RR – 85-02

RANDOM COEFFICIENT REGRESSION MODELS
FOR
MULTILEVEL ANALYSIS

Jan de Leeuw
Department of Data Theory
State University of Leiden
Ita Kreft
Department of Education
University of Amsterdam

Abstract

We propose a possible statistical model for both contextual analysis and slopes-as-outcomes analysis. These techniques have been used in multilevel analysis for quite some time, but a precise specification of the regression models has not been given before. Following Tate and Wongbudhit, we propose a random coefficient regression model, and investigate its statistical properties in some detail. Various estimation methods are reviewed, and applied to a Dutch school-career example.

Keywords

multilevel analysis, contextual analysis, random coefficient models

1. INTRODUCTION

In a recent paper Tate and Wongbundhit (1983) have argued that random coefficient regresssion models are more appropriate for multilevel analysis in educational research than fixed coefficient models. We briefly summarize their argument, which consists of four steps. In the first place within-group regressions can reflect important aspects of the multilevel mechanism. It is quite conceivable, for instance, that in some schools the regression of success on intelligence is steeper than in others, and that this steepness reflects policies, strategies, or ideologies that differentiate schools. This first step in the argument is also the starting point of the 'slopes as outcomes' analyses used by Burstein and his associates, which will be reviewed briefly below. The second step in the Tate-Wongbundhit argument is, that we can expect a great deal of variation in the within-group regressions, not only because of the policies and strategies mentioned above, but also because of a large number of differences between groups which are more difficult to isolate. Thirdly it is common practice to use random variability ('disturbances' or 'errors') to 'explain' variations that are not modelled explicitly. And, fourthly and finally, working with incompletely specified models inevitably leads to a loss of efficiency in the estimates. 'We agree with the argument that data from many educational settings are generated by random coefficient processes. Therefore, we also believe that statistical inference should be based on the same kind of model.' (Tate and Wongbundhit, 1983, page 107).

We shall illustrate the argument of Tate and Wongbundhit by analyzing a number of more specific models and techniques that have been proposed in the multilevel literature, and that seem to require random coefficient regession techniques. The first instance is the general contextual model discussed most completely by Boyd and Iversen (1979, especially chapter III). Boyd and Iversen systematically distinguish the <u>single equation approach</u> to contextual analysis from the <u>separate equations approach</u>. The separate equations approach is the more basic one. There are two types of equations in the contextual model. The first one specifies an individual-level within-group regression model, one for each separate group. The second set of equations relates the within-group regression coefficients to contextual variables describing the groups. These contextual variables are often within-group averages of individual level variables, but this is by no means necessary. We are not concerned here with general theoretical and methodological aspects of the contextual model. These aspects are reviewed admirably by Boyd and Iversen (1979) and by Blalock (1984). We concentrate on the statistical aspects of the model, a subject which is somewhat neglected. 'Unfortunately, Boyd and Iversen did not consider the question of statistical inference.' (Tate and Wongbundhit, 1983, page 107).

There is one very basic problem with the separate equations approach. This is the question what is exactly modelled in the second set of equations. There are two possible answers, based on two different assumptions. Either the regression coefficients in the within-group models are fixed parameters, or they are random variables. If they are fixed parameters, then they are estimated by ordinary within-group regression analysis. The estimates of the within-group regression coefficients, which must be distinguished from the regression coefficients themselves, are again random variables. In the second modelling step, or in the second set of equations, we can model the distribution of the estimates. But then we must remember, of course, that this distribution is already determined to a large extent by the assumptions and the calculations in the first step. If we assume directly that the within-group regression coefficients are random variables, then the above remains true. We still must distinguish the regression coefficients from their estimates, where the notion of 'estimation' is now extended to cover estimation of random variables. In fact it is clear that fixed parameters are merely a 'degenerate' special case of random coefficients. A basic problem with the contextual analysis literature is that the choice between fixed and random coeffcient models is never made explicit. Boyd and Iversen (1979, for instance section 3.2) write their equations as if they are thinking of random coefficient models. Their later discussion of the disturbance terms in the single equation approach (1979, page 55) also suggests this. But their estimation procedure is ordinary unweighted least squares for both sets of coefficients, which ignores the information provided by the random coefficient model.

A similar incomplete specification is apparent in a recent interesting paper by Van den Eeden and Saris (1984). They analyze school-career data collected in the Netherlands by a 'two-step' approach. The adjective 'two-step' has two different meanings. In the first place the model is specified by two sets of equations, the first one within-schools at the individual level and the second one between-schools at the school-level. But the approach is also called 'two-step', because the estimation is done by ordinary least squares for both sets of equations separately. In fact the most important methodological conclusion of Van den Eeden and Saris is that their two-step procedure is preferable to a one-step procedure which combines the equations into a single equation and then estimates all parameters jointly by ordinary least squares. We shall comment on this conclusion in a later section of the paper, at this point we merely remark that Van den Eeden and Saris also do not specify if their within-group regression coefficients are random variables of fixed constants. Or, to put it differently, they do not make explicit assumptions about the behaviour of the disturbance terms in the between-schools equation, and they act as if the usual linear model assumptions are true at both stages. In fact they even use standard errors associated with the usual linear model in the second step. We agree with Tate and Wongbundhit (1983) that incomplete specification usually leads to loss of efficiency. Moreover, in the case of Van den Eeden and Saris, use of ordinary least squares standard errors in the second stage is not only inefficient, it is in fact wrong. We shall illustrate this below, by analyzing the same school career data in a different and theoretically more satisfactory way.

For completeness we emphasize that Boyd and Iversen are certainly aware of the problems associated with combining the two sets of equations into a single one. In their appendix B (1979, pages 232-233) they discuss a weighted regession procedure for the second stage, which incorporates weights for the variances of the within-group regressions. Their discussion suggests a fixed coefficient model, in which the second stage provides a model for the estimates of the within-group regressions. In their appendix C (1979, pages 234-236) they discuss conditions under which separate equations and single equation ordinary least squares give the same estimates. In practice both estimates will be quite different, and Boyd and Iversen do not give explicit criteria which can be used to choose between the two.

