbf{G}_c - \mathbf{\Phi}_{(3)})\ \mathbf{\Phi}_{(3)}^{-1}$$

and

$$\mathbf{R}_i = (\mathbf{c}_i - \boldsymbol{\beta})' \boldsymbol{\Phi}_{(3)}^{-1} , \qquad (3.4.11)$$

it follows that

$$\frac{\partial \, ln \, g_i(\mathbf{c})}{\partial \Phi_{(3)rs}} = \tfrac{1}{2} \operatorname{tr} \boldsymbol{\Phi}_{(3)}^{-1} (\mathbf{G}_c - \boldsymbol{\Phi}_{(3)}) \boldsymbol{\Phi}_{(3)}^{-1} \frac{\partial \boldsymbol{\Phi}_{(3)}}{\partial \Phi_{(3)rs}}$$

$$= \tfrac{1}{2} (2 - \delta_{rs}) [\, \boldsymbol{\Phi}_{(3)}^{-1} (\mathbf{G}_c - \boldsymbol{\Phi}_{(3)}) \, \boldsymbol{\Phi}_{(3)}^{-1} \,]_{rs} \; .$$

since

$$\frac{\partial \, \boldsymbol{\Phi}_{(3)}}{\partial \Phi_{(3)rs}} = \mathbf{J}_{rs} + (1 - \delta_{rs}) \, \mathbf{J}_{sr}$$

and

$$\operatorname{tr} \mathbf{A} \, \mathbf{J}_{rs} = [\mathbf{A}]_{sr} \, .$$

Let

$$\frac{\partial \, ln \, L}{\partial \Phi_{(3)rs}} = \sum_{i=1}^{N} \mathrm{E}_{\, \mathbf{c}_i | \mathbf{y}_i} \left\{ \frac{\partial \, ln \, g_i(\mathbf{c})}{\partial \Phi_{(3)rs}} \right\} = 0, \quad r, \, s = 1, 2, \ldots, q \, .$$

Then $\hat{\boldsymbol{\Phi}}_{(3)}$ is obtained as

$$\hat{\boldsymbol{\Phi}}_{(3)} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}_{\, \mathbf{c}_i | \mathbf{y}_i} \hat{\mathbf{G}}_c$$

where

$$\mathrm{E}_{\, \mathbf{c}_i | \mathbf{y}_i} \hat{\mathbf{G}}_c = \mathrm{E}_{\, \mathbf{c}_i | \mathbf{y}_i} \, [(\mathbf{c}_i - \mathrm{E}(\mathbf{c}_i | \mathbf{y}_i)) + (\mathrm{E}(\mathbf{c}_i | \mathbf{y}_i) - \hat{\boldsymbol{\beta}})] \times$$

$$[(\mathbf{c}_i - \mathrm{E}(\mathbf{c}_i | \mathbf{y}_i)) + (\mathrm{E}(\mathbf{c}_i | \mathbf{y}_i) - \hat{\boldsymbol{\beta}})]'$$

$$= \mathrm{E}_{\, \mathbf{c} | \mathbf{y}_i} \, (\mathbf{c}_i - \mathrm{E}(\mathbf{c}_i | \mathbf{y}_i)) \, (\mathbf{c}_i - \mathrm{E}(\mathbf{c}_i | \mathbf{y}_i))' +$$

$$(\mathrm{E}(\mathbf{c}_i | \mathbf{y}_i) - \hat{\boldsymbol{\beta}})(\mathrm{E}(\mathbf{c}_i | \mathbf{y}_i) - \hat{\boldsymbol{\beta}})'$$

$$= \text{Cov}(\mathbf{c}_i|\mathbf{y}_i) + (\text{ E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\beta}\ )\ (\text{ E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\beta}\ )'.$$

Thus, $\hat{\mathbf{\Phi}}_{(3)}$ may be expressed as

$$\hat{\mathbf{\Phi}}_{(3)} = \frac{1}{N}\sum_{i=1}^{N}\left\{\text{Cov}(\mathbf{c}_i|\mathbf{y}_i) + (\text{ E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\beta}\ )\ (\text{ E}(\mathbf{c}_i|\mathbf{y}_i\ ) - \hat{\beta}\ )'\right\}. \qquad \square$$

PROPOSITION 3.4.4

The maximum likelihood estimator $\hat{\Phi}_{(1)}$ of $\Phi_{(1)}$ is given by

$$\hat{\Phi}_{(1)} = \frac{1}{NN^*}\sum_{i=1}^{N}\left\{\sum_{j=1}^{n_i}\left\{(\ \mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))\ (\ \mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\text{E}(\mathbf{b}_{ij}\,|\mathbf{y}_i))' + \text{tr}\ \mathbf{X}'_{(2)ij}\mathbf{X}_{(2)ij}\text{ Cov}(\mathbf{b}_{ij}|\mathbf{y}_i)\right\}\right\} \qquad (3.4.12)$$

where $N^* = \sum_{j=1}^{n_i} n_{ij}$ with n_{ij} the number of units of the i-th level-3 and j-th level-2 unit.

Proof

The conditional pdf of $\mathbf{y}_i$, given

$$\mathbf{b}_i = \begin{bmatrix} \mathbf{b}_{i1} \\ \vdots \\ \mathbf{b}_{ij} \\ \vdots \\ \mathbf{b}_{in_i} \end{bmatrix},$$

is

$$f(\mathbf{y}_i|\mathbf{b}) = (2\pi)^{-\frac{N^*}{2}}\,|\Phi_{(1)}|^{-\frac{N^*}{2}}\exp -\frac{1}{2\,\Phi_{(1)}}\,(\mathbf{y}_i - \mathbf{x}_i^*)'\,(\mathbf{y}_i - \mathbf{x}_i^*)$$

where

$$\mathbf{x}_i^* = \begin{bmatrix} \mathbf{X}_{(2)i1}\,\mathbf{b}_{i1} \\ \mathbf{X}_{(2)i2}\,\mathbf{b}_{i2} \\ . \\ . \\ \mathbf{X}_{(2)in_i}\,\mathbf{b}_{in_i} \end{bmatrix} .$$

The natural logarithm of the conditional pdf $f\,(\mathbf{y}_i|\,\mathbf{b})$ is

$$ln\, f\,(\mathbf{y}_i|\,\mathbf{b}) \;=\; -\,\frac{N^*}{2}\, ln\,(2\pi) \;-\; \frac{N^*}{2}\, ln\, |\Phi_{(1)}| \;-\; \frac{1}{2\Phi_{(1)}}\,\mathrm{tr}\,\mathbf{G}_y$$

where

$$\mathbf{G}_y = (\mathbf{y}_i - \mathbf{x}_i^*)\,(\mathbf{y}_i - \mathbf{x}_i^*)' .$$

From Proposition 3.4.1 it follows that

$$\frac{\partial\, ln\, L}{\partial \Phi_{(1)}} \;=\; \sum_{i=1}^{N} \mathrm{E}\;{}_{\mathbf{b}_i|\,\mathbf{y}_i}\left(\frac{\partial\; ln\;\; f\,(\mathbf{y}_i|\,\mathbf{b}_i)}{\partial \Phi_{(1)}}\right) .$$

The derivative of $ln\, f\,(\mathbf{y}_i\,|\,\mathbf{b}_i)$ with respect to $\hat{\Phi}_{(1)}$ is

$$\frac{\partial\; ln\, f\,(\mathbf{y}_i|\,\mathbf{b}_i)}{\partial \Phi_{(1)}} \;=\; -\,\frac{N^*}{2\Phi_{(1)}} + \frac{1}{2\Phi_{(1)}^2}\,\mathrm{tr}\,\mathbf{G}_y .$$

Setting $\frac{\partial\, ln\, L}{\partial \Phi_{(1)}}$ equal to zero, $\hat{\Phi}_{(1)}$ may be calculated as

$$\hat{\Phi}_{(1)} = \frac{1}{NN^*} \sum_{i=1}^{N} \mathrm{E}\;{}_{\mathbf{b}_i|\,\mathbf{y}_i}\,\mathrm{tr}\,\mathbf{G}_y \qquad (3.4.13)$$

where

$$\mathrm{tr}\,\mathbf{G}_y = \mathrm{tr}\,(\mathbf{y}_i - \mathbf{x}_i^*)'\,(\mathbf{y}_i - \mathbf{x}_i^*)$$

$$= \sum_{j=1}^{n_i} \text{tr} \; [\, \mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) + \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) - \mathbf{X}_{(2)ij}\, \mathbf{b}_{ij} \,]' \times$$

$$[\, \mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) + \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) - \mathbf{X}_{(2)ij}\, \mathbf{b}_{ij} \,]$$

and, therefore,

$$\text{E}_{\,\mathbf{b}_i|\mathbf{y}_i} \text{tr } \mathbf{G}_y = \sum_{j=1}^{n_i} \Big\{ \text{E}_{\,\mathbf{b}_{ij}|\mathbf{y}_i} \text{tr } (\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))' \times$$

$$(\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i)) +$$

$$\text{E}_{\,\mathbf{b}_{ij}|\mathbf{y}_i} \text{tr } \mathbf{X}'_{(2)ij}\, \mathbf{X}_{(2)ij}\, (\mathbf{b}_{ij} - \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))' \times$$

$$(\mathbf{b}_{ij} - \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i)\Big\}$$

$$= \sum_{j=1}^{n_i} \Big\{ (\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))\, (\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))' +$$

$$\text{tr } \mathbf{X}'_{(2)ij}\, \mathbf{X}_{(2)ij}\, \text{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) \Big\} .$$

Equation (3.4.13) can then be written as

$$\hat{\Phi}_1 = \frac{1}{NN^*} \Big\{ \sum_{j=1}^{n_i} \Big\{ (\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))\, (\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\, \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i))' +$$

$$\text{tr } \mathbf{X}'_{(2)ij}\, \mathbf{X}_{(2)ij}\, \text{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) \Big\} . \qquad \square$$

PROPOSITION 3.4.5

The maximum likelihood estimator $\hat{\mathbf{\Phi}}_{(2)}$ of $\mathbf{\Phi}_{(2)}$ can be expressed as

$$\hat{\mathbf{\Phi}}_{(2)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \Big\{ \text{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) + \Big(\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) - \mathbf{S}_i\, \hat{\boldsymbol{\beta}} \Big) \times$$

$$\Big(\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) - \mathbf{S}_i\, \hat{\boldsymbol{\beta}} \Big)' \Big\} - \frac{1}{N} \sum_{i=1}^{N} \mathbf{S}_i\, \hat{\mathbf{\Phi}}_{(3)}\, \mathbf{S}'_i , \qquad (3.4.14)$$

with $\mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i)$ defined in (3.3.13).

Proof

The prior distribution of $\mathbf{b}_i$ can be written as

$$g(\mathbf{b}_i) = \prod_{j=1}^{n_i} g(\mathbf{b}_{ij})$$

$$= (2\pi)^{-\frac{mn_i}{2}} |\mathbf{\Phi}_b|^{-\frac{n_i}{2}} \; exp - \frac{1}{2} \sum_{j=1}^{n_i} (\mathbf{b}_{ij} - \mathbf{S}_i \boldsymbol{\beta})' \mathbf{\Phi}_b^{-1} (\mathbf{b}_{ij} - \mathbf{S}_i \boldsymbol{\beta}) \tag{3.4.15}$$

where

$$\mathbf{\Phi}_b = \mathbf{\Phi}_{(2)} + \mathbf{S}_i \mathbf{\Phi}_{(3)} \mathbf{S}_i' . \tag{3.4.16}$$

From (3.4.15) it follows that

$$ln\; g(\mathbf{b}_i) = -\frac{n_i}{2}\left\{ m\; ln\,(2\pi) + ln\; |\mathbf{\Phi}_b| + \mathrm{tr}\, \mathbf{\Phi}_b^{-1} \mathbf{G}_{b_i} \right\}$$

where

$$\mathbf{G}_{b_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{b}_{ij} - \mathbf{S}_i \boldsymbol{\beta})(\mathbf{b}_{ij} - \mathbf{S}_i \boldsymbol{\beta})' .$$

Using (3.4.11) it follows that

$$\frac{\partial\; ln\; g(\mathbf{b}_i)}{\partial \Phi_{(2)rs}} = \frac{n_i}{2} \mathrm{tr}\, [\mathbf{\Phi}_b^{-1} (\mathbf{G}_{b_i} - \mathbf{\Phi}_b) \mathbf{\Phi}_b^{-1} \frac{\partial \mathbf{\Phi}_b}{\partial \Phi_{(2)rs}}]$$

$$= \frac{n_i}{2} (2 - \delta_{rs}) [\mathbf{\Phi}_b^{-1} (\mathbf{G}_{b_i} - \mathbf{\Phi}_b) \mathbf{\Phi}_b^{-1}]_{rs} ,$$

since (cf. (3.4.16))

$$\frac{\partial \mathbf{\Phi}_b}{\partial \Phi_{(2)rs}} = \mathbf{J}_{rs} + (1 - \delta_{rs}) \mathbf{J}_{sr}$$

and

$$\text{tr } \mathbf{A}\ \mathbf{J}_{rs} = [\ \mathbf{A}\ ]_{sr}\ .$$

Setting $\dfrac{\partial\ ln\ L}{\partial \Phi_{(2)rs}}$ equal to zero, $r,\ s = 1, 2, \ldots, m$, an estimator $\hat{\mathbf{\Phi}}_b$ of $\mathbf{\Phi}_b$ is obtained as follows:

$$\hat{\mathbf{\Phi}}_b \ = \ \frac{1}{N}\sum_{i=1}^{N} \text{E}_{\ \mathbf{b}_{ij}\,|\,\mathbf{y}_i}\, \hat{\mathbf{G}}_b\ .$$

Using (3.4.16), $\hat{\mathbf{\Phi}}_{(2)}$ may be written as

$$\hat{\mathbf{\Phi}}_{(2)} = \frac{1}{N}\sum_{i=1}^{N} (\text{E}_{\ \mathbf{b}_{ij}\,|\,\mathbf{y}_i}\, \hat{\mathbf{G}}_b \ - \ \mathbf{S}_i\, \hat{\mathbf{\Phi}}_{(3)}\, \mathbf{S}_i') \qquad (3.4.17)$$

where

$$\hat{\mathbf{G}}_b = \frac{1}{n_i}\sum_{j=1}^{n_i}[(\mathbf{b}_{ij} \ - \ \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) + \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) \ - \ \mathbf{S}_i\,\hat{\boldsymbol{\beta}}\ ] \ \times$$
$$[(\mathbf{b}_{ij} \ - \ \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) + \text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) \ - \ \mathbf{S}_i\,\hat{\boldsymbol{\beta}}\ ]'$$

and $\hat{\mathbf{\Phi}}_{(3)}$ is given by (3.4.8).

The expression $\frac{1}{N}\sum_{i=1}^{N} \text{E}_{\ \mathbf{b}_{ij}\,|\,\mathbf{y}_i}\, \hat{\mathbf{G}}_b$ can then be written as

$$\frac{1}{N}\sum_{i=1}^{N} \text{E}_{\ \mathbf{b}_{ij}\,|\,\mathbf{y}_i}\, \hat{\mathbf{G}}_b = \frac{1}{N}\frac{1}{n_i}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\Big\{\text{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) +$$
$$\Big(\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) \ - \ \mathbf{S}_i\,\hat{\boldsymbol{\beta}}\Big)\Big(\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) \ - \ \mathbf{S}_i\,\hat{\boldsymbol{\beta}}\ \Big)'\Big\}$$

so that (3.4.17) can be written as

$$\hat{\mathbf{\Phi}}_{(2)} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}\Big\{\text{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) + \Big(\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i) \ - \ \mathbf{S}_i\,\hat{\boldsymbol{\beta}}\ \Big) \ \times$$
$$\Big(\text{E}(\mathbf{b}_{ij}|\mathbf{y}_i\,) \ - \ \mathbf{S}_i\,\hat{\boldsymbol{\beta}}\ \Big)'\Big\} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_i\,\hat{\mathbf{\Phi}}_{(3)}\,\mathbf{S}_i'\ . \qquad \square$$

EM estimators of the unknown parameters may be obtained recursively from equations (3.3.8), (3.3.11), (3.4.3), (3.4.8), (3.3.14), (3.3.15), (3.4.14) and (3.4.12) which, for the purpose of clarity, are repeated below, with $\mathbf{A}_{ij}$ defined by (3.3.9).

(i) $$\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i) = (\mathbf{\Phi}_{(3)}^{-1} + \mathbf{S}_i' \sum_{j=1}^{n_i}(\mathbf{A}_{ij})\,\mathbf{S}_i\,)^{-1}$$

(ii) $$\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) = \mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\;\{\,\mathbf{S}_i' \sum_{j=1}^{n_i} \mathbf{b}_{ij} - \mathbf{S}_i' \sum_{j=1}^{n_i}(\mathbf{A}_{ij}\,)\,\mathbf{S}_i\,\hat{\boldsymbol\beta}\,\} + \hat{\boldsymbol\beta}$$

(iii) $$\hat{\boldsymbol\beta} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}(\mathbf{c}_i|\mathbf{y}_i)$$

(iv) $$\hat{\mathbf{\Phi}}_{(3)} = \frac{1}{N}\sum_{i=1}^{N}\Big\{\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i) + \Big(\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\boldsymbol\beta}\Big)\Big(\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) - \hat{\boldsymbol\beta}\Big)'\Big\}$$

(v) $$\mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) = (\mathbf{I} - \hat{\mathbf{\Phi}}_{(2)}\,\mathbf{A}_{ij})\,\mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i'\,(\mathbf{I} - \hat{\mathbf{\Phi}}_{(2)}\,\mathbf{A}_{ij})' + \hat{\mathbf{\Phi}}_{(2)}\,(\mathbf{I} - \hat{\mathbf{\Phi}}_{(2)}\,\mathbf{A}_{ij})$$

(vi) $$\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i) = \mathbf{S}_i\,\mathrm{E}(\mathbf{c}_i|\mathbf{y}_i) + \hat{\mathbf{\Phi}}_{(2)}\,(\mathbf{b}_{ij} - \mathbf{A}_{ij}\,\mathbf{S}_i\,\hat{\boldsymbol\beta}) + \hat{\mathbf{\Phi}}_{(2)}\,\mathbf{A}_{ij}\,\mathbf{S}_i\,\mathrm{Cov}(\mathbf{c}_i|\mathbf{y}_i)\,\mathbf{S}_i'\,\Big\{\sum_{j=1}^{n_i}\mathbf{A}_{ij}\,\mathbf{S}_i\,\hat{\boldsymbol\beta} - \sum_{j=1}^{n_i}\mathbf{b}_{ij}\Big\}$$

(vii) $$\hat{\mathbf{\Phi}}_{(2)} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}\Big\{\mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i) + \Big(\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i\,) - \mathbf{S}_i\,\hat{\boldsymbol\beta}\Big) \times \Big(\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i) - \mathbf{S}_i\,\hat{\boldsymbol\beta}\Big)'\Big\} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_i\,\hat{\mathbf{\Phi}}_{(3)}\,\mathbf{S}_i'$$

(viii) $$\hat{\Phi}_{(1)} = \frac{1}{NN^*}\sum_{i=1}^{N}\Big\{\sum_{j=1}^{n_i}\Big\{(\,\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\,\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i))\,(\,\mathbf{y}_{ij} - \mathbf{X}_{(2)ij}\,\mathrm{E}(\mathbf{b}_{ij}|\mathbf{y}_i))' + \mathrm{tr}\,\mathbf{X}_{(2)ij}'\,\mathbf{X}_{(2)ij}\,\mathrm{Cov}(\mathbf{b}_{ij}|\mathbf{y}_i)\Big\}\Big\}$$

The following steps are used:

Step 1: Set $\boldsymbol\beta = 0$, $\Phi_{(1)} = 1$, $\mathbf{\Phi}_{(2)}^{-1} = \mathbf{0}$, $\mathbf{\Phi}_{(3)}^{-1} = \mathbf{0}$.
Step 2: Evaluate expressions (i) to (viii) successively.
Step 3: Repeat Step 2, starting from (i) to (viii), until the procedure converges.

Let

$$\gamma_k = \begin{bmatrix} \beta \\ \text{vecs } \mathbf{\Phi}_{(3)} \\ \text{vecs } \mathbf{\Phi}_{(2)} \\ \Phi_{(1)} \end{bmatrix},$$

where γ_k denotes the k-th approximation to the estimator $\hat{\gamma}$ of γ. A possible convergence criterion is to require that

$$\frac{|\,[\gamma_{k+1}]_i - [\gamma_k]_i\,|}{\min\,(1,\,|\,[\gamma_k]_i\,|)} < \varepsilon \text{ for } i = 1, 2, \ldots, k$$

where ε is an arbitrary small number, for example $\varepsilon = 10^{-5}$.

The optimization algorithm described above is known as the EM algorithm. Iterations based on (iii), (iv), (vii) and (viii) provide a rapid, robust method for obtaining close approximations to the marginal maximum likelihood estimates of the parameters.

At level-3 the Bayes estimator $\widehat{\boldsymbol{\beta}}_\beta$ of $\boldsymbol{\beta}$ for a normal prior $\mathbf{g}_{(c)}$ is given by $E(\mathbf{c}_i\,|\,\mathbf{y}_i)$, defined by (*ii*). Since $\boldsymbol{\beta}$, $\mathbf{\Phi}_{(3)}$, $\mathbf{\Phi}_{(2)}$ and $\Phi_{(1)}$ are unknown and have to be estimated from the data, $E(\mathbf{c}_i\,|\,\mathbf{y}_i)$ yields a so-called empirical Bayes estimate of $\boldsymbol{\beta}$ for experimental unit i, $i = 1, 2, \ldots, N$.

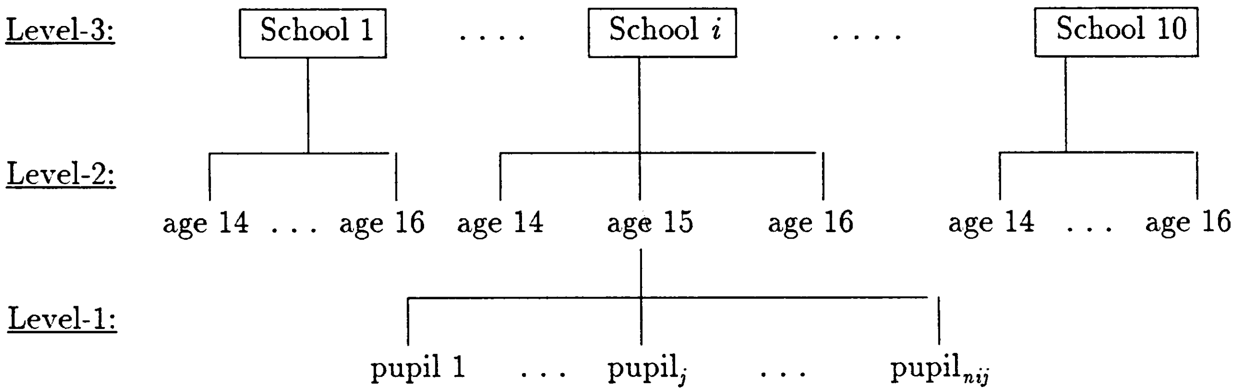
3.5 PRACTICAL APPLICATION

In this section the results given in Sections 3.2 to 3.4 are illustrated. A comparison of the empirical Bayes estimates obtained and the OLS estimates for the same level-3 units is also given.

Example 3.5.1

In this example the analysis of a data set obtained from a survey undertaken by the South African Department of National Health and Population Development is considered. As education is currently the only viable strategy to combat the spread of AIDS, the Department developed a comprehensive AIDS and life skills education package for secondary schools.

The data set used contains information on the knowledge and perception of AIDS and related issues of 1702 pupils at secondary schools and can be schematically represented as follows:



The level-3 units are the 10 schools, the level-2 units are the different age groups and the level-1 units are the individual pupils.

The response variable, y, is the total score of each pupil's perception of HIV/AIDS susceptibility. The total score of each pupil's knowledge, x, of transmission of AIDS through proven means, together with the gender of the pupil, are used as explanatory variables.

The items used to calculate the total scores used as response and explanatory variables are given in Tables 3.5.1 and 3.5.2 respectively. Note that the possible outcome of each item is either 'yes' or 'no' and that these categories are coded '1' for positive responses and '0' for negative responses prior to the calculation of the total score.

Table 3.5.1: Items used to calculate a total score for a pupil's perception of HIV/AIDS susceptability.

Description of Item	Possible outcomes
Can teenagers (people under 20 years old) get AIDS?	Yes/No
Is AIDS/HIV a disease from which mainly your own population group suffers?	Yes/No
Is AIDS/HIV a disease from which mainly other population groups suffer?	Yes/No
Do you think that all people can get AIDS/HIV?	Yes/No
Do you think that AIDS/HIV is mainly a "gay disease / homosexual disease"?	Yes/No
Do you think it is possible that even your best friend could get AIDS/HIV?	Yes/No

Maximum score: 6

Table 3.5.2: Items used to calculate a total score for a pupil's knowledge of the transmission of AIDS/HIV by proven means.

Description of Item	Possible outcomes
One cannot get AIDS/HIV by having sex with someone who has AIDS/HIV.	Yes/No
A pregnant woman can pass AIDS/HIV on to her unborn baby.	Yes/No
One can get AIDS/HIV by receiving blood that was donated by a person who has AIDS/HIV.	Yes/No
A man can get AIDS/HIV by having sex with another man who has AIDS/HIV.	Yes/No
AIDS/HIV is caused by a virus / germ	Yes/No
The body of a person with AIDS/HIV cannot defend itself against other diseases.	Yes/No
AIDS/HIV cannot be spread when two uninfected people have sex.	Yes/No
A woman cannot get AIDS/HIV by having sex with another woman who has AIDS/HIV.	Yes/No
AIDS/HIV cannot be transmitted by semen.	Yes/No
AIDS/HIV can be transmitted by vaginal fluids.	Yes/No
AIDS/HIV can only be transmitted by a person who is infected with AIDS/HIV.	Yes/No
Is it true that someone who suffers from sexually transmitted disease is more likely to get AIDS/HIV?	Yes/No
One can get AIDS/HIV by receiving blood that was donated by a person who has AIDS/HIV.	Yes/No

Maximum score: 13

The following linear model (cf. (3.2.7), (3.2.8) and (3.2.9)) is fitted to the data:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\,\mathbf{b}_{ij} + \mathbf{e}_{ij} \tag{3.5.1}$$

where

$$\mathbf{b}_{ij} = \mathbf{c}_i + \mathbf{u}_{ij} \tag{3.5.2}$$

and

$$\mathbf{c}_i = \boldsymbol{\beta} + \mathbf{v}_i\,. \tag{3.5.3}$$

A typical element of $\mathbf{y}_{ij}$ is y_{ijk} which denotes the total score with regard to HIV/AIDS susceptability of pupil k being of age j from school i. Similarly, a typical element of the $n_{ij} \times 1$ vector $\mathbf{X}_{ij}\,\mathbf{b}_{ij}$ is $\mathbf{x}'_{ijk}\,\mathbf{b}_{ij}$ where

$$\mathbf{x}'_{ijk} = (1 \quad \text{knowledge}_{ijk} \quad \text{gender}_{ijk})$$

and

$$\mathbf{b}_{ij} = \begin{bmatrix} \text{intercept}_{ij} \\ \text{regression coefficient for knowledge}_{ij} \\ \text{regression coefficient for gender}_{ij} \end{bmatrix},$$

where $\mathbf{b}_{ij}$ is a vector of stochastic coefficients for age group j and school i. The vector $\mathbf{c}_i$ (cf. (3.5.3)) is a vector of stochastic coefficients for school i. Note that for the model considered above, the matrix $\mathbf{S}_i$ (cf. (3.2.8)) is taken as the identity matrix. This implies that the covariance of $(\mathbf{b}_{ij}, \mathbf{b}'_{ij})$ (cf. Section 3.3) reduces to

$$\text{Cov}(\mathbf{b}_{ij}, \mathbf{b}'_{ij}) = \boldsymbol{\Phi}_{(3)} + \boldsymbol{\Phi}_{(2)}\,. \tag{3.5.4}$$

From the distributional assumptions given in Section 3.2, it also follows that

$$\text{Cov}(\mathbf{c}_i, \mathbf{c}'_i) = \boldsymbol{\Phi}_{(3)} \tag{3.5.5}$$

and

$$\text{Cov}(\mathbf{e}_i, \mathbf{e}_i') = \Phi_{(1)}\, \mathbf{I}_{n_i}\,. \tag{3.5.6}$$

A computer program ML3_EM was written in FORTRAN to implement the theoretical results given in Sections 3.2 to 3.4. Part of the computer output of ML3_EM is given below:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$
INTERCEPT	2.4864
SCORE	0.1848
GENDER	-0.0610

(ii) **Random part of the model:**

ESTIMATE OF $\hat{\Phi}_{(3)}$

	INTERCEPT	SCORE	GENDER
INTERCEPT	0.1682		
SCORE	-0.0135	0.0014	
GENDER	0.0090	-0.0010	0.0028

ESTIMATE OF $\hat{\Phi}_{(2)}$

	INTERCEPT	SCORE	GENDER
INTERCEPT	0.1238		
SCORE	-0.0082	0.0006	
GENDER	-0.0066	0.0004	0.0044

ESTIMATE OF $\hat{\Phi}_{(1)}$

1.4644

(iii) **Convergence details:**

Convergence (EM-algorithm in 11 iterations)

From part (i) of the output it can be seen that the estimate of the intercept is 2.4864, which may be interpreted as the expected value of y, assuming no score or gender effect. The score coefficient indicates that, on average, an increase of 5 in the predictor score is required to accomplish an increase of 1, that is 16.7 %, in the response score. The gender score of -0.0610 implies a total gender effect of 0.1220. In the case of boys (coded '-1') this leads to a positive effect of 0.0610, and for girls (coded '1') a negative effect of -0.0610, showing that boys generally obtained a slightly higher score than girls.

When considering part (ii) of the output, a comparison of the elements of $\hat{\mathbf{\Phi}}_{(3)}$ and $\hat{\mathbf{\Phi}}_{(2)}$ indicates slightly greater variation on level-3 (schools) than on level-2 (age groups). The value of 1.4644 obtained for $\hat{\Phi}_{(1)}$ (cf. (3.5.6)) can be interpreted as the variation in observation errors.

From (3.2.9) and (3.2.10) it follows that

$$\mathbf{c}_i \sim \mathrm{N}(\boldsymbol{\beta}, \mathbf{\Phi}_{(3)}). \tag{3.5.7}$$

A 95 % tolerance interval for a typical element of $\mathbf{c}_i$, say $[\mathbf{c}_i]_k$, is therefore given by

$$\beta_k \pm 2\sqrt{[\,\hat{\Phi}_{(3)}]_{k,k}}\ .$$

Since β_k and $\Phi_{(3)k,k}$ are unknown and have to be estimated from the data, an approximate 95 % tolerance interval for $[\mathbf{c}_i]_k$ is given by

$$\hat{\beta}_k \pm 2\sqrt{\hat{\Phi}_{(3)k,k}}\ . \tag{3.5.8}$$

Note that $\mathbf{c}_1$, $\mathbf{c}_2$, . . . , $\mathbf{c}_N$ are identically and independently distributed, so that the same tolerance intervals apply irrespective of the subscript of these random coefficient vectors. Approximate 95 % tolerance intervals for the level-3 random coefficients can be constructed using (3.5.8). The 95 % tolerance interval for the intercept is given by

$$\hat{\beta}_1 \pm 2\sqrt{\hat{\Phi}_{(3)11}} \ ,$$

that is,

$$(1.6662 \ ; \ 3.3066) \ .$$

The 95 % tolerance intervals for the score and gender coefficients are given by (0.1094 ; 0.2602) and (– 0.1674 ; 0.0454) respectively.

On the second level of the hierarchy it follows from (3.5.4) that

$$\mathbf{b}_{ij} \sim \mathrm{N}(\boldsymbol{\beta}, \boldsymbol{\Phi}_{(2)} + \boldsymbol{\Phi}_{(3)})$$

and hence 95 % tolerance intervals for $[\mathbf{b}_{ij}]_k$ are given by

$$\hat{\beta}_k \pm 2\sqrt{\hat{\Phi}_{(3)k,k} + \hat{\Phi}_{(2)k,k}} \quad , k = 1, 2, \ldots, m \ . \tag{3.5.9}$$

The 95 % tolerance interval for the intercept is given by

$$(1.4057 \ ; \ 3.5675) \ .$$

Similar intervals for the score and gender coefficients are given by (0.0958 ; 0.2738) and (– 0.2306 ; 0.1086) respectively. Note that the 95 % tolerance interval for gender on level-2 is much wider than on level-3. This may be caused by the aggregation over age groups on level-3.

Since the calculation of the standard errors of the coefficients does not form part of the EM algorithm, it is not possible to determine which of the estimated coefficients are significant. It would seem, however, from the small gender effect that gender is not a significant predictor of the response score.

The empirical Bayes estimators $E(\mathbf{c}_i|\mathbf{y}_i)$ of the coefficients $\mathbf{c}_i$ for school i, and corresponding estimated standard deviations (cf. (3.3.8)) can, however, be obtained from the computer output. Table 3.5.3 is a comparison of the empirical Bayes estimates with the Ordinary Least Squares estimates of the regression coefficients for each of the 10 schools.

Table 3.5.3: Comparison of Bayes estimates and estimated regression coefficients.

School no.	Multilevel model		Linear regression model	
	$\hat{E}(\mathbf{c}_i \mid \mathbf{y}_i)$	Std. dev.	Regr. coeff.	Std. error
1	2.2760	0.2452	2.2313*	0.2609
	0.1780	0.0233	0.1768*	0.0291
	-0.0471	0.0411	-0.0456	0.0720
2	2.0950	0.2770	1.6196*	0.3805
	0.2252	0.0285	0.2795*	0.0451
	-0.0949	0.0454	-0.1249	0.1016
3	2.2300	0.2841	2.2579*	0.4156
	0.1816	0.0281	0.1641*	0.0468
	-0.0674	0.0455	-0.1367	0.1122
4	2.4810	0.2848	2.5477*	0.3844
	0.1914	0.0273	0.1870*	0.0403
	-0.0836	0.0446	-0.1715	0.0970
5	2.2880	0.3012	1.8358*	0.4690
	0.2213	0.0288	0.2762*	0.0481
	-0.0902	0.0459	-0.0923	0.1167
6	2.9760	0.3445	4.3070*	0.6890
	0.1574	0.0319	0.0455*	0.0647
	-0.0397	0.0469	-0.0488	0.1081
7	2.7520	0.3376	3.2851*	0.6095
	0.1752	0.0304	0.1314*	0.0560
	-0.0604	0.0459	-0.0969	0.0988

8	2.5350	0.2920	2.5110*	0.3700
	0.1689	0.0286	0.1635*	0.0392
	-0.0278	0.0460	0.1135	0.0975
9	2.3930	0.2982	2.2516*	0.4166
	0.1978	0.0291	0.2159*	0.0442
	-0.0758	0.0452	-0.1152	0.0923
10	2.8450	0.3265	3.5915*	0.4410
	0.1497	0.0303	0.0806	0.0428
	-0.0243	0.0455	0.0230	0.0710

* Significant at 5 % level of significance

From Table 3.5.3, columns 2, 4 and 5, it can be seen that none of the regression coefficients for gender are significant at a 5 % level of significance, confirming the conclusion based on the results of the multilevel model. Standard errors of significant regression coefficients are, in general, larger than the estimated standard deviations for the estimates obtained through using the multilevel model. There is a considerable variation in the regression coefficients over schools. In the case of the regression coefficient for the intercept, a value of 1.8358 is obtained for school 5, compared to a value of 4.3070 for school 6. Figure 3.5.1 gives a graphical presentation of the predicted response scores plotted against the knowledge scores, x, for the 10 schools, given gender = 1, that is, females. In the figure FACTOR D denotes the response variable and FACTOR A denotes the predictor variable.

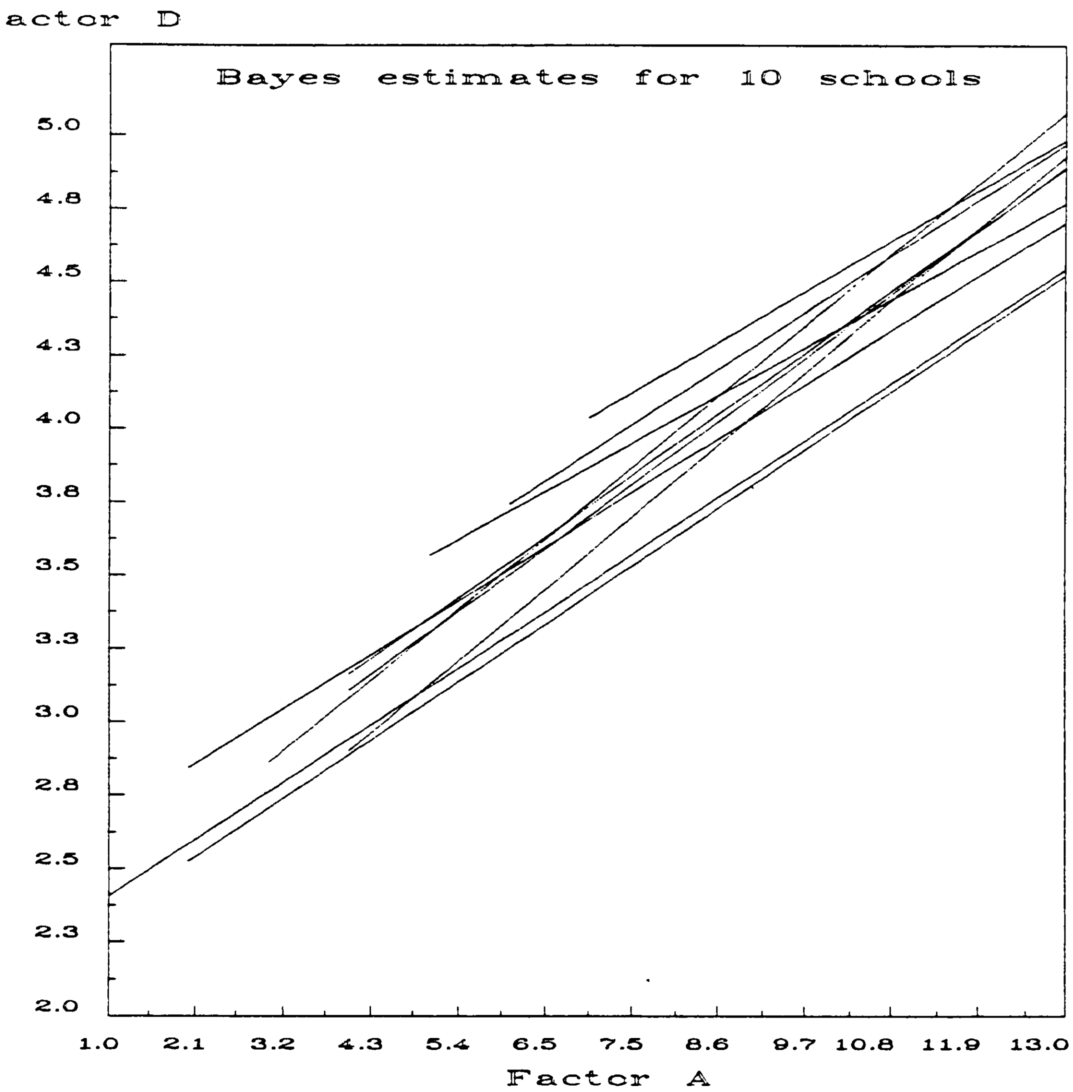


Figure 3.5.1: Predicted regression lines for 10 schools, given gender = 1.

The differences between schools is further illustrated in Figures 3.5.2 and 3.5.3 which are plots of predicted scores against x and 95 % confidence limits for the mean against x, given gender = 1.

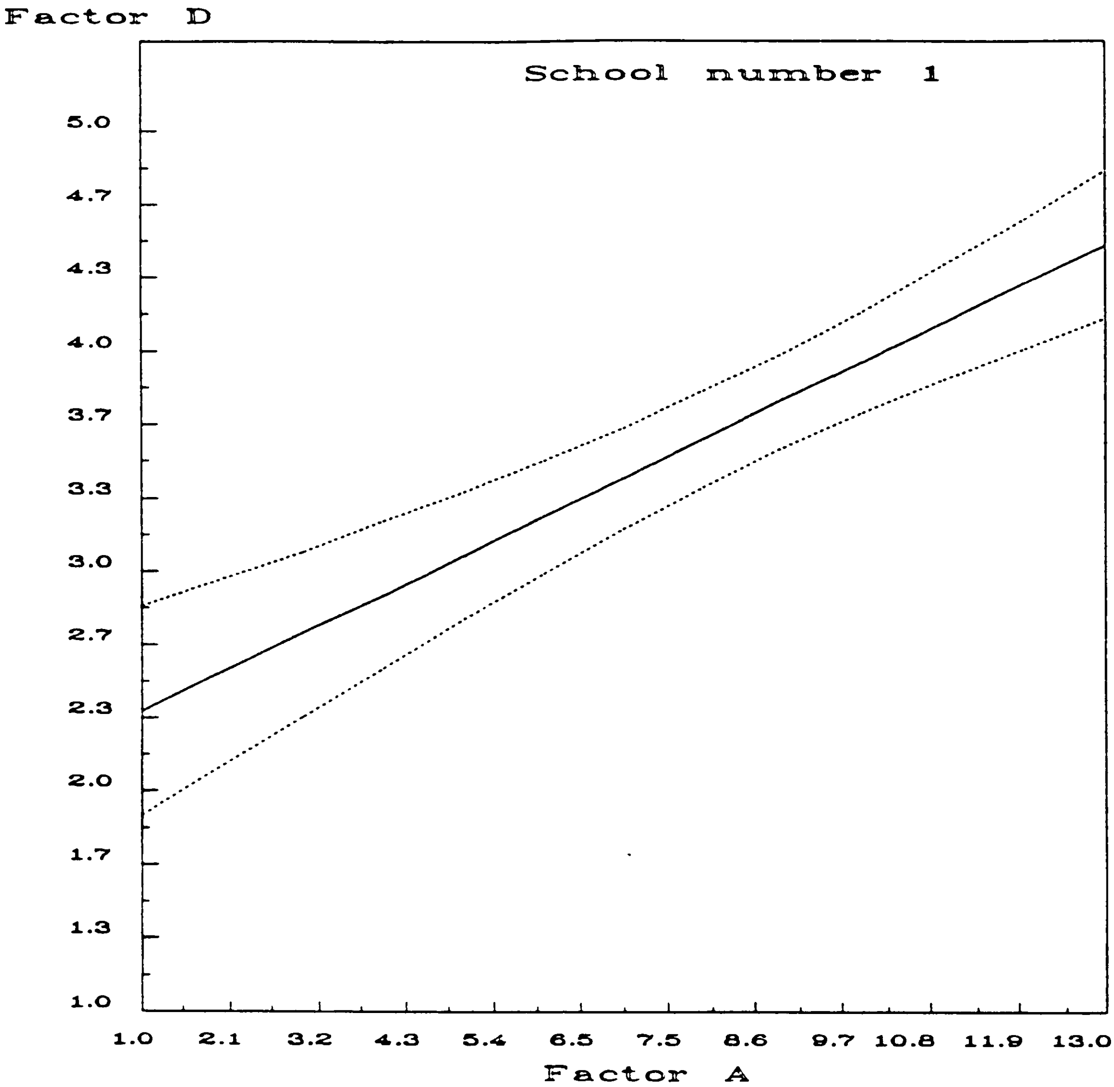


Fig. 3.5.2: Predicted scores and confidence limits for school 1.

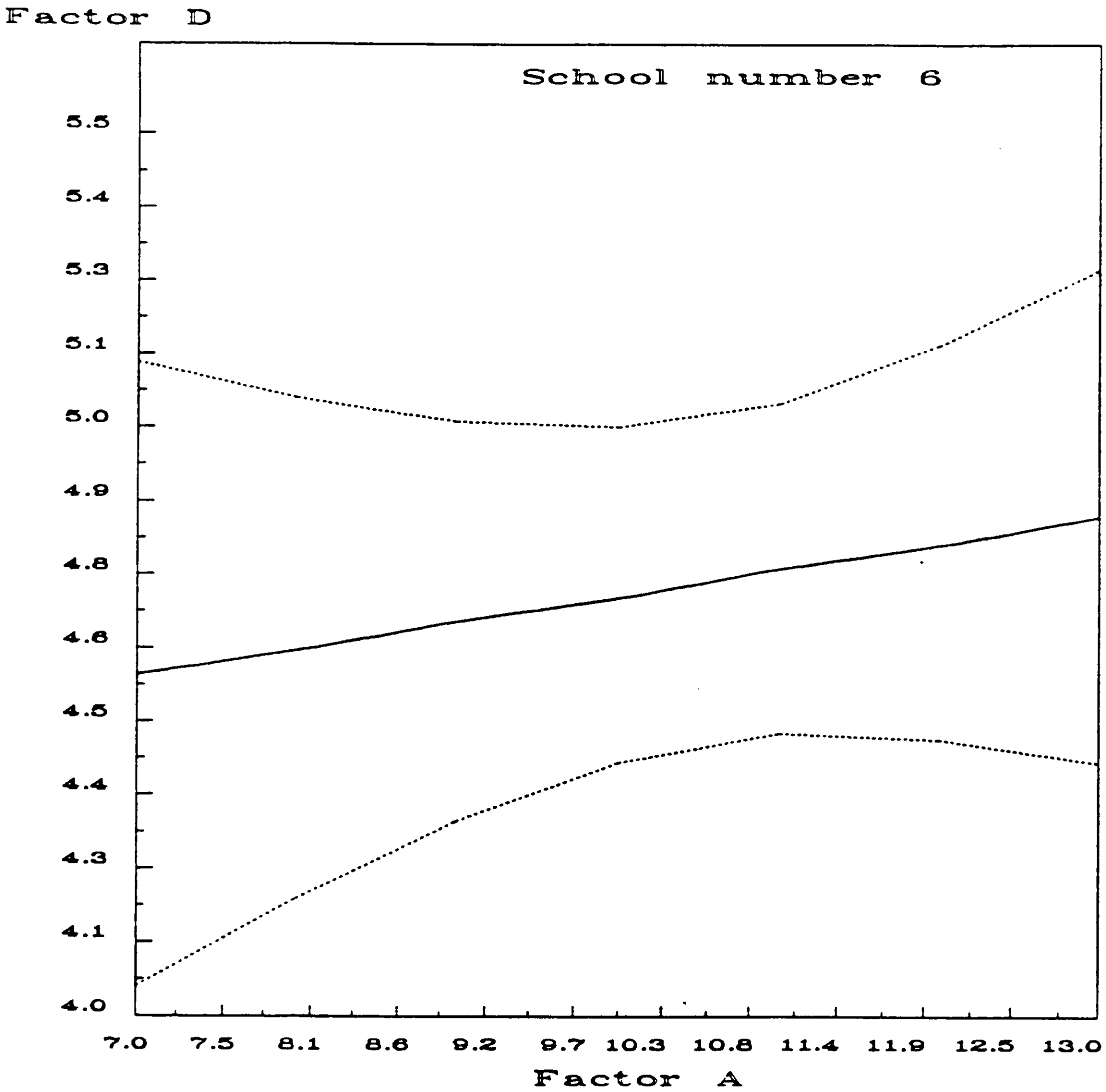


Fig. 3.5.3: Predicted scores and confidence limits for school 6.

3.6 SUMMARY

In this chapter the level-2 model considered in Section 2.3 was extended to allow for the model coefficients to be random across a third level of the hierarchy. Theoretical results required for estimation of the unknown parameters in the level-3 model were given in Sections 3.2 to 3.4. The optimization algorithm used to estimate the unknown parameters is known as the EM algorithm.

The estimation procedure in Section 3.4 was implemented in the FORTRAN program ML3_EM and a practical application was given in Section 3.5.

The EM algorithm is a fast, robust method for obtaining maximum likelihood estimates of the unknown parameters. In certain cases, however, a larger number of iterations may be required for the procedure to converge. A more serious disadvantage of the EM algorithm is that standard errors of the estimators cannot be obtained. The EM procedure also does not facilitate statistical inference such as hypothesis and contrast testing.

In the next chapter the so-called IGLS procedure will be introduced to address the problems and shortcomings associated with the EM algorithm.

APPENDIX 3.1

The terms $\mathbf{A}_{ij}$ and $\mathbf{b}_{ij}$ can be simplified as follows:

(i)
$$\mathbf{A}_{ij} = \mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\mathbf{X}_{(2)ij}$$
$$= \frac{1}{\Phi_{(1)}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}$$

(ii)
$$\mathbf{b}_{ij} = \mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\mathbf{y}_{ij}$$
$$= \frac{1}{\Phi_{(1)}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij}$$

where

$$\mathbf{R}_{ij} = \mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}$$

and

$$\Lambda_{ij}^{-1} = \frac{1}{\Phi_{(1)}}\,\mathbf{I} - \frac{1}{\Phi_{(1)}^{2}}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,.$$

Expression (i) can be derived as follows:

$$\mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\mathbf{X}_{(2)ij} = \mathbf{X}'_{(2)ij}\,\{\,\frac{1}{\Phi_{(1)}} - \frac{1}{\Phi_{(1)}^{2}}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\mathbf{X}'_{(2)ij}\}\mathbf{X}_{(2)ij}$$
$$= \frac{1}{\Phi_{(1)}}\,\{\,\mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\}\,\times$$
$$\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}$$
$$= \frac{1}{\Phi_{(1)}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,.$$

Expression (ii) can be simplified as

$$
\begin{aligned}
\mathbf{b}_{ij} &= \mathbf{X}'_{(2)ij}\, \Lambda_{ij}^{-1}\, \mathbf{y}_{ij} \\
&= \mathbf{X}'_{(2)ij} \left\{ \frac{1}{\Phi_{(1)}} - \frac{1}{\Phi^{2}_{(1)}} \mathbf{X}_{(2)ij} \left(\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\, \Phi_{(1)}^{-1}\, \mathbf{X}_{(2)ij} \right)^{-1} \mathbf{X}'_{(2)ij} \right\} \mathbf{y}_{ij} \\
&= \frac{1}{\Phi_{(1)}} \left\{ \mathbf{I} - \frac{1}{\Phi_{(1)}} \mathbf{X}'_{(2)ij}\, \mathbf{X}_{(2)ij} \left(\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\, \Phi_{(1)}^{-1}\, \mathbf{X}_{(2)ij} \right)^{-1} \right\} \mathbf{X}'_{(2)ij}\, \mathbf{y}_{ij} \\
&= \frac{1}{\Phi_{(1)}} \mathbf{R}'_{ij}\, \mathbf{X}'_{(2)ij}\, \mathbf{y}_{ij}\,.
\end{aligned}
$$

CHAPTER 4
THE IGLS ALGORITHM FOR PARAMETER ESTIMATION OF LEVEL-3 MODELS

4.1 INTRODUCTION

In this chapter attention will be given to the estimation of the unknown parameters of a level-3 model (see Section 3.2), using an IGLS algorithm. Advantages of using an IGLS procedure are that, in most instances, convergence is reached in a small number of iterations and that standard errors of the unknown parameters may be obtained. Statistical inference is also facilitated. It has also been shown (Goldstein, 1987 and Browne and du Toit, 1992), that under the assumption of multivariate normality, the IGLS estimation procedure is equivalent to normal Maximum Likelihood.

The IGLS procedure is discussed in Section 4.2. Simplified expressions of the terms involved are given in Section 4.3. Simplification of these terms ensure that the optimization algorithm is computationally efficient and that storage space requirements are greatly reduced. Statistical inference is discussed in Section 4.4 while a practical application is given in Section 4.5 and conclusions are given in Section 4.6.

4.2 PARAMETER ESTIMATION USING ITERATIVE GENERALIZED LEAST SQUARES

Let $\mathbf{X}_{(2)ij}$ denote the $n_{ij} \times m$ design matrix of the i-th level-3 and j-th level-2 experimental unit. The level-3 model (cf. Section 3.2) can be expressed in terms of the $n_i \times 1$ vector $\mathbf{y}_i$ as

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{i1} \\ \vdots \\ \mathbf{y}_{ij} \\ \vdots \\ \mathbf{y}_{in_i} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{(2)i1}\,\mathbf{b}_{i1} \\ \vdots \\ \mathbf{X}_{(2)ij}\,\mathbf{b}_{ij} \\ \vdots \\ \mathbf{X}_{(2)in_i}\,\mathbf{b}_{in_i} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{i1} \\ \vdots \\ \mathbf{e}_{ij} \\ \vdots \\ \mathbf{e}_{in_i} \end{bmatrix}, \quad i = 1, 2, \ldots, N, \qquad (4.2.1)$$

where

$$\mathbf{b}_{ij} \;=\; \mathbf{S}_i\,\mathbf{c}_i \,+\, \mathbf{u}_{ij} \;,\; j = 1, 2, \ldots, n_i$$

and where it is assumed that $\mathbf{u}_{i1}, \mathbf{u}_{i2}, \ldots, \mathbf{u}_{in_i}$ are identically and independently distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$. $\mathbf{S}_i$ is an $m \times q$ level-3 design matrix (cf. Section 3.2 and Example 3.2.1).

It is further assumed (cf. (3.2.9)) that

$$\mathbf{c}_i \;=\; \boldsymbol{\beta} \,+\, \mathbf{v}_i \,,$$

where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are identically and independently distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(3)}$. Under the distributional assumptions (cf. (3.2.10)) it follows that

$$\mathrm{E}(\mathbf{y}_i) \;=\; \mathbf{X}_{(3)i}\,\boldsymbol{\beta}\,, \qquad (4.2.2)$$

where

$$\mathbf{X}_{(3)i} \;=\; \mathbf{X}_{(2)i}\,\mathbf{S}_i \qquad (4.2.3)$$

and

$$\mathbf{X}_{(2)i} \;=\; \begin{bmatrix} \mathbf{X}_{(2)i1} \\ \mathbf{X}_{(2)i2} \\ \vdots \\ \vdots \\ \mathbf{X}_{(2)in_i} \end{bmatrix}. \qquad (4.2.4)$$

It also follows that

$$\Sigma_i = \text{Cov}(\, \mathbf{y}_i \, , \, \mathbf{y}_i' \,)$$

$$= \mathbf{X}_{(3)i} \, \boldsymbol{\Phi}_{(3)} \, \mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i \quad , \tag{4.2.5}$$

where

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \boldsymbol{\Lambda}_{i1} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{i2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & . & \boldsymbol{\Lambda}_{in_i} \end{bmatrix}$$

and where

$$\boldsymbol{\Lambda}_{ij} = \mathbf{X}_{(2)ij} \, \boldsymbol{\Phi}_{(2)} \, \mathbf{X}'_{(2)ij} + \Phi_{(1)} \, \mathbf{I}_{n_{ij}} \, . \tag{4.2.6}$$

A number of results which form the basis of the IGLS algorithm follow as Propositions 4.2.1 to 4.2.3. In order to implement these results in an efficient computer program further simplification of the mathematical expressions involved is required. This issue will dealt with in the next section.

REMARK 4.2.1

In the following propositions, new notation is introduced which requires some explanation. Let $\boldsymbol{\tau}$ be a $q^* \times 1$ vector with components the non-duplicated elements of $\boldsymbol{\Phi}_{(3)}$, $\boldsymbol{\Phi}_{(2)}$ and $\Phi_{(1)}$, where $q^* = \frac{1}{2}\, q(q+1) + \frac{1}{2}\, m(m+1) + 1$. That is,

$$\boldsymbol{\tau} = \begin{bmatrix} \text{vecs } \boldsymbol{\Phi}_{(3)} \\ \text{vecs } \boldsymbol{\Phi}_{(2)} \\ \Phi_{(1)} \end{bmatrix} . \tag{4.2.7}$$

Suppose that τ_k is the k-th approximation to the IGLS estimator $\hat{\tau}$ of τ, then it follows from (4.2.5) and (4.2.6) that $\Sigma_i(\tau_k)$ is the k-th approximation to the estimator $\hat{\Sigma}_i$ of Σ_i . In what follows it will be convenient to denote $\Sigma_i(\tau_k)$ by $\mathbf{V}_i$. Similarly, the notation β_k is used to denote the k-th approximation to the IGLS estimator $\hat{\beta}$ of the vector of fixed parameters β. □

PROPOSITION 4.2.1

The k-th approximation β_k of the iterative generalized least square estimator $\hat{\beta}$ of β is obtained from

$$\beta_k = [\sum_{i=1}^{N} \mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i}]^{-1} [\sum_{i=1}^{N} \mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{y}_i] . \tag{4.2.8}$$

Proof

Let

$$\mathbf{X}_{(3)} = \begin{bmatrix} \mathbf{X}_{(3)1} \\ \vdots \\ \mathbf{X}_{(3)i} \\ \vdots \\ \mathbf{X}_{(3)N} \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_i \\ \vdots \\ \mathbf{y}_N \end{bmatrix}$$

with $\mathbf{X}_{(3)i}$, $i = 1, 2, \ldots, N$ and $\mathbf{y}_i$, $i = 1, 2, \ldots, N$ defined by (4.2.3) and (4.2.1) respectively.

It follows that

$$E(\mathbf{y}) = \mathbf{X}_{(3)} \beta$$

and that (cf. Remark 4.2.1)

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & . & \mathbf{V}_N \end{bmatrix},$$

is the k-th approximation to the estimator $\hat{\Sigma}$ of Cov$(\mathbf{y}\ ,\ \mathbf{y}')$.

