Adjustment for Measurement Error in Multilevel Analysis

Geoffrey Michael Woodhouse

Thesis submitted for the degree of PhD

Institute of Education
University of London

Abstract

Measurements in educational research are often subject to error. Where it is desired to base conclusions on underlying characteristics rather than on the raw measurements of them, it is necessary to adjust for measurement error in the modelling process.

In this thesis it is shown how the classical model for measurement error may be extended to model the more complex structures of error variance and covariance that typically occur in multilevel models, particularly multivariate multilevel models, with continuous response. For these models parameter estimators are derived, with adjustment based on prior values of the measurement error variances and covariances among the response and explanatory variables. A straightforward method of specifying these prior values is presented.

In simulations using data with known characteristics the new procedure is shown to be effective in reducing the biases in parameter estimates that result from unadjusted estimation. Improved estimates of the standard errors also are demonstrated. In particular, random coefficients of variables with error are successfully estimated.

The estimation procedure is then used in a two-level analysis of an educational data set. It is shown how estimates and conclusions can vary, depending on the degree of measurement error that is assumed to exist in explanatory variables at level 1 and level 2. The importance of obtaining satisfactory prior estimates of measurement error variances and covariances, and of correctly adjusting for them during analysis, is demonstrated.

Contents

List of figures

List of tables

Acknowledgements

First I want to thank Harvey Goldstein, for the help, encouragement, and support he has given me as my supervisor throughout this work, and for the inspiration that he has been to me and to all the members of the Multilevel Models Project over many years. Nobody who reads this will fail to recognise his influence: I hope that where I have ventured into new territory I have not erred too wildly.

Secondly, I owe a great debt to Jon Rasbash, who has always responded to my requests for help, and for special features to be built into ML*n*, with cheerfulness and great speed. Without the macro language that he wrote, and a number of other enhancements, this thesis would not exist.

I also thank Min Yang and Huiqi Pan for their many kindnesses over a long period, as I have struggled with this work.

Finally, I must thank my wife, Kate, who has borne my abnormal working hours with long-suffering and generosity. I hope that now this is finished some balance may return to our lives.

1 Introduction

Most measurements in educational and other social research are subject to error, in the sense that an independent repetition of the measurement process does not produce an identical result. For example, measurements of cognitive outcomes in schools such as scores on standardised tests can be affected by item inconsistency, by fluctuations within individuals, and by differences in the administration of the tests and in the environment of the schools and classes where the tests take place. Measurements of non-cognitive outcomes also, such as behaviour, self-concept, and attitudes, can be similarly affected.

Of course, it is important to distinguish between measurement error in the sense just outlined and growth or change over time in the subject being measured. Longitudinal studies of educational progress, for example, depend for their validity on that distinction. In addition in this context certain forms of repeated testing, which might appeal on the ground of minimising unwanted fluctuation, are ruled out because learning and other effects associated with repetition lead to *bias* in the measurement.

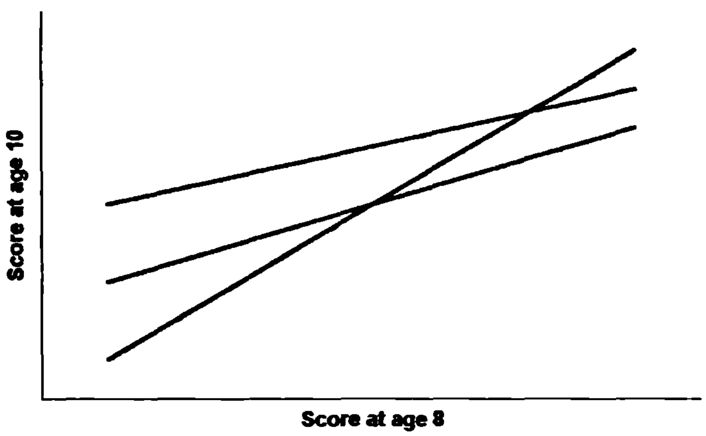
As is explained in more detail in the next chapter, this thesis does not deal with systematic bias in measurement. But it is well known that random fluctuation in measurement, if untreated in analysis, can itself lead to biased estimates of effects and hence to mistaken causal attributions (see, for example, Goldstein, 1979, p134). Biemer and Trewin (1997, pp628-9) have tabulated the effects of random errors on parameter estimators and data analysis in the context of surveys. Plewis (1985) reviewed methods of correcting for measurement error proposed by Degracie and

Fuller (1972) and by Jöreskog (1970) in the context of longitudinal studies, and explored the effects of various methods on the conclusions obtained. Fuller (1987) has given an account of methods for dealing with errors of measurement in regression models.

All of these studies are based on classical, single-level, methods of analysis. Social research data, however, often have a hierarchical structure and are most efficiently analysed by means of multilevel models, which provide a convenient way to study structures of heterogeneous variance that occur naturally in many populations. Paterson and Goldstein (1991) give a readable introduction, and Goldstein (1995, Chapters 1, 2, and part of 4) provides the background that is prerequisite to understanding this thesis.

A simple example arises when schools differ in their relative 'effectiveness' for different kinds of pupils, as illustrated in Figure 1.1 for a hypothetical data set.

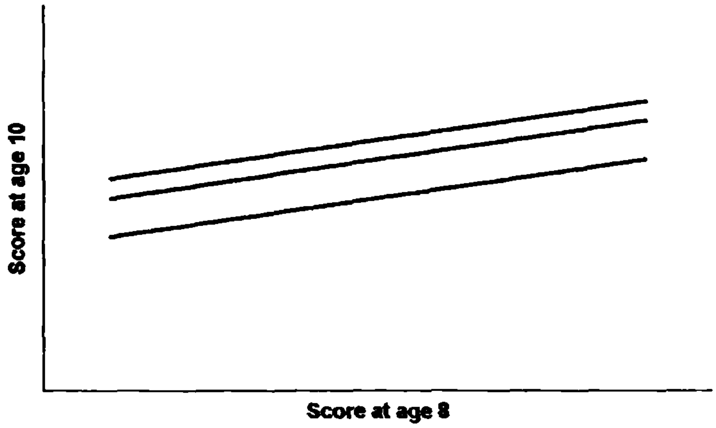
Figure 1.1 School summary lines linking reading scores at age 10 with reading scores at age 8.



Each line in Figure 1.1 is a simple summary for one school which predicts, on the basis of all the data obtained in the sample, how a pupil in that school who obtains a given reading score at age 8 years will perform (on the average) in two years' time. The graph suggests that schools vary more in their predicted scores for low-attaining pupils than for high-attaining ones. This is an example of heterogeneous variance at the level of the school and illustrates a general finding, that it is not always possible to make simple comparisons between groups. Indeed, matters are typically much more complicated than in this illustration. For example, boys may show different average patterns of progress from girls, and may vary from the average pattern in different ways. Such patterns of variation are of considerable research interest, and have policy implications also. See Woodhouse *et al.* (1996a) for more detail on the present example.

Figure 1.1 has further implications. Given a set of estimates such as those illustrated a researcher or policy-maker might conclude that some schools acted so as to 'narrow the gap' over time between the less able and the more able while others did not. But such a conclusion could be false. The measurements taken on the pupils will have been subject to error. In the absence of adjustment during the analysis, this error will have led to bias in the estimates of the slopes of the lines, and in the estimate of the variation between them. As we show in Chapter 4, in such a simple model the bias in the estimated slopes and in their estimated variation would move them towards zero: in other words the tendency would be to obtain a graph such as Figure 1.2 and fail to find a real difference in the slopes. In more complicated models the effects of failing to adjust for measurement error can be substantial but not easily predictable.

Figure 1.2 School summary lines linking reading scores at age 10 with reading scores at age 8 (with measurement error).



Absent from both the above graphs is an indication of the uncertainty associated with the estimates. It is to be expected that measurement error, by introducing noise into the data, should increase this uncertainty and this is confirmed in the analyses we describe later. The effect of such uncertainty, when correctly estimated, is sometimes to obliterate differences like those that appear in Figure 1.1, sometimes not.

Thus, the effect of measurement error on the analysis of multilevel models is pervasive. Failure to adjust for it can lead to a variety of possible misinterpretations of data. Fuller (1991, p618), in reviewing the literature of survey analysis, estimated the fraction of researchers explicitly recognising the presence of measurement error in their analyses to be 'not large'. Despite the uncertainty in the estimate it seems fair to say that this fraction is not exceeded in the population that uses multilevel models. It is also fair to say that the tools available for analysis lag behind the needs of this research community (see Chapter 2). My aim in this thesis is to make a modest contribution towards correcting this problem.

After a brief review of the literature we go on to show, in Chapter 3, how the classical model for random additive errors of measurement may be developed to model some of the more complicated error structures that typically arise in multilevel analysis. For example, errors at more than one level may occur in both response and explanatory variables; errors in response and explanatory variables may covary, again at more than one level; errors may occur in different variables of a multivariate response, some of whose values are missing; errors in multivariate response variables may covary; and so on. We develop a consistent, straightforward method by which the user may specify such complex error structures. We then show how the iterative generalised least squares (IGLS) estimation method described by Goldstein (1986) may be adapted to adjust for these errors when their variances and covariances are known. In particular, we give for the first time a method, based on IGLS, for estimating random coefficients of variables subject to error. We also define a new, clear, and concise notation for the expressions used in the estimation process, a notation that is designed to be adaptable to more complicated problems.

In Chapter 4 we describe the results of simulation studies using the new procedures, and in Chapter 5 we show in an illustrative analysis of a two-level educational data set how estimates and conclusions can vary, depending on the degree of measurement error that is assumed to exist at each level. After discussion in Chapter 6 we review in Chapter 7 the achievements and limitations of the procedures so far developed and suggest directions for further work.

Some of the limitations are self-imposed. This thesis does not deal with non-linear models. Nor are errors in categorical variables treated. Errors at all levels are assumed to have zero expectation over repeated sampling and to be Normally distributed independently of the true values of the variables. These restrictions enable the development in Chapter 3 to be kept within reasonable bounds. Further developments will be reported in due course.

2 Literature Review

Errors of measurement have an extensive literature, and this review of it will be selective. The thesis is concerned with the multilevel analysis of data that are subject to error with a known covariance structure. Accordingly I shall not cite works, important as they are, that are devoted to identifying sources of error and procedures to minimise it. Nor shall I follow the development of measurement error models further than is necessary to justify my own conception and to place it in context. The main methodological references are Fuller (1987), Lord and Novick (1968), and Carroll, Ruppert and Stefanski (1995).

The literature on multilevel models also has burgeoned, particularly in the last five years. But work that explicitly takes measurement errors into account has been rare. Main references in the area are Goldstein (1986, 1995), Bryk and Raudenbush (1992), and Longford (1993a).

2.1 Error, measurement, and true value

A review of measurement error in surveys is provided by O'Muircheartaigh (1997). He was opening the conference *Survey Measurement and Process Quality*, reported in Lyberg *et al* (1997), a conference from whose title the word 'error' had been deliberately excluded in favour of a more general characterisation of the measurement problem and procedures for handling it. His definition of error as 'work purporting to do what it does not do' was in keeping with this more general approach and, as he claimed, forces consideration of the needs for which data are being collected. It is not an operational definition, but it is a useful reminder that measurement error models and adjustment procedures are to be used with care to serve the wider purpose of improved quality in the conduct and reporting of research.

O'Muircheartaigh went on to classify what he called 'dimensions' of sources of error as 'representation, randomisation, and realism'. He also identified three 'perspectives', from government and official statistics, from academic and social research, and from commercial and market research. Following Kiaer (1897) the dominant methodology in the 'official statistics' strand became that of the sample survey, and before Mahalonobis (1946) survey statisticians were concerned mainly with measuring the effect of sample design on the precision of survey estimates. According to O'Muircheartaigh, Mahalonobis was the first statistician to emphasise the human agency in surveys, and he classified errors as those of sampling, recording, and physical fluctuations. His work was built on by Hansen *et al* (1961) in what has become known as the *U.S. Bureau of the Census* model (see also Groves, 1991; Fuller, 1991 and 1995; Biemer and Trewin, 1997). From our point of view, the important aspect of this model is its characterisation of an observation as consisting of two parts, a *true value* (relative to what they termed *essential survey conditions*) and an additive *response deviation*.

Kruskal, in his introduction to *Measurement Errors in Surveys* (Biemer *et al*, 1991), touched on the problem of true value. He contrasted the attitude to truth of the historians Parrington (in Trilling, 1951) and Handlin (1979) – briefly, that truth is absolute and the world is real – with that of the statisticians Deming (1950) and Shewhart (1939). Shewhart was 'not able even to conceive of a physical operation of observing a true length *X*' and wished to distinguish clearly between this concept and that of the limiting average of an observed length. A posited true length *X* is an example of a *Platonic* true score or true value (Sutcliffe, 1965), which in this sense means that it is an ideal measurement to which a particular practical measurement is an approximation.

In education, however, we typically study characteristics – for example, mathematical knowledge – that are not directly measurable. Following Lord and Novick (1968, sections 2.2 to 2.4), given an individual *i* we conceive of a sequence of independent observations *t* yielding *measurements* X_{ti}, and assume that over this sequence the individual's knowledge (in this example) remains constant, while other effects on X_{ti}, for example, of the subject's emotional state, or of the environment of the test, are random. We further assume the existence of a *propensity distribution* of the X_{ti}, being the probability distribution function defined over repeated statistically independent measurements on the same individual *i*. Then the *true score* of individual *i* on measurement *X* is the quantity x_i,

$$x_i = \mathrm{E}\left(X_{ti}\right), \tag{2.1}$$

where the expectation is with respect to the propensity distribution, and is assumed to be finite. Typically in our work we have only one measurement per individual, X_i (say), and we write

$$x_i = \mathrm{E}\left(X_i\right). \tag{2.2}$$

We generally use the term *true value of* X_i, rather than true score.

This is an example of a *classical* true score or value. Lord and Novick (1968, section 5.4) go on to demonstrate that x_i with the definition (2.1) can be interpreted semantically as the average value of X_{ti} over infinitely many independent repeated measurements, thus identifying for an unobservable trait the two concepts that Shewhart wished to keep separate for lengths.

The discrepancy between the true value x_i and an observed value X_i is called the *measurement error* of X for the *i*th individual, or simply the *error in* X_i. We shall denote this by ξ_i. By definition,

$$\xi_i \equiv X_i - x_i = X_i - \mathrm{E}(X_i). \tag{2.3}$$

It follows immediately that

$$\mathrm{E}(\xi_i) = 0. \tag{2.4}$$

The propensity distribution variance for individual *i*, otherwise known as the *measurement error variance of* X_i, or simply the *error variance of* X_i, is the quantity

$$\mathrm{var}(\xi_i) = \mathrm{var}(X_i) = \mathrm{E}\left[(X_i - x_i)^2\right] \tag{2.5}$$

We allow that the error variance of a measurement of a given characteristic may differ for different individuals.

In particular, Lord and Novick derived:

$$\begin{aligned} \mathrm{E}_i(\xi_i) &= 0, \\ \mathrm{cov}_i(\xi_i, x_i) &= 0, \\ \mathrm{cov}_i(\xi_{1i}, x_{2i}) &= 0, \end{aligned} \tag{2.6}$$

where E_i, cov_i denote expectation and covariance over persons, and subscripts 1 and 2 refer to different measurements. Under the further assumption that X_{1i}, X_{2i} are independently distributed for each i (conditionally on x_i) we have also $\mathrm{cov}_i\left(\xi_{1i},\xi_{2i}\right)=0$. Properties (2.6), together with the latter assumption and the definitions $X_i \equiv x_i + \xi_i$, etc., are the assumptions of the *classical model*. The assumptions are taken to hold in every (non-empty) subpopulation of a given population. We show in section 3.4 how a multilevel formulation allows us to relax some of these assumptions in order to model more complex situations.

2.2 Bias, reliability, and validity

The above definitions and properties imply that the observed values X_i are ***unbiased*** estimators of the true values x_i. Lord and Novick adapted an example from Sutcliffe (1965) of a Platonic true score x_i that necessarily yielded conditionally biased observed scores. In this example the activity is chicken-sexing and the true score x_i for the *i*th chicken is defined as:

$x_i = 0$ if *i*th chicken is a pullet,
1 if a cockerel.

The chicken-sexer briefly examines each chicken and assigns it an observed score X_i on the same scale. Clearly, the expected value of X_i for true pullets is the probability that a pullet is misclassified as a cockerel, and for true cockerels it is the probability that a cockerel is correctly classified. Thus (unless the chicken-sexer does a perfect job) X_i is positively biased for pullets and negatively biased for cockerels and, in addition to bias in the X_i, we also have $\mathrm{cov}_i\left(\xi_i,x_i\right)<0$, which contradicts another of the assumptions of the classical model. These problems may be removed in this case by transforming the (arbitrary) true-score scale to coincide with the probabilities above and retaining the observed-score scale as originally defined.

As Lord and Novick pointed out, there are certain scales (in particular, absolute scales based on counts) that do not admit such transformations. Many of the scales used in education, however, are arbitrary. In this thesis we develop procedures to adjust for measurement error in variables that may be regarded as continuous, and we shall assume that the measurement scales have been transformed if necessary so that observed values may be regarded as unbiased estimators of the corresponding true values.

The *reliability* of a test is defined by Lord and Novick as the squared correlation between observed score and true score. In our notation this is equal to the ratio $\frac{\mathrm{var}(x_i)}{\mathrm{var}(X_i)}$ where the variances are over the members of a defined population of individuals i, and this is Fuller's (1987) definition of the *reliability ratio of* X_i. It is clear from this definition that if $\mathrm{var}(\xi_i)$ is constant over a given population and we restrict attention to a subpopulation over which $\mathrm{var}(x_i)$, or equivalently $\mathrm{var}(X_i)$, is lower than over the whole population, the reliability of X_i in the subpopulation is lower than in the whole population. Thus, as Goldstein (1995) has pointed out, the use of published reliabilities of tests is often inappropriate when these tests are applied to restricted groups. Further problems of definition arise when the error variance is not constant. We use the concept of reliability for illustration only, in Chapters 4 and 5.

For *validity*, it is useful to consider Lord and Novick's distinction between *empirical* and *theoretical* validity (1968, p261), in particular *construct* validity (p278). They were applying these concepts to tests. From our perspective there are two important points. The first is that a measure may have several validities, depending on the criterion that it is supposed to measure. The second is that the notion of validity extends also to the interpretation of model estimates. As to the first point, we assume that a researcher using a substantive model has satisfied herself that the true value of any variable in the model is an adequately close approximation to the corresponding underlying characteristic of interest. As to the second, model-based inference is always open to the challenge that the model is misspecified. The misspecification may take several forms: important variables may be missing, the random structure (in a multilevel model) may be insufficiently elaborated, or the sampling method may not justify the analysis, for example if the sample is an opportunity sample. To this list must be added the possibility of misspecification, or misinterpretation, of measurement error variance and covariance. For example, if the aim is to predict performance (of an individual or school, say) on an observed measure it may not be appropriate to adjust for measurement error in the formation of all the estimators. For a general discussion about valid model inference in the context of school performance indicators see Bryk and Raudenbush (1992, p126).

2.3 Adjustment for measurement error in regression models

Hansen *et al.* (1961) and other survey analysts regarded measurement error chiefly as a component of imprecision in the estimation of a single population parameter. O'Muircheartaigh (1997) observed a subsequent shift in emphasis away from 'design-based' inference to the finite population sampled (based on the joint probabilities of selection of the elements), and towards 'model-based' estimation, in which a formal model relates the different variables in the analysis regardless of the configuration of the sample. In this respect survey analysts were moving closer to the practice of economists, psychologists, and epidemiologists, for example. He gives references to substantive work from the early 1980s. See also Degracie and Fuller (1972) and Fuller (1995).

Fuller (1987) gave a comprehensive treatment of the estimation of singe-level linear models containing measurement error. The simplest such model may be written

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,\ldots,N, \tag{2.7}$$

where $(x_1, x_2, \ldots, x_N)$ is either fixed in repeated sampling (the *functional* model) or a random sample from $\mathrm{N}(\mu_x, \sigma_{xx})$ (the Normal *structural* model) and the ε_i are independent random variables. The x_i are measured by the observed variable X_i,

$$\begin{aligned} X_i &= x_i + \xi_i, \\ \mathrm{E}(\xi_i) &= 0, \\ \mathrm{var}(\xi_i) &= \tau_i, \end{aligned} \tag{2.8}$$

where it is assumed that $\mathrm{cov}(\xi_i, \varepsilon_i) = 0$. Assuming Normality of $\mathbf{x} = \{x_i\}$ and $\boldsymbol{\xi} = \{\xi_i\}$ and writing $\mathbf{X} = \{X_i\}$, Fuller derived consistent estimators for $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$ and for the conditional covariance matrix of the parameter estimators $\mathrm{cov}(\hat{\boldsymbol{\beta}}|\mathbf{X})$, when the ratio $\dfrac{\mathrm{var}(x_i)}{\mathrm{var}(X_i)}$ is constant and known. Under the stronger assumption that $\tau_i = \tau$ where τ is constant and known he derived a consistent estimator for the variance of the first-order approximation of $\hat{\beta}_1$ as a Taylor series about the population values of the moments (the so-called delta approximation).

Fuller extended this work, using the method of maximum likelihood, to the case of vector explanatory variables (1987, Chapter 2) and to the case of unequal error variances and covariances (1987, Chapter 3), where he derived a 'sandwich' estimator for the covariance matrix of the estimators. He also derived estimators of the true x values and true residuals, suitable for model checking.

Carroll *et al.* (1995), in that part of their book that is not specific to non-linear model estimation, went over similar ground to that of Fuller (1987, Chapter 2). In addition, they described the so-called SIMEX method, which uses simulation and extrapolation to estimate parameters of a model based on true values when one of the predictors, X, is measured with known or estimable constant measurement error variance τ. In the simulation phase of this method, for a series of different positive values λ, independent pseudo-random variables with zero mean and variance $\lambda\tau$ are added to the X_i and the model is estimated without adjustment. This yields a sequence of biased estimates of the model parameters. In the extrapolation phase, each biased estimate $\hat{\beta}_p$ is regarded as a function $g_p(\lambda)$, say, with parameters γ_p to be estimated by fitting a model (often quadratic, but possibly non-linear, for example, $g_p(\lambda) = \gamma_{p,0} + \frac{\gamma_{p,1}}{\gamma_{p,2} + \lambda}$) to the estimates produced in the simulation phase. On the assumption that this model may be extrapolated, $g_p(-1)$ then provides an unbiased estimate of β_p. Further details are given in Cook and Stefanski (1994), and the asymptotic behaviour of the estimators and their covariances is described in Carroll *et al.* (1996).

The advantage of the SIMEX method is that it is relatively simple to implement. It can be applied to any model for which an estimation procedure exists. The disadvantage is that it is computer-intensive. In practice, for each value of λ the model is estimated a large number (typically 100) times and mean values of the estimates are taken. To form satisfactory estimates of the $g_p(\lambda)$, the number of values of λ also must be large. Moreover, while the method could be adapted to the case of several predictors with covarying errors, the number of separate estimations required is an exponential function of the number of error variances and covariances. Thus it is unsuitable for model exploration, and of limited use in the estimation of multilevel models, where typically there are multiple predictors, with errors covarying across levels.

Another approach to the treatment of measurement error is through the modelling of latent variables and covariances. In this approach the *manifest* variables measured with error are considered to be indicators of unobservable *latent* variables, and a stochastic model of these dependencies is proposed in order to explain the covariances between the observed variables. The approach, sometimes known as *structural modelling*, is capable of considerable generalisation. We describe in the next section an adaptation to the multilevel case.

2.4 Multilevel models and measurement error

Educational, social, and biological data frequently arise from populations that are hierarchical in structure. For example, school pupils are taught in classes within schools, people live in groups within neighbourhoods; and schools, groups, neighbourhoods and so on may be clustered into larger units, perhaps geographically or administratively. Further examples of hierarchical structure occur when individuals are measured repeatedly.

It has long been known that clustering induces correlations among the observations within a cluster, and techniques have been developed by survey analysts to adjust for such correlations in computing estimated standard errors (see for example Moser and Kalton, 1971). Where the clustering is inherent in the population, however, as in the examples just given, it is of interest to study the characteristics of, and differences between, units at more than one level in the hierarchy. Indeed one may wish to be able to generalise from a sample to a population of such units, and to model the differences between units at different levels simultaneously. Multilevel models provide the means to do this. The chief methodological references are Goldstein (1986 and 1995), Bryk and Raudenbush (1992), and Longford (1993a).

Raudenbush and Willms (1991) edited a collection of papers describing applications of multilevel models in education. It is instructive to note that only one of these papers (that of Longford, 1991) mentioned measurement error in explanatory variables despite the fact that the data being analysed were subject to it. (Another chapter, that of Rowan *et al.*, 1991, used variances estimated at different levels of a multilevel model in a *definition* of measurement error and reliability – see below.) One reason for this neglect is that the software chiefly used in the analyses – *HLM* (Bryk *et al.*, 1988) and *ML3* (Prosser *et al.*, 1991) – lacked procedures for adjusting for measurement error. The *Junior School Project* (described in Mortimore *et al.*, 1988) used the procedures developed by Fuller and his colleagues (Hidiroglou *et al.*, 1980) in the adjustment of single-level estimates, but the multilevel estimates were unadjusted, again because procedures for their adjustment were not available.

Bryk and Raudenbush (1992) mentioned measurement error at two points. The first mention was in connection with a measurement model for school climate (also described by Raudenbush *et al.*, 1991, and Rowan *et al.*, 1991). In this study questionnaires were administered to 1867 teachers in 110 schools. Each of 35 Likert-scaled items was considered to measure one of five latent constructs, principal leadership, staff cooperation, teacher control, teacher efficacy, and teacher satisfaction. They defined a 3-level model, with items at level 1, teachers at level 2, and schools at level 3. Five dummy variables, one for each construct, and each with a coefficient random at all 3 levels, were used to predict the item scores, rescaled so that the level-1 variance of each of the five coefficients was equal. Interpreting this level-1 variance as error variance in the item scores they used the model to investigate other psychometric properties of the measures.

Bryk and Raudenbush (1992) made one other mention of measurement error, in the Appendix to Chapter 9 (p225), but they did not pursue it.

Goldstein (1986), in addition to setting out the main lines for the analysis of multilevel models by the method of iterative generalised least squares (iterative GLS, or IGLS), also outlined a method of adjustment for measurement error. This outline has been developed and extended by Goldstein and co-workers within the *Multilevel Models Project* at the Institute of Education, University of London, and applied to the illustrative analysis of some data from the *Junior School Project* (see Woodhouse *et al.*, 1996b). A summary of the method has been given in Goldstein (1995, Appendix

10.1). The method does not work for random-coefficients models with error in a variable with a random coefficient, that is, where both predictors and their coefficients are random.

Muthén and Satorra (1989) commented on the paucity of research into such models. Their approach, through structural models with random coefficients, is one way of characterising the measurement error problem in multilevel analysis. They considered a standard two-level random-coefficient model, which we may express in the notation of Goldstein (1995, Chapter 2) as:

$$\begin{aligned} y_{ij} &= \alpha_j + \beta_j x_{ij} + e_{ij}, \\ \alpha_j &= \alpha + u_{0j}, \\ \beta_j &= \beta + u_{1j}, \end{aligned} \tag{2.9}$$

where $j = 1, 2, \ldots, J$ indexes groups (at level 2) and within each group $i = 1, 2, \ldots, n_j$ indexes individuals (at level 1). The y and x variables are observed, the stochastic variables u_{0j}, u_{1j} are, in Muthén and Satorra's terms, latent variable influences of group *j*, possibly covarying, and the e_{ij} are random disturbances, independent of these. They then considered the generalisation of (2.9) to a model with a latent explanatory variable ξ_{ij}, say, and listed four ways in which group membership might induce variation in the model parameters:

1. random group variation of the structural regression of y_{ij} on ξ_{ij},
2. random group variation of the structural mean parameter $\mathrm{E}(\xi)$,
3. random group variation of the structural variance parameter $\mathrm{var}(\xi)$,

4. random group variation in the measurement parameters describing the regressions of the x variables on ξ.

They did not discuss how these four might combine in a single model.

Longford (1993a, Chapter 7, and 1993b) developed these ideas further. He described a two-level model,

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_j + \varepsilon_{ij}, \tag{2.10}$$

in which the design matrix $\mathbf{x}$ of explanatory variables was partitioned into two submatrices, one ($\mathbf{x}_f$) of variables directly observed without error, the other ($\mathbf{x}_r$) of latent variables for which vectors $\mathbf{s}_{ij}$ of manifest variables were observed. These manifest variables were related to the latent variables by a linear regression formula, generalised to a two-level model, as in the following scheme:

$$\begin{aligned} y_{ij} &= \mathbf{x}_{f,ij}\boldsymbol{\beta}_{f,j} + \mathbf{x}_{r,ij}\boldsymbol{\beta}_r + \varepsilon_{ij}, \\ \boldsymbol{\beta}_{f,j} &\sim \mathrm{N}(\boldsymbol{\beta}_f, \Sigma_2), \\ \varepsilon_{ij} &\sim \mathrm{N}(0, \sigma^2), \\ \mathbf{s}_{ij} &= \Lambda\mathbf{x}_{r,ij}^{\mathrm{T}} + \boldsymbol{\xi}_j^{(2)} + \boldsymbol{\xi}_{ij}^{(1)}, \\ \boldsymbol{\xi}_j^{(2)} &\sim \mathrm{N}(\mathbf{0}, \Theta_2), \\ \boldsymbol{\xi}_{ij}^{(1)} &\sim \mathrm{N}(\mathbf{0}, \Theta_1), \\ \mathbf{x}_{r,ij} &\sim \boldsymbol{\mu}_r + \boldsymbol{\zeta}_j^{(2)} + \boldsymbol{\zeta}_{ij}^{(1)}, \\ \boldsymbol{\zeta}_j^{(2)} &\sim \mathrm{N}(\mathbf{0}, \Psi_2), \\ \boldsymbol{\zeta}_{ij}^{(1)} &\sim \mathrm{N}(\mathbf{0}, \Psi_1), \end{aligned} \tag{2.11}$$

where $\boldsymbol{\beta}_{f,j}, \varepsilon_{ij}, \boldsymbol{\xi}_j^{(2)}, \boldsymbol{\xi}_{ij}^{(1)}, \boldsymbol{\zeta}_j^{(2)}, \boldsymbol{\zeta}_{ij}^{(1)}$ are mutually independent and Λ is a constant matrix. This very general formulation can be adapted to a variety of estimation problems. For example, with no outcomes y_{ij} and unknown Λ we have two-level factor analysis, as used by Longford and Muthén (1992) in an analysis of mathematics achievement data on U.S. eighth-grade students from the Second International Mathematics Study (Crosswhite *et al.*, 1985). See also Muthén (1997). Known Λ corresponds to an assumption of varying group means for the $\mathbf{s}_{ij}$.

Longford (1993b), in an illustrative analysis of simulated data, considered the case of two exchangeable indicators s_1, s_2 of a single latent explanatory variable x in the model

$$y_{ij} = \alpha_j + \beta x_{ij} + \varepsilon_{ij}. \tag{2.12}$$

In the notation of (2.11), $\Lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and each of $\beta_f, \beta_r, \sigma^2, \Sigma_2, \Theta_2, \Theta_1, \Psi_2, \Psi_1$ is a single parameter to be estimated. The within- and between-group covariances of the manifest variables s_1, s_2 are not identified and are assumed to be zero. Note that, general as the formulation (2.11) is, it does not permit between-group variation in the coefficients β_r of the latent variables. Longford points out (1993b, p310) that to do so implies that the outcomes are not Normally distributed. Nor does (2.11) allow for errors of measurement in the response variable. Thus, the focus is on between-group variation in the latent variables and in the measurement parameters (cases 2 and 4 in Muthén and Satorra's list, above).

In the present thesis we begin the generalisation to the multilevel case of the work of Fuller that we have cited. We extend the class of multilevel models that can be estimated, within the IGLS paradigm, given prior knowledge of the measurement error variances and covariances. In particular, the procedure we develop allows the estimation of random parameters of variables subject to error. Multivariate multilevel models, with measurement error in the responses, also can be estimated. The error variances and covariances themselves may have a multilevel structure, and we develop a straightforward method for specifying this. We also derive 'adjusted sandwich' estimators of the conditional covariance matrices of the estimators for the fixed and random parameters, to reflect the adjustments for measurement error that are made in the formation of the estimators, and we give further corrections for sampling error in the fixed and random parameters.

3 Model specification and estimation

3.1 Scope

In this chapter we show how the classical model for random additive errors of measurement may be developed within a multilevel framework to model more complicated error structures. We go on to develop a method of adjustment to reduce the bias in parameter estimators for multilevel models in the presence of such error structures. These models may include multivariate response, with errors in different responses possibly covarying amongst themselves, and random coefficients of explanatory variables measured with error. A straightforward extension, described briefly below in Section 3.4.4 but not developed in detail, allows the inclusion of randomly cross-classified units. We do not consider errors in discrete variables.

The method can be adapted to the estimation of models containing imputed values for randomly missing data. Goldstein and Woodhouse (1998) describe a method for estimating variance components models with randomly missing data. That method, however, is not applicable to the random coefficients case.

We first define some notation and then review the basic multilevel model with measurement error.

3.2 Notation and terminology

We shall make no distinction in notation between a variable and a value of it. We denote observed values (vectors or matrices) by capital letters **X**, **Y**, **Z**, and true values (or estimates of them) by corresponding lower-case letters $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. Errors of measurement contained in the observed values are denoted by corresponding Greek letters ξ, η, ζ. An initial subscript *r*, *s*, etc. (or a numeral), to a data matrix denotes a column number within that matrix, that is, a particular variable. Thus, for example, $\mathbf{X}_2$ represents the column with index 2 in the matrix **X**, typically an explanatory variable. The subscripts i,i', if present, refer always to units at level 1. Subscripts j,j', k,k' refer to units at higher levels, not necessarily levels 2 and 3.

A variable is said to be *defined at level* ℓ and referred to as a *level-*ℓ *variable* if its values are separately measured or computed for each level-ℓ unit and may differ between any two of these. If $\ell > 1$, each level-1 unit within a given level-ℓ unit is assigned the value of the variable corresponding to that level-ℓ unit. An error of measurement of a variable is said to be *defined at level* ℓ if it is identically the same for all level-1 units for which a measurement exists and that are within a given level-ℓ unit, but not identically the same for level-1 units in different level-ℓ units. Thus, a measurement on a child at level 1 produces a value of a level-1 variable, but this may include separate errors defined at different levels, for example if the same measurement instrument or tester is assigned to a group of children.

A standard convention for indicating a particular instance of a variable is to use a sequence of subscripts, starting at the level of definition of the variable and ending at the highest level in the model, each subscript corresponding to a subunit within the next unit in the hierarchy. Thus the value of the level-1 variable $\mathbf{X}_2$ corresponding to level-1 unit *i* within level-2 unit *j* within level-3 unit *k* would be represented by X_{2ijk}.

This convention can be cumbersome, particularly when it is necessary to refer to an element in a cross-product matrix. Therefore we shall generally omit all unit indices after the first and denote the above-mentioned value, for example, as $X_{2,i}$. The row of $\mathbf{X}$ corresponding to the i th level-1 unit will be denoted $\mathbf{X}_i$. There will also be cases where a symbol subscripted i represents a column vector or a matrix, for example a vector of random variables for a particular unit. The subscripts j, k, etc., also may indicate rows, columns or matrices, though this will always be clear from the context. Indices at a given level take all integer values from 1 to the total number of units at that level in the data set. We shall assume that the total number of level-1 units is N, and thus $1 \le i \le N$. If j indexes units at level ℓ and k indexes units at level ℓ' we shall use J_ℓ, $K_{\ell'}$ for the total numbers of units at those levels. Throughout the remainder of this work the term *error*, when unqualified, refers to an error of measurement in a variable.

3.3 The basic model

Our purpose is to estimate the parameters of a multilevel model based on the true values of the response and explanatory variables. We may write such a model as:

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \ldots, N, \tag{3.1}$$

where y_i is the i th response, $\mathbf{x}_i$ the i th row of the design matrix $\mathbf{x}$ of explanatory variables in the *fixed part*, $\boldsymbol{\beta}$ the vector of *fixed parameters*, to be estimated, $\mathbf{z}_i$ the i th row of the design matrix $\mathbf{z}$ of explanatory variables in the *random part*, and $\boldsymbol{\varepsilon}_i$ a vector of random variables whose expectations are zero and whose variances and covariances are the *random parameters*, to be estimated subject to standard assumptions about the structure of the covariance matrix $\mathbf{v}$ of the $\mathbf{z}_i \boldsymbol{\varepsilon}_i$. If there are p fixed parameters and the vector $\boldsymbol{\varepsilon}_i$ has q elements we write

$$\mathbf{y} = (y_1 \;\; y_2 \;\; \dots \;\; y_N)^{\mathrm{T}}, \tag{3.2}$$

$$\mathbf{x} = \{\mathbf{x}_i\}_{(N \times p)}, \;\; \mathbf{x}_i = (x_{0,i} \;\; x_{1,i} \;\; \dots \;\; x_{p-1,i}), \;\; \mathbf{x}_r = (x_{r,1} \;\; x_{r,2} \;\; \dots \;\; x_{r,N})^{\mathrm{T}}, \; 0 \le r \le p-1,$$

$$\mathbf{z} = \{\mathbf{z}_i\}_{(N \times q)}, \;\; \mathbf{z}_i = (z_{0,i} \;\; z_{1,i} \;\; \dots \;\; z_{q-1,i}), \;\; \mathbf{z}_s = (z_{s,1} \;\; z_{s,2} \;\; \dots \;\; z_{s,N})^{\mathrm{T}}, \; 0 \le s \le q-1,$$

$$\boldsymbol{\beta} = (\beta_0 \;\; \beta_1 \;\; \dots \;\; \beta_{p-1})^{\mathrm{T}},$$

$$\boldsymbol{\varepsilon}_i = (\varepsilon_{0,i} \;\; \varepsilon_{1,i} \;\; \dots \;\; \varepsilon_{q-1,i})^{\mathrm{T}},$$

with $\mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}$. The variables represented in $\mathbf{z}$ may or may not include variables represented in $\mathbf{x}$. We assume for the moment that $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are defined at level 1.

