Title
Random Coefficient Models for Multilevel Analysis

Authors
Jan de Leeuw
Ita Kreft

Publication Date
2011-10-25

RANDOM COEFFICIENT MODELS FOR MULTILEVEL ANALYSIS

JAN DE LEEUW AND ITA KREFT

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. We formalize them by proposing a random coefficient regression model, and we investigate its statistical properties in some detail. Various estimation methods are reviewed and applied to a Dutch school-career example.

This paper was published previously in *Journal of Educational Statistics*, 11, 1986, 57–85. I corrected some typos, otherwise it's a faithful reproduction.

In recent years there has been an increasing awareness that many, if not most, problems in educational research have multilevel characteristics (Burstein, 1980b, Oosthoek & Van den Eeden, 1984). To make this statement a bit more precise, we introduce some terminology.

Date: August 26, 2006.

Key words and phrases. Multilevel analysis, contextual analysis, random coefficient models.

We thank the members of the Multilevel Group of NOSMO for useful comments, and we thank Nick Longford, Murray Aitkin, Harvey Goldstein, Russell Ecob, and Pieter van den Eeden for sending us preprints of unpublished papers.

A variable is a function, mapping a domain of units into a range of values. In educational research, for instance, the units can be pupils, classes, schools, school districts, national educational systems, and so on. Thus, we can also have variables of different levels, describing pupils, classes, and so on. Observe that in this example, and in many others, the units of the various domains are nested. Schools consist of classes, classes of pupils, and so on.

In a multilevel problem we want to investigate the relations between variables with different domains. A currently popular example is school effectiveness research (Averch, Carroll, Donaldson, Kiesling, & Pincus, 1974; Brookover, Flood, Schweiser, & Wisenbaker, 1979; Dreeben & Thomas, 1980; Purkey & Smith, 1983), in which we study the relationship between school characteristics and pupil achievements. A little reflection will show that teacher style research, classroom climate research, and other types of research also involve variables of different levels.

The multilevel character of many educational research problems has implications of a general methodological nature. In sociology, relating micro-variables and macro-variables and developing forms of cross-level inference have always been acknowledged as a fundamental problem (Blalock, 1979; Lazarsfeld & Menzel, 1961; Van den Eeden, 1985a). On the other hand, the multilevel characteristics of the data also have implications for statistical modeling and analysis. In this paper we will be mainly concerned with these statistical aspects of the problem.

Earlier statistical approaches have attempted to adapt uni-level techniques to multilevel situations. This can often be done by using aggregation or disaggregation. A pupil variable, such as intelligence, can be aggregated to school level by assigning to a school the average intelligence of its pupils. A school variable, such as whether it is public or private, can be disaggregated to pupil level by assigning to each pupil the type of school. By giving all variables the same domain, we can simply use classical correlation or regression techniques. Or so it seems. But the operations of aggregation and disaggregation are highly nontrivial, both from the methodological and from the statistical point of view. A change in the meaning of the variables occurs. In addition, by aggregating, we eliminate all within-school variation, we have to deal with the Robinson effect [1], and we cannot make inferences on the pupil level any more without committing the ecological fallacy (Alker, 1969; Hannan, 1971; Robinson, 1950). If we disaggregate, then we have to take into account the fact that pupils within the same school do not vary independently on disaggregated school variables on which they have, by definition, the same value.

The general outcome of the discussion (Langbein, 1977) seems to be that the effect of aggregation or disaggregation, or of any other statistical operation for that matter, can only be studied precisely within the context of a plausible statistical model. Within such a model cross-level inference becomes possible, precisely because

[1] The Robinson effect is the often dramatic increase of correlation between variables after aggregation. The ecological fallacy is the tendency to interpret correlations between aggregated variables as if they were correlations between variables measured on individuals.

some parts of the model refer to schools while other parts refer to pupils. The natural model for generalization to multilevel situations is, of course, the linear model familiar from regression analysis and analysis of variance.

In this paper we will try to review these linear model extensions and present them in a unified way. Similar attempts at integration have been made by Mason, Wong, and Entwisle (1984) and by Aitkin and Longford (in press). For simplicity of presentation, we restrict ourselves to the case of just two levels, pupils and schools. Extensions of our discussion to three or more levels are fairly straightforward (Goldstein, in press; Longford, in press).