Consider the quadratic form

$$Q_{\boldsymbol{\beta}} = [\, \mathbf{y} - \mathbf{X}_{(3)}\boldsymbol{\beta} \,]'\, \mathbf{V}^{-1} [\, \mathbf{y} - \mathbf{X}_{(3)}\boldsymbol{\beta} \,] \ .$$

The k-th approximation to the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained as the solution to the equations $\dfrac{\partial Q_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}} = 0$, and hence

$$\begin{aligned} \boldsymbol{\beta}_k &= [\, \mathbf{X}'_{(3)}\, \mathbf{V}^{-1}\ \mathbf{X}_{(3)} \,]^{-1} [\, \mathbf{X}'_{(3)}\, \mathbf{V}^{-1}\ \mathbf{y} \,] \\ &= [\sum_{i=1}^{N} \mathbf{X}'_{(3)i}\, \mathbf{V}_i^{-1}\ \mathbf{X}_{(3)i} \,]^{-1} [\sum_{i=1}^{N} \mathbf{X}'_{(3)i}\, \mathbf{V}_i^{-1}\ \mathbf{y}_i \,]. \end{aligned}$$

□

PROPOSITION 4.2.2

Let

$$\mathbf{Z}^* = [\, \mathbf{Z}_{(2)i1} \quad \cdots \quad \mathbf{Z}_{(2)ij} \quad \cdots \quad \mathbf{Z}_{(2)in_i} \,]$$

where (cf. (4.2.4))

$$\mathbf{Z}'_{(2)ij} = [\ \mathbf{0}' \quad \cdots \quad \mathbf{0}' \quad \mathbf{X}'_{(2)ij} \quad \mathbf{0}' \quad \cdots \quad \mathbf{0}' \ ] \qquad (4.2.9)$$

and

$$\Phi^* = \begin{bmatrix} \Phi & 0 & . & . & 0 \\ 0 & \Phi & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & . & \Phi \end{bmatrix} .$$

Then

$$\begin{aligned} \text{vec } \mathbf{Z}^* \, \Phi^* \, \mathbf{Z}^{*\prime} &= \mathbf{Z}^* \otimes \mathbf{Z}^* \text{ vec } (\mathbf{I}_{n_i} \otimes \Phi) \\ &= \sum_{j=1}^{n_i} (\mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \text{ vec } \Phi . \end{aligned} \qquad (4.2.10)$$

Proof

$$\mathbf{Z}^* \, \Phi^* \, \mathbf{Z}^{*\prime} = \mathbf{Z}_{(2)i1} \, \Phi \, \mathbf{Z}'_{(2)i1} + \mathbf{Z}_{(2)i2} \, \Phi \, \mathbf{Z}'_{(2)i2} + \ldots + \mathbf{Z}_{(2)in_i} \, \Phi \, \mathbf{Z}'_{(2)in_i} .$$

Using the result (see for example Browne, 1974)

$$\text{vec } \mathbf{A} \, \mathbf{B} \, \mathbf{C} = \mathbf{C}' \otimes \mathbf{A} \text{ vec } \mathbf{B} \qquad (4.2.11)$$

it follows that

$$\begin{aligned} \text{vec } \mathbf{Z}^* \, \Phi^* \, \mathbf{Z}^{*\prime} &= \text{vec } (\mathbf{Z}_{(2)i1} \, \Phi \, \mathbf{Z}'_{(2)i1} + \mathbf{Z}_{(2)i2} \, \Phi \, \mathbf{Z}'_{(2)i2} + \ldots + \mathbf{Z}_{(2)in_i} \, \Phi \, \mathbf{Z}'_{(2)in_i}) \\ &= \text{vec } (\mathbf{Z}_{(2)i1} \, \Phi \, \mathbf{Z}'_{(2)i1}) + \ldots + \text{vec } (\mathbf{Z}_{(2)in_i} \, \Phi \, \mathbf{Z}'_{(2)in_i}) \\ &= (\mathbf{Z}_{(2)i1} \otimes \mathbf{Z}_{(2)i1}) \text{ vec } \Phi + \ldots + (\mathbf{Z}_{(2)in_i} \otimes \mathbf{Z}_{(2)in_i}) \text{ vec } \Phi \\ &= \sum_{j=1}^{n_i} (\mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \text{ vec } \Phi . \qquad \square \end{aligned}$$

PROPOSITION 4.2.3

Let

$$\mathbf{Y}_i = (\mathbf{y}_i - \mathbf{X}_{(3)i}\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_{(3)i}\boldsymbol{\beta})' , \tag{4.2.12}$$

$$\mathbf{y}_i^* = \text{vecs}(\mathbf{Y}_i) , \tag{4.2.13}$$

and (cf. Remark 4.2.1)

$$\boldsymbol{\tau} = \begin{bmatrix} \text{vecs } \boldsymbol{\Phi}_{(3)} \\ \text{vecs } \boldsymbol{\Phi}_{(2)} \\ \text{vecs } \boldsymbol{\Phi}_{(1)} \end{bmatrix} .$$

Suppose that $\mathbf{W}_i$ is a consistent estimator of the covariance matrix of vecs $(\mathbf{Y}_i^*)$. The k-th approximation $\boldsymbol{\tau}_k$ to the estimator $\hat{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$ (cf. (4.2.7)) is obtained from

$$\boldsymbol{\tau}_k = [\sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{X}_i^*]^{-1} [\sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^*] \tag{4.2.14}$$

where

$$\mathbf{X}_i^* = \mathbf{H}_{n_i^*} [(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \mathbf{G}_q \quad (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \mathbf{G}_m \quad \text{vec } \mathbf{I}] \tag{4.2.15}$$

with $\mathbf{Z}_{(2)ij}$ defined by (4.2.9) and where

$$\mathbf{W}_i^{-1} = \tfrac{1}{2} \mathbf{G}_{n_i}' (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} . \tag{4.2.16}$$

In (4.2.15), $\mathbf{H}_{n_i^*} : \frac{1}{2} n_i^*(n_i^* + 1) \times n_i^{*2}$ with $n_i^* = \sum_{j=1}^{n_i} n_{ij}$ is a non-unique matrix (Browne, 1974) such that

$$\text{vecs } \mathbf{A} = \mathbf{H}_{n_i^*} \text{vec } \mathbf{A} \tag{4.2.17}$$

where $\mathbf{A}$ is symmetric while in (4.2.16) $\mathbf{G}_m : m^2 \times \frac{1}{2}\, m(m+1)$ is an unique matrix such that

$$\text{vec } \mathbf{A} = \mathbf{G}_m \text{ vecs } \mathbf{A} . \qquad (4.2.18)$$

Proof

From (4.2.12) it follows that

$$\begin{aligned} \mathrm{E}(\mathbf{Y}_i) &= \mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i') \\ &= \Sigma_i \end{aligned}$$

where (cf. (4.2.5))

$$\Sigma_i = \mathbf{X}_{(3)i}\, \mathbf{\Phi}_{(3)}\, \mathbf{X}'_{(3)i} + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\, \mathbf{\Phi}_{(2)}\, \mathbf{Z}'_{(2)ij} + \Phi_{(1)}\, \mathbf{I}_{n_i} . \qquad (4.2.19)$$

Use of (4.2.11) and Proposition 4.2.2 gives

$$\begin{aligned} \text{vec } \Sigma_i = {} & (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \text{ vec } \mathbf{\Phi}_{(3)} + \sum_{j=1}^{n_i} (\mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \text{ vec } \mathbf{\Phi}_{(2)} + \\ & \text{vec } \mathbf{I}_{n_i}\, \Phi_{(1)} \end{aligned}$$

and therefore,

$$\begin{aligned} \mathrm{E}(\mathbf{y}_i^*) &= \text{vecs } \Sigma_i \\ &= \mathbf{X}_i^* \boldsymbol{\tau} \end{aligned}$$

where $\mathbf{X}_i^*$ is defined by (4.2.15). The covariance matrix of $\mathbf{y}_i^*$ can, under the assumption of multivariate normality, be expressed as (du Toit, 1993)

$$\mathrm{Cov}(\mathbf{y}_i^*, \mathbf{y}_i^{*\prime}) = \tfrac{1}{2}\, \mathbf{G}'_{n_i} (\Sigma_i^{-1} \otimes \Sigma_i^{-1})\, \mathbf{G}_{n_i} .$$

Since $\mathbf{V}_i = \Sigma_i(\tau_k)$, it follows that

$$\mathbf{W}_i^{-1} = \tfrac{1}{2}\, \mathbf{G}'_{n_i} \left(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1} \right) \mathbf{G}_{n_i}$$

is a consistent estimator of Cov($\mathbf{y}_i^*$, $\mathbf{y}_i^{*\prime}$).

Let

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* \\ \mathbf{X}_2^* \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{X}_N^* \end{bmatrix} \quad \text{and} \quad \mathbf{y}^* = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{y}_N^* \end{bmatrix}$$

then it follows that

$$\mathrm{E}(\, \mathbf{y}^* \,) = \mathbf{X}^* \tau$$

and

$$\mathrm{Cov}(\mathbf{y}^*, \mathbf{y}^{*\prime}) = \mathbf{W}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{W}_N \end{bmatrix}.$$

Consider the following quadratic form:

$$Q_T = [\, \mathbf{y}^* - \mathbf{X}^* \tau \,]' \, \mathbf{W}^{-1} [\, \mathbf{y}^* - \mathbf{X}^* \tau \,]$$

Then

$$\frac{\partial Q_T}{\partial \tau} = \mathbf{0}$$

yields

$$\begin{aligned} \tau_k &= [\, \mathbf{X}^{*\prime}\, \mathbf{W}^{-1}\ \mathbf{X}^{*}\,]^{-1}\, [\, \mathbf{X}^{*\prime}\, \mathbf{W}^{-1}\ \mathbf{y}^{*}\,] \\ &= [\sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\ \mathbf{X}_i^{*}\,]^{-1}\, [\sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\ \mathbf{y}_i^{*}\,]\,. \end{aligned} \qquad (4.2.20)$$

□

General comments

The IGLS estimators $\hat{\beta}$ of β and $\hat{\tau}$ of τ can be obtained as follows from Propositions 4.2.1 and 4.2.3 respectively :

(i) Set $\hat{\Phi}_{(1)} = 1$; $\hat{\mathbf{\Phi}}_{(2)} = \mathbf{0}$, $\hat{\mathbf{\Phi}}_{(3)} = \mathbf{0}$ so that (cf. Remark 4.2.1) $\mathbf{V}_i = \Phi_{(1)}\, \mathbf{I}_{n_i^*}$
(ii) Calculate β_k (cf. (4.2.8))
(iii) Calculate $\mathbf{Y}_i = (\, \mathbf{y}_i - \mathbf{X}_{(3)i}\, \beta_k\,)\, (\, \mathbf{y}_i - \mathbf{X}_{(3)i}\, \beta_k\,)'$
(iv) Calculate τ_k (cf. (4.2.14))
(v) Update $\mathbf{V}_i$ (cf. Proposition 4.2.1)

Repeat Steps (ii) to (v) until convergence is obtained. The algorithm described above is known as IGLS.

Although the IGLS algorithm described above appears to be straightforward, its actual implementation in a computer program is not. The first problem encountered is that the amount of storage space required for the matrices $\mathbf{V}_i$ (cf. Proposition 4.2.1) and $\mathbf{W}_i$ (cf. (4.2.16)) may exceed the total RAM (Random Access Memory) of the computer. Note that $\mathbf{V}_i$ is an $n_i^* \times n_i^*$ matrix where $n_i^* = \sum_{j=1}^{n_i} n_{ij}$. In practical applications it is not uncommon to find situations where n_i^* is large.

For example, a sample of $n_i = 20$ schools may have been drawn from a given education department. From each of these schools, a simple random sample of size $n_{ij} = 50$, $j = 1, 2, \ldots, 20$, pupils are drawn. Therefore $n_i^* = 1000$. Hence the storage space required for $\mathbf{V}_i$ is 8 Megabytes (8 bytes for each of the 1000^2 elements). $\mathbf{W}_i$ is a $k^* \times k^*$ matrix where $k^* = \frac{1}{2}\, n_i^* \,(n_i^* + 1)$. For $n_i^* = 1000$, $\mathbf{W}_i$ is $(500 \times 1001) \times (500 \times 1001)$. Presently, most personal computers have a maximum of 4 Megabytes RAM available when executing programs.

An even more serious problem is that of the computation time required for matrix inversion. In general, n^3 operations are required to invert a matrix of order n (Press *et al*, 1988). See, for example, (4.2.8) and (4.2.14) for expressions involving matrix inversion.

4.3 COMPUTER IMPLEMENTATION OF THE IGLS ALGORITHM

In this section the problems discussed above will be addressed. It will be shown that the order of matrices to be inverted may be reduced to order m and q where m and q denote the number of variables random at level-2 and level-3 respectively. Furthermore, all the terms required to calculate the IGLS estimators are written in the form $\mathbf{X}'\,\mathbf{X}$, $\mathbf{X}'\,\mathbf{y}$ or $\mathbf{y}'\,\mathbf{y}$, that is, matrices of squares and cross products of the data.

Before simplification of the terms of $\mathbf{X}^{*\prime}\,\mathbf{W}^{-1}\,\mathbf{X}^*$ and $\mathbf{X}^{*\prime}\,\mathbf{W}^{-1}\,\mathbf{y}^*$ respectively, note that $\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^*$ can be written as follows (cf. (4.2.15) and (4.2.16)) :

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^* = \begin{bmatrix} \mathbf{G}_q'\,(\,\mathbf{X}_{(3)i}' \otimes \mathbf{X}_{(3)i}'\,) \\ \mathbf{G}_m'\,(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}'\,) \\ (\text{vec}\,\mathbf{I})' \end{bmatrix} \mathbf{H}_{n_i^*}'\,\mathbf{G}_{n_i}'\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{H}_{n_i^*}$$

$$\times\,[\,(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}\,)\,\mathbf{G}_q \quad (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}\,)\,\mathbf{G}_m \quad \text{vec}\,\mathbf{I}\,] \tag{4.3.1}$$

4.3.1: SIMPLIFICATION OF THE COMPONENTS OF THE WEIGHT MATRIX

Symbolically, (4.3.1) can be written as

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^{*} = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} & \mathbf{T}'_{(1,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} & \mathbf{T}'_{(1,2)} \\ \mathbf{T}_{(1,3)} & \mathbf{T}_{(1,2)} & \mathbf{T}_{(1,1)} \end{bmatrix}.$$

Computationally efficient expressions for each of the submatrices $\mathbf{T}_{(3,3)}$, $\mathbf{T}_{(2,3)}$, . . . $\mathbf{T}_{(1,1)}$ are given in Propositions 4.3.1 to 4.3.6 of this section. The results derived in these propositions are subsequently summarized in Theorem 4.3.1.

PROPOSITION 4.3.1 (Submatrix $\mathbf{T}_{(3,3)}$)

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q\,(\,\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}\,)\,\mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{H}_{n_i^*}\,(\,\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}\,)\,\mathbf{G}_q$$

$$= \mathbf{G}'_q\,(\,\mathbf{Q}_i\,[\,\mathbf{I} - \mathbf{C}_i\;\mathbf{Q}_i\,] \otimes \mathbf{Q}_i\,[\,\mathbf{I} - \mathbf{C}_i\;\mathbf{Q}_i\,]\,)\,\mathbf{G}_q \tag{4.3.2}$$

where

$$\mathbf{Q}_i = \mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{A}_{ij}\,\mathbf{S}_i \tag{4.3.3}$$

and

$$\mathbf{A}_{ij} = \mathbf{X}'_{(2)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(2)ij}\,. \tag{4.3.4}$$

Proof

Using the result (Magnus & Neudecker, 1988)

$$\mathbf{G}_k\,\mathbf{H}_k\,(\,\mathbf{X} \otimes \mathbf{X}\,)\,\mathbf{G}_m = (\,\mathbf{X} \otimes \mathbf{X}\,)\,\mathbf{G}_m\,, \tag{4.3.5}$$

(4.3.2) can be rewritten as follows:

$$\mathbf{G}'_q \; (\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}) \, \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \, \mathbf{G}_{n_i} \mathbf{H}_{n_i^*} (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q$$

$$= \mathbf{G}'_q \; (\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}) (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q \qquad (4.3.6)$$

From the result (Graham, 1981)

$$(\mathbf{A}' \otimes \mathbf{B}') (\mathbf{C} \otimes \mathbf{C}) (\mathbf{A} \otimes \mathbf{B}) = \mathbf{A}' \mathbf{C} \mathbf{A} \otimes \mathbf{B}' \mathbf{C} \mathbf{B} \qquad (4.3.7)$$

(4.3.6) can be rewritten as

$$\mathbf{G}'_q \; (\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}) (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q$$

$$= \mathbf{G}'_q \; (\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i}) \, \mathbf{G}_q \, . \qquad (4.3.8)$$

From (4.2.5) it follows that

$$\Sigma_i^{-1} = (\mathbf{X}_{(3)i} \, \Phi_{(3)} \, \mathbf{X}'_{(3)i} + \Lambda_i)^{-1}$$

$$= \Lambda_i^{-1} - \Lambda_i^{-1} \mathbf{X}_{(3)i} \, \mathbf{C}_i \, \mathbf{X}'_{(3)i} \Lambda_i^{-1} \qquad (4.3.9)$$

where (cf. (3.3.8))

$$\mathbf{C}_i = (\Phi_{(3)}^{-1} + \mathbf{X}'_{(3)i} \Lambda_i^{-1} \mathbf{X}_{(3)i})^{-1} . \qquad (4.3.10)$$

This follows from a general result for matrix inversion

$$(\mathbf{A} + \mathbf{B} \mathbf{C} \mathbf{B}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{B}' \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{A}^{-1} . \qquad (4.3.11)$$

Thus $\mathbf{X}'_{(3)i} \Sigma_i^{-1} \mathbf{X}'_{(3)i}$ may be written (cf. (4.3.9)) as

$$\mathbf{X}'_{(3)i} \Sigma_i^{-1} \mathbf{X}'_{(3)i} = \mathbf{X}'_{(3)i} \Lambda_i^{-1} \mathbf{X}_{(3)i} - \mathbf{X}'_{(3)i} \Lambda_i^{-1} \mathbf{X}_{(3)i} \, \mathbf{C}_i \, \mathbf{X}'_{(3)i} \Lambda_i^{-1} \mathbf{X}_{(3)i} \qquad (4.3.12)$$

where (cf. (4.3.3))

$$\begin{aligned} X'_{(3)i}\,\Lambda_i^{-1}\,X_{(3)i} &= S'_i\,X'_{(2)i}\,\Lambda_i^{-1}\,X_{(2)i}\,S_i \\ &= S'_i\sum_{j=1}^{n_i}(\,X'_{(2)ij}\,\Lambda_{ij}^{-1}\,X_{(2)ij}\,)\,S_i \\ &= Q_i \end{aligned} \qquad (4.3.13)$$

so that (cf. (4.3.12))

$$X'_{(3)i}\,\Sigma_i^{-1}\,X_{(3)i} = Q_i\,[\,I - C_i\,Q_i\,]$$

which, with Σ_i replaced by V_i, concludes the proof. □

PROPOSITION 4.3.2 (Submatrix $T_{(2,3)}$)

$$T_{(2,3)} = G'_m\,(\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}\,)\,H'_{n_i^*}\,G'_{n_i}\,(\,V_i^{-1} \otimes V_i^{-1}\,)\,G_{n_i}\,H_{n_i^*}\,(\,X_{(3)i} \otimes X_{(3)i}\,)\,G_q$$

$$= G'_m \sum_{j=1}^{n_i}\,(\,A_{ij}\,S_i\,[\,I - C_i\,Q_i\,] \otimes A_{ij}\,S_i\,[\,I - C_i\,Q_i\,]\,)\,G_q\,. \qquad (4.3.14)$$

Proof

From (4.3.5) and (4.3.7) it follows that

$$\begin{aligned} &G'_m\,(\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}\,)\,H'_{n_i^*}\,G'_{n_i}\,(\,V_i^{-1} \otimes V_i^{-1}\,)\,G_{n_i}\,H_{n_i^*}\,(\,X_{(3)i} \otimes X_{(3)i}\,)\,G_q \\ &= G'_m\,(\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}\,)\,(\,V_i^{-1} \otimes V_i^{-1}\,)\,(\,X_{(3)i} \otimes X_{(3)i}\,)\,G_q \\ &= G'_m\,(\sum_{j=1}^{n_i} Z'_{(2)ij}\,V_i^{-1}\,X_{(3)i} \otimes \sum_{j=1}^{n_i} Z'_{(2)ij}\,V_i^{-1}\,X_{(3)i}\,)\,G_q\,. \end{aligned} \qquad (4.3.15)$$

$Z'_{(2)ij}\, V_i^{-1}\, X_{(3)i}$ may be written as (cf. (4.3.9))

$$Z'_{(2)ij}\, V_i^{-1}\, X_{(3)i} = Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(3)i} - Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(3)i}\; C_i\, X'_{(3)i}\, \Lambda_i^{-1}\, X_{(3)i} \qquad (4.3.16)$$

where

$$\begin{aligned} Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(3)i} &= Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(2)i}\, S_i \\ &= X'_{(2)ij}\, \Lambda_{ij}^{-1}\, X_{(2)ij}\, S_i \\ &= A_{ij}\, S_i \qquad (4.3.17) \end{aligned}$$

and (cf. (4.3.13))

$$X'_{(3)i}\, \Lambda_i^{-1}\, X_{(3)i} = Q_i\, . \qquad (4.3.18)$$

Substitution of (4.3.17) and (4.3.18) in (4.3.16) gives

$$Z'_{(2)ij}\, V_i^{-1}\, X_{(3)i} = A_{ij}\, S_i\, [\, I - C_i\, Q_i\,]$$

which concludes the proof. □

PROPOSITION 4.3.3 (Submatrix $T_{(2,2)}$)

$$\begin{aligned} T_{(2,2)} &= G'_m\, (\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}\,)\, H'_{n_i^*}\, G'_{n_i}\, (V_i^{-1} \otimes V_i^{-1})\, G_{n_i}\, H_{n_i^*}\, (\sum_{j^*=1}^{n_i} Z'_{(2)ij^*} \otimes Z'_{(2)ij^*}\,)\, G_m \\ &= G'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} (A_{ij}\, [\, \delta_{jj^*} I - S_i\, C_i\, S'_i\; A_{ij^*}\,] \otimes A_{ij}\, [\, \delta_{jj^*} I - S_i\, C_i\, S'_i\, A_{ij^*}\,]\, G_m\, . \end{aligned} \qquad (4.3.19)$$

Proof

Using results (4.3.5) and (4.3.7), it follows that

$$G'_m\, (\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}\,)\, H'_{n_i^*}\, G'_{n_i}\, (V_i^{-1} \otimes V_i^{-1})\, G_{n_i}\, H_{n_i^*}\, (\sum_{j^*=1}^{n_i} Z'_{(2)ij^*} \otimes Z'_{(2)ij^*}\,)\, G_m$$

$$= \mathbf{G}'_m \; (\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij}) \; (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \; (\sum_{j^*=1}^{n_i} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}_{(2)ij^*}) \; \mathbf{G}_m$$

$$= \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} (\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*}) \; \mathbf{G}_m \, . \tag{4.3.20}$$

But $\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*}$ can be written as (cf. (4.3.9))

$$\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} = \mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} - \mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}'_{(3)i} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} \tag{4.3.21}$$

where

$$\mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} = \delta_{jj^*} \mathbf{X}_{(2)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}'_{(2)ij} \tag{4.3.22}$$

$$= \delta_{jj^*} \mathbf{A}_{ij}$$

with

$$\delta_{jj^*} = 1 \text{ if } j = j^*$$
$$= 0 \text{ if } j \text{ is not equal to } j^* \, .$$

Also, from (4.3.17)

$$\mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(3)i} = \mathbf{A}_{ij} \mathbf{S}_i \, . \tag{4.3.23}$$

Substitution of (4.3.22) and (4.3.23) in (4.3.21) gives

$$\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} = \mathbf{A}_{ij} [\, \delta_{jj^*} \mathbf{I} - \mathbf{S}_i \mathbf{C}_i \mathbf{S}'_i \mathbf{A}_{ij^*}]$$

which concludes the proof. □

PROPOSITION 4.3.4 (Submatrix $\mathbf{T}_{(1,3)}$)

$$\mathbf{T}_{(1,3)} = (\text{ Vec } \mathbf{I})' \; \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{H}_{n_i^*} (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \mathbf{G}_q$$

$$= [\mathbf{G}'_q \text{ vec } \{ (\mathbf{I} - \mathbf{Q}_i \mathbf{C}_i) \mathbf{S}'_i \{ \sum_{j=1}^{n_i} \mathbf{E}_{ij} \} \mathbf{S}_i (\mathbf{I} - \mathbf{Q}_i \mathbf{C}_i) \}]' . \qquad (4.3.24)$$

Proof

Using (4.3.5) and (4.3.7) it follows that

$$(\text{ Vec } \mathbf{I})' \; \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{H}_{n_i^*} (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \mathbf{G}_q$$

$$= (\text{ Vec } \mathbf{I})' \; \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{V}_i^{-1} \mathbf{X}_{(3)i}) \mathbf{G}_q . \qquad (4.3.25)$$

Using (4.3.5) and (4.2.11), (4.3.25) can be rewritten as

$$(\text{ Vec } \mathbf{I})' \; \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{V}_i^{-1} \mathbf{X}_{(3)i}) \mathbf{G}_q$$

$$= [\mathbf{G}'_q \text{ vec } (\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i})]'. \qquad (4.3.26)$$

From (4.3.9) it follows that

$$\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} = \mathbf{X}'_{(3)i} \mathbf{\Lambda}_i^{-1} - \mathbf{X}'_{(3)i} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \; \mathbf{X}_{(3)i} \mathbf{\Lambda}_i^{-1}$$

which, using (4.3.13), reduces to

$$\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} = [\mathbf{I} - \mathbf{Q}_i \mathbf{C}_i] \; \mathbf{X}'_{(3)i} \mathbf{\Lambda}_i^{-1} \; .$$

Let

$$\mathbf{E}_{ij} = \mathbf{X}'_{(2)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)ij} \qquad (4.3.27)$$

(for further simplification of $\mathbf{E}_{ij}$ see Appendix 4.1.)

From (4.3.26) it then follows that

$$\begin{aligned} X'_{(3)i} V_i^{-1} V_i^{-1} X_{(3)i} &= [I - Q_i C_i] \; X'_{(3)i} \Lambda_i^{-1} \Lambda_i^{-1} X_{(3)i} \; [I - Q_i C_i]' \\ &= [I - Q_i C_i] \sum_{j=1}^{n_i} (S'_i X'_{(2)ij} \Lambda_{ij}^{-1} \Lambda_{ij}^{-1} X_{(2)ij} S_i) \times \\ &\quad [I - Q_i C_i] \\ &= [I - Q_i C_i] S'_i \{ \sum_{j=1}^{n_i} E_{ij} \} S_i [I - Q_i C_i] . \end{aligned} \tag{4.3.28}$$

Substitution of (4.3.28) in (4.3.26) concludes the proof. □

PROPOSITION 4.3.5 (Submatrix $T_{(1,2)}$)

$$\begin{aligned} T_{(1,2)} &= (\mathrm{Vec}\, I)' \; H'_{n_i^*} G'_{n_i} (V_i^{-1} \otimes V_i^{-1}) G_{n_i} H_{n_i^*} (\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}) G_m \\ &= \sum_{j=1}^{n_i} [G'_m \mathrm{vec} \{ E_{ij} (I - S_i C_i S'_i A_{ij}) - \\ &\quad A_{ij} S_i C_i S'_i (E_{ij} - (\sum_{j=1}^{n_i} E_{ij}) S_i C_i S'_i A_{ij}) \}]' . \end{aligned} \tag{4.3.29}$$

Proof

From (4.2.11), (4.3.5) and (4.3.7), it follows that

$$\begin{aligned} &(\mathrm{Vec}\, I)' \; H'_{n_i^*} G'_{n_i} (V_i^{-1} \otimes V_i^{-1}) G_{n_i} H_{n_i^*} (\sum_{j=1}^{n_i} Z'_{(2)ij} \otimes Z'_{(2)ij}) G_m \\ &= \sum_{j=1}^{n_i} [G'_m \mathrm{vec} (Z'_{(2)ij} V_i^{-1} V_i^{-1} Z_{(2)ij})]' \end{aligned} \tag{4.3.30}$$

where $Z'_{(2)ij} V_i^{-1} V_i^{-1} Z_{(2)ij}$ can be rewritten (cf. (4.3.9)) as

$$\begin{aligned} Z'_{(2)ij} V_i^{-1} V_i^{-1} Z_{(2)ij} &= Z'_{(2)ij} \Lambda_i^{-1} \Lambda_i^{-1} Z_{(2)ij} - \\ &\quad Z'_{(2)ij} \Lambda_i^{-1} \Lambda_i^{-1} X_{(3)i} \; C_i X'_{(3)i} \Lambda_i^{-1} Z_{(2)ij} - \end{aligned}$$

$$Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(3)i}\, C_i\, X'_{(3)i}\, \Lambda_i^{-1}\, \Lambda_i^{-1}\, Z_{(2)ij} \; +$$

$$Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(3)i} \times$$

$$C_i\, X'_{(3)i}\, \Lambda_i^{-1}\, \Lambda_i^{-1}\, X_{(3)i}\, C_i\, X'_{(3)i}\, \Lambda_i^{-1}\, Z_{(2)ij}\,. \tag{4.3.31}$$

From (4.3.17), where

$$Z'_{(2)ij}\, \Lambda_i^{-1}\, X_{(3)i} = A_{ij}\, S_i \;,$$

and from (4.3.27), where

$$E_{ij} \;=\; X'_{(2)ij}\, \Lambda_{ij}^{-1}\, \Lambda_{ij}^{-1}\, X_{(2)ij}\,,$$

it follows that (4.3.31) can be written as

$$Z'_{(2)ij}\, V_i^{-1}\, V_i^{-1}\; Z_{(2)ij} \;=\; E_{ij}\,(\, I \,-\, S_i\, C_i\, S'_i\, A_{ij}\,) \;-$$

$$A_{ij}\, S_i\, C_i\, S'_i\,(\, E_{ij} \;-\; (\sum_{j=1}^{n_i} E_{ij}\,)\, S_i\, C_i\; S'_i\, A_{ij}\,)\,. \tag{4.3.32}$$

Substitution of (4.3.32) in (4.3.30) concludes the proof. □

PROPOSITION 4.3.6 (Submatrix $T_{(1,1)}$)

$$T_{(1,1)} = \mathrm{tr}\,(\, V_i^{-1}\, V_i^{-1}\,)\, G_{n_i} = \frac{n_{ij} - m}{\Phi^2_{(1)}}\; \mathrm{tr}\,\{\, \sum_{j=1}^{n_i} R'_{ij}\, R'_{ij}\,\} \;-$$

$$\frac{2}{\Phi^3_{(1)}}\; \mathrm{tr}\,\{ \sum_{j=1}^{n_i} S'_i\, X'_{(2)ij}\, X_{(2)ij}\, R'_{ij}\, R'_{ij}\, R'_{ij}\, S_i\, C_i\,\} \;+$$

$$\frac{1}{\Phi^4_{(1)}}\; \{ \sum_{j=1}^{n_i} S'_i\, X'_{(2)ij}\, X_{(2)ij}\, R'_{ij}\, R'_{ij}\, S_i\, C_i\,\} \;\times$$

$$\{ \sum_{j=1}^{n_i} S'_i\, X'_{(2)ij}\, X_{(2)ij}\, R'_{ij}\, R'_{ij}\, S_i\, C_i\,\} \tag{4.3.33}$$

where (cf. Appendix 3.1)

$$\mathbf{R}_{ij} = [\, \mathbf{I} - \frac{1}{\Phi_{(1)}} \mathbf{X}'_{(2)ij} \mathbf{X}_{(2)ij} (\, \Phi_{(2)}^{-1} + \mathbf{X}_{(2)ij} \Phi_{(1)}^{-1} \mathbf{X}'_{(2)ij})^{-1}] \tag{4.3.34}$$

and $\mathbf{C}_i$ is given by (4.3.10).

Proof

Let

$$\mathbf{M}_i = \mathbf{X}'_{(3)i} \Lambda_i^{-1} \, . \tag{4.3.35}$$

From (4.3.9) it follows that $\mathbf{V}_i^{-1} \mathbf{V}_i^{-1}$ may be written as

$$\begin{aligned} \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} &= (\, \Lambda_i^{-1} - \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i) (\, \Lambda_i^{-1} - \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i) \\ &= \Lambda_i^{-1} \Lambda_i^{-1} - \Lambda_i^{-1} \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i - \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i \Lambda_i^{-1} + \\ &\quad \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i \end{aligned}$$

so that

$$\begin{aligned} \mathrm{tr}\, (\mathbf{V}_i^{-1} \mathbf{V}_i^{-1}) = {} & \mathrm{tr}\, \{\, \Lambda_i^{-1} \Lambda_i^{-1} \} - 2\, \mathrm{tr}\, \{\, \mathbf{C}_i \mathbf{M}_i \Lambda_i^{-1} \mathbf{M}'_i \} + \\ & \mathrm{tr}\, \{\mathbf{C}_i \mathbf{M}_i \mathbf{M}'_i \mathbf{C}_i \mathbf{M}_i \mathbf{M}'_i \} \, . \end{aligned} \tag{4.3.36}$$

After simplification, the respective terms of (4.3.36) can be rewritten as (cf. Appendix 3.1)

$$\mathrm{tr}\, \{\, \Lambda_i^{-1} \Lambda_i^{-1} \} = \frac{n_{ij} - m}{\Phi_{(1)}^2} \, \mathrm{tr}\, \{ \sum_{j=1}^{n_i} \mathbf{R}'_{ij} \mathbf{R}'_{ij} \} \, , \tag{4.3.37}$$

$$\mathrm{tr}\, \{\, \mathbf{C}_i \mathbf{M}_i \Lambda_i^{-1} \mathbf{M}'_i \} = \frac{1}{\Phi_{(1)}^3} \, \mathrm{tr}\, \{ \sum_{j=1}^{n_i} \mathbf{S}'_i \mathbf{X}'_{(2)ij} \mathbf{X}_{(2)ij} \mathbf{R}'_{ij} \mathbf{R}'_{ij} \mathbf{R}'_{ij} \mathbf{S}_i \} \tag{4.3.38}$$

and

$$\text{tr}\,\{\mathbf{C}_i\,\mathbf{M}_i\,\mathbf{M}_i'\,\mathbf{C}_i\,\mathbf{M}_i\,\mathbf{M}_i'\,\} = \frac{1}{\Phi^4_{(1)}}\,\text{tr}\,\{\sum_{j=1}^{n_i}\mathbf{S}_i'\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,\mathbf{R}_{ij}'\,\mathbf{R}_{ij}'\,\mathbf{S}_i\,\mathbf{C}_i\,\} \times \{\sum_{j=1}^{n_i}\mathbf{S}_i'\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,\mathbf{R}_{ij}'\,\mathbf{R}_{ij}'\,\mathbf{S}_i\,\mathbf{C}_i\,\}\,. \tag{4.3.39}$$

Substitution of (4.3.37), (4.3.38) and (4.3.39) in (4.3.36) concludes the proof. □

THEOREM 4.3.1

Let

$$\mathbf{X}_i^* = \mathbf{H}_{n_i^*}\,[\,(\mathbf{X}_{(3)i}\otimes\mathbf{X}_{(3)i}\,)\,\mathbf{G}_q \quad (\sum_{j=1}^{n_i}\mathbf{Z}_{(2)ij}\otimes\mathbf{Z}_{(2)ij}\,)\,\mathbf{G}_m \qquad \text{vec}\,\mathbf{I}\,]$$

and

$$2\mathbf{W}_i = \mathbf{G}'_{n_i}\,(\,\mathbf{V}_i^{-1}\otimes\mathbf{V}_i^{-1}\,)\;\mathbf{G}_{n_i}$$

as defined by (4.2.15) and (4.2.16) respectively.

The matrix $\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^*$ can be written as

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^* = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} & \mathbf{T}'_{(1,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} & \mathbf{T}'_{(1,2)} \\ \mathbf{T}_{(1,3)} & \mathbf{T}_{(1,2)} & \mathbf{T}_{(1,1)} \end{bmatrix}$$

where the respective terms are

$$\mathbf{T}_{(3,3)} = \mathbf{G}_q'\,(\,\mathbf{Q}_i\,[\,\mathbf{I}-\mathbf{C}_i\;\mathbf{Q}_i\,]\otimes\mathbf{Q}_i\,[\,\mathbf{I}-\mathbf{C}_i\;\mathbf{Q}_i\,])\,\mathbf{G}_q \tag{4.3.40}$$

$$\mathbf{T}_{(2,3)} = \mathbf{G}_m'\sum_{j=1}^{n_i}(\,\mathbf{A}_{ij}\,\mathbf{S}_i\;[\,\mathbf{I}-\mathbf{C}_i\,\mathbf{Q}_i\,]\otimes\mathbf{A}_{ij}\,\mathbf{S}_i\;[\,\mathbf{I}-\mathbf{C}_i\,\mathbf{Q}_i\,])\,\mathbf{G}_q \tag{4.3.41}$$

$$\mathbf{T}_{(2,2)} = \mathbf{G}_m' \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} (\mathbf{A}_{ij} [\, \delta_{jj^*} \mathbf{I} - \mathbf{S}_i \mathbf{C}_i \mathbf{S}_i' \mathbf{A}_{ij^*}] \otimes \mathbf{A}_{ij} [\, \delta_{jj^*} \mathbf{I} - \mathbf{S}_i \mathbf{C}_i \mathbf{S}_i' \mathbf{A}_{ij^*}] \,\mathbf{G}_m \quad (4.3.42)$$

$$\mathbf{T}_{(1,3)} = [\, \mathbf{G}_q' \text{ vec } \{ [\, \mathbf{I} - \mathbf{Q}_i \mathbf{C}_i] \mathbf{S}_i' \{ \sum_{j=1}^{n_i} \mathbf{E}_{ij} \} \mathbf{S}_i [\, \mathbf{I} - \mathbf{Q}_i \mathbf{C}_i]\}]' \quad (4.3.43)$$

$$\mathbf{T}_{(1,2)} = \sum_{j=1}^{n_i} [\, \mathbf{G}_m' \text{ vec } \{ \mathbf{E}_{ij} (\, \mathbf{I} - \mathbf{S}_i \mathbf{C}_i \mathbf{S}_i' \mathbf{A}_{ij}) - \mathbf{A}_{ij} \mathbf{S}_i \mathbf{C}_i \mathbf{S}_i' (\, \mathbf{E}_{ij} - (\sum_{j=1}^{n_i} \mathbf{E}_{ij}) \mathbf{S}_i \mathbf{C}_i \mathbf{S}_i' \mathbf{A}_{ij})\}]' \quad (4.3.44)$$

$$\mathbf{T}_{(1,1)} = \frac{n_{ij} - m}{\Phi_{(1)}^2} \text{ tr } \{ \sum_{j=1}^{n_i} \mathbf{R}_{ij}' \mathbf{R}_{ij}' \} - \frac{2}{\Phi_{(1)}^3} \text{ tr } \{ \sum_{j=1}^{n_i} \mathbf{S}_i' \mathbf{X}_{(2)ij}' \mathbf{X}_{(2)ij} \mathbf{R}_{ij}' \mathbf{R}_{ij}' \mathbf{R}_{ij}' \mathbf{S}_i \mathbf{C}_i \} + \frac{1}{\Phi_{(1)}^4} \{ \sum_{j=1}^{n_i} \mathbf{S}_i' \mathbf{X}_{(2)ij}' \mathbf{X}_{(2)ij} \mathbf{R}_{ij}' \mathbf{R}_{ij}' \mathbf{S}_i \mathbf{C}_i \} \times \{ \sum_{j=1}^{n_i} \mathbf{S}_i' \mathbf{X}_{(2)ij}' \mathbf{X}_{(2)ij} \mathbf{R}_{ij}' \mathbf{R}_{ij}' \mathbf{S}_i \mathbf{C}_i \} \quad (4.3.45)$$

and where $\mathbf{A}_{ij}$, $\mathbf{Q}_i$, $\mathbf{C}_i$, $\mathbf{E}_{ij}$ and $\mathbf{R}_{ij}$ are given by (4.3.4), (4.3.3), (4.3.10), (4.3.27) and (4.3.34) respectively.

Proof

The proof follows directly from Propositions 4.3.1 to 4.3.6. □

In this section expressions required to efficiently compute the matrix $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^*$ were derived. To calculate $\hat{\tau}$ (cf. (4.2.14) and (4.2.20)) and hence $\hat{\Phi}_{(3)}$, $\hat{\Phi}_{(2)}$ and $\hat{\Phi}_{(1)}$ it is also necessary to find a computationally efficient way to calculate $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{y}_i^*$. This matter will be dealt with in the next section.

4.3.2 SIMPLIFICATION OF THE COMPONENTS OF THE COEFFICIENT VECTOR

From (4.2.15) and (4.2.16) it follows that $\mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^*$ can be written as

$$2\, \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^* = \begin{bmatrix} \mathbf{G}_q' \left(\mathbf{X}_{(3)i}' \otimes \mathbf{X}_{(3)i}' \right) \\ \mathbf{G}_m' \left(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}' \right) \\ (\text{vec } \mathbf{I})' \end{bmatrix} \mathbf{H}_{n_i^*}'\, \mathbf{G}_{n_i}' \left(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1} \right) \mathbf{G}_{n_i}\, \mathbf{H}_{n_i^*}\, \mathbf{y}_i^*. \tag{4.3.46}$$

Symbolically, (4.3.46) can be written as

$$2\, \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^* = \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \\ \mathbf{q}_{(1)} \end{bmatrix} .$$

Computationally efficient expressions for each of the subvectors $\mathbf{q}_{(3)}$, $\mathbf{q}_{(2)}$ and $\mathbf{q}_{(1)}$ are given in Propositions 4.3.7 to 4.3.9. These results are summarized in Theorem 4.3.2.

PROPOSITION 4.3.7 (subvector $\mathbf{q}_{(3)}$)

$$\begin{aligned} \mathbf{q}_{(3)} &= \mathbf{G}_q' \left(\mathbf{X}_{(3)i}' \otimes \mathbf{X}_{(3)i}' \right) \mathbf{H}_{n_i^*}'\, \mathbf{G}_{n_i}' \left(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1} \right) \mathbf{G}_{n_i}\, \mathbf{y}_i^* \\ &= \mathbf{G}_q' \text{ vec } \{ [\, \mathbf{I} - \mathbf{Q}_i\, \mathbf{C}_i \,]\, [\mathbf{S}_i \, (\sum_{j=1}^{n_i} \mathbf{b}_{ij}) - \mathbf{Q}_i\, \boldsymbol{\beta} \,] \} \times \\ &\qquad \{ [\, \mathbf{I} - \mathbf{Q}_i\, \mathbf{C}_i \,]\, [\mathbf{S}_i \, (\sum_{j=1}^{n_i} \mathbf{b}_{ij}) - \mathbf{Q}_i\, \boldsymbol{\beta} \,] \}' \end{aligned} \tag{4.3.47}$$

where (cf. (3.3.12))

$$\mathbf{b}_{ij} = \mathbf{X}_{(2)ij}'\, \boldsymbol{\Lambda}_i^{-1}\, \mathbf{y}_{ij} \tag{4.3.48}$$

and $\mathbf{Q}_i$ is given by (4.3.3). (For further simplification of $\mathbf{b}_{ij}$ see Appendix 4.1.)

Proof

From (4.3.5) and (4.3.7) it follows that

$$\mathbf{G}'_q\,(\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}\,)\,\mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*$$

$$= \mathbf{G}'_q\,(\,\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1} \otimes \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*\,. \tag{4.3.49}$$

With the use of (4.2.11), (4.2.13) and (4.2.18) equation (4.3.49) can be written as

$$\mathbf{G}'_q\,(\,\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1} \otimes \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*$$

$$= \mathbf{G}'_q\,(\,\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1} \otimes \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,)\,\text{vec}\,\mathbf{Y}_i$$

$$= \mathbf{G}'_q\,\text{vec}\,(\,\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{Y}_i\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,) \tag{4.3.50}$$

where $\mathbf{Y}_i$ is given by (4.2.12).

Let

$$\mathbf{a}_i = \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,(\mathbf{y}_i - \mathbf{X}_{(3)i}\,\boldsymbol{\beta}\,)$$

$$= \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i - \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} \tag{4.3.51}$$

with

$$\mathbf{V}_i^{-1} = \boldsymbol{\Lambda}_i^{-1} - \boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}$$

as defined by (4.3.9).

Then $\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i$ can be written as

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i = \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{y}_i - \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{y}_i \tag{4.3.52}$$

where

$$\begin{aligned} \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{y}_i &= \sum_{j=1}^{n_i} \mathbf{X}'_{(3)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \\ &= \mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{X}'_{(2)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \\ &= \mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{b}_{ij} \end{aligned} \tag{4.3.53}$$

with $\mathbf{b}_{ij}$ defined by (4.3.48).

Also, (cf. (4.3.3))

$$\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{Q}_i\ . \tag{4.3.54}$$

Substitution of (4.3.53) and (4.3.54) in (4.3.52) gives

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i = [\ \mathbf{I} - \mathbf{Q}_i\,\mathbf{C}_i\ ]\ \mathbf{S}'_i \sum_{j=1}^{n_i} (\ \mathbf{b}_{ij}\ )\ . \tag{4.3.55}$$

The second term of $\mathbf{a}_i$, $\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta}$, can be written as

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} = \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} - \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\ \mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta}$$

which, using result (4.3.54), gives

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} = [\ \mathbf{I} - \mathbf{Q}_i\,\mathbf{C}_i\ ]\ \mathbf{Q}_i\,\boldsymbol{\beta}\ . \tag{4.3.56}$$

Substitution of (4.3.55) and (4.3.56) in (4.3.51) gives

$$\mathbf{a}_i = [\ \mathbf{I} - \mathbf{Q}_i\,\mathbf{C}_i\ ]\ [\ \mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{b}_{ij} - \mathbf{Q}_i\,\boldsymbol{\beta}\ ]\ . \tag{4.3.57}$$

Finally, $\mathbf{G}_q' \operatorname{vec} (\mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{X}_{(3)i} \mathbf{V}_i^{-1})$ can be written as

$$\mathbf{G}_q' \operatorname{vec} (\mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{X}_{(3)i} \mathbf{V}_i^{-1}) = \mathbf{G}_q' \operatorname{vec} (\mathbf{a}_i \mathbf{a}_i')$$

with $\mathbf{a}_i$ given by (4.3.57). □

PROPOSITION 4.3.8 (subvector $\mathbf{q}_{(2)}$)

$$\begin{aligned} \mathbf{q}_{(2)} &= \mathbf{G}_m' (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}') \mathbf{H}_{n_i^*}' \mathbf{G}_{n_i}' (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{y}_i^* \\ &= \mathbf{G}_m' \operatorname{vec} (\{ \mathbf{b}_{ij} - \mathbf{A}_{ij} \mathbf{S}_i [\boldsymbol{\beta} + \mathbf{C}_i (\mathbf{S}_i' \sum_{j=1}^{n_i} \mathbf{b}_{ij} - \mathbf{Q}_i \boldsymbol{\beta})]\} \times \\ &\qquad \{ \mathbf{b}_{ij} - \mathbf{A}_{ij} \mathbf{S}_i [\boldsymbol{\beta} + \mathbf{C}_i (\mathbf{S}_i' \sum_{j=1}^{n_i} \mathbf{b}_{ij} - \mathbf{Q}_i \boldsymbol{\beta})]\}') . \qquad (4.3.58) \end{aligned}$$

Proof

From (4.3.5), (4.3.7) and (4.2.18) it follows that

$$\begin{aligned} &\mathbf{G}_m' (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}') \mathbf{H}_{n_i^*}' \mathbf{G}_{n_i}' (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{y}_i^* \\ &= \mathbf{G}_m' \operatorname{vec} (\mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij}) . \qquad (4.3.59) \end{aligned}$$

Let

$$\begin{aligned} \mathbf{d}_i &= \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_{(3)i} \boldsymbol{\beta}) \\ &= \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{y}_i - \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \boldsymbol{\beta} . \qquad (4.3.60) \end{aligned}$$

Using (4.3.9), the first term of (4.3.60) can be written as

$$\mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{y}_i = \mathbf{Z}_{(2)ij}' \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i - \mathbf{Z}_{(2)ij}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i$$

with (cf. (4.3.48), (4.3.17) and (4.3.53))

$$\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{y}_i = \mathbf{b}_{ij}\,,$$

$$\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{A}_{ij}\,\mathbf{S}_i$$

and

$$\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{y}_i = \mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{b}_{ij}$$

so that

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i = \mathbf{b}_{ij} - \mathbf{A}_{ij}\,\mathbf{S}_i\,\mathbf{C}_i\,\mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{b}_{ij}\,. \tag{4.3.61}$$

The second term of (4.3.60) can be written as (cf. (4.3.10) and (4.3.9))

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} = \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} - \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\;\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} \tag{4.3.62}$$

with (cf. (4.3.17))

$$\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} = \mathbf{A}_{ij}\mathbf{S}_i\,\boldsymbol{\beta} \tag{4.3.63}$$

and (cf. (4.3.56))

$$\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} = \mathbf{Q}_i\,\boldsymbol{\beta}\ . \tag{4.3.64}$$

Thus,

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\boldsymbol{\beta} = \mathbf{A}_{ij}\,\mathbf{S}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{Q}_i\,]\,\boldsymbol{\beta}\,. \tag{4.3.65}$$

Substitution of (4.3.61) and (4.3.65) in (4.3.60) gives

$$\mathbf{d}_i = \mathbf{b}_{ij} - \mathbf{A}_{ij}\,\mathbf{S}_i\,\left[\,\boldsymbol{\beta} + \mathbf{C}_i\,(\,\mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{b}_{ij} - \mathbf{Q}_i\,\boldsymbol{\beta}\,)\right]. \tag{4.3.66}$$

Finally, $\mathbf{G}_m' \operatorname{vec} (\mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij})$ may be written as

$$\mathbf{G}_m' \operatorname{vec} (\mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij})$$

$$= \mathbf{G}_m' \operatorname{vec} (\mathbf{d}_i \mathbf{d}_i')$$

with $\mathbf{d}_i$ given by (4.3.66). □

PROPOSITION 4.3.9 ($\mathbf{q}_{(1)}$)

$$\begin{aligned}
\mathbf{q}_{(1)} &= (\operatorname{vec} \mathbf{I})' \mathbf{H}_{n_i^*}' \mathbf{G}_{n_i}' (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{y}_i^* \\
&= \sum_{j=1}^{n_i} \Big(\frac{1}{\Phi_{(1)}^2} \mathbf{y}_{ij}' \mathbf{y}_{ij} - \frac{2}{\Phi_{(1)}^3} \mathbf{y}_{ij}' \mathbf{X}_{(2)ij} (\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}_{(2)ij}' \Phi_{(1)}^{-1} \mathbf{X}_{(2)ij})^{-1} \mathbf{X}_{(2)ij}' \mathbf{y}_{ij} \\
&\quad + \frac{1}{\Phi_{(1)}^4} \mathbf{y}_{ij}' \mathbf{X}_{(2)ij} (\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}_{(2)ij}' \Phi_{(1)}^{-1} \mathbf{X}_{(2)ij})^{-1} \mathbf{X}_{(2)ij}' \mathbf{X}_{(2)ij} \\
&\quad \times (\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}_{(2)ij}' \Phi_{(1)}^{-1} \mathbf{X}_{(2)ij})^{-1} \mathbf{X}_{(2)ij}' \mathbf{y}_{ij} \; - (\sum_{j=1}^{n_i} \mathbf{F}_{ij})' \mathbf{f}_i - \mathbf{f}_i' (\sum_{j=1}^{n_i} \mathbf{F}_{ij}) \\
&\quad + \mathbf{f}_i' (\sum_{j=1}^{n_i} \mathbf{E}_{ij}) \mathbf{f}_i
\end{aligned} \qquad (4.3.67)$$

where

$$\mathbf{f}_i = \mathbf{S}_i \boldsymbol{\beta} + \mathbf{S}_i \mathbf{C}_i \mathbf{S}_i' \{ \sum_{j=1}^{n_i} \mathbf{b}_{ij} - (\sum_{j=1}^{n_i} \mathbf{A}_{ij}) \mathbf{S}_i \boldsymbol{\beta} \} , \qquad (4.3.68)$$

$$\mathbf{F}_{ij} = \mathbf{X}_{(2)ij}' \mathbf{\Lambda}_{ij}^{-1} \mathbf{\Lambda}_{ij}^{-1} \mathbf{y}_{ij} \qquad (4.3.69)$$

and $\mathbf{E}_{ij}$ is given by (4.3.27). (For further simplification of $\mathbf{F}_{ij}$ see Appendix 4.1.)

Proof

From (4.3.5) and (4.2.11)

$$(\operatorname{vec} \mathbf{I})' \mathbf{H}_{n_i^*}' \mathbf{G}_{n_i}' (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{y}_i^*$$

$$= (\text{vec } \mathbf{I})' \, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \text{ vec } \mathbf{Y}$$

$$= \text{tr} \, (\mathbf{V}_i^{-1} \mathbf{Y} \mathbf{V}_i^{-1}) \, . \qquad (4.3.70)$$

Let (cf. (4.3.9))

$$\mathbf{g}_i = \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_{(3)i} \boldsymbol{\beta})$$

$$= \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i - \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}'_{(3)i} \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i - \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \boldsymbol{\beta} +$$

$$\boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}'_{(3)i} \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \boldsymbol{\beta} \, . \qquad (4.3.71)$$

Using (4.3.10) and (4.3.17),

$$\mathbf{g}_i = \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i - \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \, \mathbf{S}'_i \sum_{j=1}^{n_i} \mathbf{b}_{ij} - \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \boldsymbol{\beta} + \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{Q}_i \boldsymbol{\beta}$$

$$= \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i - \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(2)i} \mathbf{f}_i$$

where $\mathbf{f}_i$ is given by (4.3.68).

Thus,

$$\mathbf{g}'_i \mathbf{g}_i = \mathbf{y}'_i \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i + \mathbf{f}'_i \, \mathbf{X}'_{(2)i} \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(2)i} \, \mathbf{f}_i -$$

$$\mathbf{y}'_i \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(2)i} \mathbf{f}_i - \mathbf{f}'_i \, \mathbf{X}'_{(2)ij} \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Lambda}_i^{-1} \mathbf{y}_i \, . \qquad (4.3.72)$$

But (cf. (4.3.27))

$$\mathbf{E}_{ij} = \mathbf{X}'_{(2)ij} \boldsymbol{\Lambda}_{ij}^{-1} \boldsymbol{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)ij}$$

and (cf. (4.3.69))

$$\mathbf{F}_{ij} = \mathbf{X}'_{(2)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij}$$

so that

$$\mathbf{g}'_i\,\mathbf{g}_i = \mathbf{y}'_i\,\mathbf{\Lambda}_i^{-1}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{y}_i + \mathbf{f}'_i\,(\sum_{j=1}^{n_i}\mathbf{E}_{ij})\,\mathbf{f}_i - (\sum_{j=1}^{n_i}\mathbf{F}_{ij})'\,\mathbf{f}_i - \mathbf{f}'_i\,(\sum_{j=1}^{n_i}\mathbf{F}_{ij}) \tag{4.3.73}$$

with

$$\mathbf{y}_i{}'\,\mathbf{\Lambda}_i^{-1}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{y}_i = \sum_{j=1}^{n_i}\mathbf{y}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij}\,.$$

Using (4.3.11), it follows that

$$\mathbf{\Lambda}_{ij}^{-1} = \left(\frac{1}{\Phi_{(1)}}\mathbf{I}_{n_{ij}} - \frac{1}{\Phi^2_{(1)}}\,\mathbf{X}_{(2)ij}\,(\,\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\right)^{-1}$$

so that

$$\begin{aligned}\sum_{j=1}^{n_i}\mathbf{y}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} = \sum_{j=1}^{n_i}\{\,&\frac{1}{\Phi_{(1)}}\mathbf{y}'_{ij}\,\mathbf{y}_{ij} - \frac{2}{\Phi^3_{(1)}}\,\mathbf{y}'_{ij}\,\mathbf{X}_{(2)ij}\,\times\\ &(\,\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij} + \\ &\frac{1}{\Phi^4_{(1)}}\,\mathbf{y}'_{ij}\,\mathbf{X}_{(2)ij}\,(\,\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\times\\ &\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,(\,\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\times\\ &\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij}\,\}\,.\end{aligned} \tag{4.3.74}$$

Substitution of (4.3.73) and (4.3.74) in (4.3.72) gives

$$\begin{aligned}\mathbf{g}'_i\,\mathbf{g}_i = \sum_{j=1}^{n_i}\{\,&\frac{1}{\Phi_{(1)}}\mathbf{y}'_{ij}\,\mathbf{y}_{ij} - \frac{2}{\Phi^3_{(1)}}\,\mathbf{y}'_{ij}\,\mathbf{X}_{(2)ij}\,\times\\ &(\,\mathbf{\Phi}_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij} +\end{aligned}$$

$$\frac{1}{\Phi^4_{(1)}} \mathbf{y}'_{ij} \mathbf{X}_{(2)ij} (\Phi^{-1}_{(2)} + \mathbf{X}'_{(2)ij} \Phi^{-1}_{(1)} \mathbf{X}_{(2)ij})^{-1} \times$$

$$\mathbf{X}'_{(2)ij} \mathbf{X}_{(2)ij} (\Phi^{-1}_{(2)} + \mathbf{X}'_{(2)ij} \Phi^{-1}_{(1)} \mathbf{X}_{(2)ij})^{-1} \mathbf{X}'_{(2)ij} \mathbf{y}_i \} +$$

$$\mathbf{f}'_i (\sum_{j=1}^{n_i} \mathbf{E}_{ij}) \mathbf{f}_i - (\sum_{j=1}^{n_i} \mathbf{F}_{ij})' \mathbf{f}_i - \mathbf{f}'_i (\sum_{j=1}^{n_i} \mathbf{F}_{ij}) .$$

Note that

$$\text{tr} (\mathbf{g}'_i \mathbf{g}_i) = \text{tr} (\mathbf{g}_i \mathbf{g}'_i).$$

This concludes the proof. □

THEOREM 4.3.2

Let

$$\mathbf{X}^*_i = \mathbf{H}_{n^*_i} [(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \mathbf{G}_q \quad (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \mathbf{G}_m \quad \text{vec } \mathbf{I}] ,$$

$$2 \mathbf{W}_i = \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i}$$

and

$$\mathbf{y}^*_i = \text{vecs } \mathbf{Y}_i ,$$

with

$$\mathbf{Y}_i = (\mathbf{y}_i - \mathbf{X}_{(3)i} \boldsymbol{\beta}) (\mathbf{y}_i - \mathbf{X}_{(3)i} \boldsymbol{\beta})' ,$$

as defined by (4.2.15), (4.2.16), (4.2.13) and (4.2.12) respectively.