The observed values **X**, **Y**, **Z** of the variables are assumed to be related to the true values by the measurement model:

$$\mathbf{X}_r = \mathbf{x}_r + \boldsymbol{\xi}_r, \;\; \boldsymbol{\xi}_r = (\xi_{r,1} \;\; \xi_{r,2} \;\; \dots \;\; \xi_{r,N})^{\mathrm{T}}, \tag{3.3}$$

$$\mathbf{Y} = \mathbf{y} + \boldsymbol{\eta}, \;\; \boldsymbol{\eta} = (\eta_1 \;\; \eta_2 \;\; \dots \;\; \eta_N)^{\mathrm{T}},$$

$$\mathbf{Z}_s = \mathbf{z}_s + \boldsymbol{\zeta}_s, \;\; \boldsymbol{\zeta}_s = (\zeta_{s,1} \;\; \zeta_{s,2} \;\; \dots \;\; \zeta_{s,N})^{\mathrm{T}},$$

$$\mathrm{E}(\boldsymbol{\xi}_r) = \mathbf{0}, \;\; \mathrm{E}(\boldsymbol{\eta}) = \mathbf{0}, \;\; \mathrm{E}(\boldsymbol{\zeta}_s) = \mathbf{0},$$

where for $r = 0,1,\dots,p-1$, $s = 0,1,\dots,q-1$, $\boldsymbol{\xi}_r$, $\boldsymbol{\eta}$, $\boldsymbol{\zeta}_s$ are unknown random *errors* in the observed variables $\mathbf{X}_r, \mathbf{Y}, \mathbf{Z}_s$, respectively. We assume for the moment that the errors are defined at level 1 and we assume we have prior values of the *error variances and covariances* $\mathrm{E}(\xi_{r,i}\xi_{r',i})$, $\mathrm{E}(\xi_{r,i}\zeta_{s,i})$, $\mathrm{E}(\zeta_{s,i}\zeta_{s',i})$, $\mathrm{E}(\xi_{r,i}\eta_i)$, $\mathrm{E}(\zeta_{s,i}\eta_i)$ for each unit i.

We maintain the following assumption throughout:

Assumption 3.1 Errors defined at a given level do not covary between units at (or above) that level.

The standard assumptions about $\mathbf{v}$ are most easily explained in the context of a 2-level model, and one such explanation is given by Goldstein (1995, Appendix 2.1). We give an adapted version here for completeness, and give an alternative statement in Section 3.5 when we describe the full model to be estimated. The reader familiar with the basic ideas may safely skip to Section 3.4.

We assume that the responses have been sorted to reflect the multilevel hierarchy and denote by B_j the set of unit identifiers for the level-1 units in level-2 unit j, where j may take any integer value between 1 and J_2, the total number of level-2 units. Suppose that $i \in B_j$. Then for a 2-level model we may rewrite equation (3.1) as:

$$y_i = \sum_{r=0}^{p-1} x_{ri}\beta_r + \sum_{r=0}^{q_1-1} z_{ri}^{(1)}\varepsilon_{ri}^{(1)} + \sum_{r=0}^{q_2-1} z_{ri}^{(2)}\varepsilon_{rj}^{(2)}, \tag{3.4}$$

$$\boldsymbol{\varepsilon}_i^{(1)} = \left(\varepsilon_{0,i}^{(1)}\ \varepsilon_{1,i}^{(1)}\ \ldots\ \varepsilon_{q_1-1,i}^{(1)}\right)^{\mathrm{T}},\quad \boldsymbol{\varepsilon}_j^{(2)} = \left(\varepsilon_{0,j}^{(2)}\ \varepsilon_{1,j}^{(2)}\ \ldots\ \varepsilon_{q_2-1,j}^{(2)}\right)^{\mathrm{T}},$$

$$\mathbf{z}_i^{(1)} = \left(z_{0,i}^{(1)}\ z_{1,i}^{(1)}\ \ldots\ z_{q_1-1,i}^{(1)}\right),\quad \mathbf{z}_i^{(2)} = \left(z_{0,i}^{(2)}\ z_{1,i}^{(2)}\ \ldots\ z_{q_2-1,i}^{(2)}\right),$$

$$\mathbf{z}_i\boldsymbol{\varepsilon}_i = \mathbf{z}_i^{(1)}\boldsymbol{\varepsilon}_i^{(1)} + \mathbf{z}_i^{(2)}\boldsymbol{\varepsilon}_j^{(2)}.$$

Thus the design matrix $\mathbf{z}$ of explanatory variables in the random part of the model (3.1)–(3.3) has been replaced by two design matrices, $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$. The variables represented in $\mathbf{z}^{(1)}$ are said to be *random at level 1*; variables in $\mathbf{z}^{(2)}$ are said to be *random at level 2*, and may include variables in $\mathbf{z}^{(1)}$.

The level-1 random variable vectors $\boldsymbol{\varepsilon}_i^{(1)}$ are assumed not to covary with explanatory variables, either between or within level-1 units. The random variables $\varepsilon_{ri}^{(1)}$ $(r = 0, 1, \ldots, q_1 - 1)$ may covary within level-1 units and these covariances, together with the variances of the $\varepsilon_{ri}^{(1)}$, constitute the random parameters at level 1. Equivalent statements at level 2 apply to the level-2 random variable vectors $\boldsymbol{\varepsilon}_j^{(2)}$. Also, within a given level-2 unit j, $\boldsymbol{\varepsilon}_j^{(2)}$ is replicated across all the level-1 units. This induces covariances between the elements of $\boldsymbol{\varepsilon}_j^{(2)}$ across the level-1 units within level-2 unit j. None of the random variable vectors $\boldsymbol{\varepsilon}_i^{(1)}$, $\boldsymbol{\varepsilon}_j^{(2)}$ covaries with any measurement error.

We define, for $i \in B_j$, $i' \in B_{j'}$, $j, j' \in \{1, 2, \ldots, J_2\}$,

$$
\begin{aligned}
&e_i^{(1)} \equiv \mathbf{z}_i^{(1)}\boldsymbol{\varepsilon}_i^{(1)}, \quad e_{i'}^{(1)} \equiv \mathbf{z}_{i'}^{(1)}\boldsymbol{\varepsilon}_{i'}^{(1)}, \quad e_i^{(2)} \equiv \mathbf{z}_i^{(2)}\boldsymbol{\varepsilon}_j^{(2)}, \quad e_{i'}^{(2)} \equiv \mathbf{z}_{i'}^{(2)}\boldsymbol{\varepsilon}_{j'}^{(2)}, \\
&\mathrm{E}\left(e_i^{(1)}\right) = \mathrm{E}\left(e_{i'}^{(1)}\right) = \mathrm{E}\left(e_i^{(2)}\right) = \mathrm{E}\left(e_{i'}^{(2)}\right) = 0, \\
&v_{ii'}^{(1)} \equiv \mathrm{E}\left(e_i^{(1)}e_{i'}^{(1)}\right), \quad \mathbf{v}^{(1)} \equiv \left\{v_{ii'}^{(1)}\right\}, \\
&v_{ii'}^{(2)} \equiv \mathrm{E}\left(e_i^{(2)}e_{i'}^{(2)}\right), \quad \mathbf{v}^{(2)} \equiv \left\{v_{ii'}^{(2)}\right\}, \\
&\mathbf{v} \equiv \mathbf{v}^{(1)} + \mathbf{v}^{(2)},
\end{aligned}
\tag{3.5}
$$

and we assume $\mathrm{E}\left(e_i^{(1)}e_{i'}^{(2)}\right) = 0$ for all i, i'. The $e_i^{(1)}, e_{i'}^{(1)}$ are assumed to be uncorrelated across level-1 units, so $\mathbf{v}^{(1)}$ is diagonal with (i,i)th element

$$
\mathbf{z}_i^{(1)}\boldsymbol{\Omega}_{\varepsilon}^{(1)}\mathbf{z}_i^{(1)\mathrm{T}}, \quad \boldsymbol{\Omega}_{\varepsilon}^{(1)} = \mathrm{cov}\left(\boldsymbol{\varepsilon}_i^{(1)}\right),
\tag{3.6}
$$

and we assume that $\boldsymbol{\Omega}_{\varepsilon}^{(1)}$ is constant for all level-1 units in the data set. The $e_i^{(2)}, e_{i'}^{(2)}$ are assumed to be uncorrelated across level-2 units, so $\mathbf{v}^{(2)}$ is block-diagonal with jth block

$$\mathbf{z}_j^{(2)}\boldsymbol{\Omega}_\varepsilon^{(2)}\mathbf{z}_j^{(2)\mathrm{T}}, \quad \boldsymbol{\Omega}_\varepsilon^{(2)} = \mathrm{cov}\left(\boldsymbol{\varepsilon}_j^{(2)}\right), \quad \mathbf{z}_j^{(2)} = \left\{\mathbf{z}_i^{(2)}\right\}_{(n_j \times q_2)}, \tag{3.7}$$

where n_j is the number of level-1 units in the *j*th level-2 unit, and $\boldsymbol{\Omega}_\varepsilon^{(2)}$ is assumed constant for all level-2 units. Thus the *j*th block of $\mathbf{v}$ can be written

$$\mathbf{v}_j = \bigoplus_{i \in B_j}\left(\mathbf{z}_i^{(1)}\boldsymbol{\Omega}_\varepsilon^{(1)}\mathbf{z}_i^{(1)\mathrm{T}}\right) + \mathbf{z}_j^{(2)}\boldsymbol{\Omega}_\varepsilon^{(2)}\mathbf{z}_j^{(2)\mathrm{T}}. \tag{3.8}$$

The extension to higher levels is straightforward.

3.4 Extensions to the basic model

We now show how the basic model just described can be extended to apply to data with more complex structure. In this section, for ease of interpretation, we depart from the convention of using $\mathbf{x}$ for variables in the fixed part and $\mathbf{z}$ for variables in the random part, and use the same symbol for a given variable wherever in the model it appears. We use an educational example to describe the principles.

3.4.1 Errors at higher levels

The assumption in the previous section that the variables and their errors are defined at level 1 is over-restrictive. It is important to be able to include in models variables defined at higher levels, and any errors in such variables will be defined at or above the level of the variables.

Consider the following 2-level variance-components model:

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \beta_2 x_{2,j} + \varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(2)} x_{0,i}, \tag{3.9}$$
$$\mathrm{var}\left(\varepsilon_{0,i}^{(1)}\right) = \sigma_{00}^{(1)}, \quad \mathrm{var}\left(\varepsilon_{0,j}^{(2)}\right) = \sigma_{00}^{(2)},$$

where the level-1 units are pupils and the level-2 units are schools. The *i*th pupil is in the *j*th school, which we denote by the statement $i \in B_j$, and for that pupil $x_{0,i} \equiv 1$ and y_i, $x_{1,i}$, $x_{2,j}$ are the unknown true values of Y_i, $X_{1,i}$, $X_{2,j}$, respectively. Suppose

Y_i is the pupil's observed score in mathematics at age 10 years,

$X_{1,i}$ is the pupil's observed score in a test of general aptitude taken at age 8, and

$X_{2,j}$ is the mean of the $X_{1,i}$ for the pupils in school *j*, as estimated from the sample.

The $X_{2,j}$ are defined at level 2 and have error at that level arising from the errors in the $X_{1,i}$ from which they are calculated and from sampling error. Thus, if the sampled pupils from the *j*th school are n_j in number and chosen at random without replacement from a known cohort of size N_j, we have

$$X_{2,j} = \frac{1}{n_j} \sum_{i \in B_j} X_{1,i}, \tag{3.10}$$

and the error $\xi_{2,j}$ in $X_{2,j}$ is given by

$$\begin{aligned} \xi_{2,j} &= X_{2,j} - x_{2,j} = \left(X_{2,j} - \bar{x}_{1 \cdot j}\right) + \left(\bar{x}_{1 \cdot j} - x_{2,j}\right) \\ &= \frac{1}{n_j} \sum_{i \in B_j} \xi_{1,i} + \bar{x}_{1 \cdot j} - x_{2,j}, \end{aligned} \tag{3.11}$$

where $\bar{x}_{1 \cdot j}$ is the mean of the true scores for the pupils sampled, and $\mathrm{E}\left(\bar{x}_{1 \cdot j}\right) = x_{2,j}$. We have

$$\text{var}(\xi_{2,j}) = \frac{1}{n_j}\text{var}(\xi_{1,i}) + \text{var}(\bar{x}_{1\cdot j}) \qquad (3.12)$$

$$= \frac{1}{n_j}\text{var}(\xi_{1,i}) + \frac{N_j - n_j}{n_j(N_j - 1)}\sigma^2_{1,j},$$

where $\sigma^2_{1,j}$ is the variance of the true scores x_1 within the cohort in school *j*.

In practice all variables are specified as $N \times 1$ vectors and we shall often use a level-1 identifier to specify a particular instance of a higher-level variable, for example $X_{2,i}$. With this convention, for two pupils in a given school *j* the elements of $\mathbf{X}_2$ are equal and so are the elements of ξ_2. For each $i,i' \in B_j$ we have $\text{cov}(\xi_{2,i}, \xi_{2,i'}) = \text{var}(\xi_{2,j})$, and we assume that prior values of these variances for each level-1 unit are contained in an $N \times 1$ *error variance vector*, denoted by $\mathbf{C}^{(2)}_{22}$, where the superscript indicates the level and the subscripts indicate the variables (the same variable in this case).

Secondly, the level-1 error $\xi_{1,i}$ for a pupil in school *j* correlates with the level-2 error $\xi_{2,i}$. If, in addition to assumption 3.1 (p34), we assume that errors do not covary with true values then for each $i,i' \in B_j$ we have

$$\text{cov}(\xi_{1,i}, \xi_{2,i'}) = \text{cov}(\xi_{1,i}, \xi_{2,i}) \qquad (3.13)$$

$$= \text{cov}\left(\xi_{1,i}, \frac{1}{n_j}\sum_{i \in B_j}(x_{1,i} + \xi_{1,i}) - x_{2,i}\right)$$

$$= \frac{1}{n_j}\text{var}(\xi_{1,i}).$$

We assume that prior values of these covariances are contained in an *error covariance vector*, $\mathbf{C}_{12}^{(2)}$, where by convention if there is a difference in level between the errors to be included in the covariance the first subscript denotes the variable whose included errors are at the lower level. We do not require that $\text{cov}(\xi_{1ij}, \xi_{2,j})$ should be constant for all units *i* within a given level-2 unit *j*, although in many cases there will be no other information, but if these covariances differ, that is $C_{12,i}^{(2)} \neq C_{12,i'}^{(2)}$ for some $i, i' \in B_j$, we shall have $\text{cov}(\xi_{1ij}, \xi_{2,j'}) = C_{12,i}^{(2)}$, not $C_{12,i'}^{(2)}$, because it is ξ_2 that is defined at level 2.

3.4.2 Multiple sources of error

It is clearly possible for errors in a given variable to arise from more than one source. In the case just described, of a school-average score estimated from a sample of observed individual scores, it is not necessary for the purpose of model estimation to distinguish between the two sources (error in the level-1 measure and sampling error), but it may become necessary if errors in a given variable arise in distinct ways from different levels of a hierarchy. As in the previous case, the point here is to ensure that all appropriate error covariances within and across units (more properly, expected error products, which may depend on the association of variables with units) are correctly adjusted for in the estimation.

In model (3.9), for example, random errors in $\mathbf{X}_1$ may arise from factors that depend on the pupil and also from factors common to the pupils in a given school. This structure of errors may be specified by providing prior values of the error variances in $\mathbf{X}_1$ at the two levels, for each pupil, in separate vectors $\mathbf{C}_{11}^{(1)}$ and $\mathbf{C}_{11}^{(2)}$. By definition there is no covariance, either within or between pupils, between errors in $\mathbf{X}_1$ at level 1 and errors in $\mathbf{X}_1$ at level 2. The errors in $\mathbf{X}_1$ at level 2, however, will covary between pupils in the same school, and the elements of $\mathbf{C}_{11}^{(2)}$, which will be the same for pupils in a given school, are these covariances.

3.4.3 Multivariate multilevel models

Consider now a bivariate model for scores in mathematics and reading:

$$y_i = \beta_{01}z_{1,i}x_{0,j} + \beta_{02}z_{2,i}x_{0,j} + \beta_{11}z_{1,i}x_{1,j} + \beta_{12}z_{2,i}x_{1,j} + \varepsilon_{01,j}^{(2)}z_{1,i}x_{0,j} + \varepsilon_{02,j}^{(2)}z_{2,i}x_{0,j},$$
$$\text{var}\left(\varepsilon_{01,j}^{(2)}\right) = \sigma_{11}^{(2)},\ \text{var}\left(\varepsilon_{02,j}^{(2)}\right) = \sigma_{22}^{(2)},\ \text{cov}\left(\varepsilon_{01,j}^{(2)}, \varepsilon_{02,j}^{(2)}\right) = \sigma_{12}^{(2)}. \tag{3.14}$$

Here level 1 exists in order to define the bivariate structure of the response, level-2 units are pupils, and we ignore for the moment the school level. For each pupil j there exist at most 2 units, i, i': an odd-numbered unit $(i = 2j-1)$ corresponding to a mathematics score Y_i at age 10 years and an even-numbered unit $(i' = 2j)$ corresponding to a reading score $Y_{i'}$ at the same age. For each j, $x_{0,j} \equiv 1$ and $x_{1,j}$ is the unknown true value of $X_{1,j}$, which as before is the pupil's score on a test of general aptitude taken at age 8. The $z_{1,i}$ take the value 1 if i is odd and 0 if i is even, and $z_{2,i} = 1 - z_{1,i}$. Thus the model (3.14) permits the estimation of an intercept and a fixed effect of general aptitude at age 8 for each of mathematics and reading at age 10. The level-2 variances and covariance (assumed constant) are the residual between-pupil variances and covariance of these two scores.

The analysis of model (3.14) requires the use of two explanatory variables containing error: $U_{3,i} \equiv z_{1,i}X_{1,j}$ and $U_{4,i} \equiv z_{2,i}X_{1,j}$. Table 3.1 illustrates how the errors $\omega_{3,i}$, $\omega_{4,i}$ in these two variables are related, for three typical pupils.

Table 3.1 Error structure for explanatory variables in model (3.14)

Pupil (j)	Response(i)	$\mathbf{U}_3$	$\mathbf{U}_4$	ω_3	ω_4
1	1: maths	31	0	$\omega_{3,1}$	0
1	2: reading	0	31	0	$\omega_{4,2} = \omega_{3,1}$
2	4: reading	0	27	0	$\omega_{4,4}$
3	5: maths	29	0	$\omega_{3,5}$	0
3	6: reading	0	29	0	$\omega_{4,6} = \omega_{3,5}$

Pupil 2 is missing a maths score at age 10, so that record does not appear. $\mathbf{X}_1$ is defined at level 2 (pupil level) and its errors ξ_{1j} are the same for both level-1 units (response measurements) within the same level-2 unit (pupil). Because $\mathbf{U}_3 = \text{diag}(\mathbf{z}_1)\mathbf{X}_1$ and $\mathbf{U}_4 = \text{diag}(\mathbf{z}_2)\mathbf{X}_1$, errors in $\mathbf{U}_3$, $\mathbf{U}_4$ for a given pupil are the same where present and errors corresponding to zero elements in $\mathbf{z}_1$, $\mathbf{z}_2$, respectively, are zero. Assuming that we have a vector $\mathbf{C}_{11}^{(2)}$ of prior values of the error variances of $\mathbf{X}_1$ for each pupil (replicated over the level-1 units i, i' within each pupil) we may specify level-2 error variances for $\mathbf{U}_3$, $\mathbf{U}_4$ by the vectors

$$\mathbf{C}_{33}^{(2)} = \text{diag}(\mathbf{z}_1)\mathbf{C}_{11}^{(2)}, \quad \mathbf{C}_{44}^{(2)} = \text{diag}(\mathbf{z}_2)\mathbf{C}_{11}^{(2)}, \tag{3.15}$$

respectively.

We may also specify a vector

$$\mathbf{C}_{34}^{(2)} = \mathbf{C}_{11}^{(2)} \tag{3.16}$$

of level-2 error covariances between $\mathbf{U}_3$ and $\mathbf{U}_4$ for each unit. Thus, for example, $\text{E}(\omega_{3,1}\omega_{4,2}) = \text{var}(\xi_{1,1})$.

It will be important in processing these vectors not to set up spurious covariances, for example between $\omega_{3,1}$ and $\omega_{3,2}$ (=0), or between $\omega_{3,2}$ and $\omega_{4,1}$. We shall show in Section 3.6 how the use of *error incidence matrices* derived from the error variance vectors avoids this problem.

We may elaborate model (3.14) by adding a further level for schools, and model the effects on the response of the true school-average year-8 score $\mathbf{x}_2$ as follows:

$$\begin{aligned} y_i = {} & \beta_{01}z_{1,i}x_{0,j} + \beta_{02}z_{2,i}x_{0,j} + \beta_{11}z_{1,i}x_{1,j} + \beta_{12}z_{2,i}x_{1,j} + \beta_{21}z_{1,i}x_{2,k} + \beta_{22}z_{2,i}x_{2,k} \\ & + \varepsilon^{(2)}_{01,j}z_{1,i}x_{0,j} + \varepsilon^{(2)}_{02,j}z_{2,i}x_{0,j} + \varepsilon^{(3)}_{01,k}z_{1,i}x_{0,j} + \varepsilon^{(3)}_{02,k}z_{2,i}x_{0,j}, \end{aligned} \tag{3.17}$$

$$\begin{aligned} & \operatorname{var}\left(\varepsilon^{(2)}_{01,j}\right) = \sigma^{(2)}_{11}, \ \operatorname{var}\left(\varepsilon^{(2)}_{02,j}\right) = \sigma^{(2)}_{22}, \ \operatorname{cov}\left(\varepsilon^{(2)}_{01,j}, \varepsilon^{(2)}_{02,j}\right) = \sigma^{(2)}_{12}, \\ & \operatorname{var}\left(\varepsilon^{(3)}_{01,k}\right) = \sigma^{(3)}_{11}, \ \operatorname{var}\left(\varepsilon^{(3)}_{02,k}\right) = \sigma^{(3)}_{22}, \ \operatorname{cov}\left(\varepsilon^{(3)}_{01,k}, \varepsilon^{(3)}_{02,k}\right) = \sigma^{(3)}_{12}. \end{aligned}$$

We define $U_{5,i} \equiv z_{1,i}X_{2,k}$, $U_{6,i} \equiv z_{2,i}X_{2,k}$, where $X_{2,k}$ is the mean 8-year score for school k as estimated from the values $X_{1,j}$ for the pupils j sampled from school k. Where in model (3.9) we assumed the existence of a level-2 error variance vector for $\mathbf{X}_2$ we now assume a level-3 error variance vector $\mathbf{C}^{(3)}_{22}$ from which we may derive level-3 error variance vectors $\mathbf{C}^{(3)}_{55}, \mathbf{C}^{(3)}_{66}$ and covariance vector $\mathbf{C}^{(3)}_{56}$ for $\mathbf{U}_5$ and $\mathbf{U}_6$, containing the required prior values for each level-1 unit. Similarly, we may derive level-3 error covariance vectors $\mathbf{C}^{(3)}_{35}, \mathbf{C}^{(3)}_{45}, \mathbf{C}^{(3)}_{36}, \mathbf{C}^{(3)}_{46}$ from $\mathbf{C}^{(3)}_{12}$. The procedure to be described in Section 3.6 generates an expected error product at level 3 for a specific ordered pairing of units and variables if, and only if, the units are nested in the same level-3 unit, error variances exist for the variable and unit on each side of the pairing, and a covariance at level 3 has been specified.

3.4.4 Cross-classified data structure

Returning to the univariate case, if the pupils represented in model (3.9) are tested at age 8 by different examiners, this imposes further structure on the errors in $\mathbf{X}_1$. If no examiner examines in more than one school we have a cross-classification of pupils with examiners at level 1, nested within schools at level 2. If examiners are not nested within schools, as for example when a Local Education Authority or school board administers the test across schools, then we have pupils at level 1 nested within a cross-classification of examiners with schools at level 2. Rasbash and Goldstein (1994) showed how such multilevel random cross-classified structures can be specified and estimated using a purely hierarchical formulation. The method involves defining a new sequence of levels, one for each classification, together with a set of dummy variables for each classification after the first in each level of the original hierarchy. From the point of view of error specification the resulting structure is similar to that of th~ multivariate multilevel model. The additional requirement is for the estim procedure to be able to constrain the random parameters associated wit variables in the same set for a given classification to be equal. We do not ...his issue further in this thesis.

3.4.5 Errors in the response

Estimation of the model parameters must adjust also for error variances and covariances for the response, including covariances with errors in the explanatory variables. In the case of a univariate continuous response these variances and covariances can be specified as for explanatory variables. The response may be subject to error at more than one level, in which case there will be covariances between errors in different responses as well as with the errors in explanatory variables.

Multivariate and cross-classified responses may have additional error structure. For example, in model (3.17), mathematics scores and reading scores may each contain errors at both pupil and school levels. Within a given school the school-level errors for mathematics for all pupils will have a constant variance and an induced covariance across pupils. The same will be true for the school-level errors for reading, but typically the errors for reading will have different variances and covariances from the errors for mathematics, and they may or may not covary with them. Similar patterns of covarying errors may be induced by different examiners. Such error structures can be specified conveniently in a manner that we now illustrate.

In model (3.17) the response variable **y** has a sequence of pairs of values, the first element of each pair being a mathematics score and the second a reading score. (Some pupils may lack one of these.) Identifiers at level 1 distinguish mathematics from reading. We may write the observed values **Y** as the sum $\mathbf{Y}_1 + \mathbf{Y}_2$, where

$Y_{1,i}$ is the mathematics score of pupil $\frac{i+1}{2}$ if i is odd and zero if i is even,

$Y_{2,i}$ is the reading score of pupil $\frac{i}{2}$ if i is even and zero if i is odd.

Now $\mathbf{Y}_1$ has errors $\eta_1^{(2)}$, $\eta_1^{(3)}$ at pupil and school levels respectively, and $\mathbf{Y}_2$ has errors $\eta_2^{(2)}$, $\eta_2^{(3)}$. Their variances may be specified in each case as variances at level 2 or level 3, but existing only for level-1 units corresponding to the appropriate measurement. Prior values of covariances between these errors and with errors in explanatory variables can then be specified as vectors in the same way as for error covariances of explanatory variables. The extension to further response variates is straightforward, and specification for cross-classified structures is achieved similarly.

3.5 The model to be estimated

We consider an extension of the basic model described in section 3.3. We write, as before,

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \boldsymbol{\varepsilon}_i , \tag{3.18}$$

where y_i, $\mathbf{x}_i$, $\boldsymbol{\beta}$, $\mathbf{z}_i$, and $\boldsymbol{\varepsilon}_i$ have the meanings assigned to them in section 3.3. We assume that there are p fixed parameters, the vector $\boldsymbol{\varepsilon}_i$ has q elements, and there are L levels. The total response vector $\mathbf{y}$ is now the sum of m vectors, $m \geq 1$, and typically for a given value of i the ith elements of all but one of these vectors will be identically zero. The elements of $\mathbf{y}$ are sorted to reflect the multilevel hierarchy. We write:

$$\mathbf{y}=\sum_{u=1}^{m}\mathbf{y}_u\,,\quad \mathbf{y}_u=(y_{u,1}\ \ y_{u,2}\ \ \ldots\ \ y_{u,N})^{\mathrm{T}},\quad y_i=\sum_{u=1}^{m}y_{u,i}\,, \tag{3.19}$$

$$\mathbf{x}=\{\mathbf{x}_i\}_{(N\times p)},\ \mathbf{x}_i=(x_{0,i}\ \ x_{1,i}\ \ \ldots\ \ x_{p-1,i}),\quad \mathbf{x}_r=(x_{r,1}\ \ x_{r,2}\ \ \ldots\ \ x_{r,N})^{\mathrm{T}},$$

$$\mathbf{z}=\{\mathbf{z}_i\}_{(N\times q)},\ \mathbf{z}_i=(z_{p,i}\ \ z_{p+1,i}\ \ \ldots\ \ z_{p+q-1,i}),\ \mathbf{z}_s=(z_{s,1}\ \ z_{s,2}\ \ \ldots\ \ z_{s,N})^{\mathrm{T}},$$

$$\boldsymbol{\beta}=(\beta_0\ \ \beta_1\ \ldots\ \beta_{p-1})^{\mathrm{T}},$$

$$\boldsymbol{\varepsilon}_i=(\varepsilon_{p,i}\ \ \varepsilon_{p+1,i}\ \ \ldots\ \ \varepsilon_{p+q-1,i})^{\mathrm{T}},\ \varepsilon_{s,i}=\sum_{\ell=1}^{L}\varepsilon_{s,i}^{(\ell)},\ \boldsymbol{\varepsilon}_i^{(\ell)}=\left(\varepsilon_{p,i}^{(\ell)}\ \ \varepsilon_{p+1,i}^{(\ell)}\ \ \ldots\ \ \varepsilon_{p+q-1,i}^{(\ell)}\right)^{\mathrm{T}},$$

where $r\in\{0,1,\ldots,p-1\}$, $s\in\{p,p+1,\ldots,p+q-1\}$, and $\mathrm{E}\left(\boldsymbol{\varepsilon}_i^{(\ell)}\right)=\mathbf{0}$ for $\ell=1,2,\ldots,L$.

For a given ℓ, $1\le\ell\le L$, and a given j, $1\le j\le J_\ell$, we define $B_j^{(\ell)}$ to be the set of identifiers for the level-1 units nested in level-ℓ unit j. For a given $i\in B_j^{(\ell)}$ and a given $s\in\{p,p+1,\ldots,p+q-1\}$ we have by definition

$$\begin{aligned}&\varepsilon_{s,i'}^{(\ell)}\equiv\varepsilon_{s,i}^{(\ell)}\ \text{ for all } i'\in B_j^{(\ell)},\\ &\varepsilon_{s,i'}^{(\ell)}\not\equiv\varepsilon_{s,i}^{(\ell)}\ \text{ if } i'\notin B_j^{(\ell)}.\end{aligned} \tag{3.20}$$

As before, the variables represented in $\mathbf{z}$ may or may not include variables represented in $\mathbf{x}$. The variable indices for $\mathbf{x}$ run as usual from zero. The variable indices for $\mathbf{z}$ continue the sequence begun by the indices for $\mathbf{x}$. The variable indices u for $\mathbf{y}$ run from 1 and we shall often use the index $\bar{u}\equiv-u$. This indexing structure will enable us to define vectors and matrices serving similar purposes for different sets of variables without too much proliferation of symbols. Thus, negative indices correspond to y-variates, indices from 0 to $p-1$ correspond to $\mathbf{x}$-variates, and indices from p to

$p+q-1$ to z-variates. This convention implies that we may have $\mathbf{x}_r \equiv \mathbf{z}_s$ with $r \neq s$, but we shall see that in such cases no problems arise from treating $\mathbf{x}_r$ and $\mathbf{z}_s$ as distinct.

The fixed parameters to be estimated are the β_r, $r = 0,1,\ldots,p-1$. The random parameters to be estimated will be denoted by θ_h, $h = 1,2,\ldots,H$. To each h there correspond indices $r_h, s_h \in \{p, p+1,\ldots,p+q-1\}$ (where we may have $r_h = s_h$) and a level ℓ_h such that we may regard θ_h as a 'covariance' between random variables 'at level ℓ_h':

$$\theta_h \equiv \sigma_{r_h s_h}^{(\ell_h)} = \operatorname{cov}\left(\varepsilon_{r_h,i}^{(\ell_h)}, \varepsilon_{s_h,i'}^{(\ell_h)}\right),\ i,i' \in B_j^{(\ell_h)},\ \text{assumed constant for } j = 1,2,\ldots,J_{\ell_h}. \quad (3.21)$$

For $\ell = 1,2,\ldots,L$ and any $s,s' \in \{p, p+1,\ldots,p+q-1\}$, we define

$$\sigma_{ss'}^{(\ell)} \equiv \theta_h \ \text{ if } \{s,s'\} = \{r_h, s_h\} \text{ and } \ell = \ell_h \text{ for some } h,\ 1 \le h \le H, \quad (3.22)$$
$$0 \quad \text{otherwise.}$$

Further, if $i \in B_j^{(\ell)}$ and $i' \notin B_j^{(\ell)}$ then $\operatorname{cov}\left(\varepsilon_{s,i}^{(\ell)}, \varepsilon_{s',i}^{(\ell)}\right) = 0$ for all s,s' and if $\ell \neq \ell'$ then $\operatorname{cov}\left(\varepsilon_{s,i}^{(\ell)}, \varepsilon_{s',i'}^{(\ell')}\right) = 0$ for all i,i',s,s'.

The structure of the random part of the model can be restated as follows. Suppose that the fixed parameters are known. Then the true raw *residuals* $\tilde{\mathbf{y}}$ are defined by:

$$\tilde{\mathbf{y}} \equiv \mathbf{y} - \mathbf{x}\boldsymbol{\beta}, \quad (3.23)$$

and we have

$$
\begin{aligned}
\mathrm{E}(\tilde{\mathbf{y}}) &= \mathbf{0}, \\
\mathrm{E}\left(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}\right) &= \mathbf{v}, \\
\mathbf{v} &= \sum_{\ell=1}^{L} \mathbf{v}^{(\ell)}, \\
\mathbf{v}^{(\ell)} &= \bigoplus_{j=1}^{J_\ell} \mathbf{v}_j^{(\ell)}, \\
\mathbf{v}_j^{(\ell)} &= \sum_{s,s'=p}^{p+q-1} \sigma_{ss'}^{(\ell)} \mathbf{z}_{s,j} \mathbf{z}_{s',j}^{\mathrm{T}},
\end{aligned}
\tag{3.24}
$$

where $\sigma_{ss'}^{(\ell)}$ is as defined in (3.22) and $\mathbf{z}_{s,j}$ is the vector of values $z_{s,i}$ for $i \in B_j^{(\ell)}$. Thus we assume a block-diagonal structure for the residual covariance matrix $\mathbf{v}$ of the responses and the θ_h are the parameters of that structure, which are to be estimated.

The measurement error model linking the observed values **X**, **Y**, and **Z** to the true values of the variables is:

$$
\mathbf{X}_r = \mathbf{x}_r + \boldsymbol{\xi}_r,\ \boldsymbol{\xi}_r = (\xi_{r,1}\ \ \xi_{r,2}\ \ \ldots\ \ \xi_{r,N})^{\mathrm{T}},\ 0 \le r \le p-1,\ \ \boldsymbol{\xi} = \{\boldsymbol{\xi}_r\}_{(N\times p)}, \tag{3.25}
$$

$$
\mathbf{Y}_u = \mathbf{y}_u + \boldsymbol{\eta}_u,\ \boldsymbol{\eta}_u = (\eta_{u,1}\ \ \eta_{u,2}\ \ \ldots\ \ \eta_{u,N})^{\mathrm{T}},\ 1 \le u \le m,\ \boldsymbol{\eta} = \sum_{u=1}^{m} \boldsymbol{\eta}_u,\ \eta_i = \sum_{u=1}^{m} \eta_{u,i}
$$

$$
\mathbf{Z}_s = \mathbf{z}_s + \boldsymbol{\zeta}_s,\ \boldsymbol{\zeta}_s = (\zeta_{s,1}\ \ \zeta_{s,2}\ \ \ldots\ \ \zeta_{s,N})^{\mathrm{T}}, p \le s \le p+q-1,\ \ \boldsymbol{\zeta} = \{\boldsymbol{\zeta}_s\}_{(N\times q)},
$$

where each measurement error may have a component at each level, thus:

$$
\xi_{r,i} = \sum_{\ell=1}^{L} \xi_{r,i}^{(\ell)},\ \ \eta_{u,i} = \sum_{\ell=1}^{L} \eta_{u,i}^{(\ell)},\ \ \zeta_{s,i} = \sum_{\ell=1}^{L} \zeta_{s,i}^{(\ell)},\ \ 1 \le i \le N, \tag{3.26}
$$

each component having expectation zero.

By definition an error at level ℓ in the measurement of a given variable is the same for every level-1 unit in a given level-ℓ unit, provided a measurement on the variable exists for that level-1 unit. We assume that errors at level ℓ in a given variable do not covary between level-1 units in different level-ℓ units, nor with errors at different levels in the same variable within or between any level-1 units. They may, however, covary with errors at any level in a different variable. We assume finite second moments and define errors in terms of their variances and covariances within and across units at a specified level. We assume further that we have prior values of these in the form of error variance and covariance vectors as outlined in Section 3.4 and prescribed more formally in Section 3.6. We also make the following assumption:

Assumption 3.2 Errors at any level do not covary either with the true values of any of the variables or with any residuals.

The adjustment procedures developed in this chapter and tested and applied in the remainder of the thesis assume that the errors are samples from a multivariate Normal distribution (with mean vector zero). Thus, in particular, all moments of the error distribution are determined by the variances and covariances. It is in principle possible to specify other distributions, and to allow errors to covary with the true values of their variables; the methods of adjustment derived here can in principle be extended to apply to such cases. The assumptions of Normality and of independence from the true values are, however, a useful starting point. We shall see later that it is convenient also to assume multivariate Normality for the distribution of the random variables ε_i in (3.19), whose variances and covariances are estimated as the random parameters of the model. Therefore we state:

Assumption 3.3 The errors ξ, η, ζ are distributed multivariate Normally with expectation **0**.

Assumption 3.4 The random variables ε are distributed multivariate Normally with expectation **0**, and are independent of $\mathbf{x}$, $\mathbf{z}$, and the errors ξ, η, ζ.

3.6 Vectors and matrices associated with the measurement errors

We now define the *error variance and covariance vectors, error incidence matrices, error product matrices,* and *adjusted product matrices* that we shall use as the basis for the estimation of the model described in Section 3.5.

3.6.1 Error variance and covariance vectors

For each ℓ, r, s such that $1 \le \ell \le L$, $-m \le r \le p+q-1$, and $-m \le s \le p+q-1$, we assume the existence of an $N \times 1$ *error covariance vector* that is one of $\mathbf{C}_{rs}^{(\ell)}$, $\mathbf{C}_{sr}^{(\ell)}$. Not more than one such vector may be specified for a given choice of ℓ, r, s: suppose for the purpose of explanation that this is $\mathbf{C}_{rs}^{(\ell)}$. The ith element of $\mathbf{C}_{rs}^{(\ell)}$ is the *covariance at level* ℓ between the measurement errors ω_r, ω_s in the observed variables $\mathbf{U}_r, \mathbf{U}_s$ for level-1 unit i. If $r \le -1$ then $\mathbf{U}_r \equiv \mathbf{Y}_u$, $\omega_r \equiv \eta_u$ with $u = -r$; if $0 \le r \le p-1$ then $\mathbf{U}_r \equiv \mathbf{X}_r$, $\omega_r \equiv \xi_r$; if $r \ge p$ then $\mathbf{U}_r \equiv \mathbf{Z}_r$, $\omega_r \equiv \zeta_r$; with similar meanings for $\mathbf{U}_s, \omega_s$. If for a given choice of ℓ, r, s there is no prior specification of $\mathbf{C}_{rs}^{(\ell)}$ nor one of $\mathbf{C}_{sr}^{(\ell)}$ then it is assumed that $\mathbf{C}_{rs}^{(\ell)} = \mathbf{0}$.