There is another class of multilevel models in which random coefficients seen necessary. These are the 'slopes as outcomes' analyses of Burstein and his associates. The most important references are Burstein (1976, 1980a, 1980b, 1981), Burstein and Linn (1976), Burstein, Linn, and Capell (1978), Burstein and Miller (1981), and Burstein, Miller, and Linn (1981). These papers concentrate on motivation and interpretation of the results if within-group slopes are used in between-group regression analysis. Again we are not concerned with the theoretical and methodological reasons for adopting this approach, and with its usefulness in educational contexts. For this we refer to the literature we cited. We restrict ourselves to the statistical problems, which are, again, largely ignored by others. The fact that there are some non-standard problems is acknowledged by Burstein, Miller, and Linn (1981). 'The mathematical properties of slopes as outcomes are not well understood. We are essentially treating the within-group slopes as a random variable with an unknown underlying distribution function. ... The criticism that within-group slopes should not be treated as random variables is troubling, but certainly not fatal. There are too many instances in behavioural reserach where sensible analytical work has been conducted without mathematical confirmation of the appropriateness of the distributional assumptions in the measurement of a critical variable.' (1981, page 19). It seems to us that the last part of the quotation is unduly pessimistic. If we make a complete specification of the model, along the lines already indicated above, then the problems with random or fixed coefficients merely become questions of correct or incorrect specification which can, at least in principle, be investigated by standard statistical methods.

It is somewhat disappointing that Tate and Wongbundhit (1983), who seem to have a clear understanding of the problems involved, merely contribute a Monte Carlo study to show that some techniques can be misleading if a random coefficient model is true. The same thing is true for Burstein, Linn, and Capell (1978), who argue quite convincingly for the importance of assuming heterogeneous within-group slopes, but then illustrate their point by a (very) small scale Monte Carlo study. This is all quite unnecessary, because in the situations studied by these authors analytical results can easily be derived. We must emphasize, however, that the models studied by Tate and Wongbundhit (1983) and by Burstein, Linn, and Capell (1978) are more general than the models we intend to discuss. Our models have random coefficients but fixed regressors. The more general models have both random coefficients and random regressors. Of course the additional variation in the regressors introduces additional complications, that we do not want to go into in this paper. We do not want to belittle the distinction, however. It is quite important, because random regressor models seem more natural in many situations in educational research. The more general class of models also makes it possible to fit in multilevel analysis more smoothly into standard structural equation modelling practice. We hope to come back to this more general class of models at a later occasion.

In the next section of the paper we shall introduce a fairly general random coefficient regression model, with fixed regressors, which can be used in multilevel analysis to unify and extent many results obtained by contextual analysis or 'slopes as outcomes' approaches. Section 3 briefly reviews the history of this model, and relates it to other models that have been proposed before, mainly in econometrics. Section 4 discusses ordinary least squares estimation, and compares the 'one-step' and 'two-step' methods more in detail. In section 5 we introduce a class of weighted least squares estimates, and in section 6 the more complicated maximum likelihood methods are derived and discussed. The Dutch school-career data, analyzed earlier by Van den Eeden and Saris, are used to illustrate the various techniques in section 7. Section 8 concludes.

2. MODEL

The model is specified in two sets of equations, one within-groups, and one between-groups. We suppose there are m groups, n_j observations in group j, and p within-group fixed regressors. The measurements for group j on the p regressors are collected in the $n_j \times p$ fixed matrix X_j, the measurements on the dependent variable in the n_j-element vector $\underline{y}_j$. In this paper we use the convention to underline random variables (Hemelrijk, 1966). In this context, in which to be fixed or to be random is the question, such a convention is especially convenient. The model for group j is

$$\underline{y}_j = X_j \underline{\beta}_j + \underline{\varepsilon}_j. \tag{1}$$

Here $\underline{\beta}_j$ is the p-vector of within-group regression coefficients, and $\underline{\varepsilon}_j$ is the n_j-vector of disturbances. We assume, for the disturbances,

$$E(\underline{\varepsilon}_j) = 0, \tag{2a}$$

$$E(\underline{\varepsilon}_j \underline{\varepsilon}_j') = \sigma_j^2 I. \tag{2b}$$

Thus we have a standard linear model for each group, except for the fact that the $\underline{\beta}_j$ are supposed to be random vectors. Their properties are specified in the next set of equations. The m groups are characterized by q regressors, with values in the $m \times q$ matrix Z. If we collect the mp random variables $\underline{\beta}_{js}$ (j=1,...,m and s=1,...,p) in the random $m \times p$ matrix $\underline{B}$, then the second stage model is

$$\underline{B} = Z\Theta + \underline{\Delta}. \tag{3}$$

Here Θ is a $q \times p$ matrix of fixed regression coefficients, and $\underline{\Delta}$ is an $m \times p$ matrix of random disturbances. It is sometimes more convenient to write (3) as

$$\underline{\beta}_j = \Theta' z_j + \underline{\delta}_j, \tag{4}$$

where z_j contains the q elements of row j of Z, and $\underline{\delta}_j$ contains the p elements of row j of $\underline{\Delta}$. In addition we assume that

$$E(\underline{\delta}_j) = 0, \tag{5a}$$

$$E(\underline{\delta}_j \underline{\delta}_\ell') = 0 \quad \text{if } j \neq \ell, \tag{5b}$$

$$E(\underline{\delta}_j \underline{\delta}_j') = \Omega, \tag{5c}$$

$$E(\underline{\delta}_j \underline{\varepsilon}_\ell') = 0. \tag{5d}$$

Equations (1)-(5) define our basic model. The parameters are the matrices

Ω and Θ and the m values of σ_j^2. For the model to make sense the matrix Ω should be positive semi-definite.

The within-group and between-group components of the model can be combined into a single individual-level model. Substituting (4) in (1) gives

$$\underline{y}_j = X_j\Theta'z_j + \underline{\nu}_j, \tag{6}$$

with

$$\underline{\nu}_j = X_j\underline{\delta}_j + \underline{\varepsilon}_j. \tag{7}$$

Thus

$$E(\underline{\nu}_j) = 0, \tag{8a}$$

$$E(\underline{\nu}_j\underline{\nu}_\ell') = 0 \text{ if } j \neq \ell, \tag{8b}$$

$$E(\underline{\nu}_j\underline{\nu}_j') = X_j\Omega X_j' + \sigma_j^2 I. \tag{8c}$$

Model (6)-(8) is, in a sense, the reduced form of model (1)-(5). The unobservable variables $\underline{\beta}_j$ are eliminated from the model, and we find the fixed coefficient regression model (6), with correlated errors according to (8c). We can finally eliminate the disturbances, and write the model entirely in terms of observables. For this purpose it is convenient to define

$$\mu_j = X_j\Theta'z_j, \tag{9}$$

$$\Xi_j = X_j\Omega X_j' + \sigma_j^2 I. \tag{10}$$

Then our model says that

$$E(\underline{y}_j) = \mu_j, \tag{11a}$$

$$E(\underline{y}_j\underline{y}_\ell') = 0 \text{ if } j \neq \ell, \tag{11b}$$

$$E(\underline{y}_j\underline{y}_j') = \Xi_j. \tag{11c}$$

For some purposes we assume, in addition, that the $\underline{y}_j$ are jointly multivariate normal, but we do not need this assumption for most of our discussion. We shall indicate clearly where we do make it.