The vector $\mathbf{X}^{*\prime}_i \mathbf{W}_i^{-1} \mathbf{y}^*_i$ can be written as

$$2 \mathbf{X}^{*\prime}_i \mathbf{W}_i^{-1} \mathbf{y}^*_i = \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \\ \mathbf{q}_{(1)} \end{bmatrix}$$

where the respective terms are

$$q_{(3)} = G'_q \text{ vec } \{[\, I - Q_i\, C_i\,]\,[\, S_i\,(\sum_{j=1}^{n_i} b_{ij}\,) - Q_i\,\beta\,]\} \times$$

$$\{[\, I - Q_i\, C_i\,]\,[\, S_i\,(\sum_{j=1}^{n_i} b_{ij}\,) - Q_i\,\beta\,]\}'$$

$$q_{(2)} = G'_m \text{ vec } (\{\, b_{ij} - A_{ij}\, S_i\,\,[\,\beta + C_i\,(\, S'_i \sum_{j=1}^{n_i} b_{ij} - Q_i\,\beta\,)]\} \times$$

$$\{\, b_{ij} - A_{ij}\, S_i\,\,[\,\beta + C_i\,(\, S'_i \sum_{j=1}^{n_i} b_{ij} - Q_i\,\beta\,)]\}')$$

$$q_{(1)} = \sum_{j=1}^{n_i} (\, \frac{1}{\Phi^2_{(1)}}\, y'_{ij}\, y_{ij} - \frac{2}{\Phi^3_{(1)}}\, y'_{ij}\, X_{(2)ij}\,(\, \Phi^{-1}_{(2)} + X'_{(2)ij}\, \Phi^{-1}_{(1)}\, X_{(2)ij}\,)^{-1}\, X'_{(2)ij}\, y_{ij}$$

$$+ \frac{1}{\Phi^4_{(1)}}\, y'_{ij}\, X_{(2)ij}\,(\, \Phi^{-1}_{(2)} + X'_{(2)ij}\, \Phi^{-1}_{(1)}\, X_{(2)ij}\,)^{-1}\, X'_{(2)ij}\, X_{(2)ij}$$

$$\times\,(\, \Phi^{-1}_{(2)} + X'_{(2)ij}\, \Phi^{-1}_{(1)}\, X_{(2)ij}\,)^{-1}\, X'_{(2)ij}\, y_{ij} - (\sum_{j=1}^{n_i} F_{ij}\,)'\, f_i - f'_i\,(\sum_{j=1}^{n_i} F_{ij}\,)$$

$$+ f'_i\,(\sum_{j=1}^{n_i} E_{ij}\,)\, f_i$$

with Q_i, C_i, b_{ij}, A_{ij}, f_i, F_{ij} and E_{ij} given by (4.3.3), (4.3.10), (4.3.48), (4.3.4), (4.3.68), (4.3.69) and (4.3.27) respectively.

Proof

The proof follows directly from Propositions 4.3.7 to 4.3.9. □

Simplification of the components of (4.2.14) is not sufficient to ensure computational efficiency. The terms $X_{(f)i}\ V_i^{-1}\ V_{(f)i}$ and $X_{(f)i}\ V_i^{-1}\ y_i$ (cf. (4.2.8)) needed for the calculation of $\hat{\beta}$ must also be simplified. For a detailed discussion of the required simplification of these terms see Section 5.2.3.

4.4 STATISTICAL INFERENCE

In Section 4.4.1 results are given which are required for the calculation of the standard errors of the estimated parameters. Hypotheses of the form $c_1\beta_1 + c_2\beta_2 + \ldots c_q\beta_q = k$

about the elements of the fixed parameter vector β are be considered in Section 4.4.2. Section 4.4.3 deals with the calculation of residuals, while likelihood ratio tests are discussed in Section 4.4.4. This section is based closely on that in du Toit (1993).

4.4.1 STANDARD ERRORS

Consider the general 3-level model

$$\mathbf{y}_i = \mathbf{X}_{(f)i}\,\beta + \mathbf{Z}_i^*\,\mathbf{e}_i^* \qquad i = 1, 2, \ldots, N \tag{4.4.1}$$

where

$$\mathbf{X}_{(f)i} = \mathbf{X}_{(2)i}\,\mathbf{S}_i = \mathbf{X}_{(3)i}\,,$$

$$\mathbf{Z}_i^* = [\;\mathbf{X}_{(3)i}\;\;\mathbf{X}_{(2)i}\;\;\mathbf{X}_{(1)i}\;]$$

with

$$\mathbf{X}_{(2)i} = \begin{bmatrix} \mathbf{X}_{(2)i1} \\ \mathbf{X}_{(2)i2} \\ \vdots \\ \vdots \\ X_{(2)in_i} \end{bmatrix}, \quad \mathbf{X}_{(1)i} = \begin{bmatrix} \mathbf{x}'_{(1)i11} \\ \mathbf{x}'_{(1)i12} \\ \vdots \\ \vdots \\ \mathbf{x}'_{(1)in_in_{ij}} \end{bmatrix}$$

and

$$\mathbf{e}_i^* = \begin{bmatrix} \mathbf{v}_i \\ \mathbf{u}_i \\ \mathbf{e}_i \end{bmatrix}.$$

The set of models described by (4.4.1) can be combined into a single model by means of column stacking. Thus,

$$\mathbf{y} = \mathbf{X}_{(f)}\,\beta + \mathbf{Z}^*\,\mathbf{e}^* \tag{4.4.2}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \vdots \\ \mathbf{y}_N \end{bmatrix} , \quad \mathbf{X}_{(f)} = \begin{bmatrix} \mathbf{X}_{(f)1} \\ \mathbf{X}_{(f)2} \\ \vdots \\ \vdots \\ \mathbf{X}_{(f)N} \end{bmatrix}$$

and

$$\mathbf{Z}^* = \text{Diag} \; (\; \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N) \; .$$

From (4.4.2) it follows that

$$\hat{\beta} \;=\; [\mathbf{X}'_{(f)} \, \Sigma^{-1} \, \mathbf{X}_{(f)} \,]^{-1} \, \mathbf{X}'_{(f)} \, \Sigma^{-1} \, \mathbf{y} \; , \tag{4.4.3}$$

where

$$\Sigma = \text{Diag} \; (\; \Sigma_1, \Sigma_2, \ldots, \Sigma_N)$$

and

$$\Sigma_i = \text{Cov}(\mathbf{y}_i \, , \, \mathbf{y}'_i) = \mathbf{Z}^*_i \; \text{Cov}(\mathbf{e}^*_i, \, \mathbf{e}^*_i \,) \; \mathbf{Z}^{*\prime}_i ,$$

and where

$$\text{Cov}(\mathbf{e}^*_i, \, \mathbf{e}^{*\prime}_i) = \begin{bmatrix} \Phi_{(3)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi_{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Phi_{(1)} \end{bmatrix} .$$

From (4.2.8) it follows that the covariance matrix of $\hat{\beta}$ is given by

$$\text{Cov}(\hat{\beta} \, , \, \hat{\beta}') \;=\; [\mathbf{X}^{*\prime}_{(f)} \, \Sigma^{-1} \, \mathbf{X}^*_{(f)} \,]^{-1} \, . \tag{4.4.4}$$

In practice, Σ_i is unknown and is replaced by maximum likelihood estimator $\Sigma_i(\hat{\tau}) = \mathbf{V}_i$ (cf. Remark 4.2.1). Hence, a consistent estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by

$$\text{Cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}') = [\mathbf{X}^{*\prime}_{(f)} \mathbf{V}^{-1} \mathbf{X}^{*}_{(f)}]^{-1} . \tag{4.4.5}$$

Similarly, it can be shown that a consistent estimator of the covariance matrix of $\hat{\tau}$ (cf. (4.2.20)) is given by

$$\text{Cov}(\hat{\tau}, \hat{\tau}') = [\sum_{i=1}^{N} \mathbf{X}^{*\prime}_i \mathbf{W}_i^{-1} \mathbf{X}^{*}_i]^{-1} .$$

The diagonal elements of the covariance matrix (4.4.4) may be used to obtain large-sample estimates of the standard errors for the fixed parameter estimates. For large samples $\hat{\boldsymbol{\beta}}$ and $\hat{\tau}$ have approximate multivariate normal distributions. See for example Malinvaud (1970), for general results on the distribution of least squares estimators.

4.4.2 CONTRASTS

The construction of contrasts or linear functions of the parameters is a useful statistical analysis tool and enables the researcher to perform hypothesis testing concerning the equality of subsets of parameters. In this section (see also du Toit, 1993), a summary of the results required for contrast testing is given.

A $p \times q$ contrast matrix $\mathbf{C}$, where p denotes the number of contrasts, can be used to formulate a complex hypothesis about several elements of $\boldsymbol{\beta}$. The hypothesis is written in the form $\mathbf{C}\,\boldsymbol{\beta} = \mathbf{k}$, where $\mathbf{k}$ is a known $p \dot{\times} 1$ vector.

Consider as example the case where $q = 3$ and the following hypothesis is to be tested:

$$\beta_1 - \beta_2 = 0$$

$$\beta_3 - \beta_2 = 0 .$$

The null hypothesis can be formulated as

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} .$$

For large samples it can be shown that the vector variate $\mathbf{C}\,\widehat{\boldsymbol{\beta}}$ will be approximately distributed as $N(\mathbf{C}\,\boldsymbol{\beta}\ ,\ \mathbf{C}\,(\mathbf{X}'_{(f)}\,\widehat{\mathbf{V}}^{-1}\,\mathbf{X}_{(f)})^{-1}\mathbf{C}')$.

Therefore, if H_0 is true,

$$\mathbf{M} = (\ \mathbf{C}\,\widehat{\boldsymbol{\beta}} - \mathbf{k}\ )'\ \{\mathbf{C}\,(\mathbf{X}'_{(f)}\,\widehat{\mathbf{V}}^{-1}\,\mathbf{X}_{(f)})^{-1}\mathbf{C}'\}\ (\ \mathbf{C}\,\widehat{\boldsymbol{\beta}} - \mathbf{k}\ ) \qquad (4.4.6)$$

has an approximate χ^2-square distribution with p degrees of freedom.

Let $\mathbf{c}'$ denote the i-th row of $\mathbf{C}$ and $\chi^2_{q,\alpha}$ the critical value of the χ^2- distribution with q degrees of freedom. A set of 100 $(1 - \alpha)$ % simultaneous confidence intervals for the p elements of $\mathbf{C}\,\boldsymbol{\beta}$ is given by the p intervals

$$\mathbf{c}_i'\,\widehat{\boldsymbol{\beta}}\ \pm\ \{\mathbf{c}_i'\,(\mathbf{X}'_{(f)}\,\widehat{\mathbf{V}}^{-1}\,\mathbf{X}_{(f)})^{-1}\,\mathbf{c}_i\,\chi^2_{q,\alpha}\,\}^{0.5},\ p < q\ . \qquad (4.4.7)$$

The null hypothesis $H_0 : \widehat{\beta}_j,\ j = 1, 2, \ldots, q = 0$ is tested by using the test statistic

$$z = \frac{\widehat{\beta}_k}{S.E.(\widehat{\beta}_k)}$$

which, for large samples, has an approximate N(0,1) distribution if H_0 is true. (See Section 4.5 for a practical illustration.)

4.4.3 RESIDUALS

The residuals $\widehat{\mathbf{v}}_i$, $\widehat{\mathbf{u}}_i$ and $\widehat{\mathbf{e}}_i$ (cf. Section 4.4.1) may be estimated as follows:

Let

$$\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{X}_{(1)i}\,\boldsymbol{\beta}\ ,$$

then, from (4.4.1),

$$\widetilde{\mathbf{y}}_i = \mathbf{X}_{(3)i}\,\mathbf{v}_i + \mathbf{X}_{(2)i}\,\mathbf{u}_i + \mathbf{I}_i\,\mathbf{e}_i\ . \tag{4.4.8}$$

Under the assumption of multivariate normality it follows that $\widetilde{\mathbf{y}}_i \sim \mathrm{N}(\mathbf{0}\ ,\ \Sigma)$, $\mathbf{v}_i \sim \mathrm{N}(\mathbf{0}\ ,\ \boldsymbol{\Phi}_{(3)})$ and hence the joint distribution of $\begin{bmatrix} \widetilde{\mathbf{y}}_i \\ \mathbf{v}_i \end{bmatrix}$ is

$$\begin{pmatrix} \widetilde{\mathbf{y}}_i \\ \mathbf{v}_i \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},\ \begin{matrix} \Sigma & \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)} \\ \boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} & \boldsymbol{\Phi}_{(2)} \end{matrix}\right).$$

From standard results on conditional distributions (see for example Morrison, 1991) it follows that

$$\begin{aligned} \mathrm{E}(\mathbf{v}_i \mid \widetilde{\mathbf{y}}_i) &= \mathbf{0} + (\ \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\ )'\ \Sigma_i^{-1}\ (\ \widetilde{\mathbf{y}}_i\ -\ \mathbf{0}\ ) \\ &= (\ \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\ )'\ \Sigma_i^{-1}\ \ \widetilde{\mathbf{y}}_i \\ &= \boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i}\,\Sigma_i^{-1}\ \ \widetilde{\mathbf{y}}_i\ . \end{aligned}$$

Thus, the empirical Bayes estimate (see Chapter 3) of $\mathbf{V}_i$ is

$$\hat{\mathbf{v}}_i = \ \hat{\boldsymbol{\Phi}}_{(3)}\,\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,(\ \mathbf{y}_i\ -\ \mathbf{X}_{(1)i}\,\hat{\boldsymbol{\beta}}\ )\ . \tag{4.4.9}$$

Similarly,

$$\hat{\mathbf{u}}_i = \hat{\boldsymbol{\Phi}}_{(2)}\,\mathbf{X}'_{(2)i}\,\mathbf{V}_i^{-1}\,(\ \mathbf{y}_i\ -\ \mathbf{X}_{(1)i}\,\hat{\boldsymbol{\beta}}\ ) \tag{4.4.10}$$

and

$$\widehat{\mathbf{e}}_i = \widehat{\Phi}_{(1)} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_{(f)i} \widehat{\beta}) \quad . \tag{4.4.11}$$

4.4.4 LIKELIHOOD RATIO TESTS

Finally, likelihood ratio tests are considered. Tests of a null hypothesis against a restricted alternative hypothesis can be constructed, provided that two conditions are met. Firstly, the models under H_0 and H_1 should be estimable and secondly, the parameter space Ω_0 for H_0 must be a subset of the parameter space Ω of H_1 .

Use is made of the likelihood ratio test statistic

$$\lambda = \frac{L_0(\widehat{\beta}, \widehat{\tau})}{L_1(\widehat{\beta}, \widehat{\tau})} \, , \tag{4.4.12}$$

where L_0 and L_1 respectively denote the likelihood functions under H_0 and H_1 . For N large (see for example Anderson, 1984), $-2 \, ln \, \lambda$ has an approximate $\chi^2(v)$ distribution where the number of degrees of freedom v is the difference in the number of parameters estimated under H_1 and the number of parameters estimated under H_0 .

Example 4.4.1

Consider the null hypothesis

$$H_0 : \text{Cov}(\mathbf{y}_{ij} , \mathbf{y}'_{ij}) = \mathbf{X}_{(2)ij} \, \mathbf{\Phi}_{(2)} \, \mathbf{X}'_{(2)ij} + \Phi_{(1)} \, \mathbf{I}_{n_{ij}}$$

as opposed to the alternative hypothesis

$$H_1 : \text{Cov}(\mathbf{y}_{ij} , \mathbf{y}'_{ij}) \; = \; \mathbf{X}_{(3)ij} \, \mathbf{\Phi}_{(3)} \, \mathbf{X}'_{(3)ij} + \mathbf{X}_{(2)ij} \, \mathbf{\Phi}_{(2)} \, \mathbf{X}'_{(2)ij} + \Phi_{(1)} \, \mathbf{I}_{n_{ij}} \, .$$

Let

$$\lambda = \frac{L_0(\widehat{\mathbf{\Phi}}_{(1)} , \widehat{\mathbf{\Phi}}_{(2)})}{L_1(\widehat{\mathbf{\Phi}}_{(1)} , \widehat{\mathbf{\Phi}}_{(2)} , \widehat{\mathbf{\Phi}}_{(3)})} \quad .$$

For N large, $-2 \ ln \ \lambda = -2 \ (\ ln \ L_0 - \ ln \ L_1)$ has an approximate $\chi^2(v)$ distribution with the number of degrees of freedom, $v = \frac{1}{2} \ q(q+1)$, which is the number of non-duplicated elements of $\mathbf{\Phi}_{(3)}$. Note that $ln \ L$ is the log-likelihood function

$$ln \ L = - \tfrac{1}{2} \sum_{i=1}^{N} \{ \ n_i \ ln \ (2\pi) + ln \mid \Sigma_i \mid +$$

$$\text{tr} \ \Sigma_i^{-1} \ (\ \mathbf{y}_i - \mathbf{X}_{(f)i} \beta \ ) \ (\ \mathbf{y}_i - \mathbf{X}_{(f)i} \beta \ )' \}$$

with $\mathbf{X}_i^* \beta$ and Σ_i respectively the expected value and covariance of $\mathbf{y}_i$.

4.5 PRACTICAL APPLICATIONS

Example 4.5.1

As part of a test battery developed by the HSRC, Grade 1 to Standard 1 pupils are tested with regard to their ability to understand and carry out verbal instructions. In this example the test results of 5022 pupils from 139 schools belonging to 13 education departments are considered. The education departments are the level-3 units, the schools are the level-2 units and the pupils tested are the level-1 units.

The maximum possible score for the test, which is used as the response variable y, is 35. Four variables are included as predictors, namely an intercept, the gender, the reading ability and the mathematical ability of a pupil. The gender variable was coded '1' and '-1', where '1' denotes males and '-1' females. The reading and mathematical abilities of each pupil are measured on a scale from 1 to 5, where '1' indicates the minimum score and '5' the maximum score. The ages of pupils tested ranged from 6 to 10 years.

The following linear model (cf. (4.2.1)) is fitted to the data:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij} \, \mathbf{b}_{ij} + \mathbf{e}_{ij} \qquad (4.5.1)$$

where

$$\mathbf{b}_{ij} = \mathbf{c}_i + \mathbf{u}_{ij} \qquad (4.5.2)$$

and

$$\mathbf{c}_i = \boldsymbol{\beta} + \mathbf{v}_i \tag{4.5.3}$$

A typical element of $\mathbf{y}_{ij}$ is y_{ijk} which denotes the test score with regard to understanding of verbal instructions of pupil k from school j from education department i. Likewise. a typical element of the $n_{ij} \times 1$ vector $\mathbf{X}_{ij}\, \mathbf{b}_{ij}$ is $\mathbf{x}'_{ijk}\, \mathbf{b}_{ij}$ where

$$\mathbf{x}'_{ijk} = (1\ ,\ \text{gender}_{ijk}\ ,\ \text{read}_{ijk}\ ,\ \text{maths}_{ijk}\ )$$

and

$$\mathbf{b}_{ij} \;=\; \begin{bmatrix} \text{intercept}_{ij} \\ \text{regression coefficient for gender}_{ij} \\ \text{regression coefficient for reading}_{ij} \\ \text{regression coefficient for maths}_{ij} \end{bmatrix}.$$

Note that $\mathbf{b}_{ij}$ is a vector of stochastic coefficients for school j and education department i. The vector $\boldsymbol{c}_i$ is a vector of stochastic coefficients for education department i. Note that for the model above, the matrix $\mathbf{S}_i$ (cf. (4.2.3)) is taken as the identity matrix, which implies that

$$\text{Cov}(\mathbf{b}_{ij}, \mathbf{b}'_{ij}) = \boldsymbol{\Phi}_{(3)} + \boldsymbol{\Phi}_{(2)}\,. \tag{4.5.4}$$

From the distributional assumptions given in Section 3.2, it also follows that

$$\text{Cov}(\mathbf{c}_i, \mathbf{c}'_i) = \boldsymbol{\Phi}_{(3)} \tag{4.5.5}$$

and

$$\text{Cov}(\mathbf{e}_i, \mathbf{e}'_i) = \boldsymbol{\Phi}_{(1)}\, \mathbf{I}_{n_i^*} \tag{4.5.6}$$

where $n_i^* \;=\; \sum_{j=1}^{n_i} n_{ij}$. The computer program IGLS was written in FORTRAN to implement the theoretical results given in Sections 4.2 to 4.4. Part of the computer output of IGLS is given below:

(i) Fixed part of the model:

PARAMETER	$\hat{\beta}$	STD.ERR	Z-VALUE	PR>|Z|
INTERCEPT	19.42720	0.59354	32.73089	0.00000
GENDER	0.23911	0.07861	3.04161	0.00235
READ	0.61737	0.08883	6.94964	0.00000
MATHS	0.66975	0.14264	4.69547	0.00000

(ii) Random part of the model:

Level-3:

PARAMETER	$\hat{\tau}$	STD.ERR	Z-VALUE	PR>|Z|
INTER./INTER.	3.03389	1.75459	1.72911	0.08379
INTERCEPT/GENDER	– 0.24654	0.17968	– 1.37207	0.17004
GENDER/GENDER	0.03097	0.02945	1.05153	0.29301
INTERCEPT/READ	– 0.14142	0.18717	0.75558	0.44990
GENDER/READ	0.01084	0.02366	0.45844	0.64663
READ/READ	0.00672	0.03409	0.19711	0.84375
INTERCEPT/MATHS	0.31953	0.29822	1.07143	0.28397
GENDER/MATHS	– 0.02669	0.03961	– 0.67389	0.50038
READ/MATHS	– 0.01361	0.04520	– 0.30119	0.76327
MATHS/MATHS	0.16761	0.09997	1.67661	0.09362

Level-2:

INTER./INTER.	5.94295	1.43615	4.13810	0.00004
INTERCEPT/GENDER	0.16771	0.19482	0.86084	0.38933
GENDER/GENDER	0.00496	0.05217	0.09510	0.92423
INTERCEPT/READ	0.23658	0.29521	0.80138	0.42291
GENDER/READ	0.04356	0.05616	0.77558	0.43800
READ/READ	11.72362	0.11758	99.71007	0.00000
INTERCEPT/MATHS	– 0.75881	0.32808	– 2.31286	0.02073
GENDER/MATHS	– 0.05195	0.05538	– 0.93798	0.34826
READ/MATHS	– 0.18644	0.09636	– 1.93484	0.05301
MATHS/MATHS	15.90603	0.11742	135.46414	0.00000

Level-1:

INTER./INTER.	14.24395	0.29941	47.57350	0.00000

CORRELATION MATRIX: Level-3

	INTERCEPT	GENDER	READ	MATHS
INTERCEPT	1.0000			
GENDER	− 0.8043	1.0000		
READ	− 0.9905	0.7518	1.0000	
MATHS	0.4481	− 0.3705	− 0.4056	1.0000

CORRELATION MATRIX: Level-2

	INTERCEPT	GENDER	READ	MATHS
INTERCEPT	1.0000			
GENDER	0.9767	1.0000		
READ	0.0283	0.1806	1.0000	
MATHS	− 0.0780	− 0.1849	− 0.0137	1.0000

(iii) **Convergence details:**

Convergence (IGLS-algorithm in 10 iterations)

From Part (i) of the output it can be seen that the estimate of the intercept is 19.42720, which may be interpreted as the expected value of y, assuming no gender, reading ability or mathematical ability effect. The gender score of 0.23911 indicates a total gender effect of 0.47822. In the case of boys (coded '1') this leads to a positive effect of 0.23911 and for girls (coded '−1') a negative effect of −0.23911, showing that boys generally obtained a slightly higher test result than girls. All the coefficients are highly significant.

The second part of the output contains information on the random parameters and the correlation matrices on level-3 and level-2 of the model. The covariances of the coefficients on level-3 are the elements of $\mathbf{\Phi}_{(3)}$. The largest variation on level-3, that is over education departments, is for the intercept. This variation is significant at a 10 % level of significance.

From the level-3 correlation matrix it can be seen that the intercept is highly correlated with both gender and reading ability. On level-2 there is only one large correlation coefficient, namely the correlation between the intercept and gender.

Approximate 95 % tolerance intervals for the elements of $\mathbf{c}_i$ (cf. (3.5.7) and (3.5.8)) can be constructed using the standard deviations of the intercept, gender, reading ability and mathematical ability coefficients as obtained from $\hat{\boldsymbol{\Phi}}_{(3)}$. The approximate 95 % tolerance interval for the intercept is given by

$$\hat{\beta}_1 \pm \sqrt{\hat{\Phi}_{(3)_{11}}},$$

that is, (17.6854;21.1690).

The approximate 95 % tolerance intervals for the gender and for reading and mathematical abilities are given by (0.0631;0.2446), (0.5354;0.6993) and (0.2603;1.0792) respectively.

On level-2, the approximate 95 % tolerance interval for the intercept can be written as (cf. (3.5.9))

$$\hat{\beta}_1 \pm \sqrt{\hat{\Phi}_{(3)_{11}} + \hat{\Phi}_{(2)_{11}}}$$

which gives (16.4311;22.4233). Similar intervals for the coefficients for gender and for reading and mathematical abilities are given by (0.0496;0.4288), (– 2.8076;4.0424) and (– 3.3394;4.6790) respectively. Note that the approximate 95 % tolerance intervals for the coefficients are wider on level-2 than on level-3.

Example 4.5.2

The testing of contrasts is an important component of statistical inference. In Section 4.4.2 the theoretical aspects of contrast testing in a multilevel context were discussed.

To illustrate these concepts, the data set described in Example 4.5.1 is used in this example. As in Example 4.5.1, the education departments, schools and pupils are the level-3, level-2 and level-1 units respectively.

The same response variable (ability to interpret verbal instructions) is used, but 5 dummy variables, denoting 5 socio-economic categories, and the reading and mathematical abilities of a pupil are used as predictors. The dummy variables are coded as follows:

	SOCIO-1	SOCIO-2	SOCIO-3	SOCIO-4	SOCIO-5
Socio-economic category 1:	1	0	0	0	0
Socio-economic category 2:	0	1	0	0	0
Socio-economic category 3:	0	0	1	0	0
Socio-economic category 4:	0	0	0	1	0
Socio-economic category 5:	0	0	0	0	1

Note that an intercept term is not included in the model to ensure that the design matrix is of full rank.

Part of the computer output obtained by using the FORTRAN program IGLS is given below:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	STD.ERR	Z-VALUE	PR>|Z|
SOCIO-1	20.54575	0.70565	29.11608	0.00000
SOCIO-2	20.57563	0.59140	34.79148	0.00000
SOCIO-3	19.42454	0.48797	39.80650	0.00000
SOCIO-4	19.59274	0.57014	34.36473	0.00000
SOCIO-5	19.29114	0.64313	29.99589	0.00000
READ	0.55773	0.20188	2.76273	0.00573
MATHS	0.64862	0.27405	2.36678	0.01794

(ii) **Random part of the model:**

ESTIMATE OF $\hat{\Phi}_{(3)}$

	SOCIO-1	SOCIO-2	SOCIO-3	SOCIO-4	SOCIO-5
SOCIO-1	5.10055				
SOCIO-2	3.41276	2.41390			
SOCIO-3	3.03821	1.76983	3.01754		
SOCIO-4	3.03943	1.76644	1.69916	4.20558	
SOCIO-5	3.94914	2.49372	1.94700	2.04880	8.26439
READ	0.00822	0.01836	0.00891	-0.08066	-0.11406
MATHS	0.40510	0.16390	0.20618	0.24966	0.28335

	READ	MATHS
READ	0.01009	
MATHS	-0.01527	2.16769

ESTIMATE OF $\hat{\Phi}_{(2)}$

	SOCIO-1	SOCIO-2	SOCIO-3	SOCIO-4	SOCIO-5
SOCIO-1	6.98216				
SOCIO-2	6.54765	7.09487			
SOCIO-3	4.79667	5.61434	5.49784		
SOCIO-4	2.54706	4.91440	5.15002	11.30542	
SOCIO-5	1.83658	4.30411	4.46511	4.38530	16.14331
READ	-0.09995	0.06470	0.26399	0.40737	0.06119
MATHS	-0.39335	-0.73882	-0.70333	-0.80759	-0.51824

	READ	MATHS
READ	0.11960	
MATHS	-0.15285	0.40827

ESTIMATE OF $\hat{\Phi}_{(1)}$

13.9908

(iii) **Contrast testing:**

Contrast tested	χ^2-value
SOCIO-1 − SOCIO-2 = 0	0.016
SOCIO-1 − SOCIO-3 = 0	11.550
SOCIO-1 − SOCIO-4 = 0	26.679
SOCIO-1 − SOCIO-5 = 0	46.522
SOCIO-2 − SOCIO-3 = 0	11.977
SOCIO-2 − SOCIO-4 = 0	13.075
SOCIO-2 − SOCIO-5 = 0	21.990
SOCIO-3 − SOCIO-4 = 0	0.195
SOCIO-3 − SOCIO-5 = 0	0.131
SOCIO-4 − SOCIO-5 = 0	4.618

From Part (i) of the output it follows that all the elements of $\widehat{\boldsymbol{\beta}}$ are significant. The first five diagonal elements of $\widehat{\boldsymbol{\Phi}}_{(3)}$ and $\widehat{\boldsymbol{\Phi}}_{(2)}$ exhibit a tendency to increase with diminishing socio-economic status, where SOCIO-5 denotes the lowest socio-economic category. This implies that a larger variation in the response scores can be expected as socio-economic status declines.

All possible pairs of contrasts with regard to socio-economic status were tested. Each contrast tested can be written in the form (cf. Section 4.4.2)

$$\mathbf{c}' \, \boldsymbol{\beta} = 0.$$

For example, to test the contrast SOCIO-2 − SOCIO-5 = 0, $\mathbf{c}' = (0\ 1\ 0\ 0\ -1\ 0\ 0)$. To ensure an overall level of significance of 5 %, the χ^2-values are compared with $\chi^2_{0.005}$ (degrees of freedom = 1). From Part (iii) of the output it can be seen that the largest χ^2-value occurs for the contrast SOCIO-1 − SOCIO-5 = 0. It can also be concluded that, although there is no significant difference between the SOCIO-1 and SOCIO-2 effects, both these effects differ significantly from the remaining effects.

4.6 SUMMARY

The estimation of the unknown parameters in multilevel models using Iterative Generalized Least Squares was discussed in this chapter. Although the mathematical equations on which this procedure was based appear to be straightforward, simplification of these equations was necessary to ensure that the optimization algorithm is computationally efficient.

The equations implemented in the FORTRAN program IGLS are based on the sums of squares and cross products of the data and are of the form $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{y}$ and $\mathbf{y}'\mathbf{y}$ which results in a substantial reduction in computation time, especially when the number of columns (predictors) of the design matrix is substantially smaller than the number of rows (observations).

Some of the shortcomings of the EM algorithm used in Chapter 3 were addressed. Two practical applications were given to illustrate the theoretical principles involved.

The level-3 model discussed in this chapter was, however, still limited as complex level-1 structures could not be accommodated. It is also not possible to use different sets of predictors in the fixed and stochastic parts of the model. These issues will be addressed in the next chapter.

APPENDIX 4.1

The terms $\mathbf{b}_{ij}$, $\mathbf{E}_{ij}$ and $\mathbf{F}_{ij}$ can be simplified as follows:

(i)
$$\mathbf{b}_{ij} = \mathbf{X}'_{(2)ij}\, \Lambda_{ij}^{-1}\, \mathbf{y}_{ij}$$

$$= \frac{1}{\Phi_{(1)}}\, \mathbf{R}'_{ij}\, \mathbf{X}'_{(2)ij}\, \mathbf{y}_{ij}$$

(ii)
$$\mathbf{E}_{ij} = \mathbf{X}'_{(2)ij}\, \Lambda_{ij}^{-1}\, \Lambda_{ij}^{-1}\, \mathbf{X}_{(2)ij}$$

$$= \frac{1}{\Phi_{(1)}}\, \mathbf{A}_{ij}\, \mathbf{R}_{ij}$$

(iii)
$$\mathbf{F}_{ij} = \mathbf{X}'_{(2)ij}\, \Lambda_{ij}^{-1}\, \Lambda_{ij}^{-1}\, \mathbf{y}_{ij}$$

$$= \frac{1}{\Phi^2_{(1)}}\, \mathbf{R}_{ij}\, \mathbf{R}_{ij}\, \mathbf{X}'_{(2)ij}\, \mathbf{y}_{ij}$$

where

$$\mathbf{R}'_{ij} = \mathbf{I} - \frac{1}{\Phi_{(1)}}\, \mathbf{X}'_{(2)ij}\, \mathbf{X}_{(2)ij}\, (\, \Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\, \Phi_{(1)}^{-1}\, \mathbf{X}_{(2)ij}\,)^{-1}$$

and

$$\Lambda_{ij}^{-1} = \frac{1}{\Phi_{(1)}}\, \mathbf{I} - \frac{1}{\Phi^2_{(1)}}\, \mathbf{X}_{(2)ij}\, (\, \Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\, \Phi_{(1)}^{-1}\, \mathbf{X}_{(2)ij}\,)^{-1}\, \mathbf{X}'_{(2)ij}\, .$$

Expression (i) can be derived as follows:

$$\mathbf{b}_{ij} = \mathbf{X}'_{(2)ij}\, \Lambda_{ij}^{-1}\, \mathbf{y}_{ij}$$

$$= \mathbf{X}'_{(2)ij}\, \{\, \frac{1}{\Phi_{(1)}}\, \mathbf{I} - \frac{1}{\Phi^2_{(1)}}\, \mathbf{X}_{(2)ij}\, (\, \Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\, \Phi_{(1)}^{-1}\, \mathbf{X}_{(2)ij}\,)^{-1}\, \mathbf{X}'_{(2)ij}\, \}\, \mathbf{y}_{ij}$$

$$= \frac{1}{\Phi_{(1)}}\, \{\, \mathbf{I} - \frac{1}{\Phi_{(1)}}\, \mathbf{X}'_{(2)ij}\, \mathbf{X}_{(2)ij}\, (\, \Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\, \Phi_{(1)}^{-1}\, \mathbf{X}_{(2)ij}\,)^{-1}\}\, \mathbf{X}'_{(2)ij}\, \mathbf{y}_{ij}$$

$$= \frac{1}{\Phi_{(1)}}\, \mathbf{R}'_{ij}\, \mathbf{X}'_{(2)ij}\, \mathbf{y}_{ij}\, .$$

Expression (ii) can be simplified as

$$\begin{aligned}
\mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\Lambda_{ij}^{-1}\,\mathbf{X}_{(2)ij} = & \Big\{ \frac{1}{\Phi_{(1)}}\mathbf{X}'_{(2)ij} - \frac{1}{\Phi_{(1)}^{2}}\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \;\times \\
& (\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\Big\} \;\times \\
& \Big\{ \frac{1}{\Phi_{(1)}}\mathbf{X}_{(2)ij} - \frac{1}{\Phi_{(1)}^{2}}\mathbf{X}_{(2)ij} \;\times \\
& (\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\Big\} \\
= & \frac{1}{\Phi_{(1)}^{2}}\Big\{ \mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \;\times \\
& (\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\Big\}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \; - \\
& \frac{1}{\Phi_{(1)}^{3}}\Big\{ \mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \;\times \\
& (\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\Big\}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \\
= & \frac{1}{\Phi_{(1)}^{2}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} - \frac{1}{\Phi_{(1)}^{3}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \;\times \\
& (\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\Big\}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \\
= & \frac{1}{\Phi_{(1)}^{2}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\Big\{ \mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \;\times \\
& (\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\Big\}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,\Big\} \\
= & \frac{1}{\Phi_{(1)}^{2}}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,\mathbf{R}_{ij} \\
= & \frac{1}{\Phi_{(1)}}\,\mathbf{A}_{ij}\,\mathbf{R}'_{ij}\,.
\end{aligned}$$

Expression (iii) can be written as

$$\mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\Lambda_{ij}^{-1}\,\mathbf{y}_{ij} = \frac{1}{\Phi^2_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij} - \frac{1}{\Phi^3_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \times$$

$$(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij} -$$

$$\frac{1}{\Phi^3_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1} \times$$

$$\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij} + \frac{1}{\Phi^4_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \times$$

$$(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \times$$

$$(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij}$$

$$= \frac{1}{\Phi^2_{(1)}}\{\,\mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij} \times$$

$$(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1}\}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij} -$$

$$\frac{1}{\Phi^3_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij}\,)^{-1} \times$$

$$\{\,\mathbf{I} - \frac{1}{\Phi_{(1)}}\,\mathbf{X}'_{(2)ij}\,\mathbf{X}_{(2)ij}\,(\,\Phi_{(2)}^{-1} + \mathbf{X}'_{(2)ij}\,\Phi_{(1)}^{-1}\,\mathbf{X}_{(2)ij})^{-1}\}$$

$$\times\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij}$$

$$= \frac{1}{\Phi^2_{(1)}}\,\mathbf{R}'_{ij}\,\mathbf{R}'_{ij}\,\mathbf{X}'_{(2)ij}\,\mathbf{y}_{ij}\,.$$

CHAPTER 5
ANALYSIS OF GENERAL LEVEL-3 MODELS

5.1 INTRODUCTION

In the previous chapters a level-3 model was considered which was limited in a number of ways: the model made no allowance for a complex level-1 structure, the only assumption made being that level-1 error variates are uncorrelated and have constant variance $\Phi_{(1)}$. Also, variables random on level-3 were a subset or functions of variables random on level-2. (cf. (3.2.12), Section 3.2).

In this chapter attention will be given to the situation where any subset of p predictors may be random on any of the three levels of the model, thus allowing for complex level-1 error structures. Provision will also be made for the situation where there are no random variables on a particular level.

The extended general level-3 model will be discussed in detail in Section 5.2. Section 5.3 will deal with cases where, for a specific level, no random variables are present. Practical applications will be given in Section 5.4 and conclusions will be drawn in Section 5.5.

5.2 THE GENERAL LEVEL-3 MODEL

Consider the situation where a response variable y depends on a set of p predictors $x_1, x_2, \ldots, x_p$.

The general 3-level model is defined as

$$y_{ijk} = \mathbf{x}'_{(f)ijk}\,\boldsymbol{\beta} + \mathbf{x}'_{(3)ijk}\,\mathbf{v}_i + \mathbf{x}'_{(2)ijk}\,\mathbf{u}_{ij} + \mathbf{x}'_{(1)ijk}\,\mathbf{e}_{ijk}$$

where
$i = 1, 2, \ldots, N$ denotes level-3 units (for example education departments),
$j = 1, 2, \ldots, n_i$ denotes level-2 units (for example schools) and
$k = 1, 2, \ldots, n_{ij}$ denotes level-1 units (for example pupils).

$\mathbf{x}'_{(f)ijk}$: $1 \times s$ is a typical row of the design matrix of the fixed part of the model, the elements being a subset of the p predictors. $\mathbf{x}'_{(3)ijk}$: $1 \times q$, $\mathbf{x}'_{(2)ijk}$: $1 \times m$ and $\mathbf{x}'_{(1)ijk}$: $1 \times r$ are typical rows of the design matrices for the random part of the model on levels 3, 2 and 1 respectively. The elements of these vectors are also subsets of the p predictors. $\boldsymbol{\beta}$: $s \times 1$ is a vector of fixed, but unknown, parameters to be estimated.

It is assumed that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are independently and identically distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(3)}$. It is further assumed that $\mathbf{u}_{i1}, \mathbf{u}_{i2}, \ldots, \mathbf{u}_{in_i}$ are i.i.d. with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$, while $\mathbf{e}_{ij1}, \mathbf{e}_{ij2}, \ldots, \mathbf{e}_{ijn_{ij}}$ are i.i.d. with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(1)}$. Finally it is assumed that $\mathbf{v}_i$, $\mathbf{u}_{ij}$ and $\mathbf{e}_{ijk}$ are independent.

Let

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_{in_i} \end{bmatrix}$$

where $\mathbf{y}_{ij}$ denotes the $n_{ij} \times 1$ vector of responses for the i-th level-3 unit and the j-th level-2 unit.

Note that the $n_{ij} \times 1$ vector $\mathbf{y}_{ij}$ can be expressed as

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)ij}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}_{(2)ij}\,\mathbf{u}_{ij} + \begin{bmatrix} \mathbf{x}'_{(1)ij1}\,\mathbf{e}_{ij1} \\ \mathbf{x}'_{(1)ij2}\,\mathbf{e}_{ij2} \\ \vdots \\ \vdots \\ \mathbf{x}'_{(1)ijn_{ij}}\mathbf{e}_{ijn_{ij}} \end{bmatrix},$$

where

$$\mathbf{X}_{(f)ij} = \begin{bmatrix} \mathbf{x}'_{(f)ij1} \\ \mathbf{x}'_{(f)ij2} \\ \vdots \\ \vdots \\ \mathbf{x}'_{(f)ijn_{ij}} \end{bmatrix}, \quad \mathbf{X}_{(3)ij} = \begin{bmatrix} \mathbf{x}'_{(3)ij1} \\ \mathbf{x}'_{(3)ij2} \\ \vdots \\ \vdots \\ \mathbf{x}'_{(3)ijn_{ij}} \end{bmatrix}$$

and

$$\mathbf{X}_{(2)ij} = \begin{bmatrix} \mathbf{x}'_{(2)ij1} \\ \mathbf{x}'_{(2)ij2} \\ \vdots \\ \vdots \\ \mathbf{x}'_{(2)ijn_{ij}} \end{bmatrix}.$$

The model can be written as

$$\mathbf{y}_i = \mathbf{X}_{(f)i}\,\boldsymbol{\beta} + \mathbf{X}_{(3)i}\,\mathbf{v}_i + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\,\mathbf{u}_{ij} + \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \mathbf{U}_{(1)ijk}\,\mathbf{e}_{ijk} \tag{5.2.1}$$

where

$$\mathbf{X}_{(3)i} = \begin{bmatrix} \mathbf{X}_{(3)i1} \\ \mathbf{X}_{(3)i2} \\ \vdots \\ \vdots \\ \mathbf{X}_{(3)in_i} \end{bmatrix}, \tag{5.2.2}$$

$$\mathbf{Z}_{(2)ij} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{X}_{(2)ij} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \tag{5.2.3}$$

$$\mathbf{U}_{(1)ijk} = \begin{bmatrix} \mathbf{0}' \\ \vdots \\ \mathbf{x}'_{(1)ijk} \\ \vdots \\ \mathbf{0}' \end{bmatrix} \tag{5.2.4}$$

and where $\mathbf{v}_i$, $\mathbf{u}_{ij}$ and $\mathbf{e}_{ij}$ denote the random parameter vectors on level-3, level-2 and level-1 of the model. It will be convenient to replace the double subscript jk with the subscript l where $l = 1, 2, \ldots, n_i^*$ and

$$n_i^* = \sum_{j=1}^{n_i} n_{ij} .$$

Thus, (5.2.1) can be rewritten as

$$\mathbf{y}_i = \mathbf{X}_{(3)i}\,\mathbf{v}_i + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\,\mathbf{u}_{ij} + \sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il}\,\mathbf{e}_{il} . \tag{5.2.5}$$

Under the distributional assumptions given above, it follows that

$$\mathrm{E}(\mathbf{y}_i) = \mathbf{X}_{(f)i}\,\boldsymbol{\beta} ,$$

where

$$\mathbf{X}_{(f)i} = \begin{bmatrix} \mathbf{X}_{(f)i1} \\ \mathbf{X}_{(f)i2} \\ \vdots \\ \vdots \\ \mathbf{X}_{(f)in_i} \end{bmatrix} . \tag{5.2.6}$$

It also follows that

$$\mathrm{Cov}(\mathbf{y}_i , \mathbf{y}'_i) = \Sigma_i$$

$$= \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{Z}'_{(2)ij} +$$

$$\sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il} \, \mathbf{\Phi}_{(1)} \, \mathbf{U}'_{(1)il}$$

or, alternatively, as

$$\text{Cov}(\mathbf{y}_i \, , \, \mathbf{y}'_i) \;=\; \mathbf{X}_{(3)i} \, \mathbf{\Phi}_{(3)} \, \mathbf{X}'_{(3)i} \, + \, \mathbf{\Lambda}_i \, , \tag{5.2.7}$$

where

$$\mathbf{\Lambda}_i \;=\; \begin{bmatrix} \mathbf{\Lambda}_{i1} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{i2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{\Lambda}_{in_i} \end{bmatrix}, \tag{5.2.8}$$

with

$$\mathbf{\Lambda}_{ij} \;=\; \mathbf{X}_{(2)ij} \, \mathbf{\Phi}_{(2)} \, \mathbf{X}'_{(2)ij} + \mathbf{D}_{ij}$$

and where

$$\mathbf{D}_{ij} \;=\; \begin{bmatrix} \lambda_{ij1} & 0 & . & . & 0 \\ 0 & \lambda_{ij2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \lambda_{ijn_{ij}} \end{bmatrix}, \tag{5.2.9}$$

with

$$\lambda_{ijk} \;=\; \mathbf{x}'_{(1)ijk} \, \mathbf{\Phi}_{(1)} \; \mathbf{x}_{(1)ijk} \, . \tag{5.2.10}$$

Suppose that $\hat{\mathbf{\Phi}}_{(3)}$, $\hat{\mathbf{\Phi}}_{(2)}$ and $\hat{\mathbf{\Phi}}_{(1)}$ are consistent estimators of $\mathbf{\Phi}_{(3)}$, $\mathbf{\Phi}_{(2)}$ and $\mathbf{\Phi}_{(1)}$ respectively so that

$$\mathbf{V}_i \;=\; \mathbf{X}_{(3)i} \, \hat{\mathbf{\Phi}}_{(3)} \, \mathbf{X}'_{(3)i} \, + \, \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \, \hat{\mathbf{\Phi}}_{(2)} \, \mathbf{Z}'_{(2)ij} + \sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il} \, \hat{\mathbf{\Phi}}_{(1)} \, \mathbf{U}'_{(1)il}$$

is a consistent estimator of Σ_i.

The generalized least squares estimator $\widehat{\beta}$ of β is obtained as the minimum of the quadratic function (see also Section 4.2)

$$\mathbf{Q}_f = \sum_{i=1}^{N} [\, \mathbf{y}_i - \mathbf{X}_{(f)i}\, \beta \,]' \, \mathbf{V}_i^{-1} \, [\, \mathbf{y}_i - \mathbf{X}_{(f)i}\, \beta \,]$$

with solution

$$\widehat{\beta} = [\, \sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{X}_{(f)i} \,]^{-1} \, [\, \sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{y}_i \,]. \qquad (5.2.11)$$

In order to estimate $\mathbf{\Phi}_{(3)}$, $\mathbf{\Phi}_{(2)}$ and $\mathbf{\Phi}_{(1)}$ let (cf. (4.2.13))

$$\mathbf{y}_i^* = \text{vecs}\, (\mathbf{y}_i - \mathbf{X}_{(f)i}\, \beta\,)\, (\mathbf{y}_i - \mathbf{X}_{(f)i}\, \beta\,)' \,,$$

then

$$\text{E}\, (\mathbf{y}_i^*\,) = \text{vecs}\, \Sigma_i \,.$$

Using the result (Browne, 1974) on vector operations (cf. (4.2.11)), namely

$$\text{vec}\, (\, \mathbf{C}\, \mathbf{A}\, \mathbf{C}'\,) = (\, \mathbf{C} \otimes \mathbf{C}\,)\;\; \text{vec}\, \mathbf{A} \,,$$

it follows that

$$\text{vec}\, \mathbf{V}_i = \mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}\, \text{vec}\, \mathbf{\Phi}_{(3)} + \sum_{j=1}^{n_i} (\, \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}\,)\, \text{vec}\, \mathbf{\Phi}_{(2)} + \sum_{l=1}^{n_i^*} (\, \mathbf{U}_{(1)il} \otimes \mathbf{U}_{(1)il}\,)\;\; \text{vec}\, \mathbf{\Phi}_{(1)} \,.$$

There exists an unique matrix (see Browne, 1974 or McCulloch, 1982) $\mathbf{G}_p : p^2 \times \frac{1}{2}\, p(\, p + 1)$ such that

$$\text{vec}\, \mathbf{A} = \mathbf{G}_p\, \text{vecs}\, \mathbf{A} \qquad (5.2.12)$$

with $\mathbf{A}$ a symmetric $p \times p$ matrix. There is also a non-unique matrix $\mathbf{H}_p : \frac{1}{2}\, p\ (p+1) \times p^2$ so that

$$\text{vecs}\ \mathbf{A} = \mathbf{H}_p \text{ vec } \mathbf{A}. \tag{5.2.13}$$

The vector vecs Σ_i, consisting of the non-duplicated elements of Σ_i, can then be written as

$$\begin{aligned} \text{Vecs}\ \Sigma_i = {} & \mathbf{H}_{n_i^*} (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\ \mathbf{G}_q \text{ vecs } \mathbf{\Phi}_{(3)} + \\ & \mathbf{H}_{n_i^*} \Big(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij} \Big)\ \mathbf{G}_m \text{ vecs } \mathbf{\Phi}_{(2)} + \\ & \mathbf{H}_{n_i^*} \Big(\sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il} \otimes \mathbf{U}_{(1)il} \Big)\ \mathbf{G}_r \text{ vecs } \mathbf{\Phi}_{(1)} \\ = {} & \mathbf{X}_i^* \boldsymbol{\tau} \end{aligned}$$

where

$$\mathbf{X}_i^* = \mathbf{H}_{n_i^*} \Big[(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\ \mathbf{G}_q \quad \Big(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij} \Big)\ \mathbf{G}_m \quad \Big(\sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il} \otimes \mathbf{U}_{(1)il} \Big)\ \mathbf{G}_r \Big] \tag{5.2.14}$$

and

$$\boldsymbol{\tau} \quad = \begin{bmatrix} \text{vecs } \mathbf{\Phi}_{(3)} \\ \text{vecs } \mathbf{\Phi}_{(2)} \\ \text{vecs } \mathbf{\Phi}_{(1)} \end{bmatrix}. \tag{5.2.15}$$

Now consider the quadratic form

$$\mathbf{Q}_r = \sum_{i=1}^{N} \{ [\ \mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\tau}\ ]'\ \mathbf{W}_i^{-1} [\ \mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\tau}\ ]\ \}$$

where $\mathbf{W}_i$ is a consistent estimator of the covariance of $\mathbf{y}_i^*$ (see Proposition 4.2.3).

It has, for example, been shown by Browne (1974) and Goldstein (1989) that if

$$\mathbf{W}_i^{-1} = \tfrac{1}{2}\,\mathbf{G}'_{n_i}\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,, \tag{5.2.16}$$

then $\mathbf{W}_i$ is a consistent estimator of the covariance of $\mathbf{y}_i^*$.

Minimization of $\mathbf{Q}_r$ with respect to τ yields

$$\hat{\tau} = [\,\sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^*\,]^{-1}\,[\,\sum_{i=1}^{N} \mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{y}_i^*\,]. \tag{5.2.17}$$

In order to ensure computational efficiency, the components of $\hat{\tau}$ must be further simplified. Under the distributional assumptions given in the previous chapter, these components could be expressed in terms of matrices of sums of squares and cross products. For more complex error structures these results no longer hold. Although the simplifications given in this section follow a similar pattern to those in Chapter 4, the end results are quite different. For the purpose of clarity, certain basic results given in Chapter 4 are, therefore, repeated.

In Section 5.2.1 the simplification of $\mathbf{X}_i^{*\prime}\ \mathbf{W}_i^{-1}\ \mathbf{X}_i^*$ will be considered while simplification of the terms of $\mathbf{X}_i^{*\prime}\ \mathbf{W}_i^{-1}\ \mathbf{y}_i^*$ will be dealt with in Section 5.2.1. Finally, the simplification of the fixed part of the model will be discussed in Section 5.2.3.

5.2.1 SIMPLIFICATION OF THE COMPONENTS OF THE WEIGHT MATRIX

It follows from (5.2.14) and (5.2.16) that

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^* = \begin{bmatrix} \mathbf{G}'_q\,(\,\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}\,) \\ \mathbf{G}'_m\,(\,\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij}\,) \\ \mathbf{G}'_r\,(\,\sum_{l=1}^{n_i^*} \mathbf{U}'_{(1)il} \otimes \mathbf{U}'_{(1)il}\,) \end{bmatrix} \mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{H}_{n_i^*}$$

$$\times\ [(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\mathbf{G}_q \quad (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij})\mathbf{G}_m \quad (\sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il} \otimes \mathbf{U}_{(1)il})\mathbf{G}_r\,]$$

$$= \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} & \mathbf{T}'_{(1,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} & \mathbf{T}'_{(1,2)} \\ \mathbf{T}_{(1,3)} & \mathbf{T}_{(1,2)} & \mathbf{T}_{(1,1)} \end{bmatrix} \qquad (5.2.18)$$

$$= \quad \mathbf{T}_i$$

where, for example, the subscript (2,3) denotes the product obtained if a level-2 term is multiplied with a level-3 term. If, for example, there is no random term on level-1, the matrix $\mathbf{T}_i$ reduces to

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} \end{bmatrix} .$$

The submatrices $\mathbf{T}_{(1,1)}$ to $\mathbf{T}_{(3,3)}$ are given by

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q\,(\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i})\,\mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\,\mathbf{G}_{n_i}\mathbf{H}_{n_i^*}\,(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\,\mathbf{G}_q$$

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m\,(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij})\,\mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\,\mathbf{G}_{n_i}\,\mathbf{H}_{n_i^*}\,(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\,\mathbf{G}_q$$

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m\,(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij})\,\mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\,\mathbf{G}_{n_i}\,\mathbf{H}_{n_i^*} \times$$
$$(\sum_{j^*=1}^{n_i} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}_{(2)ij^*})\,\mathbf{G}_m$$

$$\mathbf{T}_{(1,3)} = \mathbf{G}'_r\,(\sum_{l=1}^{n_i^*} \mathbf{U}'_{(1)il} \otimes \mathbf{U}'_{(1)il})\,\mathbf{H}'_{n_i^*}\,\mathbf{G}'_{n_i}\,(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\,\mathbf{G}_{n_i}\,\mathbf{H}_{n_i^*}\,(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\,\mathbf{G}_q$$

$$\mathbf{T}_{(1,2)} = \mathbf{G}'_r \left(\sum_{j=1}^{n_i^*} \mathbf{U}'_{(1)il} \otimes \mathbf{U}'_{(1)il} \right) \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} \left(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1} \right) \mathbf{G}_{n_i} \mathbf{H}_{n_i^*} \times$$
$$\left(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij} \right) \mathbf{G}_m$$
$$\mathbf{T}_{(1,1)} = \mathbf{G}'_r \left(\sum_{l=1}^{n_i^*} \mathbf{U}'_{(1)il} \otimes \mathbf{U}'_{(1)il} \right) \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} \left(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1} \right) \mathbf{G}_{n_i} \mathbf{H}_{n_i^*} \times$$
$$\left(\sum_{l^*=1}^{n_i^*} \mathbf{U}_{(1)il^*} \otimes \mathbf{U}_{(1)il^*} \right) \mathbf{G}_r .$$

Using the results (Magnus & Neudecker, 1988)

$$\mathbf{G}_k \mathbf{H}_k (\mathbf{X} \otimes \mathbf{X}) \mathbf{G}_m = (\mathbf{X} \otimes \mathbf{X}) \mathbf{G}_m ,$$

and (Graham, 1981)

$$(\mathbf{A}' \otimes \mathbf{B}')(\mathbf{C} \otimes \mathbf{C})(\mathbf{A} \otimes \mathbf{B}) = \mathbf{A}'\mathbf{C}\mathbf{A} \otimes \mathbf{B}'\mathbf{C}\mathbf{B}$$

the submatrices $\mathbf{T}_{(1,1)}$ to $\mathbf{T}_{(3,3)}$ can be written as

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q \left(\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \right) \mathbf{G}_q \quad (5.2.19)$$

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m \left(\sum_{j=1}^{n_i} \{ \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \} \right) \mathbf{G}_q \quad (5.2.20)$$

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \left(\sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} \{ \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} \} \right) \mathbf{G}_m \quad (5.2.21)$$

$$\mathbf{T}_{(1,3)} = \mathbf{G}'_r \left(\sum_{l=1}^{n_i^*} \{ \mathbf{U}'_{(1)il} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{U}'_{(1)il} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \} \right) \mathbf{G}_q \quad (5.2.22)$$

$$\mathbf{T}_{(1,2)} = \mathbf{G}'_r \left(\sum_{j=1}^{n_i} \sum_{l=1}^{n_i^*} \{ \mathbf{U}'_{(1)il} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij} \otimes \mathbf{U}'_{(1)il} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij} \} \right) \mathbf{G}_m \quad (5.2.23)$$

$$\mathbf{T}_{(1,1)} = \mathbf{G}'_r \left(\sum_{l=1}^{n_i^*} \sum_{l^*=1}^{n_i^*} \{ \mathbf{U}'_{(1)il} \mathbf{V}_i^{-1} \mathbf{U}_{(1)il^*} \otimes \mathbf{U}'_{(1)il} \mathbf{V}_i^{-1} \mathbf{U}_{(1)il^*} \} \right) \mathbf{G}_r . \quad (5.2.24)$$

Computationally efficient expressions for each of the submatrices $\mathbf{T}_{(3,3)}$, $\mathbf{T}_{(2,3)}$, . . . $\mathbf{T}_{(1,1)}$ are given in Propositions 5.2.1 to 5.2.3 and Propositions 5.2.7 to 5.2.9. The results derived in these propositions are subsequently summarized in Theorem 5.2.1.

PROPOSITION 5.2.1 (Submatrix $\mathbf{T}_{(3,3)}$)

$$\begin{aligned}\mathbf{T}_{(3,3)} &= \mathbf{G}'_q\,(\,\mathbf{X}'_{(3)i}\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} \otimes \mathbf{X}'_{(3)i}\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,)\,\mathbf{G}_q \\ &= \mathbf{G}'_q\,\{\,\mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\mathbf{A}_i\,] \otimes \mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\mathbf{A}_i\,]\}\,\mathbf{G}_q \qquad (5.2.25)\end{aligned}$$

where

$$\mathbf{A}_i = \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} \qquad (5.2.26)$$

and

$$\mathbf{C}_i = (\boldsymbol{\Phi}_{(3)}^{-1} + \mathbf{A}_i)^{-1}\,. \qquad (5.2.27)$$

Proof

Using (4.3.11), it follows that $\mathbf{V}_i^{-1}$ can be written as

$$\mathbf{V}_i^{-1} = \boldsymbol{\Lambda}_i^{-1} - \boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1} \qquad (5.2.28)$$

with

$$\mathbf{C}_i = (\boldsymbol{\Phi}_{(3)}^{-1} + \mathbf{X}_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}'_{(3)i}\,)^{-1}.$$

Then

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} - \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,. \qquad (5.2.29)$$

Let

$$\mathbf{A}_i = \mathbf{X}_{(3)i}\,\Lambda_i^{-1}\,\mathbf{X}'_{(3)i}$$
$$= \sum_{j=1}^{n_i} \mathbf{X}_{(3)ij}\,\Lambda_{ij}^{-1}\,\mathbf{X}'_{(3)ij}\ .$$

Subsequently, (5.2.29) can be rewritten as

$$\mathbf{X}_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}'_{(3)i} = \mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]\ . \tag{5.2.30}$$

Substitution of (5.2.30) in (5.2.29) concludes the proof. □

PROPOSITION 5.2.2 (Submatrix $\mathbf{T}_{(2,3)}$)

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m\,(\sum_{j=1}^{n_i}\{\,\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} \otimes \mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\})\,\mathbf{G}_q$$
$$= \mathbf{G}'_m\,(\sum_{j=1}^{n_i}\{\,\mathbf{B}_{ij}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \otimes \mathbf{B}_{ij}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]\}\,\mathbf{G}_q \tag{5.2.31}$$

where

$$\mathbf{B}_{ij} = \mathbf{X}'_{(2)ij}\,\Lambda_{ij}^{-1}\,\mathbf{X}_{(3)ij} \tag{5.2.32}$$

and $\mathbf{A}_i$ and $\mathbf{C}_i$ are given by (5.2.26) and (5.2.27) respectively.