If the variables denoted by r and s are distinct we allow the possibility that an error at level $\ell > 1$ in the second-named variable may covary with errors at level $\ell' \le \ell$ in the first, for example if the second variable is derived from the first by an aggregation process. Assume that this is the case and consider a level-1 unit i, in level-ℓ unit j, for which an error $\omega_{r,i}^{(\ell')}$ exists at some level $\ell' \le \ell$. Then we define the *error covariance at level ℓ for unit i* between ω_r and ω_s by setting

$$C_{rs,i}^{(\ell)} = \mathrm{E}\left(\omega_{r,i}^{(\cdot)}\omega_{s,i'}^{(\ell)}\right), \text{ where} \qquad (3.27)$$
$$\omega_{r,i}^{(\cdot)} \equiv \sum_{\ell'=1}^{\ell} \omega_{r,i}^{(\ell')}$$

and $i' \in B_j^{(\ell)}$ denotes a level-1 unit for which $\omega_{s,i'}^{(\ell)}$ exists. Because this latter error is at level ℓ it takes the same value for all $i' \in B_j^{(\ell)}$ for which it exists and we have, for the unit i that we are considering and for each $i' \in B_j^{(\ell)}$,

$$\mathrm{E}\left(\omega_{r,i}^{(\cdot)}\omega_{s,i'}^{(\ell)}\right) = \begin{cases} C_{rs,i}^{(\ell)} & \text{if } \omega_{s,i'}^{(\ell)} \text{ exists,} \\ 0 & \text{otherwise.} \end{cases} \qquad (3.28)$$

As the components of $\omega_{r,i}^{(\cdot)}$ may differ for different level-1 units within the same level-ℓ unit j, it is possible that for some $i,i' \in B_j^{(\ell)}$ we may have non-zero $C_{rs,i}^{(\ell)}$ and $C_{rs,i'}^{(\ell)}$ with $C_{rs,i}^{(\ell)} \ne C_{rs,i'}^{(\ell)}$.

If r and s denote the same variable the corresponding vector $\mathbf{C}_{rs}^{(\ell)} = \mathbf{C}_{ss}^{(\ell)} = \mathbf{C}_{rr}^{(\ell)}$ is more properly called an *error variance vector*. We do not permit errors in the same variable at different levels to covary, thus the element $C_{rr,i}^{(\ell)}$ is the variance of $\omega_{r,i}^{(\ell)}$ alone, if this exists, otherwise $C_{rr,i}^{(\ell)} = 0$. If $C_{rr,i}^{(\ell)}, C_{rr,i'}^{(\ell)}$ are both non-zero for some $i,i' \in B_j^{(\ell)}$ $(\ell > 1)$ then we must have $C_{rr,i}^{(\ell)} = C_{rr,i'}^{(\ell)}$.

3.6.2 Error incidence matrices

For each r we use the error variance vectors $\mathbf{C}_{rr}^{(\ell)}$ for $\ell = 1,2,\ldots,L$ to form the $N \times N$ *error incidence matrix* $\mathbf{K}_r \equiv \mathrm{diag}\{K_{r,i}\}$. We define

$$K_{r,i} = \begin{cases} 1 & \text{if } C_{rr,i}^{(\ell)} \neq 0 \text{ for some } \ell, \\ 0 & \text{otherwise.} \end{cases} \tag{3.29}$$

3.6.3 Error product matrices

For each ℓ, r, s we now define the *(expected) error product matrix at level* ℓ for the ordered pair of measurement errors $\boldsymbol{\omega}_r$, $\boldsymbol{\omega}_s$:

$$\mathbf{M}_{rs}^{(\ell)} = \begin{cases} \bigoplus_{j=1}^{J_\ell}\left(\mathbf{K}_{r,j}\mathbf{C}_{rs,j}^{(\ell)}\mathbf{1}_{n_j}^{\mathrm{T}}\mathbf{K}_{s,j}\right) & \text{if } \mathbf{C}_{rs}^{(\ell)} \text{ exists,} \\ \bigoplus_{j=1}^{J_\ell}\left(\mathbf{K}_{s,j}\mathbf{C}_{sr,j}^{(\ell)}\mathbf{1}_{n_j}^{\mathrm{T}}\mathbf{K}_{r,j}\right)^{\mathrm{T}} & \text{if } \mathbf{C}_{sr}^{(\ell)} \text{ exists,} \end{cases} \tag{3.30}$$

where n_j is the number of level-1 units in the *j*th level-ℓ unit, $\mathbf{1}_{n_j}$ is an $n_j \times 1$ vector of ones, and elsewhere the subscript j denotes the *j*th level-ℓ block of a vector or matrix. This definition ensures that $M_{rs,ii'}^{(\ell)} = 0$ unless three conditions are satisfied. First, there must exist j, $1 \leq j \leq J_\ell$, for which $i,i' \in B_j^{(\ell)}$, that is, level-1 units i and i' must be in the same unit at level ℓ. Second, $\mathrm{var}(\omega_{r,i}), \mathrm{var}(\omega_{s,i'})$ must both be non-zero, that is, an error must exist for unit i in the variable denoted by r, and for unit i'

in the variable denoted by s. Thirdly we must have either $C_{rs,i}^{(\ell)}$, $C_{rs,i'}^{(\ell)}$ both non-zero or $C_{sr,i}^{(\ell)}$, $C_{sr,i'}^{(\ell)}$ both non-zero, that is, an error covariance at level ℓ for the two variables $\mathbf{U}_r, \mathbf{U}_s$ must be specified and be non-zero for the units i and i'. Thus

$$M_{rs,ii'}^{(\ell)} = \mathrm{E}\left(\omega_{r,i}^{(\cdot)}\omega_{s,i'}^{(\ell)}\right) \text{ if } \mathbf{C}_{rs}^{(\ell)} \text{ exists,} \qquad (3.31)$$
$$\mathrm{E}\left(\omega_{r,i}^{(\ell)}\omega_{s,i'}^{(\cdot)}\right) \text{ if } \mathbf{C}_{sr}^{(\ell)} \text{ exists.}$$

In either case, $M_{rs,ii'}^{(\ell)}$ is the level-ℓ contribution to the total covariance between the error $\boldsymbol{\omega}_r$ in $\mathbf{U}_r$ for level-1 unit i and the error $\boldsymbol{\omega}_s$ in $\mathbf{U}_s$ for level-1 unit i'. We define the total (expected) error product matrix for $\boldsymbol{\omega}_r$, $\boldsymbol{\omega}_s$:

$$\mathbf{M}_{rs} = \sum_{\ell=1}^{L}\mathbf{M}_{rs}^{(\ell)}. \qquad (3.32)$$

In many cases only one of the matrices in this sum will be non-zero; but the definition allows the specification of a measurement error model in which errors in given variables $\mathbf{U}_r, \mathbf{U}_s$ arise and covary at different levels. With appropriate prior specification of either $\mathbf{C}_{rs}^{(\ell)}$ or $\mathbf{C}_{sr}^{(\ell)}$ for $\ell = 1, 2, \ldots, L$, we have

$$M_{rs,ii'} = \mathrm{E}\left(\omega_{r,i}\omega_{s,i'}\right), \qquad (3.33)$$
$$\mathbf{M}_{rs} = \mathrm{E}\left(\boldsymbol{\omega}_r\boldsymbol{\omega}_s^{\mathrm{T}}\right),$$

and we note that $\mathbf{M}_{sr} = \mathbf{M}_{rs}^{\mathrm{T}}$. A worked example based on model (3.14) and the illustrative data in Table 3.1 appears in the next section.

For convenience we write

$$\mathbf{M}_{r\eta} \equiv \sum_{u=1}^{m} \mathbf{M}_{r\bar{u}}, \ \mathbf{M}_{\eta r} \equiv \sum_{u=1}^{m} \mathbf{M}_{\bar{u}r}, \ \mathbf{M}_{\eta\eta} \equiv \sum_{u,u'=1}^{m} \mathbf{M}_{\bar{u}\bar{u}'}. \tag{3.34}$$

In addition we require to define error product matrices associated with the observed raw residuals $\widetilde{\mathbf{Y}}$. If the fixed parameters $\boldsymbol{\beta}$ are known we define

$$\begin{aligned} \widetilde{\mathbf{Y}} &\equiv \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \\ &= \widetilde{\mathbf{y}} + \boldsymbol{\lambda}, \\ \boldsymbol{\lambda} &\equiv \boldsymbol{\eta} - \boldsymbol{\xi}\boldsymbol{\beta}. \end{aligned} \tag{3.35}$$

We have

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{\omega}_r \boldsymbol{\lambda}^{\mathrm{T}}\right) &= \mathrm{E}\left[\boldsymbol{\omega}_r (\boldsymbol{\eta} - \boldsymbol{\xi}\boldsymbol{\beta})^{\mathrm{T}}\right] \\ &= \mathrm{E}\left(\boldsymbol{\omega}_r \boldsymbol{\eta}^{\mathrm{T}}\right) - \mathrm{E}\left(\boldsymbol{\omega}_r \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\xi}^{\mathrm{T}}\right) \\ &= \mathbf{M}_{r\eta} - \sum_{t=0}^{p-1} \beta_t \mathbf{M}_{rt}. \end{aligned} \tag{3.36}$$

We define

$$\begin{aligned} \mathbf{M}_{r\lambda} &\equiv \mathbf{M}_{r\eta} - \sum_{t-0}^{p-1} \beta_t \mathbf{M}_{rt}, \\ \mathbf{M}_{\lambda r} &\equiv \mathbf{M}_{r\lambda}^{\mathrm{T}}. \end{aligned} \tag{3.37}$$

Thus

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{\omega}_r \boldsymbol{\lambda}^{\mathrm{T}}\right) &= \mathbf{M}_{r\lambda}, \\ \mathrm{E}\left(\boldsymbol{\lambda}\boldsymbol{\omega}_r^{\mathrm{T}}\right) &= \mathbf{M}_{\lambda r}. \end{aligned} \tag{3.38}$$

We define

$$\begin{aligned}\mathbf{M}_{\eta\lambda} &\equiv \sum_{u=1}^{m} \mathbf{M}_{u\lambda} \\ &= \mathbf{M}_{\eta\eta} - \sum_{t=0}^{p-1} \beta_t \mathbf{M}_{\eta t}, \\ \mathbf{M}_{\lambda\eta} &\equiv \mathbf{M}_{\eta\lambda}^{\mathrm{T}},\end{aligned} \tag{3.39}$$

and we have

$$\begin{aligned}\mathrm{E}\left(\eta\lambda^{\mathrm{T}}\right) &= \mathbf{M}_{\eta\lambda}, \\ \mathrm{E}\left(\lambda\eta^{\mathrm{T}}\right) &= \mathbf{M}_{\lambda\eta}.\end{aligned} \tag{3.40}$$

We also have

$$\begin{aligned}\mathrm{E}\left(\lambda\lambda^{\mathrm{T}}\right) &= \mathrm{E}\left[(\eta-\xi\beta)(\eta-\xi\beta)^{\mathrm{T}}\right] \\ &= \mathrm{E}\left(\eta\eta^{\mathrm{T}}\right) - \mathrm{E}\left[(\xi\beta)\eta^{\mathrm{T}}\right] - \mathrm{E}\left[\eta(\xi\beta)^{\mathrm{T}}\right] + \mathrm{E}\left[(\xi\beta)(\xi\beta)^{\mathrm{T}}\right], \\ \mathrm{E}\left(\eta\eta^{\mathrm{T}}\right) &= \mathbf{M}_{\eta\eta}, \\ \mathrm{E}\left[(\xi\beta)\eta^{\mathrm{T}}\right] &= \sum_{t=0}^{p-1} \beta_t \, \mathrm{E}\left(\xi_t \eta^{\mathrm{T}}\right) = \sum_{t=0}^{p-1} \beta_t \mathbf{M}_{t\eta}, \\ \mathrm{E}\left[\eta(\xi\beta)^{\mathrm{T}}\right] &= \sum_{t=0}^{p-1} \beta_t \mathbf{M}_{\eta t}, \\ \mathrm{E}\left[(\xi\beta)(\xi\beta)^{\mathrm{T}}\right] &= \sum_{t,t'=0}^{p-1} \beta_t \beta_{t'} \, \mathrm{E}\left(\xi_t \xi_{t'}^{\mathrm{T}}\right) = \sum_{t,t'=0}^{p-1} \beta_t \beta_{t'} \mathbf{M}_{tt'}\end{aligned} \tag{3.41}$$

and we define

$$\mathbf{M}_{\lambda\lambda} \equiv \mathbf{M}_{\eta\eta} - \sum_{t=0}^{p-1} \beta_t \mathbf{M}_{t\eta} - \sum_{t=0}^{p-1} \beta_t \mathbf{M}_{\eta t} + \sum_{t,t'=0}^{p-1} \beta_t \beta_{t'} \mathbf{M}_{tt'}. \tag{3.42}$$

Thus

$$\mathrm{E}\left(\lambda\lambda^{\mathrm{T}}\right) = \mathbf{M}_{\lambda\lambda}. \tag{3.43}$$

The expressions in (3.35) to (3.43) assume knowledge of $\boldsymbol{\beta}$. Corresponding expressions based on an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ are denoted by attaching a 'hat' (^), thus:

$$
\begin{aligned}
\hat{\tilde{\mathbf{y}}} &\equiv \mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}, \\
\hat{\tilde{\mathbf{Y}}} &\equiv \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}, \\
\hat{\boldsymbol{\lambda}} &\equiv \boldsymbol{\eta} - \boldsymbol{\xi}\hat{\boldsymbol{\beta}}, \\
\mathbf{M}_{r\hat{\lambda}} &\equiv \mathbf{M}_{r\eta} - \sum_{t=0}^{p-1} \hat{\beta}_t \mathbf{M}_{rt}, \\
\mathbf{M}_{\eta\hat{\lambda}} &\equiv \mathbf{M}_{\eta\eta} - \sum_{t=0}^{p-1} \hat{\beta}_t \mathbf{M}_{\eta t}, \\
\mathbf{M}_{\hat{\lambda}\hat{\lambda}} &\equiv \mathbf{M}_{\eta\eta} - \sum_{t=0}^{p-1} \hat{\beta}_t \mathbf{M}_{t\eta} - \sum_{t=0}^{p-1} \hat{\beta}_t \mathbf{M}_{\eta t} + \sum_{t,t'=0}^{p-1} \hat{\beta}_t \hat{\beta}_{t'} \mathbf{M}_{tt'}.
\end{aligned}
\tag{3.44}
$$

3.6.4 Some examples

Consider again model (3.9) in section 3.4.1, which contains a general aptitude score $\mathbf{x}_1$ and its mean $\mathbf{x}_2$ as explanatory variables, with observed values $X_{1,i}$, $X_{2,i}$ for each pupil i. A covariance between the errors in $\mathbf{X}_1$ and $\mathbf{X}_2$ exists across pupils in the same school, and in the absence of other information this covariance is assumed to be constant for all pupils in a given school. Thus, for a given school j and any $i,i' \in B_j = B_j^{(2)}$, we have $C_{12,i}^{(2)} = C_{12,i'}^{(2)}$. The matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are unit matrices because errors in both variables exist for all pupils. In this case $\mathbf{M}_{12} = \mathbf{M}_{12}^{(2)}$ is the direct sum of J_2 square submatrices, where J_2 is the number of schools, each submatrix having elements equal to the constant error covariance between $\mathbf{X}_1$ and $\mathbf{X}_2$ for pupils in the corresponding school.

If in a different case we had values on two such variables for all level-1 units but it was believed that the error covariance between the two was not constant within a level-2 unit then for some j we should have non-zero but unequal values of $C_{12,i}^{(2)}$ and $C_{12,i'}^{(2)}$ for some $i,i' \in B_j$. Once again $\mathbf{K}_1$ and $\mathbf{K}_2$ would be unit matrices and $\mathbf{M}_{12}^{(2)}$ would be the direct sum of square submatrices. In the jth submatrix the elements in each row would be constant but the values in rows i and i' would differ. In such a case $\mathbf{M}_{12}^{(2)}$ would be non-symmetric.

The analysis of the bivariate model (3.14) for scores in mathematics and reading, described in section 3.4.3, uses explanatory variables $\mathbf{U}_3$ and $\mathbf{U}_4$ that are derived from the same variable $\mathbf{X}_1$. The error variance vectors for $\mathbf{U}_3$ and $\mathbf{U}_4$ are specified at level

2 (the pupil level), and $\mathbf{C}_{33}^{(1)} = \mathbf{C}_{44}^{(1)} = \mathbf{0}$. The error covariance vector $\mathbf{C}_{34}^{(2)}$ for $\mathbf{U}_3$ and $\mathbf{U}_4$ is just the error variance vector $\mathbf{C}_{11}^{(2)}$ for $\mathbf{X}_1$, and $\mathbf{C}_{34}^{(1)} = \mathbf{0}$. If we write $\tau_j^2 \equiv \mathrm{var}(\xi_{1,j})$ for the error variance of $\mathbf{X}_1$ for pupil j $(j = 1, 2, 3)$, the illustrative data in Table 3.1 require the following error variance and covariance vectors to be specified:

$$\mathbf{C}_{33}^{(2)} = \begin{pmatrix} \tau_1^2 \\ 0 \\ 0 \\ \tau_3^2 \\ 0 \\ \vdots \end{pmatrix}, \quad \mathbf{C}_{44}^{(2)} = \begin{pmatrix} 0 \\ \tau_1^2 \\ \tau_2^2 \\ 0 \\ \tau_3^2 \\ \vdots \end{pmatrix}, \quad \mathbf{C}_{34}^{(2)} = \begin{pmatrix} \tau_1^2 \\ \tau_1^2 \\ \tau_2^2 \\ \tau_3^2 \\ \tau_3^2 \\ \vdots \end{pmatrix}. \tag{3.45}$$

They generate the following error incidence matrices:

$$\mathbf{K}_3 = \begin{pmatrix} 1 & & & & & \\ & 0 & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix}, \quad \mathbf{K}_4 = \begin{pmatrix} 0 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & 1 & \\ & & & & & \ddots \end{pmatrix}, \tag{3.46}$$

and the following error product matrices:

$$\mathbf{M}_{33} = \mathbf{M}_{33}^{(2)} = \begin{pmatrix} \tau_1^2 & 0 & & & & \\ 0 & 0 & & & & \\ & & 0 & & & \\ & & & \tau_3^2 & 0 & \\ & & & 0 & 0 & \\ & & & & & \ddots \end{pmatrix}, \quad \mathbf{M}_{44} = \mathbf{M}_{44}^{(2)} = \begin{pmatrix} 0 & 0 & & & & \\ 0 & \tau_1^2 & & & & \\ & & \tau_2^2 & & & \\ & & & 0 & 0 & \\ & & & 0 & \tau_3^2 & \\ & & & & & \ddots \end{pmatrix}, \tag{3.47}$$

$$\mathbf{M}_{34} = \mathbf{M}_{34}^{(2)} = \begin{pmatrix} 0 & \tau_1^2 & & & & \\ 0 & 0 & & & & \\ & & 0 & & & \\ & & & 0 & \tau_3^2 & \\ & & & 0 & 0 & \\ & & & & & \ddots \end{pmatrix}, \quad \mathbf{M}_{43} = \mathbf{M}_{43}^{(2)} = \begin{pmatrix} 0 & 0 & & & & \\ \tau_1^2 & 0 & & & & \\ & & 0 & & & \\ & & & 0 & 0 & \\ & & & \tau_3^2 & 0 & \\ & & & & & \ddots \end{pmatrix}.$$

The (1,2)th element of $\mathbf{M}_{34}$, for example, is $\mathrm{E}(\omega_{3,1}\omega_{4,2})$. The (1,2)th element of $\mathbf{M}_{43}$ is $\mathrm{E}(\omega_{4,1}\omega_{3,2})$. The values produced are the appropriate ones, as $\omega_{3,1} = \omega_{4,2} = \xi_{1,1}$ and $\omega_{4,1} = \omega_{3,2} = 0$. The error incidence matrices derived from the error variance vectors control the generation of the expected error products: provided the error variance vectors contain values only for units for which errors exist, and zeros elsewhere, the user may specify error *co*variance vectors as full vectors.

3.6.5 Adjusted product matrices

In addition to the error product matrices we shall require estimates of the cross-product matrices of the true values of the variables. In the notation of Section 3.6.1, if $\mathbf{u}_r, \mathbf{u}_s$ are the true values of $\mathbf{U}_r, \mathbf{U}_s$, respectively, then

$$\begin{aligned} \mathrm{E}(\mathbf{U}_r\mathbf{U}_s^{\mathrm{T}}) &= \mathrm{E}\left[(\mathbf{u}_r + \boldsymbol{\omega}_r)(\mathbf{u}_s + \boldsymbol{\omega}_s)^{\mathrm{T}}\right] \\ &= \mathrm{E}(\mathbf{u}_r\mathbf{u}_s^{\mathrm{T}} + \boldsymbol{\omega}_r\mathbf{u}_s^{\mathrm{T}} + \mathbf{u}_r\boldsymbol{\omega}_s^{\mathrm{T}} + \boldsymbol{\omega}_r\boldsymbol{\omega}_s^{\mathrm{T}}) \\ &= \mathbf{u}_r\mathbf{u}_s^{\mathrm{T}} + \mathrm{E}(\boldsymbol{\omega}_r\boldsymbol{\omega}_s^{\mathrm{T}}), \quad \text{from assumption 3.2,} \\ &= \mathbf{u}_r\mathbf{u}_s^{\mathrm{T}} + \mathbf{M}_{rs}, \end{aligned} \tag{3.48}$$

from equation (3.33), provided suitable error variance and covariance vectors have been specified.

For $-m \le r,s \le p+q-1$ we define the *adjusted product matrix*

$$\mathbf{N}_{rs} \equiv \mathbf{U}_r\mathbf{U}_s^{\mathrm{T}} - \mathbf{M}_{rs}, \tag{3.49}$$

and now

$$\mathrm{E}(\mathbf{N}_{rs}) = \mathbf{u}_r\mathbf{u}_s^{\mathrm{T}}. \tag{3.50}$$

We shall use $\mathbf{N}_{rs}$ as an estimator of the cross-product matrix $\mathbf{u}_r\mathbf{u}_s^{\mathrm{T}}$ of the true values when required.

For convenience we write

$$\mathbf{N}_{ry} \equiv \sum_{u=1}^{m}\mathbf{N}_{r\bar{u}},\ \mathbf{N}_{yr} \equiv \sum_{u=1}^{m}\mathbf{N}_{\bar{u}r},\ \mathbf{N}_{yy} \equiv \sum_{u,u'=1}^{m}\mathbf{N}_{\bar{u}\bar{u}'}. \tag{3.51}$$

For the residuals we define

$$\begin{aligned}
&\mathbf{N}_{r\tilde{y}} \equiv \mathbf{U}_r\tilde{\mathbf{Y}}^{\mathrm{T}} - \mathbf{M}_{r\lambda},\ \mathbf{N}_{\tilde{y}r} \equiv \mathbf{N}_{r\tilde{y}}^{\mathrm{T}},\\
&\mathbf{N}_{r\hat{\tilde{y}}} \equiv \mathbf{U}_r\hat{\tilde{\mathbf{Y}}}^{\mathrm{T}} - \mathbf{M}_{r\hat{\lambda}},\ \mathbf{N}_{\hat{\tilde{y}}r} \equiv \mathbf{N}_{r\hat{\tilde{y}}}^{\mathrm{T}},
\end{aligned} \tag{3.52}$$

and we shall use $\mathbf{N}_{r\tilde{y}}$, $\mathbf{N}_{\tilde{y}r}$, etc. as estimators of $\mathbf{u}_r\tilde{\mathbf{y}}^{\mathrm{T}}$, $\tilde{\mathbf{y}}\mathbf{u}_r^{\mathrm{T}}$, etc. when required.

We also define

$$\begin{aligned}
&\mathbf{N}_{\tilde{y}\tilde{y}} \equiv \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}} - \mathbf{M}_{\lambda\lambda},\ \mathbf{N}_{y\tilde{y}} \equiv \mathbf{Y}\tilde{\mathbf{Y}}^{\mathrm{T}} - \mathbf{M}_{\eta\lambda},\ \mathbf{N}_{\tilde{y}y} \equiv \tilde{\mathbf{Y}}\mathbf{Y}^{\mathrm{T}} - \mathbf{M}_{\lambda\eta},\\
&\mathbf{N}_{\hat{\tilde{y}}\hat{\tilde{y}}} \equiv \hat{\tilde{\mathbf{Y}}}\hat{\tilde{\mathbf{Y}}}^{\mathrm{T}} - \mathbf{M}_{\hat{\lambda}\hat{\lambda}},\ \mathbf{N}_{y\hat{\tilde{y}}} \equiv \mathbf{Y}\hat{\tilde{\mathbf{Y}}}^{\mathrm{T}} - \mathbf{M}_{\eta\hat{\lambda}},\ \mathbf{N}_{\hat{\tilde{y}}y} \equiv \hat{\tilde{\mathbf{Y}}}\mathbf{Y}^{\mathrm{T}} - \mathbf{M}_{\hat{\lambda}\eta},
\end{aligned} \tag{3.53}$$

and use $\mathbf{N}_{\tilde{y}\tilde{y}}$, $\mathbf{N}_{y\tilde{y}}$, $\mathbf{N}_{\tilde{y}y}$, etc. as estimators of $\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}$, $\mathbf{y}\tilde{\mathbf{y}}^{\mathrm{T}}$, $\tilde{\mathbf{y}}\mathbf{y}^{\mathrm{T}}$, etc.

3.7 Estimating the fixed parameters

3.7.1 Introduction

The procedure we propose in order to estimate the parameters of the model described in Section 3.5 is an adaptation of the iterative generalised least squares method (iterative GLS, or IGLS) described by Goldstein (1986). It begins with the estimation of $\boldsymbol{\beta}$ (the fixed parameters) by ordinary least squares (OLS) following the procedure described in Section 3.7.2. These estimators are then used to obtain initial (OLS) estimators of the random parameters as described in Section 3.8.2. Thereafter the fixed and random parameters are estimated alternately as described in the remainder of Sections 3.7 and 3.8, using existing estimators where these are required, until a convergence criterion is met.

Assume that the first two stages have been completed. The GLS estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{v}^{-1}\mathbf{x})^{-1}\mathbf{x}^T\mathbf{v}^{-1}\mathbf{y}, \tag{3.54}$$

where $\mathbf{v}$ is the residual covariance matrix of the responses and we assume $\mathbf{v}^{-1}$ exists. We do not know $\mathbf{x}$, $\mathbf{y}$, or $\mathbf{v}^{-1}$ and we seek an expression involving known (observed) values of the variables, suitably adjusted so that in expectation it approximates $(\mathbf{x}^T\mathbf{v}^{-1}\mathbf{x})^{-1}\mathbf{x}^T\mathbf{v}^{-1}\mathbf{y}$. Our general strategy, in fact, will be to work with expressions that in expectation approximate the desired, but unknown, quantities. The first step here is to find matrices $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\Delta}}$, such that

$$\mathrm{E}(\hat{\boldsymbol{\Gamma}}) = \mathbf{x}^T\mathbf{v}^{-1}\mathbf{x}, \quad \mathrm{E}(\hat{\boldsymbol{\Delta}}) = \mathbf{x}^T\mathbf{v}^{-1}\mathbf{y}. \tag{3.55}$$

In the remainder of Chapter 3 we shall use the symbol $\approx$ to mean '(which) may be estimated by' in the sense that the expectation of the expression on the right of the symbol approximates the value of the expression on the left, possibly with some bias which we may attempt to reduce by a further approximation.

Recalling the definition of $\sigma_{ss'}^{(\ell)}$ in equation (3.22) and writing $\hat{\sigma}_{ss'}^{(\ell)}$ for its current estimator we define, for each $\ell \in \{1,2,\ldots,L\}$,

$$\hat{\mathbf{V}}_j^{(\ell)} \equiv \sum_{s,s'=p}^{p+q-1} \hat{\sigma}_{ss'}^{(\ell)} \mathbf{Z}_{s,j} \mathbf{Z}_{s',j}^{\mathrm{T}}, \quad \hat{\mathbf{V}}^{(\ell)} \equiv \bigoplus_{j=1}^{J_\ell} \hat{\mathbf{V}}_j^{(\ell)}, \quad \hat{\mathbf{V}} \equiv \sum_{\ell=1}^{L} \hat{\mathbf{V}}^{(\ell)},$$
$$\hat{\mathbf{T}}_j^{(\ell)} \equiv \sum_{s,s'=p}^{p+q-1} \hat{\sigma}_{ss'}^{(\ell)} \mathbf{M}_{ss',j}, \quad \hat{\mathbf{T}}^{(\ell)} \equiv \bigoplus_{j=1}^{J_\ell} \hat{\mathbf{T}}_j^{(\ell)}, \quad \hat{\mathbf{T}} \equiv \sum_{\ell=1}^{L} \hat{\mathbf{T}}^{(\ell)}, \tag{3.56}$$

where the subscript j denotes the jth level-ℓ block of a vector or matrix. Note that for given s, s' with $s \neq s'$, for given ℓ, and for each $j \in \{1,2,\ldots,J_\ell\}$, $\hat{\mathbf{V}}_j^{(\ell)}$ contains separate terms $\hat{\sigma}_{ss'}^{(\ell)} \mathbf{Z}_{s,j} \mathbf{Z}_{s',j}^{\mathrm{T}}$ and $\hat{\sigma}_{s's}^{(\ell)} \mathbf{Z}_{s',j} \mathbf{Z}_{s,j}^{\mathrm{T}}$ and $\hat{\mathbf{T}}_j^{(\ell)}$ contains separate terms $\hat{\sigma}_{ss'}^{(\ell)} \mathbf{M}_{ss'}$ and $\hat{\sigma}_{s's}^{(\ell)} \mathbf{M}_{s's}$. If $s = s'$ only one such term exists in each case.

By assumption 3.2 (p49), and ignoring error in the values from which the $\mathbf{M}_{ss'}$ are derived, we have

$$\mathbf{v} = \mathrm{E}(\hat{\mathbf{V}}) - \hat{\mathbf{T}}, \tag{3.57}$$

and we write

$$\hat{\mathbf{v}} \equiv \hat{\mathbf{V}} - \hat{\mathbf{T}}. \tag{3.58}$$

Assuming the existence of $\hat{\mathbf{v}}^{-1}$ a first approximation to $\mathbf{x}_r^{\mathrm{T}} \mathbf{v}^{-1} \mathbf{x}_{r'}$, that is the (r,r')th element of $\mathbf{x}^{\mathrm{T}} \mathbf{v}^{-1} \mathbf{x}$ for $r,r' \in \{0,1,\ldots,p-1\}$, is given by

$$\hat{\Gamma}_{rr'} \equiv \mathrm{tr}\left(\mathbf{N}_{rr'}\hat{\mathbf{v}}^{-1}\right), \tag{3.59}$$

for we have

$$\begin{aligned} &\mathbf{x}_r^{\mathrm{T}}\hat{\mathbf{v}}^{-1}\mathbf{x}_{r'} = \mathrm{tr}\left(\mathbf{x}_r^{\mathrm{T}}\hat{\mathbf{v}}^{-1}\mathbf{x}_{r'}\right) = \mathrm{tr}\left(\mathbf{x}_{r'}\mathbf{x}_r^{\mathrm{T}}\hat{\mathbf{v}}^{-1}\right) = \mathrm{tr}\left(\mathbf{x}_r\mathbf{x}_{r'}^{\mathrm{T}}\hat{\mathbf{v}}^{-1}\right), \text{ and} \\ &\mathrm{E}\left(\mathbf{x}_r\mathbf{x}_{r'}^{\mathrm{T}}\right) = \mathbf{N}_{rr'}, \end{aligned} \tag{3.60}$$

from (3.50). Likewise a first approximation to $\mathbf{x}_r^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{y}$ is

$$\hat{\Delta}_r \equiv \mathrm{tr}\left(\mathbf{N}_{ry}\hat{\mathbf{v}}^{-1}\right), \tag{3.61}$$

and, writing $\hat{\mathbf{\Gamma}} = \left\{\hat{\Gamma}_{rr'}\right\}_{r,r'}$, $\hat{\mathbf{\Delta}} = \left\{\hat{\Delta}_r\right\}_r$, we have, as a first approximation,

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{\Gamma}}^{-1}\hat{\mathbf{\Delta}}. \tag{3.62}$$

But, although $\mathbf{N}_{rr'}$, $\mathbf{N}_{ry}$, $\hat{\mathbf{v}}$ are unbiased estimators of $\mathbf{x}_r\mathbf{x}_{r'}^{\mathrm{T}}$, $\mathbf{x}_r\mathbf{y}^{\mathrm{T}}$, $\mathbf{v}$, respectively, they each typically contain measurement error and the correlations between these errors will typically produce bias in one or both of the products $\mathbf{N}_{rr'}\hat{\mathbf{v}}^{-1}$, $\mathbf{N}_{ry}\hat{\mathbf{v}}^{-1}$, for example if an explanatory variable with error has a coefficient with fixed and random components as in the standard 'random-coefficients' model. We now show how these biases may be reduced.

If we again ignore error in the prior values on which the $\mathbf{M}_{ss'}$ are based, the matrix $\hat{\mathbf{v}}$ contains error from two sources: measurement error in the $\mathbf{Z}_s$ and sampling error in the $\hat{\sigma}_{ss'}^{(\ell)}$. We may write:

$\hat{\mathbf{v}} = \mathbf{v} + \mathbf{w} + \mathbf{u}$, where

$$\mathbf{w} \equiv \sum_{\ell=1}^{L} \mathbf{w}^{(\ell)},\ \ \mathbf{w}^{(\ell)} \equiv \bigoplus_{j=1}^{J_\ell} \mathbf{w}_j^{(\ell)},\ \ \mathbf{w}_j^{(\ell)} \equiv \sum_{s,s'=p}^{p+q-1} \hat{\sigma}_{ss'}^{(\ell)} \left(\mathbf{z}_{s,j} \boldsymbol{\zeta}_{s',j}^{\mathrm{T}} + \boldsymbol{\zeta}_{s,j} \mathbf{z}_{s',j}^{\mathrm{T}} + \boldsymbol{\zeta}_{s,j} \boldsymbol{\zeta}_{s',j}^{\mathrm{T}} - \mathbf{M}_{ss',j} \right),$$

$$\mathbf{u} \equiv \sum_{\ell=1}^{L} \mathbf{u}^{(\ell)},\ \ \mathbf{u}^{(\ell)} \equiv \bigoplus_{j=1}^{J_\ell} \mathbf{u}_j^{(\ell)},\ \ \mathbf{u}_j^{(\ell)} \equiv \sum_{s,s'=p}^{p+q-1} \hat{\varsigma}_{ss'}^{(\ell)} \mathbf{z}_{s,j} \mathbf{z}_{s',j}^{\mathrm{T}},$$

$$\hat{\varsigma}_{ss'}^{(\ell)} \equiv \hat{\sigma}_{ss'}^{(\ell)} - \sigma_{ss'}^{(\ell)}.$$

(3.63)

Thus, $\mathbf{w}$ is the error in $\hat{\mathbf{v}}$ resulting from the measurement errors $\boldsymbol{\zeta}_s$ etc., including their contributions in respect of the sampling error in the $\hat{\sigma}_{ss'}^{(\ell)}$; $\mathbf{u}$ is the error due to sampling which remains in $\hat{\mathbf{v}}$ after all measurement errors are accounted for. We assume $\mathrm{E}(\mathbf{w}) = \mathrm{E}(\mathbf{u}) = \mathbf{0}$.

It is convenient to express $\mathbf{w}^{(\ell)}$, $\mathbf{u}^{(\ell)}$ also as double sums of $N \times N$ matrices:

$$\mathbf{w}^{(\ell)} = \sum_{j=1}^{J_\ell} \sum_{s,s'=p}^{p+q-1} \hat{\sigma}_{ss'}^{(\ell)} \left(\mathbf{z}_{s(j)} \boldsymbol{\zeta}_{s'(j)}^{\mathrm{T}} + \boldsymbol{\zeta}_{s(j)} \mathbf{z}_{s'(j)}^{\mathrm{T}} + \boldsymbol{\zeta}_{s(j)} \boldsymbol{\zeta}_{s'(j)}^{\mathrm{T}} - \mathbf{M}_{ss'(jj)} \right), \tag{3.64}$$

$$\mathbf{u}^{(\ell)} = \sum_{j=1}^{J_\ell} \sum_{s,s'=p}^{p+q-1} \hat{\varsigma}_{ss'}^{(\ell)} \mathbf{z}_{s(j)} \mathbf{z}_{s'(j)}^{\mathrm{T}},$$

where $\mathbf{z}_{s(j)}$, for example, is defined as the $N \times 1$ vector whose jth level-ℓ block is equal to $\mathbf{z}_{s,j}$ and which is zero elsewhere and $\mathbf{M}_{ss'(jj)}$ is the $N \times N$ matrix whose jth level-ℓ block is equal to $\mathbf{M}_{ss',j}$ and which is zero elsewhere.