It is also clear that the fixed coefficient model is the special case in which $\underline{\Delta} \equiv 0$, i.e. in which $\Omega = 0$ and $\Xi_j = \sigma_j^2 I$. In the 'slopes as outcomes' approach, and also in the usual contextual models, X_j has only two columns of which the first one is identically equal to +1. The two elements of $\underline{\beta}_j$ are the (random) intercept and the (random) slope. If we want to include mixed models, in which for instance intercepts are fixed but slopes are random, then this can be done by requiring certain elements of Ω and/or Θ to be zero. We shall consider such restricted versions of our general model to be additional specifications, which can be tested within our general model. Another additional specification, which is also of some importance, is that the σ_j^2 are the same. We have already indicated that a basic limitation of our model is that all regressors (both Z and the X_j) are fixed. Another limitation is that we only deal with a single dependent variable $\underline{y}$. This last restriction of generality could easily be removed, but this would merely complicate the notation without giving essentially different results from the univariate case.

The interpretation of our model in the multi-level context is clear, because it is a straightforward generalization of the general contextual model of Boyd and Iversen (1979), which it merely makes more explicit and more precise. In the same way it generalizes the 'slopes as outcomes' approach, showing in what sense regression coefficients 'are' random variables. We shall see below, that our estimation procedures generalize the one-step and two-step procedures of Boyd and Iversen (1979) and Van den Eeden and Saris (1984). In fact they generalize them, correct them where necessary, and put them on a more solid statistical basis. But first we shall indicate that our model is far from new, and has been studied in great detail already in the econometric and statistical literature.

3. SOME HISTORY

Random coefficient models, or more generally variable coefficient models, have a long history in econometrics. There was pioneering work by Rubin, Klein, Wald, and Theil in the late forties and early fifties, but this had little practical impact and was ignored for some time. More comprehensive papers, oriented towards practical applications, were written in the late sixties by Rao, Fisk, Hildreth and Houk, and Swamy. The pre-1970 literature is reviewed almost completely in the monograph by Swamy (1971). In the seventies a substantial body of theory was developed, and a number of useful review papers appeared. We mention Rosenberg (1973), Spjøtvoll (1977), Mundlak (1978). Chapter 17 in the book by Maddala (1977), and the very recent chapter by Chow (1984) in the Handbook of Econometrics are also very useful. Annotated bibliographies have been published by Johnson (1977, 1980).

Although these econometric papers often discuss models which are more general than our (1)-(5) in some respects, they are usually less general in one important respect. In the second stage specification (4) most econometric models use $q = 1$ and $z_{j1} = 1$ for all j. This means that we can rewrite (4) more simply as $\underline{\beta}_j = \theta + \underline{\delta}_j$, and thus there effectively is no second stage model of independent interest. This makes the standard random coefficient models as such quite useless for multilevel research, although there are some exceptions. The first exception is Hanushek (1974). He only considers the case $p = 1$, but for this case he presents a two-stage model which is very similar to our model. Unfortunately Hanushek does not distinguish random and fixed variables clearly in his paper, and as a consequence the statistical analysis of his model is confused. Another two-stage model has been proposed by Amemiya (1978), in the context of pooling cross-section and time-series data. The model assume $\underline{y}_j = X_j\underline{\beta}_j + \underline{\varepsilon}_j$, as we do, but the second-stage model is $\underline{\beta}_j = Z_j\theta + \underline{\delta}_j$, which is quite different. The difference arises, of course, because the models are designed for different types of applications. The assumptions on the distributions of the disturbances used by Amemiya are, again, different from our assumptions. A two-stage model very similar to Amemiya's has been studied very recently by Pfefferman (1984). Pfefferman works in the Gauss-Markov framework, and supposes that the dispersions of the disturbances are essentially known.

The fact that the random coefficient models in econometrics are either not specific enough, or just a little bit different, need not bother us at all. The estimators that have been proposed in the literature can be adapted without too much trouble to our two-stage multilevel model, and this is exactly what we shall do in the sequel. Moreover there are many results in statistics that deal with general mixed linear models. They can be used for our model too. And finally our two-stage models are closely related to Bayesian and empirical Bayes methods for the linear model. These results are discussed most completely in Lindley and Smith (1972), and in the contributions of the discussants of that paper.

There are also two important developments in the econometric random coefficient literature that we have not incorporated in our model, although they could very well be useful in multilevel research. The first one, already discussed in connection with Burstein, Linn, and Capell (1978), is the use of random regressors. In a basic paper Mundlak (1978) discusses random regressor-random coefficient models in which there may be 'transmitted errors', i.e. correlations between coefficients and independent random variables. Our second omission, somewhat less serious perhaps, is the modelling of the first-level error variances as random variables as well. This could be useful in 'residual-as-outcome' research. Models which allow for random variances are discussed by Aragon (1985). Both extensions of our basic model would have led us into many complications, and into largely uncharted territory.

4. LEAST SQUARES ESTIMATION

In this section we discuss various aspects of least squares estimation in our model (5). The first aspect is 'estimation' of the $\underline{\beta}_j$. We must realize, of course, that we estimate a random variable here, and not a fixed constant. Nevertheless the notions of bias and variance apply to estimation of random variables as well. Gauss-Markov theory for random coefficient models was developed by Rao (1965a), compare also section 4a.11 of Rao (1965b), by Swamy (1970, 1971), and by Pfefferman (1984). The relevant result for our model is that the minimum variance unbiased linear estimate of $\underline{\beta}_j$ is $\hat{b}_j = (X_j'X_j)^{-1}X_j'\underline{y}_j$. The expectation of $\hat{b}_j$ is $E(\hat{b}_j) = \Theta' z_j$, and its variance is $\Omega + \sigma_j^2(X_j'X_j)^{-1}$. If $\underline{r}_j = \underline{y}_j - X_j\hat{b}_j$ is the residual, then $E(\underline{r}_j) = 0$, and

$$E(\underline{r}_j\underline{r}_j') = \sigma_j^2(I - X_j(X_j'X_j)^{-1}X_j'). \tag{12}$$

It follows that $E(\underline{r}_j'\underline{r}_j) = \sigma_j^2(n_j - p)$, and thus $\hat{\sigma}_j^2 = \underline{r}_j'\underline{r}_j/(n_j - p)$ is unbiased for σ^2. This is the first step of our estimation process, which gives us unbiased estimates of the random variables $\underline{\beta}_j$ and the error variances σ_j^2.