Proof

Using (5.2.28), it follows that

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{Z}'_{(2)ij}\,\Lambda_i^{-1}\,\mathbf{X}_{(3)i} - \mathbf{Z}'_{(2)ij}\,\Lambda_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\Lambda_i^{-1}\,\mathbf{X}_{(3)i} \tag{5.2.33}$$

where (cf. (5.2.26))

$$\mathbf{A}_i = \mathbf{X}'_{(3)i}\,\Lambda_i^{-1}\,\mathbf{X}_{(3)i}$$

and

$$\mathbf{Z}_{(2)ij}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \begin{bmatrix} \mathbf{0}' & \cdots & \mathbf{X}'_{(2)ij} & \mathbf{0}' & \cdots & \mathbf{0}' \end{bmatrix} \times$$

$$\begin{bmatrix} \boldsymbol{\Lambda}_{i1}^{-1} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{i2}^{-1} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & \mathbf{0} & \boldsymbol{\Lambda}_{in_i}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(3)i1} \\ \mathbf{X}_{(3)i2} \\ \vdots \\ \vdots \\ \mathbf{X}_{(3)in_i} \end{bmatrix}$$

$$= \mathbf{X}'_{(2)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij}\ .$$

Let

$$\mathbf{B}_{ij} = \mathbf{X}'_{(2)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij}\ , \tag{5.2.34}$$

then

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{B}_{ij}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \tag{5.2.35}$$

which concludes the proof. □

PROPOSITION 5.2.3 (Submatrix $\mathbf{T}_{(2,2)}$)

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m\,(\sum_{j=1}^{n_i}\sum_{j^*=1}^{n_i}\{\,\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*}\,\})\ \mathbf{G}_m$$

$$= \mathbf{G}'_m\,(\sum_{j=1}^{n_i}\sum_{j^*=1}^{n_i}\{\,(\,\delta_{jj^*}\,\mathbf{E}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,)\otimes(\,\delta_{jj^*}\,\mathbf{E}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,)\})\ \mathbf{G}_m \tag{5.2.36}$$

with

$$\mathbf{E}_{ij} = \mathbf{X}'_{(2)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(2)ij} \tag{5.2.37}$$

and $\mathbf{B}_{ij}$ as given by (5.2.34), Proposition 5.2.2.

Proof

From (5.2.21) it follows that

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \left(\sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} \{ \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} \} \right) \mathbf{G}_m \,.$$

Using (5.2.28), $\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*}$ can be written as

$$\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} = \mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} - \mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}'_{(3)i} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} \tag{5.2.38}$$

where (cf. (5.2.32))

$$\begin{aligned} \mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(3)i} &= \mathbf{X}'_{(2)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(3)ij} \\ &= \mathbf{B}_{ij} \,, \end{aligned} \tag{5.2.39}$$

$$\begin{aligned} \mathbf{X}'_{(3)i} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} &= \mathbf{X}'_{(3)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)ij^*} \\ &= \mathbf{B}'_{ij^*} \,, \end{aligned} \tag{5.2.40}$$

and

$$\begin{aligned} \mathbf{Z}'_{(2)ij} \mathbf{\Lambda}_i^{-1} \mathbf{Z}_{(2)ij^*} &= \mathbf{X}'_{(2)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)ij^*} \\ &= \mathbf{X}'_{(2)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)ij} \quad \text{if } j = j^* \\ &= \mathbf{0} \quad \text{otherwise.} \end{aligned}$$

Let

$$\mathbf{E}_{ij} = \mathbf{X}'_{(2)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(2)ij} \,.$$

Using (5.2.37), and substituting (5.2.39) and (5.2.40) in (5.2.38), gives

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} = \delta_{jj^*}\,\mathbf{E}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}$$

which concludes the proof. □

PROPOSITION 5.2.4

Suppose $\mathbf{X}_\alpha$ is an $n \times r$ matrix with all rows $\mathbf{0}'$ except the α-th row, that is

$$\mathbf{X}_\alpha = \begin{bmatrix} \mathbf{0}' \\ \vdots \\ \mathbf{0}' \\ \mathbf{x}'_\alpha \\ \mathbf{0}' \\ \vdots \\ \mathbf{0}' \end{bmatrix}$$

and $\mathbf{A}$ is an $n \times n$ symmetric matrix with typical row $\mathbf{a}'_\gamma$, $\gamma = 1, 2, \ldots, n$. Then

$$\mathbf{X}'_\alpha\,\mathbf{A}\,\mathbf{X}_\beta = \mathrm{a}_{\alpha\beta}\,\mathrm{x}_\alpha\,\mathrm{x}'_\beta\,. \tag{5.2.41}$$

Proof

$$[\,\mathbf{X}'_\alpha\,\mathbf{A}\,]_{i,k} = \sum_{l=1}^{n}[\,\mathbf{X}'_\alpha\,]_{i,l}\,[\mathbf{A}]_{l,k}$$

$$= \mathrm{x}_{\alpha\,i}\,\mathrm{a}_{\alpha\,k}\,,$$

since all other terms are equal to zero. A typical element of the matrix $\mathbf{X}'_\alpha\,\mathbf{A}\,\mathbf{X}_\beta$ is

$$[\,\mathbf{X}'_\alpha\,\mathbf{A}\,\mathbf{X}_\beta\,]_{i,j} = \sum_{k=1}^{n}[\,\mathbf{X}_\alpha\,\mathbf{A}\,]_{i,k}\,[\,\mathbf{X}_\beta\,]_{k,j}$$

$$= \sum_{k=1}^{n}[\,\mathrm{x}_{\alpha i}\ \mathrm{a}_{\alpha,k}\,]\,[\,\mathbf{X}_\beta\,]_{k,j}$$

$$= x_{\alpha i}\, a_{\alpha\beta}\, x_{\beta j}$$

$$= a_{\alpha\beta}\, x_{\alpha i}\, x_{\beta j}\,.$$

Thus,

$$\mathbf{X}'_{\alpha}\, \mathbf{A}\, \mathbf{X}_{\beta} \;=\; a_{\alpha\beta}\, \mathbf{x}_{\alpha}\, \mathbf{x}'_{\beta}\,. \qquad \square$$

PROPOSITION 5.2.5

$$\mathbf{U}'_{(1)ijk}\, \boldsymbol{\Lambda}_i^{-1}\, \mathbf{Z}_{(2)ij} = \mathbf{x}_{(1)ijk}\, \mathbf{f}'_k \qquad (5.5.42)$$

where $\mathbf{f}'_k$ denotes the k-th row of $\mathbf{F} = \boldsymbol{\Lambda}_{ij}^{-1}\, \mathbf{X}_{(2)ij}$.

Proof

Let (cf. (5.2.4))

$$\mathbf{U}_{(1)ijk} \;=\; \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{V}_{ijk} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \qquad (5.2.43)$$

where each matrix in the partitioning of $\mathbf{U}_{(1)jk}$ is a $n_{i\alpha} \times r$ null matrix, $\alpha = 1, 2, \ldots, n_i$; $\alpha = j$, except the $n_{ij} \times r$ matrix $\mathbf{V}_{ijk}$ where

$$\mathbf{V}_{ijk} = \begin{bmatrix} \mathbf{0}' \\ \vdots \\ \mathbf{0}' \\ \mathbf{x}'_{(1)ijk} \\ \mathbf{0}' \\ \vdots \\ \mathbf{0}' \end{bmatrix}. \tag{5.2.44}$$

From (5.2.3) and (5.2.8) it follows that

$$\begin{aligned} \mathbf{U}'_{(1)ijk}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij} &= \mathbf{V}'_{ijk}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(2)ij} \\ &= [\,\mathbf{0}\ \dots\ \mathbf{0}\ \ \mathbf{x}_{(1)ijk}\ \ \mathbf{0}\ \ \dots\ \mathbf{0}\,] \begin{bmatrix} \mathbf{f}'_1 \\ \vdots \\ \mathbf{f}'_k \\ \vdots \\ \mathbf{f}'_{nij} \end{bmatrix} \end{aligned}$$

which concludes the proof. □

PROPOSITION 5.2.6

$$\mathbf{U}'_{(1)ijk}\ \mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k \tag{5.2.45}$$

where $\mathbf{g}'_k$ denotes the k-th row of $\mathbf{G} = \mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij}$.

Proof

The proof follows directly from (5.2.4) and Proposition 5.2.5. □

PROPOSITION 5.2.7 (Submatrix $\mathbf{T}_{(1,3)}$)

$$\mathbf{T}_{(1,3)} = \mathbf{G}_r' \left(\sum_{l=1}^{n_i^*} \{ \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \otimes \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} \} \right) \mathbf{G}_q$$
$$= \mathbf{G}_r' \left(\sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \{ \mathbf{x}_{(1)ijk} \mathbf{g}_k' [\mathbf{I} - \mathbf{C}_i \mathbf{A}_i] \otimes \mathbf{x}_{(1)ijk} \mathbf{g}_k' [\mathbf{I} - \mathbf{C}_i \mathbf{A}_i] \} \right) \mathbf{G}_q \tag{5.2.46}$$

where (cf. (5.2.27) and (5.2.26))

$$\mathbf{C}_i = (\boldsymbol{\Phi}_{(3)}^{-1} + \mathbf{A}_i)^{-1}, \tag{5.2.47}$$

$$\mathbf{A}_i = \mathbf{X}_{(3)i} \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \tag{5.2.48}$$

and $\mathbf{g}_k$ is as defined in Proposition 5.2.4.

Proof

From (5.2.28) it follows that

$$\mathbf{U}_{(1)ijk}' \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} = \mathbf{U}_{(1)ijk}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} - \mathbf{U}_{(1)ijk}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} . \tag{5.2.49}$$

Also (cf. (5.2.26))

$$\mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} = \mathbf{A}_i \tag{5.2.50}$$

and (cf. (5.2.45), Proposition 5.2.6)

$$\mathbf{U}_{(1)ijk}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} = \mathbf{x}_{(1)ijk} \mathbf{g}_k' . \tag{5.2.51}$$

Substitution of (5.2.50) and (5.2.51) in (5.2.49) gives

$$\mathbf{U}'_{(1)ijk}\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,[\,\mathbf{I}\,-\,\mathbf{C}_i\,\mathbf{A}_i\,]$$

which concludes the proof. □

PROPOSITION 5.2.8 (Submatrix $\mathbf{T}_{(1,2)}$)

$$\mathbf{T}_{(1,2)} = \mathbf{G}'_r\,(\,\sum_{l=1}^{n_i^*}\,\sum_{j^*=1}^{n_i}\,\{\,\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} \otimes \mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*}\,\})\,\mathbf{G}_m$$
$$= \mathbf{G}'_r\,(\,\sum_{j=i}^{n_i}\,\sum_{j^*=1}^{n_i}\,\sum_{k=1}^{n_{ij}}\,\{\,\mathbf{x}_{(1)ijk}\,[\,\mathbf{f}'_k\,-\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,] \otimes \mathbf{x}_{(1)ijk}\,[\,\mathbf{f}'_k\,-\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,]\})\,\mathbf{G}_m\,. \qquad (5.2.52)$$

Proof

From (5.2.28) it follows that

$$\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} = \mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij^*} - \mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\;\;\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij^*} \qquad (5.2.53)$$

where (cf. Proposition 5.2.5 and Proposition 5.2.6)

$$\mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,,$$

$$\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij^*} = \mathbf{X}'_{(3)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(2)ij^*}$$
$$= \mathbf{B}'_{ij^*}$$

and

$$\mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij} = \mathbf{x}_{(1)ijk}\,\mathbf{f}'_k\;\;.$$

Thus, (5.2.53) can be rewritten as

$$\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij} = \mathbf{x}_{(1)ijk}\,\mathbf{f}'_k - \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}$$

$$= \mathbf{x}_{(1)ijk}\,[\,\mathbf{f}'_k - \mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,]$$

which concludes the proof. □

PROPOSITION 5.2.9 (Submatrix $\mathbf{T}_{(1,1)}$)

$$\mathbf{T}_{(1,1)} = \mathbf{G}'_r\,(\sum_{l=1}^{n_i^*}\sum_{l^*=1}^{n_i^*}\{\,\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{U}_{(1)il^*}\otimes\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{U}_{(1)il^*}\})\,\mathbf{G}_r$$

$$= \mathbf{G}'_r\,(\sum_{j^*=1}^{n_i}\sum_{j=1}^{n_i}\sum_{k^*=1}^{n_{ij}}\sum_{k=1}^{n_{ij}}\{\,(\,(\delta_{jj^*})\,[\,\Lambda_{ij}^{-1}\,]_{kk^*}\,\mathbf{x}_{(1)ijk}\,\mathbf{x}'_{(1)ij^*k^*} -$$

$$\mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{g}_{k^*}\,\mathbf{x}'_{(1)ij^*k^*}\otimes(\delta_{jj^*})\,[\,\Lambda_{ij}^{-1}\,]_{kk^*}\,\mathbf{x}_{(1)ijk}\,\mathbf{x}'_{(1)ij^*k^*} -$$

$$\mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{g}_{k^*}\,\mathbf{x}'_{(1)ij^*k^*}\,\}\,)\,\mathbf{G}_r\,. \tag{5.2.54}$$

Proof

From (5.2.28) it follows that

$$\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{U}_{(1)il^*} = \mathbf{U}'_{(1)il}\,\Lambda_i^{-1}\,\mathbf{U}_{(1)il^*} -$$

$$\mathbf{U}'_{(1)il}\,\Lambda_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\Lambda_i^{-1}\,\mathbf{U}_{(1)il^*}$$

with

$$\mathbf{U}'_{(1)il}\,\Lambda_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,. \tag{5.2.55}$$

It also follows (cf. Proposition 5.2.6) that

$$\begin{aligned} \mathbf{U}'_{(1)il}\,\Lambda_i^{-1}\,\mathbf{U}_{(1)il^*} &= [\,\Lambda_{ij}^{-1}\,]_{kk^*}\,\mathbf{x}_{(1)ijk}\,\mathbf{x}'_{(1)ij^*k^*}\ ,\ j = j^* \\ &= \mathbf{0}\ ,\ j \neq j^* \end{aligned} \tag{5.2.56}$$

so that

$$\mathbf{U}'_{(1)il}\ \mathbf{V}_i^{-1}\,\mathbf{U}_{(1)il^*} = (\delta_{jj^*})\,[\,\Lambda_{ij}^{-1}\,]_{kk^*}\,\mathbf{x}_{(1)ijk}\,\mathbf{x}'_{(1)ij^*k^*} - \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{g}_{k^*}\,\mathbf{x}'_{(1)ij^*k^*}$$

which concludes the proof. □

THEOREM 5.2.1

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^* = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} & \mathbf{T}'_{(1,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} & \mathbf{T}'_{(1,2)} \\ \mathbf{T}_{(1,3)} & \mathbf{T}_{(1,2)} & \mathbf{T}_{(1,1)} \end{bmatrix}$$

where

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q\,\{\,\mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \otimes \mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\mathbf{A}_i\,]\}\,\mathbf{G}_q \tag{5.2.57}$$

$$\mathbf{T}_{(2,\,3)} = \mathbf{G}'_m\,\{\,\sum_{j=1}^{n_i}\,(\,\mathbf{B}_{ij}\,(\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,) \otimes \mathbf{B}_{ij}\,(\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,))\}\,\mathbf{G}_q \tag{5.2.58}$$

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m\,(\,\sum_{j=1}^{n_i}\,\sum_{j^*=i}^{n_i}\,\{\,(\,(\delta_{jj^*})\,\mathbf{E}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,) \otimes (\,(\delta_{jj^*})\,\mathbf{E}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,)\})\,\mathbf{G}_m \tag{5.2.59}$$

$$\mathbf{T}_{(1,3)} = \mathbf{G}'_r\,(\,\sum_{j=1}^{n_i}\,\sum_{k=1}^{n_{ij}}\,\{\,\mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \otimes \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]\})\,\mathbf{G}_q \tag{5.2.60}$$

$$\mathbf{T}_{(1,2)} = \mathbf{G}'_r\,(\,\sum_{j=1}^{n_i}\,\sum_{j^*=1}^{n_i}\,\sum_{k=1}^{n_{ij}}\,\{\,\mathbf{x}_{(1)ijk}\,[\,\mathbf{f}'_k - \mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}\,] \otimes \mathbf{x}_{(1)ijk}\,[\,\mathbf{f}'_k - \mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{B}'_{ij^*}]\})\,\mathbf{G}_m \tag{5.2.61}$$

$$\mathbf{T}_{(1,\,1)} = \mathbf{G}_r' \Big(\sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} \sum_{k=1}^{n_{ij}} \sum_{k^*=1}^{n_{ij}} \{ (\delta_{jj^*}) [\Lambda_{ij}^{-1}]_{kk^*} \mathbf{x}_{(1)ijk} \mathbf{x}_{(1)ij^*k^*}' - \mathbf{x}_{(1)ijk} \mathbf{g}_k' \mathbf{C}_i \mathbf{g}_{k^*} \times$$

$$\mathbf{x}_{(1)ij^*k^*}' \otimes (\delta_{jj^*}) [\Lambda_{ij}^{-1}]_{kk^*} \mathbf{x}_{(1)ijk} \mathbf{x}_{(1)ij^*k^*}' - \mathbf{x}_{(1)ijk} \mathbf{g}_k' \mathbf{C}_i \mathbf{g}_{k^*} \mathbf{x}_{(1)ij^*k^*}' \} \Big) \mathbf{G}_r \tag{5.2.62}$$

with (cf. (5.2.26), (5.2.27), (5.2.32) and (5.2.37))

$$\mathbf{A}_i = \mathbf{X}_{(3)i}' \Lambda_i^{-1} \mathbf{X}_{(3)i}\,,$$

$$\mathbf{C}_i = (\Phi_{(3)}^{-1} + \mathbf{A}_i)^{-1},$$

$$\mathbf{B}_{ij} = \mathbf{X}_{(2)ij}' \Lambda_{ij}^{-1} \mathbf{X}_{(3)ij}\,,$$

$$\mathbf{E}_{ij} = \mathbf{X}_{(2)ij}' \Lambda_{ij}^{-1} \mathbf{X}_{(2)ij}$$

and $\mathbf{g}'$ and $\mathbf{f}_k'$ as defined in Propositions 5.2.4 and 5.2.5 respectively.

Proof

The proof follows directly from Propositions 5.2.1 to 5.2.9. □

In this subsection expressions required for the efficient computation of the weight matrix $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^*$ were derived. To calculate $\hat{\tau}$ (cf. (5.2.15) and (5.2.17)) and hence $\hat{\Phi}_{(3)}$, $\hat{\Phi}_{(2)}$ and $\hat{\Phi}_{(1)}$ it is also necessary to find a computationally efficient way to calculate the coefficient vector $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{y}_i^*$. This matter will be dealt with in the next section.

5.2.2 SIMPLIFICATION OF THE COMPONENTS OF THE COEFFICIENT VECTOR

It follows from (5.2.14) and (5.2.16) that the vector $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-\prime} \mathbf{y}_i^*$ can be written as

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{y}_i^* = \begin{bmatrix} \mathbf{G}_q'\,(\,\mathbf{X}_{(3)i}' \otimes \mathbf{X}_{(3)i}'\,) \\ \mathbf{G}_m'\,(\,\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}'\,) \\ \mathbf{G}_r'\,(\,\sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il}' \otimes \mathbf{U}_{(1)il}'\,) \end{bmatrix} \mathbf{H}_{n_i^*}'\,\mathbf{G}_{n_i}'\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*$$

$$= \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \\ \mathbf{q}_{(1)} \end{bmatrix} = \mathbf{q}_i \tag{5.2.63}$$

where the subscripts (3), (2) and (1) denote the level-3, level-2 and level-1 terms respectively. The vectors $\mathbf{q}_{(3)}$, $\mathbf{q}_{(2)}$ and $\mathbf{q}_{(1)}$ are given by

$$\mathbf{q}_{(3)} = \mathbf{G}_q'\,(\,\mathbf{X}_{(3)i}' \otimes \mathbf{X}_{(3)i}'\,)\,\mathbf{H}_{n_i^*}'\,\mathbf{G}_{n_i}'\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*\,,$$

$$\mathbf{q}_{(2)} = \mathbf{G}_m'\,(\,\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}'\,)\,\mathbf{H}_{n_i^*}'\,\mathbf{G}_{n_i}'\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}_r'\,(\,\sum_{l=1}^{n_i^*} \mathbf{U}_{(1)il}' \otimes \mathbf{U}_{(1)il}'\,)\,\mathbf{H}_{n_i^*}'\,\mathbf{G}_{n_i}'\,(\,\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}\,)\;\mathbf{G}_{n_i}\,\mathbf{y}_i^*\;.$$

Using result (4.3.5) and (cf. (4.3.7))

$$(\,\mathbf{A}' \otimes \mathbf{B}'\,)\,(\,\mathbf{C} \otimes \mathbf{C}\,)\,(\,\mathbf{A} \otimes \mathbf{B}\,) = \mathbf{A}'\,\mathbf{C}\,\mathbf{A} \otimes \mathbf{B}'\,\mathbf{C}\,\mathbf{B}\;,$$

the vectors $\mathbf{q}_{(1)}$ to $\mathbf{q}_{(3)}$ can be expressed as

$$\mathbf{q}_{(3)} = \mathbf{G}_q'\,(\,\mathbf{X}_{(3)i}'\,\mathbf{V}_i^{-1} \otimes \mathbf{X}_{(3)i}'\,\mathbf{V}_i^{-1}\,)\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*\,,$$

$$\mathbf{q}_{(2)} = \mathbf{G}_m'\,(\,\sum_{j=1}^{n_i} \{\,\mathbf{Z}_{(2)ij}'\,\mathbf{V}_i^{-1} \otimes \mathbf{Z}_{(2)ij}'\,\mathbf{V}_i^{-1}\,\})\,\mathbf{G}_{n_i}\,\mathbf{y}_i^*$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}_r' \left(\sum_{j=1}^{n_i^*} \{ \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \otimes \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \} \right) \mathbf{G}_{n_i} \mathbf{y}_i^* \ .$$

Note that

$$\mathbf{Y}_i = (\mathbf{y}_i - \mathbf{X}_{(f)i} \boldsymbol{\beta}) (\mathbf{y}_i - \mathbf{X}_{(f)i} \boldsymbol{\beta})', \tag{5.2.64}$$

and

$$\mathbf{y}_i^* = \text{vecs} \, (\mathbf{Y}_i) \ .$$

Using (5.2.12), the vectors $\mathbf{q}_{(1)}$ to $\mathbf{q}_{(3)}$ can then be simplified as follows:

$$\mathbf{q}_{(3)} = \mathbf{G}_q' (\mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \otimes \mathbf{X}_{(3)i}' \mathbf{V}_i^{-1}) \text{ vec } (\mathbf{Y}_i) \ ,$$

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \left(\sum_{j=1}^{n_i} \{ \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \otimes \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \} \right) \text{ vec } (\mathbf{Y}_i)$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}_r' \left(\sum_{l=1}^{n_i^*} \{ \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \otimes \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \} \right) \text{ vec } (\mathbf{Y}_i) \ .$$

Finally, use of (4.2.11) gives

$$\mathbf{q}_{(3)} = \mathbf{G}_q' \text{ vec } (\mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}^{-1} \mathbf{X}_{(3)i}) \tag{5.2.65}$$

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \text{ vec } \left(\sum_{j=1}^{n_i} \{ \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}^{-1} \mathbf{Z}_{(2)ij} \} \right) \tag{5.2.66}$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}_r' \text{ vec } \left(\sum_{l=1}^{n_i^*} \{ \mathbf{U}_{(1)il}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{U}_{(1)il} \} \right) . \tag{5.2.67}$$

Computationally efficient expressions for each of the subvectors $\mathbf{q}_{(3)}$, $\mathbf{q}_{(2)}$ and $\mathbf{q}_{(1)}$ are given in Propositions 5.2.10 to 5.2.13. These results are summarized in Theorem 5.2.2.

PROPOSITION 5.2.10 (Subvector $\mathbf{q}_{(3)}$)

$$\mathbf{q}_{(3)} = \mathbf{G}_q' \operatorname{vec} (\mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{X}_{(3)i})$$

$$= \mathbf{G}_q' \operatorname{vec} (\mathbf{a}_i \mathbf{a}_i') \tag{5.2.68}$$

where

$$\mathbf{a}_i = [\mathbf{I} - \mathbf{A}_i \mathbf{C}_i] \mathbf{h}_i \tag{5.2.69}$$

with

$$\mathbf{h}_i = \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{e}_i . \tag{5.2.70}$$

Proof

Let

$$\mathbf{a}_i = \mathbf{X}_{(3)i} \mathbf{V}_i^{-1} \mathbf{e}_i ,$$

where

$$\mathbf{e}_i = (\mathbf{y}_i - \mathbf{X}_{(f)i} \boldsymbol{\beta}) . \tag{5.2.71}$$

Then $\mathbf{a}_i \mathbf{V}_i^{-1} \mathbf{e}_i$ can be written as (cf. (5.2.28))

$$\mathbf{a}_i = \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{e}_i - \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{e}_i$$

$$= [\mathbf{I} - \mathbf{A}_i \mathbf{C}_i] \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}^{-1} \mathbf{e}_i \tag{5.2.72}$$

with $\mathbf{A}_i$ as defined in (5.2.26) and

$$\mathbf{h}_i = \mathbf{X}_{(3)i}' \boldsymbol{\Lambda}_i^{-1} \mathbf{e}_i = \sum_{j=1}^{n_i} \mathbf{X}_{(3)ij}' \boldsymbol{\Lambda}_{ij}^{-1} \mathbf{e}_{ij} . \tag{5.2.73}$$

From (5.2.73), (5.2.72) can be rewritten as

$$\mathbf{a}_i = [\, \mathbf{I} - \mathbf{A}_i \, \mathbf{C}_i \,] \, \mathbf{h}_i$$

which concludes the proof. □

PROPOSITION 5.2.11 (Subvector $\mathbf{q}_{(2)}$)

$$\begin{aligned} \mathbf{q}_{(2)} &= \mathbf{G}'_m \operatorname{vec} \left(\sum_{j=1}^{n_i} \{ \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{Y}_i \, \mathbf{V}_i^{-1} \, \mathbf{Z}_{(2)ij} \} \right) \\ &= \mathbf{G}'_m \operatorname{vec} \sum_{j=1}^{n_i} (\mathbf{b}_{ij} \, \mathbf{b}'_{ij}) \end{aligned} \tag{5.2.74}$$

with

$$\mathbf{b}_{ij} = \mathbf{k}_{ij} - \mathbf{B}_{ij} \, \mathbf{C}_i \, \mathbf{h}_i \tag{5.2.75}$$

where

$$\mathbf{k}_{ij} = \mathbf{X}'_{(2)ij} \, \boldsymbol{\Lambda}_{ij}^{-1} \, \mathbf{e}_{ij} \tag{5.2.76}$$

and $\mathbf{B}_{ij}$, $\mathbf{C}_i$ and $\mathbf{h}_i$ are given by (5.2.34), (5.2.27) and (5.2.73) respectively.

Proof

Let

$$\mathbf{b}_{ij} = \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{e}_i$$

with $\mathbf{e}_i$ given by (5.2.71). From (5.2.28) it follows that

$$\mathbf{b}_{ij} = \mathbf{Z}'_{(2)ij} \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{e}_i - \mathbf{Z}'_{(2)ij} \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{X}_{(3)i} \, \mathbf{C}_i \, \mathbf{X}'_{(3)i} \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{e}_i \tag{5.2.77}$$

with (cf. (5.2.33) and (5.2.70))

$$\mathbf{Z}'_{(2)ij}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{X}'_{(2)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij}$$

$$= \mathbf{B}_{ij}\,,$$

and

$$\mathbf{Z}'_{(2)ij}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{e}_i = \mathbf{X}'_{(2)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,.$$

Using (5.2.76), (5.2.77) can be written as

$$\mathbf{b}_{ij} = \mathbf{k}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{h}_i$$

which concludes the proof. □

PROPOSITION 5.2.12 (Subvector $\mathbf{q}_{(1)}$)

$$\begin{aligned}\mathbf{q}_{(1)} &= \mathbf{G}'_r \operatorname{vec}\Big(\sum_{l=1}^{n_i^*}\{\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{Y}_i\,\mathbf{V}_i^{-1}\,\mathbf{U}_{(1)il}\}\Big)\\ &= \mathbf{G}'_r \operatorname{vec}\sum_{l=1}^{n_i^*}(\mathbf{c}_{il}\,\mathbf{c}'_{il}) \qquad (5.2.78)\end{aligned}$$

with

$$\mathbf{c}'_{il} = \mathbf{x}_{(1)il}\,[\,[\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,]_{k,1} - \mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{h}_i\,] \qquad (5.2.79)$$

and $\mathbf{g}'_k$, $\mathbf{C}_i$ and $\mathbf{h}_i$ as given by (5.2.45), (5.2.27) and (5.2.73) respectively. The subscript l is as given in (5.2.5).

Proof

Let

$$\mathbf{c}_{il} = \mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{e}_i$$

with $\mathbf{e}_i$ given by (5.2.71). From (5.2.28) it follows that

$$\mathbf{U}'_{(1)il}\,\mathbf{V}_i^{-1}\,\mathbf{e}_i = \mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i - \mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i \tag{5.2.80}$$

where (cf. (5.2.51) and (5.2.70))

$$\mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{x}_{(1)ijk}\,\mathbf{g}'_k\,,$$

and

$$\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i = \mathbf{h}_i\ .$$

It also follows that (cf. (5.2.8) and (5.2.43))

$$\mathbf{U}'_{(1)il}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i = \mathbf{U}'_{(1)il}\begin{bmatrix}\mathbf{\Lambda}_{i1}^{-1}\,\mathbf{e}_{i1}\\ \mathbf{\Lambda}_{i2}^{-1}\,\mathbf{e}_{i2}\\ \cdot\\ \cdot\\ \cdot\\ \cdot\end{bmatrix}$$

$$= [\,\mathbf{0}'\ \ldots\ \mathbf{0}'\ \mathbf{V}'_{ijk}\ \mathbf{0}'\ldots]\begin{bmatrix}\mathbf{\Lambda}_{i1}^{-1}\,\mathbf{e}_{i1}\\ \mathbf{\Lambda}_{i2}^{-1}\,\mathbf{e}_{i2}\\ \cdot\\ \cdot\\ \cdot\\ \cdot\end{bmatrix}$$

$$= \mathbf{V}'_{ijk}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,.$$

From (5.2.44) it follows that

$$\mathbf{V}'_{ijk}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij} = \mathbf{x}_{(1)il}\,[\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,]_{k,1} \quad ,$$

so that (5.2.80) can be written as

$$\begin{aligned}\mathbf{U}'_{il}\,\mathbf{V}_i^{-1}\,\mathbf{e}_i &= \mathbf{x}_{(1)il}\,[\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,]_{k,1} - \mathbf{x}_{(1)il}\,\mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{h}_i \\ &= \mathbf{x}_{(1)il}\,[\,[\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,]_{k,1} - \mathbf{g}'_k\,\mathbf{C}_i\,\mathbf{h}_i\,]\end{aligned}$$

which concludes the proof. □

THEOREM 5.2.2

Let

$$2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{y}_i^* = \begin{bmatrix}\mathbf{q}_{(3)}\\ \mathbf{q}_{(2)}\\ \mathbf{q}_{(1)}\end{bmatrix}$$

where

$$\mathbf{q}_{(3)} = \mathbf{G}'_q\,\text{vec}\,(\mathbf{a}_i\,\mathbf{a}'_i\,) \quad ,$$

$$\mathbf{q}_{(2)} = \mathbf{G}'_m\,\text{vec}\,\sum_{j=1}^{n_i}(\mathbf{b}_{ij}\,\mathbf{b}'_{ij})$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}'_r\,\sum_{l=1}^{n_i^*}\text{vec}\,(\mathbf{c}_{il}\,\mathbf{c}'_{il}) \quad ,$$

with

$$\mathbf{a}_i = [\,\mathbf{I} - \mathbf{A}_i\,\mathbf{C}_i\,]\,\mathbf{h}_i \quad ,$$

$$\mathbf{b}_{ij} = [\,\mathbf{k}_{ij} - \mathbf{B}_{ij}\,\mathbf{C}_i\,\mathbf{h}_i\,]$$

and

$$\mathbf{c}_{il} = \mathbf{x}_{(1)il} \, [\, [\, \boldsymbol{\Lambda}_{ij}^{-1} \, \mathbf{e}_{ij} \,]_{k,1} - \mathbf{g}_k' \, \mathbf{C}_i \, \mathbf{h}_i \,] \quad ,$$

and where

$$\mathbf{h}_i = \mathbf{X}_{(3)i}' \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{e}_i \quad ,$$

$$\mathbf{B}_{ij} = \mathbf{X}_{(2)ij}' \, \boldsymbol{\Lambda}_{ij}^{-1} \, \mathbf{X}_{(3)ij} \quad ,$$

$$\mathbf{C}_i = (\boldsymbol{\Phi}_{(3)}^{-1} + \mathbf{A}_i)^{-1},$$

$$\mathbf{A}_i = \mathbf{X}_{(3)i} \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{X}_{(3)i} \quad ,$$

$$\mathbf{k}_{ij} = \mathbf{X}_{(2)ij}' \, \boldsymbol{\Lambda}_{ij}^{-1} \, \mathbf{e}_{ij}$$

and $\mathbf{g}_k'$ as defined in Proposition 5.2.6.

Proof

The proof follows directly from Propositions 5.2.10 to 5.2.12. □

5.2.3 SIMPLIFICATION OF THE WEIGHT MATRIX AND COEFFICIENT VECTOR (FIXED PART OF THE MODEL)

Simplification of (5.2.18) and (5.2.63) is not sufficient to ensure optimal use of computer storage and program execution time. To achieve this, the terms $\mathbf{X}_{(f)i}' \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(f)i}$ and $\mathbf{X}_{(f)i}' \, \mathbf{V}_i^{-1} \, \mathbf{y}_i$ (cf. (5.2.11)) must also be written in terms of computationally efficient equations. These two matrix expressions will now be considered briefly.

The first of these terms can be written as

$$\mathbf{X}_{(f)i}' \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(f)i} = \mathbf{X}_{(f)i}' \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{X}_{(f)i} - \mathbf{X}_{(f)i}' \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{X}_{(3)i} \, \mathbf{C}_i \, \mathbf{X}_{(3)i}' \, \boldsymbol{\Lambda}_i^{-1} \, \mathbf{X}_{(f)i} \tag{5.2.81}$$

where

$$X'_{(f)i}\,\Lambda_i^{-1}\,X_{(f)i} = \sum_{j=1}^{n_i} X'_{(f)ij}\,\Lambda_{ij}^{-1}\,X_{(f)ij}$$

and

$$X'_{(f)i}\,\Lambda_i^{-1}\,X_{(g)i} = \sum_{j=1}^{n_i} X'_{(f)ij}\,\Lambda_{ij}^{-1}\,X_{(g)ij}\ .$$

Let

$$L_i = \sum_{j=1}^{n_i} L_{ij} = \sum_{j=1}^{n_i} X'_{(f)ij}\,\Lambda_{ij}^{-1}\,X_{(f)ij} \qquad (5.2.82)$$

and

$$M_i = \sum_{j=1}^{n_i} M_{ij} = \sum_{j=1}^{n_i} X'_{(f)ij}\,\Lambda_{ij}^{-1}\,X_{(g)ij}\,. \qquad (5.2.83)$$

Then (5.2.81) can be rewritten as

$$X'_{(f)i}\,V_i^{-1}\,X_{(f)i} = L_i - M_i\,C_i\,M_i\,. \qquad (5.2.84)$$

Similarly, $X'_{(f)i}\,V_{(f)i}^{-1}\,y_i$ can be written as

$$\begin{aligned} X'_{(f)i}\,V_i^{-1}\,y_i &= X'_{(f)i}\,\Lambda_i^{-1}\,y_i - X'_{(f)i}\,\Lambda_i^{-1}\,X_{(g)i}\,C_i\,X'_{(g)i}\,\Lambda_i^{-1}\,y_i \\ &= X'_{(f)i}\,\Lambda_i^{-1}\,y_i - M_i\,C_i\,X_{(g)i}\,\Lambda_i^{-1}\,y_i \qquad (5.2.85) \end{aligned}$$

where

$$X'_{(f)i}\,\Lambda_i^{-1}\,y_i = \sum_{j=1}^{n_i} X'_{(f)ij}\,\Lambda_{ij}^{-1}\,y_{ij}$$

and

$$X'_{(g)i}\,\Lambda_i^{-1}\,y_i = \sum_{j=1}^{n_i} X'_{(g)ij}\,\Lambda_{ij}^{-1}\,y_{ij}\ .$$

Let

$$\mathbf{n}_{ij} = \mathbf{X}'_{(f)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \qquad (5.2.86)$$

and

$$\mathbf{p}_{ij} = \mathbf{X}'_{(3)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij}\,. \qquad (5.2.87)$$

Using (5.2.85) and (5.2.86), expression (5.2.84) can be rewritten as

$$\mathbf{X}'_{(f)i}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i = \sum_{j=1}^{n_i}\mathbf{n}_{ij} - \mathbf{M}_i\,\mathbf{C}_i\sum_{j=1}^{n_i}\mathbf{p}_{ij} \qquad (5.2.88)$$

which concludes the derivation of terms needed to calculate $\hat{\boldsymbol{\beta}}$.

5.3 SPECIAL CASES OF THE GENERAL LEVEL-3 MODEL

In this section four special cases of the general level-3 model defined in the previous sections of this chapter will be discussed. The absence of random components on different levels of the model and the implication thereof on the weight matrix and random coefficient vector will be investigated.

Case 1 : Random components at level-2 and level-3 only

With no random component on level-1, the model (cf. (5.2.1)) can be written as

$$\mathbf{y}_i = \mathbf{X}_{(f)i}\,\boldsymbol{\beta} + \mathbf{X}_{(3)i}\,\mathbf{v}_i + \sum_{j=1}^{n_i}\mathbf{Z}_{(2)ij}\,\mathbf{u}_{ij}\,,$$

where

$$\mathrm{Cov}(\mathbf{y}_i\,,\,\mathbf{y}'_i) = \boldsymbol{\Sigma}_i = \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i\;,$$

and where

$$\Lambda_i = \begin{bmatrix} \mathbf{X}_{(2)i1}\,\Phi_{(2)}\,\mathbf{X}'_{(2)i1} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & . & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \mathbf{X}_{(2)in_i}\,\Phi_{(2)}\,\mathbf{X}'_{(2)in_i} \end{bmatrix} .$$

In this case the matrix $\mathbf{T}'_i = 2\,\mathbf{X}^{*\prime}_i\,\mathbf{W}_i^{-1}\,\mathbf{X}^*_i$ (cf. (5.2.18)) reduces to

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} \\ T_{(2,3)} & \mathbf{T}_{(2,2)} \end{bmatrix}$$

with the expressions for the submatrices $\mathbf{T}_{(3,3)}$, $\mathbf{T}_{(2,3)}$ and $\mathbf{T}_{(2,2)}$ as given in Theorem 5.2.1, Section 5.2 and

$$\Lambda_{ij}^{-1} = (\mathbf{X}_{(2)ij}\,\Phi_{(2)}\,\mathbf{X}'_{(2)ij}\,)^{-1} .$$

The vector $2\,\mathbf{X}^{*\prime}_i\,\mathbf{W}_i^{-1}\,\mathbf{y}^*_i$ is given by (cf. (5.2.63))

$$\mathbf{q}_i = \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \end{bmatrix}$$

with expressions for vectors $\mathbf{q}_{(3)}$ and $\mathbf{q}_{(2)}$ as given in Theorem 5.2.2.

Case 2 : Random components at level-1 and level-3 only

With no random component on level-2, the model (cf. (5.2.1)) becomes

$$\mathbf{y}_i = \mathbf{X}_{(f)i}\,\beta + \mathbf{X}_{(3)i}\,\mathbf{v}_i + \sum_{j=1}^{n_i}\ \sum_{k=1}^{n_{ij}} \mathbf{U}_{(1)ijk}\,\mathbf{e}_{ijk}$$

where

$$\text{Cov}(\mathbf{y}_i\,,\ \mathbf{y}_i') = \mathbf{V}_i\ = \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i\,.$$

In this case

$$\boldsymbol{\Lambda}_i\ = \begin{bmatrix} \mathbf{D}_{i1} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{i2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{D}_{in_i} \end{bmatrix},$$

where

$$\mathbf{D}_{ij}\ = \begin{bmatrix} \lambda_{ij1} & 0 & . & . & 0 \\ 0 & \lambda_{ij2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & 0 & \lambda_{ijn_i} \end{bmatrix}$$

and

$$\boldsymbol{\lambda}_{ijk} = \mathbf{x}'_{(1)ijk}\,\boldsymbol{\Phi}_{(1)}\,\mathbf{x}_{(1)ijk}\,.$$

The weight matrix $\mathbf{T}_i = 2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{X}_i^*$ (cf. (5.2.18)) reduces to

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(1,3)} \\ \mathbf{T}_{(1,3)} & \mathbf{T}_{(1,1)} \end{bmatrix}$$

with expressions for the submatrices $\mathbf{T}_{(3,3)}$, $\mathbf{T}_{(1,3)}$ and $\mathbf{T}_{(1,1)}$ as given in Theorem 5.2.1, Section 5.2 and $\boldsymbol{\Lambda}_{ij}^{-1} = \mathbf{D}_{ij}^{-1}$ where $\mathbf{D}_{ij}$ is given above.

The coefficient vector $\mathbf{q}_i = 2\ \mathbf{X}_i^{*\prime}\ \mathbf{W}_i^{-1}\ \mathbf{y}_i^*$ is given by

$$\mathbf{q}_i = \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(1)} \end{bmatrix}$$

with $\mathbf{q}_{(3)}$ and $\mathbf{q}_{(2)}$ as given in Theorem 5.2.2.

Case 3 : Random components at level-1 and level-2 only (level-2 model)

When no level-3 random component is specified, one of the following situations apply :
(i) No level-3 information is available or
(ii) Information on level-3 units is available but it is assumed that these units are homogeneous with respect to the characteristic being studied.

Thus, instead of considering the vectors of responses $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N$, information on all the level-2 units is summarized in a single vector $\mathbf{y}_1$. Therefore, if $N = 1$, a level-2 model is obtained as a special case of the level-3 model.

The model (cf. (5.2.1)) becomes

$$\mathbf{y}_1 = \mathbf{X}_{(f)1}\,\beta + \sum_{j=1}^{n_1} \mathbf{Z}_{(2)1j}\,\mathbf{u}_{1j} + \sum_{j=1}^{n_1}\ \sum_{k=1}^{n_{1j}} \mathbf{U}_{(1)1jk}\,\mathbf{e}_{1jk}$$

where

$$\mathrm{Cov}(\mathbf{y}_1\,,\mathbf{y}_1') = \boldsymbol{\Lambda}_1 = \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{12} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{0} & . & . & \mathbf{0} & \boldsymbol{\Lambda}_{1n_i} \end{bmatrix}$$

and

$$\boldsymbol{\Lambda}_{1j} = \mathbf{X}_{(2)1j}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{X}'_{(2)1j} + \mathbf{D}_{1j}\,,$$

where

$$\mathbf{D}_{ij} = \begin{bmatrix} \lambda_{1j1} & 0 & . & . & 0 \\ 0 & \lambda_{1j2} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \lambda_{1jn_i} \end{bmatrix}$$

and

$$\lambda_{1jk} = \mathbf{x}_{(1)1jk}\,\boldsymbol{\Phi}_{(1)}\,\mathbf{x}'_{(1)1jk}\,.$$

In this case the matrix $\mathbf{T}_i = 2\,\mathbf{X}_1^{*\prime}\,\mathbf{W}_1^{-1}\,\mathbf{X}_1^*$ (cf. (5.2.18)) is given by

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{T}_{(2,2)} & \mathbf{T}'_{(1,2)} \\ \mathbf{T}_{(1,2)} & \mathbf{T}_{(1,1)} \end{bmatrix}$$

where (cf. (5.2.21), (5.2.23) and (5.2.24))

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m\,(\sum_{j^*=1}^{n_1}\sum_{j=1}^{n_1}\mathbf{Z}'_{(2)1j}\,\boldsymbol{\Lambda}_1^{-1}\,\mathbf{Z}_{(2)1j^*}\otimes\mathbf{Z}'_{(2)1j}\,\boldsymbol{\Lambda}_1^{-1}\,\mathbf{Z}_{(2)1j^*})\,\mathbf{G}_m\,,$$

$$\mathbf{T}_{(1,2)} = \mathbf{G}'_r\,(\sum_{l=1}^{n_1^*}\sum_{j=1}^{n_1}\mathbf{U}'_{(1)1l}\,\boldsymbol{\Lambda}_1^{-1}\,\mathbf{Z}_{(2)1j}\otimes\mathbf{U}'_{(1)1l}\,\boldsymbol{\Lambda}_1^{-1}\,\mathbf{Z}_{(2)1j})\,\mathbf{G}_m$$

and

$$\mathbf{T}_{(1,1)} = \mathbf{G}'_r\,(\sum_{l=1}^{n_1^*}\sum_{l^*=1}^{n_1^*}\mathbf{U}'_{(1)1l}\,\boldsymbol{\Lambda}_1^{-1}\,\mathbf{U}_{(1)1l^*}\otimes\mathbf{U}'_{(1)1l}\,\boldsymbol{\Lambda}_1^{-1}\,\mathbf{U}_{(1)1l^*})\,\mathbf{G}_r\,.$$

From Propositions 5.2.3, 5.2.8 and 5.2.9 it follows that the submatrices $\mathbf{T}_{(2,2)}$, $\mathbf{T}_{(2,1)}$ and $\mathbf{T}_{(1,1)}$ reduce to

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m\,(\sum_{j=1}^{n_1}\sum_{j^*=1}^{n_1}(\delta_{jj^*})\,\mathbf{E}_{1j}\otimes(\delta_{jj^*})\,\mathbf{E}_{1j})\,\mathbf{G}_m\;,$$

$$\mathbf{T}_{(1,2)} = \mathbf{G}_r' \left(\sum_{j=1}^{n_1} \sum_{j^*=1}^{n_1} \sum_{k=1}^{n_{1j}} \{ \mathbf{x}_{(1)1jk} \mathbf{f}_k' \otimes \mathbf{x}_{(1)1jk} \mathbf{f}_k' \} \right) \mathbf{G}_m$$

and

$$\mathbf{T}_{(1,1)} = \mathbf{G}_r' \left(\sum_{j=1}^{n_1} \sum_{j^*=1}^{n_1} \sum_{k=1}^{n_{1j}} \sum_{k^*=1}^{n_{1j}} \{ \delta_{jj^*} [\boldsymbol{\Lambda}_{1j}^{-1}]_{kk^*} \mathbf{x}_{(1)1jk} \mathbf{x}_{(1)1j^*k^*}' \otimes \right.$$

$$\left. \delta_{jj^*} [\boldsymbol{\Lambda}_{1j}^{-1}]_{kk^*} \mathbf{x}_{(1)1jk} \mathbf{x}_{(1)1j^*k^*} \} \right) \mathbf{G}_r \, .$$

The vector $\mathbf{q}_1 = 2\, \mathbf{X}_1^{*\prime} \mathbf{W}_1^{-1} \mathbf{y}_1^*$ (cf. (5.2.63)) is given by

$$\mathbf{q}_1 = \begin{bmatrix} \mathbf{q}_{(2)} \\ \mathbf{q}_{(1)} \end{bmatrix}$$

where (cf. (5.2.66) and (5.2.67))

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \operatorname{vec} \left(\sum_{j=1}^{n_1} \mathbf{Z}_{(2)1j}' \boldsymbol{\Lambda}_1^{-1} \mathbf{Y}_1 \; \boldsymbol{\Lambda}_1^{-1} \mathbf{Z}_{(2)1j} \right)$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}_r' \operatorname{vec} \left(\sum_{j=1}^{n_1} \sum_{k=1}^{n_{1j}} \mathbf{U}_{(1)1jk}' \boldsymbol{\Lambda}_1^{-1} \mathbf{Y}_1 \, \boldsymbol{\Lambda}_1^{-1} \mathbf{U}_{(1)1jk} \right) .$$

From Propositions 5.2.11 and 5.2.12 it follows that $\mathbf{q}_{(2)}$ and $\mathbf{q}_{(1)}$ can be rewritten as

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \sum_{j=1}^{n_i} \operatorname{vec} (\mathbf{b}_{1j} \mathbf{b}_{1j}')$$

and

$$\mathbf{q}_{(1)} = \mathbf{G}_r' \sum_{l=1}^{n_i^*} \operatorname{vec} (\mathbf{c}_{1l} \mathbf{c}_{1l}')$$

with (cf. (5.2.75))

$$\mathbf{b}_1 = \mathbf{k}_{1j}$$

and (cf. (5.2.79))

$$\mathbf{c}_{1l} = \mathbf{x}_{(1)1l} [\, \boldsymbol{\Lambda}_{1j}^{-1} \, \mathbf{e}_{1j} \,]_{k,1} \,.$$

where l denotes the double subscript jk, $j = 1, 2, \ldots, n_1$ and $k = 1, 2, \ldots, n_{1j}$.

Case 4 : Random component at level-1 only (level-1 model)

The level-1 model is another special case of the level-3 model and is obtained by assuming that there is only one level-3 unit ($N = 1$) and one level-2 unit ($n_1 = 1$) . The model (cf. (5.2.1)) is given by

$$\mathbf{y}_1 = \mathbf{X}_{(f)1} \,\boldsymbol{\beta} + \sum_{k=1}^{n_{11}} \mathbf{U}_{(1)11k} \, \mathbf{e}_{11k}$$

where

$$\mathrm{Cov}(\mathbf{y}_1 \,,\, \mathbf{y}_i') = \mathbf{D}_{11} \; = \begin{bmatrix} \lambda_{111} & . & . & . & 0 \\ 0 & \lambda_{112} & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & \lambda_{11n_{11}} \end{bmatrix}$$

with

$$\lambda_{11k} = \mathbf{x}'_{(1)11k} \, \boldsymbol{\Phi}_{(1)} \, \mathbf{x}_{(1)11k} \,.$$

The matrix $\mathbf{T}_1 = 2\, \mathbf{X}_1^{*\prime} \, \mathbf{W}_1^{-1} \, \mathbf{X}_1^{*}$ (cf. (5.2.18)) is then

$$\mathbf{T}_1 = [\, \mathbf{T}_{(1,1)} \,]$$

where (cf. Proposition 5.2.9)

$$\mathbf{T}_{(1,1)} = \mathbf{G}_r' \, (\sum_{k=1}^{n_{11}} \sum_{k^*=1}^{n_{11}} \mathbf{U}_{(1)11k} \, \mathbf{D}_{11}' \, \mathbf{U}_{(1)11k^*} \otimes \mathbf{U}_{(1)11k} \, \mathbf{D}_{11}^{-1} \, \mathbf{U}_{(1)11k^*} \,) \, \mathbf{G}_r$$

$$= \mathbf{G}_r' \, (\sum_{k=1}^{n_{11}} \sum_{k^*=1}^{n_{11}} \lambda_{11k}^{-1} \, \mathbf{x}_{(1)11k} \mathbf{x}'_{(1)11k^*} \otimes \lambda_{11k}^{-1} \, \mathbf{x}_{(1)11k} \mathbf{x}'_{(1)11k^*} \,) \, \mathbf{G}_r$$

From (5.2.63) it follows that $\mathbf{q}_1 = 2\,\mathbf{X}_i^{*\prime}\,\mathbf{W}_i^{-1}\,\mathbf{y}_1^*$ is given by

$$\mathbf{q}_1 = [\,\mathbf{q}_{(1)}\,]$$

where (cf. (5.2.67))

$$\mathbf{q}_{(1)} = \mathbf{G}_r'\,\text{vec}\,\Big(\sum_{k=1}^{n_{11}} \mathbf{U}_{(1)11k}'\,\mathbf{D}_{11}^{-1}\,\mathbf{Y}_1\,\mathbf{D}_{11}^{-1}\,\mathbf{U}_{(1)11k}\Big)\,.$$

From Proposition 5.2.12 it follows that $\mathbf{q}_1$ can be written as

$$\mathbf{q}_{(1)} = \mathbf{G}_r' \sum_{l=1}^{n_i^*} \text{vec}\,(\mathbf{c}_{1l}\,\mathbf{c}_{1l}')$$

where (cf. (5.2.79))

$$\mathbf{c}_{1l} = \lambda_{11k}^{-1}\,\mathbf{x}_{(1)11k}\,\mathbf{e}_{11}\,.$$

5.4 PRACTICAL APPLICATIONS

Three practical applications are given in this section. All applications are based on the same data set, which is described below.

During 1993, the Transvaal Provincial Administration (TPA) conducted a survey in anticipation of majority rule, using their employees as respondents. The questions asked dealt mainly with perceptions regarding the implementation of affirmative action programs, discrimination in the workplace and criteria to be used for the selection, training and promotion of employees. Respondents were required to indicate their responses to statements relating to the above-mentioned issues on a 5-point scale, where a score of '1' indicated strong disagreement, '2' disagreement, '3' do not know, '4' agreement and '5' strong agreement with the particular statement. Biographical variables included gender, age, qualification, home language and job level within the organization.

In the case of questions with a 5-point scale response, the 5-point scale was recoded so as to consist of three categories, that is 'Negative', 'Do not know' and 'Positive'.

Optimal scaling (du Toit & Strasheim, 1987), was used to assign numerical values to the three categories. A factor analysis with VARIMAX rotation (SAS/STAT, 1990), was done with respect to the scaled items in order to find subgroups of mutually correlated items. Total scores were then calculated for the different subgroups. In this example, the total scores pertaining to the criteria for affirmative-action recruiting in the work place and to discrimination are used as dependent and predictor variables respectively. The items and scaled values used for the construction of these two variables, which will be referred to as FACTOR1 and FACTOR3 respectively, are given in Tables 5.4.1 and 5.4.2.

Table 5.4.1 : FACTOR1 - Affirmative-action recruitment criteria

Description of item	**Scaled values**		
	Disagree	**Do not know**	**Agree**
Priority should be given to the recruitment of staff from groups that are currently under-represented in the TPA.	3	1.6176	0
Recruitment advertisements should be formulated in such a way that all groups will have an equal opportunity to apply.	3	1.6535	0
During the selection process, preference should be given to applicants from under-represented groups.	3	1.5933	0

In order to obtain a more representative work force,			
appointments should be made additional to the establishment, and	3	1.6017	0
offers should be made to worthy persons outside the TPA.	3	1.7926	0
Unions and Staff Associations should be involved in the implementation of the Equal Job-opportunities Program.	3	1.6406	0
How do you feel about the Equal Job-opportunity Program?	3	1.6775	0

Alpha coefficient (Kuder-Richardson's K20 coefficient of reliability, Huynh (1986))	0.6689
Mean of Total	7.5010
Standard deviation	5.0023
Minimum value	0
Maximum value	21

A total score close to 21 would thus indicate strong agreement with the statements as given in Table 5.4.1, while a total score close to 0 would indicate strong disagreement with these statements.

Table 5.4.2: FACTOR3 - Discrimination in the TPA

Description of item	Scaled values		
	Disagree	Do not know	Agree
In the section where you work, is there discrimination against staff on the following grounds:			
Religion	0.9710	3	0
Gender	0.9553	3	0
Race	0.7287	3	0
Age	1.0764	3	0

Alpha coefficient (Kuder Richardson's K20 coefficient of reliability)	0.6528
Mean of total	3.6841
Standard deviation	2.1718
Minimum value	0
Maximum value	12

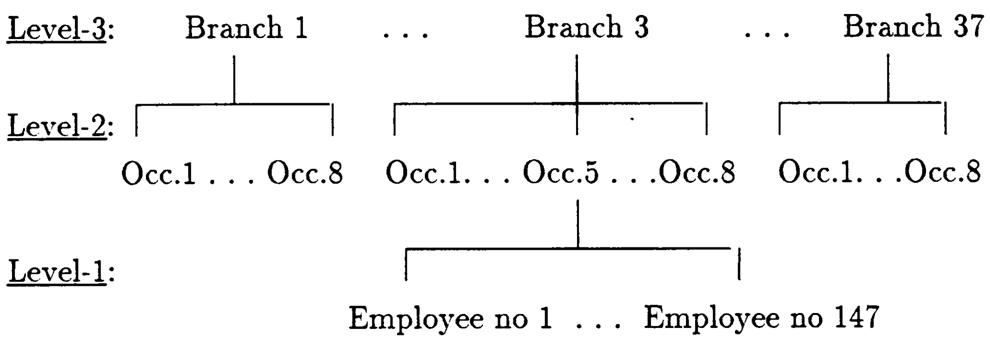
The maximum value that can be obtained for FACTOR3 is 12, indicating no opinion with regard to the presence of discrimination. A score close to 0 indicates agreement with the statements as given in Table 5.4.2. A respondent disagreeing with all four statements will obtain a score of 3.7314.

The TPA can be divided into five units, namely Road Construction and Maintenance, Community Development, Financial Management and Operations, General Provincial Services and Health Services. Each of these units can be subdivided into branches. In these examples, the data for the Health Services unit are used. The sample from this unit consists of 37 branches with a total number of 5207 employees. The number of employees per branch and in each of the eight job categories differs dramatically, as can be seen from an excerpt of a frequency table for occupation by branch.

Table 5.4.3: Excerpt from frequency table for occupation category by branch number

Occupation		Branch	
Description	**Category**	**3**	**51**
Medical (eg doctor)	1	4	4
Supplementary health services (eg radiologist)	2	16	0
Nursing (eg matron)	3	267	6
General Assistant	4	2	0
Administrative	5	147	0
Technical (eg electrician)	6	0	0
Professional (eg engineer)	7	15	1
Other	8	7	0
Total		458	11

The data can be schematically represented as follows:



The 37 branches represent the level-3 units while the eight occupation categories represents the level-2 units. Note that in the case of branch 3, given in the schematical representation above, there are employees in all eight occupation categories. In the case of branch 51 (see Table 5.4.3) there are employees in categories 1, 3 and 7 only, indicating the presence of only 3 level-2 units for this branch. The level-1 units are the total scores for FACTOR1 for each of the employees in a particular level-2 unit.

To investigate the structural relationships between variables, several models were fitted to the data. In each case a data set was created. Three of the data sets and the models fitted will now be considered.

Example 5.4.1

In this example a model without random coefficients at level-3 and with a simple level-1 variance structure is considered. This type of model is commonly referred to as a level-2 model, the level-2 units in this case being the (branch 1, occupation 1), (branch 1, occupation 3) . . . (branch 37, occupation 8) subgroups. There are 233 level-2 units and 5 207 level-1 units.