Now

$$\begin{aligned}(\mathbf{v} + \mathbf{w} + \mathbf{u})^{-1} &= \left(\mathbf{I} + \mathbf{v}^{-1}(\mathbf{w} + \mathbf{u})\right)^{-1} \mathbf{v}^{-1} \\ &= \mathbf{v}^{-1} - \mathbf{v}^{-1}(\mathbf{w} + \mathbf{u})\mathbf{v}^{-1} + \mathbf{v}^{-1}\left((\mathbf{w} + \mathbf{u})\mathbf{v}^{-1}\right)^2 - \mathbf{v}^{-1}\left((\mathbf{w} + \mathbf{u})\mathbf{v}^{-1}\right)^3 + \ldots,\end{aligned} \tag{3.65}$$

provided this series converges, and therefore

$$
\begin{aligned}
&\mathrm{E}\,\mathrm{tr}\left(\mathbf{N}_{rr'}\hat{\mathbf{v}}^{-1}\right) \qquad (3.66)\\
&= \mathrm{E}\,\mathrm{tr}\left[\left(\mathbf{x}_r\mathbf{x}_{r'}^T + \mathbf{x}_r\xi_{r'}^T + \xi_r\mathbf{x}_{r'}^T + \xi_r\xi_{r'}^T - \mathbf{M}_{rr'}\right)\left(\mathbf{v}^{-1} - \mathbf{v}^{-1}(\mathbf{w}+\mathbf{u})\mathbf{v}^{-1} + \mathbf{v}^{-1}\left((\mathbf{w}+\mathbf{u})\mathbf{v}^{-1}\right)^2 - \ldots\right)\right]\\
&= \mathrm{tr}\left(\mathbf{x}_r\mathbf{x}_{r'}^T\mathbf{v}^{-1}\right)\\
&-\mathrm{E}\,\mathrm{tr}\left(\left(\mathbf{x}_r\xi_{r'}^T + \xi_r\mathbf{x}_{r'}^T\right)\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\right)\\
&+\mathrm{E}\,\mathrm{tr}\left(\mathbf{x}_r\mathbf{x}_{r'}^T\mathbf{v}^{-1}\left(\left(\mathbf{w}\mathbf{v}^{-1}\right)^2 + \left(\mathbf{u}\mathbf{v}^{-1}\right)^2\right)\right)\\
&+\mathrm{E}\,\mathrm{tr}\left(\left(\mathbf{x}_r\xi_{r'}^T + \xi_r\mathbf{x}_{r'}^T\right)\mathbf{v}^{-1}\left(\mathbf{w}\mathbf{v}^{-1}\mathbf{u}\mathbf{v}^{-1} + \mathbf{u}\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\right)\right)\\
&+\mathrm{E}\,\mathrm{tr}\left(\left(\xi_r\xi_{r'}^T - \mathbf{M}_{rr'}\right)\mathbf{v}^{-1}\left(\mathbf{w}\mathbf{v}^{-1}\right)^2\right)\\
&-\mathrm{E}\,\mathrm{tr}\left(\left(\mathbf{x}_r\xi_{r'}^T + \xi_r\mathbf{x}_{r'}^T\right)\mathbf{v}^{-1}\left(\left(\mathbf{w}\mathbf{v}^{-1}\right)^3 + \mathbf{w}\mathbf{v}^{-1}\left(\mathbf{u}\mathbf{v}^{-1}\right)^2 + \mathbf{u}\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\mathbf{u}\mathbf{v}^{-1} + \left(\mathbf{u}\mathbf{v}^{-1}\right)^2\mathbf{w}\mathbf{v}^{-1}\right)\right)\\
&-\mathrm{E}\,\mathrm{tr}\left(\left(\mathbf{x}_r\mathbf{x}_{r'}^T + \xi_r\xi_{r'}^T - \mathbf{M}_{rr'}\right)\mathbf{v}^{-1}\left(\left(\mathbf{w}\mathbf{v}^{-1}\right)^2\mathbf{u}\mathbf{v}^{-1} + \mathbf{w}\mathbf{v}^{-1}\mathbf{u}\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1} + \mathbf{u}\mathbf{v}^{-1}\left(\mathbf{w}\mathbf{v}^{-1}\right)^2\right)\right)\\
&-\mathrm{E}\,\mathrm{tr}\left(\mathbf{x}_r\mathbf{x}_{r'}^T\mathbf{v}^{-1}\left(\mathbf{u}\mathbf{v}^{-1}\right)^3\right)\\
&+\ldots,
\end{aligned}
$$

where we have used assumption 3.3 (p50) to eliminate terms of odd degree in the measurement errors. We may now write

$$
\begin{aligned}
\mathbf{x}_r^T\mathbf{v}^{-1}\mathbf{x}_{r'} &= \mathrm{tr}\left(\mathbf{x}_r\mathbf{x}_{r'}^T\mathbf{v}^{-1}\right) \qquad (3.67)\\
&\approx \mathrm{tr}\left(\mathbf{N}_{rr'}\hat{\mathbf{v}}^{-1}\right)\\
&+\hat{\mathrm{E}}\,\mathrm{tr}\left(\left(\mathbf{x}_r\xi_{r'}^T + \xi_r\mathbf{x}_{r'}^T\right)\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\right)\\
&-\hat{\mathrm{E}}\,\mathrm{tr}\left(\mathbf{x}_r\mathbf{x}_{r'}^T\mathbf{v}^{-1}\left(\left(\mathbf{w}\mathbf{v}^{-1}\right)^2 + \left(\mathbf{u}\mathbf{v}^{-1}\right)^2\right)\right)\\
&-\hat{\mathrm{E}}\,\mathrm{tr}\left(\left(\mathbf{x}_r\xi_{r'}^T + \xi_r\mathbf{x}_{r'}^T\right)\mathbf{v}^{-1}\left(\mathbf{w}\mathbf{v}^{-1}\mathbf{u}\mathbf{v}^{-1} + \mathbf{u}\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\right)\right)\\
&+\ldots,
\end{aligned}
$$

where $\hat{\mathrm{E}}$ in each case denotes an estimator of the corresponding expectation.

The convergence of the right-hand side of (3.67) and the terms to be included in the estimators depend on assumptions about the moments of the measurement errors and of the sampling errors. We recall that $\mathbf{M}_{rr'}$ may be non-symmetric, and this complicates any general theoretical treatment of the convergence problem. Simulation may be used to find practical limits to the validity of the procedure which we now outline. An alternative procedure, which avoids use of the expansion in equation

(3.65) at the expense of some loss of efficiency in the resulting estimators, is given in Section 3.7.2.

Consider the term $\hat{\mathrm{E}}\,\mathrm{tr}\left(\left(\mathbf{x}_r\boldsymbol{\xi}_{r'}^{\mathrm{T}}+\boldsymbol{\xi}_r\mathbf{x}_{r'}^{\mathrm{T}}\right)\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\right)$ in equation (3.67). We write

$$\mathrm{E}\,\mathrm{tr}\left(\left(\mathbf{x}_r\boldsymbol{\xi}_{r'}^{\mathrm{T}}+\boldsymbol{\xi}_r\mathbf{x}_{r'}^{\mathrm{T}}\right)\mathbf{v}^{-1}\mathbf{w}\mathbf{v}^{-1}\right)=\sum_{\ell_1=1}^{L}\mathrm{E}\,\mathrm{tr}\left(\left(\mathbf{x}_r\boldsymbol{\xi}_{r'}^{\mathrm{T}}+\boldsymbol{\xi}_r\mathbf{x}_{r'}^{\mathrm{T}}\right)\mathbf{v}^{-1}\mathbf{w}^{(\ell_1)}\mathbf{v}^{-1}\right) \tag{3.68}$$

and the contribution from $\mathbf{w}^{(\ell_1)}$ for given ℓ_1 is

$$\sum_{j_1=1}^{J_{\ell_1}}\sum_{s_1,s_1'=p}^{p+q-1}\hat{\sigma}_{s_1s_1'}^{(\ell_1)}\,\mathrm{E}\,\mathrm{tr}\left[\left(\mathbf{x}_r\boldsymbol{\xi}_{r'}^{\mathrm{T}}+\boldsymbol{\xi}_r\mathbf{x}_{r'}^{\mathrm{T}}\right)\mathbf{v}^{-1}\left(\mathbf{z}_{s_1(j_1)}\boldsymbol{\zeta}_{s_1'(j_1)}^{\mathrm{T}}+\boldsymbol{\zeta}_{s_1(j_1)}\mathbf{z}_{s_1'(j_1)}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right]. \tag{3.69}$$

We simplify notation by writing

$$a\equiv r,\ b\equiv r',\ \breve{p}\equiv s_1(j_1),\ \breve{q}\equiv s_1'(j_1). \tag{3.70}$$

The trace in (3.69) now becomes

$$\begin{aligned}
&\mathrm{E}\,\mathrm{tr}\left[\left(\mathbf{x}_a\boldsymbol{\xi}_b^{\mathrm{T}}+\boldsymbol{\xi}_a\mathbf{x}_b^{\mathrm{T}}\right)\mathbf{v}^{-1}\left(\mathbf{z}_{\breve{p}}\boldsymbol{\zeta}_{\breve{q}}^{\mathrm{T}}+\boldsymbol{\zeta}_{\breve{p}}\mathbf{z}_{\breve{q}}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right]\\
&=\mathrm{E}\,\mathrm{tr}\left(\mathbf{x}_a\boldsymbol{\xi}_b^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{p}}\boldsymbol{\zeta}_{\breve{q}}^{\mathrm{T}}\mathbf{v}^{-1}\right)+\mathrm{E}\,\mathrm{tr}\left(\mathbf{x}_a\boldsymbol{\xi}_b^{\mathrm{T}}\mathbf{v}^{-1}\boldsymbol{\zeta}_{\breve{p}}\mathbf{z}_{\breve{q}}^{\mathrm{T}}\mathbf{v}^{-1}\right)\\
&\quad+\mathrm{E}\,\mathrm{tr}\left(\boldsymbol{\xi}_a\mathbf{x}_b^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{p}}\boldsymbol{\zeta}_{\breve{q}}^{\mathrm{T}}\mathbf{v}^{-1}\right)+\mathrm{E}\,\mathrm{tr}\left(\boldsymbol{\xi}_a\mathbf{x}_b^{\mathrm{T}}\mathbf{v}^{-1}\boldsymbol{\zeta}_{\breve{p}}\mathbf{z}_{\breve{q}}^{\mathrm{T}}\mathbf{v}^{-1}\right)\\
&=\mathrm{tr}\left[\mathbf{x}_a\mathbf{z}_{\breve{p}}^{\mathrm{T}}\mathbf{v}^{-1}\,\mathrm{E}\left(\boldsymbol{\xi}_b\boldsymbol{\zeta}_{\breve{q}}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right]+\mathrm{tr}\left(\mathbf{x}_a\mathbf{z}_{\breve{q}}^{\mathrm{T}}\mathbf{v}^{-1}\right)\mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{\xi}_b^{\mathrm{T}}\mathbf{v}^{-1}\boldsymbol{\zeta}_{\breve{p}}\right)\right]\\
&\quad+\mathrm{tr}\left(\mathbf{x}_b^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{p}}\right)\mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{\xi}_a\boldsymbol{\zeta}_{\breve{q}}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right]+\mathrm{tr}\left[\mathbf{x}_b\mathbf{z}_{\breve{q}}^{\mathrm{T}}\mathbf{v}^{-1}\,\mathrm{E}\left(\boldsymbol{\xi}_a\boldsymbol{\zeta}_{\breve{p}}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right]\\
&\approx\mathrm{tr}\left(\mathbf{N}_{a\breve{p}}\hat{\mathbf{v}}^{-1}\mathbf{M}_{b\breve{q}}\hat{\mathbf{v}}^{-1}\right)+\mathrm{tr}\left(\mathbf{M}_{a\breve{p}}\hat{\mathbf{v}}^{-1}\mathbf{N}_{b\breve{q}}\hat{\mathbf{v}}^{-1}\right)\\
&\quad+\mathrm{tr}\left(\mathbf{N}_{b\breve{p}}\hat{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{M}_{a\breve{q}}\hat{\mathbf{v}}^{-1}\right)+\mathrm{tr}\left(\mathbf{M}_{b\breve{p}}\hat{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{N}_{a\breve{q}}\hat{\mathbf{v}}^{-1}\right),
\end{aligned} \tag{3.71}$$

where, for example,

$$\mathbf{N}_{a\breve{p}}\equiv\mathbf{N}_{a,s_1(j_1)}\equiv\mathbf{N}_{as_1(\cdot j_1)} \tag{3.72}$$

is defined as the $N \times N$ matrix whose ith column matches that of $\mathbf{N}_{as_1}$ when $i \in B_{j_1}^{(\ell_1)}$ and which is zero elsewhere.

Using (3.67), we can now begin to write down the second approximation to $\mathbf{x}_r^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{x}_{r'}$:

$$
\begin{aligned}
&\mathbf{x}_r^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{x}_{r'} \\
&\approx \mathrm{tr}\left(\mathbf{N}_{rr'}\hat{\mathbf{v}}^{-1}\right) \\
&+\sum_{\ell_1=1}^{L}\sum_{j_1-1}^{J_{\ell_1}}\sum_{s_1,s_1'=p}^{p+q-1}\hat{\sigma}_{s_1s_1'}^{(\ell_1)}\left\{\begin{array}{l}\mathrm{tr}\left(\mathbf{N}_{rs_1(\cdot j_1)}\hat{\mathbf{v}}^{-1}\mathbf{M}_{r's_1'(\cdot j_1)}\hat{\mathbf{v}}^{-1}\right)+\mathrm{tr}\left(\mathbf{M}_{rs_1(\cdot j_1)}\hat{\mathbf{v}}^{-1}\mathbf{N}_{r's_1'(\cdot j_1)}\hat{\mathbf{v}}^{-1}\right) \\ +\mathrm{tr}\left(\mathbf{N}_{r's_1(\cdot j_1)}\hat{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{M}_{rs_1'(\cdot j_1)}\hat{\mathbf{v}}^{-1}\right)+\mathrm{tr}\left(\mathbf{M}_{r's_1(\cdot j_1)}\hat{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{N}_{rs_1'(\cdot j_1)}\hat{\mathbf{v}}^{-1}\right)\end{array}\right\} \\
&-\ldots.
\end{aligned}
\tag{3.73}
$$

The second approximation still contains bias, which we should remove. For example, the product $\mathbf{N}_{rs_1(\cdot j_1)}\hat{\mathbf{v}}^{-1}\mathbf{M}_{r's_1'(\cdot j_1)}\hat{\mathbf{v}}^{-1}$ contains terms of order m^4u^{-4}, where m^4 indicates a product of measurement errors of order 4 and u^{-4} is the reciprocal of a 4th-degree product of true values of explanatory variables. Such terms are generally not negligible. Thus, a third approximation is required, and so on, until all such non-negligible biases have been removed.

3.7.2 Estimation of $\boldsymbol{\beta}$ with an error-free weight matrix

The complexity of the procedure outlined in the last section stems mainly from the presence of measurement error in the matrix $\hat{\mathbf{v}}$. But instead of using $\hat{\mathbf{v}}$ we may use the matrix $\bar{\mathbf{v}}$ obtained from $\hat{\mathbf{V}}$ (equation 3.56) by removing all terms in it containing measurement error. This does not affect the consistency or inconsistency of the resulting estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\theta}}$, though these will not be fully efficient. (Note that we do not demonstrate consistency of any estimator that we propose: our first priority is to find estimators that appear to be effective in practice in reducing bias.) We shall use the term *purged-V estimators* to describe estimators obtained by using such a weight matrix $\bar{\mathbf{v}}$. *OLS estimators* are obtained by setting $\bar{\mathbf{v}} = \mathbf{I}_N$, the identity matrix.

We first develop the estimators ignoring all moments of the sampling errors and then consider a further adjustment for sampling error in $\hat{\boldsymbol{\beta}}$. With the notation of Section 3.7.1, we write:

$$\hat{\Gamma}_{ab} = \mathrm{tr}\left(\mathbf{N}_{ab}\bar{\mathbf{v}}^{-1}\right), \ \hat{\Delta}_a = \mathrm{tr}\left(\mathbf{N}_{ay}\bar{\mathbf{v}}^{-1}\right), \tag{3.74}$$

and now the estimator

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{\Delta}} \tag{3.75}$$

provides a first approximation to $\boldsymbol{\beta}$. The estimator in this form is in theory biased on account of the correlation of the measurement errors still present in $\hat{\boldsymbol{\Gamma}}^{-1}$ and $\hat{\boldsymbol{\Delta}}$, although in the simulations that we have carried out, and which we report in Chapter 4, relative bias has been found to be small. We outline a possible further adjustment in Section 3.7.3.

We now develop a robust, 'sandwich' estimator for $\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right)$ with $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{\Delta}}$ as in (3.74)–(3.75). Our model is based on true values of the variables **x**, **y**, and **z**, and we estimate $\mathrm{cov}\left(\hat{\boldsymbol{\beta}}|\mathbf{x},\bar{\mathbf{v}}\right)$, the covariance of $\hat{\boldsymbol{\beta}}$ that would occur in repeated sampling from a population characterised by those true values. If we knew the true values **x**, **y**, **z** and used these in the estimation of $\boldsymbol{\beta}$ we should have:

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}|\mathbf{x},\overline{\mathbf{v}}\right) \equiv \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}-\mathrm{E}\left(\hat{\boldsymbol{\beta}}|\mathbf{x},\overline{\mathbf{v}}\right)\right)\left(\hat{\boldsymbol{\beta}}-\mathrm{E}\left(\hat{\boldsymbol{\beta}}|\mathbf{x},\overline{\mathbf{v}}\right)\right)^{\mathrm{T}}\right]$$
$$= \mathrm{E}\left\{\left[\left(\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\right]\left[\left(\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\right]^{\mathrm{T}}\right\}. \tag{3.76}$$

In fact we have used estimators $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\Delta}}$ for $\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x}$, $\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{y}$, respectively. If there were no error in the observed variables **X** and **Y** we could rewrite (3.76) as

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}|\mathbf{X},\overline{\mathbf{v}}\right) = \left(\mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\,\mathrm{E}\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}|\mathbf{X},\overline{\mathbf{v}}\right]\overline{\mathbf{v}}^{-1}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{X}\right)^{-1}, \tag{3.77}$$

and form the sandwich estimator by replacing the expectation $\mathrm{E}\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\right]$ by a matrix of the same form using the observed raw residuals $\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}$. Since $\mathrm{E}\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\right]$ is block-diagonal at level L the sandwich replacement would be

$$\sum_{j=1}^{J_L}\left(\mathbf{Y}_j-\mathbf{X}_j\hat{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_j-\mathbf{X}_j\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}. \tag{3.78}$$

We should then write

$$\mathbf{P} = \mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\left[\sum_{j=1}^{J_L}\left(\mathbf{Y}_j-\mathbf{X}_j\hat{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_j-\mathbf{X}_j\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\right]\overline{\mathbf{v}}^{-1}\mathbf{X}, \tag{3.79}$$
$$\hat{\mathrm{cov}}\left(\hat{\boldsymbol{\beta}}|\mathbf{X},\overline{\mathbf{v}}\right) = \left(\mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{X}\right)^{-1}\mathbf{P}\left(\mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{X}\right)^{-1}.$$

For the case with measurement error in **X** and **Y** one approach is to seek a $p \times p$ matrix $\hat{\mathbf{Q}}$, based on **P** and suitably adjusted, and then write

$$\hat{\boldsymbol{\Lambda}} \equiv \hat{\mathrm{cov}}\left(\hat{\boldsymbol{\beta}}|\mathbf{x},\overline{\mathbf{v}}\right) = \hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{Q}}\hat{\boldsymbol{\Gamma}}^{-1}. \tag{3.80}$$

The adjustment to $\mathbf{P}$ should reflect the adjustments made in the estimation of $\boldsymbol{\beta}$, in this case the use of $\hat{\boldsymbol{\Gamma}}$ to estimate $\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x}$ and $\hat{\boldsymbol{\Delta}}$ to estimate $\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{y}$. Variances and covariances of errors remaining in $\hat{\boldsymbol{\beta}}$ should be present in the expression for $\hat{\boldsymbol{\Lambda}}$.

Let $a,b \in \{0,1,\ldots,p-1\}$ and consider again equation (3.76). We have

$$\begin{aligned}\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) &= (\mathbf{X}-\boldsymbol{\xi})^{\mathrm{T}}\overline{\mathbf{v}}^{-1}(\widetilde{\mathbf{Y}}-\boldsymbol{\lambda}) \\ &= \mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\widetilde{\mathbf{Y}} - \mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\boldsymbol{\lambda} - \boldsymbol{\xi}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\widetilde{\mathbf{Y}} + \boldsymbol{\xi}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\boldsymbol{\lambda} \\ &= \mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\widetilde{\mathbf{Y}} - \mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\boldsymbol{\lambda} - \boldsymbol{\xi}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\widetilde{\mathbf{y}} - \boldsymbol{\xi}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\boldsymbol{\lambda}\end{aligned} \tag{3.81}$$

and therefore, using assumption 3.2 (p49),

$$\begin{aligned}\mathbf{x}_a^{\mathrm{T}}\overline{\mathbf{v}}^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) &\approx \mathbf{X}_a^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\widetilde{\mathbf{Y}} - \mathrm{tr}\left(\mathbf{M}_{a\lambda}\overline{\mathbf{v}}^{-1}\right) \\ &= \mathrm{tr}\left(\mathbf{N}_{a\tilde{y}}\overline{\mathbf{v}}^{-1}\right).\end{aligned} \tag{3.82}$$

This suggests that we should write the (a,b)th element of $\hat{\mathbf{Q}}$, given $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{\Delta}}$, as:

$$\hat{Q}_{ab} = \sum_{j=1}^{J_L} \mathrm{tr}_j\left(\mathbf{N}_{a\hat{\tilde{y}}}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}_j\left(\mathbf{N}_{b\hat{\tilde{y}}}\overline{\mathbf{v}}^{-1}\right), \tag{3.83}$$

where tr_j indicates that only the *j*th blocks of the matrices in the product are to be used in computing the trace and we use $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$ throughout. The 'adjusted sandwich' estimator $\hat{\boldsymbol{\Lambda}}$ for the covariance of $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{\Delta}}$ follows from equation (3.80).

The formation of the expression in (3.83) entailed separating the factors in the summand of $\left[\sum_{j=1}^{J_L}\left(\mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\right]$, premultiplying the first factor by $\mathbf{X}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}$ and postmultiplying the second by $\overline{\mathbf{v}}^{-1}\mathbf{X}$, adjusting the resulting products for measurement error, and then recombining. It could be argued that separating the factors in this way is not appropriate for a sandwich estimator. If we followed this line of argument then no adjustment to **P** would be admissible, for it is not appropriate to remove error variances and covariances either from $\left(\mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}$ or from $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, since these adjustments were not made in the formation of $\hat{\boldsymbol{\beta}}$. In that case we should have

$$\hat{\mathbf{Q}} = \mathbf{X}^{\mathrm{T}}\overline{\mathbf{V}}^{-1}\left[\sum_{j=1}^{J_L}\left(\mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\right]\overline{\mathbf{V}}^{-1}\mathbf{X}, \tag{3.84}$$

$$\hat{Q}_{ab} = \sum_{j=1}^{J_L}\mathrm{tr}_j\left(\mathbf{X}_a\hat{\tilde{\mathbf{Y}}}^{\mathrm{T}}\overline{\mathbf{V}}^{-1}\right)\mathrm{tr}_j\left(\mathbf{X}_b\hat{\tilde{\mathbf{Y}}}^{\mathrm{T}}\overline{\mathbf{V}}^{-1}\right),$$

but this is conditioning on the *observed* values.

A more substantial objection to the estimator $\hat{Q}_{ab}$ in (3.83) is that the use of $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$ introduces bias. With no error in **X** and no error in **Y** the expectation in (3.77), for which we seek an estimator based on raw residuals, may be written:

$$\begin{aligned}\mathrm{E}\left(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}}\right) &= \mathrm{E}\left[\left\{\hat{\tilde{\mathbf{Y}}} + \mathbf{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right\}\left\{\hat{\tilde{\mathbf{Y}}} + \mathbf{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right\}^{\mathrm{T}}\right] \\ &= \mathrm{E}\left(\hat{\tilde{\mathbf{Y}}}\hat{\tilde{\mathbf{Y}}}^{\mathrm{T}}\right) + \mathrm{E}\left[\mathbf{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right].\end{aligned} \tag{3.85}$$

Given an existing sandwich estimator $\hat{\boldsymbol{\Lambda}} \equiv \widehat{\mathrm{cov}}\left(\hat{\boldsymbol{\beta}}\right)$, initially zero, we should form a new estimator by replacing this expectation in (3.77) by:

$$\sum_{j=1}^{J_L}\left\{\hat{\tilde{\mathbf{Y}}}_{(j)}\hat{\tilde{\mathbf{Y}}}_{(j)}^{\mathrm{T}} + \mathbf{X}_{(j)}\hat{\boldsymbol{\Lambda}}\mathbf{X}_{(j)}^{\mathrm{T}}\right\}. \tag{3.86}$$

When there is measurement error in **X** or **Y** we have:

$$\mathbf{x}_a^{\mathrm{T}}\overline{\mathbf{v}}^{-1}(\mathbf{y}-\mathbf{x}\beta)=\mathbf{x}_a^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\left\{\mathbf{y}-\mathbf{x}\hat{\beta}+\mathbf{x}\left(\hat{\beta}-\beta\right)\right\} \tag{3.87}$$
$$\approx \mathrm{tr}\left(\mathbf{N}_{a\hat{\tilde{y}}}\overline{\mathbf{v}}^{-1}\right)+\sum_{r=0}^{p-1}\mathrm{tr}\left(\mathbf{N}_{ar}\overline{\mathbf{v}}^{-1}\right)\left(\hat{\beta}_r-\beta_r\right),$$

where the error remaining in the right-hand side of (3.87) is $\xi_a^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\tilde{\mathbf{y}}+\mathbf{x}_a^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\lambda$.

This suggests that we should write, in place of (3.83):

$$\hat{Q}_{ab}=\sum_{j=1}^{J_L}\left\{\mathrm{tr}_j\left(\mathbf{N}_{a\hat{\tilde{y}}}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}_j\left(\mathbf{N}_{b\hat{\tilde{y}}}\overline{\mathbf{v}}^{-1}\right)+\sum_{r,s=0}^{p-1}\hat{\Lambda}_{rs}\,\mathrm{tr}_j\left(\mathbf{N}_{ar}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}_j\left(\mathbf{N}_{bs}\overline{\mathbf{v}}^{-1}\right)\right\}, \tag{3.88}$$

where $\hat{\Lambda}_{rs}$ is the current estimator, initially set to zero, of $\mathrm{cov}\left(\hat{\beta}_r,\hat{\beta}_s\right)$. Equation (3.80) then yields a new estimator $\hat{\Lambda}$ by iteration.

3.7.3 Further approximations

We now return to the issue of possible bias in the estimator $\hat{\beta}=\hat{\Gamma}^{-1}\hat{\Delta}$ as defined in (3.74)–(3.75). The magnitude of the bias is clearly dependent on prior assumptions about the measurement error variances and covariances, and simulations have shown that there are cases where the measurement error is substantial and the relative bias in $\hat{\beta}=\hat{\Gamma}^{-1}\hat{\Delta}$ is negligible. See Chapter 4. We outline here a possible further adjustment.

We may write

$$\gamma\equiv \mathrm{E}\left(\hat{\Gamma}\right)=\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x},\ \ \tilde{\gamma}\equiv\hat{\Gamma}-\gamma,\ \ \tilde{\gamma}_{ab}=\mathrm{tr}\left[\mathbf{x}_a\xi_b^{\mathrm{T}}+\xi_a\mathbf{x}_b^{\mathrm{T}}+\xi_a\xi_b^{\mathrm{T}}-\mathbf{M}_{ab}\right]\overline{\mathbf{v}}^{-1}, \tag{3.89}$$
$$\delta\equiv \mathrm{E}\left(\hat{\Delta}\right)=\mathbf{x}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{y},\ \ \tilde{\delta}\equiv\hat{\Delta}-\delta,\ \ \tilde{\delta}_a=\mathrm{tr}\left[\mathbf{x}_a\eta^{\mathrm{T}}+\xi_a\mathbf{y}^{\mathrm{T}}+\xi_a\eta^{\mathrm{T}}-\mathbf{M}_{a\eta}\right]\overline{\mathbf{v}}^{-1},$$

and we have:

$$
\begin{aligned}
&\mathrm{E}\left(\hat{\Gamma}^{-1}\hat{\Delta}\right) \\
&= \gamma^{-1}\delta - \mathrm{E}\left(\gamma^{-1}\tilde{\gamma}\gamma^{-1}\tilde{\delta}\right) + \mathrm{E}\left(\gamma^{-1}\left(\tilde{\gamma}\gamma^{-1}\right)^2\left(\delta+\tilde{\delta}\right)\right) - \mathrm{E}\left(\gamma^{-1}\left(\tilde{\gamma}\gamma^{-1}\right)^3\left(\delta+\tilde{\delta}\right)\right) + \mathrm{E}\left(\gamma^{-1}\left(\tilde{\gamma}\gamma^{-1}\right)^4\delta\right) \\
&\quad + O\left(m^6u^{-6}\right),
\end{aligned}
\tag{3.90}
$$

where m^6 indicates a product of moments of the measurement errors of total order 6 and u^{-6} is the reciprocal of a 6th-degree product of values of $\mathbf{x}$, $\mathbf{y}$, or $\mathbf{z}$. If, then, we can find estimators $\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \hat{\mathbf{E}}_3, \hat{\mathbf{E}}_4$ for the four expectations in (3.90), each with expected error of order m^6u^{-6}, we can form a second approximation

$$
\hat{\beta} = \hat{\Gamma}^{-1}\hat{\Delta} + \hat{\mathbf{E}}_1 - \hat{\mathbf{E}}_2 + \hat{\mathbf{E}}_3 - \hat{\mathbf{E}}_4, \tag{3.91}
$$

with expected error of the same order of magnitude. The assumption here is that this magnitude will be small in comparison to the bias in $\hat{\beta} = \hat{\Gamma}^{-1}\hat{\Delta}$, and this assumption would need to be tested in simulation.

Subject to this assumption, we illustrate the method by considering the contribution of the term $\mathrm{E}\left(\gamma^{-1}\tilde{\gamma}\gamma^{-1}\tilde{\delta}\right)$ to the element $\hat{\beta}_a$. The a-th element of $\mathrm{E}\left(\gamma^{-1}\tilde{\gamma}\gamma^{-1}\tilde{\delta}\right)$ is

$$
\sum_{r,s,t=0}^{p-1} \mathrm{E}\left(\bar{\gamma}_{ar}\tilde{\gamma}_{rs}\bar{\gamma}_{st}\tilde{\delta}_t\right), \tag{3.92}
$$

where $\bar{\gamma}_{ar}$, $\bar{\gamma}_{st}$ are respectively the (a,r)th and (s,t)th elements of γ^{-1}. We approximate this by

$$
\begin{aligned}
&\sum_{r,s,t=0}^{p-1} \mathrm{E}\left(\overline{\Gamma}_{ar}\widetilde{\gamma}_{rs}\overline{\Gamma}_{st}\widetilde{\delta}_t\right) \\
&= \sum_{r,s,t=0}^{p-1} \overline{\Gamma}_{ar}\overline{\Gamma}_{st}\, \mathrm{E}\begin{Bmatrix} \mathrm{tr}\left(\mathbf{x}_r\xi_s^{\mathrm{T}} + \xi_r\mathbf{x}_s^{\mathrm{T}} + \xi_r\xi_s^{\mathrm{T}} - \mathbf{M}_{rs}\right)\overline{\mathbf{v}}^{-1} \\ \times\,\mathrm{tr}\left(\mathbf{x}_t\eta^{\mathrm{T}} + \xi_t\mathbf{y}^{\mathrm{T}} + \xi_t\eta^{\mathrm{T}} - \mathbf{M}_{t\eta}\right)\overline{\mathbf{v}}^{-1} \end{Bmatrix} \\
&= \sum_{r,s,t=0}^{p-1} \overline{\Gamma}_{ar}\overline{\Gamma}_{st} \begin{Bmatrix} \mathrm{E}\,\mathrm{tr}\left(\mathbf{x}_t\mathbf{x}_r^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\xi_s\eta^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right) + \mathrm{E}\,\mathrm{tr}\left(\xi_t\xi_r^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x}_s\mathbf{y}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right) \\ +\mathrm{E}\,\mathrm{tr}\left(\mathbf{x}_t\mathbf{x}_s^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\xi_r\eta^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right) + \mathrm{E}\,\mathrm{tr}\left(\xi_t\xi_s^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{x}_r\mathbf{y}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right) \\ +\mathrm{E}\left[\mathrm{tr}\left(\xi_r\xi_s^{\mathrm{T}} - \mathbf{M}_{rs}\right)\overline{\mathbf{v}}^{-1}\,\mathrm{tr}\left(\xi_t\eta^{\mathrm{T}} - \mathbf{M}_{t\eta}\right)\overline{\mathbf{v}}^{-1}\right] \end{Bmatrix} \\
&\approx \sum_{r,s,t=0}^{p-1} \overline{\Gamma}_{ar}\overline{\Gamma}_{st} \begin{Bmatrix} \mathrm{tr}\left(\mathbf{N}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{M}_{s\eta}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{N}_{sy}\overline{\mathbf{v}}^{-1}\right) \\ +\mathrm{tr}\left(\mathbf{N}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{M}_{r\eta}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{N}_{ry}\overline{\mathbf{v}}^{-1}\right) \\ +\hat{\mathrm{E}}\left[\mathrm{tr}\left(\xi_r\xi_s^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\xi_t\eta^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right)\right] - \mathrm{tr}\left(\mathbf{M}_{rs}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{M}_{t\eta}\overline{\mathbf{v}}^{-1}\right) \end{Bmatrix} \\
&\approx \sum_{r,s,t=0}^{p-1} \overline{\Gamma}_{ar}\overline{\Gamma}_{st} \begin{Bmatrix} \mathrm{tr}\left(\mathbf{N}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{M}_{s\eta}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{N}_{sy}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{M}_{s\eta}\overline{\mathbf{v}}^{-1}\right) \\ +\mathrm{tr}\left(\mathbf{N}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{M}_{r\eta}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{N}_{ry}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{M}_{r\eta}\overline{\mathbf{v}}^{-1}\right) \end{Bmatrix},
\end{aligned}
\tag{3.93}
$$

where $\overline{\Gamma}_{ar}$, $\overline{\Gamma}_{st}$ are the corresponding elements of $\hat{\mathbf{\Gamma}}^{-1}$.

Now, using (3.91), the second approximation to β_a begins:

$$
\hat{\beta}_a = \sum_{r=0}^{p-1} \overline{\Gamma}_{ar}\hat{\Delta}_r + \sum_{r,s,t=0}^{p-1} \overline{\Gamma}_{ar}\overline{\Gamma}_{st} \begin{Bmatrix} \mathrm{tr}\left(\mathbf{N}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{M}_{s\eta}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{N}_{sy}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{tr}\overline{\mathbf{v}}^{-1}\mathbf{M}_{s\eta}\overline{\mathbf{v}}^{-1}\right) \\ +\mathrm{tr}\left(\mathbf{N}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{M}_{r\eta}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{N}_{ry}\overline{\mathbf{v}}^{-1}\right) + \mathrm{tr}\left(\mathbf{M}_{ts}\overline{\mathbf{v}}^{-1}\mathbf{M}_{r\eta}\overline{\mathbf{v}}^{-1}\right) \end{Bmatrix} - \ldots,
$$

.

(3.94)

where the second sum is an approximation to the a-th element of $\hat{\mathbf{E}}_1$. Since we are constrained to use elements of $\hat{\mathbf{\Gamma}}^{-1}$, $\hat{\mathbf{\Delta}}$ in the correction terms these still contain bias: a third approximation can remove such biases up to order m^4u^{-4}. For each such approximation there are corresponding adjustments to be made to the expression for $\hat{\mathbf{\Lambda}} \equiv \mathrm{c\hat{o}v}\left(\hat{\boldsymbol{\beta}}\middle|\mathbf{x},\overline{\mathbf{v}}\right)$.

3.7.4 Summary

In Section 3.7.2 we derived an approximate estimator $\hat{\boldsymbol{\beta}}$ of the fixed parameters of the multilevel model with measurement error, as defined in equations (3.18) to (3.26) in Section 3.5. We also derived a sandwich estimator $\hat{\Lambda}$ of $\text{cov}(\hat{\boldsymbol{\beta}})$, adjusted both for measurement error and for the sampling error in $\hat{\boldsymbol{\beta}}$. We repeat here, for convenience, the assumptions on which the method is based.

Assumption 3.1 Errors defined at a given level do not covary between units at (or above) that level.

Assumption 3.2 Errors at any level do not covary either with the true values of any of the variables or with any residuals.

Assumption 3.3 The errors ξ, η, ζ are distributed multivariate Normally with expectation $\mathbf{0}$.

Assumption 3.4 The random variables ε are distributed multivariate Normally with expectation $\mathbf{0}$, and are independent of $\mathbf{x}$, $\mathbf{z}$, and the errors ξ, η, ζ.

We further assume the existence of prior values of the error variances and covariances at all levels.

The usual GLS estimator $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$ for the fixed parameters (where $\mathbf{V}$ is the residual covariance matrix based on the observed random-part explanatory variables $\mathbf{Z}$) is biased as a result of the variances and covariances of the measurement errors in $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{V}$. We showed in Section 3.7.1 how to obtain an unbiased estimator $\hat{\mathbf{v}}$ of $\mathbf{V}$. In Section 3.6.5 we defined the unbiased estimators $\mathbf{N}_{rr'}$, $\mathbf{N}_{ry}$ of the cross-product matrices $\mathbf{x}_r\mathbf{x}_r^{\mathrm{T}}$, $\mathbf{x}_r\mathbf{y}^{\mathrm{T}}$, respectively, of true values of the response and fixed-part explanatory variables. But these unbiased estimators still contain error, and as a result the first approximation to $\hat{\boldsymbol{\beta}}$, based on equation (3.62), is biased. In Section 3.7.1 we went on to show how this bias can be removed by successive approximation subject to the assumptions above, and assuming also the existence of $\hat{\mathbf{v}}^{-1}$ and the convergence of the series expansion in (3.65).

That method, however, is complicated and lengthy, and we did not describe it in full. Instead we developed in Section 3.7.2 a simpler method based on the use of a weight matrix $\bar{\mathbf{v}}^{-1}$ containing no measurement error. There are several possible choices for $\bar{\mathbf{v}}$. For variance components models, and other models whose random-part explanatory variables contain no error, we may use $\bar{\mathbf{v}} = \hat{\mathbf{v}}$. For other models we may choose, for example, $\bar{\mathbf{v}} = \mathbf{I}_N$, the identity matrix, yielding OLS estimators, or we may choose to remove from $\hat{\mathbf{V}}$ (the estimator of $\mathbf{V}$ based on the current estimators of the random parameters and the observed values of the random-part explanatory variables) all terms involving variables with error. We term the resulting parameter estimators in this case 'purged-V' estimators. (In Section 4.3 we compare the results of these two strategies for the estimation of a 2-level model with a random coefficient on a variable with measurement error.)

This method, too, is not exact, and we may expect some biases in the resulting estimators. The most important of these is the bias in the sandwich estimator of the covariance matrix of the parameter estimators (and hence in their estimated standard errors) which results from using $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$ in the computation of the residuals on which the sandwich estimator is based. In Section 3.7.2 we gave details of an adjustment for this bias, and we shall demonstrate its effect in Section 4.2.

We indicated in Section 3.7.3 some further adjustments for the remaining biases in $\hat{\beta}$ and $\hat{\Lambda}$. We do not in this thesis develop these, but rely on the expressions (3.74), (3.75), (3.80), and (3.88).

3.8 Estimating the random parameters

3.8.1 Introduction

Adapting the notation of Goldstein (1986), for known $\boldsymbol{\beta}$ we write:

$$\begin{aligned} \tilde{\mathbf{y}} &\equiv \mathbf{y} - \mathbf{x}\boldsymbol{\beta}, \\ \mathbf{y}^{*} &\equiv \tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}, \\ \mathbf{y}^{**} &\equiv \operatorname{vec}\mathbf{y}^{*} = \tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}}. \end{aligned} \tag{3.95}$$

Thus $\mathbf{y}^{*}$ is the cross-product matrix of true residuals and we have $\mathrm{E}(\mathbf{y}^{*}) = \mathbf{v}$. We require to estimate the linear model

$$\mathbf{y}^{**} = \mathbf{z}^{**}\boldsymbol{\theta} + \mathbf{e}^{**}, \tag{3.96}$$

where $\mathbf{z}^{**}{}_{(N^2 \times H)}$ is the design matrix, based on the true values $\mathbf{z}$, for the random parameters, and $\mathbf{e}^{**}$ is an $N^2 \times 1$ vector of residuals.