In the second step we compute a preliminary estimate of Θ. This is done by collecting the $\hat{b}_j$ in an m x p matrix $\hat{B}$. This matrix has expected value $Z\Theta$, and thus $(Z'Z)^{-1}Z'\hat{B}$ is unbiased for Θ. This is the usual two-step estimate computed in contextual analysis. Our second step ends with the computation of an estimate for Ω. This requires some thinking, because the unbiased estimates developed in Rao (1965a) and Swamy (1970) will not work for our model. They are based on the econometric one-step model in which $E(\hat{b}_j) = E(\underline{\beta}_j) = \theta$ for all j. But if we compute the matrix of residuals $\hat{B} - Z\Theta$,

and define $\Phi = I - Z(Z'Z)^{-1}Z'$, then

$$E(\hat{B}'\Phi\hat{B}) = (m - q)\Omega + \sum_{j=1}^{m} \phi_{jj}\sigma_j^2(X_j'X_j)^{-1}. \tag{13}$$

It follows that, using ϕ_{jj} for the diagonal elements of Φ,

$$\hat{\Omega} = (m - q)^{-1}\{\hat{B}'\Phi\hat{B} - \sum_{j=1}^{m} \phi_{jj}\hat{\sigma}_j^2(X_j'X_j)^{-1}\} \tag{14}$$

is unbiased. Thus we can estimate the variance of the dispersion on both levels from the least squares residuals. It is somewhat unfortunate that estimate (14) may not be positive semi-definite. Compare the discussion in Swamy (1971, page 107-111).

After two steps we have unbiased estimates of all parameters. We conclude this section by considering an alternative way of computing least squares estimates of Θ. For this purpose it is convenient to write out model in a more compact form. We define an n x pq matrix U in the following way. U consists of matrices U_{jr}, with j=1,...,m and r=1,...,q. Each U_{jr} is n_j x p. The U_{jr} with the same first index are next to each other in U, and the U_{jr} with the same second index are below each other. We have $U_{jr} = z_{jr}X_j$. We also string out Θ to a pq element vector θ, by placing the q rows of Θ on top of each other in a single column. The $\underline{y}_j$ are collected in an n-vector $\underline{y}$, and the $\underline{\nu}_j$ in an n-vector $\underline{\nu}$. Then (6) becomes

$$\underline{y} = U\theta + \underline{\nu}, \tag{15}$$

and (8) can be written simply as

$$E(\underline{\nu}) = 0, \tag{16a}$$

$$E(\underline{\nu}\underline{\nu}') = V. \tag{16b}$$

Here V is n x n and block-diagonal, with the $V_j = X_j\Omega X_j' + \sigma_j^2 I$ as diagonal blocks. It is possible to show that the two-step least squares estimate $\hat{\theta}$ developed above is of the form $\hat{\theta} = U^{-}\underline{y}$, where U^{-} is a generalized inverse of U that fully utilizes its block structure. As Boyd and Iversen have pointed out (1979, page 53-55) it is also possible to study the single-step (or single-equation) estimate of θ. In our notation this is $\hat{\theta} = (U'U)^{-1}U'\underline{y} = U^{+}\underline{y}$, with U^{+} the Moore-Penrose inverse of U. Observe that we use the same symbol for the single-equation estimate, although it generally differs from the separate equation estimate (Boyd and Iversen, 1979, appendix C). It is also unbiased, but it has some disadvantages. We must invert a much larger matrix, which could very well be ill-conditioned, and we do not get auxilary results which make it possible to estimate the remaining parameters of the model. Thus separate equations estimates seem preferable to single equation estimates. The same conclusion is reached by Van den Eeden and Saris (1984, page 176-178), although their arguments are quite different. They also point to multicollinearity of single equation estimates, but they point out in addition that separate equations estimates give results which are easier to interpret. We agree with this evaluation. We do <u>not</u> agree with their argument that separate equations estimates are bothered less by non-homogeneous variance of the disturbances. We have shown above that the $\hat{b}_j$ of the separate equations procedure are non-homogeneous as well. Thus using ordinary least squares as both Boyd and Iversen and Van den Eeden and Saris do is suboptimal in terms of variance. Moreover computing standard errors of the estimated second stage regressions by convential formulae, as Van den Eeden and Saris do, is incorrect. It is also not correct, as Van den Eeden and Saris suggest, that the single equation procedure leads to biased estimates. Because the disturbances have zero expectation, unbiasedness is guaranteed. The difference is mainly one of precision, computational ease, and interpretability.

Tate and Wongbundhit (1983) also reach the conclusion that the procedures they have compared (single equation, separate equations, and mixed) all produce unbiased estimates. It is somewhat peculiar to use a Monte Carlo study to confirm this. If the theory directly shows unbiasedness, and a Monte Carlo study does not confirm this, then the result merely indicates that either computational errors have been made or the random number generator did not work properly. In Tate and Wongbundhit no bias was found, so their random number generator worked correctly. There are some curious passages in their paper on 'actual' and 'adjusted' separate equations procedures, with the 'actual' procedure being badly biased. Since the estimates must be unbiased, by our results above, we conjecture that Tate and Wongbundhit have made a systematic error in their computations. In their basic tables III and IV they compare computed standard errors over Monte Carlo replications with expected values of the estimate ordinary least squares standard errors. The actual SE's are interesting, although they indicate very little difference between the procedures. The apparent SE's are not very interesting, because they are all theoretically in error. There seems to be little point in computing quantities that are known to be wrong. As we have indicated above Tate and Wongbundhit could have computed the theoretical expected values and dispersions of their estimates very easily, given the parameter values in table I. The trip to Monte Carlo was entirely unnecessary and, in fact, a bit misleading. It must be emphasized that in the random coefficient-random regressor model of Tate and Wongbundhit the regressors were independent of the coefficients. Thus there are no 'transmitted errors' in the sense of Mundlak (1978), and the model is rather close to a fixed regressor model.