1. Regressing Regression Coefficients

The basic idea of multilevel linear models is very simple. There is a micro-model, defined separately for each macro-unit. This is a linear model, with pupil-level regressors or predictors, and with a pupil-level dependent variable. Each school has its own model. The macro-model relates the parameters of the micro-models, which are the regression coefficients and the error variances, to macro-level regressors. Thus, within-school regression coefficients are regressed on school variables in the macro-model.

This general approach is already quite old. According to Mason, Wong, and Entwisle (1984), "Although its origins are uncertain, the notion of a regression in which the dependent variable consists of regression coefficients from other regressions has long been attractive to social scientists and statisticians" (p. 73). But even within this general idea, a number of specific choices have to be made.

The most important choice, for our purposes, is whether we want to model the regression coefficients in the micro-models as random variables or as fixed constants.

Tate and Wongbundhit (1983) argued that random coefficient regression models are more appropriate than fixed coefficient models for multilevel analysis in educational research. We briefly summarize their argument, which consists of four steps. First, 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 degree of 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" analysis used by Burstein and his associates, which will be reviewed later in this section.

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 other differences between schools, which are more difficult to isolate. Third, it is common practice to use random variability (disturbances or errors) to “explain” variations that are not modeled explicitly. Fourth, working with incompletely specified models inevitably leads to a loss of efficiency in the estimates. To quote Tate and Wongbundhit

> 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" (p. 107).

It is possible to add more arguments to this list. Random coefficient models are more general, because fixed constants are special random variables. Whether something is random or fixed should be decided by considering what would happen if we replicated the experiment. Would it be realistic to suppose that the regression coefficients stayed the same under replication? If not, then random coefficients are appropriate. It is also possible to think of the distribution of the random coefficients as a prior distribution. This line of reasoning shows that Bayesian or empirical-Bayesian approaches lead naturally to random coefficient models.

We will illustrate the arguments of Tate and Wongbundhit (1983) by analyzing a number of specific models and techniques that have been proposed in the multilevel literature and that seem to require random coefficient regression techniques. The first instance is the general contextual model discussed most completely by Boyd and Iversen (1979, see especially chapter 111). Boyd and Iversen systematically distinguish the single equation approach to contextual analysis from the separate equations approach, which is the more basic one. In the general contextual model there are two types of equations, as in the micro-macro models mentioned previously. The first type specifies an individual level within-group regression model, one for each separate group. The second set of equations relates 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 that is somewhat neglected. As Tate and Wongbundhit point out, "Unfortunately, Boyd and Iversen did not consider the question of statistical inference" (p. 107).

One basic problem with the separate equations approach is that we must decide what exactly we are modeling 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 can be estimated (optimally) 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 modeling step, or in the second set of equations, we can model the distribution of the estimates. We must remember, however, that this distribution is already determined to a large extent by the assumptions and calculations in the first step.

If we assume directly that the regression coefficients are random variables, then much of the above remains true. We must continue to distinguish between regression coefficients and their estimates, where the notion of "estimation" is now extended to cover estimation of random variables. A basic problem with the contextual

analysis literature is that the choice between fixed and random coefficient models is never made explicit. Boyd and Iversen (1979, e.g., 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 (p. 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 Van den Eeden and Saris (1984) and Van den Eeden (1985b). Van den Eeden and Saris analyze school-career data by a two-step approach. The adjective two-step has two different meanings. First, 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. 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 data analytical 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 will comment on this conclusion in a later section of the paper; for now, we merely remark that Van den Eeden and Saris also do not specify if their within-group regression coefficients are random variables or fixed constants. 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. They even use the 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 wrong. We illustrate this by analyzing the same school career data in a different and theoretically more satisfactory way.

For completeness, we emphasize that Boyd and Iversen (1979) are certainly aware of the problems associated with combining two sets of equations into a single set. In their appendix B (pp. 232-233) they discuss a weighted regression 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 (pp. 234-236) they discuss conditions under which separate equations and single equation ordinary least squares give the same estimates. In practice, estimates will be quite different, and Boyd and Iversen give no explicit criteria that can be used to choose between the two.