In the notation of (3.21), the hth column $\mathbf{z}_h^{**}$ of $\mathbf{z}^{**}$ $(h = 1, 2, \ldots, H)$ is given by

$$\begin{aligned} \mathbf{z}_h^{**} &= \operatorname{vec}\mathbf{z}_h^{*}, \\ \mathbf{z}_h^{*} &= \frac{1}{2}\left(2 - \delta_{r_h s_h}\right) \bigoplus_{j=1}^{J_{t_h}} \left(\mathbf{z}_{r_h,j}\mathbf{z}_{s_h,j}^{\mathrm{T}} + \mathbf{z}_{s_h,j}\mathbf{z}_{r_h,j}^{\mathrm{T}}\right), \\ \delta_{r_h s_h} &= 1 \text{ if } r_h = s_h, \\ &\quad\ 0 \text{ otherwise}, \end{aligned} \tag{3.97}$$

and

$$\begin{aligned} \operatorname{cov}(\mathbf{e}^{**}) &= \mathrm{E}\left[(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})^{\mathrm{T}}\right] - \left[\mathrm{E}(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})\right]\left[\mathrm{E}(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})\right]^{\mathrm{T}} \\ &= \mathrm{E}\left(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}} \otimes \tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}\right) - \operatorname{vec}\mathbf{v}\left(\operatorname{vec}\mathbf{v}\right)^{\mathrm{T}} \\ &= \left(\mathbf{I} + \mathbf{s}_N\right)\left(\mathbf{v} \otimes \mathbf{v}\right), \end{aligned} \tag{3.98}$$

where $\mathbf{s}_N$ is the vec permutation matrix of order N^2. We give in the Appendix what we believe to be a new derivation of this latter result, which clarifies the role of the Normality assumptions.

Now the GLS estimator of $\boldsymbol{\theta}$ in (3.96) is

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \left(\mathbf{z}^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}\mathbf{z}^{**}\right)^{-1}\mathbf{z}^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}\mathbf{y}^{**} \\ &= \left(\mathbf{z}^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{z}^{**}\right)^{-1}\mathbf{z}^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{y}^{**}.\end{aligned} \tag{3.99}$$

The (h,h')th element of $\mathbf{z}^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{z}^{**}$, for each $h,h'\in\{1,2,\dots,H\}$, is given by:

$$\mathbf{z}_h^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{z}_{h'}^{**} = \left(\mathrm{vec}\,\mathbf{z}_h^{*}\right)^{\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\,\mathrm{vec}\,\mathbf{z}_{h'}^{*}, \tag{3.100}$$

where $\mathbf{z}_h^{*}, \mathbf{z}_{h'}^{*}$ are symmetric, block-diagonal $N\times N$ matrices as defined in (3.97). Therefore, using standard results (see, for example, Searle *et al.*, 1992, App. M.2 and M.9, and Searle, 1982, Section 12.9),

$$\mathbf{s}_N\,\mathrm{vec}\,\mathbf{z}_{h'}^{*} = \mathrm{vec}\,\mathbf{z}_{h'}^{*}\,, \tag{3.101}$$

so that

$$2\,\mathrm{vec}\,\mathbf{z}_{h'}^{*} = \left(\mathbf{I}+\mathbf{s}_N\right)\mathrm{vec}\,\mathbf{z}_{h'}^{*}, \tag{3.102}$$

and

$$\begin{aligned}\mathbf{z}_h^{**\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{z}_{h'}^{**} &= \left(\mathrm{vec}\,\mathbf{z}_h^{*}\right)^{\mathrm{T}}(\mathbf{v}\otimes\mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}(\mathbf{I}+\mathbf{s}_N)\,\mathrm{vec}\,\mathbf{z}_{h'}^{*} \\ &= \left(\mathrm{vec}\,\mathbf{z}_h^{*}\right)^{\mathrm{T}}\left(\mathbf{v}^{-1}\otimes\mathbf{v}^{-1}\right)\mathrm{vec}\,\mathbf{z}_{h'}^{*} \\ &= \mathrm{tr}\left(\mathbf{z}_h^{*}\mathbf{v}^{-1}\mathbf{z}_{h'}^{*}\mathbf{v}^{-1}\right).\end{aligned} \tag{3.103}$$

Similarly, the hth element of $\mathbf{z}^{**\mathrm{T}}(\mathbf{v} \otimes \mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{y}^{**}$ is

$$\mathbf{z}_h^{**\mathrm{T}}(\mathbf{v} \otimes \mathbf{v})^{-1}(\mathbf{I}+\mathbf{s}_N)^{-1}2\mathbf{y}^{**} = \mathrm{tr}\left(\mathbf{z}_h^* \mathbf{v}^{-1}\mathbf{y}^* \mathbf{v}^{-1}\right). \tag{3.104}$$

We do not know $\mathbf{z}^*, \mathbf{y}^*$, or $\mathbf{v}$ and, in similarity with the case of the fixed parameters, we seek matrices $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$ such that to a sufficiently close approximation

$$\begin{aligned} \mathrm{tr}\left(\mathbf{z}_h^* \mathbf{v}^{-1}\mathbf{z}_{h'}^* \mathbf{v}^{-1}\right) &= \mathrm{E}\left(\hat{\Phi}_{hh'}\right), \\ \mathrm{tr}\left(\mathbf{z}_h^* \mathbf{v}^{-1}\mathbf{y}^* \mathbf{v}^{-1}\right) &= \mathrm{E}\left(\hat{\Psi}_h\right). \end{aligned} \tag{3.105}$$

A first approximation to $\boldsymbol{\theta}$ will then be the estimator

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Phi}}^{-1}\hat{\boldsymbol{\Psi}}. \tag{3.106}$$

In Section 3.8.2 we derive approximations for $\hat{\Phi}_{hh'}$, $\hat{\Psi}_h$ based on the use of a weight matrix $\overline{\mathbf{v}}^{-1}$ in place of $\mathbf{v}^{-1}$, where $\overline{\mathbf{v}}^{-1}$ contains no term with measurement error.

3.8.2 Estimation of $\boldsymbol{\theta}$ with an error-free weight matrix

Consider $\mathrm{tr}\left(\mathbf{z}_h^* \mathbf{v}^{-1}\mathbf{z}_{h'}^* \mathbf{v}^{-1}\right)$ first. The first point to note is that by the definition in (3.97) $\mathbf{z}_h^*$ and $\mathbf{z}_{h'}^*$ are block-diagonal at levels ℓ_h and $\ell_{h'}$ respectively. Suppose $\theta_h = \sigma_{ab}^{(\ell_h)}$, $\theta_{h'} = \sigma_{cd}^{(\ell_{h'})}$, for some $a,b,c,d \in \{p, p+1, \ldots, p+q-1\}$ not necessarily distinct, and use $j_h, j_{h'}$ to index units at levels $\ell_h, \ell_{h'}$, respectively, with $1 \le j_h \le J_h, 1 \le j_{h'} \le J_{h'}$. Then

$$\begin{aligned} &\mathrm{tr}\left(\mathbf{z}_h^* \mathbf{v}^{-1}\mathbf{z}_{h'}^* \mathbf{v}^{-1}\right) \\ &= \frac{1}{4}(2-\delta_{ab})(2-\delta_{cd})\sum_{j_h=1}^{J_h}\sum_{j_{h'}=1}^{J_{h'}} \mathrm{tr}\left[\left(\mathbf{z}_{a(j_h)}\mathbf{z}_{b(j_h)}^{\mathrm{T}} + \mathbf{z}_{b(j_h)}\mathbf{z}_{a(j_h)}^{\mathrm{T}}\right)\mathbf{v}^{-1}\left(\mathbf{z}_{c(j_{h'})}\mathbf{z}_{d(j_{h'})}^{\mathrm{T}} + \mathbf{z}_{d(j_{h'})}\mathbf{z}_{c(j_{h'})}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right], \end{aligned} \tag{3.107}$$

where $\mathbf{z}_{a(j_h)}$, for example, is the $N \times 1$ vector whose j_hth level-ℓ_h block is equal to $\mathbf{z}_{a,j_h}$ and which is zero elsewhere.

Now write:

$$\breve{a} \equiv a(j_h),\;\; \breve{b} \equiv b(j_h),\;\; \breve{c} \equiv c(j_{h'}),\;\; \breve{d} \equiv d(j_{h'}). \tag{3.108}$$

The trace on the right-hand side of (3.107) becomes

$$\operatorname{tr}\left[\left(\mathbf{z}_{\breve{a}}\mathbf{z}_{\breve{b}}^{\mathrm{T}} + \mathbf{z}_{\breve{b}}\mathbf{z}_{\breve{a}}^{\mathrm{T}}\right)\mathbf{v}^{-1}\left(\mathbf{z}_{\breve{c}}\mathbf{z}_{\breve{d}}^{\mathrm{T}} + \mathbf{z}_{\breve{d}}\mathbf{z}_{\breve{c}}^{\mathrm{T}}\right)\mathbf{v}^{-1}\right] \tag{3.109}$$

and we consider the term $\operatorname{tr}\left(\mathbf{z}_{\breve{a}}\mathbf{z}_{\breve{b}}^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{c}}\mathbf{z}_{\breve{d}}^{\mathrm{T}}\mathbf{v}^{-1}\right)$. As a first approximation we have:

$$\begin{aligned}\operatorname{tr}\left(\mathbf{z}_{\breve{a}}\mathbf{z}_{\breve{b}}^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{c}}\mathbf{z}_{\breve{d}}^{\mathrm{T}}\mathbf{v}^{-1}\right) &= \operatorname{tr}\left(\mathbf{z}_{\breve{b}}^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{c}}\right)\operatorname{tr}\left(\mathbf{z}_{\breve{a}}^{\mathrm{T}}\mathbf{v}^{-1}\mathbf{z}_{\breve{d}}\right)\\ &\approx \operatorname{tr}\left(\mathbf{N}_{\breve{b}\breve{c}}\overline{\mathbf{v}}^{-1}\right)\operatorname{tr}\left(\mathbf{N}_{\breve{a}\breve{d}}\overline{\mathbf{v}}^{-1}\right).\end{aligned} \tag{3.110}$$

Using the Normality assumption,

$$\begin{aligned}&\mathrm{E}\left[\operatorname{tr}\left(\mathbf{N}_{\breve{b}\breve{c}}\overline{\mathbf{v}}^{-1}\right)\operatorname{tr}\left(\mathbf{N}_{\breve{a}\breve{d}}\overline{\mathbf{v}}^{-1}\right)\right]\\ &= \mathrm{E}\left\{\operatorname{tr}\left[\left(\mathbf{z}_{\breve{b}}\mathbf{z}_{\breve{c}}^{\mathrm{T}} + \mathbf{z}_{b}\zeta_{\breve{c}}^{\mathrm{T}} + \zeta_{\breve{b}}\mathbf{z}_{\breve{c}}^{\mathrm{T}} + \zeta_{\breve{b}}\zeta_{\breve{c}}^{\mathrm{T}} - \mathbf{M}_{\breve{b}\breve{c}}\right)\overline{\mathbf{v}}^{-1}\right]\operatorname{tr}\left[\left(\mathbf{z}_{\breve{a}}\mathbf{z}_{\breve{d}}^{\mathrm{T}} + \mathbf{z}_{\breve{a}}\zeta_{\breve{d}}^{\mathrm{T}} + \zeta_{\breve{a}}\mathbf{z}_{\breve{d}}^{\mathrm{T}} + \zeta_{\breve{a}}\zeta_{\breve{d}}^{\mathrm{T}} - \mathbf{M}_{\breve{a}\breve{d}}\right)\overline{\mathbf{v}}^{-1}\right]\right\}\\ &= \operatorname{tr}\left(\mathbf{z}_{a}\mathbf{z}_{\breve{b}}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{z}_{\breve{c}}\mathbf{z}_{d}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right)\\ &\quad + \mathrm{E}\left\{\operatorname{tr}\left[\left(\mathbf{z}_{\breve{b}}\zeta_{\breve{c}}^{\mathrm{T}} + \zeta_{\breve{b}}\mathbf{z}_{\breve{c}}^{\mathrm{T}}\right)\overline{\mathbf{v}}^{-1}\right]\operatorname{tr}\left[\left(\mathbf{z}_{\breve{a}}\zeta_{\breve{d}}^{\mathrm{T}} + \zeta_{\breve{a}}\mathbf{z}_{\breve{d}}^{\mathrm{T}}\right)\overline{\mathbf{v}}^{-1}\right]\right\}\\ &\quad + \mathrm{E}\left\{\operatorname{tr}\left[\left(\zeta_{\breve{b}}\zeta_{\breve{c}}^{\mathrm{T}} - \mathbf{M}_{\breve{b}\breve{c}}\right)\overline{\mathbf{v}}^{-1}\right]\operatorname{tr}\left[\left(\zeta_{\breve{a}}\zeta_{\breve{d}}^{\mathrm{T}} - \mathbf{M}_{\breve{a}\breve{d}}\right)\overline{\mathbf{v}}^{-1}\right]\right\}.\end{aligned}$$

$$\tag{3.111}$$

This yields the second approximation:

$$
\begin{aligned}
&\mathrm{tr}\left(\mathbf{z}_{\breve{a}}\mathbf{z}_{\breve{b}}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\mathbf{z}_{\breve{c}}\mathbf{z}_{\breve{d}}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right) \\
&= \mathrm{E}\left[\mathrm{tr}\left(\mathbf{N}_{\breve{b}\breve{c}}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{N}_{a\breve{d}}\overline{\mathbf{v}}^{-1}\right)\right] \\
&\quad -\mathrm{E}\left\{\mathrm{tr}\left[\left(\mathbf{z}_{\breve{b}}\zeta_{\breve{c}}^{\mathrm{T}}+\zeta_{\breve{b}}\mathbf{z}_{\breve{c}}^{\mathrm{T}}\right)\overline{\mathbf{v}}^{-1}\right]\mathrm{tr}\left[\left(\mathbf{z}_{\breve{a}}\zeta_{\breve{d}}^{\mathrm{T}}+\zeta_{\breve{a}}\mathbf{z}_{\breve{d}}^{\mathrm{T}}\right)\overline{\mathbf{v}}^{-1}\right]\right\} \\
&\quad -\mathrm{E}\left\{\mathrm{tr}\left[\left(\zeta_{\breve{b}}\zeta_{\breve{c}}^{\mathrm{T}}-\mathbf{M}_{\breve{b}\breve{c}}\right)\overline{\mathbf{v}}^{-1}\right]\mathrm{tr}\left[\left(\zeta_{a}\zeta_{\breve{d}}^{\mathrm{T}}-\mathbf{M}_{\breve{a}\breve{d}}\right)\overline{\mathbf{v}}^{-1}\right]\right\} \\
&\approx \mathrm{tr}\left(\mathbf{N}_{\breve{b}\breve{c}}\overline{\mathbf{v}}^{-1}\right)\mathrm{tr}\left(\mathbf{N}_{\breve{a}\breve{d}}\overline{\mathbf{v}}^{-1}\right) \\
&\quad -\mathrm{tr}\left(\mathbf{N}_{\breve{a}\breve{b}}\overline{\mathbf{v}}^{-1}\mathbf{M}_{\breve{c}\breve{d}}\overline{\mathbf{v}}^{-1}\right)-\mathrm{tr}\left(\mathbf{M}_{\breve{a}\breve{b}}\overline{\mathbf{v}}^{-1}\mathbf{N}_{\breve{c}\breve{d}}\overline{\mathbf{v}}^{-1}\right) \\
&\quad -\mathrm{tr}\left(\mathbf{N}_{\breve{a}\breve{c}}\overline{\mathbf{v}}^{-1}\mathbf{M}_{\breve{b}\breve{d}}\overline{\mathbf{v}}^{-1}\right)-\mathrm{tr}\left(\mathbf{M}_{\breve{a}\breve{c}}\overline{\mathbf{v}}^{-1}\mathbf{N}_{\breve{b}\breve{d}}\overline{\mathbf{v}}^{-1}\right) \\
&\quad -\mathrm{tr}\left(\mathbf{M}_{\breve{a}\breve{b}}\overline{\mathbf{v}}^{-1}\mathbf{M}_{\breve{c}\breve{d}}\overline{\mathbf{v}}^{-1}\right)-\mathrm{tr}\left(\mathbf{M}_{\breve{a}\breve{c}}\overline{\mathbf{v}}^{-1}\mathbf{M}_{\breve{b}\breve{d}}\overline{\mathbf{v}}^{-1}\right).
\end{aligned}
\tag{3.112}
$$

We write the sum of these traces over all sub-block combinations (as implied by the double sum in equation 3.107) as:

$$
\begin{aligned}
\hat{F}_{abcd} \equiv\ & \mathbf{N}_{ad} \times \mathbf{N}_{bc} \\
& - \mathbf{M}_{ac} \cdot \mathbf{N}_{bd} - \mathbf{M}_{bd} \cdot \mathbf{N}_{ac} - \mathbf{M}_{cd} \cdot \mathbf{N}_{ab} - \mathbf{M}_{ab} \cdot \mathbf{N}_{cd} \\
& - \mathbf{M}_{ac} \cdot \mathbf{M}_{bd} - \mathbf{M}_{ab} \cdot \mathbf{M}_{cd}.
\end{aligned}
\tag{3.113}
$$

Each **N** and **M** in (3.113) is assumed to be post-multiplied by $\overline{\mathbf{v}}^{-1}$. We use the 'dot' product to indicate that the resulting matrices are to be multiplied before extraction of the trace, and the 'cross' product to indicate that the traces are extracted before multiplication. The products are understood to be computed for each sub-block at level ℓ_h (for a and b) and level $\ell_{h'}$ (for c and d) and then summed over all combinations of these. In practice, because of the block-diagonal structure of $\overline{\mathbf{v}}^{-1}$, the only non-zero products in the sum are those for sub-blocks which share a level-L block.

We now define

$$\hat{\Phi}_{hh'} \equiv \frac{1}{4}(2-\delta_{ab})(2-\delta_{cd})\left(\hat{F}_{abcd}+\hat{F}_{abdc}+\hat{F}_{bacd}+\hat{F}_{badc}\right). \qquad (3.114)$$

Given an estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ we define $\hat{F}_{ab\hat{\tilde{\mathbf{y}}}\hat{\tilde{\mathbf{y}}}}$ in the obvious way, by substituting $\hat{\tilde{\mathbf{y}}}$ for c or d whenever either of these occurs as a subscript of N, and substituting $\hat{\boldsymbol{\lambda}}$ for c or d when either occurs as a subscript of **M**, in the right-hand side of (3.113), with implied blocking at level L. We now define

$$\hat{\Psi}_h \equiv \frac{1}{2}(2-\delta_{ab})\left(\hat{F}_{ab\hat{\tilde{\mathbf{y}}}\hat{\tilde{\mathbf{y}}}}+\hat{F}_{ba\hat{\tilde{\mathbf{y}}}\hat{\tilde{\mathbf{y}}}}\right) \qquad (3.115)$$

and an estimator of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\Phi}}^{-1}\hat{\boldsymbol{\Psi}}$.

This estimator is in theory biased on account of the correlation of the measurement errors still present in $\hat{\boldsymbol{\Phi}}^{-1}$ and $\hat{\boldsymbol{\Psi}}$ and we may adjust for these biases in a manner analogous to that indicated in Section 3.7.3. It is biased also because the use of $\hat{\boldsymbol{\beta}}$ in (3.115) introduces variances and covariances of the sampling errors in $\hat{\boldsymbol{\beta}}$. A method for removing this latter bias known as *restricted* unbiased iterative generalised least squares (or RIGLS) estimation has been described by Goldstein (1989, and 1995, p40), and we now adapt it to the measurement error case.

We require that $\hat{\Psi}_h$ should be an unbiased estimator of $\frac{1}{2}(2-\delta_{ab})\,\mathrm{tr}\left[\left(\mathbf{z}_a\mathbf{z}_b^{\mathrm{T}}+\mathbf{z}_b\mathbf{z}_a^{\mathrm{T}}\right)\overline{\mathbf{V}}^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{\mathrm{T}}\overline{\mathbf{V}}^{-1}\right]$. If the true values $\mathbf{x}$ and $\mathbf{z}$ are known and fixed in repeated sampling and $\mathbf{Y}=\mathbf{y}$ is measured without error then

$$\mathrm{E}\left[\left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}|\mathbf{x},\bar{\mathbf{v}}\right]+\mathbf{x}\,\mathrm{cov}\left(\hat{\boldsymbol{\beta}}|\mathbf{x},\bar{\mathbf{v}}\right)\mathbf{x}^{\mathrm{T}}$$
$$=\mathrm{E}\left[\left\{\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}+\mathbf{x}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\}\left\{\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}+\mathbf{x}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\}^{\mathrm{T}}|\mathbf{x},\bar{\mathbf{v}}\right]$$
$$=(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{\mathrm{T}}. \tag{3.116}$$

Thus, for the case with measurement error, we should add to the expression for $\hat{\Psi}_h$ in (3.115) an expression which approximates

$$\frac{1}{2}\left(2-\delta_{ab}\right)\mathrm{tr}\left[\left(\mathbf{z}_a\mathbf{z}_b^{\mathrm{T}}+\mathbf{z}_b\mathbf{z}_a^{\mathrm{T}}\right)\bar{\mathbf{v}}^{-1}\mathbf{x}\,\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x}^{\mathrm{T}}\bar{\mathbf{v}}^{-1}\right],\ \text{or} \tag{3.117}$$
$$\frac{1}{2}\left(2-\delta_{ab}\right)\mathrm{tr}\left[\left(\mathbf{z}_a\mathbf{z}_b^{\mathrm{T}}+\mathbf{z}_b\mathbf{z}_a^{\mathrm{T}}\right)\bar{\mathbf{v}}^{-1}\left(\sum_{r,s=0}^{p-1}\hat{\Lambda}_{rs}\mathbf{x}_r\mathbf{x}_s^{\mathrm{T}}\right)\bar{\mathbf{v}}^{-1}\right].$$

The adjusted estimator for Ψ_h becomes

$$\hat{\Psi}_h=\frac{1}{2}\left(2-\delta_{ab}\right)\left\{\hat{F}_{ab\hat{y}\hat{y}}+\hat{F}_{ba\hat{y}\hat{y}}+\sum_{r,s=0}^{p-1}\hat{\Lambda}_{rs}\left(\hat{F}_{abrs}+\hat{F}_{bars}\right)\right\}, \tag{3.118}$$

and now the product $\hat{\boldsymbol{\Phi}}^{-1}\hat{\boldsymbol{\Psi}}$ yields a RIGLS estimator for $\boldsymbol{\theta}$.

Turning to the covariance matrix of $\hat{\boldsymbol{\theta}}$, we consider IGLS estimation first. If there is no error in **X**, **Y**, or **Z**, then equation (3.99) yields

$$\hat{\boldsymbol{\theta}}=\left(\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathbf{Z}^{**}\right)^{-1}\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathbf{Y}^{**},$$
$$\mathrm{cov}\left(\hat{\boldsymbol{\theta}}|\mathbf{X},\mathbf{Z},\bar{\mathbf{v}}\right)=\left(\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathbf{Z}^{**}\right)^{-1}$$
$$\times\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\,\mathrm{E}\left[\left(\mathbf{Y}^{**}-\mathbf{Z}^{**}\boldsymbol{\theta}\right)\left(\mathbf{Y}^{**}-\mathbf{Z}^{**}\boldsymbol{\theta}\right)^{\mathrm{T}}|\mathbf{X},\mathbf{Z},\bar{\mathbf{v}}\right](\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathbf{Z}^{**}$$
$$\times\left(\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathbf{Z}^{**}\right)^{-1}, \tag{3.119}$$

and the sandwich estimator of $\mathrm{cov}(\hat{\theta})$ is formed by replacing the expectation $\mathrm{E}\left[(\mathbf{Y}^{**}-\mathbf{Z}^{**}\theta)(\mathbf{Y}^{**}-\mathbf{Z}^{**}\theta)^{\mathrm{T}}\right]$ by a matrix of the same form using the raw residuals $\mathbf{Y}^{**}-\mathbf{Z}^{**}\hat{\theta}$ of the linear model for the random parameters.

The expectation $\mathrm{E}\left[(\mathbf{Y}^{**}-\mathbf{Z}^{**}\theta)(\mathbf{Y}^{**}-\mathbf{Z}^{**}\theta)^{\mathrm{T}}\right]$ is equal to $(\mathbf{I}+\mathbf{s}_N)(\mathbf{V}\otimes\mathbf{V})$. A matrix of the same form may be constructed as the sum of two cross-product matrices:

$$\sum_{j=1}^{J_L}\left[\mathrm{vec}\left(\hat{\mathbf{Y}}_{(j)}\hat{\mathbf{Y}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\right]\left[\mathrm{vec}\left(\hat{\mathbf{Y}}_{(j)}\hat{\mathbf{Y}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\right]^{\mathrm{T}} \tag{3.120}$$

and

$$\sum_{j=1}^{J_L-1}\sum_{j'=j+1}^{J_L}\left[\mathrm{vec}\left(\hat{\mathbf{Y}}_{(j)}\hat{\mathbf{Y}}_{(j')}^{\mathrm{T}}+\hat{\mathbf{Y}}_{(j')}\hat{\mathbf{Y}}_{(j)}^{\mathrm{T}}\right)\right]\left[\mathrm{vec}\left(\hat{\mathbf{Y}}_{(j)}\hat{\mathbf{Y}}_{(j')}^{\mathrm{T}}+\hat{\mathbf{Y}}_{(j')}\hat{\mathbf{Y}}_{(j)}^{\mathrm{T}}\right)\right]^{\mathrm{T}}. \tag{3.121}$$

The second of these arises from terms such as $\hat{Y}_{i_1}\hat{Y}_{i_2}$, $\hat{Y}_{i_2}\hat{Y}_{i_1}$, where i_1, i_2 represent level-1 units in different level-L blocks. Such terms, though zero in expectation, still have variances and a covariance. Neglecting sampling error in $\hat{\mathbf{V}}$, we replace the product

$$\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathrm{E}\left[(\mathbf{Y}^{**}-\mathbf{Z}^{**}\theta)(\mathbf{Y}^{**}-\mathbf{Z}^{**}\theta)^{\mathrm{T}}|\mathbf{X},\mathbf{Z},\bar{\mathbf{v}}\right](\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\mathbf{Z}^{**}$$

in (3.119) by the $H\times H$ matrix

$$\mathbf{R} \equiv \sum_{j=1}^{J_L}\left[\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\right]\left[\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\right]^{\mathrm{T}} \tag{3.122}$$
$$+\sum_{j=1}^{J_L-1}\sum_{j'=j+1}^{J_L}\left\{\begin{array}{l}\left[\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j')}^{\mathrm{T}}+\hat{\bar{\mathbf{Y}}}_{(j')}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}\right)\right]\\ \times\left[\mathbf{Z}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j')}^{\mathrm{T}}+\hat{\bar{\mathbf{Y}}}_{(j')}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}\right)\right]^{\mathrm{T}}\end{array}\right\},$$

and for each $h,h' \in \{1,2,\ldots,H\}$ we have

$$\begin{aligned}
R_{hh'} &= \sum_{j=1}^{J_L}\left[\mathbf{Z}_h^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\right]\left[\mathbf{Z}_{h'}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\right]^{\mathrm{T}}\\
&\quad+\sum_{j=1}^{J_L-1}\sum_{j'=j+1}^{J_L}\left\{\begin{array}{l}\left[\mathbf{Z}_h^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j')}^{\mathrm{T}}+\hat{\bar{\mathbf{Y}}}_{(j')}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}\right)\right]\\ \times\left[\mathbf{Z}_{h'}^{**\mathrm{T}}(\bar{\mathbf{v}}\otimes\bar{\mathbf{v}})^{-1}\operatorname{vec}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j')}^{\mathrm{T}}+\hat{\bar{\mathbf{Y}}}_{(j')}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}\right)\right]^{\mathrm{T}}\end{array}\right\}\\
&= \sum_{j=1}^{J_L}\operatorname{tr}\left[\mathbf{Z}_h^{*}\bar{\mathbf{v}}^{-1}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\bar{\mathbf{v}}^{-1}\right]\operatorname{tr}\left[\mathbf{Z}_{h'}^{*}\bar{\mathbf{v}}^{-1}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\bar{\mathbf{v}}^{-1}\right]\\
&\quad+\sum_{j=1}^{J_L-1}\sum_{j'=j+1}^{J_L}\operatorname{tr}\left[\mathbf{Z}_h^{*}\bar{\mathbf{v}}^{-1}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j')}^{\mathrm{T}}+\hat{\bar{\mathbf{Y}}}_{(j')}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}\right)\bar{\mathbf{v}}^{-1}\right]\operatorname{tr}\left[\mathbf{Z}_{h'}^{*}\bar{\mathbf{v}}^{-1}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j')}^{\mathrm{T}}+\hat{\bar{\mathbf{Y}}}_{(j')}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}\right)\bar{\mathbf{v}}^{-1}\right]\\
&= \sum_{j=1}^{J_L}\operatorname{tr}\left[\mathbf{Z}_h^{*}\bar{\mathbf{v}}^{-1}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\bar{\mathbf{v}}^{-1}\right]\operatorname{tr}\left[\mathbf{Z}_{h'}^{*}\bar{\mathbf{v}}^{-1}\left(\hat{\bar{\mathbf{Y}}}_{(j)}\hat{\bar{\mathbf{Y}}}_{(j)}^{\mathrm{T}}-\hat{\mathbf{V}}_{(jj)}\right)\bar{\mathbf{v}}^{-1}\right],
\end{aligned} \tag{3.123}$$

since $\mathbf{Z}_h^{*}$, $\mathbf{Z}_{h'}^{*}$, and $\bar{\mathbf{v}}^{-1}$ are block-diagonal.

Adopting the approach used for the fixed parameters in Section 3.7.2, we seek a matrix $\hat{\mathbf{S}}$, based on $\mathbf{R}$ and suitably adjusted, so that we can write

$$\hat{\Pi} \equiv \hat{\operatorname{cov}}\left(\hat{\theta}\,|\,\mathbf{x},\mathbf{z},\bar{\mathbf{v}}\right) = \hat{\Phi}^{-1}\hat{\mathbf{S}}\hat{\Phi}^{-1}. \tag{3.124}$$

The final expression for $R_{hh'}$ in (3.123) is the sum of products of traces. It is appropriate to adjust each trace by removing measurement error variances and covariances since this is done in the estimation of $\boldsymbol{\theta}$. Errors that remain will in expectation generate variances and covariances when the traces are multiplied together, but these properly form part of the covariance matrix of $\hat{\boldsymbol{\theta}}$.

Consider, therefore, the expression

$$\sum_{j=1}^{J_L} \mathrm{tr}\left[\mathbf{z}_h^{*}\overline{\mathbf{v}}^{-1}\left(\hat{\tilde{\mathbf{y}}}_{(j)}\hat{\tilde{\mathbf{y}}}_{(j)}^{\mathrm{T}} - \hat{\mathbf{v}}_{(jj)}\right)\overline{\mathbf{v}}^{-1}\right]. \tag{3.125}$$

By definition,

$$\hat{\mathbf{v}}_{(jj)} = \sum_{\ell_\alpha=1}^{L} \sum_{j_\alpha : B_{j_\alpha}^{(\ell_\alpha)} \subset B_j^{(L)}} \sum_{s_\alpha, s'_\alpha = p}^{p+q-1} \hat{\sigma}_{s_\alpha s'_\alpha}^{(\ell_\alpha)} \mathbf{z}_{s_\alpha(j_\alpha)} \mathbf{z}_{s'_\alpha(j_\alpha)}^{\mathrm{T}}. \tag{3.126}$$

Using the notation of (3.108) and writing $\breve{p} \equiv s_\alpha(j_\alpha)$, $\breve{q} \equiv s'_\alpha(j_\alpha)$, we have:

$$\mathrm{tr}\left(\mathbf{z}_h^{*}\overline{\mathbf{v}}^{-1}\hat{\mathbf{v}}_{(jj)}\overline{\mathbf{v}}^{-1}\right) = \frac{1}{2}\left(2-\delta_{ab}\right)\sum_{\ell_\alpha=1}^{L} \sum_{j_\alpha : B_{j_\alpha}^{(\ell_\alpha)} \subset B_j^{(L)}} \sum_{s_\alpha, s'_\alpha = p}^{p+q\ 1} \hat{\sigma}_{s_\alpha s'_\alpha}^{(\ell_\alpha)} \mathrm{tr}\left[\left(\mathbf{z}_{\breve{a}}\mathbf{z}_{\breve{b}}^{\mathrm{T}} + \mathbf{z}_{\breve{b}}\mathbf{z}_{\breve{a}}^{\mathrm{T}}\right)\overline{\mathbf{v}}^{-1}\mathbf{z}_{\breve{p}}\mathbf{z}_{\breve{q}}^{\mathrm{T}}\overline{\mathbf{v}}^{-1}\right] \tag{3.127}$$

Now suppose that, for given $k, k' \in \{1, 2, \ldots, H\}$, $\theta_k \equiv \sigma_{r_k s_k}^{(\ell_k)}$, $\theta_{k'} \equiv \sigma_{r_{k'} s_{k'}}^{(\ell_{k'})}$. We define

$$\hat{G}_{abk} \equiv \frac{1}{2}\left(2-\delta_{r_k s_k}\right)\left(\hat{F}_{abr_k s_k} + \hat{F}_{abs_k r_k}\right), \tag{3.128}$$

$$\hat{G}_{cdk'} \equiv \frac{1}{2}\left(2-\delta_{r_{k'} s_{k'}}\right)\left(\hat{F}_{cdr_{k'} s_{k'}} + \hat{F}_{cds_{k'} r_{k'}}\right), \text{ etc.}$$

Then

$$\sum_{j=1}^{J_L} \mathrm{tr}\left[\mathbf{z}_h^{*}\overline{\mathbf{v}}^{-1}\left(\hat{\tilde{\mathbf{y}}}_{(j)}\hat{\tilde{\mathbf{y}}}_{(j)}^{\mathrm{T}} - \hat{\mathbf{v}}_{(jj)}\right)\overline{\mathbf{v}}^{-1}\right] \tag{3.129}$$
$$\approx \frac{1}{2}\left(2-\delta_{ab}\right)\left[\hat{F}_{ab\hat{\tilde{y}}\hat{\tilde{y}}} + \hat{F}_{ba\hat{\tilde{y}}\hat{\tilde{y}}} - \sum_{k=1}^{H}\hat{\theta}_k\left(\hat{G}_{abk} + \hat{G}_{bak}\right)\right].$$

We now write the (h,h') the element of $\hat{\mathbf{S}}$ as:

$$\hat{S}_{hh'} = \frac{1}{4}\left(2-\delta_{ab}\right)\left(2-\delta_{cd}\right)\sum_{j=1}^{L}\left\{\begin{array}{l}\left[\hat{F}_{ab\hat{\tilde{y}}\hat{\tilde{y}}} + \hat{F}_{ba\hat{\tilde{y}}\hat{\tilde{y}}} - \sum_{k=1}^{H}\hat{\theta}_k\left(\hat{G}_{abk} + \hat{G}_{bak}\right)\right]_j \\ \times\left[\hat{F}_{cd\hat{\tilde{y}}\hat{\tilde{y}}} + \hat{F}_{dc\hat{\tilde{y}}\hat{\tilde{y}}} - \sum_{k'=1}^{H}\hat{\theta}_{k'}\left(\hat{G}_{cdk'} + \hat{G}_{dck'}\right)\right]_j\end{array}\right\}, \tag{3.130}$$

where the subscripts j indicate that the square brackets are to be evaluated for the jth level-L block. This estimator ignores sampling error in $\hat{\mathbf{v}}$, that is, in the current estimators of the random parameters. With no error in **X**, **Y**, or **Z**, the expectation in (3.119) can be written:

$$\mathrm{E}\left[\left\{\mathbf{Y}^{**} - \mathbf{Z}^{**}\hat{\boldsymbol{\theta}} + \mathbf{Z}^{**}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\right\}\left\{\mathbf{Y}^{**} - \mathbf{Z}^{**}\hat{\boldsymbol{\theta}} + \mathbf{Z}^{**}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\right\}^{\mathrm{T}}\right] \tag{3.131}$$
$$= \mathrm{E}\left[\left(\mathbf{Y}^{**} - \mathbf{Z}^{**}\hat{\boldsymbol{\theta}}\right)\left(\mathbf{Y}^{**} - \mathbf{Z}^{**}\hat{\boldsymbol{\theta}}\right)^{\mathrm{T}}\right] + \mathrm{E}\left[\mathbf{Z}^{**}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{Z}^{**\mathrm{T}}\right].$$

Given an estimator $\hat{\Pi} \equiv \hat{\mathrm{cov}}\left(\hat{\boldsymbol{\theta}}\right)$, initially set to zero, we adjust the expression for $\hat{S}_{hh'}$ in (3.130) by adding the quantity

$$\frac{1}{4}\left(2-\delta_{ab}\right)\left(2-\delta_{cd}\right)\sum_{j=1}^{L}\sum_{k=1}^{H}\sum_{k'=1}^{H}\hat{\Pi}_{kk'}\left[\hat{G}_{abk} + \hat{G}_{bak}\right]_j\left[\hat{G}_{cdk'} + \hat{G}_{dck'}\right]_j, \tag{3.132}$$

and a new adjusted sandwich estimator of the covariance matrix of the IGLS estimators $\hat{\boldsymbol{\theta}}$ follows from (3.124) by iteration.

For the RIGLS estimators we write, instead of (3.130):

$$\hat{S}_{hh'} = \frac{1}{4}\left(2-\delta_{ab}\right)\left(2-\delta_{cd}\right) \tag{3.133}$$

$$\times\sum_{j=1}^{L}\left\{\begin{array}{l}\left[\hat{F}_{ab\hat{\mathbf{y}}\hat{\mathbf{y}}}+\hat{F}_{ba\hat{\mathbf{y}}\hat{\mathbf{y}}}+\sum_{r,s=0}^{p-1}\hat{A}_{rs}\left(\hat{F}_{abrs}+\hat{F}_{bars}\right)-\sum_{k=1}^{H}\hat{\theta}_{k}\left(\hat{G}_{abk}+\hat{G}_{bak}\right)\right]_{j}\\ \times\left[\hat{F}_{cd\hat{\mathbf{y}}\hat{\mathbf{y}}}+\hat{F}_{dc\hat{\mathbf{y}}\hat{\mathbf{y}}}+\sum_{r',s'=0}^{p-1}\hat{A}_{r's'}\left(\hat{F}_{cdr's'}+\hat{F}_{dcr's'}\right)-\sum_{k'=1}^{H}\hat{\theta}_{k'}\left(\hat{G}_{cdk'}+\hat{G}_{dck'}\right)\right]_{j}\end{array}\right\}.$$

The further adjustment for sampling error in $\hat{\mathbf{v}}$ is, as before, to add the expression (3.132).