5. WEIGHTED LEAST SQUARES

In the previous section we have discussed both separate equations and single equation least squares methods. We have seen that the single equation method has little to recommend it, and that the separate equations method is preferable from a computational and interpretational point of view. In this section, and in the next section, we shall develop procedures which are more satisfactory from a statistical point of view, and that maintain the interpretational advantages of the separate equations method.

From (15) and (16) we know, using the Gauss-Markov theorem, that the best linear unbiased estimate of θ is given by

$$\hat{\theta} = (U'V^{-1}U)^{-1}U'V^{-1}\underline{y}. \tag{17}$$

This result, as such, is quite useless, because V is generally unknown. It has been suggested by Swamy (1970, 1971) to substitute the estimates $\hat{\sigma}_j^2$ and $\hat{\Omega}$ computed in the previous section in the definition of V. This gives an estimated $\hat{V}$. We then compute

$$\hat{\theta} = (U'\hat{V}^{-1}U)^{-1}U'\hat{V}^{-1}\underline{y}. \tag{18}$$

This is, of course, a natural idea. Because estimates are no longer linear in the observations the simple calculus of bias does not apply any more, and we have to resort to asymptotic methods to evaluate our estimates. Before we do this, we first simplify the expression (18) for the weighted least squares estimates.

Remember that V is the direct sum of m matrices $V_j = X_j\Omega X_j' + \sigma_j^2 I$. Thus V^{-1} is the direct sum of m matrices V_j^{-1}. But (Swamy, 1971, page 101)

$$V_j^{-1} = \sigma_j^{-2}(I - X_j(X_j'X_j)^{-1}X_j') + X_j(X_j'X_j)^{-1}\{\Omega + \sigma_j^2(X_j'X_j)^{-1}\}^{-1}(X_j'X_j)^{-1}X_j'. \tag{19}$$

The matrix $U'V^{-1}V$ can be thought of as being build up from q^2 matrices of dimension p x p. Submatrix $(U'V^{-1}U)_{rs}$ has the form

$$(U'V^{-1}U)_{rs} = \sum_{j=1}^{m} z_{jr}z_{js}\{\Omega + \sigma_j^2(X_j'X_j)^{-1}\}^{-1}, \tag{20a}$$

and subvector $(U'V^{-1}\underline{y})_r$ has the form

$$(U'V^{-1}\underline{y})_r = \sum_{j=1}^{m} z_{jr}\{\Omega + \sigma_j^2(X_j'X_j)^{-1}\}^{-1}\hat{b}_j. \tag{20b}$$

This can be written more elegantly by using Kronecker products. If

$W_j = \Omega + \sigma_j^2(X_j'X_j)^{-1}$, and W is the direct sum of the W_j (i.e. the matrix with the W_j as diagonal blocks, and zero everywhere else), then we can rewrite (17) as

$$\hat{\theta} = \{(I \otimes Z)'W^{-1}(I \otimes Z)\}^{-1}(I \otimes Z)'W^{-1}\hat{b}, \tag{21}$$

with $\otimes$ the Kronecker product. Of course (18) can be written in exactly the same way, with $\hat{W}$ substituted for W. We can also derive (21) immediately from the fact that $\hat{b}$, interpreted as a vector of length mp, satisfies

$$E(\hat{b}) = (I \otimes Z)\theta, \tag{22a}$$

$$V(\hat{b}) = W. \tag{22b}$$

These results were already derived, in a slightly different notation, in section 4. A further relation with section 4 becomes apparent when we write the separate equations unweighted least squares estimate as

$$\hat{\theta} = \{(I \otimes Z)'(I \otimes Z)\}^{-1}(I \otimes Z)'\hat{b}. \tag{23}$$

Comparing (21) and (23) suggests that perhaps the separate equations procedure is closer to the optimal weighted procedure than the single equation least squares method, which cannot be written in a comparable form. The optimal procedure is, in a sense, also a two-step procedure, because it first computes $\hat{b}$ and the regresses this on $I \otimes Z$.

Again we emphasize that the development above generalizes that of Rao, Swamy, and others, because they only study the simple case in which the design matrix Z consists of a single column of ones. Then (21), for instance, simplifies to

$$\hat{\theta} = \{\sum_{j=1}^{m} W_j^{-1}\}^{-1} \sum_{j=1}^{m} W_j^{-1}\hat{b}_j, \tag{24}$$

which shows that in this case $\hat{\theta}$ is a 'matrix-weighted average' of the $\hat{b}_j$. The asymptotics for studying $\hat{\theta}$ has also been worked out only for this interesting, but highly restricted special case.

A careful study in the asymptotics of weighted least squares for the restricted model has been published by Johansen (1982, compare also 1984, chapter IV). We shall not adapt all his results, but merely discuss his improvement of an earlier result of Swamy (1970). This assumes, in the original form, that both m and the n_j tend to infinity. The matrices $n_j^{-1} X_j'X_j$ and $m^{-1}Z'Z$ also tend to limits. Let C be the limit of $m^{-1}Z'Z$. Then Swamy's result, translated to our context, says that $m^{\frac{1}{2}}(\hat{\theta} - \theta)$ is asymptotically normal with mean zero and dispersion $C^{-1} \otimes \Omega$.

Johansen (1982) improves the conditions under which the theorem holds, but also points out that the result may not be very satisfactory in some situations. In fact under Swamy's conditions the weighted least squares estimate has the same asymptotic distribution as the unweighted separate equations estimate, and thus (asymptotically at least) there was no reason to weight in the first place. Johansen proves a much more general result, which allows for an asymptotic effect of the weights. The result depends critically, however, on assume Gaussian disturbances, and it is not very easy to apply. Thus we do not discuss it in detail, and we do not try to extend it to our multilevel model, although this can in principle be done.

If we summarize the developments in this chapter we think that the weighted estimate of will generally improve upon the unweighted estimate, although this is by no means a certainty. Weighting will introduce some bias in small samples, with a small number of groups, but will presumably improve the precision. The asymptotic behaviour of weighted and unweighted estimates, for a large number of groups, depends on the relative speed with which m and the n_j tend to their limits. It seems clear that what we need is expansions, not limit theorems, to make more definite statements.