Consider the first 10 rows of the following data set:

ID3	ID2	ID1	F.1	CONS	S1	Q1	Q2	L1	L2	F.3	CONS	Q1	Q2	G1	G2	G3	G4	L1	L2	F.3
2	3	1	0.00	1	-1	-1	-1	-1	-1	3.97	1	-1	-1	-1	0	0	1	-1	-1	3.97
2	3	1	6.00	1	1	0	1	1	0	3.73	1	0	1	1	1	0	0	1	0	3.73
2	5	1	4.62	1	1	0	1	1	0	0.97	1	0	1	1	1	0	0	1	0	0.97
3	5	1	0.00	1	-1	-1	-1	-1	-1	3.97	1	-1	-1	-1	0	0	1	-1	-1	3.97
3	5	1	1.79	1	1	0	1	-1	-1	3.00	1	0	1	1	1	0	0	-1	-1	3.00
3	5	1	6.00	1	-1	0	1	0	1	6.97	1	0	1	1	0	0	1	0	1	6.97
3	5	1	9.00	1	1	0	1	-1	-1	3.73	1	0	1	1	1	0	0	-1	-1	3.73
3	5	1	6.27	1	-1	0	1	1	0	3.73	1	0	1	1	0	0	1	1	0	3.73
3	5	1	6.45	1	1	0	1	-1	-1	3.73	1	0	1	1	1	0	0	-1	-1	3.73
3	5	1	0.00	1	1	0	1	-1	-1	12.00	1	0	1	1	1	0	0	-1	-1	12.00
Identification variables			(y)	Predictors for fixed part of the model $\mathbf{X}_{(f)}$							Predictors for random part of the model : Level-1			Level-2				Level-3		

The combination of the variables ID3 and ID2 denote the branch-occupation subgroups. The first row, for instance, belongs to the subgroup (branch 2, occupation 3) while the third row belongs to the subgroup (branch 2, occupation 5). FACTOR1 is the dependent or response variable.

The predictors for the fixed part of the model are the variables CONS, S1, Q1, L1, L2 and FACTOR3. The inclusion of an intercept term (CONS) allows for the estimation of a general constant effect, given that the values of the remaining coefficients are 0. The gender of a respondent is given by the variable S1, where a value of '1' indicates a male respondent, and a value of '-1' a female respondent. The coding used to create the dummy variables Q1 and Q2, pertaining to the qualification of a respondent, is given below.

Qualification	**Value of Q1**	**Value of Q2**
Lower than Std 10	1	0
Std 10 plus 1-2 years of tertiary education	0	1
Std 10 plus 3 years or more tertiary education	-1	-1

The coding for the dummy variables indicating a respondent's home language, that is L1 and L2, is done in a similar way. In this case the variable L1 assumes a value of '1' if the respondent is Afrikaans-speaking, '0' if English-speaking and '-1' in all other cases. The coefficient of each dummy variable, that is S1, Q1, Q2, L1, and L2 is a measure of the additional effect, over and above the effect attributed to the intercept term. Together with the total score for perception of discrimination given by FACTOR3, these variables form the columns of the fixed parameter design matrix $\mathbf{X}_{(f)}$ (cf. (5.2.1)).

On the first level of the hierarchy, only the coefficient of the variable CONS is allowed to vary randomly across respondents. The coefficients of the variables G1 to G4, indicating the gender and type of appointment held by an employee, are allowed to be random on level-2. Coding for these variables is as follows:

Gender	**Type of appointment**	**Value of**			
		G1	**G2**	**G3**	**G4**
Male	Permanent	1	0	0	0
Male	Temporary	0	1	0	0
Female	Permanent	0	0	1	0
Female	Temporary	0	0	0	1

The four columns representing the variables G1, G2, G3, and G4 are the columns of the random parameter design matrix $\mathbf{X}_{(2)}$. The covariance of $\mathbf{y}_{ij}$ is thus given by

$$\text{Cov}(\mathbf{y}_{ij}\,,\,\mathbf{y}'_{ij}) = \mathbf{X}_{(2)ij}\,\mathbf{\Phi}_{(2)}\,\mathbf{X}'_{(2)ij} + \Phi_{(1)}\,\mathbf{I}_{n_{ij}}\,.$$

Since the gender by type of appointment groups form 4 independent subpopulations, no allowance is made for association between the coefficients of the associated dummy variables. Hence, the corresponding elements of $\mathbf{\Phi}_{(2)}$ are constrained to be 0. See Section 7.5, Chapter 7 for more information on constraints.

The following computer output is obtained:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	**STD.ERR.**	**Z-VALUE**	**PR>|Z|**
CONS	6.7920	0.1646	41.2637	0.0000
S1	-0.1969	0.1119	-1.7596	0.0785
Q1	-0.3098	0.0996	-3.1104	0.0019
Q2	0.0392	0.1024	0.3828	0.7019

L1	1.7580	0.1002	17.5449	0.0000
L2	0.5474	0.1502	3.6445	0.0027
FACTOR3	0.1796	0.0287	6.2578	0.0000

(ii) **Random part of the model:**

Level-2:

PARAMETER	$\hat{\tau}$	**STD.ERR**	**Z-VALUE**	**PR>|Z|**
G1/G1	2.420	0.7322	3.3051	0.0009
G2/G2	3.645	1.5200	2.3980	0.0165
G3/G3	1.424	0.3271	4.3534	0.0000
G4/G4	1.442	0.5921	2.4354	0.0149

Level-1:

CONS/CONS	19.110	0.3889	49.1386	0.0000

(iii) **Likelihood function:**

$$-\ 2\ ln\ \mathrm{L} = 30397.1$$

From Part (i) of the output it can be seen that the estimate of the intercept is 6.7920. This estimate may be interpreted as the expected value of the response variable FACTOR1, assuming no gender, qualification, language or FACTOR3 effect. The coefficient for gender, given by S1, is significant at a 10 % level of significance and indicates a total gender effect of 0.3938. In the case of the male employees (coded '1'), this gives a negative effect of -0.1969 and for female employees (coded '-1') a positive effect of 0.1969. This indicates that female employees generally obtained a slightly higher score than male employees.

The expected value of the response variable for different levels of education, given that there is no gender, language or FACTOR3 effect, can be written as

$$\hat{y} = 6.7920\ -\ 0.3098\ \mathrm{Q1} + 0.0392\ \mathrm{Q2}\ .$$

The expected FACTOR1 score for an employee with an educational level lower than Standard 10 is then obtained as

$$\hat{y}_{<10} = 6.7920 - 0.3098 = 6.4822.$$

The expected FACTOR1 score for the employees with Standard 10 and up to two years of tertiary education is 6.8312, while the expected score for employees with 3 years or more of tertiary education is 7.0626. There is a positive relationship between employees' level of education and expected FACTOR1 scores.

In the case of home language, the expected value of FACTOR1, given no effect bar that of home language, can be expressed as

$$\hat{y} = 6.7920 + 1.7580 \text{ L1} + 0.5474 \text{ L2}$$

which leads to the following expected scores: 8.5500 for Afrikaans-speaking employees, 7.3394 for English-speaking employees and 4.4866 for employees with other home languages. The Afrikaans-speaking group thus has, on average, a higher expected FACTOR1 score, indicating a stronger tendency to agree with the statements given in Table 5.4.1. The third language group, which includes Black languages, appears to disagree with the suggested criteria for affirmative action recruitment.

Finally, the estimate of 0.1796 for the coefficient of FACTOR3 indicates an increase of 0.1796 (1.5 %) in the expected value of FACTOR1 for each unit increase in the FACTOR3 score.

From Part (ii) of the output it follows that all the estimates are significant at a 5 % level of significance. The largest coefficient, that is 3.645, is obtained for G2/G2, indicating that the largest variation on level-2 of the hierarchy is for the temporary-employed males subgroup. The variation over employees, that is on level-1, is quite high and highly significant. This variation indicates a large amount of variation in the data still unexplained by this model.

Example 5.4.2

In this example the model discussed in Example 5.4.1 is extended to include a random component on a third level of the hierarchy. The dummy variables for language, that is L1 and L2, and the FACTOR3 score are allowed to be random at level-3, that is on the branch level. The first 10 rows of the data set are given below:

ID3	ID2	ID1	F.1	CONS	S1	Q1	Q2	L1	L2	F.3	CONS	G1	G2	G3	G4	L1	L2	F.3
2	3	1	0.00	1	-1	-1	-1	-1	-1	3.97	1	-1	0	0	1	-1	-1	3.97
2	3	1	6.00	1	1	0	1	1	0	3.73	1	1	1	0	0	1	0	3.73
2	5	1	4.62	1	1	0	1	1	0	0.97	1	1	1	0	0	1	0	0.97
3	5	1	0.00	1	-1	-1	-1	-1	-1	3.97	1	-1	0	0	1	-1	-1	3.97
3	5	1	1.79	1	1	0	1	-1	-1	3.00	1	1	1	0	0	-1	-1	3.00
3	5	1	6.00	1	-1	0	1	0	1	6.97	1	1	0	0	1	0	1	6.97
3	5	1	9.00	1	1	0	1	-1	-1	3.73	1	1	1	0	0	-1	-1	3.73
3	5	1	6.27	1	-1	0	1	1	0	3.73	1	1	0	0	1	1	0	3.73
3	5	1	6.45	1	1	0	1	-1	-1	3.73	1	1	1	0	0	-1	-1	3.73
3	5	1	0.00	1	1	0	1	-1	-1	12.00	1	1	1	0	0	-1	-1	12.00
Identification variables			(y)	Predictors for fixed part of the model $\mathbf{X}_{(f)}$							Predictors for random part of model							
											L-1	L-2				L-3		

In all cases the coding used for the creation of dummy variables is identical to that discussed in Example 5.4.1. The last three columns of the data as given above, form the random parameter design matrix $\mathbf{X}_{(3)}$ (cf. (5.2.3)) and the covariance of $\mathbf{y}_{ij}$ is

$$\text{Cov}(\mathbf{y}_{ij},\mathbf{y}'_{ij}) = \mathbf{X}_{(3)ij}\,\mathbf{\Phi}_{(3)}\mathbf{X}'_{(3)ij} + \mathbf{X}_{(2)ij}\,\mathbf{\Phi}_{(2)}\mathbf{X}'_{(2)ij} + \mathbf{\Phi}_{(1)}\mathbf{I}_{n_{ij}}\,.$$

The existence of two or more random coefficients at the same level is referred to as complex variation. In this model complex variation exists on level-2 and level-3 of the model.

Since the different language groups form independent subpopulations, the corresponding elements of $\mathbf{\Phi}_{(3)}$ were constrained to be 0. Thus, no allowance was made for an association between the coefficients of the dummy variables L1 and L2.

The computer output obtained for this model is as follows:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	STD.ERR.	Z-VALUE	PR>|Z|
CONS	6.6160	0.1604	63.3716	0.0000
S1	-0.1686	0.1044	-1.6149	0.1063
Q1	-0.3810	0.1002	-3.8024	0.0001
Q2	0.1101	0.1012	1.0879	0.2766
L1	1.7060	0.1588	10.7431	0.0000
L2	0.5535	0.1479	3.7424	0.0002
FACTOR3	0.2055	0.0374	5.4947	0.0000

(ii) **Random part of the model:**

Level-3:

PARAMETER	$\hat{\tau}$	STD.ERR.	Z-VALUE	PR>|Z|
L1/L1	0.4826	0.1767	2.7312	0.0063
L2/L2	0.0190	0.1110	0.1712	0.8641
FACTOR3/L1	0.0156	0.0264	0.5909	0.5546
FACTOR3/L2	-0.0339	0.0223	1.5202	0.1285
FACTOR3/FACTOR3	0.0164	0.0071	2.3099	0.0209

Level-2:

G1/G1	1.7880	0.6421	2.7846	0.0054
G2/G2	2.5560	1.2570	2.0334	0.0420
G3/G3	0.9640	0.2710	3.5572	0.0004
G4/G4	1.0900	0.5348	2.0381	0.0415

Level-1:

CONS/CONS	18.7800	0.3822	49.0084	0.0000

(iii) **Likelihood function:**

$- 2 \, ln \, L = 30325.5$

As far as the fixed part of the model is concerned, it can be seen that the estimates, their significance and interpretation are much the same as described in the previous example. The intercept term and the dummy variable L1 have the largest z-values in both models.

From the output for the random part of the model, it follows that only the coefficients for L1/L1 and FACTOR3/FACTOR3 are significant at a 5 % level of significance. The largest variation on level-3 is for the subgroup of Afrikaans-speaking employees, denoted by L1.

When the random coefficients on level-2 of this model are compared with those obtained in the previous example, it is seen that the inclusion of a component on level-3 had caused an overall drop in variation on level-2, that is over occupation categories. The largest variation on level-2 is still the variation of the temporary-employed males subgroup. All coefficients on level-2 are significant at a 5 % level of significance.

From the intercept coefficient on level-1, it can be deduced that there is still a large amount of variation unexplained by this model.

Example 5.4.3

From the results of the previous example, it can be seen that the fitting of a level-3 model with complex variance structures on level-3 and level-2 of the model still left a large amount of the variation in the data unexplained. We now extend the model by allowing the coefficients of the dummy variables Q1 and Q2 to vary randomly over respondents, thus obtaining a level-3 model with different random components on all levels of the hierarchy.

The first 10 rows of the data set are given below.

ID3	ID2	ID1	F.1	CONS	S1	Q1	Q2	L1	L2	F.3	CONS	Q1	Q2	G1	G2	G3	G4	L1	L2	F.3
2	3	1	0.00	1	-1	-1	-1	-1	-1	3.97	1	-1	-1	-1	0	0	1	-1	-1	3.97
2	3	1	6.00	1	1	0	1	1	0	3.73	1	0	1	1	1	0	0	1	0	3.73
2	5	1	4.62	1	1	0	1	1	0	0.97	1	0	1	1	1	0	0	1	0	0.97
3	5	1	0.00	1	-1	-1	-1	-1	-1	3.97	1	-1	-1	-1	0	0	1	-1	-1	3.97
3	5	1	1.79	1	1	0	1	-1	-1	3.00	1	0	1	1	1	0	0	-1	-1	3.00
3	5	1	6.00	1	-1	0	1	0	1	6.97	1	0	1	1	0	0	1	0	1	6.97
3	5	1	9.00	1	1	0	1	-1	-1	3.73	1	0	1	1	1	0	0	-1	-1	3.73
3	5	1	6.27	1	-1	0	1	1	0	3.73	1	0	1	1	0	0	1	1	0	3.73
3	5	1	6.45	1	1	0	1	-1	-1	3.73	1	0	1	1	1	0	0	-1	-1	3.73
3	5	1	0.00	1	1	0	1	-1	-1	12.00	1	0	1	1	1	0	0	-1	-1	12.00
Identification variables			(y)	Predictors for fixed part of the model $\mathbf{X}_{(f)}$							Predictors for random part of the model : Level-1			Level-2				Level-3		

The coding of the dummy variables is the same as that discussed in Example 5.4.1. The columns representing the predictors for the random part of the model, as given above, form the random parameter design matrices $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$ and $\mathbf{X}_{(3)}$. The covariance of $\mathbf{y}_{ij}$ can be expressed as:

$$\text{Cov}(\mathbf{y}_{ij},\mathbf{y}'_{ij}) = \mathbf{X}_{(3)ij}\, \boldsymbol{\Phi}_{(3)}\, \mathbf{X}'_{(3)ij} + \mathbf{X}_{(2)ij}\, \boldsymbol{\Phi}_{(2)}\, \mathbf{X}'_{(2)ij} + \mathbf{X}_{(1)ij}\, \boldsymbol{\Phi}_{(1)}\, \mathbf{X}'_{(1)ij}\,.$$

Note that the level-1 covariance matrix $\boldsymbol{\Phi}_{(1)}$ can be written as the symmetric matrix

$$\boldsymbol{\Phi}_{(1)} = \begin{bmatrix} \Phi_{(1)_{11}} & & \\ \Phi_{(1)_{21}} & & \\ \Phi_{(1)_{31}} & \Phi_{(1)_{32}} & \Phi_{(1)_{33}} \end{bmatrix}$$

containing six non-duplicated elements. The variance on level-1 is given by the expression

$$\text{var}_{(1)} = \Phi_{(1)11} + (Q1)^2\,\Phi_{(1)22} + (Q2)^2\,\Phi_{(1)33} + 2\,(Q1)\,\Phi_{(1)12} +$$

$$2\,(Q2)\,\Phi_{(1)13} + 2\,(Q1)(Q2)\,\Phi_{(1)23}\,.$$

The variance assumes three distinct values, as shown below:

Value of Q1	**Value of Q2**	**Value of var$_{(1)}$**
1	0	$(\Phi_{(1)11} + \Phi_{(1)22}) + 2\,\Phi_{(1)12}$
0	1	$(\Phi_{(1)11} + \Phi_{(1)33}) + 2\,\Phi_{(1)13}$
-1	-1	$(\Phi_{(1)11} + \Phi_{(1)22} + \Phi_{(1)33}) + 2\,\Phi_{(1)12} - 2\,\Phi_{(1)13} + 2\,\Phi_{(1)23}$

To obtain an unique solution for these variance components, $\Phi_{(1)22}$, $\Phi_{(1)33}$ and $\Phi_{(1)23}$ are set equal to 0. Constraints on the elements of $\Phi_{(2)}$ and $\Phi_{(3)}$ are as given in Example 5.4.2.

Fitting of the model as described above produces the following computer output:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	**STD.ERR.**	**Z-VALUE**	**PR>|Z|**
CONS	6.6280	0.1607	41.2446	0.0000
S1	-0.1651	0.1046	-1.5784	0.1145
Q1	-0.3794	0.1005	-3.7751	0.0002
Q2	0.1124	0.1031	1.0902	0.2756
L1	1.7000	0.1588	10.7053	0.0000
L2	0.5674	0.1458	3.8916	0.0001
FACTOR3	0.2045	0.0377	5.4244	0.0000

(ii) **Random part of the model:**

Level-3:

PARAMETER	$\hat{\tau}$	STD.ERR	Z-VALUE	PR>|Z|
L1/L1	0.4862	0.1768	2.7500	0.0060
L2/L2	0.0139	0.1086	0.1280	0.8982
FACTOR3/L1	0.0145	0.0267	0.5431	0.5871
FACTOR3/L2	-0.0323	0.0225	-1.4356	0.1511
FACTOR3/FACTOR3	0.0169	0.0072	2.3472	0.0189

Level-2:

G1/G1	1.8460	0.6532	2.8261	0.0047
G2/G2	2.5170	1.2510	2.0120	0.0442
G3/G3	0.9605	0.2706	3.5495	0.0004
G4/G4	1.1730	0.5535	2.1192	0.0341

Level-1:

CONS/CONS	18.8500	0.3951	47.7094	0.0000
Q1/CONS	-0.0289	0.2589	0.1116	0.9111
Q2/CONS	0.6563	0.3035	2.1624	0.0306

(iii) **Likelihood function:**

$$-\ 2\ ln\ \mathrm{L} = 30318.8$$

From Part (i) of the output it follows that the coefficients for the intercept (CONS) and the dummy variables for language group, L1 and L2, will have the largest influence on the expected FACTOR1 score, given no other effect is present. From the coefficient for the dummy variables L1 and L2 it follows that the expected FACTOR1 score for Afrikaans-speaking employees is 8.328. For English-speaking employees the expected value of the response variable is 7.1954 and for employees with other home languages it is 4.3606. The Afrikaans-speaking employee is thus more likely to agree with the statements as given in Table 5.4.1.

The highest variation in the random component on level-3 of the model is for the dummy variable L1. The only coefficients significant at a 5 % level of significance are the coefficients for L1/L1 and FACTOR3/FACTOR3.

All coefficients on level-2 are significant at a 5 % level of significance, indicating significant variation for all four gender by appointment type subgroups over the occupational categories.

From Part (iii) of the output it follows that a discrepancy function value ($-2 \ln L$) of 30 318.8 is obtained as opposed to the discrepancy function value of 30 325.5 reported for the previous model.

It follows from Section 4.4.4 that the difference of 6.7 between the discrepancy function values may be regarded as an observation from a χ^2- distribution with 2 degrees of freedom. It can therefore be concluded that the addition of the two additional random coefficients on level-1 of the hierarchy is statistically justified (5 % level of significance).

5.5 SUMMARY

A general level-3 model, allowing for complex variance structures on all levels of the hierarchy, was introduced in this chapter. Cases where no coefficients were random at a specific level were considered. A FORTRAN program GENIGLS had been written to implement the theoretical results.

Three examples were given. The examples ranged from a case which was similar in structure to the cases discussed in the previous chapters, to an example of a level-3 model with complex variance structures on all levels of the hierarchy.

Note that the estimation procedure discussed in this chapter yields estimation results which was computationally less efficient than the procedures discussed in the previous chapters. The chief advantage of the procedure discussed in this chapter, however, is the ability to handle a wide variety of models, as provision is also made for the situation where there are no random coefficients on a particular level of the model. In the next chapter attention will be given to a computationally efficient estimation procedure for analysing multivariate multilevel models with continuous response variables.

CHAPTER 6
MULTIVARIATE MULTILEVEL MODELS

6.1 INTRODUCTION

In this chapter a multivariate multilevel model will be considered, that is where the outcomes of two or more correlated continuous response variables are assumed to depend on the same set of predictors. An example of such a model was given by Cresswell (1991), who examined educational achievement across schools, with each pupil having two response scores. A level-3 hierarchy was assumed, with the scores at level-1, pupils at level-2 and schools at level-3.

While it may be easier to analyse different responses via separate models, there are the following advantages in doing one multivariate analysis (Prosser, Rasbash & Goldstein, 1991):

(i) This approach is particularly suited to the analysis of data with missing or incomplete responses.
(ii) This type of analysis allows for random coefficients on the various levels of the hierarchy, thus enabling one to quantify the amount of between-unit-variation at a specific level.
(iii) Direct comparison of the way in which measurements relate to the explanatory variables are facilitated.
(iv) By taking the hierarchical structure of the data into account, the responses may be estimated more efficiently.

A general level-3 model will be defined in Section 6.2. In Section 6.3 the estimation of the unknown parameters will be discussed while simplification of the terms needed for parameter estimation in the case of complete data will be given in Section 6.4. The handling of missing data will be considered in Section 6.5 and two practical applications given in Section 6.6.

6.2 THE MULTIVARIATE MULTILEVEL MODEL

A multivariate multilevel model is a model in which there are two or more response variables. An example, which will be discussed in detail in Example 6.6.1, is multiple electricity consumption measurements per day for a number of days (the level-2 units) for a given household (the level-3 unit). The consumption measurements may then be regarded as level-1 units nested within a particular day.

Suppose that there are q response variables and let $\mathbf{y}_{ijk}$ denote the k-th response for the (i,j)-th unit.

The model to be considered in this section is defined by

$$\mathbf{y}_{ijk} = \mathbf{x}'_{(f)ijk}\,\beta + [\ \mathbf{v}_i\ ]_k + [\ \mathbf{u}_{ij}\ ]_k\ , \tag{6.2.1}$$

where

$$i = 1, 2, \ldots, N\,;\ j = 1, 2, \ldots, n_i\,;\ k \in \{1, 2, \ldots, q\}.$$

Assume that the $q \times 1$ random vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are independently and identically distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Phi}_{(3)}$, independently distributed of the $q \times 1$ i.i.d random vectors $\mathbf{u}_{i1}, \mathbf{u}_{i2}, \ldots, \mathbf{u}_{in_i}$ which have mean $\mathbf{0}$ and covariance matrix $\mathbf{\Phi}_{(2)}$.

$\mathbf{x}'_{(f)ijk}$: $1 \times s$ is a typical row of the design matrix of the fixed part of the model, the elements being values of the s predictors. The elements of $\mathbf{v}_i$ and $\mathbf{u}_{ij}$ make provision for variation of responses over level-3 and level-2 units respectively.

Note that no allowance is made for level-1 variation, the reason for this being twofold. Firstly, the model as formulated in (6.2.1) does not make provision for the unique identification of level-2 and level-1 variance components. Secondly, there are no true experimental units below level-2.

The set of equations given in (6.2.1) can be written in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_{(1)i}\,\boldsymbol{\beta} + \mathbf{X}_{(3)i}\,\mathbf{v}_i + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\,\mathbf{u}_{ij}$$

where $\mathbf{X}_{(1)i}$ has typical row $\mathbf{x}'_{(1)ijk}$,

$$\mathbf{X}_{(3)i} = \begin{bmatrix} \mathbf{S}_{i1} \\ \vdots \\ \mathbf{S}_{ij} \\ \vdots \\ \mathbf{S}_{in_i} \end{bmatrix}, \tag{6.2.2}$$

$\mathbf{Z}_{(2)ij}$ is a $(\sum_{j=1}^{n_i} n_{ij}) \times m$ matrix partitioned as

$$\mathbf{Z}_{(2)ij} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ \mathbf{S}_{ij} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{6.2.3}$$

and $\mathbf{S}_{ij}$ is a selection matrix consisting of a subset of the rows of the $q \times q$ identity matrix $\mathbf{I}_q$ where the rows of $\mathbf{S}_{ij}$ correspond to the response measurements available for the (i,j)-th unit.

As an example of how the $\mathbf{S}_{ij}$ matrices are constructed, consider the measurement of peak-hour electricity consumption where responses were measured at 7h00, 8h00, 9h00, 19h00, 20h00 and 21h00 respectively. For the case where all six response measurements are available, $\mathbf{S}_{ij} = \mathbf{I}_6$. If, however, only the 8h00 and 19h00 measurements are available,

$$\mathbf{S}_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

6.3 PARAMETER ESTIMATION

Under the distributional assumptions given in Section 6.2, it follows that

$$\mathrm{E}(\mathbf{y}_i) = \mathbf{X}_{(f)i}\,\boldsymbol{\beta}$$

and

$$\begin{aligned}\mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i') &= \boldsymbol{\Sigma}_i \\ &= \mathbf{X}_{(3)i}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)i} + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{Z}'_{(2)ij}\,, \end{aligned} \qquad (6.3.1)$$

with $\mathbf{X}_{(f)i}$, $\mathbf{X}_{(3)i}$ and $\mathbf{Z}_{(2)ij}$ as defined by (6.2.2) and (6.2.3).

It is supposed that $\widehat{\boldsymbol{\Phi}}_{(3)}$ and $\widehat{\boldsymbol{\Phi}}_{(2)}$ are consistent estimators of $\boldsymbol{\Phi}_{(3)}$ and $\boldsymbol{\Phi}_{(2)}$ respectively so that

$$\begin{aligned}\mathbf{V}_i &= \mathbf{X}_{(3)i}\,\widehat{\boldsymbol{\Phi}}_{(3)}\,\mathbf{X}'_{(3)i} + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}\,\widehat{\boldsymbol{\Phi}}_{(2)}\,\mathbf{Z}'_{(2)ij} \\ &= \mathbf{X}_{(3)i}\,\widehat{\boldsymbol{\Phi}}_{(3)}\,\mathbf{X}'_{(3)i} + \boldsymbol{\Lambda}_i \end{aligned} \qquad (6.3.2)$$

is a consistent estimator of $\boldsymbol{\Sigma}_i$.

The generalized least squares estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained as the minimum of the quadratic function

$$\mathbf{Q}_f = \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_{(f)i}\,\boldsymbol{\beta})'\,\mathbf{V}_i^{-1}\,(\mathbf{y}_i - \mathbf{X}_{(f)i}\,\boldsymbol{\beta})$$

with solution

$$\hat{\beta} = [\ \sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{X}_{(f)i}\ ]^{-1}\ [\ \sum_{i=1}^{N} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{y}_i\ ]\ . \tag{6.3.3}$$

In order to estimate $\mathbf{\Phi}_{(3)}$ and $\mathbf{\Phi}_{(2)}$, let

$$\mathbf{y}_i^* = \text{vecs}\ (\mathbf{y}_i - \mathbf{X}_{(f)i}\,\beta)(\mathbf{y}_i - \mathbf{X}_{(f)i}\,\beta)'\ , \tag{6.3.4}$$

then

$$\text{E}(\mathbf{y}_i^*) = \text{vecs}\ \Sigma_i\ . \tag{6.3.5}$$

Using the result (cf. (4.2.11))

$$\text{vec}\ (\mathbf{A}\ \mathbf{B}\ \mathbf{C}') = \mathbf{C}' \otimes \mathbf{A}\ \ \text{vec}\ \mathbf{B}\ , \tag{6.3.6}$$

it follows that

$$\text{vec}\ \mathbf{V}_i = (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i})\ \text{vec}\ \mathbf{\Phi}_{(3)} + \sum_{j=1}^{n_i} (\mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij})\ \text{vec}\ \mathbf{\Phi}_{(2)}\ . \tag{6.3.7}$$

There exists an unique matrix (cf. (5.2.12)) $\mathbf{G}_p$: $p^2 \times \frac{1}{2}\, p(p+1)$ so that

$$\text{vec}\ \mathbf{A} = \mathbf{G}_p\ \text{vecs}\ \mathbf{A}$$

with $\mathbf{A}$ a symmetric $p \times p$ matrix. There is also (cf. (5.2.13)) a non-unique matrix $\mathbf{H}_p : \frac{1}{2}\, p(p+1) \times p^2$ so that

$$\text{vecs}\ \mathbf{A} = \mathbf{H}_p\ \text{vec}\ \mathbf{A}\ . \tag{6.3.8}$$

Vecs $\mathbf{V}_i$ can therefore be written as

$$\text{vecs } \mathbf{V}_i = \mathbf{H}_{n_i^*} (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q \text{ vecs } \mathbf{\Phi}_{(3)} + \mathbf{H}_{n_i^*} (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \, \mathbf{G}_m \text{ vecs } \mathbf{\Phi}_{(2)}$$

$$= \mathbf{X}_i^* \boldsymbol{\tau} \qquad (6.3.9)$$

where

$$\mathbf{X}_i^* = \mathbf{H}_{n_i^*} [(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q \qquad (\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij} \otimes \mathbf{Z}_{(2)ij}) \, \mathbf{G}_m] \qquad (6.3.10)$$

and

$$\boldsymbol{\tau} = \begin{bmatrix} \text{vecs} & \mathbf{\Phi}_{(3)} \\ \text{vecs} & \mathbf{\Phi}_{(2)} \end{bmatrix}. \qquad (6.3.11)$$

Now consider the quadratic form

$$\mathbf{Q}_\tau = \sum_{i=1}^{N} (\mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\tau})' \, \mathbf{W}_i^{-1} (\mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\tau})$$

where (cf. (5.2.16))

$$2 \, \mathbf{W}_i^{-1} = \mathbf{G}_{n_i}' (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \, \mathbf{G}_{n_i} .$$

Minimization of $\mathbf{Q}_\tau$ with respect to $\boldsymbol{\tau}$ yields

$$\hat{\boldsymbol{\tau}} = [\sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \, \mathbf{W}_i^{-1} \mathbf{X}_i^*]^{-1} [\sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \, \mathbf{W}_i^{-1} \mathbf{y}_i^*] . \qquad (6.3.12)$$

To obtain the IGLS estimates of the unknown parameters proceed as follows:

(i) Set $\mathbf{V}_i = \mathbf{I}$.
(ii) Calculate $\hat{\boldsymbol{\beta}}$, $\mathbf{y}_i^*$ and $\hat{\boldsymbol{\tau}}$ using expressions (6.3.3), (6.3.4) and (6.3.12).
(iii) Obtain a revised estimate of $\mathbf{V}_i$.
Repeat steps (ii) and (iii) until convergence is obtained.

Approximate standard errors of the elements of $\widehat{\beta}$ are obtained as the square roots of the diagonal elements of the matrix $[\sum_{i=1}^{N} \mathbf{X}'_{(f)i} \mathbf{V}_i^{-1} \mathbf{X}_{(f)i}]^{-1}$. Likewise, approximate standard errors of the elements of $\widehat{\tau}$ are obtained as the square roots of the diagonal elements of the matrix $[\sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^{*}]^{-1}$.

6.4 COMPUTATIONAL CONSIDERATIONS

As has been pointed out in Chapters 4 and 5, matrix expressions such as (6.3.3) and (6.3.12) require further analytical evaluation prior to their implementation in a computer program. In this section computationally efficient expressions will be derived for each of a number of submatrices. It is assumed that observations on each of the response variables are available. The handling of missing responses is considered in Section 6.5.

6.4.1 SIMPLIFICATION OF THE COMPONENTS OF THE WEIGHT MATRIX

In this section it is assumed that, for a given (i,j)-combination, where i denotes a level-3 and j a level-2 unit, observations were made on each of the q response variables. Thus (cf. (6.2.2)), $\mathbf{S}_{ij} = \mathbf{I}_q$, $i = 1, 2, \ldots, N$; $j = 1, 2, \ldots, n_i$. Expressions for the simplified terms of $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^{*}$ in the presence of missing data will be derived in Section 6.5.1.

It follows from (6.3.10) and (5.2.16) that the matrix $\mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^{*}$ can be written as

$$2\, \mathbf{X}_i^{*\prime} \mathbf{W}_i^{-1} \mathbf{X}_i^{*} = \begin{bmatrix} \mathbf{G}'_q \,(\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}) \\ \mathbf{G}'_m \,(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij}) \end{bmatrix} \mathbf{H}'_{n_i^*} \mathbf{G}'_{n_i} (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \mathbf{G}_{n_i} \mathbf{H}_{n_i^*} \times$$

$$[(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \mathbf{G}_q \quad (\sum_{j^*=1}^{n_i} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}_{(2)ij^*}) \mathbf{G}_m]$$

$$= \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} \end{bmatrix} \qquad (6.4.1)$$

$$= \mathbf{T}_i \, ,$$

where

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q \, (\mathbf{X}'_{(3)i} \otimes \mathbf{X}'_{(3)i}) \, \mathbf{H}'_{n_i^*} \, \mathbf{G}'_{n_i} \, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \, \mathbf{G}_{n_i} \, \mathbf{H}_{n_i^*} \, [(\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q \, ,$$

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m \, (\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij}) \, \mathbf{H}'_{n_i^*} \, \mathbf{G}'_{n_i} \, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \, \mathbf{G}_{n_i} \, \mathbf{H}_{n_i^*} \, (\mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}) \, \mathbf{G}_q$$

and

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \, (\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \otimes \mathbf{Z}'_{(2)ij}) \mathbf{H}'_{n_i^*} \, \mathbf{G}'_{n_i} \, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1}) \, \mathbf{G}_{n_i} \, \mathbf{H}_{n_i^*} \times$$
$$(\sum_{j^*=1}^{n_i} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}_{(2)ij^*}) \, \mathbf{G}_m \, .$$

Using the results (Magnus & Neudecker, 1988)

$$\mathbf{G}_k \, \mathbf{H}_k \, (\mathbf{X} \otimes \mathbf{X}) \, \mathbf{G}_m = (\mathbf{X} \otimes \mathbf{X}) \, \mathbf{G}_m \, , \qquad (6.4.2)$$

and (Graham, 1981)

$$(\mathbf{A}' \otimes \mathbf{B}')(\mathbf{C} \otimes \mathbf{C})(\mathbf{A} \otimes \mathbf{B}) = \mathbf{A}' \, \mathbf{C} \, \mathbf{A} \otimes \mathbf{B}' \, \mathbf{C} \, \mathbf{B} \, , \qquad (6.4.3)$$

the submatrices $\mathbf{T}_{(3,3)}$, $\mathbf{T}_{(2,3)}$ and $\mathbf{T}_{(2,2)}$ can be expressed as

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q \, (\mathbf{X}'_{(3)i} \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i} \otimes \mathbf{X}'_{(3)i} \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i}) \, \mathbf{G}_q \, , \qquad (6.4.4)$$

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m \, (\sum_{j=1}^{n_i} \{ \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i} \otimes \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i} \}) \, \mathbf{G}_q \qquad (6.4.5)$$

and

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} (\mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{Z}'_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{Z}_{(2)ij^*}) \, \mathbf{G}_m \, . \qquad (6.4.6)$$

Computationally efficient expressions for each of the submatrices in $\mathbf{T}_i$ are given in Propositions 6.4.1 to 6.4.3. The results of these propositions are summarized in Theorem 6.4.1.

PROPOSITION 6.4.1 (Submatrix $\mathbf{T}_{(3,3)}$)

$$\begin{aligned} \mathrm{T}_{(3,3)} &= \mathbf{G}'_q\,(\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} \otimes \mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,)\,\mathbf{G}_q \\ &= \mathbf{G}'_q\,\{\mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \otimes \mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]\}\,\mathbf{G}_q \end{aligned} \qquad (6.4.7)$$

where

$$\begin{aligned} \mathbf{A}_i &= \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} \\ &= n_i\,\boldsymbol{\Phi}_{(2)}^{-1} \end{aligned} \qquad (6.4.8)$$

and

$$\mathbf{C}_i = (\boldsymbol{\Phi}_{(3)}^{-1} + \mathbf{A}_i)^{-1}\,. \qquad (6.4.9)$$

Proof

Let (cf. (5.2.26))

$$\mathbf{A}_i = \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\;. \qquad (6.4.10)$$

Using (6.2.2), it follows that (6.4.10) can be written as

$$\mathbf{A}_i = \sum_{j=1}^{n_i} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{S}_{ij}\,. \qquad (6.4.11)$$

For the case considered in this section $\mathbf{S}_{ij} = \mathbf{I}_q$, so that (6.4.11) reduces to (cf. (6.2.3))

$$\mathbf{A}_i = \sum_{j=1}^{n_i} \boldsymbol{\Lambda}_{ij}^{-1}$$

$$= \sum_{j=1}^{n_i} \mathbf{S}_{ij}\, \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{S}'_{ij}$$

$$= n_i\, \mathbf{\Phi}_{(2)}^{-1}\,. \qquad (6.4.12)$$

Using the matrix identity (see for example Browne, 1991)

$$(\mathbf{A} + \mathbf{B}\,\mathbf{C}\,\mathbf{B}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\,\mathbf{B}\,(\mathbf{C}^{-1} + \mathbf{B}'\,\mathbf{A}^{-1}\,\mathbf{B}\,)^{-1}\,\mathbf{B}'\,\mathbf{A}^{-1} \qquad (6.4.13)$$

it follows that

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} - \mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\;\mathbf{C}_i\;\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} \qquad (6.4.14)$$

with

$$\mathbf{C}_i = (\mathbf{\Phi}_{(3)}^{-1} + \mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,)^{-1}\,. \qquad (6.4.15)$$

Using (6.4.12), (6.4.14) can be written as

$$\mathbf{X}'_{(3)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{A}_i\,[\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]$$

with (cf. (6.4.15))

$$\mathbf{C}_i = (\mathbf{\Phi}_{(3)}^{-1} + \mathbf{A}_i\,)^{-1}$$

which concludes the proof. □

PROPOSITION 6.4.2 (Submatrix $\mathbf{T}_{(2,3)}$)

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m\,(\,\sum_{j=1}^{n_i}\{\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} \otimes \mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i}\,\})\,\mathbf{G}_q$$

$$= n_i\,\mathbf{G}'_m\,(\mathbf{\Phi}_{(2)}^{-1}\,[\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,])\otimes(\mathbf{\Phi}_{(2)}^{-1}\,[\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,])\,\mathbf{G}_q \qquad (6.4.16)$$

with $\mathbf{A}_i$ and $\mathbf{C}_i$ defined by (6.4.8) and (6.4.9) respectively.

Proof

From (6.3.2) and (6.4.13) it follows that

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} - \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,, \qquad (6.4.17)$$

where (cf. (6.4.8))

$$\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = n_i\,\mathbf{\Phi}_{(2)}^{-1} \qquad (6.4.18)$$

and

$$\begin{aligned}\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} &= \mathbf{S}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij} \\ &= \mathbf{S}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{S}_{ij} \\ &= \mathbf{\Phi}_{(2)}^{-1} \qquad (6.4.19)\end{aligned}$$

as $\mathbf{S}_{ij} = \mathbf{I}_q$.

Substitution of (6.4.18) and (6.4.19) in (6.4.17) gives

$$\mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]$$

so that

$$\begin{aligned}\mathbf{T}_{(2,3)} &= \mathbf{G}'_m\,\{\,\sum_{j=1}^{n_i}\,(\mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]) \otimes (\mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,])\}\,\mathbf{G}_q \\ &= n_i\,\mathbf{G}'_m\,(\mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \otimes \mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,])\,\mathbf{G}_q\,. \qquad \square\end{aligned}$$

PROPOSITION 6.4.3 (Submatrix $\mathbf{T}_{(2,2)}$)

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} (\mathbf{Z}'_{(2)ij}\, \mathbf{V}_i^{-1}\, \mathbf{Z}'_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij}\, \mathbf{V}_i^{-1}\, \mathbf{Z}'_{(2)ij^*})\, \mathbf{G}_m$$

$$= n_i\, \mathbf{G}'_m \{(\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i) \otimes (\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i) + n_i (n_i - 1)(\mathbf{B}_i \otimes \mathbf{B}_i)\} \mathbf{G}_m \tag{6.4.20}$$

where

$$\mathbf{B}_i = \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{C}_i\, \mathbf{\Phi}_{(2)}^{-1} \tag{6.4.21}$$

and $\mathbf{C}_i$ defined by (6.4.9).

Proof

From (6.3.2) and (6.4.13) it follows that

$$\mathbf{Z}'_{(2)ij}\, \mathbf{V}_i^{-1}\, \mathbf{Z}_{(2)ij^*} = \mathbf{Z}'_{(2)ij}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{Z}_{(2)ij^*} - \mathbf{Z}'_{(2)ij}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{X}_{(3)i}\, \mathbf{C}_i\, \mathbf{X}'_{(3)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{Z}_{(2)ij^*}\,, \tag{6.4.22}$$

where (cf. (6.4.19))

$$\mathbf{Z}'_{(2)ij}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{X}_{(3)i} = \mathbf{\Phi}_{(2)}^{-1} \tag{6.4.23}$$

and thus

$$\mathbf{X}'_{(3)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{Z}_{(2)ij^*} = \mathbf{\Phi}_{(2)}^{-1}\,. \tag{6.4.24}$$

The term $\mathbf{Z}'_{(2)ij}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{Z}_{(2)ij^*}$ can be written as (cf. (6.2.3))

$$\mathbf{Z}'_{(2)ij}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{Z}_{(2)ij^*} = \mathbf{\Phi}_{(2)}^{-1} \text{ if } j = j^*$$

$$= 0 \quad \text{otherwise}\,. \tag{6.4.25}$$

Substitution of (6.4.23) to (6.4.25) in (6.4.22) gives (cf. (6.4.21))

$$\begin{aligned} \mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} &= \delta_{jj^*}\,\mathbf{\Phi}_{(2)}^{-1} - \mathbf{\Phi}_{(2)}^{-1}\,\mathbf{C}_i\,\mathbf{\Phi}_{(2)}^{-1} \\ &= \delta_{jj^*}\,\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,. \end{aligned} \qquad (6.4.26)$$

The term $\mathbf{T}_{(2,2)}$ can then be written as

$$\begin{aligned} \mathbf{T}_{(2,2)} = {} & \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} \{(\,\delta_{jj^*}\,\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,) \otimes (\,\delta_{jj^*}\,\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,)\}\,\mathbf{G}_m \\ = {} & n_i\,\mathbf{G}'_m\{(\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,) \otimes (\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,)\}\,\mathbf{G}_m + \\ & n_i\,\mathbf{G}'_m\,\{n_i\,(n_i - 1)(\,\mathbf{B}_i \otimes \mathbf{B}_i\,)\}\,\mathbf{G}_m \end{aligned}$$

which concludes the proof. □

THEOREM 6.4.1

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} \end{bmatrix}$$

where

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q\,\{\mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,] \otimes \mathbf{A}_i\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]\}\,\mathbf{G}_q$$

$$\mathbf{T}_{(2,3)} = n_i\,\mathbf{G}'_m\,(\mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]) \otimes \mathbf{\Phi}_{(2)}^{-1}\,[\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,])\,\mathbf{G}_q$$

$$\mathbf{T}_{(2,2)} = n_i\,\mathbf{G}'_m\,\{(\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,) \otimes (\mathbf{\Phi}_{(2)}^{-1} - \mathbf{B}_i\,) + n_i\,(n_i - 1)(\,\mathbf{B}_i \otimes \mathbf{B}_i\,)\}\,\mathbf{G}_m$$

with (cf. (6.4.8), (6.4.9) and (6.4.21))

$$\mathbf{A}_i = n_i\,\mathbf{\Phi}_{(2)}^{-1}$$

$$\mathbf{C}_i = (\boldsymbol{\Phi}_{(3)}^{-1} + \mathbf{A}_i)^{-1}$$

and

$$\mathbf{B}_i = \boldsymbol{\Phi}_{(2)}^{-1}\, \mathbf{C}_i\, \boldsymbol{\Phi}_{(2)}^{-1}\,.$$

Proof

The proof follows directly from Propositions 6.4.1 to 6.4.3. □

6.4.2 SIMPLIFICATION OF THE COMPONENTS OF THE COEFFICIENT VECTOR

To calculate $\hat{\tau}$ (cf. (6.3.11) and (6.3.12)) it is also necessary to find a computationally efficient way to calculate $\mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^*$. In this subsection, expressions for the efficient computation of $\mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^*$ in the case of complete data are derived. Derivation of simplified expressions of $\mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^*$ in the case of missing data will be given in Section 6.5.2.

The vector $\mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^*$ is given by

$$\begin{aligned} 2\, \mathbf{X}_i^{*\prime}\, \mathbf{W}_i^{-1}\, \mathbf{y}_i^* &= \begin{bmatrix} \mathbf{G}_q' \,(\, \mathbf{X}_{(3)i}' \otimes \mathbf{X}_{(3)i}' \,)\, \mathbf{H}_{n_i^*}'\, \mathbf{G}_{n_i}'\, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\, \mathbf{G}_{n_i}\, \mathbf{y}_i^* \\ \mathbf{G}_m' \,(\, \sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \otimes \mathbf{Z}_{(2)ij}' \,)\, \mathbf{H}_{n_i^*}'\, \mathbf{G}_{n_i}'\, (\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\, \mathbf{G}_{n_i}\, \mathbf{y}_i^* \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \end{bmatrix} \\ &= \mathbf{q}_i\,. \end{aligned} \tag{6.4.27}$$

Using the results (6.4.2) and (6.4.3), the vectors $\mathbf{q}_{(3)}$ and $\mathbf{q}_{(2)}$ can be written as

$$\mathbf{q}_{(3)} = \mathbf{G}_q'\, (\mathbf{X}_{(3)i}'\, \mathbf{V}_i^{-1} \otimes \mathbf{X}_{(3)i}'\, \mathbf{V}_i^{-1}\,)\, \mathbf{G}_{n_i}\, \mathbf{y}_i^*$$

and

$$\mathbf{q}_{(2)} = \mathbf{G}'_m \left(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \right) \mathbf{G}_{n_i} \mathbf{y}_i^* .$$

Let

$$\mathbf{Y}_i = (\mathbf{y}_i - \mathbf{X}_{(1)i} \beta)(\mathbf{y}_i - \mathbf{X}_{(1)i} \beta)' , \tag{6.4.28}$$

then

$$\mathbf{y}_i^* = \text{vecs} \, (\mathbf{Y}_i) . \tag{6.4.29}$$

From (6.3.8) and (6.4.2) it follows that

$$\mathbf{q}_{(3)} = \mathbf{G}'_q \, (\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \otimes \mathbf{X}'_{(3)i} \mathbf{V}_i^{-1}) \, \text{vec} \, \mathbf{Y}_i$$

and

$$\mathbf{q}_{(2)} = \mathbf{G}'_m \left(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \right) \text{vec} \, \mathbf{Y}_i . \tag{6.4.30}$$

Using (6.3.6), the vectors $\mathbf{q}_{(3)}$ and $\mathbf{q}_{(2)}$ can be expressed as

$$\mathbf{q}_{(3)} = \mathbf{G}'_q \, \text{vec} \, (\mathbf{X}'_{(3)i} \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{X}_{(3)i}) \tag{6.4.31}$$

and

$$\mathbf{q}_{(2)} = \mathbf{G}'_m \, \text{vec} \left(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij} \right) . \tag{6.4.32}$$

Computationally efficient expressions for each of the subvectors in $\mathbf{q}_i$ are given in Propositions 6.4.5 and 6.4.6. The results of these propositions are summarized in Theorem 6.4.2.

PROPOSITION 6.4.5 (Subvector $\mathbf{q}_{(3)}$)

$$\mathbf{q}_{(3)} = \mathbf{G}'_q \text{ vec } (\mathbf{a}_i \, \mathbf{a}'_i) \tag{6.4.33}$$

where

$$\mathbf{a}_i = [\, \mathbf{I} - \mathbf{A}_i \, \mathbf{C}_i \,] \, \mathbf{\Phi}_{(2)}^{-1} \, \{ \sum_{j=1}^{n_i} \mathbf{e}_{ij} \} \tag{6.4.34}$$

with

$$\mathbf{e}_{ij} = \mathbf{y}_{ij} - \mathbf{X}_{(f)ij} \, \boldsymbol{\beta} \tag{6.4.35}$$

and $\mathbf{A}_i$ and $\mathbf{C}_i$ as defined by (6.4.8) and (6.4.9) respectively.

Proof

Let

$$\mathbf{a}_i = \mathbf{X}'_{(3)i} \, \mathbf{V}_i^{-1} \, \mathbf{e}_i \tag{6.4.36}$$

with $\mathbf{e}_{ij}$ defined by (6.4.35). Using (6.3.2) and (6.4.13), (6.4.36) can be written as

$$\mathbf{a}_i = \mathbf{X}'_{(3)i} \, \mathbf{\Lambda}_i^{-1} \, \mathbf{e}_i - \mathbf{X}'_{(3)i} \, \mathbf{\Lambda}_i^{-1} \, \mathbf{X}_{(3)i} \, \mathbf{C}_i \, \mathbf{X}'_{(3)i} \, \mathbf{\Lambda}_i^{-1} \, \mathbf{e}_i$$

where (cf. (6.4.8))

$$\mathbf{X}'_{(3)i} \, \mathbf{\Lambda}_i^{-1} \, \mathbf{X}_{(3)i} = \mathbf{A}_i \tag{6.4.37}$$

and (cf. (6.2.2))

$$\begin{aligned} \mathbf{X}'_{(3)i} \, \mathbf{\Lambda}_i^{-1} \, \mathbf{e}_i &= \sum_{j=1}^{n_i} \mathbf{X}'_{(3)ij} \, \mathbf{\Lambda}_{ij}^{-1} \, \mathbf{e}_{ij} \\ &= \sum_{j=1}^{n_i} \mathbf{S}'_{ij} \, \mathbf{\Lambda}_{ij}^{-1} \, \mathbf{e}_{ij} \\ &= \mathbf{\Phi}_{(2)}^{-1} \sum_{j=1}^{n_i} \mathbf{e}_{ij} \end{aligned}$$

as

$$\mathbf{S}_{ij} = \mathbf{I}_q \, . \tag{6.4.38}$$

Substitution of (6.4.37) and (6.4.38) in (6.4.36) concludes the proof. □

PROPOSITION 6.4.6 (Subvector $\mathbf{q}_{(2)}$)

$$\mathbf{q}_{(2)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \{\mathrm{vec}\,(\mathbf{b}_{ij}\,\mathbf{b}'_{ij}\,)\} \tag{6.4.39}$$

where

$$\mathbf{b}_{ij} = \mathbf{\Phi}_{(2)}^{-1}\,[\mathbf{e}_{ij} \; - \; \mathbf{C}_i \;\mathbf{\Phi}_{(2)}^{-1} \sum_{j=1}^{n_i} \mathbf{e}_{ij}\,] \tag{6.4.40}$$

with $\mathbf{e}_{ij}$ defined by (6.4.35).

Proof

From (6.4.32)

$$\mathbf{q}_{(2)} = \mathbf{G}'_m \;\mathrm{vec}\,(\sum_{j=1}^{n_i} \mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{Y}_i\;\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij}\,)\,.$$

Let

$$\mathbf{b}_{ij} = \mathbf{Z}'_{(2)ij}\,\mathbf{V}_i^{-1}\,\mathbf{e}_i\,. \tag{6.4.41}$$

Using (6.3.2) and (6.4.13), (6.4.41) can be written as

$$\mathbf{b}_{ij} = \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i \; - \; \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i \tag{6.4.42}$$

where (cf. (6.4.19))

$$\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{\Phi}_{(2)}^{-1} \tag{6.4.43}$$

and (cf. (6.2.3))

$$\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i = \mathbf{S}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}$$

$$= \mathbf{\Phi}_{(2)}^{-1}\,\mathbf{e}_{ij}\,. \tag{6.4.44}$$

Finally,

$$\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{e}_i = \sum_{j=1}^{n_i} \mathbf{S}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}$$

$$= \mathbf{\Phi}_{(2)}^{-1} \sum_{j=1}^{n_i} \mathbf{e}_{ij}\,. \tag{6.4.45}$$

Substitution of (6.4.43) to (6.4.45) in (6.4.42) gives

$$\mathbf{b}_{ij} = \mathbf{\Phi}_{(2)}^{-1}\,[\mathbf{e}_{ij} - \mathbf{C}_i\,\mathbf{\Phi}_{(2)}^{-1} \sum_{j=1}^{n_i} \mathbf{e}_{ij}\,]$$

which concludes the proof. □

THEOREM 6.4.2

$$\mathbf{q}_i = \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \end{bmatrix}$$

where

$$\mathbf{q}_{(3)} = \mathbf{G}'_q \operatorname{vec}\,(\mathbf{a}_i\,\mathbf{a}'_i)\ ,$$

$$\mathbf{q}_{(2)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \{\operatorname{vec}\,(\mathbf{b}_{ij}\,\mathbf{b}'_{ij}\,)\}$$

with

$$\mathbf{a}_i = [\,\mathbf{I} - \mathbf{C}_i\,\mathbf{A}_i\,]\ \ \mathbf{\Phi}_{(2)}^{-1}\,\{\sum_{j=1}^{n_i} \mathbf{e}_{ij}\,\}\ ,$$

$$\mathbf{b}_{ij} = \mathbf{\Phi}_{(2)}^{-1} [\mathbf{e}_{ij} \ - \ \mathbf{C}_i \mathbf{\Phi}_{(2)}^{-1} \sum_{j=1}^{n_i} \mathbf{e}_{ij}]$$

and $\mathbf{A}_i$, $\mathbf{C}_i$ and $\mathbf{e}_{ij}$ as defined by (6.4.8), (6.4.9) and (6.4.35) respectively.

Proof

The proof follows directly from Propositions 6.4.5 and 6.4.6. □

6.4.3 SIMPLIFICATION OF THE WEIGHT MATRIX AND COEFFICIENT VECTOR (FIXED PART OF THE MODEL)

Finally, the terms $\mathbf{X}'_{(f)i} \mathbf{V}_i^{-1} \mathbf{X}_{(f)i}$ and $\mathbf{X}'_{(f)i} \mathbf{V}_i^{-1} \mathbf{y}_i$, required for the calculation of $\widehat{\boldsymbol{\beta}}$ (cf. (6.3.3)), are simplified.

Using (6.3.2) and (6.4.13), the matrix $\mathbf{X}'_{(f)i} \mathbf{V}_i^{-1} \mathbf{X}_{(f)i}$ can be written as

$$\mathbf{X}'_{(f)i} \mathbf{V}_i^{-1} \mathbf{X}_{(f)i} = \mathbf{X}'_{(f)i} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(f)i} - \mathbf{X}'_{(f)i} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(s)i} \mathbf{C}_i \mathbf{X}'_{(s)i} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(f)i} \ , \tag{6.4.46}$$

where

$$\begin{aligned} \mathbf{X}'_{(f)i} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(f)i} &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(f)ij} , \\ &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \{\mathbf{S}_{ij} \mathbf{\Phi}_{(2)}^{-1} \mathbf{S}'_{ij} \}^{-1} \mathbf{X}_{(f)ij} \\ &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \mathbf{\Phi}_{(2)}^{-1} \mathbf{X}_{(f)ij} , \end{aligned} \tag{6.4.47}$$

under the assumption that

$$\mathbf{S}_{ij} = \mathbf{I}_q \ , \ i = 1, 2 \ , \ . \ . \ . \ , N \ ; \ j = 1, 2, \ . \ . \ . \ , n_i \ .$$

Also,

$$\mathbf{X}'_{(f)i} \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(s)i} \ = \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \mathbf{\Lambda}_{ij}^{-1} \mathbf{X}_{(s)ij}$$

$$= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\, \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{S}_{ij}$$

$$= \{ \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \}\, \mathbf{\Phi}_{(2)}^{-1} \,. \tag{6.4.48}$$

Substitution of (6.4.47) and (6.4.48) in (6.4.46) gives (cf. (6.4.21))

$$\begin{aligned} \mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{X}_{(f)i} &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\, \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{X}_{(f)ij} - \\ &\quad \{ \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \}\, \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{C}_i\, \mathbf{\Phi}_{(2)}^{-1} \{ \sum_{j=1}^{n_i} \mathbf{X}_{(f)ij} \} \\ &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\, \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{X}_{(f)ij} - \{ \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \}\, \mathbf{B}_i \{ \sum_{j=1}^{n_i} \mathbf{X}_{(f)ij} \} \,. \end{aligned}$$

The vector $\mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{y}_i$ can be written (cf. (6.3.2) and (6.4.13)) as

$$\mathbf{X}'_{(f)i}\, \mathbf{V}_i^{-1}\, \mathbf{y}_i = \mathbf{X}'_{(f)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{y}_i - \mathbf{X}'_{(f)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{X}_{(3)i}\, \mathbf{C}_i\, \mathbf{X}'_{(3)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{y}_i \tag{6.4.49}$$

where (cf. (6.4.48))

$$\mathbf{X}'_{(f)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{X}_{(3)i} = \{ \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij} \}\, \mathbf{\Phi}_{(2)}^{-1} \,, \tag{6.4.50}$$

$$\begin{aligned} \mathbf{X}'_{(f)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{y}_i &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\, \mathbf{\Lambda}_{ij}^{-1}\, \mathbf{y}_{ij} \\ &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\, \mathbf{\Phi}_{(2)}^{-1}\, \mathbf{y}_{ij} \,, \end{aligned} \tag{6.4.51}$$

and

$$\begin{aligned} \mathbf{X}'_{(3)i}\, \mathbf{\Lambda}_i^{-1}\, \mathbf{y}_i &= \sum_{j=1}^{n_i} \mathbf{X}'_{(3)ij}\, \mathbf{\Lambda}_{ij}^{-1}\, \mathbf{y}_{ij} \\ &= \sum_{j=1}^{n_i} \mathbf{S}'_{ij}\, \mathbf{\Lambda}_{ij}^{-1}\, \mathbf{y}_{ij} \\ &= \mathbf{\Phi}_{(2)}^{-1} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \,. \end{aligned} \tag{6.4.52}$$

Substitution of (6.4.50) to (6.4.52) in (6.4.49) gives

$$\begin{aligned} \mathbf{X}'_{(f)i}\,\mathbf{V}_i^{-1}\,\mathbf{y}_i &= \sum_{j=1}^{n_i}\mathbf{X}'_{(f)ij}\,\boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{y}_{ij} - \{\sum_{j=1}^{n_i}\mathbf{X}'_{(f)ij}\}\,\boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{C}_i\,\boldsymbol{\Phi}_{(2)}^{-1}\,\{\sum_{j=1}^{n_i}\mathbf{y}_{ij}\} \\ &= \sum_{j=1}^{n_i}\mathbf{X}'_{(f)ij}\,\boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{y}_{ij} - \{\sum_{j=1}^{n_i}\mathbf{X}'_{(f)ij}\}\,\mathbf{B}_i\,\{\sum_{j=1}^{n_i}\mathbf{y}_{ij}\} \end{aligned}$$

with $\mathbf{B}_i$ and $\mathbf{C}_i$ defined in (6.4.21) and (6.4.9) respectively.