.9 Summary

n Section 3.4 we showed how the classical model of measurement error, defined in hapter 2, could be extended to model more complicated error variance and ovariance structures, such as those arising in multivariate multilevel models. We used n educational example to show how these error structures might arise in practice. his led us to define in Section 3.5 a multilevel model in which the response was the um of m variates, $m \geq 1$, with $m = 1$ representing the univariate case.

n Section 3.6 we defined the error variance and covariance vectors (the **C** vectors) to e used to specify to the estimation procedure the prior values of the error variances nd covariances. We also gave the convention to be followed when specifying cross-evel covariances (which occur, for example, when a variable and its aggregate are oth present). In multivariate models the error covariances between dummy variables enerated from the same explanatory variable can have a complex structure. The error ariances and covariances used in the estimation are held in what are termed error roduct matrices (the **M** matrices), and we showed how to derive these from the **C** ectors. The required adjusted product matrices (the **N** matrices), being estimators of ross-product matrices of true values of the variables, then followed.

n Section 3.7 we showed how to use the **M** and **N** matrices to derive an IGLS stimator of $\boldsymbol{\beta}$, using a weight matrix $\overline{\mathbf{V}}^{-1}$ containing no measurement error, subject to e assumptions which we summarised in Section 3.7.4. We also derived an adjusted andwich estimator of $\text{cov}(\hat{\boldsymbol{\beta}})$, conditional on the true values of the variables, and rther corrected for sampling error in $\hat{\boldsymbol{\beta}}$. There are a number of possible choices for

$\bar{v}$. OLS estimation is provided by $\bar{\mathbf{v}} = \mathbf{I}$, and what we have termed 'purged-V' estimation by removing from the current estimator $\hat{\mathbf{V}}$ (see equation 3.56) all terms containing measurement error. Although we did not give details, it would be quite straightforward to enable the user to specify a weight matrix.

The reason for using an error-free weight matrix $\bar{\mathbf{v}}^{-1}$, rather than the unbiased estimator $\hat{\mathbf{v}}^{-1}$ derived from $\hat{\mathbf{V}}$ by removing error variances and covariances, is that in any given case $\hat{\mathbf{v}}^{-1}$ will still contain errors of measurement which covary with the errors of measurement remaining in the **N** matrices in the expression for $\hat{\boldsymbol{\beta}}$. Although it is possible to remove these (up to, say, 4th moments) the procedure for doing so, which we indicated in Section 3.7.1, is complex, and subject to further assumptions that would need to be tested.

In Section 3.8 we derived IGLS and RIGLS estimators of the random parameters $\boldsymbol{\theta}$ and corresponding adjusted sandwich estimators of $\mathrm{cov}\left(\hat{\theta}\right)$, the latter corrected also for sampling error in $\hat{\boldsymbol{\theta}}$, again using a weight matrix based on $\bar{\mathbf{v}}^{-1}$.

We have not demonstrated the consistency of any of these estimators, and there are further adjustments that we have indicated but not pursued. In the next chapter we explore for some simple models how well the estimators that we have defined perform in practice.

For convenience, we now summarise the steps for IGLS estimation in the general case:

1. Using the $\mathbf{C}$ vectors specified by the user, form $\mathbf{M}$ and $\mathbf{N}$ matrices for the explanatory and response variables as defined in Section 3.6.

2. With $\overline{\mathbf{v}} = \mathbf{I}_N$, use equations (3.74) and (3.75) to form OLS estimates of the fixed parameters $\boldsymbol{\beta}$.

3. Using the existing estimates of $\boldsymbol{\beta}$, form the $\mathbf{M}$ and $\mathbf{N}$ matrices that concern residuals.

4. Use equations (3.80) and (3.88), with $\hat{\boldsymbol{\Lambda}} = \mathbf{0}$ as input, to form an initial estimate of $\boldsymbol{\Lambda} \equiv \mathrm{cov}\left(\hat{\boldsymbol{\beta}} \middle| \mathbf{x}, \overline{\mathbf{v}}\right)$.

5. Again with $\overline{\mathbf{v}} = \mathbf{I}_N$, use equations (3.113)–(3.115) to form OLS estimates of the random parameters $\boldsymbol{\theta}$, and equations (3.124), (3.128), (3.130), and (3.132) with $\hat{\boldsymbol{\Pi}} = \mathbf{0}$ as input, to form an initial estimate of $\boldsymbol{\Pi} \equiv \mathrm{cov}\left(\hat{\boldsymbol{\theta}} \middle| \mathbf{x}, \mathbf{z}, \overline{\mathbf{v}}\right)$.

6. Using the existing estimates of $\boldsymbol{\theta}$, form $\hat{\mathbf{V}}$ as defined in (3.56). Then remove all terms containing measurement error to form $\overline{\mathbf{v}}$.

7. Use equations (3.74), (3.75) to form a new estimate of $\boldsymbol{\beta}$; form new versions of the $\mathbf{M}$ and $\mathbf{N}$ matrices for the residuals; use (3.80) and (3.88), with the existing estimates $\hat{\Lambda}_{rs}$ as input to (3.88), to form a new estimate of $\boldsymbol{\Lambda}$.

8. Use equations (3.113)–(3.115), (3.124), (3.128), (3.130), and (3.132), with the existing estimates $\hat{\Pi}_{kk'}$ as input to (3.132), to form new estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\Pi}$.

9. Repeat steps 6 to 8 until a convergence criterion is met, for example, that current estimates of all parameters are within 1% of the previous estimates.

10. Iterate equation (3.88) until convergence to form a final estimate of Λ.

11. Iterate equation (3.130), with the addition of (3.132), until convergence to form a final estimate of Π.

This method yields what we have termed purged-V estimators. These are true IGLS estimators for the case where there is no measurement error in the random part of the model, for example, the case of simple variance components. OLS estimators are obtained by executing steps 1, 2, 3, 4, 10, and 11.

4 Simulation studies

4.1 Introduction

The estimation algorithm developed in Chapter 3, with limitations which we shall describe, has been implemented as a suite of macros to be executed by the program *MLwiN* (Goldstein *et al.*, 1998). An earlier algorithm, suitable for estimating models with no error in the random part explanatory variables (and with other restrictions which we do not describe here), has been implemented in ML*n*, Version 1.0B (Rasbash and Woodhouse, 1996). That version conditions on the observed values of variables when estimating covariances of the parameter estimators. It has been summarised by Goldstein (1995, Appendix 10.1).

The limitations of the new implementation are as follows. Only IGLS, and not RIGLS, estimation of the random parameters has been implemented. Secondly, estimation of the covariances of the parameter estimators has been implemented only for models with no error in the random-part explanatory variables. Thirdly, the covariances for the random parameters are not adjusted for sampling error in the random parameter estimators: the expression in (3.130) for S_{W} is used without the addition of the expression in (3.132). Thus we may anticipate some bias in the estimates we obtain for the random parameters and for their standard errors.

The macros were designed to handle arbitrary numbers of levels and to be easily enhanced, for example to incorporate RIGLS estimation and the procedures outlined in Sections 3.7.1 and 3.7.3 and their analogues for the random parameters. The macros are not pre-compiled but interpreted during execution by *MLwiN*. In consequence they run rather slowly, and simulation studies have been more limited in scope than would have been ideal. The purpose of the simulations has been to indicate the general behaviour of the estimation procedure under conditions frequently met in practice, and to explore whether it offers a sound basis for further development. The simulations have also shown the effects on the model estimates of ignoring measurement error when it is present.

The models described in this chapter are as follows:

i. a 2-level variance components model with one predictor with measurement error at level 1 and one other continuous level-1 predictor (see Section 4.2.1),

ii. a 2-level variance components model with two predictors with measurement error, one at level 1 and one (correlated) at level 2, and one other continuous level-1 predictor (see Section 4.2.2),

iii. a 2-level random coefficient model with one predictor with measurement error at level 1 and a coefficient random at level 2, and one other level-1 predictor (see Section 4.3).

The data sets for all the simulation studies were balanced, and contained 30 level-2 units each with 20 level-1 units. With this size of data set the estimation of a simple model using the macros takes between 20 and 30 minutes on the fastest available machine (300MHz Pentium Pro). Simulation for larger data sets with the current implementation is not feasible on the present generation of personal computers.

In the descriptions of the models in this chapter we use j to index a level-2 unit. Whenever the context requires it, level-1 unit i is assumed to belong to level-2 unit j.

Superscripts (1) and (2) indicate the level of a random parameter, and these superscripts are applied also to the corresponding random variables.

4.2 Variance components models

For variance components models with univariate response and no error in the response variable the parameter estimates given by the new macros are the same as those given by the ML*n* implementation (version 1.0B, Rasbash and Woodhouse, 1995). The weight matrix $\hat{\mathbf{v}}^{-1}$ in this case is already free of measurement error, and action such as 'purging' or replacement by the identity matrix, as described in Sections 3.7.2 and 3.7.4, is unnecessary.

The new macros differ in the estimation of the standard errors of the parameter estimates. Thus, for the two models described in this section, we used ML*n* directly to study the characteristics of the estimates themselves, including their empirical standard deviations. Then the macros were used in a more limited simulation to obtain mean values of the sandwich estimates of the standard errors and compare these with the empirical standard deviations from the simulation with ML*n*. Two sets of standard error estimates for the fixed parameters were compared, one set including the correction for sampling error in $\hat{\boldsymbol{\beta}}$ given in equation (3.88) – these are termed *sandwich, corrected* – and the other not including that correction – termed *sandwich, uncorrected*. We also compared the standard error estimates obtained using ML*n*, which as we indicated earlier are conditioned on the observed values of the variables. These estimates are based on model estimators of $\text{cov}(\mathbf{y})$ and $\text{cov}(\mathbf{y}^{**})$ and we term them *model-based* standard errors.

We carried out analyses using measurement error variances τ_1^2 equal to 0%, 10%, 20%, and 30% of the total variance of the variable with error.

4.2.1 The simplest case

The first model to be studied was

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(2)} x_{0,i}, \qquad (4.1)$$
$$\mathrm{var}\left(\varepsilon_{0,i}^{(1)}\right) = \sigma_{00}^{(1)}, \quad \mathrm{var}\left(\varepsilon_{0,j}^{(2)}\right) = \sigma_{00}^{(2)},$$

where $x_{0,i} = 1$ is constant, $x_{1,i}$ is the true value for unit i of $\mathbf{X}_1$, a level-1 variable measured with random error distributed as $\mathrm{N}\left(0, \tau_1^2\right)$, and $x_{2,i}$ is the true value for unit i of $\mathbf{x}_2$, a level-1 variable measured without error. The designed values of β_0, β_1, and β_2 were each 1. The designed value of $\sigma_{00}^{(1)}$ was 5, and that of $\sigma_{00}^{(2)}$ was 1.

True values $\mathbf{x}_1$ and $\mathbf{x}_2$ were formed by generating two independent samples of 600 from $\mathrm{N}(0,1)$. These values remained the same throughout the simulation for this model. To estimate the parameters and obtain empirical standard deviations of the estimates we used the following scheme.

For a given measurement error variance τ_1^2,

1. form the error variance vector $\mathbf{C}_{11}^{(1)}$ as a column containing 600 copies of τ_1^2,
2. take an independent sample of 600 from $\mathrm{N}\left(0, \tau_1^2\right)$ and add to $\mathbf{x}_1$ to form $\mathbf{X}_1$,
3. take an independent sample of 600 from $\mathrm{N}(0,5)$ for the $\varepsilon_{0,i}^{(1)}$, and one of 30 from $\mathrm{N}(0,1)$ for the $\varepsilon_{0,j}^{(2)}$,

4. for each i generate y_i by adding to the fixed linear predictor $x_{0,i} + x_{1,i} + x_{2,i}$ the appropriate member from each set of random variables $\varepsilon_{0,i}^{(1)}$ and $\varepsilon_{0,j}^{(2)}$,

5. estimate the parameters of model (4.1) using $\mathbf{x}_0 = \mathbf{1}$, $\mathbf{X}_1$, and $\mathbf{x}_2$ as explanatory variables and $\mathbf{y} = \{y_i\}$ as the response, using $\mathbf{C}_{11}^{(1)}$ to adjust for the measurement error in $\mathbf{X}_1$, as described in Chapter 3 and summarised in Section 3.9,

6. repeat steps 2 to 5 a total of 10,000 times, and calculate the mean and standard deviation of each parameter estimate over these 10,000 replications.

The bias and mean squared error statistics that we tabulate below are in relation to the designed values. Finally, we estimated model (4.1) over 200 repetitions of steps 2 to 5 to obtain similar statistics for the standard errors of the estimators, using the empirical standard deviations from the first set of simulations as the basis for comparison.

Error variances approximately 10%, 20%, and 30% of the total variance of $\mathbf{X}_1$ were specified by setting τ_1^2 equal to $\frac{1}{9}$, $\frac{2}{8}$, and $\frac{3}{7}$, respectively. If we define the reliability R_1 of $\mathbf{X}_1$ as the ratio of the variances of $\mathbf{x}_1$ and $\mathbf{X}_1$, these values of τ_1^2 correspond to values of R_1 approximately 0.9, 0.8, and 0.7, respectively.

We report first on the parameter estimates and then on the standard errors. Table 4.1 summarises the parameter estimates obtained with $\mathbf{X}_1 = \mathbf{x}_1$, that is, $R_1 = 1.0$. Recall that these estimates were obtained using ML*n*.

Table 4.1 **Parameter estimates for model (4.1).**
$R_1 = 1.0$.

	designed value	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	0.997	−0.003	0.202	0.000
β_1	1.000	1.001	+0.001	0.097	0.000
β_2	1.000	1.000	−0.000	0.088	0.000
$\sigma_{00}^{(2)}$	1.000	0.960	−0.040	0.317	0.016
$\sigma_{00}^{(1)}$	5.000	4.982	−0.004	0.059	0.004

*as a proportion of the designed value

(We display values to at least 3 d.p. in most tables in order to illustrate patterns of behaviour of the adjusted estimates as measurement error increases.) We note the relative bias −0.040 in the estimate of the level-2 intercept variance, which is characteristic of the IGLS estimation procedure when applied to a data set of this size. This bias is corrected (to −0.003) for this data set by the RIGLS procedure of ML*n*.

Table 4.2 compares the estimates for model (4.1) that were obtained by repeating steps 2 to 5 200 times for $\tau_1^2 = \frac{1}{9}, \frac{2}{8}, \frac{3}{7}$ ($R_1 \approx 0.9, 0.8, 0.7$) but with no adjustment for the measurement error variance of $\mathbf{X}_1$. Note that with relatively few repetitions the mean values of the residual variances actually achieved in the simulated data, which we have labelled 'designed value', differ somewhat from 1 and 5 respectively.

The main purpose of the table is to illustrate the biases in the estimates of β_1, the parameter associated with the variable with error, which are similar to those expected in the single-level case. We note also a corresponding increase in the estimated residual variance at level 1. Estimates of the other parameters are affected only slightly.

Table 4.2 Unadjusted parameter estimates for model (4.1).
R_1 = 0.9, 0.8, 0.7. 200 replications.

	designed value	R_1	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.00	0.9	1.02	+0.02	0.23	0.01
		0.8	1.03	+0.03	0.23	0.01
		0.7	1.03	+0.03	0.23	0.02
β_1	1.00	0.9	0.88	**−0.12**	0.15	0.63
		0.8	0.78	**−0.22**	0.24	0.87
		0.7	0.67	**−0.33**	0.34	0.94
β_2	1.00	0.9	1.01	+0.01	0.10	0.02
		0.8	1.02	+0.02	0.10	0.03
		0.7	1.02	+0.02	0.10	0.03
$\sigma_{00}^{(2)}$	1.02	0.9	0.98	−0.05	0.33	0.02
		0.8	0.97	−0.05	0.33	0.02
		0.7	0.97	−0.05	0.33	0.02
$\sigma_{00}^{(1)}$	4.97	0.9	**5.05**	+0.02	0.06	0.07
		0.8	**5.15**	+0.04	0.07	0.26
		0.7	**5.24**	+0.05	0.08	0.45

*as a proportion of the designed value

Table 4.3 summarises the parameter estimates obtained for decreasing values of the reliability R_1, using the procedure described in steps 1 to 6 above. The biases in the estimates of β_1 and $\sigma_{00}^{(1)}$ are almost fully corrected. The bias in $\hat{\sigma}_{00}^{(2)}$ remains, but is no greater than for the case $R_1 = 1.0$. There are slight increases in the RMS errors of the estimates reflecting increasing uncertainty as measurement error increases. For each estimate the squared bias remains a very small proportion of the mean squared error throughout this range of measurement error variances.

Table 4.3 Adjusted parameter estimates for model (4.1).
R_1 = 0.9, 0.8, 0.7.

	designed value	R_1	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	0.9	0.998	−0.002	0.203	0.000
		0.8	0.997	−0.003	0.203	0.000
		0.7	0.997	−0.003	0.204	0.000
β_1	1.000	0.9	1.001	+0.001	0.104	0.000
		0.8	1.002	+0.002	0.113	0.000
		0.7	1.003	+0.003	0.124	0.000
β_2	1.000	0.9	1.000	−0.000	0.089	0.000
		0.8	1.000	−0.000	0.091	0.000
		0.7	1.000	−0.000	0.092	0.000
$\sigma_{00}^{(2)}$	1.000	0.9	0.960	−0.040	0.319	0.016
		0.8	0.959	−0.041	0.321	0.016
		0.7	0.959	−0.041	0.324	0.016
$\sigma_{00}^{(1)}$	5.000	0.9	4.982	−0.004	0.060	0.004
		0.8	4.981	−0.004	0.062	0.004
		0.7	4.978	−0.004	0.064	0.005

*as a proportion of the designed value

The achieved variance t_1^2 of the measurement errors sampled at step 2 in each replication itself varies about the designed value τ_1^2. We should expect from the results in Table 4.2 that the estimator $\hat{\beta}_1$ should be negatively correlated, and $\hat{\sigma}_{00}^{(1)}$ positively correlated, with the achieved measurement error variance for a given value of τ_1^2. The estimator $\hat{\sigma}_{00}^{(1)}$ is affected also by the achieved variance of the values $\varepsilon_{0j}^{(1)}$ sampled at step 3 of the scheme, so in Table 4.4 we correlate the biases in the estimators, which we denote by $\tilde{\beta}_1$ and $\tilde{\sigma}_{00}^{(1)}$.

Table 4.4 Correlations between the estimator biases $\tilde{\beta}_1$ and $\tilde{\sigma}_{00}^{(1)}$ and the achieved measurement error variance t_1^2. Model (4.1).

	$R_1 = 0.9$	$R_1 = 0.8$	$R_1 = 0.7$
$\mathrm{corr}\left(\tilde{\beta}_1, t_1^2\right)$	−0.06	−0.13	−0.21
$\mathrm{corr}\left(\tilde{\sigma}_{00}^{(1)}, t_1^2\right)$	+0.08	+0.13	+0.18
$\mathrm{corr}\left(\tilde{\beta}_1, \hat{\sigma}_{00}^{(1)}\right)$	−0.22	−0.37	−0.49

The results confirm our expectations.

We turn now to the estimates of the standard errors of the parameter estimators. Tables 4.5 to 4.7 summarise these for the case $R_1 = 1.0$. For the fixed-parameter standard errors we compare the model-based estimators provided by ML*n* with two forms of sandwich estimator, one of them corrected for sampling error in $\hat{\boldsymbol{\beta}}$ using equation (3.88), the other uncorrected. The corresponding correction for the random-parameter standard errors, given in equation (3.132), was not implemented, so for these we compare the model-based estimators with the uncorrected sandwich estimators. The 'empirical s.d.' for a given standard error estimate is the standard deviation over 10,000 replications of the corresponding parameter estimate.

Table 4.5 Standard error estimates for model (4.1). $R_1 = 1.0$. Model-based.

	empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
$\text{s.e.}(\hat{\beta}_0)$	0.202	0.200	−0.009	0.135	0.005
$\text{s.e.}(\hat{\beta}_1)$	0.097	0.096	−0.009	0.030	0.097
$\text{s.e.}(\hat{\beta}_2)$	0.088	0.088	−0.001	0.028	0.002
$\text{s.e.}(\hat{\sigma}_{00}^{(2)})$	0.315	0.315	+0.001	0.270	0.000
$\text{s.e.}(\hat{\sigma}_{00}^{(1)})$	0.294	0.293	−0.001	0.058	0.000

*as a proportion of the empirical s.d.

Table 4.6 Standard error estimates for model (4.1). $R_1 = 1.0$. Sandwich, uncorrected.

	empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
$\text{s.e.}(\hat{\beta}_0)$	0.202	0.200	−0.009	0.135	0.005
$\text{s.e.}(\hat{\beta}_1)$	0.097	0.092	−0.051	0.135	0.144
$\text{s.e.}(\hat{\beta}_2)$	0.088	0.086	−0.029	0.135	0.046
$\text{s.e.}(\hat{\sigma}_{00}^{(2)})$	0.315	0.289	−0.083	0.316	0.068
$\text{s.e.}(\hat{\sigma}_{00}^{(1)})$	0.294	0.285	−0.028	0.153	0.034

*as a proportion of the empirical s.d.

Table 4.7 **Standard error estimates for the fixed parameters of model (4.1). $R_1 = 1.0$. Sandwich, corrected.**

	empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
$\text{s.e.}(\hat{\beta}_0)$	0.202	0.204	+0.008	0.137	0.003
$\text{s.e.}(\hat{\beta}_1)$	0.097	0.094	−0.033	0.131	0.064
$\text{s.e.}(\hat{\beta}_2)$	0.088	0.088	−0.009	0.135	0.005

*as a proportion of the empirical s.d.

The uncorrected sandwich estimates in Table 4.6 matched those obtained using ML*n* directly with IGLS estimation of the parameters and sandwich estimation of the standard errors. In other words, when there is no measurement error, the sandwich estimation method of the macros, uncorrected for sampling error in the parameter estimates, gives the same results as ML*n*. Table 4.6 confirms the expected downward bias in that estimate of $\text{s.e.}(\hat{\sigma}^{(2)}_{00})$. Correction for sampling error in $\hat{\boldsymbol{\beta}}$ reduced the negative bias in the sandwich estimate of all fixed-parameter standard errors, as is shown in Table 4.7. (This procedure was not implemented for the random-parameter standard errors.) The model-based estimators appear to perform somewhat better for these data than the sandwich estimators, which generally are downwardly biased. The data in this case were constructed in such a way that the distributional assumptions underlying the model-based estimators were correct. Moreover, the tabulated estimates are based on only 200 replications. In particular, in the absence of measurement error the model and the data are symmetrical in β_1 and β_2 and there is no reason to expect a significant difference in the estimates of their standard errors.

Tables 4.8 to 4.10 summarise the standard error estimates of each type, for increasing levels of measurement error in $\mathbf{X}_1$.

Table 4.8 Standard error estimates for model (4.1). $R_1 = 0.9, 0.8, 0.7$. Model-based.

	R_1	empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
$\text{s.e.}(\hat{\beta}_0)$	0.9	0.203	0.201	−0.009	0.135	0.005
	0.8	0.203	0.201	−0.010	0.135	0.005
	0.7	0.204	0.202	−0.010	0.136	0.005
$\text{s.e.}(\hat{\beta}_1)$	0.9	0.104	0.108	+0.040	0.053	0.559
	0.8	0.113	0.122	+0.080	0.092	0.750
	0.7	0.124	0.137	+0.110	0.126	0.758
$\text{s.e.}(\hat{\beta}_2)$	0.9	0.089	0.089	−0.001	0.028	0.000
	0.8	0.091	0.090	−0.000	0.028	0.000
	0.7	0.092	0.092	−0.001	0.028	0.001
$\text{s.e.}(\hat{\sigma}_{00}^{(2)})$	0.9	0.316	0.318	+0.005	0.270	0.000
	0.8	0.318	0.320	+0.005	0.272	0.000
	0.7	0.321	0.322	+0.003	0.271	0.000
$\text{s.e.}(\hat{\sigma}_{00}^{(1)})$	0.9	0.300	0.301	+0.003	0.058	0.003
	0.8	0.308	0.309	+0.003	0.058	0.002
	0.7	0.319	0.319	+0.000	0.056	0.000

*as a proportion of the empirical s.d.

Table 4.9 **Standard error estimates for model (4.1). $R_1 = 0.9, 0.8, 0.7$. Sandwich, uncorrected.**

	R_1	empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
s.e.$(\hat{\beta}_0)$	0.9	0.203	0.201	−0.010	0.135	0.005
	0.8	0.203	0.201	−0.010	0.135	0.006
	0.7	0.204	0.202	−0.010	0.135	0.006
s.e.$(\hat{\beta}_1)$	0.9	0.104	0.099	−0.051	0.134	0.146
	0.8	0.113	0.107	−0.049	0.132	0.138
	0.7	0.124	0.118	−0.047	0.133	0.127
s.e.$(\hat{\beta}_2)$	0.9	0.089	0.087	−0.029	0.135	0.045
	0.8	0.091	0.088	−0.029	0.135	0.046
	0.7	0.092	0.089	−0.029	0.135	0.047
s.e.$(\hat{\sigma}^{(2)}_{00})$	0.9	0.316	0.291	−0.081	0.316	0.066
	0.8	0.318	0.293	−0.081	0.315	0.065
	0.7	0.321	0.295	−0.080	0.315	0.065
s.e.$(\hat{\sigma}^{(1)}_{00})$	0.9	0.300	0.292	−0.026	0.154	0.028
	0.8	0.308	0.300	−0.025	0.155	0.026
	0.7	0.319	0.311	−0.025	0.155	0.026

*as a proportion of the empirical s.d.

Table 4.10 **Standard error estimates for the fixed parameters of model (4.1). $R_1 = 0.9, 0.8, 0.7$. Sandwich, corrected.**

	R_1	empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
s.e.$(\hat{\beta}_0)$	0.9	0.203	0.204	+0.007	0.137	0.003
	0.8	0.203	0.205	+0.007	0.137	0.003
	0.7	0.204	0.205	+0.007	0.137	0.003
s.e.$(\hat{\beta}_1)$	0.9	0.104	0.101	−0.032	0.130	0.062
	0.8	0.113	0.109	−0.030	0.128	0.054
	0.7	0.124	0.120	−0.027	0.130	0.044
s.e.$(\hat{\beta}_2)$	0.9	0.089	0.088	−0.009	0.135	0.004
	0.8	0.091	0.090	−0.009	0.135	0.004
	0.7	0.092	0.091	−0.009	0.135	0.004

*as a proportion of the empirical s.d.

The empirical standard deviations of all parameter estimators increase with increasing measurement error, as we should expect, and this effect is seen to be most marked for $\hat{\beta}_1$ which is associated with the variable with error. In Table 4.8 we see also progressive over-estimation of $\text{s.e.}\left(\hat{\beta}_1\right)$ by the model-based procedure of ML*n*, resulting from the assumption of fixed *observed* values $\mathbf{X}_1$. The biases in each of the sandwich estimators remain virtually unchanged across the range of measurement error variances we have considered, and the correction for sampling error in $\hat{\boldsymbol{\beta}}$ appears to give satisfactory estimates for the fixed parameter standard errors. The biases in the sandwich standard error estimators for the random parameters appear to be unaffected by measurement error in this range.

Another measure of the performance of a standard error estimator in simulation is the proportion of replications in which a coverage interval around the parameter estimate constructed using the estimated standard error contains the true value of the parameter. Table 4.11 provides this information for the corrected sandwich estimators of the fixed parameter standard errors, based on 200 replications.

Table 4.11 Proportion of 200 trials in which true values of fixed parameters lay within ±2 estimated standard errors of their estimates. Model (4.1). Sandwich standard errors, corrected.

	$R_1 = 1.0$	$R_1 = 0.9$	$R_1 = 0.8$	$R_1 = 0.7$
β_0	0.920	0.920	0.920	0.920
β_1	0.940	0.935	0.930	0.930
β_2	0.920	0.925	0.930	0.930

Although apparently somewhat low, these proportions are based on only 200 trials. There is no evidence of substantial deterioration in performance on this indicator as measurement error increases.

4.2.2 Errors at level 2

The second model to be studied was:

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,j} + \varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(2)} x_{0,i}, \qquad (4.2)$$
$$\operatorname{var}\left(\varepsilon_{0,i}^{(1)}\right) = \sigma_{00}^{(1)}, \quad \operatorname{var}\left(\varepsilon_{0,j}^{(2)}\right) = \sigma_{00}^{(2)},$$

which is model (4.1) with the additional predictor $\mathbf{x}_3$, defined at level 2. The variable $\mathbf{x}_3$ is the true level-2 unit mean on $\mathbf{x}_1$. It is measured by $\mathbf{X}_3$ where, for given j, $X_{3,j}$ is the mean of the values $X_{1,i}$ over the sample units i which belong to level-2 unit j. In this situation $\mathbf{X}_3$ is subject to error from two sources: first, the measurement error in $\mathbf{X}_1$, and second, sampling error. Thus model (4.2) is similar to model (3.9), with an additional level-1 predictor. In particular, the errors in $\mathbf{X}_1$ covary with those in $\mathbf{X}_3$. The designed values of β_0, β_1, and β_2 were each 1, and that of β_3 was –0.3. The designed values of $\sigma_{00}^{(1)}$ and $\sigma_{00}^{(2)}$ were 0.3 and 0.06, respectively. These values are of a similar order of magnitude to those for model (5.1), which we use for the substantive application in Chapter 5.

The $\mathbf{x}_1$ were formed by first sampling 600 values independently from $\mathrm{N}(0,1)$, then adding level-2 variance 0.125 by sampling 30 values independently from $\mathrm{N}(0,0.125)$ and adding the appropriate value from this set to each $x_{1,i}$. This level-2 variance proportion also is similar to that of the data for model (5.1). The values $\mathbf{x}_2$ were generated as in the data for model (4.1). For $\mathbf{x}_3$ the difficulty arises that if $\mathbf{x}_1$ and $\mathbf{x}_3$

remain fixed in repeated sampling then any sampling error in $\mathbf{X}_3$ also remains fixed. To avoid possible resulting bias in the estimation we adopted the following scheme.

For each $j \in \{1,2,\ldots,30\}$, $x_{3,j}$ was calculated as the mean of the $x_{1,i}$ for $i \in B_j^{(2)}$: thus, for the purpose of generating the y_i, we assumed that our sampled values of $x_{1,i}$ gave an accurate estimate of the corresponding level-2 means, that is, the sampled members were assumed for this purpose to be perfectly representative of their respective level-2 units. The values $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, and hence the linear predictor of **y**, remained the same throughout. For each $j \in \{1,2,\ldots,30\}$, random 'sampling error' was added to $x_{3,j}$ to simulate the error that would occur in a 50% simple random sample without replacement (SRSWOR) from a level-2 unit of total size 40. The method we used was to take a random sample of 30 from $\mathrm{N}(0,\tau_2^2)$, where $\tau_2^2 = \dfrac{40-20}{20(40-1)}\varsigma_1^2 = \dfrac{\varsigma_1^2}{39}$, and ς_1^2 was the within-level-2-unit variance of the $x_{1,i}$, assumed constant and estimated as 0.931. Random measurement errors were then added to the $x_{1,i}$ and their level-2 unit means added to $x_{3,j}$.

Specifically, after formation of the values $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as above, the procedure followed for each measurement error variance $\tau_1^2 \in \left\{0,\frac{1}{9},\frac{1}{8},\frac{3}{7}\right\}$ was:

1. form the error variance vectors $\mathbf{C}_{11}^{(1)}$, $\mathbf{C}_{33}^{(2)}$ and the error covariance vector $\mathbf{C}_{13}^{(2)}$ such that for each i, $C_{11,i}^{(1)} = \tau_1^2$, $C_{13,i}^{(2)} = \dfrac{\tau_1^2}{20}$, $C_{33,i}^{(2)} = \dfrac{\tau_1^2}{20} + \dfrac{\varsigma_1^2}{39}$, where the latter two expressions are derived from equations (3.12) and (3.13),

2. take an independent sample of 30 from $N(0,\tau_2^2)$ as above to form sampling errors v_j,

3. take an independent sample of 600 from $N(0,\tau_1^2)$ and add to $\mathbf{x}_1$ to form $\mathbf{X}_1$,

4. for each $j \in \{1,2,\ldots,30\}$, calculate the mean μ_j of the measurement errors sampled for the *j*th level-2 unit and form $X_{3,j} = x_{3,j} + \mu_j + v_j$,

5. take an independent sample of 600 from $N(0,0.3)$ for the $\varepsilon_{0,i}^{(1)}$, and one of 30 from $N(0,0.06)$ for the $\varepsilon_{0,j}^{(2)}$,

6. for each i generate y_i by adding to the fixed linear predictor $x_{0,i} + x_{1,i} + x_{2,i} - 0.3x_{3,j}$ the appropriate member from each set of random variables $\varepsilon_{0,i}^{(1)}$ and $\varepsilon_{0,j}^{(2)}$,

7. estimate the parameters of model (4.1) using $\mathbf{x}_0 = \mathbf{1}$, $\mathbf{X}_1$, $\mathbf{x}_2$, and $\mathbf{X}_3$ as explanatory variables and $\mathbf{y} = \{y_i\}$ as the response, using $\mathbf{C}_{11}^{(1)}$, $\mathbf{C}_{33}^{(2)}$, and $\mathbf{C}_{13}^{(2)}$ to adjust for the measurement error in $\mathbf{X}_1$ and the measurement and sampling error in $\mathbf{X}_3$, as summarised in Section 3.9,

8. repeat steps 3 to 7 200 times, and calculate the mean and standard deviation of each parameter estimate over these 200 replications,

9. repeat steps 2 and 8 200 times, and calculate the grand mean and the mean standard deviation of each parameter estimate over the 200 distinct realisations of the sampling errors.

This 'nested loop' gave 200 realisations of the 'sampling errors' in $\mathbf{X}_3$ via step 2. For each such realisation 200 sets of measurement errors were generated for $\mathbf{X}_1$ and $\mathbf{X}_3$ in steps 3 and 4, and 200 sets of residuals in step 5 for addition to the fixed linear predictor to form $\mathbf{y}$ in step 6. Thus, in all, 40,000 replications of the data were generated for each value of τ_1^2 and analysed at step 7. Bias and mean squared error statistics for the parameter estimators were calculated for each realisation of the sampling errors, using for comparison the designed values of the fixed parameters and the mean achieved values of the variances $\sigma_{00}^{(2)}$ and $\sigma_{00}^{(1)}$ over the 200 replications for that realisation. These statistics then were averaged over the 200 realisations of the sampling errors, and it is these averages that are tabulated.

The point of this scheme was to simulate the variation that would occur in repeated sampling keeping fixed $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, and hence also the sampling error in $\mathbf{X}_3$, without incurring the risk of bias from using only a single set of sampling errors. An alternative strategy would have been to form an initial 'pool' of 1,200 values in 30 sets of 40, from which to take 50% samples for the x_{1ij}. This, however, would have entailed forming a new linear predictor for each such sample. Although this scheme would have mirrored the practical *application* of model (4.2) more accurately, we took the view that it might introduce unwanted uncertainty into the error variances and covariances actually realised in the simulated data. Since our concern was to evaluate the performance of the estimation procedure in the presence of *known* error variances and covariances we preferred the former scheme.

For the estimated standard errors we carried out a separate series of estimations. For each measurement error variance $\tau_1^2 \in \left\{0, \frac{1}{9}, \frac{2}{8}, \frac{3}{7}\right\}$ step 1 above was performed, followed by 200 repetitions of steps 2 to 7. Thus, to obtain statistics for the estimated standard errors, we used a single set of measurement errors and residuals for each realisation of the sampling errors. Each realisation of the sampling errors in the first set had yielded an empirical standard deviation for each parameter estimator. The means of these empirical standard deviations over the 200 realisations of the sampling errors in the first set were used as true values in the calculation of bias and mean squared error statistics for the standard error estimators in the second set.

Table 4.12 summarises the estimates obtained for model (4.2) with no measurement error and no sampling error in the data.

Table 4.12 Parameter estimates for model (4.2). R_1=1.0. No sampling error.

	designed value	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	1.000	−0.000	0.050	0.004
β_1	1.000	1.000	−0.000	0.024	0.005
β_2	1.000	1.000	−0.000	0.022	0.006
β_3	−0.300	−0.300	−0.001	0.391	0.005
$\sigma_{00}^{(2)}$	0.060	0.055	−0.083	0.323	0.068
$\sigma_{00}^{(1)}$	0.300	0.299	−0.003	0.059	0.004

*as a proportion of the designed value

Note again the negative bias in the IGLS estimate of $\sigma_{00}^{(2)}$.

Table 4.13 shows the effect on these estimates of the addition of sampling error and measurement error, when no adjustment for these is made.

Table 4.13 Unadjusted parameter estimates for model (4.2). R_1 = 1.0, 0.9, 0.8, 0.7. Sampling error as for 50% SRSWOR from each level-2 unit. 200 replications.

	designed value	R_1	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.00	1.0	1.00	+0.00	0.05	0.00
		0.9	1.00	+0.00	0.05	0.00
		0.8	1.00	+0.00	0.05	0.00
		0.7	1.00	+0.00	0.05	0.00
β_1	1.00	1.0	1.00	**−0.00**	0.02	0.00
		0.9	0.89	**−0.11**	0.12	0.95
		0.8	0.79	**−0.21**	0.21	0.99
		0.7	0.68	**−0.32**	0.32	0.99
β_2	1.00	1.0	1.00	+0.00	0.02	0.01
		0.9	1.00	+0.00	0.02	0.04
		0.8	1.01	+0.01	0.03	0.08
		0.7	1.01	+0.01	0.03	0.13
β_3	−0.30	1.0	−0.25	**−0.16**	0.38	0.17
		0.9	−0.17	**−0.43**	0.56	0.58
		0.8	−0.10	**−0.68**	0.77	0.77
		0.7	−0.03	**−0.91**	0.98	0.85
$\sigma_{00}^{(2)}$	0.063	1.0	0.059	**−0.06**	0.31	0.03
		0.9	0.056	**−0.12**	0.33	0.13
		0.8	0.053	**−0.16**	0.36	0.20
		0.7	0.051	**−0.19**	0.38	0.24
$\sigma_{00}^{(1)}$	0.30	1.0	0.30	**−0.00**	0.06	0.00
		0.9	0.40	**+0.33**	0.34	0.94
		0.8	0.50	**+0.66**	0.67	0.98
		0.7	0.59	**+0.98**	0.99	0.99

*as a proportion of the designed value

The relative biases in $\hat{\beta}_1$ are similar to the single-level case. The estimates of β_3 become more severely attenuated, such that at $R_1 = 0.7$ this estimate is less than 10%

of its true value. There are concomitant biases in the estimates of the random parameters.

Table 4.14 shows the average results when the same model is analysed with adjustment for the measurement and sampling error variances, as in the procedure described above.