6. MAXIMUM LIKELIHOOD

Maximum likelihood estimation methods for mixed analysis of variance models were first discussed systematically by Hartley and Rao (1967). Recent state-of-the-art reviews are by Harville (1977) and Thomson (1980). Compare also Rao and Kleffe (1980). Recent computational developments are often based on the EM-algorithm of Dempster, Laird, and Rubin (1977). Applications of this algorithm to various classes of mixed ANOVA problems are outlined in Dempster, Rubin, and Tsutakawa (1981), Rubin and Szatrowski (1982), Laird and Ware (1982), and Adrade and Helms (1984). We have computed maximum likelihood estimates for our model with a special version of the EM-algorithm, but we shall not discuss it in detail because we have strong reservations about its computational efficiency. Alternative estimates for the dispersions of the residuals, at both stages, could be based on Rao's MINQUE theory, which is reviewed by Rao (1979), Kleffe (1980), Rao and Kleffe (1980). We merely note this here, we do not apply MINQUE to our random coefficient model. For the possibilities we refer to the dissertations of Streitberg (1977) and Infante (1978).

We have seen in the previous two sections that the most natural unweighted least squares method, and the weighted least squares method both worked in two computational steps. In the first step within-class regressions were computed by ordinary least squares, together with the within class residuals. In the second step the within-class regressions were used as dependent variables for the between-class analysis. This is an important property, because it implies that in the second step we do not work with the original $\underline{y}_j$ and X_j any more, but with much smaller aggregations. This makes computation in the second step relatively cheap. In this section we show that a similar aggregation mechanism works in the case of maximum likelihood estimation. Although this method will generally be much more complicated than the least squares methods, it does share this basic simplifying property with them.

For the method of maximum likelihood to apply, we must make explicit distributional assumptions. Thus we assume multivariate normality here. We emphasize, however, that the estimates also make sense if multinormality is not satisfied. In that case they are 'quasimaximum-likelihood', but they may still be quite good. The form of the log-likelihood function follows directly from (9)(10)(11), together with multivariate normality. We ignore some irrelevant constants, and we conclude that we must minimize the function

$$L(\Theta,\Omega,\Sigma) = \sum_{j=1}^{m} \ln \det(X_j\Omega X_j' + \sigma_j^2 I) + (\underline{y}_j - X_j\Theta z_j)'(X_j\Omega X_j' + \sigma_j^2 I)^{-1}(\underline{y}_j - X_j\Theta z_j).$$

We now need some simplifications to get the order of the matrices down to the aggregated level. The first simplification results from formula (19). It follows from (19) that

$$(\underline{y}_j - X_j\Theta z_j)'(X_j\Omega X_j' + \sigma_j^2 I)^{-1}(\underline{y}_j - X_j\Theta z_j) =$$

$$(n_j - p)\hat{\sigma}_j^2/\sigma_j^2 + (\hat{b}_j - \Theta z_j)'(\Omega + \sigma_j^2(X_j'X_j)^{-1})^{-1}(\hat{b}_j - \Theta z_j). \qquad (26)$$

This brings the order of the matrices down from n_j to p, which is of course a large gain. It also shows how the first stage within-class regression statistics are used. For the log-determinant in (25) we use the result

$$\ln \det(X_j\Omega X_j' + \sigma_j^2 I) = \ln \det (X_j'X_j) + (n_j - p) \ln \sigma_j^2 + \ln \det(\Omega + \sigma_j^2(X_j'X_j)^{-1}).$$

Again this is highly satisfactory, for the same reasons. If we combine (25) and (27) we see that we must minimize the sum of m terms of the form

$$(n_j - p)(\ln \sigma_j^2 + \hat{\sigma}_j^2/\sigma_j^2) + \ln \det(\Omega + \sigma_j^2(X_j'X_j)^{-1}) +$$
$$+ (\hat{b}_j - \Theta z_j)'(\Omega + \sigma_j^2(X_j'X_j)^{-1})^{-1}(\hat{b}_j - \Theta z_j). \qquad (28)$$

For fixed values of the dispersion parameters σ_j^2 and Ω the optimal Θ is computed easily by (21). Computation of the optimal dispersion parameters is a bit more complicated, and we do not have a completely satisfactory algorithm yet. The estimates used in the example in the next section we computed by a version of the EM-method, but this turned out to be very expensive indeed.

The asymptotic properties of maximum likelihood estimates in mixed analysis of variance situations, which include our random coefficient model as a special case, have been investigated most thoroughly by Miller (1977). He proves consistency, asymptotic normality and efficiency of the maximum likelihood estimates by using an increasing sequence of designs (bothe the number of schools and the number of pupils in the schools tend to infinity). As we have already indicated in the discussion of the weighted least squares estimates it is not entirely clear which particular form of asymptotics we need in multilevel situations. Most of the results seem a bit contrived, and it is probably safe practice to use Monte Carlo methods next to asymptotic results as long as satisfactory expansions are not available. Of course there is no need to use Monte Carlo if exact results are available, as in the one-step and two-step ordinary least squares case with fixed regressors.

7. A SCHOOL CAREER EXAMPLE

We illustrate some aspects of the techniques developed in this paper by analyzing the GALO-data, described by Peschar (1975), and analyzed previously with multilevel analysis by Van den Eeden and Saris (1984) and Dronkers and Schijf (1984). The GALO-data contain information about primary school leavers in the city of Groningen during 1959 and 1960. Following Dronkers and Schijf, we selected 30 schools which had pupils in both the 1959 and the 1960 cohort. For each pupil the individual-level independent variables we used were sex, IQ, and occupational level of the father. The individual-level dependent variable is teachers advice on secondary education. Thus our example had $p = 4$ (constant term, SEX, IQ, SES) and $m = 60$ (30 schools in 2 cohorts). The total number of pupils was 2058, and on the average there are approximately 35 pupils in each school-year combination (the actual n_j varied between 15 and 68). IQ was coded as a 'continuous' variable, it had values between 58 and 148. fathers occupation had six possible values, and teachers advice had seven. Optimal scaling techniques have indicated that integer scoring of the categories leads to regressions which do not deviate too much from linearity. Thus we have treated occupation and advice as numerical variables, although this is clearly not the most rational way to proceed. For more information on this topic we refer to Meester and De Leeuw (1983). They perform the optimal scalings and analyse the GALO-data by techniques appropriate for categorical variables. Their analysis is single-level, because they aggregate over schools and cohorts first. We complete our description of the data by describing the independent variables for the school-level analysis. We used the fact that the 30 schools were situated in 10 neighborhoods. This was coded by using 10 dummy variables. The year (1959 or 1960) was used as the additional group-level variable, thus $q = 11$. Our choice of variables differs considerably from that of Van den Eeden and Saris, who use $p = q = 2$, and have IQ and average IQ as only predictors.