This concludes the simplification of all terms for the situation where no data on the response variables are missing. In the following section, the situation where response variable observations are missing will be considered.

6.5 MISSING MEASUREMENTS

Consider the situation where, for some (i,j)-combinations, measurements on one or more of the response variables are incomplete. Let ω_1 denote the set of m_i^* (i,j)-combinations for which $\mathbf{S}_{ij} = \mathbf{I}_q$ and ω_2 denote the set of $n_i - m_i^*$ (i,j)-combinations for which each $\mathbf{S}_{ij}$ consists of a subset of the rows of $\mathbf{I}_q$ corresponding to the measurements available for that particular (i,j)-combination. Also, let $\boldsymbol{\Phi}_{(2)ij}$ denote a matrix obtained by deleting rows and corresponding columns of $\boldsymbol{\Phi}_{(2)}$ as a result of the matrix product $\mathbf{S}_{ij}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{S}'_{ij}$. For example, suppose that only the first and second measurements of the q response variables are available, then

$$\mathbf{S}_{ij} = \begin{bmatrix} 1 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & 0 \end{bmatrix}$$

and hence

$$\mathbf{S}_{ij}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{S}'_{ij} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} = \boldsymbol{\Phi}_{(2)ij}\,. \tag{6.5.1}$$

6.5.1 SIMPLIFICATION OF THE COMPONENTS OF THE WEIGHT MATRIX

Derivation of modified expressions for the matrix $\mathbf{T}_i$ (cf. (6.4.1)) in the case of missing data is now considered.

PROPOSITION 6.5.1 (Submatrix $\mathbf{T}_{(3,3)}$)

$$\begin{aligned}\mathbf{T}_{(3,3)} &= \mathbf{G}_q' \, (\mathbf{X}_{(3)i}' \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i} \otimes \mathbf{X}_{(3)i}' \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i}) \, \mathbf{G}_q \\ &= \mathbf{G}_q' \, \{\mathbf{A}_i^* [\, \mathbf{I} - \mathbf{C}_i^* \, \mathbf{A}_i^* \,] \otimes \mathbf{A}_i^* [\, \mathbf{I} - \mathbf{C}_i^* \, \mathbf{A}_i^* \,]\} \, \mathbf{G}_q \qquad (6.5.2)\end{aligned}$$

where

$$\mathbf{A}_i^* = m_i^* \, \mathbf{\Phi}_{(2)}^{-1} + \sum_{j \epsilon \omega_2} \mathbf{S}_{ij}' \, \mathbf{\Phi}_{(2)ij}^{-1} \, \mathbf{S}_{ij} \qquad (6.5.3)$$

and

$$\mathbf{C}_i^* = (\mathbf{\Phi}_{(3)}^{-1} + \mathbf{A}_i^*)^{-1} \, . \qquad (6.5.4)$$

Proof

From (6.4.7) it follows that

$$\mathbf{T}_{(3,3)} = \mathbf{G}_q' \, \{\mathbf{A}_i [\, \mathbf{I} - \mathbf{C}_i \, \mathbf{A}_i \,] \otimes \mathbf{A}_i [\, \mathbf{I} - \mathbf{C}_i \, \mathbf{A}_i \,]\} \, \mathbf{G}_q$$

with (cf. (6.4.10))

$$\mathbf{A}_i = \mathbf{X}_{(3)i}' \, \mathbf{\Lambda}_i^{-1} \, \mathbf{X}_{(3)i} \, ,$$

which (cf. (6.4.11)) can be written as

$$\mathbf{A}_i = \sum_{j=1}^{n_i} \mathbf{S}_{ij}' \, \mathbf{\Lambda}_{ij}^{-1} \, \mathbf{S}_{ij} \, .$$

Let

$$\mathbf{A}_i^* = \sum_{j \epsilon \omega_1} \mathbf{A}_i + \sum_{j \epsilon \omega_2} \mathbf{A}_i \qquad (6.5.5)$$

where ω_1 denotes the set of (i,j)-combinations for which complete data is available and ω_2 denotes the (i,j)-combinations for which some response-variable measurements are missing. In the case of complete data, $\mathbf{S}_{ij} = \mathbf{I}_q$ and (cf. (6.4.12))

$$\sum_{j \epsilon \omega_1} \mathbf{A}_i = m_i^* \, \mathbf{\Phi}_{(2)}^{-1} \, . \qquad (6.5.6)$$

When dealing with the case of missing data, $\mathbf{S}_{ij} \neq \mathbf{I}_q$ and

$$\sum_{j \epsilon \omega_2} \mathbf{A}_i = \sum_{j \epsilon \omega_2} \mathbf{S}'_{ij} \, \mathbf{\Lambda}_{ij}^{-1} \, \mathbf{S}_{ij}$$

$$= \sum_{j \epsilon \omega_2} \mathbf{S}'_{ij} \, \{ \, \mathbf{S}_{ij} \, \mathbf{\Phi}_{(2)}^{-1} \, \mathbf{S}'_{ij} \, \}^{-1} \, \mathbf{S}_{ij}$$

which (cf. (6.5.1)) reduces to

$$\sum_{j \epsilon \omega_2} \mathbf{A}_i = \sum_{j \epsilon \omega_2} \mathbf{S}'_{ij} \, \mathbf{\Phi}_{(2)ij}^{-1} \, \mathbf{S}_{ij} \, . \qquad (6.5.7)$$

Substitution of (6.5.6) and (6.5.7) in (6.5.5) gives

$$\mathbf{A}_i^* = m_i^* \, \mathbf{\Phi}_{(2)}^{-1} + \sum_{j \epsilon \omega_2} \mathbf{S}'_{ij} \, \mathbf{\Phi}_{(2)ij}^{-1} \, \mathbf{S}_{ij}$$

which concludes the proof. □

PROPOSITION 6.5.2 (Submatrix $\mathbf{T}_{(2,3)}$)

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m \, (\sum_{j=1}^{n_i} \{ \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i} \otimes \mathbf{Z}'_{(2)ij} \, \mathbf{V}_i^{-1} \, \mathbf{X}_{(3)i} \}) \, \mathbf{G}_q$$

$$= \mathbf{G}'_m \sum_{j=1}^{n_i} \{ \mathbf{B}_{ij}^* \, [\mathbf{I} - \mathbf{C}_i^* \, \mathbf{A}_i^*] \otimes \mathbf{B}_{ij}^* \, [\mathbf{I} - \mathbf{C}_i^* \, \mathbf{A}_i^*] \} \, \mathbf{G}_q \qquad (6.5.8)$$

with

$$\mathbf{B}^*_{ij} = m^*_i \, \mathbf{\Phi}^{-1}_{(2)} + \sum_{j \epsilon \omega_2} \mathbf{S}'_{ij} \, \mathbf{\Phi}^{-1}_{(2)} \, \mathbf{S}_{ij} \tag{6.5.9}$$

and $\mathbf{A}^*_i$ and $\mathbf{C}^*_i$ as defined by (6.5.3) and (6.5.4).

Proof

From (6.4.17) it follows that $\mathbf{Z}'_{(2)ij} \, \mathbf{V}^{-1}_i \, \mathbf{X}_{(3)i}$ can be written as

$$\mathbf{Z}'_{(2)ij} \, \mathbf{V}^{-1}_i \, \mathbf{X}_{(3)i} = \mathbf{Z}'_{(2)ij} \, \mathbf{\Lambda}^{-1}_i \, \mathbf{X}_{(3)i} - \mathbf{Z}'_{(2)ij} \, \mathbf{\Lambda}^{-1}_i \, \mathbf{X}_{(3)i} \, \mathbf{C}_i \, \mathbf{X}'_{(3)i} \, \mathbf{\Lambda}^{-1}_i \, \mathbf{X}_{(3)i} \, . \tag{6.5.10}$$

From (6.5.3) and (6.5.5) it follows that

$$\mathbf{X}'_{(3)i} \, \mathbf{\Lambda}^{-1}_i \, \mathbf{X}_{(3)i} = \mathbf{A}^*_i \, . \tag{6.5.11}$$

Let

$$\begin{aligned} \mathbf{Z}'_{(2)ij} \, \mathbf{\Lambda}^{-1}_i \, \mathbf{X}_{(3)i} &= \mathbf{B}^*_{ij} \\ &= \sum_{j \epsilon \omega_1} \mathbf{B}_{ij} + \sum_{j \epsilon \omega_2} \mathbf{B}_{ij} \, . \end{aligned} \tag{6.5.12}$$

The first term of (6.5.12) reduces to

$$\begin{aligned} \sum_{j \epsilon \omega_1} \mathbf{B}_{ij} &= \sum_{j \epsilon \omega_1} \mathbf{S}'_{ij} \, \{ \, \mathbf{S}_{ij} \, \mathbf{\Phi}^{-1}_{(2)} \, \mathbf{S}'_{ij} \, \}^{-1} \, \mathbf{X}_{(3)ij} \\ &= \sum_{j \epsilon \omega_1} \mathbf{S}'_{ij} \, \mathbf{\Phi}^{-1}_{(2)} \, \mathbf{S}_{ij} \\ &= \sum_{j \epsilon \omega_1} \mathbf{\Phi}^{-1}_{(2)} \\ &= m^*_i \, \mathbf{\Phi}^{-1}_{(2)}, \end{aligned} \tag{6.5.13}$$

since $\mathbf{S}_{ij} = \mathbf{I}_q$.

The second term of (6.5.12) can be written as

$$\sum_{j\epsilon\omega_2} \mathbf{B}_{ij} = \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij} \{ \mathbf{S}_{ij} \mathbf{\Phi}_{(2)}^{-1} \mathbf{S}'_{ij} \}^{-1} \mathbf{X}_{(3)ij}$$

$$= \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij} \mathbf{\Phi}_{(2)ij}^{-1} \mathbf{S}_{ij} \,. \qquad (6.5.14)$$

Substitution of (6.5.13) and (6.5.14) in (6.5.12) gives

$$\mathbf{B}^*_{ij} = m^*_i \mathbf{\Phi}_{(2)}^{-1} + \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij} \mathbf{\Phi}_{(2)ij}^{-1} \mathbf{S}_{ij} \,. \qquad (6.5.15)$$

Using (6.5.11), (6.5.10) can be written as

$$\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{X}_{(3)i} = \mathbf{B}^*_{ij} [\mathbf{I} - \mathbf{C}^*_i \mathbf{A}^*_i]$$

which concludes the proof. □

PROPOSITION 6.5.3 (Submatrix $\mathbf{T}_{(2,2)}$)

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} (\mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*} \otimes \mathbf{Z}'_{(2)ij} \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij^*}) \mathbf{G}_m$$

$$= \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} \{ \mathbf{B}^*_{ij} (\delta_{jj^*} \mathbf{I} - \mathbf{C}^*_i \mathbf{B}^{*\prime}_{ij^*}) \otimes \mathbf{B}^*_{ij} (\delta_{jj^*} \mathbf{I} - \mathbf{C}^*_i \mathbf{B}^{*\prime}_{ij^*}) \} \mathbf{G}_m \qquad (6.5.16)$$

with

$$\delta_{jj^*} = 1 \text{ if } j = j^*$$
$$= 0 \text{ otherwise}$$

and $\mathbf{B}^*_{ij}$ and $\mathbf{C}^*_i$ as defined in (6.5.9) and (6.5.4).

Proof

From (6.4.22)

$$\mathbf{Z}'_{(2)ij}\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} = \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij^*} - \mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij^*} \tag{6.5.17}$$

with (cf. (6.5.12))

$$\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{B}^*_{ij}$$

and

$$\begin{aligned}\mathbf{Z}'_{(2)ij}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{Z}_{(2)ij^*} &= \mathbf{S}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{S}_{ij} \text{ if } \quad j = j^* \\ &= 0 \text{ otherwise .}\end{aligned} \tag{6.5.18}$$

But (cf. (6.5.13))

$$\mathbf{S}'_{ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{S}_{ij} = \mathbf{B}_{ij} \tag{6.5.19}$$

so that (cf. (6.5.12), (6.5.18) and (6.5.19)) equation (6.5.18) can be rewritten as

$$\begin{aligned}\mathbf{Z}'_{(2)ij}\mathbf{V}_i^{-1}\,\mathbf{Z}_{(2)ij^*} &= \delta_{jj^*}\,\mathbf{B}^*_{ij} - \mathbf{B}^*_{ij}\,\mathbf{C}^*_i\,\mathbf{B}^{*\prime}_{ij^*} \\ &= \mathbf{B}^*_{ij}\,[\delta_{jj^*}\,\mathbf{I} - \mathbf{C}^*_i\,\mathbf{B}^{*\prime}_{ij^*}]\end{aligned}$$

which concludes the proof. □

THEOREM 6.5.1

In the case of missing measurements on the q dependent variables

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{T}_{(3,3)} & \mathbf{T}'_{(2,3)} \\ \mathbf{T}_{(2,3)} & \mathbf{T}_{(2,2)} \end{bmatrix}$$

where

$$\mathbf{T}_{(3,3)} = \mathbf{G}'_q \{\mathbf{A}^*_i [\mathbf{I} - \mathbf{C}^*_i \mathbf{A}^*_i] \otimes \mathbf{A}^*_i [\mathbf{I} - \mathbf{C}^*_i \mathbf{A}^*_i]\} \mathbf{G}_q \,,$$

$$\mathbf{T}_{(2,3)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \{\mathbf{B}^*_{ij} [\mathbf{I} - \mathbf{C}^*_i \mathbf{A}^*_i] \otimes \mathbf{B}^*_{ij} [\mathbf{I} - \mathbf{C}^*_i \mathbf{A}^*_i]\} \mathbf{G}_q$$

and

$$\mathbf{T}_{(2,2)} = \mathbf{G}'_m \sum_{j=1}^{n_i} \sum_{j^*=1}^{n_i} \{\mathbf{B}^*_{ij} (\delta_{jj^*} \mathbf{I} - \mathbf{C}^*_i \mathbf{B}^{*\prime}_{ij^*}) \otimes \mathbf{B}^*_{ij} (\delta_{jj^*} \mathbf{I} - \mathbf{C}^*_i \mathbf{B}^{*\prime}_{ij^*})\}$$

with

$$\mathbf{A}^*_i = m^*_i \, \boldsymbol{\Phi}^{-1}_{(2)} + \sum_{j \epsilon \omega_2} \mathbf{S}'_{ij} \, \boldsymbol{\Phi}^{-1}_{(2)ij} \, \mathbf{S}_{ij} \,,$$

$$\mathbf{C}^*_i = (\boldsymbol{\Phi}^{-1}_{(3)} + \mathbf{A}^*_i)^{-1}$$

and

$$\mathbf{B}^*_{ij} = m^*_i \, \boldsymbol{\Phi}^{-1}_{(2)} + \mathbf{S}'_{ij} \, \boldsymbol{\Phi}^{-1}_{(2)ij} \, \mathbf{S}_{ij} \,.$$

Proof

The proof follows directly from Propositions 6..5.1 to 6.5.3. □

6.5.2: SIMPLIFICATION OF THE COMPONENTS OF THE COEFFICIENT VECTOR

Modified expressions for the vectors $\mathbf{q}_{(3)}$ and $\mathbf{q}_{(2)}$ (cf. (6.4.27)) for the situation in which some response-variable measurements are missing, are given in Propositions 6.5.4 and 6.5.5. Results are summarized in Theorem 6.5.2.

PROPOSITION 6.5.4 (Subvector $\mathbf{q}_{(3)}$)

$$\mathbf{q}_{(3)} = \mathbf{G}_q' \ \text{vec}\ (\mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \mathbf{Y}_i \ \mathbf{V}_i^{-1} \mathbf{X}_{(3)i})$$

$$= \mathbf{G}_q' \,\text{vec}\ (\mathbf{a}_i^* \ \mathbf{a}_i^{*\prime}) \tag{6.5.20}$$

where

$$\mathbf{a}_i^* = [\mathbf{I} - \mathbf{A}_i^* \mathbf{C}_i^*] \ \mathbf{d}_i \tag{6.5.21}$$

with

$$\mathbf{d}_i = \mathbf{\Phi}_{(2)}^{-1} \sum_{j \epsilon \omega_1} \mathbf{e}_{ij} + \sum_{j \epsilon \omega_2} \mathbf{S}_{ij}' \mathbf{\Phi}_{(2)ij}^{-1} \mathbf{e}_{ij} \tag{6.5.22}$$

and $\mathbf{Y}_i$ is given by (6.4.28).

Proof

Let

$$\mathbf{a}_i^* = \mathbf{X}_{(3)i}' \mathbf{V}_i^{-1} \mathbf{e}_i \tag{6.5.23}$$

with $\mathbf{e}_{ij}$ defined by (6.4.35). From (6.3.2) and (6.4.13) it follows that

$$\mathbf{a}_i^* = \mathbf{X}_{(3)i}' \mathbf{\Lambda}_i^{-1} \mathbf{e}_i - \mathbf{X}_{(3)i}' \mathbf{\Lambda}_i^{-1} \mathbf{X}_{(3)i} \mathbf{C}_i \ \mathbf{X}_{(3)i}' \mathbf{\Lambda}_i^{-1} \mathbf{e}_i \tag{6.5.24}$$

where (cf. (6.5.3) and (6.5.5))

$$\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} = \mathbf{A}_i^* \tag{6.5.25}$$

and

$$\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{e}_i = \sum_{j\epsilon\omega_1} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij} + \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij}\,. \tag{6.5.26}$$

The first term of (6.5.26) reduces to

$$\begin{aligned} \sum_{j\epsilon\omega_1} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij} &= \sum_{j\epsilon\omega_1} \mathbf{S}'_{ij}\,\{\,\mathbf{S}_{ij}\,\boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{S}'_{ij}\,\}^{-1}\,\mathbf{e}_{ij} \\ &= \sum_{j\epsilon\omega_1} \boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{e}_{ij} \\ &= \boldsymbol{\Phi}_{(2)}^{-1} \sum_{j\epsilon\omega_1} \mathbf{e}_{ij}\,. \end{aligned} \tag{6.5.27}$$

The second term of (6.5.26) can be written as

$$\begin{aligned} \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{e}_{ij} &= \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij}\,\{\,\mathbf{S}_{ij}\,\boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{S}'_{ij}\,\}^{-1}\,\mathbf{e}_{ij} \\ &= \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij}\,\boldsymbol{\Phi}_{(2)ij}^{-1}\,\mathbf{e}_{ij}\,. \end{aligned} \tag{6.5.28}$$

Substitution of (6.5.27) and (6.5.28) in (6.5.26) gives

$$\begin{aligned} \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_i^{-1}\,\mathbf{e}_i &= \boldsymbol{\Phi}_{(2)}^{-1} \sum_{j\epsilon\omega_1} \mathbf{e}_{ij} + \sum_{j\epsilon\omega_2} \mathbf{S}'_{ij}\,\boldsymbol{\Phi}_{(2)ij}^{-1}\,\mathbf{e}_{ij} \\ &= \sum_{j\epsilon\omega_1} \mathbf{d}_{ij} + \sum_{j\epsilon\omega_2} \mathbf{d}_{ij}\,. \end{aligned} \tag{6.5.29}$$

with

$$\begin{aligned} \mathbf{d}_{ij} &= \boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{e}_{ij} \qquad \text{if } j\epsilon\omega_1 \\ &= \mathbf{S}'_{ij}\,\boldsymbol{\Phi}_{(2)ij}^{-1}\,\mathbf{e}_{ij} \quad \text{if } j\epsilon\omega_2\,. \end{aligned} \tag{6.5.30}$$

Let

$$\mathbf{d}_i = \sum_{j=1}^{n_i} \mathbf{d}_{ij} \tag{6.5.31}$$

then (6.5.24) can be rewritten as (cf. (6.5.25) and (6.5.31))

$$\mathbf{a}_i^* = \mathbf{d}_i - \mathbf{A}_i^* \mathbf{C}_i^* \mathbf{d}_i$$

$$= [\, \mathbf{I} - \mathbf{A}_i^* \mathbf{C}_i^* \,] \, \mathbf{d}_i$$

which concludes the proof. □

PROPOSITION 6.5.5 (Subvector $\mathbf{q}_{(2)}$)

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \text{ vec } \left(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij} \right)$$

$$= \mathbf{G}_m' \sum_{j=1}^{n_i} \{\text{vec } (\mathbf{b}_{ij}^* \mathbf{b}_{ij}^{*\prime})\} \tag{6.5.32}$$

where

$$\mathbf{b}_{ij}^* = \mathbf{d}_{ij} - \mathbf{B}_{ij}^* \mathbf{C}_i^* \mathbf{d}_i \tag{6.5.33}$$

and $\mathbf{d}_{ij}$ and $\mathbf{d}_i$ are defined by (6.5.30) and (6.5.31) respectively.

Proof

From (6.4.32)

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \text{ vec } \left(\sum_{j=1}^{n_i} \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Z}_{(2)ij} \right) . \tag{6.5.34}$$

Let

$$\mathbf{b}_{ij}^* = \mathbf{Z}_{(2)ij}' \mathbf{V}_i^{-1} \mathbf{e}_i \tag{6.5.35}$$

with $\mathbf{e}_{ij}$ defined by (6.4.35).

From (6.3.2) and (6.4.13) it follows that $Z'_{(2)ij} V_i^{-1} e_i$ can be written as

$$Z'_{(2)ij} V_i^{-1} e_i = Z'_{(2)ij} \Lambda_i^{-1} e_i - Z'_{(2)ij} \Lambda_i^{-1} X_{(3)i} C_i X'_{(3)i} \Lambda_i^{-1} e_i \tag{6.5.36}$$

with (cf. (6.5.12))

$$Z'_{(2)ij} \Lambda_i^{-1} X_{(3)i} = B^*_{ij} \tag{6.5.37}$$

and, in the case of complete data,

$$\begin{aligned} Z'_{(2)ij} \Lambda_i^{-1} e_i &= S'_{ij} \{S_{ij} \Phi_{(2)}^{-1} S'_{ij}\}^{-1} e_{ij} \\ &= \Phi_{(2)}^{-1} e_{ij} \end{aligned} \tag{6.5.38}$$

since $S_{ij} = I_q$. When some of the response measurements are missing, $S_{ij} \neq I_q$ and

$$\begin{aligned} Z'_{(2)ij} \Lambda_i^{-1} e_i &= S'_{ij} \{S_{ij} \Phi_{(2)}^{-1} S'_{ij}\}^{-1} e_{ij} \\ &= S'_{ij} \Phi_{(2)ij}^{-1} e_{ij} . \end{aligned} \tag{6.5.39}$$

$Z'_{(2)ij} \Lambda_i^{-1} e_i$ can then be written as (cf. (6.5.38) and (6.5.39))

$$\begin{aligned} Z'_{(2)ij} \Lambda_i^{-1} e_i &= \Phi_{(2)}^{-1} e_{ij} && \text{if } j \epsilon \omega_1 \\ &= S'_{ij} \Phi_{(2)ij}^{-1} e_{ij} && \text{if } j \epsilon \omega_2 \\ &= d_{ij} \end{aligned}$$

so that (cf. (6.5.31)), (6.5.36) can be rewritten as

$$b^*_{ij} = d_{ij} - B^*_{ij} C^*_i d_i$$

which concludes the proof. □

THEOREM 6.5.2

$$\mathbf{q}_i = \begin{bmatrix} \mathbf{q}_{(3)} \\ \mathbf{q}_{(2)} \end{bmatrix}$$

where

$$\mathbf{q}_{(3)} = \mathbf{G}_q' \text{ vec } (\mathbf{a}_i^* \ \mathbf{a}_i^{*\prime})$$

and

$$\mathbf{q}_{(2)} = \mathbf{G}_m' \sum_{j=1}^{n_i} \{\text{vec } (\mathbf{b}_{ij}^* \ \mathbf{b}_{ij}^{*\prime}),$$

with

$$\mathbf{a}_i^* = [\ \mathbf{I} \ - \ \mathbf{A}_i^* \ \mathbf{C}_i^*] \ \mathbf{d}_i \ ,$$

$$\mathbf{b}_{ij}^* = \mathbf{d}_{ij} - \mathbf{B}_{ij}^* \ \mathbf{C}_i^* \ \mathbf{d}_i$$

and

$$\mathbf{d}_i = \boldsymbol{\Phi}_{(2)}^{-1} \sum_{j \epsilon \omega_1} \mathbf{e}_{ij} + \sum_{j \epsilon \omega_2} \mathbf{S}_{ij}' \ \boldsymbol{\Phi}_{(2)ij}^{-1} \ \mathbf{e}_{ij} \ .$$

The matrices $\mathbf{A}_i^*$, $\mathbf{B}_{ij}^*$ and $\mathbf{C}_i^*$ are defined in (6.5.3), (6.5.15) and (6.5.4) respectively.

Proof

The proof follows directly from Propositions 6.5.4 and 6.5.5. □

6.5.3 SIMPLIFICATION OF THE WEIGHT MATRIX AND COEFFICIENT VECTOR (FIXED PART OF THE MODEL)

In this section attention is focused on the calculation of $\hat{\boldsymbol{\beta}}$ in the case of missing response variable measurements.

From (6.3.2) and (6.4.13) it follows that $\mathbf{X}'_{(f)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(f)i}$ can be written as

$$\mathbf{X}'_{(f)i}\,\mathbf{V}_i^{-1}\,\mathbf{X}_{(f)i} = \mathbf{X}'_{(f)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(f)i} - \mathbf{X}'_{(f)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\;\mathbf{X}'_{(3)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(f)i} \tag{6.5.40}$$

where

$$\mathbf{X}'_{(f)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(f)i} = \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(f)ij}\,.$$

It further follows that

$$\begin{aligned} \mathbf{X}'_{(f)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(f)ij} &= \mathbf{X}'_{(f)ij}\,\mathbf{\Phi}_{(2)}^{-1}\,\mathbf{X}_{(f)ij} && \text{if } j\epsilon\omega_1 \\ &= \mathbf{X}'_{(f)ij}\,\mathbf{\Phi}_{(2)ij}^{-1}\,\mathbf{X}_{(f)ij} && \text{if } j\epsilon\omega_2\,. \end{aligned} \tag{6.5.41}$$

Similarly,

$$\begin{aligned} \mathbf{X}'_{(f)i}\,\mathbf{\Lambda}_i^{-1}\,\mathbf{X}_{(3)i} &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij} \\ &= \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{S}_{ij}\,, \end{aligned}$$

from which it follows that

$$\begin{aligned} \mathbf{X}'_{(f)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(3)ij} &= \mathbf{X}'_{(f)ij}\,\mathbf{\Phi}_{(2)}^{-1} && \text{if } j\epsilon\omega_1 \\ &= \mathbf{X}'_{(f)ij}\,\mathbf{\Phi}_{(2)ij}^{-1}\,\mathbf{S}_{ij} && \text{if } j\epsilon\omega_2\,. \end{aligned} \tag{6.5.42}$$

Let

$$\mathbf{F}^*_{ij} = \mathbf{X}'_{(f)ij}\,\mathbf{\Lambda}_{ij}^{-1}\,\mathbf{X}_{(f)ij} \tag{6.5.43}$$

and

$$\mathbf{F}^*_i = \sum_{j=1}^{n_i} \mathbf{F}^*_{ij}\,. \tag{6.5.44}$$

Also let

$$\mathbf{G}^*_{ij} = \mathbf{X}'_{(f)ij}\,\boldsymbol{\Lambda}^{-1}_{ij}\,\mathbf{X}_{(3)ij}$$

$$= \mathbf{X}'_{(f)ij}\,\boldsymbol{\Phi}^{-1}_{(2)} \qquad \text{if } j\epsilon\omega_1$$

$$= \mathbf{X}'_{(f)ij}\,\boldsymbol{\Phi}^{-1}_{(2)ij}\,\mathbf{S}_{ij} \quad \text{if } j\epsilon\omega_2 \tag{6.5.45}$$

and

$$\mathbf{G}^*_i = \sum_{j=1}^{n_i} \mathbf{G}^*_{ij}\,. \tag{6.5.46}$$

Equation (6.5.40) can then be written as (cf. (6.5.41) to (6.5.46))

$$\mathbf{X}'_{(f)i}\,\mathbf{V}^{-1}_i\,\mathbf{X}_{(f)i} = \mathbf{F}^*_i - \mathbf{G}^*_i\,\mathbf{C}^*_i\,\mathbf{G}^*_i\,. \tag{6.5.47}$$

Finally, the simplification of the term $\mathbf{X}'_{(f)i}\,\mathbf{V}^{-1}_i\,\mathbf{y}_i$ is considered (cf. (6.3.3)). Using (6.3.2) and (6.4.13), it follows that

$$\mathbf{X}'_{(f)i}\,\mathbf{V}^{-1}_i\,\mathbf{y}_i = \mathbf{X}'_{(f)i}\,\boldsymbol{\Lambda}^{-1}_i\,\mathbf{y}_i - \mathbf{X}'_{(f)i}\,\boldsymbol{\Lambda}^{-1}_i\,\mathbf{X}_{(3)i}\,\mathbf{C}_i\,\mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}^{-1}_i\,\mathbf{y}_i \tag{6.5.48}$$

with (cf. (6.5.46))

$$\mathbf{X}'_{(f)i}\,\boldsymbol{\Lambda}^{-1}_i\,\mathbf{X}_{(3)i} = \mathbf{G}^*_i$$

and

$$\mathbf{X}'_{(f)i}\,\boldsymbol{\Lambda}^{-1}_i\,\mathbf{y}_i = \sum_{j=1}^{n_i} \mathbf{X}'_{(f)ij}\,\boldsymbol{\Lambda}^{-1}_{ij}\,\mathbf{y}_{ij} \tag{6.5.49}$$

where

$$\mathbf{X}'_{(f)ij}\,\boldsymbol{\Lambda}^{-1}_{ij}\,\mathbf{y}_{ij} = \mathbf{X}'_{(f)ij}\,\boldsymbol{\Phi}^{-1}_{(2)}\,\mathbf{y}_{ij} \qquad \text{if } j\epsilon\omega_1$$

$$= \mathbf{X}'_{(f)ij}\,\boldsymbol{\Phi}^{-1}_{(2)ij}\,\mathbf{y}_{ij} \qquad \text{if } j\epsilon\omega_2\,. \tag{6.5.50}$$

Let

$$\mathbf{h}_{ij}^{*} = \mathbf{X}'_{(f)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \tag{6.5.51}$$

and

$$\mathbf{h}_{i}^{*} = \sum_{j=1}^{n_i} \mathbf{h}_{ij}^{*} \ . \tag{6.5.52}$$

It also follows that

$$\begin{aligned} \mathbf{X}'_{(3)i}\,\boldsymbol{\Lambda}_{i}^{-1}\,\mathbf{y}_{i} &= \sum_{j=1}^{n_i} \mathbf{X}'_{(3)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \\ &= \sum_{j=1}^{n_i} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \end{aligned} \tag{6.5.53}$$

where

$$\begin{aligned} \mathbf{S}'_{ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} &= \boldsymbol{\Phi}_{(2)}^{-1}\,\mathbf{y}_{ij} && \text{if } j\epsilon\omega_1 \\ &= \mathbf{S}'_{ij}\,\boldsymbol{\Phi}_{(2)ij}^{-1}\,\mathbf{y}_{ij} && \text{if } j\epsilon\omega_2 \ . \end{aligned} \tag{6.5.54}$$

Let

$$\mathbf{k}_{ij}^{*} = \mathbf{X}'_{(3)ij}\,\boldsymbol{\Lambda}_{ij}^{-1}\,\mathbf{y}_{ij} \tag{6.5.55}$$

and

$$\mathbf{k}_{i}^{*} = \sum_{j=1}^{n_i} \mathbf{k}_{ij}^{*} \ . \tag{6.5.56}$$

Using (6.5.46) and (6.5.49) to (6.5.56), (6.5.48) can be written as

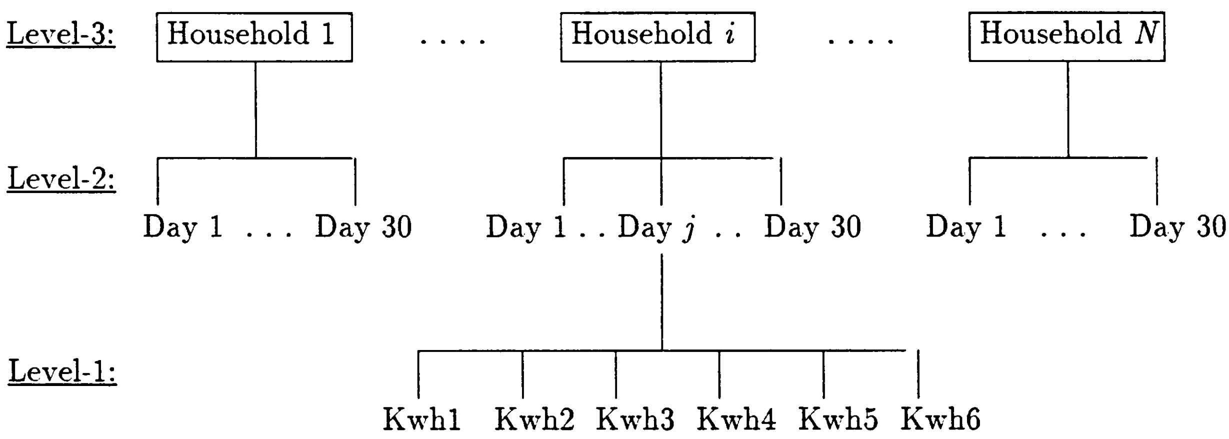
$$\mathbf{X}'_{(f)i}\,\mathbf{V}_{i}^{-1}\,\mathbf{y}_{i} = \mathbf{h}_{i}^{*} - \mathbf{G}_{i}^{*}\,\mathbf{C}_{i}^{*}\,\mathbf{k}_{i}^{*} \ . \tag{6.5.57}$$

6.6 PRACTICAL APPLICATIONS

In this section two examples are given. The first example is concerned with the analysis of electricity consumption patterns. In the second example the data described in Example 3.5.1 is examined in the context of a multivariate level-3 model.

Example 6.6.1

In this example the analysis of a large data set from an Electricity Supply Commission survey of 293 black households in Naledi, Soweto is considered (with the kind permission of Rob Surtees, Load Research, ESKOM, South Africa). The sample is schematically represented as follows:



The 293 households represent the level-3 units. The level-2 units are the days on which the measurements were taken, while the level-1 units are the hourly Kwh readings, taken at 7h00, 8h00, 9h00, 19h00, 20h00 and 21h00 respectively. Clusters of level-1 units are present within each level-2 unit and clusters of level-2 units are present within each level-3 unit. Thus, a maximum of $6 \times 30 \times 293 = 52\ 740$ units are included.

Additional information available includes information on the electrical appliances utilized by each household and, as the exact dates during June 1993 on which measurements were taken are available, it is also known which measurements refer to weekdays and which to weekend days.

The purpose of the analysis is to estimate the variation in peak-hour electricity consumption over households. Estimation is based on the type of appliance utilized and whether a particular response measurement was made on a weekday or a weekend day. A multivariate multilevel model approach is adopted to accommodate the hierarchical structure of the data which is evident from the schematic representation given above.

The FORTRAN program MULTVAR was used for the analysis. The diskette which accompanies the dissertation contains the program MULTVAR. Also included on the diskette is an ASCII file named 'README.DOC'. This file contains information on how to run the examples. See Chapter 10 for additional information on this program.

To estimate the variation in peak-hour electricity consumption at 7h00, 8h00, 9h00, 19h00, 20h00 and 21h00 respectively the following predictors were included as dummy variables:

(i) Utilization of stove (STOVEKW*k*)
(ii) Utilization of geyser or urn (GURNKW*k*)
(iii) Utilization of hotplate (HPLATKW*k*)
(iv) Utilization of heater (HEATRKW*k*)
(v) Utilization of freezer (FREEZKW*k*)
(vi) Day of week on which measurement was made (DAYWEKW*k*)

Note that a value of '1' is assigned to STOVEKW*k*, *k*=1, 2, . . . , 6, provided that the electricity consumption measurement was made on occasion *k* and a stove was utilized. Otherwise, a value of '0' is assigned. Coding for variables (ii) to (v) is similar. The predictor DAYWEKW*k*, k=1, 2, . . . , 6, assumes a value of '1' if the *k*-th measurement was made on a weekday and '0' otherwise.

Part of the computer output is given below and is briefly discussed with regard to the fixed and random parts of the model respectively.

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	STD. ERR.	Z – VALUE	PR > |Z|
STOVEKW1	0.5476	0.1188	4.6110	0.0000
STOVEKW2	0.7631	0.1326	5.7532	0.0000
STOVEKW3	0.7810	0.1281	6.0986	0.0000
STOVEKW4	1.1489	0.1613	7.1238	0.0000
STOVEKW5	1.2927	0.1745	7.4067	0.0000
STOVEKW6	1.2555	0.1676	7.4916	0.0000
GURNKW1	0.0834	0.1170	0.7125	0.4761
GURNKW2	0.1756	0.1309	1.3420	0.1796
GURNKW3	0.3166	0.1263	2.5063	0.0122
GURNKW4	0.5561	0.1591	3.4959	0.0005
GURNKW5	0.4354	0.1723	2.5272	0.0115
GURNKW6	0.4168	0.1655	2.5191	0.0118
HPLATKW1	0.2229	0.0880	2.5344	0.0113
HPLATKW2	0.2827	0.0981	2.8818	0.0040
HPLATKW3	0.4370	0.0947	4.6136	0.0000
HPLATKW4	0.7324	0.1193	6.1380	0.0000
HPLATKW5	0.5833	0.1290	4.5230	0.0000
HPLATKW6	0.5349	0.1238	4.3206	0.0000
HEATRKW1	0.2450	0.0856	2.8613	0.0042
HEATRKW2	0.2782	0.0958	2.9034	0.0037
HEATRKW3	0.2163	0.0925	2.3383	0.0194
HEATRKW4	0.6080	0.1165	5.2211	0.0000
HEATRKW5	0.7940	0.1262	6.2911	0.0000
HEATRKW6	0.7153	0.1212	5.9019	0.0000
FREEZKW1	– 0.0818	0.0859	– 0.9532	0.3405
FREEZKW2	– 0.0244	0.0960	– 0.2542	0.7993
FREEZKW3	0.1140	0.0927	1.2304	0.2185

FREEZKW4	0.1726	0.1167	1.4791	0.1391
FREEZKW5	0.0913	0.1264	0.7223	0.4701
FREEZKW6	0.0496	0.1214	0.4085	0.6829
DAYWEKW1	0.5811	0.0252	23.0661	0.0000
DAYWEKW2	0.4295	0.0253	16.9778	0.0000
DAYWEKW3	– 0.0443	0.0247	– 1.7981	0.0722
DAYWEKW4	0.0449	0.0323	1.3932	0.1635
DAYWEKW5	0.1993	0.0312	6.3795	0.0000
DAYWEKW6	0.1275	0.0292	4.3747	0.0000

(ii) **Random part of the model:**

PARAMETER	$\hat{\tau}$	**STD.ERR.**	**Z – VALUE**	**PR > |Z|**
KW1,1(3)	0.4697	0.0297	15.8193	0.0000
KW2,1(3)	0.4320	0.0297	14.5572	0.0000
KW2,2(3)	0.5971	0.0372	16.0577	0.0000
KW3,1(3)	0.3507	0.0269	13.0207	0.0000
KW3,2(3)	0.5234	0.0336	15.5659	0.0000
KW3,3(3)	0.5559	0.0346	16.0490	0.0000
KW4,1(3)	0.3900	0.0328	11.8955	0,0000
KW4,2(3)	0.4878	0.0378	12.9214	0.0000
KW4,3(3)	0.4913	0.0369	13.2984	0.0000
KW4,4(3)	0.8756	0.0549	15.9458	0.0000
KW5,1(3)	0.4260	0.0356	11.9545	0.0000
KW5,2(3)	0.5327	0.0411	12.9700	0.0000
KW5,3(3)	0.4945	0.0392	12.6109	0.0000
KW5,4(3)	0.8806	0.0558	15.7780	0.0000
KW5,5(3)	1.0433	0.0645	16.1762	0.0000
KW6,1(3)	0.4244	0.0345	12.2848	0.0000
KW6,2(3)	0.5222	0.0397	13.1644	0.0000
KW6,3(3)	0.4950	0.0381	12.9971	0.0000
KW6,4(3)	0.8215	0.0531	15.4773	0.0000
KW6,5(3)	0.9661	0.0598	16.1665	0.0000

KW6,6(3)	0.9660	0.0595	16.2376	0.0000
KW1,1(2)	0.8585	0.0105	81.9444	0.0000
KW2,1(2)	0.1432	0.0076	18.8183	0.0000
KW2,2(2)	0.8825	0.0108	81.9498	0.0000
KW3,1(2)	0.0181	0.0073	2.4935	0.0127
KW3,2(2)	0.2234	0.0076	29.2990	0.0000
KW3,3(2)	0.8284	0.0101	81.9457	0.0000
KW4,1(2)	– 0.0012	0.0096	– 0.1213	0.9034
KW4,2(2)	– 0.0241	0.0098	– 2.4665	0.0136
KW4,3(2)	0.0206	0.0095	2.1761	0.0295
KW4,4(2)	1.4532	0.0177	81.9507	0.0000
KW5,1(2)	0.0390	0.0094	4.1635	0.0000
KW5,2(2)	0.0387	0.0095	4.0769	0.0000
KW5,3(2)	0.0472	0.0092	5.1342	0.0000
KW5,4(2)	0.1318	0.0122	10.7830	0.0000
KW5,5(2)	1.3700	0.0167	81.9550	0.0000
KW6,1(2)	0.0389	0.0087	4.4936	0.0000
KW6,2(2)	0.0236	0.0088	2.6966	0.0070
KW6,3(2)	0.0237	0.0085	2.7881	0.0053
KW6,4(2)	0.2421	0.0114	21.1597	0.0000
KW6,5(2)	0.3811	0.0114	33.4098	0.0000
KW6,6(2)	1.1697	0.0143	81.9493	0.0000

HOUSEHOLD LEVEL: ESTIMATE OF $\Phi_{(3)}$

	KW1-LEV3	KW2-LEV3	KW3-LEV3	KW4-LEV3	KW5-LEV3	KW6-LEV3
KW1-LEV3	0.4697					
KW2-LEV3	0.4320	0.5971				
KW3-LEV3	0.3507	0.5234	0.5559			
KW4-LEV3	0.3900	0,4878	0.4913	0.8756		
KW5-LEV3	0.4260	0.5327	0.4945	0.8806	1.0433	
KW6-LEV3	0.4244	0.5222	0.4950	0.8215	0.9661	0.9660

HOUSEHOLD LEVEL: CORRELATION MATRIX

	KW1-LEV3	KW2-LEV3	KW3-LEV3	KW4-LEV3	KW5-LEV3	KW6-LEV3
KW1-LEV3	1.0000					
KW2-LEV3	0.8158	1.0000				
KW3-LEV3	0.6862	0.9084	1.0000			
KW4-LEV3	0.6082	0.6747	0.7042	1.0000		
KW5-LEV3	0.6086	0.6749	0.6493	0.9214	1.0000	
KW6-LEV3	0.6300	0.6876	0.6755	0.8932	0.9624	1.0000

DAY LEVEL: ESTIMATE OF $\Phi_{(2)}$

	KW1-LEV2	KW2-LEV2	KW3-LEV2	KW4-LEV2	KW5-LEV2	KW6-LEV2
KW1-LEV2	0.8585					
KW2-LEV2	0.1432	0.8825				
KW3-LEV2	0.0181	0.2234	0.8284			
KW4-LEV2	-0.0012	-0.0241	0.0206	1.4532		
KW5-LEV2	0.0390	0.0387	0.0472	0.1318	1.3700	
KW6-LEV2	0.0389	0.0236	0.0237	0.2421	0.3811	1.1697

DAY LEVEL: CORRELATION MATRIX

	KW1-LEV2	KW2-LEV2	KW3-LEV2	KW4-LEV2	KW5-LEV2	KW6-LEV2
KW1-LEV2	1.0000					
KW2-LEV2	0.1646	1.0000				
KW3-LEV2	0.0215	0.2613	1.0000			
KW4-LEV2	-0.0010	0.0213	0.0188	1.0000		
KW5-LEV2	0.0359	0.0352	0.0443	0.0934	1.0000	
KW6-LEV2	0,0388	0.0233	0.0241	0.1857	0.3011	1.0000

(iii) **Convergence details:**

Convergence (IGLS – Algorithm) in 3 iterations

From the fixed part of the model it follows that the majority of coefficients, with the exception of the regression coefficients of the dummy variables FREEZKW1 to FREEZKW6, are highly significant. STOVEKW1 to STOVEKW6 have the largest coefficients. There are significant differences between weekday and weekend electricity consumption, except at 9h00 and 19h00. A study of the appliance coefficients indicate a peak demand for electricity at 19h00 (KW4).

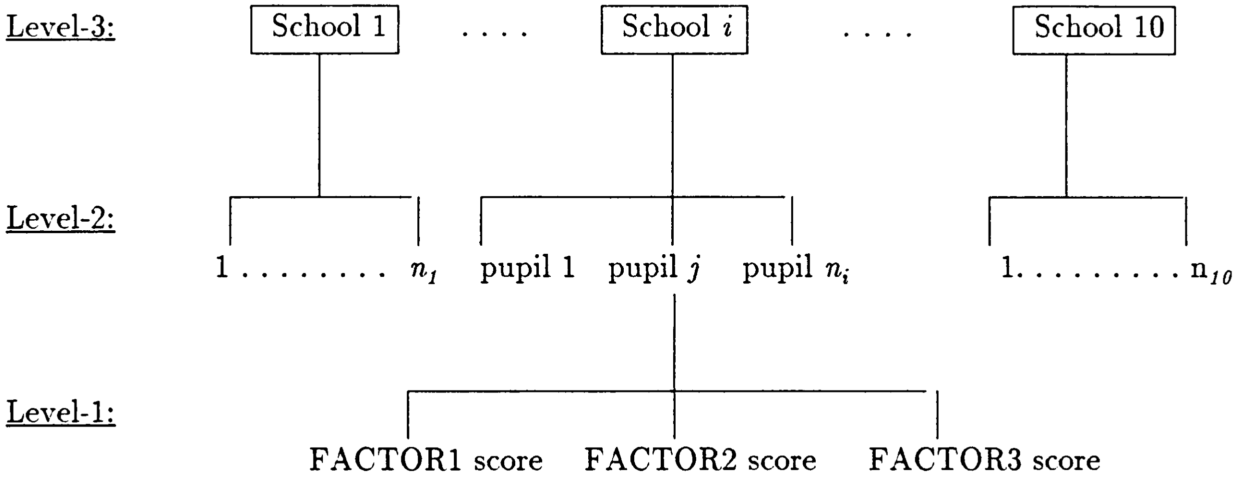
From the random part of the model it follows that, on level-3, there is highly significant variation in domestic demand for electricity and high correlations between the variables KW1-LEV3 to KW6-LEV3. Households which have a high demand for electricity at any given hour tend to also have a high demand on the other hours.

On a daily level (level-2) there is an even larger variation in electricity consumption. Correlations between the variables KW1-LEV2 to KW6-LEV2 are relatively small, indicating that the amount of electricity used on, for example, a Monday morning is not correlated with the amount of electricity used on a Wednesday evening.

The variances of the level-3 and level-2 matrices can be used to construct 95 % confidence intervals for the average electricity consumption during any of the peak hours.

Example 6.6.2

In this example the analysis of a data set obtained from a survey undertaken by the South African Department of National Health and Population Development is considered. The data set used contains information on the knowledge and perceptions of AIDS and related issues of 1702 pupils at 10 secondary schools and can be schematically represented as follows:



The 10 schools represent the level-3 units. The level-2 units are the pupils, while the level-1 units are the scores obtained by a pupil for FACTOR1, FACTOR2 and FACTOR3. FACTOR1 indicates a pupil's knowledge of proven means by which AIDS can be transmitted, FACTOR2 indicates a pupil's perception of HIV/AIDS susceptibility and FACTOR3 indicates a pupil's knowledge regarding the transmission of AIDS through casual contact.

The majority of questions answered by pupils were of a categorical nature with possible outcomes 'Yes' and 'No' with assigned values '1' and '0' respectively. Where a third category 'Do not know' was present, optimal scaling (Du Toit & Strasheim, 1987), was used to assign numerical values to the three categories. Subsequently a factor analysis (SAS/STAT, 1990), with VARIMAX rotation was done with respect to the scaled items in order to find subgroups of mutually correlated items. Tables 6.6.1 to 6.6.3 below give the items and corresponding factor loadings belonging to each of the three factors.

Table 6.6.1 FACTOR1: Knowledge of transmission of HIV/AIDS by proven means

Description of item	Factor loadings
1. One cannot get AIDS/HIV by having sex with someone who has AIDS/HIV.	0,2184
2. A pregnant woman can pass AIDS/HIV on to her unborn baby.	0,2885
3. One can get AIDS/HIV by using the same injection needles that other persons have used when injecting yourself with drugs.	0,1861
4. One can get AIDS/HIV by receiving blood that was donated by a person who has AIDS/HIV.	0,1568
5. A man can get AIDS/HIV by having sex with another man who has AIDS/HIV.	0,3442
6. AIDS/HIV is caused by a virus/germ.	0,2409
7. The body of a person with AIDS/HIV cannot defend itself against other diseases.	0,4547
8. AIDS/HIV cannot be spread when two uninfected people have sex.	0.2377
9. A woman cannot get AIDS/HIV by having sex with another woman who has AIDS/HIV .	0,3109
10. AIDS/HIV cannot be transmitted by semen.	0,3928
11. AIDS/HIV can be transmitted by vaginal fluids.	0,4225
12. AIDS/HIV can be transmitted by a person who is infected with AIDS/HIV.	0,2071
13. Is it true that someone who suffers from a sexually transmitted disease is more likely to get AIDS/HIV?	0,1688

Minimum value = 0
Maximum value = 13

Table 6.6.2 FACTOR2: Perceptions of HIV/AIDS susceptibility

Description of item	Factor loading
1. Can teenagers get AIDS?	0,3057
2. Is AIDS/HIV a disease from which mainly your population group suffers?	0,2961
3. Is AIDS/HIV a disease from which mainly other population groups suffer?	0,2246
4. Do you think that all people get AIDS/HIV?	0,3002
5. Do you think that AIDS/HIV is mainly a "gay disease/ homosexual disease"?	0,2038
6. Do you think it is possible that even your best friend could get AIDS/HIV?	0,1878

Minimum value = 0
Maximum value = 8

Table 6.6.3 FACTOR3 : Knowledge of HIV/AIDS transmission via casual contact

Description of item	Factor loading
1. One can get AIDS/HIV when someone who has AIDS/HIV coughs or sneezes on you.	0,6408
2. One can get AIDS/HIV by coming into contact with the perspiration of a person who has AIDS/HIV.	0,5831
3. One cannot get AIDS/HIV by swimming in the same swimming pool with someone who has AIDS/HIV.	0,5583
4. One can get AIDS/HIV by sharing food with someone	0,5420

who has AIDS/HIV.	
5. One cannot get AIDS/HIV from wearing clothes worn by a person who has AIDS/HIV.	0,5528
6. One cannot get AIDS/HIV by sitting on a toilet seat that has been used by a person who has AIDS/HIV.	0,4394
7. One can get AIDS/HIV by touching someone who has AIDS/HIV.	0,3448
8. One can get AIDS/HIV from mosquitoes that can pass it from one person to another.	0,3121

Minimum value = 0
Maximum value = 8

Additional information available includes the gender and standard or grade of each pupil.

The purpose of the analysis was to estimate the variation in factor scores over pupils and to determine the extent to which schools vary with respect to the response variable (this is particularly important since schools were selected to represent different segments of the South African population). It was also required to predict mean factor scores for the different standard gender combinations and to determine the significance of gender and standard with respect to the different factors.

Part of the computer output of the FORTRAN program MULTVAR is given below.

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	STD.ERR	Z-VALUE	PR > |Z|
CONS_FA1	-3.3184	1.4967	-2.2172	0.0266
CONS_FA2	1.2740	0.7797	1.6338	0.1023
CONS_FA3	-2.4094	1.4096	-1.7092	0.0874
FA1_ SEX	-0.6338	0.1095	-5.7867	0.0000

FA2_ SEX	-0.1592	0.0624	-2.5531	0.0107
FA3_ SEX	-0.1962	0.1031	-1.9027	0.0571
FA1_ STD	1.3991	0.2057	6.8020	0.0000
FA2_ STD	0.4267	0.1083	3.9397	0.0001
FA3_ STD	1.0355	0.1940	5.3377	0.0000

(ii) **Random part of the model:**

PARAMETER	$\hat{\tau}$	**STD.ERR**	**Z-VALUE**	**PR > |Z|**
FA1/1(3)	1.1359	0.5219	2.1764	0.0295
FA2/1(3)	0.4067	0.1896	2.1454	0.0319
FA2/2(3)	0.1524	0.0726	2.0994	0.0358
FA3/1(3)	0.9521	0.4557	2.0894	0.0367
FA3/2(3)	0.3610	0.1712	2.1086	0.0350
FA3/3(3)	0.9586	0.4410	2.1734	0.0298
FA1/1(2)	5.0074	0.1722	29.0861	0.0000
FA2/1(2)	0.8718	0.0727	11.9948	0.0000
FA2/2(2)	1.6332	0.0561	29.0875	0.0000
FA3/1(2)	1.5817	0.1209	13.0824	0.0000
FA3/2(2)	0.5994	0.0671	8.9376	0.0000
FA3/3(2)	4.4396	0.1526	29.0861	0.0000

SCHOOL LEVEL : ESTIMATE OF $\Phi_{(3)}$

	CONS_FA1	CONS_FA2	CONS_FA3
CONS_FA1	1.1359		
CONS_FA2	0.4067	0.1524	
CONS_FA3	0.9521	0.3610	0.9586

SCHOOL LEVEL : CORRELATION MATRIX

	CONS_FA1	CONS_FA2	CONS_FA3
CONS_FA1	1.0000		
CONS_FA2	0.9776	1.0000	
CONS_FA3	0.9124	0.9447	1.0000

PUPIL LEVEL: ESTIMATE OF $\Phi_{(2)}$

	CONS_FA1	CONS_FA2	CONS_FA3
CONS_FA1	5.0074		
CONS_FA2	0.8718	1.6332	
CONS_FA3	1.5817	0.5994	4.4396

PUPIL LEVEL : CORRELATION MATRIX

	CONS_FA1	CONS_FA2	CONS_FA3
CONS_FA1	1.0000		
CONS_FA2	0.3048	1.0000	
CONS_FA3	0.3355	0.2226	1.0000

From the fixed part of the output it follows that, with the exception of the coefficients of FA3_SEX and of the dummy variables CONS_FA2 and CONS_FA3, all coefficients are significant. From the coefficients of the variables FA1_SEX and FA2_SEX it can be seen that boys (coded '0') have an average higher score than girls (coded '1').

The expected FACTOR1 score for a male pupil in Standard 7 is calculated as

$$\hat{fa1} = -\ 3.3184\ -\ 0.6338(0) + 1.3391(7) = 6.4753$$

while for a female pupil in Standard 7

$$\hat{fa1} = -\ 3.3184\ -\ 0.6338(1) + 1.3391(7) = 5.8415$$

which illustrates both the difference between the two sexes and a lack of knowledge with regard to the transmission of AIDS/HIV through proven means. From Table 6.6.1 the maximum score for FACTOR1 is 13.

The coefficients of FA1_STD, FA2_STD and FA3_STD are all positive, implying a higher level of knowledge with respect to any of the factors for a pupil in a higher standard.

As far as the random part of the output is concerned, it can be seen that all random coefficients are significant at the 5 % level of significance. The significant variation over schools (level-3) was anticipated, as schools were selected at the sampling stage to represent different segments of the South African population. Variation over pupils (level-2) is larger than in the case of schools (level-3). The high variation over pupils shows a varying level of knowledge over pupils, indicating the need for further education with respect to AIDS/HIV related issues.

From the level-3 correlation matrix CONS_FA1, CONS_FA2 and CONS_FA3 are highly correlated. If a school has a high Factor 1 score, it would thus also have a high score on the other factors. Correlations between the variables CONS_FA1, CONS_FA2 and CONS_FA3 on a pupil level are fairly low, underlining gaps in their knowledge of AIDS/HIV.

6.7 SUMMARY

The analysis of models with two or more continuous response variables were considered in this chapter by the introduction of a multivariate multilevel model. The mathematical implications of missing data were illustrated. The theoretical results derived were implemented in the FORTRAN program MULTVAR included on the accompanying diskette.

Two practical examples were given. In these examples the coefficients of variables included in the fixed part of the model were not allowed to vary across the levels of the hierarchy. If models with complex variance structures were to be fitted to the data, the general model discussed in Chapter 5 could be used. Note, however, that the estimation procedure given in Chapter 5 was computationally less efficient than the procedure described in Chapter 6.

CHAPTER 7
MULTILEVEL MODELS FOR CATEGORICAL RESPONSE VARIABLES

7.1 INTRODUCTION

In the last few decades a wide variety of methods for the analysis of categorical data have been proposed. Many of these are generalizations of continuous data analysis methods (see for example Bishop, Fienberg & Holland, 1975, and Agresti, 1990) .

Analysis of variance (ANOVA) refers to the analysis of means and the partitioning of variation into various sources. In the analysis of categorical data, using the continuous data analysis approach, 'analysis of variance' is used to denote the analysis of response functions and the partitioning of variation among those functions into various sources (see Introduction to PROC CATMOD, SAS/STAT, 1990).

The response functions may be functions of the marginal probabilities, cumulative logits (see McCullagh, 1980), or other functions that incorporate the essential information from the dependent variables.

In this chapter it will be assumed that the response functions are the natural logarithms of the ratio of cell frequencies in a contingency table. The derived results, however, may be extended to include the analysis of categorical data with ordered categories. See for example Goodman (1979) and Stoker (1982) for a review of methods for the analysis of ordinal categorical data.

In the preceding chapters the analysis of data with a continuous response variable was considered. A general level-3 model which allowed for complex variation on all three levels of the hierarchy was discussed in Chapter 5. The theoretical framework developed in Chapter 5 facilitates the analysis of hierarchical data with categorical response variables.