Another class of multilevel models in which random coefficients seem necessary are the slopes as outcomes analyses of Burstein and his associates: Burstein (1976, 1980a, 1980b, 1981), Burstein and Linn (1976), Burstein, Linn, and Capell (1978), Burstein and Miller (1981), and Burstein, Miller, and Linn (1979). These papers concentrate on motivation and interpretation of the results when

within-group slopes are used as dependent variables in between-group regression analysis. Again, we are not concerned with the theoretical and methodological reasons for adapting this approach, or with its usefulness in educational contexts. For this we refer to the cited literature. We restrict our attention to the statistical problems that are largely ignored by others. The fact that there are some nonstandard problems is acknowledged by Burstein, Miller, and Linn:

> 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 research where sensible analytical work has been conducted without mathematical confirmation of the appropriateness of the distributional assumptions in the measurement of a critical variable. (p. 19)

It seems to us that the last part of the quotation is unduly pessimistic (and a bit muddled). 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 that 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 multilevel techniques can be quite misleading if a random coefficient model is true. The same thing is true for Burstein, Linn, and Capell (1978), who argue convincingly for the importance of assuming heterogeneous within-group slopes, but then illustrate their point by a very small-scale Monte Carlo study. The Monte Carlo results serve well as illustrations, but they give the papers a more limited scope than is really necessary. In situations studied by these authors it is possible to derive analytical results for expectations and standard errors. We must emphasize, however, that the models studied by Tate and Wongbundhit and by Burstein, Linn, and Capell 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 further 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 multilevel analysis more smoothly into standard structural equation modeling practice. For a first version of more general models, we refer to De Leeuw (1985).

In the next section, we will introduce a fairly general random coefficient regression model, with fixed regressors, that can be used in multilevel analysis to unify and extend many results obtained

by contextual analysis or slopes as outcomes approaches. Following the next section, we will review the history of this model and relate it to other models that have been proposed before, mainly in econometrics. Topics discussed later in the paper include the following: (a) ordinary least-squares estimation and a comparison of one-step and two-step approaches; (b) a class of weighted least-squares estimators; and (c) more complicated maximum likelihood estimates. The Dutch school-career data, analyzed earlier by Van den Eeden and Saris, are used to illustrate the various techniques in the section on a school effects example.

2. Model

As indicated, 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 on the p regressors for group j are collected in the $n_j \times p$ 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 the question is whether to treat the within-group regression coefficients as fixed or random, such a convention is especially convenient. The model for group (or school) j is

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

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

disturbances

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

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

Thus, we have a standard linear model for each group, except that the $\underline{\beta}_j$ are supposed to be random vectors. Their properties are specified in the next set of equations.

For each of the random variables $\underline{\beta}_{js}$ ($j = 1, \cdots, m$ and $s = 1, \cdots, p$), we have a model of the form

$$\underline{\beta}_{js} = z_{js}'\theta_s + \underline{\delta}_{js}. \tag{3}$$

The vector z_{js} has q_s elements. Equation (3) is clarified in matrix notation. For micro-variable s we can write the m-vector $\underline{\beta}_s$ in the form

$$\underline{\beta}_s = Z_s\theta_s + \underline{\delta}_s. \tag{4}$$

Thus, there is a separate regression model for the regression coefficients corresponding with each micro-variable. We allow for the possibility that the regression coefficients for IQ in the various schools are regressed on a different set of school variables than the regression coefficients for sex. In this we follow Mason, Wong, and Entwisle (1984). For the disturbances in Equation (3) we assume

$$\mathbf{E}(\underline{\delta}_{js}) = 0, \tag{5a}$$

$$\mathbf{E}(\underline{\delta}_{js}\underline{\delta}_{\ell t}) = 0 \text{ if } j \neq \ell, \tag{5b}$$

$$\mathbf{E}(\underline{\delta}_{js}\underline{\delta}_{jt}) = \omega_{st}, \tag{5c}$$

$$\mathbf{E}(\underline{\delta}_{js}\underline{\epsilon}_{ij}) = 0. \tag{5d}$$

Equation (5b) tells us that disturbances in different groups are uncorrelated, while Equation (5c) tells us that the dispersion of the regression coefficients is the same in each group. We see from Equation (5d) that disturbances of Equations (1) and (3) are uncorrelated. This will become clearer if we rewrite the model in matrix notation.

A useful notation in this context is the direct sum of matrices (MacDuffee, 1946, p. 81). If $A_1, \cdots, A_s$ are matrices, with matrix A_r having n_r rows and m_r columns, then the direct sum $A_1 \dot{+} \cdots \dot{+} A_s$ is an $(n_1 + \cdots + n_r) \times (m_1 + \cdots + m_r)$ block diagonal matrix, with the $A_1, \cdots, A_s$ as the diagonal blocks. Thus $X = X_1 \dot{+} \cdots \dot{+} X_m$ is a matrix with $n = \sum n_j$ rows and with mp columns. If we stack the m vectors $\underline{y}_j$ on top of each other to form the n-vector $\underline{y}$, and in the same way form the mp-vector $\underline{\beta}$ and the n-vector $\underline{\epsilon}$, then we can write Equation (1) as

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}, \tag{6}$$

with $\mathbf{E}(\underline{\epsilon}) = 0$ and $\mathbf{E}(\underline{\epsilon}\underline{\epsilon}') = \sigma_1^2 I \dot{+} \cdots \dot{+} \sigma_m^2 I$.