Our first analysis step is to perform the 60 within-class regressions, which estimate $\hat{b}_j$ and $\hat{\sigma}_j^2$. All regressions were done in raw score form. There is an enormous variation both in the regression coefficients and in the residual variances, and the second step in the analysis is certainly needed to model at least some of this variation.We intend to publish interpretational details elsewhere. It suffices to say here, that advice was somewhat higher in the second cohort, that SEX and SES were somewhat less important in the second cohort, that intelligence was a very predominant factor in forming advice and did not vary much in importance over the neighborhoods or years. Neighborhood of a school did not predict the regression coefficient of SEX or SES in the individual equations very well, the corresponding elements of Θ did not deviate much from zero. As a general rule the results of the second-stage analysis were difficult to relate to neighborhood-characteristics such as quality of housing, percentage of manual workers, and average income. Of course this may be due to contamination with the individual variable SES.

In this paper we are more interested, however, in some technical results. We have computed both one-step and two-step ordinary least squares estimates of Θ, a matrix which contains pq = 44 elements in this example. The most interesting result is, perhaps, that the correlation between the two matrices of estimates is .9978. Deviations are usually very small, and impossible to relate to external information. For all practical purposes the two sets of estimates are the same. If we try to estimate the standard errors of the two sets of estimates we do find some systematic differences. Two-step unweighted least squares is about 5 % more efficient that one-step. In order to compute these standard errors we must have estimates of Ω and σ_j^2, of course. They were computed by using (12) and (14). We can use the same estimates of the dispersions to compute the weighted least squares estimates. We find that weighted least squares estimates of Θ correlate .9991 with two-step unweighted least squares estimates, and are 10% more efficient. Again the second stage regression coefficients cannot be distinguished from the earlier solutions, but there is slight gain in precision indicated. It is not directly clear that the extra computation is worth its while. We have also computed the maximum likelihood estimates by our ad-hoc EM-algorithm. They correlate .9999 with the weighted least squares estimates. Of course the maximum likelihood method also gives us estimates of the dispersions. We first compare the estimates of σ_j^2 with those computed from the least squares residuals. The maximum likelihood estimates are, as a rule, a bit smaller. But not much, because the average of the 60 ratios $\hat{\sigma}^2_{ML}/\hat{\sigma}^2_{LS}$ is .9932. Again the differences are very small, and insignificant for interpretational purposes. It is not true that everything is the same for the two solutions, however. The log-likelihood function decreases from 1700 to 1500 in about 10 iteration cycles, which is a considerable improvement from least squares to maximum likelihood. The estimate of Ω changes a great deal. We have compute the eigenvalues of $\hat{\Omega}_{ML}^{-1}\hat{\Omega}_{LS}$, and they are 3.26, 2.33, 1.01, and -0.22. This shows that the least squares estimate has a small negative eigenvalue, and that it is much larger than the maximum likelihood estimate in some directions. This instability of Ω is the only possible reason why one should want to do maximum likelihood in this example, the other parameters are almost perfectly stable with respect to estimation methods choice. Because the maximum likelihood method, as currently implemented, is about 50 times as expensive for this example, we think that the additional computation was not worthwhile. Of course this may be due to a peculiarity of this example, and the conclusion certainly depends on the algorithm that is used. On the whole we think that our analysis of the GALO-example shows that combining the two-step unweighted least squares method with the weighted least squares method is by far the best overall method. Because the weighted least squares method needs the statistics computed in the unweighted analysis they combine very naturally. In fact they can be interpreted as the first two steps in an iterative procedure, in which the residuals from step k are used to improve the variance estimates, which are then used to define new weights for step k + 1. Similar iterative estimation methods are discussed, for example, by Streitberg (1977), who relates them to maximum likelihood and MINQUE theory.

8. CONCLUSIONS AND RECOMMENDATIONS

Our first, and perhaps most important, recommendation is that if one uses contextual analysis, or 'slopes-as-outcomes' analysis, then one should try to specify the statistical model as completely as possible. This does not necessarily mean that one must adopt the specification we have investigated here, there are indeed many other possibilities. In fact it seems to us that our model, although it is certainly a step ahead, is not quite general enough. It must be generalized in such a way that it can deal with recursive causal models, in which there are several dependent variables and in which the regressors are random. Moreover for many school-career analysis situations it must have provisions for incorporating categorical variables. These seem to be the developments that are needed from the modelling point of view.

From the algorithmic point of view we have seen that a better method to compute maximum likelihood estimates is certainly needed. Otherwise the (possible) statistical advantages of the method will never be large enough to offset the compuattional disadvantages, and that would be a pity. Computation of weighted and unweighted least squares estimates is simple enough, although the simplicity vanishes rapidly if we generalize our fixed regressor model in the way indicated above. In the fixed regressor model there seem to be no reasons why one would want to use the one-step method.

In a statistical sense our methods are still far from complete. If we assume multivariate normality, we can derive the exact sampling distribution of the unweighted least squares estimates. But, of course, there is hardly any situation in educational research in which the assumption of multivariate normality applies. If we drop it, we have to use asymptotic results. It is not clear yet, what the precise properties of weighted least squares and maximum likelihood estimates are, even asymptotically. This must be investigated in the future. Another possibility, that we have not mentioned at all so far, is that tests of hypotheses can be carried out in various ways. In our model we can be interested in the hypotheses that $\Omega = 0$, for instance, or that the σ_j^2 are equal, that some rows of Θ are zero, and so on. Because we have concentrated on estimation, and not on testing and interpretation, we have developed these possibilities, but they seem indispensable for a more satisfactory data analysis.

Although a lot of work remains to be done, our most important conclusion is that the fixed-regressor random-coefficient model we have studied seems an interesting specification of contextual analysis models, and that the Rao-Swamy-Johansen weighted least squares method is an excellent method to estimate the unknown parameters of the model.

REFERENCE NOTES

Burstein, L. (1976), Assessing the difference of between-group and individual regression coefficients, Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Burstein, L. (1981), Explanatory models using between and within class regression: basic concepts and an example, Paper presented at the Second International Mathematics Study Data Analysis Workshop, Toronto, Canada.

Burstein, L., Miller, M.D., and Linn, R.L. (1979), The use of within-group slopes as indices of group outcomes, Los Angeles, Center for the Study of Evaluation, UCLA.