In this chapter the multilevel analysis of categorical response variables will be discussed. A review of logit modelling will be given in Section 7.2, while a level-2 logit model will be introduced in Section 7.3. In Section 7.4 a level-3 logit model will be introduced and a multivariate multilevel model for handling more that one categorical response variable will be developed in Section 7.5.

Three practical applications will be given in Section 7.6 with conclusions in Section 7.7.

7.2 THE LOGIT MODEL

7.2.1 DEFINITION OF THE LOGIT MODEL

Consider the following two-way table in which the response variable has 'yes' and 'no' categories and where the i-th row of the table contains the number of 'yes' and the number of 'no' responses of respondents from subpopulation i, $i = 1, 2, \ldots, s$.

Subpopulation	**Response**		**Total**
	yes	no	
1	$f_{1,1}$	$f_{1,2}$	$f_{1.}$
2	$f_{2,1}$	$f_{2,2}$	$f_{2.}$
.	.	.	.
.	.	.	.
.	.	.	.
s	$f_{s,1}$	$f_{s,2}$	$f_{s.}$

For each subpopulation i the probability π_{ij} of the j-th response ($j = 1,2$) occurring is estimated by $p_{ij} = \frac{f_{i,j}}{f_{i}}$. These estimates are used to obtain the elements of the response vector $\mathbf{y}$, where the elements of $\mathbf{y}$ are assumed to be a function of the p_{ij}'s. It is assumed (see for example du Toit & Lamprecht, 1986) that the response vector $\mathbf{y}$ can be expressed in the form

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \mathbf{e} \tag{7.2.1}$$

where the elements of $\mathbf{X}$ depend on whether provision is made for the inclusion of an intercept or constant effect and on the way in which subpopulations are formed.

Example 7.2.1 illustrates the creation of such a matrix for a given set of p predictors. The elements of the vector $\boldsymbol{\beta}$ are fixed, but unknown, parameters to be estimated.
The theoretical cell probabilities are given by

Subpopulation	**Response**		**Total**
	yes	no	
1	π_{11}	π_{12}	$\pi_{1.}$
2	π_{21}	π_{22}	$\pi_{2.}$
.	.	.	.
.	.	.	.
.	.	.	.
s	π_{s1}	π_{s2}	$\pi_{s.}$

and satisfy the condition

$$\sum_{j=1}^{2} \pi_{ij} = 1 \qquad i = 1, 2, \ldots, n\,;\, j = 1, 2\,. \tag{7.2.2}$$

Denote the $(2s \times 1)$ vector of cell probabilities by $\boldsymbol{\pi}$, where

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \cdots\cdots \\ \pi_{21} \\ \pi_{22} \\ \cdots\cdots \\ \cdot \\ \cdot \\ \cdot \\ \cdots\cdots \\ \pi_{s1} \\ \pi_{s2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{\pi}_s \end{bmatrix}. \tag{7.2.3}$$

The elements of the corresponding estimator $\mathbf{p}$ of $\boldsymbol{\pi}$ are based on the sample proportions $\mathrm{p}_{ij} = \dfrac{f_{i,j}}{f_{i.}}$, and hence $\mathbf{p}$ is given by

$$\mathbf{p} = \begin{bmatrix} \frac{f_{1,1}}{f_{1.}} \\ \frac{f_{1,2}}{f_{1.}} \\ \cdots\cdots \\ \frac{f_{2,1}}{f_{2.}} \\ \frac{f_{2,2}}{f_{2.}} \\ \cdots\cdots \\ \cdot \\ \cdot \\ \cdot \\ \cdots\cdots \\ \frac{f_{s,1}}{f_{s.}} \\ \frac{f_{s,2}}{f_{s.}} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{p}_s \end{bmatrix}. \qquad (7.2.4)$$

Under the assumption that the underlying distribution is product multinomial (Agresti, 1990), it follows that

$$\mathrm{E}\left[\frac{f_{i,j}}{f_{i.}}\right] = \pi_{ij} \qquad i = 1,2, \ldots, s\,;\, j = 1, 2$$

$$\mathrm{Var}\left[\frac{f_{i,j}}{f_{i.}}\right] = \pi_{ij}(1 - \pi_{ij})/f_{i.} \qquad i = 1, 2, \ldots, s\,;\, j = 1, 2$$

$$\mathrm{Cov}\left[\frac{f_{i,j}}{f_{i.}}, \frac{f_{i,k}}{f_{i.}}\right] = -\pi_{ij}\,\pi_{ik}/f_{i.} \qquad i = 1, 2, \ldots, s\,;\, j = 1, 2$$

$$\mathrm{Cov}\left[\frac{f_{i,j}}{f_{i.}}, \frac{f_{l,k}}{f_{l.}}\right] = 0 \qquad \begin{array}{l} i = 1, 2, \ldots, s\,;\, j = 1, 2 \\ l = 1, 2, \ldots, s\,;\, k = 1, 2 \\ i \neq l;\; j \neq k\,. \end{array}$$

From the above it follows that

$$E(\mathbf{p}_i) = \boldsymbol{\pi}_i \qquad i = 1, 2, \ldots, s \qquad (7.2.5)$$

and

$$\begin{aligned}\mathrm{Cov}(\mathbf{p}_i, \mathbf{p}_i') &= \frac{1}{f_{i\cdot}} (\mathrm{Diag}(\boldsymbol{\pi}_i - \boldsymbol{\pi}_i \boldsymbol{\pi}_i') \\ &= \mathbf{V}_i \ . \end{aligned} \qquad (7.2.6)$$

Thus,

$$\mathrm{Cov}(\mathbf{p}, \mathbf{p}') = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \mathbf{V}_n \end{bmatrix} = \mathbf{V} \qquad (7.2.7)$$

with $\mathbf{V}_i$ defined above.

For the logit model, the vector of responses is

$$\mathbf{y} = \begin{bmatrix} ln \frac{p_{11}}{p_{12}} \\ ln \frac{p_{21}}{p_{22}} \\ . \\ . \\ . \\ ln \frac{p_{s1}}{p_{s2}} \end{bmatrix} \qquad (7.2.8)$$

with

$$E(\mathbf{y}) = \mathbf{X}\,\boldsymbol{\beta} \ ,$$

where **X** is a $s \times k$ design matrix and β a $k \times 1$ parameter vector.

Example 7.2.1

Residents of Johannesburg were asked whether they were satisfied with the public transportation system in the city. The responses for subpopulations of residents are given below.

Subpopulation	**Response**		**Total**
	yes	no	
Male inner-city residents	59	47	106
Male suburban residents	43	63	106
Female inner-city residents	71	75	146
Female suburban residents	47	89	136

Suppose that the expected value of a typical element $y_i = ln \frac{p_{i1}}{p_{i2}}$ of the response vector **y** can be expressed as

$$E(y_i) = \beta_{INTERCEPT} + \beta_{GENDER}\,\text{GENDER} + \beta_{AREA}\,\text{AREA} \qquad (7.2.9)$$

where GENDER = '1' if the respondent is male and GENDER = '−1' if the respondent is female. If a respondent lives in the inner-city, a value of '1' is assigned to the variable AREA. For a suburban resident AREA = '−1'.

The expected value of **y** is given by

$$E(\mathbf{y}) = \mathbf{X}\,\beta$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{INTERCEPT} \\ \beta_{GENDER} \\ \beta_{AREA} \end{bmatrix} . \qquad \square$$

The generalized least squares estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by minimizing the quadratic function

$$Q_{\beta} = (\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta})'\,\Sigma^{-1}(\,\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}) \tag{7.2.10}$$

where

$$\Sigma = \mathrm{Cov}(\mathbf{y}, \mathbf{y}') \,. \tag{7.2.11}$$

An approximate expression for Σ is obtained through use of the following first order Taylor expansion of $\mathbf{y}$ evaluated in the neighbourhood of $\boldsymbol{\pi}$

$$\mathbf{y} \simeq f(\boldsymbol{\pi}) + \mathbf{J}(\mathbf{p} - \boldsymbol{\pi}) \tag{7.2.12}$$

where

$$\mathbf{J} : s \times cs = \frac{\partial \mathbf{y}}{\partial \mathbf{p}'} \mid \mathbf{p} = \boldsymbol{\pi} \,, \tag{7.2.13}$$

and c is the number of categories of the response variable.

The covariance of $\mathbf{y}$, to the first order of approximation, can be written as

$$\Sigma = \mathbf{J}\ \mathrm{Cov}(\mathbf{p},\ \mathbf{p}')\ \mathbf{J}'$$

$$= \mathbf{J}\ \mathbf{V}\ \mathbf{J}'\ . \tag{7.2.14}$$

In general it can be shown that, if the c-th category of the response variable is used as the reference category, Σ has the following form

$$\Sigma = \begin{bmatrix} \frac{1}{f_{1.}}\Psi_{11} & 0 & . & 0 \\ 0 & \frac{1}{f_{2.}}\Psi_{22} & . & . \\ . & . & . & 0 \\ 0 & . & 0 & \frac{1}{f_{n.}}\Psi_{ss} \end{bmatrix} . \tag{7.2.15}$$

If there are two response categories (eg. 'yes' and 'no'), Ψ_{ii} for the i-th subpopulation is given by

$$\Psi_{ii} = \left(\frac{1}{\pi_{i1}} + \frac{1}{\pi_{i2}}\right) \tag{7.2.16}$$

where π_{i1} and π_{i2} denote the cell probabilities of the first and second category respectively.

In the case of three response categories (for example 'yes', 'no' and 'don't know') the following expression for $\mathbf{\Psi}_{ii}$ is obtained if the third category is used as a reference category:

$$\mathbf{\Psi}_{ii} = \begin{bmatrix} \frac{1}{\pi_{i1}} + \frac{1}{\pi_{i3}} & \frac{1}{\pi_{i3}} \\ \frac{1}{\pi_{i3}} & \frac{1}{\pi_{i2}} + \frac{1}{\pi_{i3}} \end{bmatrix} \tag{7.2.17}$$

Finally, the case with four response categories is considered. If the fourth category is used as reference category, $\mathbf{J\ V\ J'}$ for the i-th population is given by

$$\boldsymbol{\Psi}_{ii} = \begin{bmatrix} \frac{1}{\pi_{i1}} + \frac{1}{\pi_{i4}} & \frac{1}{\pi_{i4}} & \frac{1}{\pi_{i4}} \\ \frac{1}{\pi_{i4}} & \frac{1}{\pi_{i2}} + \frac{1}{\pi_{i4}} & \frac{1}{\pi_{i4}} \\ \frac{1}{\pi_{i4}} & \frac{1}{\pi_{i4}} & \frac{1}{\pi_{i3}} + \frac{1}{\pi_{i4}} \end{bmatrix}. \tag{7.2.18}$$

In general, for a c category response variable

$$\boldsymbol{\Psi}_{ii} = \mathbf{D}_{\boldsymbol{\pi}_i} + \mathbf{j}\ \mathbf{j}'/\pi_{ic}\ , \tag{7.2.19}$$

where π_{ic} is the probability that respondents in the i-th subpopulation will select the c-th category of the response variable.

By using the above results, the logit model can be written as

$$\mathbf{y} = \mathbf{X}\ \boldsymbol{\beta} + \mathbf{D}\ \mathbf{e} \tag{7.2.20}$$

where

$$\mathrm{E}(\mathbf{e}) = \mathbf{0}\ ,$$

$$\mathrm{Cov}(\mathbf{e}, \mathbf{e}') = \begin{bmatrix} \boldsymbol{\Psi}_{11} & 0 & . & 0 \\ 0 & \boldsymbol{\Psi}_{22} & . & . \\ . & . & . & 0 \\ 0 & . & 0 & \boldsymbol{\Psi}_{ss} \end{bmatrix} \tag{7.2.21}$$

and

$$\mathbf{D} : n \times n = \begin{bmatrix} \frac{1}{\sqrt{f_{1.}}} \mathbf{I}_{c-1} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{f_{2.}}} \mathbf{I}_{c-1} & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \frac{1}{\sqrt{f_{s.}}} \mathbf{I}_{c-1} \end{bmatrix} \qquad (7.2.22)$$

where $\mathbf{I}_{c-1}$ denotes an identity matrix of order $c-1$ and $n = s \times (c-1)$. The expected value of $\mathbf{y}$ is given by

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\,\boldsymbol{\beta}$$

with (cf. (7.2.14))

$$\mathrm{Cov}(\mathbf{y}, \mathbf{y}') = \mathbf{D}\ \mathrm{Cov}(\mathbf{e}, \mathbf{e}')\ \mathbf{D}' \simeq \Sigma\,.$$

7.2.2 ESTIMATION OF THE UNKNOWN PARAMETERS

The IGLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained as the minimum of (7.2.10), with solution

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\,\Sigma^{-1}\,\mathbf{X})^{-1}\mathbf{X}'\,\Sigma^{-1}\,\mathbf{y}\,. \qquad (7.2.23)$$

Since Σ is unknown, it is initially calculated through the substitution of the population probabilities π_{ij} (cf. (7.2.3)) with their corresponding sample probabilities p_{ij} (cf. (7.2.4)). A first estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is subsequently obtained.

Estimators $\hat{\pi}_{ij}$ of π_{ij} are calculated from the elements of $\hat{\boldsymbol{\beta}}$. These estimators are used to obtain a revised estimator $\hat{\Sigma}$ of Σ where $\hat{\Sigma} = \Sigma(\hat{\pi}_{ij})$. This procedure is repeated until convergence is obtained.

From (7.2.9), Example 7.2.1 it follows that

$$ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i2}}\right) = k \tag{7.2.24}$$

where

$$k = \hat{\beta}_{INTERCEPT} + \hat{\beta}_{GENDER}\ \text{GENDER} + \hat{\beta}_{AREA}\ \text{AREA}.$$

Since $\hat{\pi}_{i2} = 1 - \hat{\pi}_{i1}$, it follows from (7.2.24) that

$$\hat{\pi}_{i1} = \frac{e^k}{1 + e^k} \quad .$$

If the response variable has c categories, $c-1$ equations of the form

$$ln\,\frac{\hat{\pi}_{ij}}{\hat{\pi}_{ic}} = k_j \qquad\qquad j = 1, 2, \ldots, c-1$$

are obtained with solution

$$\hat{\pi}_{ij} = \frac{\exp(k_j)}{1 + \sum_{j=1}^{c-1} \exp(k_j)} \,. \tag{7.2.25}$$

In this section an overview of the standard logit model and the estimation of the unknown parameters was given. This logit model can be regarded as a level-1 model with $n = s \times (c-1)$ units and can be written in the form

$$\mathbf{y} = \mathbf{X}_{(f)}\,\boldsymbol{\beta} + \mathbf{X}_{(1)}\,\mathbf{e}$$

where $\mathbf{X}_{(f)} = \mathbf{X}$ (cf. (7.2.20)) and the random parameter design matrix on level-1 is $\mathbf{X}_{(1)}$, which is equal to $\mathbf{D}$ (cf. (7.2.22)). In the following sections this logit model will be extended to allow for hierarchically structured data.

7.3 LEVEL-2 LOGIT MODELS

In this section an overview is given of level-2 logit models (Prosser, Rasbash & Goldstein, 1991 and du Toit, 1993).

Suppose that there are N level-2 units and that for the i-th level-2 unit s_i subpopulations are formed by the categories of p predictor variables. Suppose further that the response variable under consideration has c categories and that the variation in frequency of responses can be adequately described by a logit model. The contingency table for the i-th level-2 unit is as follows:

Subpopulation no	Response category				Total
	1	2	. . .	c	
1	$f_{i1,1}$	$f_{i1,2}$	. . .	$f_{i1,c}$	$f_{i1.}$
2	$f_{i2,1}$	$f_{i2,2}$	. . .	$f_{i2,c}$	$f_{i2.}$
.	.	.	. . .	.	.
.	.	.	. . .	.	.
.	.	.	. . .	.	.
s	$f_{is_i,1}$	$f_{is_i,2}$	. . .	$f_{is_i,c}$	$f_{is_i.}$

For each subpopulation j , $j = 1, 2, \ldots, s_i$, the probability π_{ijk} of the k-th response being selected is estimated by $p_{ijk} = \frac{f_{ij,k}}{f_{ij.}}$. These estimates are used to obtain the elements of the response vector $\mathbf{y}_i$, $i = 1, 2, \ldots, N$, where

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{i1} \\ . \\ . \\ \mathbf{y}_{ij} \\ . \\ . \\ \mathbf{y}_{is_i} \end{bmatrix} \qquad (7.3.1)$$

Note that the number of subpopulations, s_i , may vary from one level-2 unit to another. In this case, s_i varies between 1 and 6. The dimension of $\mathbf{y}_i$ is $(c-1)s_i$ and may thus vary from 2 to 12 for this particular example. The model for the i-th level-2 model is given by

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{b}_i + \mathbf{D}_i \mathbf{e}_i \, , \tag{7.3.3}$$

where (cf. (7.2.22))

$$\mathbf{D}_i = \begin{bmatrix} \frac{1}{\sqrt{f_{i1.}}}\mathbf{I}_{c-1} & 0 & . & 0 \\ 0 & \frac{1}{\sqrt{f_{i2.}}}\mathbf{I}_{c-1} & . & . \\ . & . & . & 0 \\ 0 & . & 0 & \frac{1}{\sqrt{f_{is_i.}}}\mathbf{I}_{c-1} \end{bmatrix}. \tag{7.3.4}$$

Suppose that the first element of the coefficient vector $\mathbf{b}_i$ denotes the intercept, while the remaining $k-1$ coefficients are subpopulation effects. Random variation over the level-2 units may be allowed for by defining $\mathbf{b}_i$, for example, as

$$\mathbf{b}_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ . \\ . \\ . \\ b_{ik} \end{bmatrix} = \begin{bmatrix} \beta_1 + u_{i1} \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}, \qquad i = 1, 2, \ldots, N. \tag{7.3.5}$$

In general, $\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{u}_i$, and it is assumed that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$ are identically and independently distributed with mean $\mathbf{0}$ and variance $\boldsymbol{\Phi}_{(2)}$.

The model given by (7.3.3) can then be rewritten as

$$\mathbf{y}_i = \mathbf{X}_{(2)i}\, \mathbf{u}_{i1} + \mathbf{X}_{(f)i}\, \boldsymbol{\beta} + \mathbf{X}_{(1)i}\, \mathbf{e}_i \qquad (7.3.6)$$

where (cf. (7.3.4))

$$\mathbf{X}_{(1)i} = \mathbf{D}_i\,, \qquad (7.3.7)$$

and $\mathbf{X}_{(2)i} : (c-1)\, s_i \times 1$ is a vector with all elements equal to 1, that is,

$$\mathbf{X}_{(2)i} = \mathbf{j}\,. \qquad (7.3.8)$$

Under the assumption that $\mathrm{E}(\mathbf{e}_i) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{e}_i, \mathbf{e}_i') = \boldsymbol{\Phi}_{(1)}$, it follows that

$$\mathrm{E}\,(\mathbf{y}_i) = \mathbf{X}_{(f)i}\, \boldsymbol{\beta}\,, \qquad (7.3.9)$$

$$\Sigma_i = \mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i') = \mathbf{X}_{(2)i}\, \boldsymbol{\Phi}_{(2)}\, \mathbf{X}_{(2)i}' + \mathbf{X}_{(1)i}\, \boldsymbol{\Phi}_{(1)}\, \mathbf{X}_{(1)i}' \qquad (7.3.10)$$

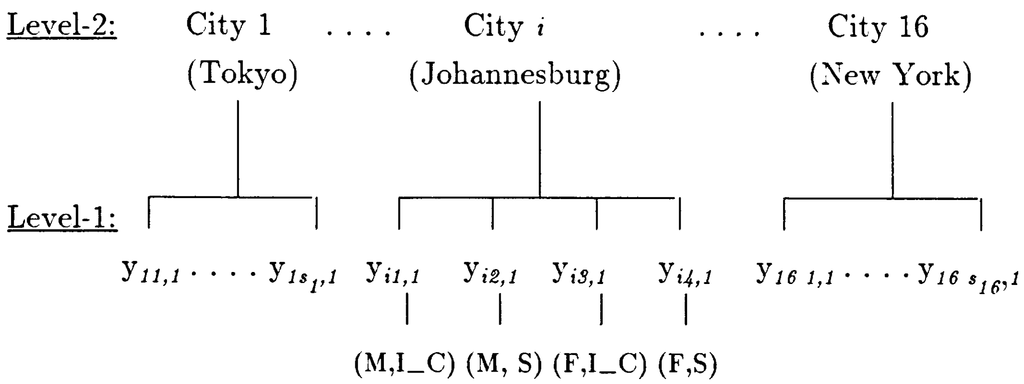
where (cf. (7.2.15) to (7.2.18))

$$\boldsymbol{\Phi}_{(1)} = \begin{bmatrix} \boldsymbol{\Phi}_{(1)_{11}} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{(1)_{22}} & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0}. & . & \mathbf{0} & \boldsymbol{\Phi}_{(1)_{s_i s_i}} \end{bmatrix} \qquad (7.3.11)$$

and $\boldsymbol{\Phi}_{(1)jj}$, $j = 1, 2, \ldots, s_i$ are positive definite matrices of order $(c-1)$.

Example 7.3.2

Suppose the survey regarding public transportation mentioned in Example 7.2.1 is extended to include 15 other cities. A schematic representation of the data is given by



where (M, I_C), for example, indicates the response of male inner-city respondents. For the i-th level-2 unit, that is Johannesburg, it follows from Example 7.2.1 that (cf. (7.3.3) and (7.3.6))

$$\mathbf{y}_i = \mathbf{X}_{(2)i}\, u_{i1} + \mathbf{X}_{(f)i}\, \boldsymbol{\beta} + \mathbf{X}_{(1)i}\, \mathbf{e}_i$$

with

$$\mathbf{y}_i = \begin{bmatrix} ln\, \frac{59}{47} \\ ln\, \frac{43}{63} \\ ln\, \frac{71}{75} \\ ln\, \frac{47}{89} \end{bmatrix},$$

$$\mathbf{X}_{(2)i} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{X}_{(f)i} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix},$$

$$\mathbf{X}_{(1)i} = \begin{bmatrix} \frac{1}{\sqrt{47}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{63}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{75}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{89}} \end{bmatrix}$$

and

$$\beta = \begin{bmatrix} \beta_{INTERCEPT} \\ \beta_{GENDER} \\ \beta_{AREA} \end{bmatrix}.$$

□

In this section the logit model was generalized to accommodate random variation across level-2 units. In the next section this theory will be extended to allow for random variation on a third level of the hierarchy.

7.4 LEVEL-3 LOGIT MODELS

7.4.1 INTRODUCTION

If there exists a third level to the hierarchy in the data to be modelled, the model discussed in Section 7.3 can be extended.

Suppose s subpopulations are formed from the set of predictors available. For example, given the predictors GENDER (2 categories), MARITAL_STATUS (5 categories) and INCOME (4 categories) it is possible to form 40 subpopulations. Let subpopulation ijk refer to the k-th subpopulation from the j-th level-2 unit, $j = 1, 2, \ldots, n_i$ and i-th level-3 unit, $i = 1, 2, \ldots, N$. An example of such a subpopulation is married males in the middle income group from district j (level-2) in province i (level-3).

In the case of a level-3 hierarchy the contingency table for the j-th level-2 unit from the i-th level-3 unit is given by

Subpopulation	Response category				Total
	1	2	...	c	
1	$f_{ij1,1}$	$f_{ij1,2}$	...	$f_{ij1,c}$	$f_{ij1.}$
2	$f_{ij2,1}$	$f_{ij2,2}$	...	$f_{ij2,c}$	$f_{ij2.}$
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
s_j	$f_{ijs_j,1}$	$f_{ijs_j,2}$	...	$f_{ijs_j,c}$	$f_{ijs_j.}$

.

The cell frequencies for the response categories for the k-th subpopulation of the j-th level-2 and i-th level-3 unit are denoted by $f_{ijk,1}$, $f_{ijk,2}$, ... , $f_{ijk,c}$. The row total $f_{ijk.}$ is the sample size for the subpopulation.

For each level-3 unit a number of tables such as the one given above may be constructed. The number of tables constructed for a given level-3 unit will be equal to the number of level-2 units contained within that specific level-3 unit.

The vector of responses for the j-th level-2 unit from the i-th level-3 unit is given by

$$\mathbf{y}_{ij} = \begin{bmatrix} ln\,\dfrac{f_{ij1,1}}{f_{ij1,c}} \\ \cdot \\ \cdot \\ ln\,\dfrac{f_{ij1,c-1}}{f_{ij1,c}} \\ \cdot \\ \cdot \\ ln\,\dfrac{f_{ijs_j,1}}{f_{ijs_j,c}} \\ \cdot \\ \cdot \\ \cdot \\ ln\,\dfrac{f_{ijs_j,c-1}}{f_{ijs_j,c}} \end{bmatrix} . \quad (7.4.1)$$

The model for the ij-th unit is given by (cf. (7.3.3))

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\,\mathbf{b}_{ij} + \mathbf{D}_{ij}\,\mathbf{e}_{ij} \tag{7.4.2}$$

with

$$\mathbf{D}_{ij} = \begin{bmatrix} \frac{1}{\sqrt{f_{ij1.}}}\mathbf{I}_{c-1} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{f_{ij2.}}}\mathbf{I}_{c-1} & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & . & \mathbf{0} & \frac{1}{\sqrt{f_{ijs_j.}}}\mathbf{I}_{c-1} \end{bmatrix} \tag{7.4.3}$$

and the parameter vector $\mathbf{b}_{ij}$ includes the components

$$\mathbf{b}_{ij} = \boldsymbol{\beta} + \mathbf{S}_{(3)}\,\mathbf{v}_i + \mathbf{S}_{(2)}\,\mathbf{u}_{ij}\,. \tag{7.4.4}$$

$\mathbf{S}_{(3)}$ is a $t \times q$ $(q \leq t)$ matrix formed by the selection of columns from the identity matrix of order t. These columns correspond with the elements of $\mathbf{b}_{ij}$ allowed to be random on level-3. Likewise, $\mathbf{S}_{(2)} : t \times m$ is a selection of m columns from $\mathbf{I} : t \times t$.

If, for example, $t = 4$ and only the first and fourth coefficients are allowed to vary randomly on level-2, then

$$\mathbf{S}_{(2)} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The vectors $\mathbf{v}_i$ and $\mathbf{u}_{ij}$ represent the coefficients of variables allowed to be random on level-3 and level-2 of the model respectively. The vector $\boldsymbol{\beta}$ denotes the fixed parameter vector.

Using (7.4.2) to (7.4.4), the model can be written as

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)ij}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}_{(2)ij}\,\mathbf{u}_{ij} + \mathbf{X}_{(1)ij}\,\mathbf{e}_{ij} \qquad (7.4.5)$$

where $\mathbf{X}_{(f)ij}$ is the fixed parameter design matrix and $\mathbf{X}_{(1)ij} = \mathbf{D}_{ij}$. The matrices $\mathbf{X}_{(3)ij}$, $\mathbf{X}_{(2)ij}$ and $\mathbf{X}_{(1)ij}$ denote the random parameter design matrices for level-3, level-2 and level-1 respectively. The columns of $\mathbf{X}_{(3)ij}$ and $\mathbf{X}_{(2)ij}$ are usually subsets of the columns of $\mathbf{X}_{(f)ij}$, that is

$$\mathbf{X}_{(3)ij} = \mathbf{X}_{(f)ij}\,\mathbf{S}_{(3)} \qquad (7.4.6)$$

and

$$\mathbf{X}_{(2)ij} = \mathbf{X}_{(f)ij}\,\mathbf{S}_{(2)}\,. \qquad (7.4.7)$$

$\mathbf{v}_1$, $\mathbf{v}_2$, . . . are assumed to be i.i.d. with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(3)}$ while $\mathbf{u}_{11}$, $\mathbf{u}_{12}$, . . . are assumed to be i.i.d with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$. The $\mathbf{e}_{ijk}$'s are similarly assumed to be i.i.d with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(1)}$, where $\boldsymbol{\Phi}_{(1)}$ is defined by (7.3.11). It is further assumed that the vectors of random coefficients $\mathbf{v}_i$, $\mathbf{u}_{ij}$ and $\mathbf{e}_{ijk}$ are uncorrelated.

Under the distributional assumptions given above, it follows that

$$\mathrm{E}(\mathbf{y}_{ij}) = \mathbf{X}_{(f)ij}\,\boldsymbol{\beta} \qquad (7.4.8)$$

and

$$\mathrm{Cov}(\mathbf{y}_{ij},\mathbf{y}'_{ij}) = \mathbf{X}_{(3)ij}\,\boldsymbol{\Phi}_{(3)}\,\mathbf{X}'_{(3)ij} + \mathbf{X}_{(2)ij}\,\boldsymbol{\Phi}_{(2)}\,\mathbf{X}'_{(2)ij} + \mathbf{X}_{(1)ij}\,\boldsymbol{\Phi}_{(1)}\,\mathbf{X}'_{(1)ij}\,,$$

where

$$\Phi_{(1)} = \begin{bmatrix} \Phi_{(1)_{11}} & 0 & . & 0 \\ 0 & \Phi_{(1)_{22}} & . & . \\ . & . & . & 0 \\ 0 & . & 0 & \Phi_{(1)_{kk}} \end{bmatrix}.$$

Example 7.4.1

A survey was undertaken in a large company. Data was collected on employees' attitudes to changes in company policy. Biographical information available included the gender and home language of the respondents.

Consider the following level-3 hierarchical structure:

Level-3: Department 1 . . . . . Department i . . . . . . Department N

Level-2: Div. 1 . . Div. n_1 Div. 1 . . Div. j . . Div. n_i Div. 1 . . Div. n_N

Level-1:

Subpopulation	Response			Total
	Yes	No	Don't know	
Male, English	$f_{ij1,1}$	$f_{ij1,2}$	$f_{ij1,3}$	$f_{ij1.}$
Male, Other	$f_{ij2,1}$	$f_{ij2,2}$	$f_{ij2,3}$	$f_{ij2.}$
Female, English	$f_{ij3,1}$	$f_{ij3,2}$	$f_{ij3,3}$	$f_{ij3.}$
Female, Other	$f_{ij4,1}$	$f_{ij4,2}$	$f_{ij4,3}$	$f_{ij4.}$

Suppose that there are only 2 divisions, namely clothing and sanitary ware, in the 10-th department. Employees' responses to a question regarding the advisability of more stringent safety measures in the workplace are given in the following contingency table:

Division	Subpopulation	Response			Total
		Yes	No	Don't know	
1	Male, English	26	39	24	89
	Male, Other	39	8	21	68
	Female, English	32	24	37	93
	Female, Other	22	13	45	80
2	Male, English	23	33	46	102
	Male, Other	22	24	15	61
	Female, English	–	–	–	–
	Female, Other	44	25	8	77

The vector of observed responses $\mathbf{y}_i$ for the 10-th department is then given by

$$\mathbf{y}_i = \begin{bmatrix} ln\,\frac{26}{24} \\ ln\,\frac{39}{24} \\ ln\,\frac{39}{21} \\ ln\,\frac{8}{21} \\ ln\,\frac{32}{37} \\ ln\,\frac{24}{37} \\ ln\,\frac{22}{45} \\ ln\,\frac{13}{45} \\ ln\,\frac{23}{46} \\ ln\,\frac{33}{46} \\ ln\,\frac{22}{15} \\ ln\,\frac{24}{15} \\ ln\,\frac{44}{8} \\ ln\,\frac{25}{8} \end{bmatrix} = \begin{bmatrix} 0.0800 \\ 0.4855 \\ 0.6190 \\ -0.9651 \\ -0.1452 \\ -0.4329 \\ -0.7156 \\ -1.2417 \\ -0.6931 \\ -0.3321 \\ 0.3830 \\ 0.4700 \\ 1.7047 \\ 1.1394 \end{bmatrix}.$$

The design matrix for the fixed part of the model is

$$\mathbf{X}_{(f)i} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}$$

and the fixed parameter vector $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{INTERCEPT,YES} \\ \beta_{GENDER,YES} \\ \beta_{LANGUAGE,YES} \\ \beta_{INTERCEPT,NO} \\ \beta_{GENDER,NO} \\ \beta_{LANGUAGE,NO} \end{bmatrix} .$$

If only the intercept terms are allowed to vary randomly over the level-2 units, the matrix $\mathbf{X}_{(2)i}$ is given by

$$\mathbf{X}_{(2)i} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

where the first column corresponds to the constant term for the modelling of the log-odds ratio 'yes/don't know', while the second column corresponds to the constant term of the log-odds ratio 'no/don't know'.

The vector $\mathbf{u}_{ij}$ is given by

$$\mathbf{u}_{ij} = \begin{bmatrix} u_{ij,YES} \\ u_{ij,NO} \end{bmatrix}.$$

If the variable GENDER is allowed to vary across level-3 units, the vector $\mathbf{v}_i$ and random parameter design matrix $\mathbf{X}_{(3)i}$ are respectively given by

$$\mathbf{v}_i = \begin{bmatrix} v_{i,MALE} \\ v_{i,FEMALE} \end{bmatrix}$$

and

$$\mathbf{X}_{(3)i} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ -1 & 0 \\ -1 & 0 \\ -1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} . \qquad \square$$

In this section the level-3 logit model was considered. It may, however, be desirable to simultaneously analyse more than one response variable. In Section 7.5 a multivariate multilevel logit model will be discussed.

7.4.2 ESTIMATION CONSIDERATIONS

In this section it is shown that parameter estimation of a level-3 logit model is achieved by fitting a general level-3 model with random coefficients on level-3 and level-2 of the hierarchy.

Let

$$\mathbf{y}_{ijk} = \begin{bmatrix} y_{ijk,1} \\ y_{ijk,2} \\ \vdots \\ y_{ijk,c-1} \end{bmatrix}$$

denote a set of $c-1$ responses from subpopulation k, $k = 1, 2, \ldots, s_j$, from the j-th level-2 and i-th level-3 unit.

It follows from (7.4.5) that

$$\begin{aligned}
y_{ijk,1} &= \mathbf{x}'_{(f)ijk,1}\,\boldsymbol{\beta} + \mathbf{x}'_{(3)ijk,1}\,\mathbf{v}_i + \mathbf{x}'_{(2)ijk,1}\,\mathbf{u}_{ij} + \mathbf{x}'_{(1)ijk,1}\,\mathbf{e}_{ijk} \\
y_{ijk,2} &= \mathbf{x}'_{(f)ijk,2}\,\boldsymbol{\beta} + \mathbf{x}'_{(3)ijk,2}\,\mathbf{v}_i + \mathbf{x}'_{(2)ijk,2}\,\mathbf{u}_{ij} + \mathbf{x}'_{(1)ijk,2}\,\mathbf{e}_{ijk} \\
&\ \ \vdots \\
y_{ijk,c-1} &= \mathbf{x}'_{(f)ijk,c-1}\,\boldsymbol{\beta} + \mathbf{x}'_{(3)ijk,c-1}\,\mathbf{v}_i + \mathbf{x}'_{(2)ijk,c-1}\,\mathbf{u}_{ij} + \mathbf{x}'_{(1)ijk,c-1}\,\mathbf{e}_{ijk}\,.
\end{aligned} \qquad (7.4.9)$$

Under the assumption of a multinomial level-1 error structure (cf. Section 7.2) it follows that the elements of $\mathbf{e}_{ijk}$ are correlated.

In previous chapters the assumption was made that the error vectors on level-1 of the hierarchy are independently and identically distributed. A multilevel analysis of logit models requires that

(i) $\mathbf{e}_{ijk}$, $k = 1, 2, \ldots, s_j$ are independently distributed and that
(ii) $\mathrm{Cov}(\mathbf{e}_{ijk}, \mathbf{e}'_{ijk}) = \boldsymbol{\Phi}_{(1)_{kk}}$ (cf. (7.2.19))

which implies that $c-1$ consecutive error terms are allowed to be correlated.

It will now be shown that the level-3 logit model can be expressed as a model with random components on level-3 and level-2 of the hierarchy.

Let $\mathbf{X}_{(f)ijk}$, $\mathbf{X}_{(3)ijk}$, $\mathbf{X}_{(2)ijk}$ and $\mathbf{X}_{(1)ijk}$ have typical rows (cf. (7.4.9)) $\mathbf{x}'_{(f)ijk,l}$, $\mathbf{x}'_{(3)ijk,l}$, $\mathbf{x}'_{(2)ijk,l}$ and $\mathbf{x}'_{(1)ijk,l}$, $l = 1, 2, \ldots, c-1$, respectively.

From (7.4.9) it follows that

$$\mathbf{y}_{ijk} = \mathbf{X}_{(f)ijk}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ijk}\,\mathbf{v}_i + \mathbf{X}_{(2)ijk}\,\mathbf{u}_{ij} + \mathbf{X}_{(1)ijk}\,\mathbf{e}_{ijk}\,,\ k = 1, 2, \ldots, s. \qquad (7.4.10)$$

The set of regression equations given by (7.4.10) can be written as

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)ij}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}^*_{(2)ij}\,\mathbf{u}^*_{ij} \tag{7.4.11}$$

where

$$\mathbf{X}^*_{(2)ij} = \begin{bmatrix} \mathbf{X}_{(2)ij1} & \mathbf{X}_{(1)ij1} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{X}_{(2)ij2} & \mathbf{0} & \mathbf{X}_{(1)ij2} & \mathbf{0} & . \\ . & . & . & . & \mathbf{0} \\ \mathbf{X}_{(2)ijs} & \mathbf{0} & . & \mathbf{0} & \mathbf{X}_{(1)ijs} \end{bmatrix},$$

and

$$\mathbf{u}^*_{ij} = \begin{bmatrix} \mathbf{u}_{ij} \\ \mathbf{e}_{ij1} \\ \mathbf{e}_{ij2} \\ \vdots \\ \mathbf{e}_{ijs} \end{bmatrix},$$

where s is the maximum number of subpopulations.

Note that

$$\mathbf{X}^*_{(2)ij} = \begin{bmatrix} \mathbf{X}_{(2)ij} & \mathbf{X}_{(1)ij} \end{bmatrix},$$

where (cf. (7.4.3))

$$\mathbf{X}_{(1)ij} = \mathbf{D}_{ij}\,.$$

Also note that when $s_j < s$, where s indicates the maximum number of subpopulations, $\mathbf{X}_{(2)ij}$ has $(c-1)\times s_j$ rows, but the number of columns remain $m + (c-1)\times s$, which is the dimension of the random coefficient vector $\mathbf{u}^*_{ij}$.

The covariance of $\mathbf{u}^*_{ij}$ can then be written as

$$\boldsymbol{\Phi}^*_{(2)} = \text{Cov}(\mathbf{u}^*_{ij},\mathbf{u}^{*\prime}_{ij}) = \begin{bmatrix} \boldsymbol{\Phi}_{(2)} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{(1)_{11}} & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \boldsymbol{\Phi}_{(1)_{ss}} \end{bmatrix}.$$

Suppose that the number of random coefficients on level-2, denoted by m, is equal to 4, that the number of categories, denoted by c, is also equal to 4 and that the number of subpopulations, s, is equal to 10.

In this case the dimension of the matrix $\boldsymbol{\Phi}^*_{(2)}$ is 34×34. Since $\boldsymbol{\Phi}_{(2)}$ is a 4×4 matrix and $\boldsymbol{\Phi}_{(1)_{11}}, \ldots, \boldsymbol{\Phi}_{(1)_{ss}}$ are 3×3 matrices, the number of non-duplicated zero elements of $\boldsymbol{\Phi}^*_{(2)}$ is equal to $\frac{1}{2}(34)(35)-\frac{1}{2}(4)(5)-10(\frac{1}{2}(3)(4)) = 525$.

Since (cf. (5.2.15)),

$$\boldsymbol{\tau} = \begin{bmatrix} \text{vecs } \boldsymbol{\Phi}_{(3)} \\ \text{vecs } \boldsymbol{\Phi}^*_{(2)} \end{bmatrix},$$

it follows that in the solution of (cf. (5.2.17))

$$\hat{\boldsymbol{\tau}} = [\mathbf{X}^{*\prime}\, \mathbf{W}^{-1}\, \mathbf{X}^*]^{-1}\, [\mathbf{X}^{*\prime}\, \mathbf{W}^{-1}\, \mathbf{y}^*] \qquad (7.4.12)$$

a large numbers of rows and corresponding columns of $\mathbf{X}^{*\prime}\mathbf{W}^{-1}\mathbf{X}^*$ and rows of $\mathbf{X}^{*\prime}\mathbf{W}^{-1}\mathbf{y}^*$ need not be calculated. This is best illustrated by means of the following example.

Example 7.4.2

Consider the set of equations

$$\begin{aligned}
a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 + a_{14}\,x_4 &= b_1 \\
a_{21}\,x_1 + a_{22}\,x_2 + a_{23}\,x_3 + a_{24}\,x_4 &= b_2 \\
a_{31}\,x_1 + a_{32}\,x_2 + a_{33}\,x_3 + a_{34}\,x_4 &= b_3 \\
a_{41}\,x_1 + a_{42}\,x_2 + a_{43}\,x_3 + a_{44}\,x_4 &= b_4\ ,
\end{aligned}$$

or $\mathbf{A}\,\mathbf{x} = \mathbf{b}$, that is, $\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b}$, which is similar in form to (7.4.12).

If x_2 and x_4 are constrained to be equal to 0, elimination of rows 2 and 4 and columns 2 and 4 from the coefficient matrix $\mathbf{A}$ gives

$$\begin{aligned}
a_{11}\,x_1 + a_{13}\,x_3 &= b_1 \\
a_{31}\,x_1 + a_{33}\,x_3 &= b_3\ .
\end{aligned}$$

If, on the other hand, we constrain x_2 to be equal to x_3, then

$$\begin{aligned}
&a_{11}\,x_1 + (a_{12} + a_{13})\,x_2 + a_{14}\,x_4 = b_1 \\
&(a_{21} + a_{31})\,x_1 + (a_{22} + a_{23} + a_{32} + a_{33})\,x_2 + (a_{24} + a_{34})\,x_4 = b_2 + b_3 \\
&a_{41}\,x_1 + (a_{42} + a_{43})\,x_2 + a_{44}\,x_4 = b_4\ . \qquad \square
\end{aligned}$$

Constraining elements of $\boldsymbol{\tau}$ to be zero thus involves the elimination of rows and corresponding columns from the weight matrix $\mathbf{X}^{*\prime}\ \mathbf{W}^{-1}\ \mathbf{X}^{*}$ and rows from the coefficient vector $\mathbf{X}^{*\prime}\ \mathbf{W}^{-1}\ \mathbf{y}^{*}$. Constraining elements of $\boldsymbol{\tau}$ to be equal involves the addition of rows and corresponding columns of $\mathbf{X}^{*\prime}\mathbf{W}^{-1}\,\mathbf{X}^{*}$ and the addition of rows of $\mathbf{X}^{*\prime}\ \mathbf{W}^{-1}\ \mathbf{y}^{*}$.

7.5 MULTIVARIATE LEVEL-3 LOGIT MODELS

Survey data usually consist of a mixture of biographical, geographical and response or attitude variables. Quite often these response variables are categorical in nature.

In dealing with continuous response variables, researchers will most likely use a multivariate analysis (MANOVA) approach to model structural relationships between the dependent variables and the predictors. When the dependent variables are of a categorical nature, data analysts tend to resort to a one-at-a-time loglinear modelling approach.

In this section an approach to the analysis of multivariate level-3 logit models is proposed.

Let $f_{ijkl,m}$ denote the number of respondents from the (ijk)-th subpopulation who selected category m when answering question l, $l = 1, 2, \ldots, p$ and $m = 1, 2, \ldots, c$. The table below gives the frequency counts for the i-th level-3 and the j-th level-2 unit.

Subpopulation	**Response variable**	**Response category** 1	2	$\ldots$	c	**Total**
1	1	$f_{ij11,1}$	$f_{ij11,2}$	$\cdots$	$f_{ij11,c}$	$f_{ij11.}$
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	p	$f_{ij1p,1}$	$f_{ij1p,2}$	$\cdots$	$f_{ij1p,c}$	$f_{ij1p.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
k	1	$f_{ijk1,1}$	$f_{ijk1,2}$	$\cdots$	$f_{ijk1,c}$	$f_{ijk1.}$
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	l	$f_{ijkl,1}$	$f_{ijkl,2}$	$\cdots$	$f_{ijkl,c}$	$f_{ijkl.}$
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	p	$f_{ijkp,1}$	$f_{ijkp,2}$	$\cdots$	$f_{ijkp,c}$	$f_{ijkp.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
s_j	1	$f_{ijs_j1,1}$	$f_{ijs_j1,2}$	$\cdots$	$f_{ijs_j1,c}$	$f_{ijs_j1.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
s_j	p	$f_{ijs_jp,1}$	$f_{ijs_jp,2}$	$\cdots$	$f_{ijs_jp,c}$	$f_{ijs_jp.}$

Let $\mathbf{y}_{ijk}$ be a $p(c-1)$ vector defined as follows:

$$\mathbf{y}_{ijk} = \begin{bmatrix} ln\,\dfrac{f_{ijk1,1}}{f_{ijk1,c}} \\ ln\,\dfrac{f_{ijk1,2}}{f_{ijk1,c}} \\ \vdots \\ ln\,\dfrac{f_{ijk1,c-1}}{f_{ijk1,c}} \\ \vdots \\ ln\,\dfrac{f_{ijkl,1}}{f_{ijkl,c}} \\ ln\,\dfrac{f_{ijkl,2}}{f_{ijkl,c}} \\ \vdots \\ ln\,\dfrac{f_{ijkl,c-1}}{f_{ijkl,c}} \\ \vdots \\ ln\,\dfrac{f_{ijkp,c-1}}{f_{ijkp,c}} \end{bmatrix} \qquad (7.5.1)$$

Suppose that the following model provides an adequate description of the data:

$$\mathbf{y}_{ijk} = \mathbf{X}_{(f)ijk}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ijk}\,\mathbf{v}_i + \mathbf{X}_{(2)ijk}\,\mathbf{u}_{ij} + \mathbf{X}_{(1)ijk}\,\mathbf{e}_{ijk}\,,\ k = 1,\,2,\,\ldots,\,s_j\,, \qquad (7.5.2)$$

where s_j is the number of subpopulations for the (i,j)-th unit and the columns of $\mathbf{X}_{(3)ijk}$ and $\mathbf{X}_{(2)ijk}$ are subsets of the columns of $\mathbf{X}_{(f)ijk}$ (cf. (7.4.6) and (7.4.7)). The fixed parameter design matrix can be written as (see Example 7.6.3)

$$\mathbf{X}_{(f)ijk} = \mathbf{Z}_{ij} \otimes \mathbf{I}_{q(c-1)}\,. \qquad (7.5.3)$$

Example 7.5.1

The data described in Example 7.4.1, concerning employees' attitudes to company policy, shows that the fixed-parameter design matrix for the i-th unit has columns containing the dummy variables for the intercept term, gender and language. Assuming that two questions concerning safety measures are to be analysed simultaneously, each with three possible outcomes, it follows that the fixed-parameter design matrix will have pairs of dummy variables for each of the response variables. For the first response variable the dummy variables created are CONS1_1, CONS1_2, GENDER1_1, GENDER1_2, LANG1_1 and LANG1_2. The construction of dummy variables for the second response variable will be similar to those of the first, giving 12 dummy variables in total.

For the third subpopulation from the j-th division the matrix $\mathbf{X}_{(f)ijk}$, $k = 3$, is given by

$$\mathbf{X}_{(f)ij3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

or

$$\mathbf{X}_{(f)ij3} = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix} \otimes \mathbf{I}_4 \, . \qquad \square$$

The vector of response for the (i,j)-th unit can be written as

$$\mathbf{y}_{ij} = \begin{bmatrix} \mathbf{y}_{ij1} \\ \vdots \\ \mathbf{y}_{ijk} \\ \vdots \\ \mathbf{y}_{ijs} \end{bmatrix}, \qquad (7.5.4)$$

where

$$\mathbf{y}_{ijk} = \begin{bmatrix} y_{ij1} \\ \cdot \\ \cdot \\ y_{ijc\text{-}1} \end{bmatrix} = ln \begin{bmatrix} \frac{f_{ij1}}{f_{ijc}} \\ \cdot \\ \cdot \\ \frac{f_{ijc\text{-}1}}{f_{ijc}} \end{bmatrix}.$$

and f_{ijk} denotes the number of respondents in subpopulation (i,j) who selected category k of the response variable, $k = 1, 2, \ldots, c$.

From (7.5.4) it follows that

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)ij}\, \beta + \mathbf{X}_{(3)ij}\, \mathbf{v}_i + \mathbf{X}_{(2)ij}\, \mathbf{u}_{ij} + \mathbf{X}_{(1)ij}\, \mathbf{e}_{ij} \qquad (7.5.5)$$

where

$$\mathbf{e}_{ij} = \begin{bmatrix} \mathbf{e}_{ij1} \\ \vdots \\ \mathbf{e}_{ijk} \\ \vdots \\ \mathbf{e}_{ijs} \end{bmatrix}$$

and (cf. (7.4.3))

$$\mathbf{X}_{(1)ij} = \begin{bmatrix} \frac{1}{\sqrt{f_{ij1.}}}\mathbf{I}_{c\text{-}1} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{f_{ij2.}}}\mathbf{I}_{c\text{-}1} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \cdot & \mathbf{0} & \frac{1}{\sqrt{f_{ijs.}}}\mathbf{I}_{c\text{-}1} \end{bmatrix}. \qquad (7.5.6)$$

Suppose that for a given (i,j)-combination the number of subpopulations is s_j, with $s_j < s$. The $s(c-1)\times 1$ vector of responses $\mathbf{y}_{ij}$ then becomes a $s_j(c-1)\times 1$ vector through the omission of the rows pertaining to those subpopulations which are not present. Corresponding rows (cf. Section 7.4.2) are also omitted from the design matrices $\mathbf{X}_{(f)ij}$, $\mathbf{X}_{(3)ij}$, $\mathbf{X}_{(2)ij}$ and $\mathbf{X}_{(1)ij}$.

It is assumed that the $(c-1)\times 1$ dimensional vectors $\mathbf{e}_{ij1}$, $\mathbf{e}_{ij2}$, . . . , $\mathbf{e}_{ijs}$ are independently distributed with covariance matrices $\mathbf{\Phi}_{(1)11}$, $\mathbf{\Phi}_{(1)22}$, . . . , $\mathbf{\Phi}_{(1)ss}$ respectively.

From (7.5.5) it follows that

$$\mathrm{E}(\mathbf{y}_{ij}) = \mathbf{X}_{(f)ij}\,\boldsymbol{\beta}$$

and

$$\mathrm{Cov}(\mathbf{y}_{ij},\mathbf{y}_{ij}') = \mathbf{X}_{(3)ij}\,\mathbf{\Phi}_{(3)}\,\mathbf{X}'_{(3)ij} + \mathbf{X}_{(2)ij}\,\mathbf{\Phi}_{(2)}\,\mathbf{X}'_{(2)ij} + \mathbf{X}_{(1)ij}\,\mathbf{\Phi}_{(1)}\,\mathbf{X}'_{(1)ij}$$

where

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} \mathbf{\Phi}_{(1)_{11}} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{(1)_{22}} & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \mathbf{\Phi}_{(1)_{ss}} \end{bmatrix}. \qquad (7.5.7)$$

To estimate the unknown parameters, model (7.5.5) is rewritten as (cf. Section 7.4)

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}^*_{(2)ij}\,\mathbf{u}^*_{ij}\,,$$

where

$$\mathbf{X}^*_{(2)ij} = \begin{bmatrix} \mathbf{X}_{(2)ij} & \mathbf{X}_{(1)ij} \end{bmatrix}$$

and

$$\mathbf{u}^*_{ij} = \begin{bmatrix} \mathbf{u}_{ij} \\ \mathbf{e}_{ij} \end{bmatrix}. \tag{7.5.8}$$

From (7.5.8) it follows that

$$\text{Cov}(\mathbf{u}^*_{ij}, \mathbf{u}'^*_{ij}) = \begin{bmatrix} \mathbf{\Phi}_{(2)} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{(1)_{11}} & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \mathbf{\Phi}_{(1)_{ss}} \end{bmatrix}. \tag{7.5.9}$$

Suppose that the number of response variables is equal to two. For the first response variable the multinomial covariance structure for the level-1 error terms is of the form $\mathbf{\Psi}_{11,_{kk}}$, as given in (7.2.17). The covariance structure for the second response variable similarly has the form $\mathbf{\Psi}_{22,_{kk}}$.

By allowing for correlation between these response variables, the following error structure is obtained

$$\mathbf{\Psi}_{kk} = \begin{bmatrix} \mathbf{\Psi}_{11,_{kk}} & \mathbf{\Psi}_{12,_{kk}} \\ \mathbf{\Psi}_{21,_{kk}} & \mathbf{\Psi}_{22,_{kk}} \end{bmatrix},$$

where $\mathbf{\Psi}_{12,_{kk}}$ denotes the covariances between the elements of the $(c-1)\times 1$ error vector belonging to the first response variable and the $(c-1)\times 1$ error vector of the second response variable. From (7.5.7) it follows that $\widehat{\mathbf{\Phi}}_{(1)_{kk}}$ is an estimator of $\mathbf{\Psi}_{kk}$ (cf. (7.2.17)).

The covariance matrix given in (7.5.9) once again contains a large number of zero elements. In Section 7.4.2 it was shown how constraints on these elements may be handled.

7.6 PRACTICAL APPLICATIONS

In this section three practical applications are given. The data set used for these examples is the TPA survey described in Section 5.4. The focus in the following examples is on employees' perceptions regarding the existence of discrimination against individuals in the TPA. The questions dealing with the issue of discrimination on the basis of gender and race are used as response variables in the examples that follow. There are five possible outcomes in both cases. A score of '1' indicated strong disagreement, '2' disagreement, '3' do not know, '4' agreement and '5' strong agreement with a particular statement.

Example 7.6.1

The first example is a level-3 model with the perception of discrimination on the basis of gender as response variable. The five possible outcomes for this question were recoded to the following two categories: negative responses (scores 1 and 2) and positive or uncertain responses (scores 3 to 5).

The level-3 units are the 37 branches in the Health Services unit of the TPA. Within each level-3 unit, two occupational groups are formed, that is medical personnel and other personnel. These occupational groups are the level-2 units. There are, therefore, a maximum of two level-2 units within each level-3 unit.

Within each level-2 unit, a maximum of six subpopulations are formed according to the gender and education level of employees, as shown below for the (i,j)-th unit :

Gender	Educational level	Number of responses		Total
		Negative	Don't know/yes	
Male	Lower than Std 10	$f_{ij1,1}$	$f_{ij1,2}$	$f_{ij1.}$
Male	Std 10 plus 1-2 years tertiary education	$f_{ij2,1}$	$f_{ij2,2}$	$f_{ij2.}$
Male	Std 10 plus 3 years or more of tertiary education	$f_{ij3,1}$	$f_{ij3,2}$	$f_{ij3.}$
Female	Lower than Std 10	$f_{ij4,1}$	$f_{ij4,2}$	$f_{ij4.}$
Female	Std 10 plus 1-2 years tertiary education	$f_{ij5,1}$	$f_{ij5,2}$	$f_{ij5.}$
Female	Std 10 plus 3 years or more of tertiary education	$f_{ij6,1}$	$f_{ij6,2}$	$f_{ij6.}$

The frequencies $f_{ijk,1}$ and $f_{ijk,2}$ denote the number of negative and don't know/positive responses respectively. The frequency $f_{ijk.}$ is the total number of responses for the k-th subpopulation from the (i,j)-th unit.

The vector of responses for the (i,j)-th unit is formed by using the second category as reference category and is given by

$$\mathbf{y}_{ij} = \begin{bmatrix} ln\,\frac{f_{ij1,1}}{f_{ij1,2}} \\ ln\,\frac{f_{ij2,1}}{f_{ij2,2}} \\ . \\ . \\ . \\ ln\,\frac{f_{ij6,1}}{f_{ij6,2}} \end{bmatrix}.$$

The model thus obtained is a level-3 logit model where each (i,j) combination has a maximum of six values of the form $y_{ijk} = ln\,\frac{f_{ijk,1}}{f_{ijk,2}}$. To avoid the occurrence of undefined y_{ijk}-values, a value of 0.25 is added to each observed frequency (Fienberg, 1980).

The gender of an employee (variable GENDER) and the level of education (variables QUAL1 and QUAL2) are used as predictors in this model. The predictor GENDER is coded '1' for males and '-1' for females. In the case of educational level, the variables QUAL1 and QUAL2 are constructed as shown below:

Level of education	**Value of QUAL1**	**Value of QUAL2**
Lower than Std 10.	1	0
Std 10 plus 1 to 2 years of tertiary education.	0	1
Std 10 plus 3 years or more of tertiary education.	-1	-1

The level-3 model can then be written as (cf. (7.4.5))

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)ij}\,\beta + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}_{(2)ij}\,\mathbf{u}_{ij} + \mathbf{X}_{(1)ij}\,\mathbf{e}_{ij}$$

where $\mathbf{X}_{(f)ij}$ denotes the fixed parameter design matrix for the (i,j)-th unit and consists of columns representing a constant term (CONS) and the variables GENDER, QUAL1 and QUAL2. If all subpopulations are present within the (i,j)-th unit, then

$$\mathbf{X}_{(f)ij} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}.$$

From (7.4.6) it follows that

$$\mathbf{X}_{(3)ij} = \mathbf{X}_{(f)ij}\,\mathbf{S}_{(3)}\,.$$

In the model fitted, the constant term CONS and GENDER are allowed to vary randomly across level-3 of the model. Thus $\mathbf{X}_{(3)ij}$ contains the following subset of columns of $\mathbf{X}_{(f)ij}$:

$$\mathbf{X}_{(3)ij} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

The variables QUAL1 and QUAL2 are allowed to vary randomly over the level-2 of the hierarchy. The matrix $\mathbf{X}_{(2)ij}$, which is the random-parameter design matrix on level-2 , contains a subset of the columns of $\mathbf{X}_{(f)ij}$ (cf. (7.4.7))

$$\mathbf{X}_{(2)ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}.$$

Since $c = 2$, it follows from (7.2.22) that the matrix $\mathbf{X}_{(1)ij}$ can be written as

$$\mathbf{X}_{(1)ij} = \begin{bmatrix} \frac{1}{\sqrt{f_{ij1.}}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{f_{ij2.}}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{f_{ij3.}}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{f_{ij4.}}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{f_{ij5.}}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{f_{ij6.}}} \end{bmatrix}.$$

The columns of $\mathbf{X}_{(1)ij}$ are denoted by the variables WEIGHT1, WEIGHT2 to WEIGHT6 respectively.