Table 4.14 Adjusted parameter estimates for model (4.2).
R_1 = 1.0, 0.9, 0.8, 0.7. Sampling error as for 50% SRSWOR from each level-2 unit.

	designed value	R_1	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	1.0	0.999	−0.001	0.051	0.034
		0.9	0.999	−0.001	0.052	0.033
		0.8	0.999	−0.001	0.053	0.031
		0.7	0.999	−0.001	0.054	0.029
β_1	1.000	1.0	1.000	+0.000	0.024	0.005
		0.9	1.001	+0.001	0.030	0.005
		0.8	1.001	+0.001	0.038	0.006
		0.7	1.003	+0.003	0.049	0.008
β_2	1.000	1.0	1.000	−0.000	0.022	0.006
		0.9	1.000	−0.000	0.025	0.006
		0.8	1.000	−0.000	0.029	0.006
		0.7	1.000	−0.000	0.034	0.006
β_3	−0.300	1.0	−0.302	+0.006	0.424	0.029
		0.9	−0.300	+0.001	0.442	0.027
		0.8	−0.298	−0.006	0.466	0.025
		0.7	−0.296	−0.015	0.502	0.024
$\sigma_{00}^{(2)}$	0.060	1.0	0.054	−0.091	0.334	0.077
		0.9	0.054	−0.095	0.348	0.078
		0.8	0.054	−0.101	0.364	0.080
		0.7	0.053	−0.108	0.385	0.082
$\sigma_{00}^{(1)}$	0.300	1.0	0.299	−0.003	0.059	0.003
		0.9	0.299	−0.005	0.082	0.006
		0.8	0.298	−0.007	0.110	0.008
		0.7	0.297	−0.011	0.145	0.010

*as a proportion of the designed value

The biases in the fixed parameter estimators are virtually eliminated as is the bias in $\hat{\sigma}_{00}^{(1)}$. The downward bias in $\hat{\sigma}_{00}^{(2)}$ increases from approximately 8%, its value when there is no error in the data (see Table 4.12), to approximately 11%. It would be worth while to investigate the extent to which a RIGLS procedure could reduce this bias. We note again the progressive increases in the relative RMS errors of all estimators. The proportion of mean squared error accounted for by squared bias remains small except in the case of $\hat{\sigma}_{00}^{(2)}$, where it remains roughly constant at 8%.

We now tabulate the correlations, that are of interest, among the parameter estimator biases and between these and the achieved values of the measurement error variances. We use t_1^2 to denote the achieved measurement error variance at level 1 and t_2^2 for the achieved total of measurement error and sampling error variance at level 2, and as before we denote parameter estimator biases by a tilde. For t_1^2 we show in Table 4.15 the correlations within a realisation of the sampling errors, averaged over the 200 realisations. For t_2^2 we show in Table 4.16 the correlations across these 200 realisations between the mean biases and achieved measurement error variances for each realisation. We omit from the tables any correlation less than 0.15 in absolute value. On this criterion the correlations with the achieved measurement error correlation between the two levels were not of interest.

Table 4.15 Mean correlations between the estimator biases $\tilde{\beta}_1$, $\tilde{\beta}_3$, and $\tilde{\sigma}_{00}^{(1)}$ and the achieved measurement error variance t_1^2, for fixed sampling error in X_3. Model (4.2).

	$R_1 = 0.9$	$R_1 = 0.8$	$R_1 = 0.7$
$\mathrm{corr}\left(\tilde{\beta}_1, t_1^2\right)$	−0.23	−0.41	−0.54
$\mathrm{corr}\left(\tilde{\sigma}_{00}^{(1)}, t_1^2\right)$	+0.37	+0.52	+0.63
$\mathrm{corr}\left(\tilde{\beta}_1, \tilde{\beta}_3\right)$	−0.25	−0.30	−0.35
$\mathrm{corr}\left(\tilde{\beta}_1, \tilde{\sigma}_{00}^{(1)}\right)$	−0.33	−0.49	−0.61

The designed value of β_1 is positive. The negative correlation between $\tilde{\beta}_1$ and the achieved measurement error variance t_1^2 reflects the attenuation that occurs when measurement error is inadequately adjusted for (and the inflation which typically arises from over-adjustment). Within a given realisation of the sampling errors, a high value of t_1^2 is associated with a high value of t_2^2 and hence with attenuation of $\hat{\beta}_3$ towards zero. Since β_3 is negative this corresponds to positive bias, hence the negative correlation between $\tilde{\beta}_1$ and $\tilde{\beta}_3$. Also, with $\beta_1 > 0$, negative bias in its estimator is associated with greater unexplained level-1 variance. This accounts for the negative correlation between $\tilde{\beta}_1$ and $\tilde{\sigma}_{00}^{(1)}$ and the positive correlation between $\tilde{\sigma}_{00}^{(1)}$ and t_1^2. The correlations in Table 4.16 below likewise have the signs expected.

Table 4.16 Correlations between mean estimator biases $\tilde{\beta}_1$, $\tilde{\beta}_3$, and $\tilde{\sigma}_{00}^{(2)}$ and the mean achieved total level-2 error variance t_2^2, over 200 realisations of the sampling error in X_3. Model (4.2).

	$R_1 = 1.0$	$R_1 = 0.9$	$R_1 = 0.8$	$R_1 = 0.7$
$\mathrm{corr}(\tilde{\beta}_3, t_2^2)$	+0.39	+0.37	+0.35	+0.33
$\mathrm{corr}(\tilde{\sigma}_{00}^{(2)}, t_2^2)$	+0.46	+0.44	+0.43	+0.41
$\mathrm{corr}(\tilde{\beta}_1, \tilde{\beta}_3)$	−0.29	−0.28	−0.28	−0.28
$\mathrm{corr}(\tilde{\beta}_3, \tilde{\sigma}_{00}^{(2)})$	+0.40	+0.35	+0.29	+0.23

We now compare the estimated standard errors, for model (4.2) with no error in the data, with the mean empirical standard deviations of the estimators obtained earlier. Table 4.17 summarises the model-based estimates obtained using ML*n*. Table 4.18 shows the sandwich estimates, corrected for the fixed parameters in respect of sampling error in $\hat{\boldsymbol{\beta}}$ but not corrected for the random parameters.

Table 4.17 **Standard error estimates for model (4.2).**
R_1=1.0. No sampling error.
Model-based.

	mean empirical s.d.	mean estimate	relative bias*	relative RMS error*	bias²/MSE
$\text{s.e.}(\hat{\beta}_0)$	0.050	0.049	−0.020	0.126	0.025
$\text{s.e.}(\hat{\beta}_1)$	0.024	0.024	−0.004	0.031	0.014
$\text{s.e.}(\hat{\beta}_2)$	0.022	0.022	+0.005	0.032	0.026
$\text{s.e.}(\hat{\beta}_3)$	0.117	0.114	−0.025	0.121	0.042
$\text{s.e.}(\hat{\sigma}_{00}^{(2)})$	0.019	0.019	−0.004	0.250	0.000
$\text{s.e.}(\hat{\sigma}_{00}^{(1)})$	0.018	0.018	−0.005	0.062	0.005

*as a proportion of the mean empirical s.d.

Table 4.18 **Standard error estimates for model (4.2).**
R_1=1.0. No sampling error.
Sandwich, corrected for sampling error for fixed parameters.

	mean empirical s.d.	mean estimate	relative bias*	relative RMSE*	bias²/MSE
$\text{s.e.}(\hat{\beta}_0)$	0.050	0.050	+0.014	0.131	0.011
$\text{s.e.}(\hat{\beta}_1)$	0.024	0.024	−0.001	0.134	0.000
$\text{s.e.}(\hat{\beta}_2)$	0.022	0.022	+0.007	0.134	0.003
$\text{s.e.}(\hat{\beta}_3)$	0.117	0.116	−0.007	0.187	0.013
$\text{s.e.}(\hat{\sigma}_{00}^{(2)})$	0.019	0.017	−0.069	0.307	0.050
$\text{s.e.}(\hat{\sigma}_{00}^{(1)})$	0.018	0.018	−0.011	0.155	0.005

*as a proportion of the mean empirical s.d.

As with model (4.1), the sandwich estimator of $\text{s.e.}\left(\hat{\sigma}_{00}^{(2)}\right)$, which was uncorrected for sampling error in $\hat{\boldsymbol{\theta}}$, was appreciably downwardly biased. For these data the corrected sandwich estimators of the fixed-parameter standard errors are preferable to the model-based estimators, even in the absence of measurement error.

Tables 4.19 and 4.20 summarise the standard error estimates of each type for increasing levels of measurement error in $\mathbf{X}_1$, together with sampling error and measurement error in $\mathbf{X}_3$.

Table 4.19 Standard error estimates for model (4.2). R_1 = 1.0, 0.9, 0.8, 0.7. Sampling error as for 50% SRSWOR from each level-2 unit. Model-based.

	R_1	mean empirical s.d.	mean estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
$\text{s.e.}\left(\hat{\beta}_0\right)$	1.0	0.050	0.050	−0.004	0.130	0.001
	0.9	0.051	0.051	−0.005	0.130	0.001
	0.8	0.052	0.052	−0.005	0.131	0.001
	0.7	0.053	0.053	−0.004	0.132	0.001
$\text{s.e.}\left(\hat{\beta}_1\right)$	1.0	0.024	0.024	−0.001	0.032	0.002
	0.9	0.030	0.031	+0.020	0.041	0.238
	0.8	0.038	0.038	+0.004	0.045	0.010
	0.7	0.049	0.048	−0.030	0.065	0.212
$\text{s.e.}\left(\hat{\beta}_2\right)$	1.0	0.022	0.022	+0.005	0.032	0.028
	0.9	0.025	0.025	+0.005	0.030	0.025
	0.8	0.029	0.029	+0.003	0.026	0.010
	0.7	0.034	0.034	+0.001	0.026	0.000
$\text{s.e.}\left(\hat{\beta}_3\right)$	1.0	0.125	0.132	+0.055	0.174	0.101
	0.9	0.131	0.139	+0.065	0.192	0.114
	0.8	0.138	0.148	+0.070	0.207	0.116
	0.7	0.149	0.159	+0.070	0.222	0.099
$\text{s.e.}\left(\hat{\sigma}_{00}^{(2)}\right)$	1.0	0.019	0.019	+0.002	0.255	0.000
	0.9	0.020	0.020	−0.005	0.253	0.000
	0.8	0.021	0.021	−0.007	0.253	0.001
	0.7	0.022	0.022	−0.008	0.251	0.001
$\text{s.e.}\left(\hat{\sigma}_{00}^{(1)}\right)$	1.0	0.018	0.018	−0.004	0.062	0.004
	0.9	0.024	0.024	−0.002	0.059	0.002
	0.8	0.033	0.033	−0.002	0.053	0.001
	0.7	0.043	0.043	−0.002	0.052	0.001

*as a proportion of the mean empirical s.d.

Table 4.20 Standard error estimates for model (4.2). R_1 = 1.0, 0.9, 0.8, 0.7. Sampling error as for 50% SRSWOR from each level-2 unit. Sandwich, corrected for sampling error for fixed parameters.

	R_1	mean empirical s.d.	mean estimate	relative bias*	relative RMSE	$\frac{\text{bias}^2}{\text{MSE}}$
$\text{s.e.}(\hat{\beta}_0)$	1.0	0.050	0.050	+0.034	0.140	0.059
	0.9	0.051	0.052	+0.033	0.141	0.053
	0.8	0.052	0.054	+0.033	0.143	0.055
	0.7	0.053	0.055	+0.036	0.146	0.061
$\text{s.e.}(\hat{\beta}_1)$	1.0	0.024	0.024	−0.000	0.134	0.000
	0.9	0.030	0.030	−0.005	0.128	0.002
	0.8	0.038	0.038	−0.008	0.131	0.004
	0.7	0.049	0.049	−0.010	0.144	0.005
$\text{s.e.}(\hat{\beta}_2)$	1.0	0.022	0.022	+0.007	0.134	0.003
	0.9	0.025	0.025	−0.003	0.136	0.000
	0.8	0.029	0.029	−0.009	0.135	0.004
	0.7	0.034	0.033	−0.013	0.134	0.010
$\text{s.e.}(\hat{\beta}_3)$	1.0	0.125	0.130	+0.038	0.210	0.032
	0.9	0.131	0.135	+0.033	0.217	0.023
	0.8	0.138	0.143	+0.032	0.225	0.020
	0.7	0.149	0.154	+0.033	0.237	0.019
$\text{s.e.}(\hat{\sigma}_{00}^{(2)})$	1.0	0.019	0.018	−0.063	0.324	0.038
	0.9	0.020	0.019	−0.072	0.307	0.055
	0.8	0.021	0.019	−0.074	0.301	0.060
	0.7	0.022	0.020	−0.073	0.295	0.061
$\text{s.e.}(\hat{\sigma}_{00}^{(1)})$	1.0	0.018	0.018	−0.011	0.155	0.005
	0.9	0.024	0.024	−0.021	0.145	0.022
	0.8	0.033	0.032	−0.027	0.142	0.036
	0.7	0.043	0.042	−0.030	0.142	0.045

*as a proportion of the mean empirical s.d.

As in model (4.1), the corrected sandwich standard error estimators for the fixed parameters were virtually unaffected by increasing measurement error in this range, and appear to be satisfactory. The model-based estimators, which are wrongly conditioned, progressively over-estimated $\text{s.e.}\left(\hat{\beta}_3\right)$. The downward bias in the uncorrected sandwich estimator for $\text{s.e.}\left(\hat{\sigma}_{00}^{(2)}\right)$ remained at roughly 7%, which is comparable to the bias with no measurement or sampling error obtained using ML*n* directly, with IGLS estimation of the parameters and sandwich estimation of the standard errors.

Finally in this section we show for the fixed parameters the proportion of the 200 trials in which their true values lay within ± 2 estimated standard errors of their estimates.

Table 4.21 Proportion of 200 trials in which true values of the fixed parameters lay within ±2 estimated standard errors of their estimates.
Model (4.2). Sampling error as for 50% SRSWOR from each level-2 unit.
Sandwich standard errors, corrected.

	$R_1 = 1.0$	$R_1 = 0.9$	$R_1 = 0.8$	$R_1 = 0.7$
β_0	0.945	0.940	0.940	0.945
β_1	0.965	0.955	0.945	0.925
β_2	0.960	0.935	0.955	0.965
β_3	0.940	0.945	0.925	0.940

4.3 A random coefficient model

The final model to be studied was

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(1)} x_{0,i} + \varepsilon_{1,j}^{(2)} x_{1,i}, \tag{4.3}$$
$$\mathrm{var}\left(\varepsilon_{0,i}^{(1)}\right) = \sigma_{00}^{(1)}, \quad \mathrm{var}\left(\varepsilon_{0,j}^{(2)}\right) = \sigma_{00}^{(2)}, \quad \mathrm{var}\left(\varepsilon_{1,j}^{(2)}\right) = \sigma_{11}^{(2)}, \quad \mathrm{cov}\left(\varepsilon_{0,j}^{(2)}, \varepsilon_{1,j}^{(2)}\right) = \sigma_{01}^{(2)},$$

where $x_{0,i}$, $x_{1,i}$, and $x_{2,i}$ are as in model (4.1). Macros have not yet (November 1997) been written to compute estimates for the covariances of the parameter estimators for the random coefficients case, and for model (4.3) we record the characteristics of the parameter estimators only.

The designed values of β_0, β_1, and β_2 in model (4.3) were each 1. The designed value of $\sigma_{00}^{(1)}$ was 15. The designed values of $\sigma_{00}^{(2)}$, $\sigma_{11}^{(2)}$, and $\sigma_{01}^{(2)}$ were 0.6, 0.7, and -0.162, respectively, reflecting a correlation -0.25 between $\varepsilon_{0,j}^{(2)}$ and $\varepsilon_{1,j}^{(2)}$. For this model with measurement error the faster ML*n* implementation does not work, and each set of statistics that we tabulate is derived from 200 replications only of the data. (We retain 3 decimal places in most tables to illustrate patterns of change.) For each replication the level-1 residuals and the measurement errors were generated from univariate Normal distributions with zero mean and appropriate variance. The level-2 residuals were generated from the bivariate distribution $\mathrm{N}\left(0, \begin{bmatrix} 0.6 & -0.162 \\ -0.162 & 0.7 \end{bmatrix}\right)$. The y_i were then generated by adding to the fixed linear predictor the quantity $\varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(2)} x_{0,i} + \varepsilon_{1,j}^{(2)} x_{1,i}$ computed directly from the appropriate residuals and values of $x_{1,i}$.

Table 4.22 summarises the parameter estimates for this model, obtained using the IGLS procedure of ML*n* with no measurement error in the data.

Table 4.22 IGLS estimates for model (4.3).
$R_1 = 1.0$.

	designed value	mean estimate	s.d. of estimate	relative bias*	relative RMS error*	$\frac{bias^2}{MSE}$
β_0	1.000	1.005	0.217	+0.005	0.217	0.001
β_1	1.000	0.993	0.240	−0.007	0.240	0.001
β_2	1.000	1.003	0.158	+0.003	0.157	0.001
$\sigma_{00}^{(2)}$	0.600	0.567	0.366	−0.054	0.611	0.008
$\sigma_{01}^{(2)}$	−0.161	−0.175	0.259	+0.087	1.605	0.003
$\sigma_{11}^{(2)}$	0.705	0.666	0.366	−0.055	0.521	0.011
$\sigma_{00}^{(1)}$	14.973	14.939	0.974	−0.002	0.065	0.001

*as a proportion of the designed value

Note that with only 200 repetitions the mean values of the random parameters actually achieved in the simulated data, and which we have labelled 'designed value', depart slightly from the true designed values. In all tables the relative bias and RMS statistics are relative to the values of the parameters actually achieved in the data.

The standard deviations of the estimates give an empirical estimate of the standard errors and we note, in particular, that the level-2 random parameters are imprecisely estimated for this small data set. There are non-negligible biases, too, in these estimators. RIGLS estimation reduces the biases in the level-2 variance estimators (to +0.022 and +0.019 respectively), though not in the covariance estimator.

Table 4.23 compares the estimates for model (4.3) obtained for $R_1 = 0.9$, 0.8, 0.7 using ML*n* with IGLS estimation and no adjustment for the measurement error in $\mathbf{X}_1$.

Table 4.23 Unadjusted parameter estimates for model (4.3). R_1 = 0.9, 0.8, 0.7.

	designed value	R_1	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.00	0.9	1.01	+0.01	0.22	0.00
		0.8	1.02	+0.02	0.22	0.01
		0.7	1.02	+0.02	0.22	0.01
β_1	1.00	0.9	0.88	**−0.12**	0.25	0.22
		0.8	0.78	**−0.22**	0.30	0.54
		0.7	0.67	**−0.33**	0.38	0.76
β_2	1.00	0.9	1.00	+0.00	0.16	0.00
		0.8	1.01	+0.01	0.16	0.00
		0.7	1.01	+0.01	0.16	0.00
$\sigma_{00}^{(2)}$	0.60	0.9	0.56	−0.06	0.61	0.01
		0.8	0.56	−0.07	0.61	0.01
		0.7	0.55	−0.08	0.61	0.02
$\sigma_{01}^{(2)}$	−0.16	0.9	−0.16	**−0.02**	1.52	0.00
		0.8	−0.15	**−0.09**	1.43	0.00
		0.7	−0.13	**−0.20**	1.32	0.02
$\sigma_{11}^{(2)}$	0.71	0.9	0.53	**−0.25**	0.53	0.22
		0.8	0.41	**−0.42**	0.58	0.53
		0.7	0.30	**−0.57**	0.67	0.75
$\sigma_{00}^{(1)}$	14.98	0.9	15.12	+0.01	0.07	0.02
		0.8	15.29	+0.02	0.07	0.09
		0.7	15.46	+0.03	0.07	0.19

*as a proportion of the designed value

The relative biases in the estimates for β_1 are again similar to those expected in the single-level case. The biases in the estimates of the level-2 variance $\sigma_{11}^{(2)}$ also become substantial, such that for $R_1 = 0.7$ this estimate is only 43% of the true value, and there are concomitant biases in the estimates of $\sigma_{01}^{(2)}$.

The adjustment procedure developed in Sections 3.7.2 and 3.8.2 and summarised in Section 3.9 uses a weight matrix $\bar{\mathbf{V}}^{-1}$ that is entirely free from measurement error. We study the results of using two such weight matrices. In one case we use $\bar{\mathbf{V}} = \mathbf{I}$: this gives OLS estimators. In the second case we form $\bar{V}$ by removing from $\hat{V}$ all terms involving X_1. These we term purged-V estimators. We first compare the OLS and purged-V estimators with the IGLS estimators for the model without error. See Tables 4.22, 4.24, and 4.25.

Table 4.24 OLS estimates for model (4.3).
$R_1 = 1.0$.

	designed value	mean estimate	s.d. of estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	1.006	0.218	+0.006	0.217	0.001
β_1	1.000	0.993	0.236	−0.007	0.236	0.001
β_2	1.000	1.004	0.158	+0.004	0.158	0.001
$\sigma_{00}^{(2)}$	0.600	0.564	0.361	−0.059	0.603	0.009
$\sigma_{01}^{(2)}$	−0.161	−0.171	0.271	+0.064	1.679	0.001
$\sigma_{11}^{(2)}$	0.705	0.667	0.371	−0.054	0.527	0.010
$\sigma_{00}^{(1)}$	14.973	14.939	0.981	−0.002	0.065	0.001

*as a proportion of the designed value

Table 4.25 **Purged-V estimates for model (4.3).**
$R_1 = 1.0$.

	designed value	mean estimate	s.d. of estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	1.006	0.218	+0.006	0.217	0.001
β_1	1.000	0.992	0.238	−0.008	0.238	0.001
β_2	1.000	1.003	0.158	+0.003	0.157	0.001
$\sigma_{00}^{(2)}$	0.600	0.568	0.363	−0.053	0.606	0.008
$\sigma_{01}^{(2)}$	−0.161	−0.175	0.266	+0.087	1.646	0.003
$\sigma_{11}^{(2)}$	0.705	0.666	0.368	−0.056	0.524	0.011
$\sigma_{00}^{(1)}$	14.973	14.938	0.979	−0.002	0.065	0.001

*as a proportion of the designed value

For the model without error all three sets of estimators show virtually identical characteristics.

In Tables 4.26 and 4.27 on the following pages we compare the OLS and purged-V estimators for decreasing values of R_1, obtained using the adjustment summarised in Section 3.9. Such differences as there are between these two sets of results are very slight. In both the bias in $\hat{\beta}_1$ observed in Table 4.23 has been almost eliminated. All estimates become progressively less precise as measurement error increases. This effect is most marked in the case of $\hat{\sigma}_{11}^{(2)}$, the variance of the coefficient of the variable with error. The bias also in this estimator increases somewhat over the range, but remains of the order of 0.1 standard deviation. The estimates for $\sigma_{01}^{(2)}$ also become progressively more biased, leading to large relative RMS errors. This estimator is the most sensitive to errors in the estimates of the other random parameters.

Table 4.26 Adjusted OLS estimates for model (4.3). $R_1 = 0.9, 0.8, 0.7$.

	designed value	R_1	mean estimate	s.d. of estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	0.9	1.006	0.220	+0.006	0.220	0.001
		0.8	1.006	0.222	+0.006	0.221	0.001
		0.7	1.006	0.224	+0.006	0.223	0.001
β_1	1.000	0.9	0.988	0.248	−0.012	0.248	0.002
		0.8	0.985	0.259	−0.015	0.259	0.003
		0.7	0.984	0.273	−0.016	0.273	0.004
β_2	1.000	0.9	1.003	0.157	+0.003	0.157	0.000
		0.8	1.003	0.157	+0.003	0.157	0.000
		0.7	1.001	0.157	+0.002	0.157	0.000
$\sigma_{00}^{(2)}$	0.600	0.9	0.564	0.362	−0.060	0.605	0.010
		0.8	0.563	0.363	−0.061	0.608	0.010
		0.7	0.562	0.366	−0.063	0.612	0.011
$\sigma_{01}^{(2)}$	−0.161	0.9	−0.181	0.289	+0.124	1.793	0.005
		0.8	−0.185	0.306	+0.149	1.899	0.006
		0.7	−0.188	0.326	+0.168	2.023	0.007
$\sigma_{11}^{(2)}$	0.705	0.9	0.668	0.420	−0.054	0.597	0.008
		0.8	0.662	0.468	−0.061	0.665	0.008
		0.7	0.654	0.528	−0.073	0.751	0.009
$\sigma_{00}^{(1)}$	14.973	0.9	14.945	1.010	−0.002	0.067	0.001
		0.8	14.951	1.039	−0.001	0.069	0.000
		0.7	14.958	1.076	−0.001	0.072	0.000

*as a proportion of the designed value

Table 4.27 Adjusted purged-V estimates for model (4.3). R_1 = 0.9, 0.8, 0.7.

	designed value	R_1	mean estimate	s.d. of estimate	relative bias*	relative RMSE*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.000	0.9	1.006	0.220	+0.006	0.220	0.001
		0.8	1.006	0.222	+0.006	0.221	0.001
		0.7	1.006	0.224	+0.006	0.223	0.001
β_1	1.000	0.9	0.988	0.249	−0.014	0.248	0.003
		0.8	0.984	0.259	−0.016	0.259	0.004
		0.7	0.983	0.272	−0.018	0.272	0.004
β_2	1.000	0.9	1.002	0.157	+0.002	0.157	0.000
		0.8	1.002	0.157	+0.002	0.157	0.000
		0.7	1.001	0.157	+0.002	0.157	0.000
$\sigma_{00}^{(2)}$	0.600	0.9	0.567	0.363	−0.054	0.607	0.008
		0.8	0.567	0.365	−0.055	0.610	0.008
		0.7	0.566	0.367	−0.056	0.614	0.008
$\sigma_{01}^{(2)}$	−0.161	0.9	−0.184	0.282	+0.145	1.753	0.007
		0.8	−0.188	0.298	+0.167	1.851	0.008
		0.7	−0.191	0.317	+0.185	1.969	0.009
$\sigma_{11}^{(2)}$	0.705	0.9	0.665	0.413	−0.057	0.586	0.009
		0.8	0.659	0.459	−0.065	0.652	0.010
		0.7	0.650	0.517	−0.079	0.735	0.011
$\sigma_{00}^{(1)}$	14.973	0.9	14.945	1.005	−0.002	0.067	0.001
		0.8	14.951	1.033	−0.001	0.069	0.000
		0.7	14.959	1.068	−0.001	0.071	0.000

*as a proportion of the designed value

Correlations of the estimator biases with the achieved measurement error variances in no case reached 0.15 in absolute value.

5 A substantive application

The analyses in the previous chapter were of data sets that conformed to known models. Residuals of known variance and covariance were added to the response variables and measurement errors of known variance and covariance to the explanatory variables. The point of the analyses was to judge the effectiveness of the estimation procedure of Chapter 3 in retrieving the model parameters. We now apply that procedure to an educational data set whose characteristics are, *a priori*, unknown. The purpose here is to show how substantive conclusions may change as we change assumptions about the extent of measurement error in explanatory variables.

5.1 The data and the model

The data are derived from the Junior School Project (JSP), a longitudinal study of an age cohort of pupils who entered junior classes in September 1980 and transferred to secondary school in September 1984. (The first junior year in English and Welsh primary schools is now called Year 3 of the National Curriculum: pupils typically reach their eighth birthday during this year.) The schools used by the JSP were a random sample from the 636 primary schools that were maintained by the Inner London Education Authority at the start of the project. A full account of the project is given in Mortimore *et al.* (1988).

For our illustrative application we use a sub-sample of 1,075 pupils from 48 schools. We consider the data to have a two-level structure, with pupils (indexed by i) at level 1 and schools (indexed by j) at level 2. Table 5.1 shows for each school j the number n_j

of pupils in the sample, the total N_j in the age cohort within the school, and the sampling fraction.

Table 5.1 Sample size n_j, cohort size N_j, and sampling fraction for each school j.

j	n_j	N_j	n_j/N_j	j	n_j	N_j	n_j/N_j
1	6	20	0.30	25	19	32	0.59
2	7	25	0.28	26	21	25	0.84
3	7	43	0.16	27	21	37	0.57
4	11	22	0.50	28	21	45	0.47
5	12	22	0.55	29	21	51	0.41
6	12	29	0.41	30	23	34	0.68
7	14	29	0.48	31	23	40	0.58
8	14	29	0.48	32	24	43	0.56
9	14	32	0.44	33	24	47	0.51
10	14	33	0.42	34	25	31	0.81
11	15	26	0.58	35	25	41	0.61
12	15	31	0.48	36	26	53	0.49
13	15	39	0.38	37	28	55	0.51
14	15	43	0.35	38	29	53	0.55
15	15	44	0.34	39	29	58	0.50
16	16	31	0.52	40	32	54	0.59
17	16	38	0.42	41	32	82	0.39
18	16	59	0.27	42	33	65	0.51
19	18	28	0.64	43	35	49	0.71
20	18	29	0.62	44	38	53	0.72
21	18	37	0.49	45	39	52	0.75
22	18	51	0.35	46	44	61	0.72
23	18	53	0.34	47	52	92	0.57
24	19	31	0.61	48	68	116	0.59

It can be seen that there is considerable variability both in the size of the level-2 samples and in the sampling fractions. The total of the N_j is 2,093, giving an overall sampling fraction of 0.51.

For the analysis we have four variables $\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2,$ and $\mathbf{X}_3$ where, for $i = 1, 2, \ldots, 1075$ and $j = 1, 2, \ldots, 48$,

- Y_i is the pupil's observed score on a reading test at age 10 years, transformed by using Normal scores to have a standard Normal distribution,
- $X_{1,i}$ is the pupil's observed score on a reading test taken at age 8, similarly Normalised,
- $X_{2,i}$ is the pupil's family socio-economic status (SES), coded 1 if the father is in non-manual work and 0 otherwise,
- $X_{3,j}$ is the mean of the $X_{1,i}$ for the pupils in school *j*, as estimated from the sample.

We illustrate the use of these four variables to model the effects on a pupil's true reading score at age 10 of that pupil's true score at age 8, the pupil's true family SES, and the true mean reading score at age 8 of the pupil's cohort in the school. It is well known that a pupil's prior achievement on a related test is a powerful predictor of subsequent performance. Family socio-economic status, too, has been shown to have a positive association with progress in reading over this age range. The effect, if any, of the mean reading score for the pupil's cohort in the school is sometimes called a 'context' effect: such effects are of considerable educational interest.

Since our concern is to show the effects of measurement error we omit details of data description and pass straight to multilevel modelling. The simplest linear model for our purpose is:

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,j} + \varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(2)} x_{0,i}, \qquad (5.1)$$
$$\mathrm{var}\left(\varepsilon_{0,i}^{(1)}\right) = \sigma_{00}^{(1)}, \quad \mathrm{var}\left(\varepsilon_{0,j}^{(2)}\right) = \sigma_{00}^{(2)},$$

where $x_{0,i} \equiv 1$, $x_{1,i}$, $x_{3,j}$ are the unknown true values of $X_{1,i}$, $X_{3,j}$, respectively, and $Y_i = y_i$, $X_{2,i} = x_{2,i}$ are both assumed to have been measured without error. We assume

$$\begin{aligned} X_{1,i} &= x_{1,i} + \xi_{1,i}, \\ X_{3,j} &= x_{3,j} + \xi_{3,j}, \end{aligned} \tag{5.2}$$

where the measurement errors ξ_1, ξ_3 obey assumptions 3.1 to 3.3 of Chapter 3 (see pp69, 70). The random variables $\varepsilon_i^{(1)}, \varepsilon_j^{(2)}$ obey assumption 3.4 and we assume further that the measurement errors ξ_1 have constant variance.

Thus, model (5.1)–(5.2) is formally the same as model (4.2), and similar to model (3.9) with an additional explanatory variable at level 1. In its present form it cannot be proposed as a satisfactory model for the data. For example, it is unrealistic to assume complete reliability of either the response variable or the SES variable. It is probable that the coefficient of $\mathbf{x}_1$ varies from school to school. It is also a simplification to assume that each pupil's score at age 10 will be affected equally by the mean score at age 8 for the full cohort in the school. Our purpose is not to arrive at the best possible model for the data, but to illustrate in a simple multilevel context the effects of adjusting, and of failing to adjust, for measurement error.

5.2 Measurement error variances and covariances, and reliability

Before carrying out the analysis, we require a prior value for the measurement error variance of $\mathbf{X}_1$, and for each school j an estimate of the variance of the total error in $X_{3,j}$ due to sampling and measurement error and of its covariance with the errors $\xi_{1,i}$

for pupils within the school. Ecob and Goldstein (1983) discussed the difficulty of obtaining satisfactory estimates of measurement error variances and covariances. They cast doubt on the assumptions underlying standard procedures and proposed an alternative procedure based on the use of instrumental variables. In the present case, no dependable prior estimate is available either for the measurement error variance τ_1^2 of $\mathbf{X}_1$ or for its reliability in the population under study. Nor have we developed a method for using instrumental variables. Accordingly, we study the effects on the analysis of different assumed values of τ_1^2. We also show some of the effects of incomplete adjustment for the sampling and measurement error in $\mathbf{X}_3$.

From equation (3.12), we have:

$$\operatorname{var}\left(\xi_{3,j}\right) = \frac{\tau_1^2}{n_j} + \frac{N_j - n_j}{n_j\left(N_j - 1\right)}\sigma_{1,j}^2, \tag{5.3}$$

where $\sigma_{1,j}^2$ is the variance of the true scores $x_{1,i}$ within the cohort for school j. We also have, from equation (3.13),

$$\operatorname{cov}\left(\xi_{1,i}, \xi_{3,j}\right) = \frac{\tau_1^2}{n_j}. \tag{5.4}$$

We can estimate model (5.1)–(5.2) provided that we have prior estimates of, or make assumptions about, τ_1^2 and $\sigma_{1,j}^2$.

We assume that $\sigma_{1,j}^2 = \varsigma_1^2$ holds for all schools, where ς_1^2 is constant. Since the measurement error variance τ_1^2 in $\mathbf{X}_1$ is constant it follows that the within-school variance of $\mathbf{X}_1$ also is constant, with value $\varsigma_1^2 + \tau_1^2$. We may obtain a reasonable estimate of this value as the level-1 variance $\hat{\sigma}^2(\mathbf{X}_1)$ obtained when fitting $\mathbf{X}_1$ to its mean in a two-level variance components model. We define the *level-1 reliability* R_1 of $\mathbf{X}_1$ by:

$$R_1 \equiv \frac{\varsigma_1^2}{\hat{\sigma}^2(\mathbf{X}_1)} = 1 - \frac{\tau_1^2}{\hat{\sigma}^2(\mathbf{X}_1)}, \qquad (5.5)$$

and now, given a series of assumed level-1 reliabilities R_1, we can derive corresponding values of τ_1^2 and ς_1^2 and hence the prior values required for the estimation of the model.

5.3 Model estimation

We describe three analyses of model (5.1)–(5.2) to show the effects of progressively more complete adjustment for measurement error:

i. analysis A, adjustment for measurement error at level 1 only, that is, no adjustment for measurement or sampling error in $\mathbf{X}_3$,

ii. analysis B, adjustment for errors in $\mathbf{X}_1$ and $\mathbf{X}_3$, but not for the covariance between these errors,

iii. analysis C, full adjustment.

Each analysis was carried out for values of R_1 ranging from 1.0 to 0.7. Unadjusted results were provided by analysis A with $R_1 = 1.0$.

The level-1 variance estimator $\hat{\sigma}^2(\mathbf{X}_1)$ was found by variance components analysis to have the value 0.89. For each value of R_1 equation (5.5) gives the corresponding values of τ_1^2 and of $\sigma_{1,j}^2 = \varsigma_1^2$. The total error variance $\tau_{3,j}^2$ at level 2 for school *j* and the error covariance $\tau_{13,j}$ follow from (5.3) and (5.4). For analysis C (full adjustment) the error variance vectors $\mathbf{C}_{11}^{(1)}, \mathbf{C}_{33}^{(2)}$ and the error covariance vector $\mathbf{C}_{13}^{(2)}$ were formed such that, for pupil *i* in school *j*, $C_{11,i}^{(1)} = \tau_1^2$, $C_{13,i}^{(2)} = \dfrac{\tau_1^2}{n_j}$, and $C_{33,i}^{(2)} = \dfrac{\tau_1^2}{n_j} + \dfrac{N_j - n_j}{n_j(N_j - 1)}\varsigma_1^2$. For analysis B the vector $\mathbf{C}_{13}^{(2)}$ was omitted, and for analysis A both $\mathbf{C}_{13}^{(2)}$ and $\mathbf{C}_{33}^{(2)}$ were omitted.

The estimates that were affected by adjustment for measurement error at level 2 were those of the level-2 variance $\sigma_{00}^{(2)}$ and the fixed parameter β_3 associated with the level-2 aggregate variable $\mathbf{x}_3$. The other parameter estimates were essentially the same in each of analyses A, B, and C, for a given value of R_1. The results are summarised in Tables 5.2 and 5.3, where we give in parentheses the sandwich estimates of the standard errors, corrected (for the fixed parameters) for sampling error in $\hat{\boldsymbol{\beta}}$.

Table 5.2 **Adjusted estimates of parameters that were essentially unaffected by measurement error at level 2. Model (5.1)–(5.2). Standard errors in parentheses.**

R_1	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_{00}^{(1)}$
1.0	–0.037 (0.04)	0.77 (0.02)	0.17 (0.05)	0.30 (0.02)
0.9	–0.026 (0.04)	0.86 (0.03)	0.13 (0.05)	0.24 (0.02)
0.8	–0.012 (0.04)	0.97 (0.03)	0.08 (0.04)	0.17 (0.02)
0.7	0.006 (0.04)	1.12 (0.05)	0.00 (0.05)	0.07 (0.02)

In Table 5.2 we see the progressive disattenuation of $\hat{\beta}_1$ and accompanying attenuation of $\hat{\sigma}_{00}^{(1)}$ as the assumed reliability of $\mathbf{X}_1$ decreases, which we should expect from the analysis of model (4.2) above. The estimate $\hat{\beta}_1$ varied approximately in inverse proportion to R_1, and its precision decreased slightly as R_1 decreased (although the estimated effect remained statistically highly significant).

The parameter β_2 is the predicted difference in outcome score between two pupils in the same school and with the same true score at age 8, where one pupil's father is in non-manual employment and the other's is not. With no adjustment for measurement error in $\mathbf{X}_1$, the benefit to the pupil with a non-manual background was estimated to be 0.17 on the outcome scale (that is, 0.17 standard deviation), and to be statistically significant. When $R_1 = 0.8$ and adjustment was made, that estimate was attenuated by a factor 0.44. When $R_1 = 0.7$ no SES effect on reading progress was found with this model.

Table 5.3 Adjusted estimates of parameters that were affected by measurement error at level 2. Model (5.1)–(5.2). Standard errors in parentheses.

$\hat{\beta}_3$: *mean score at age 8*			
R_1	*Analysis A*	*Analysis B*	*Analysis C*
1.0	0.00 (0.10)	0.01 (0.11)	0.01 (0.11)
0.9	–0.08 (0.10)	–0.10 (0.12)	–0.07 (0.12)
0.8	–0.19 (0.10)	–0.23 (0.13)	–0.17 (0.13)
0.7	–0.33 (0.11)	–0.41 (0.14)	–0.30 (0.14)
$\hat{\sigma}_{00}^{(2)}$: *level-2 variance*			
R_1	*Analysis A*	*Analysis B*	*Analysis C*
1.0	0.059 (0.012)	0.059 (0.012)	0.059 (0.012)
0.9	0.060 (0.012)	0.059 (0.012)	0.060 (0.012)
0.8	0.059 (0.012)	0.058 (0.013)	0.061 (0.012)
0.7	0.056 (0.013)	0.051 (0.013)	0.063 (0.012)

The estimation of β_3 and, to a lesser extent, of the level-2 residual variance $\sigma_{00}^{(2)}$ were affected by measurement error at level 2. The parameter β_3 predicts the difference between pupils' outcomes in different schools that is attributable to the difference between the true mean scores at age 8 in the schools. Its negative sign indicates that, according to model (5.1)–(5.2), if two pupils have equal true scores at age 8 and are equal on the family SES indicator, the pupil in the school with the lower true mean score is predicted to do better two years later.