Burstein, L. & Linn, R.L. (1976), Detecting the effects of education in the analysis of mutlilevel data: the problem of hetrogeneous within-class regressions, Paper presented at the Conference on Methodology for Aggregating Data in Educational Research, Stanford, California.

Dronkers, J. & Schijf, H. (1984), Neighborhoods, schools and individual educational attainment; a better model for analyzing unequal educational opportunities, Paper presented at the conference on Cultural Basis of Education, Paris, France.

Infante, A. (1978), Die MINQUE-Schätzung bei Verlaufskurvenmodellen mit zufälligen Regressionskoeffizienten, Unpublished dissertaion, University of Dortmund, FRG.

Meester, A.C. & De Leeuw, J. (1983), Intelligentie, sociaal milieu, en de schoolloopbaan, Leiden, The Netherlands, Department of Data Theory.

Streitberg, B. (1977), Schätzung von Kovarianzstrukturen in linearen Zwei- und Mehrebenenmodellen, Unpublished dissertataion, Free University of Berlin, FRG.

REFERENCES

Andrade, D.F., & Helms, R.W. (1984), Maximum likelihood estimates in the mutlivariate normal with patterned mean and covariance via the EM algorithm, Communications in Statistics, 13, 2239-2251.

Amemiya, T. (1978), A note on a random coefficients model, International Economic Review, 19, 793-796.

Aragon, Y. (1984), Random variance linear models: estimation, Computational Statistics Quarterly (in press).

Blalock, H.M. (1984), Contextual-effects models: theoretical and methodological issues, Annual Review of Sociology, 10, 353-372.

Boyd, L.H. & Iversen, G.R. (1977), Contextual analysis: concepts and statistical techniques, Belmont, California, Wadsworth Publishing Company.

Burstein, L. (1980a), The role of levels of analysis in the specification of educational effects, in R. Dreeben & J.A. Thomas (eds), The analysis of educational productivity, volume I: Issues in micro analysis, Cambridge, Ballinger.

Burstein, L. (1980b), The analysis of multilevel data in educational research and evaluation, in D. Berliner (ed), Review of research in education, volume 8, Washington, American Educational Research Association.

Burstein, L., Linn, R.L., & Capell, F.J. (1978), Analyzing multilevel data in the presence of heterogeneous within-class regressions, Journal of Educational Statistics, 3, 347-383.

Burstein, L. & Miller, M.D. (1981), Regression-based analysis of multilevel educational data. In R.F. Boruch, P.M. Wortman, D.S. Cordray, a.o. (eds), Reanalyzing Program Evaluations, San Francisco, Jossey-Bass.

Chow, G.C. (1984), Random and changing coefficient models. In Z. Griliches & M.D. Intriligator (eds), Handbook of Econometrics, volume II, Amsterdam, North Holland Publishing Company.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977), Maximum likelihood from incomplete data using the EM algorithm, Journal of the Royal Statistical Society, B39, 1-38.

Dempster, A.P., Rubin, D.R., & Tsutakawa, R.K. (1981), Estimation in covariance components models, Journal of the American Statistical Association, 76, 341-353.

Hanushek, E.A. (1974), Efficient estimators for regressing regression coefficients, The American Statistician, 28, 66-67.

Hartley, H.O. & Rao, J.N.K. (1967), Maximum likelihood analysis for the mixed analysis of variance model, Biometrika, 54, 93-108.

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

Hemelrijk, J. (1966), Underlining random variables, Statistica Neerlandica, 20, 1-8.

Johansen, S. (1982), Asymptotic inference in random coefficient regression models, Scandinavian Journal of Statistics, 9, 201-207.

Johansen, S. (1984), Functional relations, random coefficients, and nonlinear regression, with application to kinetic data, New York, Springer Verlag.

Johnson, L.W. (1977), Stochastic parameter regression: an annotated bibliography, International Statistical Review, 45, 257-272.

Johnson, L.W. (1980), Stochastic parameter regression: an additional annotated bibliography, International Statistical Review, 48, 95-102.

Kleffe, J. (1980), On recent progress of MINQUE theory: nonnegative estimation, consistency, asymptotic normality and explicit formulae, Mathematische Operationsforschung und Statistik, 11, 563-588.

Laird, N.M. & Ware, J.H. (1982), Random-effects models for longitudinal data, Biometrics, 38, 963-974.

Lindley, D.V. & Smith, A.F.M. (1972), Bayes estimates for the linear model, Journal of the Royal Statistical Society, B34, 1-41.

Maddala, G.S. (1977), Econometrics, New York, McGraw-Hill.

Miller, J.J. (1977), Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance, Annals of Statistics, 5, 746-762.

Mundlak, Y. (1978), Models with variable coefficients: integration and extension, Annales de l'INSEE, 30-31, 483-509.

Peschar, J. (1975), Milieu, school, beroep, Groningen, Tjeenk Willink.

Pfefferman, D. (1984), On extensions of the Gauss-Markov theorem to the case of stochastic regression coefficients, Journal of the Royal Statistical Society, B46, 139-148.

Rao, C.R. (1965a), The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, Biometrika, 52, 447-458.

Rao, C.R. (1965b), Linear statistical inference and its applications, New York, Wiley.

Rao, C.R. (1979), MINQE theory and its relation to ML and MML estimation of variance components, Sankhya, B41, 138-153.

Rao, C.R. & Kleffe, J. (1980), Estimation of variance components. In P.R. Krishnaiah (ed), Handbook of Statistics, volume I, Amsterdam, North Holland Publishing Company.

Rosenberg, B. (1973), A survey of stochastic parameter regression, Annals of Economic and Social measurement, 2, 381- 397.

Rubin, D.B. & Szatrowski, T.H. (1982), Finding maximimum likelihood estimates of patterned covariance matrices by the EM algorithm, Biometrika, 69, 657-660.

Spjøtvoll, E. (1977), Random coefficients regression models: a review, Mathematische Operationsforschung und Statistik, 8, 69-93.

Swamy, P.A.V.B. (1970), Efficient inference in a random coefficient regression model, Econometrica, 38, 311-323.

Swamy, P.A.V.B. (1971), Statistical inference in random coefficients regression models, New York, Springer-Verlag.

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

Thomson, R. (1980), Maximum likelihood estimation of variance components, Mathematische Operationsforschung und Statistik, 11, 545- 561.

Van den Eeden, P. & Saris, W.E. (1984), Empirisch onderzoek naar multilevel uitspraken, Mens en Maatschappij, 59, 165-178.