Using the following constraints on elements of the covariance matrix $\mathbf{\Phi}_{(1)}$ of the level-1 coefficients, namely

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} \Phi_{(1)_{11}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \Phi_{(1)_{22}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \Phi_{(1)_{33}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \Phi_{(1)_{44}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \Phi_{(1)_{55}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \Phi_{(1)_{66}} \end{bmatrix},$$

it follows that the level-1 error structure $\mathbf{X}_{(1)}\,\mathbf{\Phi}_{(1)}\,\mathbf{X}'_{(1)ij}$ is equivalent to the covariance structure of a logit model under the multinomial assumption as given by (7.2.15), with $\hat{\Phi}_{(1)_{kk}}$ an estimate of (cf. (7.2.17)) Ψ_{kk}, $k = 1, 2, \ldots, 6$.

The vector of fixed parameters, $\boldsymbol{\beta}$, can be written as

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{CONSTANT} \\ \beta_{GENDER} \\ \beta_{QUAL1} \\ \beta_{QUAL2} \end{bmatrix}$$

while the vector of random coefficients for level-3 of the hierarchy is given by

$$\mathbf{v}_i = \begin{bmatrix} v_{i,CONS} \\ v_{i,GENDER} \end{bmatrix}.$$

Finally, $\mathbf{u}_{ij}$ can be written as

$$\mathbf{u}_{ij} = \begin{bmatrix} u_{ij,QUAL1} \\ u_{ij,QUAL2} \end{bmatrix}.$$

No allowance is made for association between the coefficients of the variables CONS and GENDER on level-3 of the model.

The following computer output is obtained:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	STD.ERR	Z-VALUE	PR>|Z|
GENDER	-0.1962	0.0536	-3.6604	0.0002
QUAL1	-0.0871	0.0789	-1.1039	0.2696
QUAL2	-0.1802	0.0794	-2.2695	0.0232
CONS	0.8483	0.0576	14.7274	0.0000

(ii) **Random part of the model:**

Level-3:

PARAMETER	$\hat{\tau}$	STD.ERR	Z-VALUE	PR>|Z|
GENDER/GENDER	0.0746	0.0312	2.3910	0.0168
CONS/CONS	0.0805	0.0354	2.2740	0.0229

Level-2:

QUAL1/QUAL1	0.3759	0.0996	3.7742	0.0002
QUAL2/QUAL1	-0.2898	0.0814	-3.5602	0.0004
QUAL2/QUAL2	0.3409	0.1021	3.3389	0.0008

Level-1:

WEIGHT1/WEIGHT1	4.755	1.012	4.699	0.0000
WEIGHT2/WEIGHT2	4.631	0.951	4.870	0.0000
WEIGHT3/WEIGHT3	5.496	1.015	5.415	0.0000

WEIGHT4/WEIGHT4	6.190	1.171	5.286	0.0000
WEIGHT5/WEIGHT5	6.311	1.128	5.595	0.0000
WEIGHT6/WEIGHT6	9.934	1.657	5.995	0.0000

From Part (i) of the output it follows that the estimate of the intercept is 0.8483. This estimate may be interpreted as the expected value of the response variable y, assuming no gender or qualification effect. The coefficient of -0.1962 for GENDER indicates a total gender effect of 0.3924. A negative effect of -0.1962 is obtained for male employees and a positive effect of 0.1962 is obtained for female employees. This indicates a slightly higher value of the response variable for females.

The expected value of the response variable for different levels of education, given that there is no gender effect, can be written as

$$ln \frac{\widehat{f}_{ijk,1}}{\widehat{f}_{ijk,2}} = 0.8483 - 0.0871 \text{ QUAL1} - 0.1802 \text{ QUAL2} \qquad (k = 1, 2, 3, 4) .$$

Similarly, the expected value of the response variable for an employee with an education level lower than Std 10 is then obtained as

$$ln \frac{\widehat{f}_{ijk,1}}{\widehat{f}_{ijk,2}} = 0.8483 - 0.0871 = 0.7612 .$$

Since

$$\frac{\widehat{f}_{ijk,1}}{\widehat{f}_{ijk,2}} = \frac{\widehat{f}_{ijk,1}}{f_{ijk.} - \widehat{f}_{ijk,1}} ,$$

it follows that

$$\widehat{f}_{ijk,1} = f_{ijk.} \left(\frac{e^{0.7612}}{1 + e^{0.7612}} \right)$$

$$= 0.6816 \, f_{ijk.}$$

and

$$\widehat{f}_{ijk,2} = 0.3184\, f_{ijk.}\ .$$

It also follows from the above that, in the case of employees with the lowest level of education, 68 % of the responses are expected to occur in the first category, that is negative response with respect to the presence of gender discrimination in the workplace.

All the estimates of the fixed coefficients, with the exception of the coefficient for QUAL1, are significant at a 5 % level of significance.

From Part (ii) of the output it can be seen that the estimates of the variances/covariances are significant at a 5 % level of significance. It thus follows that the coefficients of GENDER and of CONS vary significantly over the branches, that is the level-3 units. The same conclusion is reached with regard to the variables QUAL1 and QUAL2 on level-2. It can be concluded that the estimated responses obtained through fitting of this model will be more accurate than those obtained by using an aggregated logit model. Furthermore, by allowing for random variation over the different levels of the hierarchy, more accurate standard errors of the logit model parameters (the elements of $\hat{\beta}$) are obtained. Values of standard errors and p-values are derived under the assumption that $\widehat{\beta}$ is approximately normally distributed for large samples.

Example 7.6.2

In this example a level-3 model with the perception of racial discrimination as response variable, is considered. The five possible outcomes for this question were recoded to three categories. The first recoded category consists of negative outcomes (scores 1 and 2), the second category contains the 'don't know' responses and the third category includes the positive responses (scores 4 and 5).

The level-3 units and level-2 units are the 37 branches in the Health Services unit and the two occupational groups, as given in Example 7.6.1, respectively.

Within each level-2 unit, a maximum of six subpopulations are formed according to the gender and age of employees, as shown below for the (i,j)-th unit.

Gender	Age	Number of responses			Total
		Negative	Don't know	Yes	
Male	18 to 29 years	$f_{ij1,1}$	$f_{ij1,2}$	$f_{ij1,3}$	$f_{ij1.}$
Male	30 to 49 years	$f_{ij2,1}$	$f_{ij2,2}$	$f_{ij2,3}$	$f_{ij2.}$
Male	50 years and older	$f_{ij3,1}$	$f_{ij3,2}$	$f_{ij3,3}$	$f_{ij3.}$
Female	18 to 29 years	$f_{ij4,1}$	$f_{ij4,2}$	$f_{ij4,3}$	$f_{ij4.}$
Female	30 to 49 years	$f_{ij5,1}$	$f_{ij5,2}$	$f_{ij5,3}$	$f_{ij5.}$
Female	50 years and older	$f_{ij6,1}$	$f_{ij6,2}$	$f_{ij6,3}$	$f_{ij6.}$

The frequencies $f_{ijk,1}$, $f_{ijk,2}$ and $f_{ijk,3}$ denote the number of negative, don't know and positive responses respectively while $f_{ijk.}$ is the total number of responses for the k-th subpopulation, $k = 1, 2, \ldots, 6$.

The vector of responses is formed by using the third category, that is the number of positive responses, as reference category. For each subpopulation, two elements of the vector of responses are formed as follows:

$$\mathbf{y}_{ij} = \begin{bmatrix} ln\,\dfrac{f_{ij1,1}}{f_{ij1,3}} \\ ln\,\dfrac{f_{ij1,2}}{f_{ij1,3}} \\ \vdots \\ \vdots \\ ln\,\dfrac{f_{ij6,1}}{f_{ij6,3}} \\ ln\,\dfrac{f_{ij6,2}}{f_{ij6,3}} \end{bmatrix} .$$

The model obtained is a level-3 logit model where each (i,j)-combination has a maximum of 12 values of the form $y_{ijk} = ln\,\frac{f_{ijk,l}}{f_{ijk,3}}$, $l = 1, 2$. A value of 0.25 was added to each frequency in order to avoid the occurrence of undefined y_{ijk}-values (Fienberg, 1980).

The gender of employees is used as one of the predictors in this model. Two dummy variables, GENDER1 and GENDER2 are created, with values as given in the following table:

Gender	Response variable	Value of GENDER1	Value of GENDER2
Male	$ln\,\frac{f_{ijk,1}}{f_{ijk,3}}$	1	0
Male	$ln\,\frac{f_{ijk,2}}{f_{ijk,3}}$	0	1
Female	$ln\,\frac{f_{ijk,1}}{f_{ijk,3}}$	-1	0
Female	$ln\,\frac{f_{ijk,2}}{f_{ijk,3}}$	0	-1

Two dummy variables for the intercept term are also created, denoted by CONS1 and CONS2, where CONS1 assumes a value of 1 when the response variable is of the form $y_{ijk} = ln\,\frac{f_{ijk,1}}{f_{ijk,3}}$ and 0 otherwise. The variable CONS2 assumes a value of 1 when the response variable is of the form $y_{ijk} = ln\,\frac{f_{ijk,2}}{f_{ijk,3}}$ and 0 otherwise.

An additional four dummy variables are also formed, indicating the age group to which an employee belongs within a particular gender group. Coding for the variables AGE1_1 , AGE1_2, AGE2_1 and AGE2_2 is as shown below for the (i,j)-th unit for the male subpopulations. The variables AGE1_1 and AGE1_2 are the dummy variables for the first response category while the dummy variables for the second response category are AGE2_1 and AGE2_2. The coding of these variables for the female subpopulations are similar.

Gender	Age	Response variable	AGE1_1	AGE1_2	AGE2_1	AGE2_2
Male	20-29 years	$ln\,\frac{f_{ij1,1}}{f_{ij1,3}}$	1	0	0	0
		$ln\,\frac{f_{ij1,2}}{f_{ij1,3}}$	0	0	1	0
Male	30-49 years	$ln\,\frac{f_{ij2,1}}{f_{ij2,3}}$	0	1	0	0
		$ln\,\frac{f_{ij2,2}}{f_{ij2,3}}$	0	0	0	1
Male	50 + years	$ln\,\frac{f_{ij3,1}}{f_{ij3,3}}$	−1	−1	0	0
		$ln\,\frac{f_{ij3,2}}{f_{ij3,3}}$	0	0	−1	−1

The level-3 model can then be written as (cf. (7.4.5))

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)ij}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}_{(2)ij}\,\mathbf{u}_{ij} + \mathbf{X}_{(1)ij}\,\mathbf{e}_{ij}$$

where $\mathbf{X}_{(f)ij}$ denotes the fixed parameter design matrix and the matrices $\mathbf{X}_{(3)ij}$, $\mathbf{X}_{(2)ij}$ and $\mathbf{X}_{(1)ij}$ are the random-parameter design matrices on level-3, level-2 and level-1 of the hierarchy respectively. The matrix $\mathbf{X}_{(f)ij}$ contains eight columns representing the variables CONS1, CONS2, GENDER1, GENDER2 and AGE1_1 to AGE2_2. If it is assumed that all the subpopulations are present within the (i,j)-th unit, $\mathbf{X}_{(f)ij}$ is given by

$$
\mathbf{X}_{(1)ij} = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 \\
1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 1 & 0 \\
1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\
1 & 0 & -1 & 0 & -1 & -1 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & -1 & -1
\end{bmatrix}.
$$

In this model the intercept terms, denoted by CONS1 and CONS2, are allowed to vary randomly on level-3. From (7.4.6) it follows that

$$
\mathbf{S}_{(3)} = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0
\end{bmatrix} \quad \text{and} \quad \mathbf{X}_{(3)ij} = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1
\end{bmatrix}.
$$

The variables CONS1, CONS2, GENDER1 and GENDER2 are allowed to vary randomly across level-2 of the model. From (7.4.7) it follows that the selection matrix $\mathbf{S}_{(2)}$ and random parameter design matrix $\mathbf{X}_{(2)ij}$ are given by

$$S_{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_{(2)ij} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Since the number of response categories is three, it follows from (7.2.22) that the matrix $\mathbf{X}_{(1)ij}$ can be written as the symmetric matrix

$$\mathbf{X}_{(1)ij} = \begin{bmatrix} \frac{1}{\sqrt{f_{ij1.}}}\mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{f_{ij2.}}}\mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{f_{ij3.}}}\mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{f_{ij4.}}}\mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{f_{ij5.}}}\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{f_{ij6.}}}\mathbf{I}_2 \end{bmatrix}.$$

Each column of $\mathbf{X}_{(1)ij}$ is denoted by WT_K_M, where k denotes the subpopulation and m the response category, $k = 1, 2, \ldots, 6$ and $m = 1, 2$. Thus, 12 WT variables are formed. From (7.2.21) and (7.2.17) it follows that the following constraints may be imposed on the level-1 covariance matrix $\boldsymbol{\Phi}_{(1)}$:

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} \mathbf{\Phi}_{(1)_{11}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{(1)_{22}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Phi}_{(1)_{33}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Phi}_{(1)_{44}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Phi}_{(1)_{55}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Phi}_{(1)_{66}} \end{bmatrix} \tag{7.6.1}$$

where, for example,

$$\mathbf{\Phi}_{(1)_{11}} = \begin{bmatrix} \Phi_{(1)_{11,11}} & \Phi_{(1)_{11,12}} \\ \Phi_{(1)_{11,21}} & \Phi_{(1)_{11,22}} \end{bmatrix}. \tag{7.6.2}$$

From the above it follows that the level-1 error structure $\mathbf{X}_{(1)ij}\,\mathbf{\Phi}_{(1)}\,\mathbf{X}'_{(1)ij}$ is equivalent to the covariance structure of a logistic model with three response categories under the multinomial assumption as given by (7.2.15), with $\hat{\mathbf{\Phi}}_{(1)_{kk}}$ an estimate of $\mathbf{\Psi}_{kk}$, k =1, 2, . . . , 6, as given by (7.2.17).

The vector of fixed parameters, $\boldsymbol{\beta}$, is given by

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{CONS1} \\ \beta_{CONS2} \\ \beta_{GENDER1} \\ \beta_{GENDER2} \\ \beta_{AGE1\_1} \\ \beta_{AGE1\_2} \\ \beta_{AGE2\_1} \\ \beta_{AGE2\_2} \end{bmatrix}.$$

The vector of random coefficients for level-3 of the model is

$$\mathbf{v}_i = \begin{bmatrix} v_{i,CONS1} \\ v_{i,CONS2} \end{bmatrix}$$

while the vector of random coefficients on level-2 of the hierarchy can be written as

$$\mathbf{u}_{ij} = \begin{bmatrix} u_{ij,CONS1} \\ u_{ij,CONS2} \\ u_{ij,GENDER1} \\ u_{ij,GENDER2} \end{bmatrix}.$$

No allowance is made for association between the coefficients of the variables CONS1 and CONS2 on level-3 of the model. On level-2 of the model, no allowance is made for association between the coefficients of CONS1 and CONS2, GENDER1 and GENDER2, GENDER1 and CONS2, and GENDER2 and CONS1.

For this model the following computer output is obtained:

(i) **Fixed part of the model:**

PARAMETER	$\hat{\beta}$	**STD.ERR**	**Z-VALUE**	**PR>|Z|**
CONS1	-0.6901	0.0678	-10.1744	0.0000
CONS2	-1.5864	0.0507	-31.2909	0.0000
GENDER1	0.1467	0.0545	2.6927	0.0071
GENDER2	0.2256	0.0506	4.4626	0.0000
AGE1_1	0.0704	0.0590	1.1935	0.2327
AGE1_2	0.1372	0.0587	2.3359	0.0195
AGE2_1	0.1402	0.0511	2.7401	0.0061
AGE2_2	-0.1053	0.0528	-1.9928	0.0463

(ii) **Random part of the model:**

Level-3:

PARAMETER	$\hat{\tau}$	STD.ERR	Z-VALUE	PR>|Z|
CONS1/CONS1	0.1529	0.0727	2.1035	0.0354
CONS2/CONS2	0.0037	0.0307	0.1199	0.9045

Level-2:

CONS1/CONS1	0.1754	0.0719	2.4399	0.0147
CONS2/CONS2	0.0666	0.0441	1.5108	0.1308
GENDER1/CONS1	-0.0344	0.0364	-0.9455	0.3444
GENDER1/GENDER1	0.1000	0.0395	2.5273	0.0115
GENDER2/CONS2	0.0411	0.0252	1.6352	0.1020
GENDER2/GENDER2	0.0768	0.0336	2.2889	0.0221

Level-1:

WT_1_1/WT_1_1	6.9265	1.2184	5.6850	0.0000
WT_1_1/WT_1_2	3.7459	0.8848	4.2337	0.0000
WT_1_2/WT_1_2	5.7470	1.0063	5.7109	0.0000
WT_2_1/WT_2_1	7.8782	1.2282	6.4142	0.0000
WT_2_1/WT_2_2	4.0118	0.8424	4.7623	0.0000
WT_2_2/WT_2_2	6.0423	0.9489	6.3673	0.0000
WT_3_1/WT_3_1	6.3786	1.1059	5.7678	0.0000
WT_3_1/WT_3_2	3.5482	0.7771	4.5659	0.0000
WT_3_2/WT_3_2	4.6148	0.8098	5.6990	0.0000
WT_4_1/WT_4_1	6.6944	1.0035	6.6712	0.0000
WT_4_1/WT_4_2	3.4392	0.7910	4.3482	0.0000
WT_4_2/WT_4_2	7.8660	1.0743	7.3217	0.0000
WT_5_1/WT_5_1	8.6822	1.4113	6.1519	0.0000
WT_5_1/WT_5_2	3.1477	1.1323	2.7799	0.0054
WT_5_2/WT_5_2	13.4097	1.7892	7.4949	0.0000
WT_6_1/WT_6_1	5.4754	0.8167	6.7038	0.0000
WT_6_1/WT_6_2	2.5519	0.6155	4.1458	0.0000
WT_6_2/WT_6_2	5.9221	0.8122	7.2918	0.0000

From Part (*i*) of the output it can be seen, for example, that the coefficients for GENDER1 and GENDER2 are 0.1467 and 0.2256 respectively. The estimated values of the response variable can be written as

$$ln \frac{\hat{f}_{ijk,1}}{\hat{f}_{ijk,3}} = -0.6901 + 0.1467 \text{ GENDER1} + 0.0704 \text{ AGE1\_1} + 0.1372 \text{ AGE1\_2} \qquad (k = 1, 2, \ldots, 6)$$

and

$$ln \frac{\hat{f}_{ijk,2}}{\hat{f}_{ijk,3}} = -1.5864 + 0.2256 \text{ GENDER2} + 0.1402 \text{ AGE2\_1} - 0.1053 \text{ AGE2\_2} \qquad (k = 1, 2, \ldots, 6)\ .$$

The estimated values of the response variable for the (Female; 30 to 49 years) combination can then be obtained as

$$ln \frac{\hat{f}_{ij5,1}}{\hat{f}_{ij5,3}} = -0.6901 - 0.1467 + 0.1372 = -0.6996$$

and

$$ln \frac{\hat{f}_{ij5,2}}{\hat{f}_{ij5,3}} = -1.5864 - 0.2256 - 0.1053 = -1.9173\ ,$$

so that

$$\hat{f}_{ij5,1} = \hat{f}_{ij5,3}\, e^{-0.6996} = 0.4968\, \hat{f}_{ij5,3} \qquad (7.6.3)$$

$$\hat{f}_{ij5,2} = \hat{f}_{ij5,3}\, e^{-1.9173} = 0.1470\, \hat{f}_{ij5,3}\ . \qquad (7.6.4)$$

From the fact that

$$f_{ij5.} = \hat{f}_{ij5,1} + \hat{f}_{ij5,2} + \hat{f}_{ij5,3}\ ,$$

it follows that the following solution to equations (7.6.3) and (7.6.4) is obtained:

$$\hat{f}_{ij5,1} = 0.3022\, f_{ij5.}\ ,$$

$$\widehat{f}_{ij5,2} = 0.0894\, f_{ij5.}$$

and

$$\widehat{f}_{ij5,3} = 0.6083\, f_{ij5.} \quad .$$

It follows that over 61 % of the respondents in the (Female; 30 to 49 years of age) subpopulation are expected to indicate the presence of racial discrimination in their work environment, while only 30 % of the respondents are likely to indicate the absence of such discrimination in their work environment. Estimated response variable values for the other subpopulations may be obtained in a similar way.

From the random part of the model it follows that the coefficient of CONS1 on level-3 of the hierarchy varies significantly over the level-3 units. On level-2 of the model the coefficients of CONS1/CONS1, GENDER1/GENDER1 and GENDER2/GENDER2 are significant at a 5 % level of significance.

All variances and covariances on level-1 of the model are highly significant. The largest coefficient on level-1 is obtained for WT_5_2/WT_5_2. Note that the coefficients WT_1_1/WT_1_1, WT_1_1/WT_1_2 and WT_1_2/WT_1_2 denote the components of the matrix $\mathbf{\Phi}_{(1)_{11}}$. Coefficients for the covariance matrices $\mathbf{\Phi}_{(1)_{kk}}$ are also given in the computer output.

Example 7.6.3

A multivariate level-3 logit model is considered in this example. Two response variables are used, namely the perception of racial and gender discrimination, which are denoted by QUESTION1 and QUESTION2 respectively. The recoding of the five possible outcomes to these two questions and the level-3 and level-2 units used for this analysis are as described in Example 7.6.2.

Within each level-2 unit, a maximum of six subpopulations are formed as defined in the previous example. A frequency table for the (i,j)-th unit is given below.

Gender	Age	Question	Negative	Don't know	Yes	Total
Male	18 to 29 years	Question1	$f_{ij11,1}$	$f_{ij11,2}$	$f_{ij11,3}$	$f_{ij11.}$
		Question2	$f_{ij12,1}$	$f_{ij12,2}$	$f_{ij12,3}$	$f_{ij12.}$
Male	30 to 49 years	Question1	$f_{ij21,1}$	$f_{ij21,2}$	$f_{ij21,3}$	$f_{ij21.}$
		Question2	$f_{ij22,1}$	$f_{ij22,2}$	$f_{ij22,3}$	$f_{ij22.}$
Male	50 years and older	Question1	$f_{ij31,1}$	$f_{ij31,2}$	$f_{ij31,3}$	$f_{ij31.}$
		Question2	$f_{ij32,1}$	$f_{ij32,2}$	$f_{ij32,3}$	$f_{ij32.}$
Female	18 to 29 years	Question1	$f_{ij41,1}$	$f_{ij41,2}$	$f_{ij41,3}$	$f_{ij41.}$
		Question2	$f_{ij42,1}$	$f_{ij42,2}$	$f_{ij42,3}$	$f_{ij42.}$
Female	30 to 49 years	Question1	$f_{ij51,1}$	$f_{ij51,2}$	$f_{ij51,3}$	$f_{ij51.}$
		Question2	$f_{ij52,1}$	$f_{ij52,2}$	$f_{ij52,3}$	$f_{ij52.}$
Female	50 years and older	Question1	$f_{ij61,1}$	$f_{ij61,2}$	$f_{ij61,3}$	$f_{ij61.}$
		Question2	$f_{ij62,1}$	$f_{ij62,2}$	$f_{ij62,3}$	$f_{ij62.}$

The frequencies $f_{ijk1,1}$, $f_{ijk1,2}$ and $f_{ijk1,3}$, for example, are the number of negative, don't know and positive responses of the k-th subpopulation from the (i,j)-th unit to QUESTION1. The third response category, that is the number of positive responses, is used as reference category. The vector of responses for the k-th subpopulation from the (i,j)-th unit is given by

$$\mathbf{y}_{ijk} = \begin{bmatrix} ln\,\dfrac{f_{ijk1,1}}{f_{ijk1,3}} \\ ln\,\dfrac{f_{ijk1,2}}{f_{ijk1,3}} \\ ln\,\dfrac{f_{ijk2,1}}{f_{ijk2,3}} \\ ln\,\dfrac{f_{ijk2,2}}{f_{ijk2,3}} \end{bmatrix}, \; k = 1, 2, \ldots, 6.$$

The vector of responses for the (i,j)-th unit is then given by

$$\mathbf{y}_{ij} = \begin{bmatrix} \mathbf{y}_{ij1} \\ \mathbf{y}_{ij2} \\ \mathbf{y}_{ij3} \\ \mathbf{y}_{ij4} \\ \mathbf{y}_{ij5} \\ \mathbf{y}_{ij6} \end{bmatrix}$$

which has a maximum of 24 elements if all six subpopulations are present for the (i,j)-th unit. In this example a value of 0.25 (Fienberg, 1980) is added to each observed frequency in order to avoid the occurrence of undefined y_{ijkl}-values.

For each response variable, two dummy variables for gender are created as described in Example 7.6.2. The dummy variables for the first response variable (QUESTION1) are GENDER1_1 and GENDER1_2, while the dummy variables GENDER2_1 and GENDER2_2 are used to indicate gender for the second response variable, that is QUESTION2. The values of these four variables for the fourth subpopulation (Females; 18 to 29 years) are given below as an illustration.

Question	**Response**	**GEN.1_1**	**GEN.1_2**	**GEN.2_1**	**GEN.2_2**
QUESTION1	$ln \frac{f_{ij41,1}}{f_{ij41,3}}$	-1	0	0	0
QUESTION1	$ln \frac{f_{ij41,2}}{f_{ij41,3}}$	0	-1	0	0
QUESTION2	$ln \frac{f_{ij42,1}}{f_{ij42,3}}$	0	0	-1	0
QUESTION2	$ln \frac{f_{ij42,2}}{f_{ij42,3}}$	0	0	0	-1

The intercept term is given by the variables C1_1, C1_2 (for QUESTION1) and C2_1 and C2_2 (for QUESTION2). The coding of these variables and of the dummy variables denoting age for QUESTION1, given by AGE1_Q1_1, AGE1_Q1_2, AGE1_Q2_1, AGE1_Q2_2 are as described in Example 7.6.2. There are two dummy variables for age for each response of the form $ln \frac{f_{ijk1,m}}{f_{ijk1,3}}$, $m = 1, 2$. Similarly, four dummy variables are created for QUESTION2.

The level-3 model can be written as (cf. (7.4.11))

$$\mathbf{y}_{ij} = \mathbf{X}_{(f)}\,\boldsymbol{\beta} + \mathbf{X}_{(3)ij}\,\mathbf{v}_i + \mathbf{X}^*_{(2)ij}\,\mathbf{u}^*_{ij}$$

where $\mathbf{X}_{(f)ij}$ is the fixed parameter design matrix which contains 16 columns representing the variables C1_1 to C2_2, GENDER1_1 to GENDER2_2 and AGE1_Q1_1 to AGE2_Q2_2. Assuming that all six subpopulations are present in the (i,j)-th unit, $\mathbf{X}_{(f)ij}$ is given by

$$\mathbf{X}_{(f)ij} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix} \otimes \mathbf{I}_4 \,.$$

The intercept term, denoted by C1_1, C1_2, C2_2 and C2_2, is allowed to vary randomly on level-3 of the model, so that

$$\mathbf{S}_{(3)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes \mathbf{I}_4$$

and therefore

$$\mathbf{X}_{(3)ij} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \otimes \mathbf{I}_4 \, .$$

If the intercept term and gender are allowed to vary randomly across level-2 of the model, the random parameter design matrix $\mathbf{X}_{(2)ij}$ is given by

$$\mathbf{X}_{(2)ij} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \otimes \mathbf{I}_4$$

while the random parameter design matrix $\mathbf{X}_{(1)ijk}$ is given by

$$\mathbf{X}_{(1)ijk} = \begin{bmatrix} \frac{1}{\sqrt{f_{ijk1.}}}\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{f_{ijk2.}}}\mathbf{I}_2 \end{bmatrix} \, .$$

where $\mathbf{I}_2$ is an identity matrix of order $c-1 = 2$. Each column of $\mathbf{X}_{(1)ij}$ is denoted by WT_K_LM, where k denotes the subpopulation, l the response variable and m the response category, k = 1, 2, . . . , 6; l = 1, 2 and m = 1, 2. The matrix $\mathbf{X}^*_{(2)ij}$ (cf. (7.4.11)) is formed from $\mathbf{X}_{(2)ij}$ and $\mathbf{X}_{(1)ij}$ where

$$\mathbf{X}_{(1)ij} = \begin{bmatrix} \mathbf{X}_{ij1} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{ij2} & . & . \\ . & . & . & \mathbf{0} \\ \mathbf{0} & . & \mathbf{0} & \mathbf{X}_{ijs} \end{bmatrix}. \tag{7.5.9}$$

Constraints, as given by (7.6.1) and (7.6.2) are imposed on the elements of the level-1 covariance matrix $\mathbf{\Phi}_{(1)}$. In order to illustrate the flexibility of the computer program, these additional equality constraints are imposed

$$\mathbf{\Phi}_{(1)_{11}} = \mathbf{\Phi}_{(1)_{22}} = \mathbf{\Phi}_{(1)_{33}} = \mathbf{\Phi}_{(1)_{44}} = \mathbf{\Phi}_{(1)_{55}} = \mathbf{\Phi}_{(1)_{66}} .$$

The vector of fixed parameters, $\boldsymbol{\beta}$, is given by

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{C1\_1} \\ \beta_{C1\_2} \\ \beta_{C2\_1} \\ \beta_{C2\_2} \\ \beta_{GENDER1\_1} \\ \beta_{GENDER1\_2} \\ \beta_{GENDER2\_1} \\ \beta_{GENDER2\_2} \\ \beta_{AGE1\_Q1\_1} \\ \beta_{AGE1\_Q1\_2} \\ \beta_{AGE1\_Q2\_1} \\ \beta_{AGE1\_Q2\_2} \\ \beta_{AGE2\_Q1\_1} \\ \beta_{AGE2\_Q1\_2} \\ \beta_{AGE2\_Q2\_1} \\ \beta_{AGE2\_Q2\_2} \end{bmatrix} .$$

while the vector of random coefficients on level-3 of the model can be written as

$$\mathbf{v}_i = \begin{bmatrix} v_{i,C1\_1} \\ v_{i,C1\_2} \\ v_{i,C2\_1} \\ v_{i,C2\_2} \end{bmatrix}.$$

The vector of random parameters on level-2 of the hierarchy is given by

$$\mathbf{u}_{ij} = \begin{bmatrix} u_{ij,C1\_1} \\ u_{ij,C1\_2} \\ u_{ij,C2\_1} \\ u_{ij,C2\_2} \\ u_{ij,GENDER1\_1} \\ u_{ij,GENDER1\_2} \\ u_{ij,GENDER2\_1} \\ u_{ij,GENDER2\_2} \end{bmatrix}.$$

No allowance is made for association between the coefficients of the variables C1_1, C1_2, C2_1 and C2_2. Allowance is made, however, for association between the coefficients of C1_1 and GENDER1_1, C1_2 and GENDER1_2, C2_1 and GENDER2_1 and C2_2 and GENDER2_2 on level-2 of the hierarchy.

The following computer output is obtained when fitting the model described above to the data:

(i) Fixed part of the model:

PARAMETER	$\hat{\beta}$	STD. ERR.	Z-VALUE	PR>|Z|
C1_1	-0.8081	0.0862	-9.3759	0.0000
C1_2	-1.9958	0.0719	-27.7434	0.0000
C2_1	-1.8213	0.0858	-21.2264	0.0000

C2_2	-2.5609	0.0792	-32.3419	0.0000
GENDER1_1	0.1662	0.0661	2.5138	0.0119
GENDER1_2	0.1798	0.0744	2.4155	0.0157
GENDER2_1	0.1449	0.0767	1.8893	0.0589
GENDER2_2	0.1062	0.0658	1.6138	0.1066
AGE1_Q1_1	0.1062	0.0762	1.3932	0.1636
AGE1_Q1_2	0.2149	0.0796	2.6997	0.0069
AGE1_Q2_1	0.0435	0.0727	0.5986	0.5495
AGE1_Q2_2	0.2031	0.0740	2.7429	0.0061
AGE2_Q1_1	0.1937	0.0637	3.0437	0.0023
AGE2_Q1_2	-0.0267	0.0668	-0.3999	0.6892
AGE2_Q2_1	0.2466	0.0606	4.0681	0.0000
AGE2_Q2_2	-0.0189	0.0619	-0.3050	0.7604

(ii) Random part of the model:

Level-3:

PARAMETER	$\hat{\tau}$	STD. ERR.	Z-VALUE	PR>|Z|
C1_1/C1_1	0.2612	0.1033	2.5277	0.0114
C1_2/C1_2	0.0185	0.0538	0.3438	0.7310
C2_1/C2_1	-0.0384	0.0967	-0.3972	0.6912
C2_2/C2-2	0.0502	0.0770	0.6529	0.5138

Level-2:

C1_1/C1_1	0.2302	0.1009	2.2821	0.0224
C1_2/C1_2	0.0859	0.0736	1.1672	0.2431
C2_1/C2_1	0.4241	0.1341	3.1630	0.0016
C2_2/C2_2	0.2423	0.0980	2.4729	0.0134
C1_1/GEN1_1	0.0149	0.0525	0.2842	0.7762
GEN1_1/GEN1_1	0.1036	0.0541	1.9153	0.0555
C1_2/GEN1_2	0.0003	0.0446	0.0062	0.9951
GEN1_2/GEN1_2	0.2070	0.0703	2.9425	0.0033
C2_1/GEN2_1	-0.1024	0.0598	-1.7142	0.0865
GEN2_1/GEN2_1	0.2994	0.0782	3.8302	0.0001

C2_2/GEN2_2	0.0245	0.0366	0.6709	0.5023
GEN2_2/GEN2_2	0.1208	0.0521	2.3196	0.0204

Level-1:

WT_K_11/WT_K_11	10.6453	0.6823	15.6011	0.0000
WT_K_11/WT_K_12	4.7568	0.5227	9.1009	0.0000
WT_K_12/WT_K_12	11.8300	0.7459	15.8603	0.0000
WT_K_11/WT_K_21	4.2081	0.4817	8.7355	0.0000
WT_K_12/WT_K_21	2.3353	0.4761	4.9052	0.0000
WT_K_21/WT_K_21	9.5360	0.6240	15.2827	0.0000
WT_K_11/WT_K_22	1.8528	0.4555	4.0677	0.0000
WT_K_12/WT_K_22	4.8164	0.5109	9.4273	0.0000
WT_K_21/WT_K_22	2.8109	0.4497	6.2502	0.0000
WT_K_22/WT_K_22	10.0752	0.6445	15.6315	0.0000

From Part (i) of the output it can be seen that the intercept terms, denoted by the variables C1_1, C1_2, C2_1 and C2_2 are highly significant. The gender group to which a subpopulation belongs is significant in the case of QUESTION1, that is the presence of racial discrimination. It is, surprisingly enough, not significant where gender discrimination is concerned. Of the eight coefficients for age, four are significant at a 5 % level of significance.

For QUESTION1, the estimated values of the response variable for the first two response categories are given by

$$y_{ijk1,1} = -0.8081 + 0.1662\ \text{GENDER1\_1} + 0.1062\ \text{AGE1\_Q1\_1} + 0.1937\ \text{AGE2\_Q1\_1}$$

and

$$y_{ijk1,2} = -1.9958 + 0.1798\ \text{GENDER1\_2} + 0.2149\ \text{AGE1\_Q1\_2} - 0.0267\ \text{AGE2\_Q1\_2},$$

while the estimated values for QUESTION2 are given by

$$y_{ijk2,1} = -1.8213 + 0.1449\ \text{GENDER2\_1} + 0.0435\ \text{AGE1\_Q2\_1} + 0.2466\ \text{AGE2\_Q2\_1}$$

and

$$y_{ijk2,2} = -2.5609 + 0.1062\ \text{GENDER2\_2} + 0.2031\ \text{AGE1\_Q2\_2} - 0.0189\ \text{AGE2\_Q2\_2}\ .$$

If, for example, the (Female; 30 to 49 years) subpopulation ($k = 5$) is considered, the respective estimated values are

$$ln\ \frac{\hat{f}_{ij51,1}}{\hat{f}_{ij51,3}} = -\ 0.8081\ -\ 0.1662 + 0.1937 = -\ 0.7806$$

$$ln\ \frac{\hat{f}_{ij51,2}}{\hat{f}_{ij51,3}} = -\ 1.9958\ -\ 0.1798\ -\ 0.0267 = -\ 2.2023$$

$$ln\ \frac{\hat{f}_{ij52,1}}{\hat{f}_{ij52,3}} = -\ 1.8213\ -\ 0.1449 + 0.2466 = -\ 1.7196$$

$$ln\ \frac{\hat{f}_{ij52,2}}{\hat{f}_{ij52,3}} = -\ 2.5609\ -\ 0.1062\ -\ 0.0189 = -\ 2.6860\ ,$$

so that

$$\hat{f}_{ij51,1} = \hat{f}_{ij51,3}\ e^{-\ 0.7806} = 0.4581\ \hat{f}_{ij51,3}$$

$$\hat{f}_{ij51,2} = \hat{f}_{ij51,3}\ e^{-\ 2.2023} = 0.1105\ \hat{f}_{ij51,3}$$

and

$$\hat{f}_{ij52,1} = \hat{f}_{ij52,3}\ e^{-\ 1.7196} = 0.1791\ \hat{f}_{ij52,3}$$

$$\hat{f}_{ij52,2} = \hat{f}_{ij52,3}\ e^{-\ 2.6860} = 0.0682\ \hat{f}_{ij52,3}\ .$$

From the fact that

$$f_{ij51.} = \hat{f}_{ij51,1} + \hat{f}_{ij51,2} + \hat{f}_{ij51,3}$$

and

$$f_{ij52.} = \hat{f}_{ij52,1} + \hat{f}_{ij52,2} + \hat{f}_{ij52,3} \,,$$

it follows that

$$\hat{f}_{ij51,1} = 0.2920 \, f_{ij51.} \,,$$

$$\hat{f}_{ij51,2} = 0.0704 \, f_{ij51.} \,,$$

$$\hat{f}_{ij51,3} = 0.6375 \, f_{ij51.} \,,$$

$$\hat{f}_{ij52,1} = 0.1436 \, f_{ij52.} \,,$$

$$\hat{f}_{ij52,2} = 0.0547 \, f_{ij52.} \,,$$

and

$$\hat{f}_{ij52,3} = 0.8017 \, f_{ij52.} \,.$$

From the above it can be seen that, for QUESTION1, 64 % of the responses are estimated to be positive, compared with 80 % of the responses for QUESTION2. It can thus be concluded that, in general, employees regard gender discrimination to be a bigger problem than racial discrimination. 29 % of the estimated response to QUESTION1 is expected to be negative responses, compared with half that (14 %) in the case of QUESTION2. Estimated response values for other subpopulations may be calculated in the same way.

From the random part of the model, it follows that only the coefficient for the intercept term C1_1/C1_1 is significant at a 5 % level of significance.

Three of the four coefficients for the intercept term on level-2 of the model are significant. The variances of the gender variables are all significant at a 5 % level of significance, but none of the covariances are significant. The largest coefficient on this level is the covariance on the intercept term C2_1/C2_1.

All variances and covariances of level-1 coefficients are highly significant. Highly significant covariances between the response variables show that these variables are associated and that incorrect results may therefore be obtained if separate logit models are fitted for these variables.

7.7 SUMMARY

In this chapter the analysis of models with categorical response variables was considered. A level-3 logit model was introduced and it was shown how this type of model may be analysed using the theoretical framework of general level-3 models given in Chapter 5.

A multivariate level-3 logit model for the simultaneous analysis of more than one categorical response variable was discussed. It was also shown how equality constraints may be used to obtain a specified level-1 error covariance structure.

The theory was implemented in the FORTRAN program GENIGLS and a number of practical applications were given. The importance of simultaneous analysis of associated categorical response variables was illustrated in the last of these applications.

CHAPTER 8
SUGGESTIONS FOR FURTHER RESEARCH

The importance of multilevel modelling cannot be underestimated, as is evident from the fact that a number of software companies have added multilevel analysis to their statistical analysis packages. In the previous chapters an attempt was made to make a contribution to the analysis of data with a hierarchical structure from a complex sampling design. It stands to reason that there is ample opportunity for further research in this regard. There are also a number of problem areas that need to be addressed and computational aspects that require improvement. In the remainder of this chapter a number of areas will be discussed as possible topics for further research in the field of multilevel modelling.

All the models considered in the previous chapters are so-called linear models. Consider for example the following model (cf. (4.2.1)):

$$\mathbf{y} = \mathbf{X}\,\mathbf{b} + \mathbf{e} \tag{8.1}$$

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\mathbf{X}$ the $n \times m$ design matrix, $\mathbf{b}$ the $m \times 1$ vector of stochastic coefficients and $\mathbf{e}$ the $n \times 1$ vector denoting the error terms. From (8.1) it follows that the i-th response is given by

$$\mathrm{y}_i = \mathbf{x}_i'\,\mathbf{b} + \mathrm{e}_i\,, \quad i = 1, 2, \ldots, n.$$

An example of the regression function $\mathbf{x}_i'\,\mathbf{b}$ is the parabola

$$\mathrm{b}_0 + \mathrm{b}_1\,\mathrm{t}_i + \mathrm{b}_2\,\mathrm{t}_i^2$$

where, in this case,

$$\mathbf{x}_i' = \left[\begin{array}{ccc} 1 & \mathrm{t}_i & \mathrm{t}_i^2 \end{array}\right].$$

In many practical situations it may not be realistic to fit a function that is linear in the parameters. This is especially true in the case of time-dependent data, for example the height of children at various ages or the reaction time of rats over a period of time. This data can best be described by means of an asymptotic growth function. Research on repeated measures data with nonlinear expectation functions has been done by, amongst others, Berkey, (1982), Bates and Watts, (1988), Bock (1990) and Herbst (1994).

An example of an asymptotic growth model is the logistic function

$$y_i = f(\mathbf{b}, t_i) + e_i\,, \quad i = 1, 2, \ldots, n \tag{8.2}$$

where

$$f(\mathbf{b}, t_i) = \frac{b_1}{1 + \exp(b_2 b_3^{\,t_i})}\,.$$

The set of regression equations in (8.2) can be written in vector notation as

$$\mathbf{y} = \mathbf{f}(\mathbf{b}, \mathbf{t}) + \mathbf{e} \tag{8.3}$$

where y_i, $f(\mathbf{b}, t_i)$ and e_i denote typical elements of the $n \times 1$ vectors $\mathbf{y}$, $\mathbf{f}(\mathbf{b}, \mathbf{t})$ and $\mathbf{e}$ respectively. For data with an unequal number of measurements per experimental unit, model (8.3) can be written as

$$\mathbf{y}_i = \mathbf{f}(\mathbf{b}_i, \mathbf{t}_i) + \mathbf{e}_i\,, \quad i = 1, 2, \ldots, N \tag{8.4}$$

where $\mathbf{y}_i$, $\mathbf{f}(\mathbf{b}_i, \mathbf{t}_i)$, and $\mathbf{e}_i$ are $n_i \times 1$ vectors with typical elements y_{ij}, $f(\mathbf{b}_i, t_i)$ and e_{ij} respectively, with $j = 1, 2, \ldots, n_i$.

There are various approaches to the analysis of the nonlinear random coefficient model given in (8.4). These approaches include the use of numerical integration techniques (see for example du Toit, 1979), Gibbs sampling (see Boraine, 1995), and first order Taylor linearisation of the response function (du Toit, 1979, Lindstrom & Bates, 1990 and Davidian and Giltinan, 1995).

In order to accommodate the analysis of nonlinear models within the framework outlined in previous chapters, the following approach is suggested. Consider a first order Taylor expansion of $\mathbf{f}(\mathbf{b}_i, \mathbf{t}_i)$ about the mean $\boldsymbol{\beta}$ of the vector of random coefficients $\mathbf{b}_i$, where $\mathbf{b}_i$ is an $m \times 1$ vector

$$\mathbf{y}_i \simeq \mathbf{f}(\boldsymbol{\beta}, \mathbf{t}_i) + \mathbf{J}_i (\mathbf{b}_i - \boldsymbol{\beta}) + \mathbf{e}_i \,. \qquad (8.5)$$

The $n_i \times m$ matrix of first-order derivatives $\mathbf{J}_i$ has typical element

$$[\mathbf{J}_i]_{j,k} = \frac{\partial \mathrm{f}(\mathbf{b}_i, t_{ij})}{\partial \mathrm{b}_{ik}} \mid \mathbf{b}_i = \boldsymbol{\beta} \qquad j = 1, 2, \ldots, n_i \,;\, k = 1, 2, \ldots, m \,.$$

Note that (8.5) can be rewritten as

$$\mathbf{y}_i^* = \mathbf{X}_i \, \mathbf{b}_i^* + \mathbf{e}_i \qquad (8.6)$$

where

$$\mathbf{y}_i^* = \mathbf{y}_i - \mathbf{f}(\boldsymbol{\beta}, \mathbf{t}_i) \,, \qquad (8.7)$$

$$\mathbf{X}_i = \mathbf{J}_i \qquad (8.8)$$

and

$$\mathbf{b}_i^* = (\mathbf{b}_i - \boldsymbol{\beta}) \,. \qquad (8.9)$$

From (8.9) it follows that

$$\mathrm{E}(\mathbf{b}_i^*) = \boldsymbol{\beta}^* = \mathbf{0} \,.$$

After each iteration of the EM algorithm or the IGLS algorithm, the vector $\mathbf{y}_i^*$ and the design matrix $\mathbf{J}_i$ (cf. (8.8)) are updated. A new estimate $\boldsymbol{\beta}_{k+1}$ of $\boldsymbol{\beta}$, is obtained from

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k + \widehat{\boldsymbol{\beta}}^*_{k+1} \,, \quad k = 0, 1, 2, \ldots \,. \qquad (8.10)$$

Initially, $\widehat{\beta}^*$ is set equal to the ordinary least squares estimator $\widehat{\beta}$ of the unknown parameters β assuming a fixed parameter model $\mathbf{f}(\beta, t_i)$.

Du Toit, 1994 has implemented the above procedure for the analysis of nonlinear level-2 models. The extension and implementation of this procedure within the framework of the theoretical results derived in the previous chapters should provide sufficient material for a research project.

Another topic for further research is the incorporation of the imposition of constraints on the elements of the random coefficient matrices $\mathbf{\Phi}_{(1)}$, $\mathbf{\Phi}_{(2)}$ and $\mathbf{\Phi}_{(3)}$. Examples of possible constraints are the following:

(i) The factor analysis structure

$$\mathbf{\Phi}_{(2)} = \mathbf{\Lambda}\,\mathbf{\Lambda}' + \mathbf{D}_{\psi}$$

(ii) A time series structure

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \ddots & \ddots \\ \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \ddots \\ \gamma_2 & \gamma_1 & \gamma_0 & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \gamma_0 \end{bmatrix}$$

(iii) An intra-class correlation structure

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \rho & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \dots & \dots & \rho & 1 \end{bmatrix}$$

The successful implementation of the constraints mentioned above implies that existing time series and factor analysis models can be extended to models for the analysis of hierarchically structured data which will provide more efficient and accurate estimates of the unknown parameters concerned.

It has been pointed out by Bryk and Raudenbush (1987) that inferences based directly on the estimated variances and covariances may be problematical, as these estimates depend heavily on the normality assumption and are also likely to be imprecise in the case of small samples. They also noted that more research is needed on the robustness of estimates to non-normality and on the sample sizes needed for stable estimation. Thus, a further topic for research is an investigation into the small-sample distribution of the standard errors. It may be necessary to perform large simulation studies or to use the Gibbs sampling methodology (see Boraine, 1995, Van der Merwe & Botha, 1993 and Smith & Roberts, 1993) to study the distributional behaviour of the standard errors under non-normality distributional assumptions of the random coefficients.

CHAPTER 9
REFERENCES

AITKIN, M. & LONGFORD, N. (1986). Statistical modelling issues in school effectiveness studies (with discussion). *J. Royal Statist. Soc.*, *A*, **149(1)**, 1-43.

AGRESTI, A. (1990). *Categorical Data Analysis.* New York: Wiley

ANDERSON, T.W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2nd Ed. New York: Wiley.

BATES, D.M. & WATTS, D.G. (1988). *Nonlinear Regression Analysis and its Applications.* New York: Wiley.

BERKEY, C.S. (1982). Bayesian approach for a Nonlinear Growth Model. *Biometrics*, **38**, 953-961.

BISHOP, Y.M., FIENBERG S.E. & HOLLAND, P.W. (1975). *Discrete Multivariate Analysis.* The M.I.T. Press, Cambridge, Mass.

BOCK, R.D. (1983). Within-subject experimentation in psychiatric research. In: Gibbons, R.D. and Dysken, M.W. eds. *Statistical and methodological advances in Psychiatric Research.* New York: Spectrum.

BOCK, R.D. (1990). *Multilevel analysis of educational data.* R.D. Bock (Ed.). San Diego: Academic Press.

BORAINE, H. (1995). *The analysis of repeated measures models under non-standard distributional assumptions.* Unpublished Ph. D. thesis. Pretoria: University of Pretoria.

BROWNE, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *SA Statistical Journal*, **8(1)**.

BROWNE, M.W. (1991). STA-401: Matrix methods in statistics. Supplementary Notes. University of South Africa, Pretoria.

BROWNE, M.W. & DU TOIT, S.H.C. (1992). Automated fitting of non-standard models for mean vectors and covariance matrices. *Multivariate Research*, **27**, 2, 269-300.

BRYK, A.S. & RAUDENBUSH, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. California: Sage.

BRYK, A.S. & RAUDENBUSH, S.W. (1987). Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin*, **101**, 1, 147-158.

DAVIDIAN, M. & GILTINAN, D.M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.

DEGROOT, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.B. (1977). Maximum Likelihood Estimation from Incomplete Data via the EM algorithm. *J. Royal Statist. Soc.*, *B*, **39**, 1-38.

DEMPSTER, A.P., RUBIN, D.B. & TSUTAKAWA, R.K. (1981). Estimation in covariance component models. *J. Amer. Statist. Assoc.*, **76**, 341-353.

DU TOIT, S.H.C. (1979). *The analysis of growth curves*. Unpublished Ph. D. thesis. Pretoria: University of South Africa.

DU TOIT, S.H.C. (1993). *The analysis of multilevel models*. Pretoria: HSRC.

DU TOIT, S.H.C. (1994). *An overview of methods for the analysis of nonlinear random parameter models*. SASA Annual Conference, Kruger National Park, South Africa.

DU TOIT, S.H.C. & LAMPRECHT, T. (1986). *Analysing data with an ordinal dependent variable*. Occasional paper no 12. Pretoria: HSRC.

DU TOIT, S.H.C. & STRASHEIM, C. (1987). *The Optimal Scaling Technique for the analysis of Multivariate Categorical Data.* Office report. Pretoria: HSRC.

FIENBERG, S.E. (1980). *The analysis of Cross-Classified Categorical Data.* 2nd ed. The M.I.T. Press, Cambridge, Mass.

GIBBONS, R.D. & BOCK, R.D. (1987). Trend in correlated proportions. *Psychometrika*, **52**, 113-124.

GOLDSTEIN, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, **73**, 1, 43-56.

GOLDSTEIN, H. (1987). *Multilevel models in educational and social research.* C. Griffin and Co., London, Oxford University Press, New York.

GOLDSTEIN, H. & MACDONALD, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, **531**, 455-467.

GOLDSTEIN, H. (1989). Restricted unbiased iterative generalized least squares estimation. *Biometrika*, **76**, 662-3.

GOLDSTEIN, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, **78**, 1, 45-51.

GOLDSTEIN, H. (1991). Multilevel modelling of survey data. *The Statistician*, **40**, 235-244.

GOODMAN, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.*, **74**, 537-552.

GRAHAM, A. (1981). *Kronecker products and Matrix Calculus: with applications.* Ellis Horwood Limited, New York: Wiley.

HERBST, A. (1994). *Statistical modelling of growth.* Unpublished Ph. D. thesis. Pretoria: University of Pretoria.

HOLT, D., SCOTT, A.J. & EWINGS, P.D. (1980). Chi-squared tests with survey data. *J. Royal Statist. Soc., A*, **143**, 303-320.

HUYNH, H. (1986). Estimation of the KR20 reliability coefficient when the data are incomplete. *British Journal of the Mathematical and Statistical Psychology*, **39**, 69-78.

KHATRI, C.G. (1966). A note on a MANOVA model applied to problems in growth *curves. Ann. Inst. Statist. Math.*, **181**, 75-86.

KREFT, I.G.G., DE LEEUW, J. & VAN DER LEEDEN, R. (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *The American Statistician*, **48 (4)** , 324-335.

LINDSTROM, M.J. & BATES, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673-687.

LONGFORD, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika*, **741**, 4, 817-827.

MAGNUS, J.R. & NEUDECKER, H. (1988). *Matrix differential calculus with applications in statistics and econometrics.* New York: Wiley.

MALINVAUD, E. (1970). *Statistical Methods of Econometrics.* Amsterdam: North Holland Publishing Company.

MASON, W.M., WONG, G.M. & ENTWISTLE, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed). *Sociological methodology*, 72-103. San Francisco: Jossey-Bass.

McCULLAGH, P. (1980). Regression models for ordinal data. *J. Royal Statist. Soc., B*, **42**, 109-142.

McCULLOCH, C.E. (1982). Symmetric matrix derivatives with applications. *J. Amer. Statist. Assoc.*, **77**, 679-682.

MORRISON, D.F. (1991). *Multivariate Statistical Methods.* New York: McGraw-Hill.

POTTHOFF, R.F. & ROY, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313-26.

PRESS, W.H., FLANNERY, B.P., TEUKOLSKY, S.A. & VETTERING, W.T. (1988). *Numerical recipes in Fortran. The art of scientific computing.* USA: Cambridge University Press.

PROSSER, R., RASBASH, J. & GOLDSTEIN, H. (1991). *ML3 Software for three-level analysis.* London: Institute of Education.

RASBASH, J. (1993). *ML3E Version 2.3 Manual Supplement.* London: Institute of Education.

RAUDENBUSH, S.W. & BRYK, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, **59**, 1-17.

SAS/STAT Users Guide Version 6.4 Volume 1 (1990). SAS Institute Inc., Carey, North Carolina.

SMITH, A.F.M. & ROBERTS, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo studies. *J. Royal Statist. Soc.*, *B*, **55**, 1, 3-23.

STOKER, D.J. (1982). *The analysis of categorical data with ordered categories: a review.* Lecture/Paper PR-82. Pretoria: HSRC.

VAN DER MERWE, A.J. & BOTHA, T.J. (1993). Bayesian estimation in mixed linear models using the Gibbs sampler. *SA Statistical Journal*, **27**, 149-180.

APPENDIX
THE COMPUTER PROGRAM MULTVAR

10.1 INTRODUCTION

This appendix consists of two sections. In Section 10.2 the computer program MULTVAR, used in Chapter 6 for the analysis of data with more than one continuous response variable, is compared with the program ML3E (Prosser, Rasbash and Goldstein, 1991). Information concerning the installation and running of MULTVAR is given in Section 10.3. The program MULTVAR was compiled with FORTRAN LAHEY F77.

10.2 COMPARISON OF MULTVAR AND ML3E

In this section the computer program MULTVAR, used in Chapter 6, is compared with the commercially available program ML3E (see Section 2.4). Use is made of the data sets discussed in Examples 6.5.1 and 6.5.2. The output files produced by these programs are given on the accompanying diskette. Also included on the diskette are the input files and the data sets used for this comparison. In order to carry out these comparisons it was necessary to create separate data sets for ML3E, since this program does not create dummy variables automatically.

As noted in Chapter 2, the size of the data set which can be analysed using ML3E is limited. In the case of Example 6.5.1, it is therefore necessary to use a subset of the data. The data set used for the comparison has four response variables and consists of the first 505 records (the first 18 households) of the data set discussed in Example 6.5.1. In the case of Example 6.5.2 the complete data set is used.

Although the convergence criteria of the two programs differ, the values of the estimated parameters and large-sample standard errors are essentially the same. The comparison between the programs is thus based on a comparison of the time used per iteration. The time (in seconds) required to carry out one iteration of the IGLS procedure for each of the two examples are given in Table 10.2.1.

Table 10.2.1: Comparison of MULTVAR and ML3E

Example	Multvar	ML3E
Example 6.5.1 (subset)	2.5	40
Example 6.5.2	3.0	30

The times reported in the table above were obtained by running the programs on a 486-DX33 computer with 4 MB Random Access Memory.

In the first example there are 1 702 records and two response variables and in the second example 505 records and four response variables. From the table above it can be seen that the time used per iteration by the program MULTVAR is significantly less than that required by ML3E. This trend is even more noticeable if, for a given number of observations, the number of response variables is increased.

10.3 INSTALLATION OF MULTVAR

The file README.DOC on the diskette contains the necessary information to install and run the program MULTVAR. For the sake of completeness, the contents of this file is given below:

```
Contents of diskette:

There are three files on the diskette, namely
README.DOC
PKUNZIP.EXE
MULTVAR.ZIP
```

The file MULTVAR.ZIP contains the following files:

Type of file	Description of contents of file
Program files:	
RUN386.EXE	System file required by MULTVAR.BAT
MULTVAR.BAT	Batch file required to run MULTVAR
MULTVAR.EXP	MULTVAR executable file
Data sets:	
MULTESK.OBS	Data subset (Example 6.5.1) for program MULTVAR
ML3_ESK.OBS	Data subset (Example 6.5.1) for program ML3E
MULTAIDS.OBS	Complete data (Example 6.5.2) for program MULTVAR
ML3_AIDS.OBS	Complete data (Example 6.5.2) for program ML3E
Input files:	
MULTESK.IN	Input file (Example 6.5.1) for program MULTVAR
ML3_ESK.COM	Input file (Example 6.5.1) for program ML3E
MULTAIDS.IN	Input file (Example 6.5.2) for program MULTVAR
ML3_AIDS.COM	Input file (Example 6.5.2) for program ML3E
Output files:	
MULTESK.OUT	Output file (Example 6.5.1) for program MULTVAR
ML3_ESK.LOG	Output file (Example 6.5.1) for program ML3E
MULTAIDS.OUT	Output file (Example 6.5.2) for program MULTVAR
ML3_AIDS.LOG	Output file (Example 6.5.2) for program ML3E

System requirements:

The minimum system requirements to run the program MULTVAR are:

(1) a 386 IBM-compatible system or higher with
(2) a maths coprocessor and
(3) at least 4 MB Random Access Memory (RAM)

Installation of program:

In order to run the examples given on the diskette, the following installation procedure is suggested:

(1) Make the directory DEMO on the hard drive by typing the following command at the C:> prompt:
    C:> md demo
(2) Change the path to this new directory by typing at the C:> prompt:
    C:> cd demo

(3) Copy the three files on the diskette to the hard drive by typing:
C:\DEMO> copy a:*.*
(4) Extract the files listed in the table above by typing:
C:\DEMO> pkunzip multvar.zip

Running of examples:

The program can now be run by typing: C:\DEMO> multvar. The name of the input file, either MULTESK.IN or MULTAIDS.IN, is to be entered as input file. Any name may be used for the output file. If no output file is specified on the first screen, output will be written to the default output file MLEVEL.OUT. To proceed, press the ESC key.

Note that the raw data files on the diskette contain confidential information and may thus only be used for examination purposes.