We may hypothesise a number of possible reasons for this finding. First, if it is believed that being in a high-scoring group tends, on average, to enhance individual performance rather than to depress it, then it may be argued that a child in a lowscoring group who nevertheless achieves score x by the age of 8 is likely to be more able than a child in a high-scoring group who also scores x by the same age. If the supposed group effect decreases over time, the former child will be likely to do better eventually than the latter. An alternative hypothesis focuses on the difference between the individual's performance and the average performance of the group. Under this hypothesis a pupil with a low score (and below the average for the group) will tend to do worse if the difference is large (perhaps through loss of confidence), while a pupil with a high score is relatively unaffected by the group score, and possibly motivated by a large difference. Further analysis would be needed to test either hypothesis, and this is beyond the scope of the present study.

Table 5.3 shows how it is possible to be misled by insufficient, or incomplete, adjustment for measurement error. Of the three analyses, analysis C is clearly to be preferred. Analysis A, which did not adjust at all for the errors at level 2, overestimated the precision of estimation of β_3. Analysis B, which adjusted for the variance of the combined sampling and measurement errors in $\mathbf{X}_3$ but not for their covariance with the errors in $\mathbf{X}_1$, overestimated the magnitude of β_3.

Analysis C turned out to be the most conservative of the three in its estimates of the precision of $\hat{\beta}_3$ for each assumed value of R_1. Because the coefficients β_1 and β_3 are of opposite sign, the effect of adjusting for the positive covariance between the errors in $\mathbf{X}_1$ and $\mathbf{X}_3$ is to reduce in magnitude the estimates $\hat{\beta}_3$. (If β_1 and β_3 were of like sign, the effect of this adjustment would be to increase the magnitude of $\hat{\beta}_3$.) With an assumption of only slight measurement error in $\mathbf{X}_1$, analysis C found no statistically significant effect on reading progress of the school context as measured by the mean reading score at age 8. If, however, R_1 was assumed to have the value 0.7, the school context effect was found to be statistically significant, though it was still poorly estimated and we must enter a caveat about possibly underestimated standard error.

We may expect that the IGLS estimates of the level-2 residual variance shown for analysis C are downwardly biased, but that this bias does not increase as measurement error increases. On this assumption, the estimates of analyses A and B are more biased, for moderate values of R_1. The estimates of the standard error also are likely to be downwardly biased, though this is unlikely to affect model interpretation here.

6 Discussion

This thesis has addressed two main problems. The first was to devise a method for estimating random-coefficient multilevel models with error in a variable with a random coefficient. The second was to provide a means to specify error structures sufficiently general for use, for example, with multivariate responses and with cross-classified data. The first task proved the more troublesome. Clearly, some form of adjustment was necessary (our analysis in Section 4.3 demonstrates this) but the existing method, devised for variance components models, not only gave the wrong estimates: in most cases the procedure failed to converge at all. In early simulations not documented here, when known values of the random parameters were fed to the fixed-parameter estimation routine, the resulting estimates of the fixed parameters were found to be biased by amounts in some cases several times the magnitude of the parameter itself, when there was error in the weight matrix $\hat{\mathbf{V}}^{-1}$.

Section 3.7.1 above gives an idea of the work needed in order to correct this problem. It is not sufficient simply to find an unbiased estimator $\hat{\mathbf{V}}$ and unbiased estimators $\mathbf{N}_{rr'}$, $\mathbf{N}_{ry}$ for the cross-product matrices appearing in the IGLS estimator for $\boldsymbol{\beta}$. If, say, $\mathbf{X}_r$ is subject to error and $\mathbf{x}_r$ has a fixed and a random coefficient, the errors that remain in $\hat{\mathbf{V}}^{-1}$ in any particular instance correlate with the errors remaining in $\mathbf{N}_{rr'}$ and continue to produce seriously biased fixed-parameter estimates.

In an earlier version of the macros the method outlined in Section 3.7.1 was implemented so as to remove from the IGLS expressions all moments of the measurement errors up to the 4th (the sampling errors in the random parameter estimates were neglected). The resulting routine produced unbiased estimates of the fixed parameters when true values of the random parameters were fed into it. We have not recorded details of this work, since the corresponding work for the random parameter estimation, which is much more complicated, was unsuccessful. We developed instead procedures using an error-free weight matrix, as described in Sections 3.7.2 and 3.8.2. These procedures should give the same parameter estimates for variance components models as the existing method, since in these models the weight matrix $\hat{\mathbf{v}}^{-1}$ does not contain measurement error.

Our analyses in Section 4.2 confirmed this. Furthermore, the analysis of the two-level random-coefficient model without measurement error showed that, for this simple model, the estimates produced by OLS and by use of a purged weight matrix (in this case a weight matrix with 2-level variance-components structure) were equivalent to those produced by IGLS with the full random-coefficients weight matrix, in terms of both bias and precision (see Tables 4.22, 4.24, and 4.25). We went on to use both OLS and 'purged-V' estimation with increasing amounts of measurement error, and they continued to give similar results (Tables 4.26, 4.27).

There are grounds for cautious optimism here. The bias in the unadjusted estimate of β_1 seen in Table 4.23 was reduced by adjustment to less than 2%. Biases in the other fixed parameter estimates also were reduced, to trivial amounts. Level-2 variances were always underestimated, in this model and in the others of Chapter 4, even with no error in the data. With the adjustment for measurement error that we have implemented these biases generally became only slightly larger as measurement error increased. They seem more likely to be due to failure to adjust for sampling error in $\hat{\boldsymbol{\beta}}$, as in RIGLS estimation, than to correlations between errors of measurement remaining in $\hat{\boldsymbol{\Phi}}^{-1}$ and $\hat{\boldsymbol{\Psi}}$. Implementation of adjusted RIGLS as in equation (3.118) should, therefore, be the next step.

It may be that OLS or purged-V estimation will incur serious loss of efficiency in more complicated models with random coefficients. In that case a possible way forward is to allow the user to specify the weight matrix. Such a weight matrix might be derived from a simpler model of the data, excluding from the random part all variables with error, but still be closer to the true weight matrix than the one we are able to produce by the relatively crude procedure of 'purging'. Alternatively, the weight matrix might use predictions of explanatory variables with error using other 'instrumental' variables if these are available. The implementation of the full adjustment procedure (up to 4th moments) for errors in $\hat{\mathbf{v}}^{-1}$, with which we were unsuccessful earlier, remains a possibility, though a considerable improvement in computing efficiency would be needed to make it practicable.

A by-product of the work on the full adjustment procedure was the development of the notation introduced in equation (3.113). This notation, with its implied summation convention, can be extended to give economical and intelligible expansions of the expressions in (3.130) and (3.133), for example. Similar notational conventions might find application in descriptions of other estimation algorithms.

Standard errors of the parameter estimators have been estimated for the variance components models we have studied. Correction in respect of sampling errors, as described in equation (3.88), was implemented for the fixed parameters. The corresponding correction for the random parameter standard errors, described in equation (3.132), was not implemented in time for these studies. The estimates of relative bias in Tables 4.7, 4.10, 4.18, and 4.20, and of coverage in Tables 4.11 and 4.21, should be treated with some caution as they are based on only 200 replications. For example, Table 4.7 suggests a downward bias of 3% in the estimate of $\text{s.e.}\left(\hat{\beta}_1\right)$, but of less than 1% in that of $\text{s.e.}\left(\hat{\beta}_2\right)$, yet the model with no measurement error is symmetrical with respect to the two parameters.

The corrected estimates of standard errors for the fixed parameters are an improvement on the uncorrected ones, which in the absence of measurement error match those obtained using ML*n* directly with IGLS estimation of parameters and sandwich estimation of standard errors. In the simple model (4.1) the biases stay roughly constant as measurement error increases. Interestingly, in model (4.2) the introduction of sampling error into $\mathbf{X}_3$ occasions an immediate increase in the relative bias of the standard error of the associated parameter, and this bias then increases no further as measurement error increases. Further work is needed to establish whether the initial increase is due to the estimation of the standard error or to a fault in the (admittedly complicated) sampling scheme. At all events, the biases in the fixed-parameter standard errors are small. The biases in the standard errors for the level-2 random parameters are not negligible, and we propose to implement RIGLS estimators, further corrected for sampling error in $\hat{\theta}$, based on equations (3.132) and (3.133).

Even within this limited range of models we can discern a variety of interactive effects on the parameter estimates of failing to adjust for the presence of measurement error. In the variance-components model (4.1) with error in a level-1 variable only, the effect was confined mainly to the estimate of the fixed parameter associated with this variable. See Table 4.2. The bias in the estimate of level-1 variance increased to about 5% over the range. Thus, a decrease in the estimate of a fixed parameter associated with a level-1 variable with error was accompanied by a slight increase in the level-1 variance estimate.

The second variance-components model (4.2) contained a level-1 variable with error together with its mean at level 2, estimated with sampling error in addition to the measurement error in the first variable. The estimate of β_1, the coefficient of $\mathbf{X}_1$, was attenuated towards zero by failure to adjust for these errors and this was accompanied by a marked increase in the level-1 variance estimate. See Table 4.13. The estimate of β_3 increased, and the level-2 variance estimate decreased substantially. If model (4.2) were a model of reading progress, like model (5.1)–(5.2), these biases would mask the effect of the mean score and lead to underestimation of the residual between-school variance also. We present in Table 6.1 the corresponding results for the model

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,j} + \varepsilon_{0,i}^{(1)} x_{0,i} + \varepsilon_{0,j}^{(2)} x_{0,i}, \tag{6.1}$$
$$\operatorname{var}\left(\varepsilon_{0,i}^{(1)}\right) = \sigma_{00}^{(1)}, \quad \operatorname{var}\left(\varepsilon_{0,j}^{(2)}\right) = \sigma_{00}^{(2)},$$

which is formally the same as model (4.2), but this time with a designed value +0.3 in place of −0.3 for β_3.

Table 6.1 **Unadjusted parameter estimates for model (6.1). R_1 = 1.0, 0.9, 0.8, 0.7. Sampling error as for 50% SRSWOR from each level-2 unit. 200 replications.**

	designed value	R_1	mean estimate	relative bias*	relative RMS error*	$\frac{\text{bias}^2}{\text{MSE}}$
β_0	1.00	1.0 0.9 0.8 0.7	1.00 1.00 1.00 1.00	+0.00 +0.00 +0.00 +0.00	0.05 0.05 0.06 0.06	0.00 0.00 0.00 0.00
β_1	1.00	1.0 0.9 0.8 0.7	1.00 0.89 0.79 0.69	**+0.00** **−0.11** **−0.21** **−0.31**	0.02 0.11 0.21 0.31	0.00 0.95 0.99 0.99
β_2	1.00	1.0 0.9 0.8 0.7	1.00 1.00 1.01 1.01	+0.00 +0.00 +0.01 +0.01	0.02 0.02 0.03 0.03	0.00 0.03 0.07 0.11
β_3	0.30	1.0 0.9 0.8 0.7	0.28 0.35 0.41 0.46	**−0.05** **+0.18** **+0.37** **+0.54**	0.36 0.42 0.55 0.69	0.02 0.17 0.46 0.62
$\sigma_{00}^{(2)}$	0.063	1.0 0.9 0.8 0.7	0.059 0.063 0.069 0.078	**−0.07** **−0.00** **+0.10** **+0.24**	0.30 0.34 0.41 0.52	0.05 0.00 0.06 0.21
$\sigma_{00}^{(1)}$	0.30	1.0 0.9 0.8 0.7	0.30 0.40 0.50 0.59	**−0.00** **+0.33** **+0.66** **+0.98**	0.06 0.34 0.67 0.99	0.00 0.94 0.98 0.99

*as a proportion of the designed value

Here the effects on $\hat{\beta}_1$ and $\hat{\sigma}_{00}^{(1)}$ are similar to those observed for model (4.2), and again the effect on $\hat{\beta}_3$ is to increase it. But the interpretation in a model of progress would be different. This time the true context effect of $\mathbf{x}_3$ is positive, and the positive bias in $\hat{\beta}_3$ would exaggerate this effect. The residual between-school variance also would be exaggerated.

In the random-coefficient model (4.3), with error in the variable with the random coefficient, failure to adjust for this error caused the associated fixed-parameter estimate to be attenuated as in the other two models. See Table 4.23. The level-2 variance of this coefficient was attenuated more substantially, and this was accompanied by attenuation of the level-2 covariance of this coefficient with the intercept.

Thus the analyses in Chapter 4 confirmed that failure to adjust for random measurement error in explanatory variables at either level of a 2-level model results in biased estimators of their fixed coefficients, and associated biases in the random parameter estimators also. The effects on interpretation depend on the true values of the parameters. In these models the effects on the other fixed parameters were slight: the data in our examples were balanced, and the explanatory variables (apart from the aggregate variable in model 4.2) were uncorrelated.

We turn now to the substantive example in Chapter 5. For this data set, unlike those of Chapter 4, we did not have prior values of the measurement error variances and covariances. Thus, we reversed the process of Chapter 4 and used a series of assumed values of reliability of the reading score at age 8, together with known characteristics of the data, to derive a series of prior values for the error variances and covariances. In this way we explored the sensitivity of model inferences to different assumptions about the underlying measurement error. We also showed the effects of incomplete adjustment of two kinds. Analysis A ignored all measurement error at level 2; analysis B ignored the covariance between the errors in the observed individual age-8 scores and those in the school mean scores.

We found that adjustment for measurement error at level 2 was important chiefly in the estimation of parameters associated with level 2 (Tables 5.2 and 5.3), in particular the context effect β_3. We proposed two possible interpretations of the negative sign of this parameter, but interpretation is worthless if the estimation is faulty. Failure to adjust for the errors at level 2, as in Analysis A, caused the precision of $\hat{\beta}_3$ to be over-estimated. Incomplete adjustment at level 2, as in analysis B, led to over-estimating the magnitude of β_3. We stated in Chapter 5 that if the true aggregate score effect were positive (of like sign with the individual score effect), full adjustment would produce the larger estimate: this was confirmed in further simulations using model (6.1). The estimation and interpretation of such effects depend crucially on good estimates of the measurement error variances and covariances and on appropriate adjustment in analysis.

Some 25% of the pupils in the sample came from non-manual backgrounds. The mean observed reading score of these pupils at age 8 on the Normalised scale was +0.37. The mean observed score for pupils from other backgrounds was –0.13. Thus, there was a significant association between the variables $\mathbf{X}_1$ and $\mathbf{X}_2$. We found that as the assumed reliability of the prior attainment score $\mathbf{X}_1$ decreased, so the model estimate of the SES effect on progress also decreased, to vanish at $R_1 = 0.7$. Like that of the context effect, the estimate of the SES effect in this model was sensitive to assumptions about the measurement errors.

The models we have studied have all been simple ones. In more complex models we may expect more complex distortions to arise from failure to adjust appropriately for measurement error. It is important in any study to obtain suitable estimates of measurement error variances and covariances, and to investigate as we have done the effect of varying the estimates. It will also be important to establish that the estimation procedure is effective in reducing these distortions when the measurement error variances and covariances have been adequately specified.

The method of specification we defined in Section 3.6 is suitable for a wide variety of models, including models with multivariate response and cross-classified data. Moreover, in our development of the estimation algorithm, we have taken care to allow for the non-symmetric expected error product matrices that typically occur in these more complicated models. The vehicle that we have used for the implementation, namely the macro facility of ML*n*, is computationally inefficient and it has not so far been practicable to explore the functioning of these more complex parts of the estimation procedure in the analysis of data sets of suitable size. The results we have obtained with the simpler models are, however, encouraging, and fuller exploration is now proceeding.

7 Concluding remarks

Multilevel models are now used routinely in the analysis of educational and other social data, but adjustment for measurement error in such analysis continues to be the exception. We have demonstrated the sensitivity of conclusions to such adjustment and there is a clear need for suitable tools for this purpose.

Tools are needed both for measurement error estimation and for adjustment in subsequent analysis. As Muthén and Satorra (1989) showed, multilevel variation can affect model parameters in at least four ways (listed on pp26–7 of this thesis). Multilevel developments in structural modelling to date, however, appear to have been focused on the random group variation in the latent variable structure and in the measurement model. I have found no reference in the literature to the estimation of random regression parameters linking outcomes to latent predictors.

It is clearly important to continue to develop measurement error models. Different measurement processes call for different models, and the nature of the models available limits the choice of possible designs to take account of those processes. In this connection the developments in the framework of Bayesian analysis, using graphical models for formulation and stochastic simulation techniques (*Gibbs sampling*) for estimation, reported by Richardson (1996), are worthy of mention. Richardson demonstrated the flexibility of this approach in the context of epidemiology, at the same time pointing to the need for further research into the effects on regression results of misspecification. She did not address the multilevel problem specifically.

In any study, the estimation of a measurement error model requires either repeated measurements on a subgroup of subjects, or a 'gold standard', that is, a group of subjects in the study on whom accurate measurements have been made. Yet it is often the case that we wish to make inferences without either of these, and with only vague prior knowledge of the reliability of the measurements available. A case in point is the use of public examination results in models of the effectiveness of schools. Clearly, for some purposes, it is the results themselves that count and errors in their measurement are by definition zero. Where, however, interest lies in underlying capacities and patterns of development and the effects on these of different practices, or in comparing different modes of assessment leading to the same nominal qualifications, then the measures used will not be fully reliable indicators of the characteristics of interest. In the case of public examinations it is unlikely that suitable estimates of reliability will ever be available, partly because such estimates require special procedures at the time of examination that are costly and inconvenient, and partly because the high stakes attached to the results make it desirable to foster a general belief in their complete reliability. For studies using these results as measures of underlying characteristics it will be necessary to fall back on the kind of sensitivity analysis that we described in Chapter 5.

Such studies require efficient estimation procedures. Both the measurement models and the regressions of responses on covariates will typically be the subject of extensive exploration. Estimation methods relying on simulation, such as SIMEX, Markov Chain Monte Carlo (MCMC), and parametric bootstrap, are likely to be impracticable for this purpose for the foreseeable future. A fast procedure giving approximate results is what is needed for this phase, with validation of the final model(s) by another method, for example, parametric bootstrap.

The procedure described in this thesis, though computationally inefficient in its current implementation, is not inherently so, and the preliminary results that we have documented in Chapter 4 suggest that further development along the lines we have begun will yield an estimation method that is practically useful for multilevel model exploration in the presence of measurement error.

Much work remains to be done. It is necessary to verify that the proposed RIGLS estimators of the random parameters and the corrected standard error estimators for these parameters achieve satisfactory reductions in bias (see equations 3.118, 3.132, and 3.133). Once a more efficient routine is in place further development and more comprehensive testing will become practicable. For example, the procedure makes strong distributional assumptions (see p75) and it is necessary to test robustness in the face of violations of those assumptions. Also of high priority is to provide facilities for model diagnosis and comparison of units: as a minimum, estimates of true values and of residuals, and of the log-likelihood, will be developed.

If this work is successful it is hoped to incorporate it into the program *MLwiN* (Goldstein *et al.*, 1998). Thereafter it will be necessary to develop procedures for errors of misclassification, errors correlated with true values, constraints, and non-linear models of fixed and random parameters. There is a natural extension to the estimation of models with imputed values for randomly missing data and a prototype procedure exists for variance-components models (Goldstein and Woodhouse, 1998).

References

Biemer, P. and Trewin, D. (1997) A Review of Measurement Error Effects on the Analysis of Survey Data, in Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds) *Survey Measurement and Process Quality* (Chichester: Wiley), 603–32.

Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S. (eds) (1991) *Measurement Errors in Surveys* (Chichester: Wiley).

Bock, R.D. (ed.) (1989) *Multilevel Analysis of Educational Data* (London: Academic Press).

Bryk, A.S., Raudenbush, S.W., Seltzer, M., and Congdon, R. (1988) *An Introduction to HLM: Computer Program and User's Guide (2nd Edn)* (Chicago: University of Chicago Department of Education).

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models* (Newbury Park: Sage).

Carroll, R.J., Küchenhoff, H., Lombard, F., and Stefanski, L.A. (1996) Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models. *Journal of the American Statistical Association*, **91**, 242–50.

Carroll, R.J., Ruppert, D., and Stefanski, L.A. (1995) *Measurement Error in Nonlinear Models* (London: Chapman and Hall).

Cook, J.R. and Stefanski, L.A. (1994) Simulation–Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association*, **89**, 1314–28.

Crosswhite, F.J., Dossey, J.A., Swafford, J.O., McKnight, C.C., and Cooney, T.J. (1985) *Second International Mathematics Study: Summary Report for the United States* (Champaign: Stipes).

Deming, W.E. (1950) *Some Theory of Sampling* (New York: Wiley).

Degracie, J.S. and Fuller, W.A. (1972) Estimation of the Slopes and Analysis of Covariance when the Concomitant Variable is Measured with Error. *Journal of the American Statistical Association*, **67**, 930–7.

Ecob, R. and Goldstein, H. (1983) Instrumental Variable Methods for the Estimation of Test Score Reliability. *Journal of Educational Statistics*, **8**, 223–41.

Fuller, W.A. (1987) *Measurement Error Models* (Chichester: Wiley).

Fuller, W.A. (1991) Regression Estimation in the Presence of Measurement Error, in Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S. (eds) *Measurement Errors in Surveys* (Chichester: Wiley), 617-35.

Fuller, W.A. (1995) Estimation in the Presence of Measurement Error: the second Morris Hansen Lecture. *International Statistical Review*, **63**, *2*, 121–47.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (eds) (1996) *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall).

Goldstein, H. (1979) *The Design and Analysis of Longitudinal Studies* (London: Academic Press).

Goldstein, H. (1986) Multilevel Mixed Linear Model Analysis using Iterative Generalized Least Squares. *Biometrika*, **73**, 43–56.

Goldstein, H (1995) *Multilevel Statistical Models* (London: Arnold).

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G., and Healy, M. (eds) (1998) *A User's Guide To MLwiN* (London: Multilevel Models Project, Institute of Education, University of London).

Goldstein, H. and Woodhouse, G.M. (1998) Efficient Estimation with Missing Data in Multilevel Models. (To appear).

Groves, R. (1991) Measurement Error across Disciplines, in Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S. (eds) *Measurement Errors in Surveys* (Chichester: Wiley), 1–25.

Handlin, O. (1979) *Truth in History* (Cambridge, MA: Belknap).

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1961) Measurement Errors in Censuses and Surveys. *Bulletin of the International Statistical Institute*, **38**, *2*, 359–74.

Hidiroglou, M.A., Fuller, W.A., and Hickman, R.D. (1980) *SUPER CARP* (Ames, Iowa: Department of Statistics, Iowa State University).

Jöreskog, K.G. (1970) A General Method for Analysis of Covariance Structures. *Biometrika*, **57**, 239–51.

Kiaer, A.N. (1897) *The Representative Method of Statistical Surveys* (Oslo: Kristiana).

Longford, N.T. (1991) Searching for Multivariate Outcomes in Education, in Raudenbush, S.W. and Willms, J.D. (1991) *Schools, Classrooms, and Pupils* (London: Academic Press), 115–30.

Longford (1993a) *Random Coefficient Models* (Oxford: Clarendon Press).

Longford, N.T. (1993b) Regression analysis of multilevel data with measurement error. *British Journal of Mathematical and Statistical Psychology*, **46**, 301–11.

Longford, N.T. and Muthén, B.O. (1992) Factor Analysis for Clustered Observations. *Psychometrika*, **57**, 581–97.

Lord, F.M. and Novick, M.R. (1968) *Statistical Theories of Mental Test Scores* (London: Addison–Wesley).

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds) (1997) *Survey Measurement and Process Quality* (Chichester: Wiley).

Mahalonobis, P.C. (1946) Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, **109**, 326–78.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988) *School Matters: the Junior Years* (Wells: Open Books).

Moser, C.A. and Kalton, G. (1971) *Survey Methods in Social Investigation, 2nd Edn* (London: Heinemann Educational).

Muthén, B.O. (1997) Latent Variable Modeling of Longitudinal and Multilevel Data, in Raftery, A. (ed.) *Sociological Methodology 1997* (Boston: Blackwell), 453–80.

Muthén, B.O. and Satorra, A. (1989) Multilevel Aspects of Varying Parameters in Structural Models, in Bock, R.D. (ed.) *Multilevel Analysis of Educational Data* (London: Academic Press), 87–99.

O'Muircheartaigh, C. (1997) Measurement Error in Surveys: A Historical Perspective, in Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds) *Survey Measurement and Process Quality* (Chichester: Wiley), 1–25.

Paterson, L. and Goldstein, H. (1991) New Statistical Methods for Analysing Social Structures: An Introduction to Multilevel Models. *British Educational Research Journal*, **17**, *4*, 387–93.

Plewis, I. (1985) *Analysing Change: Measurement and Explanation using Longitudinal Data* (New York: Wiley).

Prosser, M., Rasbash, J., and Goldstein, H. (1991) *ML3 Software for Three-level Analysis: User's Guide for V.2* (London: Institute of Education, University of London).

Raftery, A. (ed.) (1997) *Sociological Methodology 1997* (Boston: Blackwell).

Rasbash, J. and Goldstein, H. (1994) Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model. *Journal of Educational and Behavioural Statistics*, **19**, 337–50.

Rasbash, J. and Woodhouse, G.M. (1996) *MLn Command Reference* (London: Multilevel Models Project, Institute of Education, University of London).

Raudenbush, S.W., Rowan, B., and Kang, S.J. (1991) A Multilevel, Multivariate Model for Studying School Climate with Estimation via the EM Algorithm and Application to U.S. High-School Data. *Journal of Educational Statistics*, **16**, 295–330.

Raudenbush, S.W. and Willms, J.D. (eds, 1991) *Schools, Classrooms, and Pupils* (London: Academic Press).

Richardson, S. (1996) Measurement Error, in Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (eds) *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall), 401–17.

Rowan, B., Raudenbush, S.W., and Kang, S.J. (1991) School Climate in Secondary Schools, in Raudenbush, S.W. and Willms, J.D. (eds) *Schools, Classrooms, and Pupils* (London: Academic Press).

Searle, S.R. (1982) *Matrix Algebra Useful for Statistics* (New York: Wiley).

Searle, S.R., Casella, G., and McCulloch, C.E. (1992) *Variance Components* (New York: Wiley).

Shewhart, W.A. (1939) *Statistical Method from the Viewpoint of Quality Control* (Graduate School, U.S. Department of Agriculture).

Sutcliffe, J.P. (1965) A Probability Model for Errors of Classification. I: General Considerations. *Psychometrika*, **30**, 73–96.

Trilling, L. (1951) Reality in America, in Trilling, L. (ed.) (1951) *The Liberal Imagination* (New York: Viking).

Woodhouse, G.M., Rasbash, J., Goldstein, H., and Yang, M. (1996a) Introduction to Multilevel Modelling, in Woodhouse, G. (ed.) *Multilevel Modelling Applications* (London: Institute of Education, University of London).

Woodhouse, G.M., Yang, M., Goldstein, H., and Rasbash, J.R. (1996b) Adjusting for Measurement Error in Multilevel Analysis, *Journal of the Royal Statistical Society, Series A*, **159**, *2*, 201–12.

Appendix Derivation of the covariance matrix $\operatorname{cov}\left(\mathbf{e}^{**}\right)$

Consider the linear model

$$\mathbf{y}^{**} = \mathbf{z}^{**}\boldsymbol{\theta} + \mathbf{e}^{**}, \tag{A.1}$$

where $\mathbf{z}^{**}{}_{(N^2 \times H)}$ is the design matrix, based upon the true values $\mathbf{z}$, for the random parameters, and $\mathbf{e}^{**}$ is an $N^2 \times 1$ vector of residuals. Equation (A.1) is equation (3.96) repeated.

We now derive an expression for $\operatorname{cov}\left(\mathbf{e}^{**}\right)$ in terms of $\mathbf{v}$. It is desirable to place the development of the measurement error adjustment for the random parameters on a firm foundation, and in particular to clarify the role of the Normality assumptions. This derivation is believed to be new.

We have:

$$\begin{aligned} \operatorname{cov}\left(\mathbf{e}^{**}\right) &= \mathrm{E}\left[(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})^{\mathrm{T}}\right] - \left[\mathrm{E}(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})\right]\left[\mathrm{E}(\tilde{\mathbf{y}} \otimes \tilde{\mathbf{y}})\right]^{\mathrm{T}} \\ &= \mathrm{E}\left(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}} \otimes \tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}\right) - \operatorname{vec}\mathbf{v}\left(\operatorname{vec}\mathbf{v}\right)^{\mathrm{T}}. \end{aligned} \tag{A.2}$$

Each $N^2 \times N^2$ matrix in (A.2) has a natural partitioning into N^2 submatrices, each submatrix being $N \times N$. For $\ell = 1, 2, \ldots, L$, $s, s' \in \{p, p+1, \ldots, p+q-1\}$, $i_1, i_2 \in \{1, 2, \ldots, N\}$, define

$$\begin{aligned} \sigma_{ss'}^{(\ell)}(i_1, i_2) &\equiv \sigma_{ss'}^{(\ell)} \text{ if } \exists\, j,\ 1 \le j \le J_\ell, \text{ such that } i_1, i_2 \in B_j^{(\ell)}, \\ &\quad 0 \text{ otherwise.} \end{aligned} \tag{A.3}$$

Then, for $i_1, i_2, i_3, i_4 \in \{1, 2, \ldots, N\}$, the (i_3, i_4) th element of the (i_1, i_2) th submatrix of $\mathrm{vec}\,\mathbf{v}(\mathrm{vec}\,\mathbf{v})^{\mathrm{T}}$, which we shall call the (i_1, i_2, i_3, i_4) th element, is

$$\sum_{\ell=1}^{L} \sum_{s,t=p}^{p+q-1} z_{s,i_1} z_{t,i_3} \sigma_{st}^{(\ell)}(i_1, i_3) \sum_{\ell'=1}^{L} \sum_{s',t'=p}^{p+q-1} z_{s',i_2} z_{t',i_4} \sigma_{s't'}^{(\ell')}(i_2, i_4) \tag{A.4}$$

and the corresponding element of $\mathrm{E}\left(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}} \otimes \tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}\right)$ is

$$\mathrm{E}\left[\sum_{\ell_1,\ell_2=1}^{L} \sum_{s,s'=p}^{p+q-1} z_{s,i_1} \varepsilon_{s,i_1}^{(\ell_1)} z_{s',i_2} \varepsilon_{s',i_2}^{(\ell_2)} \sum_{\ell_3,\ell_4=1}^{L} \sum_{t,t'=p}^{p+q-1} z_{t,i_3} \varepsilon_{t,i_3}^{(\ell_3)} z_{t',i_4} \varepsilon_{t',i_4}^{(\ell_4)}\right]. \tag{A.5}$$

Consider the case where i_1, i_2, i_3, i_4 are all different. Then for given s, s', t, t' the coefficient of $z_{s,i_1} z_{s',i_2} z_{t,i_3} z_{t',i_4}$ in the (i_1, i_2, i_3, i_4) th element of $\mathrm{vec}\,\mathbf{v}(\mathrm{vec}\,\mathbf{v})^{\mathrm{T}}$ is

$$\sum_{\ell,\ell'=1}^{L} \sigma_{st}^{(\ell)}(i_1, i_3) \sigma_{s't'}^{(\ell')}(i_2, i_4) \tag{A.6}$$

and the corresponding coefficient in $\mathrm{E}\left(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}} \otimes \tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathrm{T}}\right)$ is

$$\sum_{\ell_1,\ell_2,\ell_3,\ell_4=1}^{L} \mathrm{E}\left(\varepsilon_{s,i_1}^{(\ell_1)} \varepsilon_{s',i_2}^{(\ell_2)} \varepsilon_{t,i_3}^{(\ell_3)} \varepsilon_{t',i_4}^{(\ell_4)}\right). \tag{A.7}$$

From the assumption that the random variables $\boldsymbol{\varepsilon}$ have zero expectation and do not correlate across different levels, the expectations in (A.7) are non-zero only if the levels $\ell_1, \ell_2, \ell_3, \ell_4$ are either all equal or equal in pairs. Consider the terms in which they are equal in pairs to either ℓ or ℓ', where ℓ, ℓ' are given and $\ell \neq \ell'$. These terms have the following sum:

$$\mathrm{E}\begin{pmatrix} \varepsilon_{s,i_1}^{(\ell)}\varepsilon_{s',i_2}^{(\ell)}\varepsilon_{t,i_3}^{(\ell')}\varepsilon_{t',i_4}^{(\ell')} + \varepsilon_{s,i_1}^{(\ell)}\varepsilon_{s',i_2}^{(\ell')}\varepsilon_{t,i_3}^{(\ell)}\varepsilon_{t',i_4}^{(\ell')} + \varepsilon_{s,i_1}^{(\ell')}\varepsilon_{s',i_2}^{(\ell)}\varepsilon_{t,i_3}^{(\ell)}\varepsilon_{t',i_4}^{(\ell')} \\ +\varepsilon_{s,i_1}^{(\ell)}\varepsilon_{s',i_2}^{(\ell')}\varepsilon_{t,i_3}^{(\ell')}\varepsilon_{t',i_4}^{(\ell)} + \varepsilon_{s,i_1}^{(\ell')}\varepsilon_{s',i_2}^{(\ell)}\varepsilon_{t,i_3}^{(\ell')}\varepsilon_{t',i_4}^{(\ell)} + \varepsilon_{s,i_1}^{(\ell')}\varepsilon_{s',i_2}^{(\ell')}\varepsilon_{t,i_3}^{(\ell)}\varepsilon_{t',i_4}^{(\ell)} \end{pmatrix} \tag{A.8}$$

$$= \sigma_{ss'}^{(\ell)}(i_1,i_2)\sigma_{tt'}^{(\ell')}(i_3,i_4) + \sigma_{st}^{(\ell)}(i_1,i_3)\sigma_{s't'}^{(\ell')}(i_2,i_4) + \sigma_{s't}^{(\ell)}(i_2,i_3)\sigma_{st'}^{(\ell')}(i_1,i_4)$$
$$+ \sigma_{st'}^{(\ell)}(i_1,i_4)\sigma_{s't}^{(\ell')}(i_2,i_3) + \sigma_{s't'}^{(\ell)}(i_2,i_4)\sigma_{st}^{(\ell')}(i_1,i_3) + \sigma_{tt'}^{(\ell)}(i_3,i_4)\sigma_{ss'}^{(\ell')}(i_1,i_2).$$

The terms in the corresponding coefficient in $\operatorname{vec}\mathbf{v}(\operatorname{vec}\mathbf{v})^{\mathrm{T}}$ which arise from the same pair of levels ℓ,ℓ' have the sum

$$\sigma_{st}^{(\ell)}(i_1,i_3)\sigma_{s't'}^{(\ell')}(i_2,i_4) + \sigma_{s't'}^{(\ell)}(i_2,i_4)\sigma_{st}^{(\ell')}(i_1,i_3). \tag{A.9}$$

Now consider the term in (A.7) for which $\ell_1 = \ell_2 = \ell_3 = \ell_4 = \ell$:

$$\mathrm{E}\left(\varepsilon_{s,i_1}^{(\ell)}\varepsilon_{s',i_2}^{(\ell)}\varepsilon_{t,i_3}^{(\ell)}\varepsilon_{t',i_4}^{(\ell)}\right) \tag{A.10}$$
$$= \mathrm{E}\left(\varepsilon_{s,i_1}^{(\ell)}\varepsilon_{s',i_2}^{(\ell)}\right)\mathrm{E}\left(\varepsilon_{t,i_3}^{(\ell)}\varepsilon_{t',i_4}^{(\ell)}\right) + \mathrm{E}\left(\varepsilon_{s,i_1}^{(\ell)}\varepsilon_{t,i_3}^{(\ell)}\right)\mathrm{E}\left(\varepsilon_{s',i_2}^{(\ell)}\varepsilon_{t',i_4}^{(\ell)}\right) + \mathrm{E}\left(\varepsilon_{s,i_1}^{(\ell)}\varepsilon_{t',i_4}^{(\ell)}\right)\mathrm{E}\left(\varepsilon_{s',i_2}^{(\ell)}\varepsilon_{t,i_3}^{(\ell)}\right)$$
$$= \sigma_{ss'}^{(\ell)}(i_1,i_2)\sigma_{tt'}^{(\ell)}(i_3,i_4) + \sigma_{st}^{(\ell)}(i_1,i_3)\sigma_{s't'}^{(\ell)}(i_2,i_4) + \sigma_{st'}^{(\ell)}(i_1,i_4)\sigma_{s't}^{(\ell)}(i_2,i_3),$$

by the Normality assumption 3.4. The corresponding term in (A.6) is

$$\sigma_{st}^{(\ell)}(i_1,i_3)\sigma_{s't'}^{(\ell)}(i_2,i_4). \tag{A.11}$$

It follows from (A.1) that if i_1,i_2,i_3,i_4 are all different the (i_3,i_4)th element of the (i_1,i_2)th submatrix of $\operatorname{cov}(\mathbf{e}^{**})$ is

$$\sum_{\ell,\ell'=1}^{L}\sum_{s,s',t,t'=p}^{p+q-1}\begin{bmatrix} z_{s,i_1}z_{s',i_2}\sigma_{ss'}^{(\ell)}(i_1,i_2)z_{t,i_3}z_{t',i_4}\sigma_{tt'}^{(\ell')}(i_3,i_4) \\ +z_{s',i_2}z_{t,i_3}\sigma_{s't}^{(\ell)}(i_2,i_3)z_{s,i_1}z_{t',i_4}\sigma_{st'}^{(\ell')}(i_1,i_4) \end{bmatrix}. \tag{A.12}$$

If i_1,i_2,i_3,i_4 are not all different suppose, for example, that $i_3 = i_4$. Then for given s,s',t,t' the coefficient of $z_{s,i_1}z_{s',i_2}z_{t,i_3}z_{t',i_4}$ will combine with that for $z_{s,i_1}z_{s',i_2}z_{t',i_3}z_{t,i_4}$, but we can if we wish identify them separately. Such separate identification is possible for all cases of equality between two or more of i_1, i_2, i_3, i_4 and therefore the expression (A.12) holds for the (i_1, i_2, i_3, i_4)th element of $\mathrm{cov}(\mathbf{e}^{**})$ for all $i_1, i_2, i_3, i_4 \in \{1, 2, \ldots, N\}$. This element is equal to the sum of the (i_1, i_2, i_3, i_4)th and (i_1, i_4, i_2, i_3)th elements of $\mathbf{v} \otimes \mathbf{v}$ and thus

$$\mathrm{cov}(\mathbf{e}^{**}) = (\mathbf{I} + \mathbf{s}_N)(\mathbf{v} \otimes \mathbf{v}), \qquad \text{(A.13)}$$

where $\mathbf{s}_N$ is the vec permutation matrix of order N^2.