Translating Equation (3) into matrix notation requires a bit more thought. We first define the $p \times q$ matrix Z_j, with $q = \sum q_j$, by $Z_j = z_{j1}' \dot{+} \cdots \dot{+} z_{jp}'$. Now Equation (3) can be written as

$$\underline{\beta}_j = Z_j\theta + \underline{\delta}_j, \tag{7}$$

where $\mathbf{E}(\underline{\delta}_j) = 0$, $\mathbf{E}(\underline{\delta}_j\underline{\delta}_j') = \Omega$ and $\mathbf{E}(\underline{\delta}_j\underline{\delta}_\ell') = 0$ for $j \neq \ell$. Stack the m equations (7) on top of each other, and we obtain

$$\underline{\beta} = Z\theta + \underline{\delta}, \tag{8}$$

with $\mathbf{E}(\underline{\delta}) = 0$ and $\mathbf{E}(\underline{\delta}\underline{\delta}') = \Omega \dotplus \cdots \dotplus \Omega$ (m times). We have now replaced Equations (1) and (3) by the much more compact Equations (6) and (8). The next logical step is to combine Equations (6) and (8) into a single equation. Define the $n \times q$ matrix $U = XZ$. Then

$$\underline{Y} = U\theta + X\underline{\delta} + \underline{\epsilon}. \tag{9}$$

Alternatively we can also set, by letting $\underline{\nu} = X\underline{\delta} + \underline{\epsilon}$, $\underline{y} = U\theta + \underline{\nu}$. Now $\mathbf{E}(\underline{\nu}) = 0$ and $\mathbf{E}(\underline{\nu}\underline{\nu}) = V$, where $V = V_1 \dotplus \cdots \dotplus V_m$, and $V_j = X_j \Omega X_j' + \sigma_j^2 I$. Equation (9) shows clearly that our random coefficient model is a special mixed linear model – special because of the assumed structure for the error dispersion and because of the relation between U and X.

It is useful to take a short look at the matrices X, Z, and U we have constructed. Matrix X is $n \times mp$. If we assume, as we do in the sequel, that each X_j has rank p, then X has rank mp. Matrix Z looks a bit peculiar, but its algebraic properties become clear if we define it in terms of the matrices Z_s used in Equation (**??**). We can obtain Z by suitably rearranging the rows of $Z_1 \dotplus \cdots \dotplus Z_p$. Here, the matrices in the direct sum are the Z_s, that is, they are $m \times q_s$. If we assume, as we do, that Z_s has rank q_s, then it follows directly that Z has rank q. Thus, both X and Z are of full column rank. Matrix $U = XZ$ has a rather interesting structure. It is of order $n \times q$, and it is build up out of mp matrices $U_{js} = x_{js} z_{js}'$ of orders $n_j \times q_j$ Observe that all U_{js} are of rank one. U itself is of full column rank q.

It is also clear from these developments what a fixed coefficient model is. This is the special case in which there are no second stage disturbances – in which $\Omega = 0$. In the slopes as outcomes approach, and also in the usual contextual models, X_j has only two columns, the first of which is identically equal to +1.The two elements of $\underline{\beta}_j$ are the random intercept and the random slope. It is possible to include models in which intercepts are fixed and slopes are random by requiring certain elements of the parameter vectors to be zero. In simple covariance analysis, for instance, we have another special case in which $\Omega = 0$; the design matrix for the intercepts Z_1 is the identity and the design matrix for the slopes is a vector Z_2 with all elements equal to +1. Such restricted versions of our general model can all be considered as additional specifications whose appropriateness can be tested within the general model.

The interpretation of our model in the multilevel context is clear because it is a straightforward generalization of the contextual model of Boyd and Iversen (1979). In the same way, our model generalizes the slopes as outcomes approach, showing in what sense regression coefficients are random variables. We will see that our estimation procedures generalize the one-step and two-step procedures of Boyd and Iversen 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 will indicate that our model is far from new and has already been studied in great detail in the econometric and statistical literature.

3. The History of Variable Coefficient Models

Random coefficient models, or more generally, variable coefficient models, have a long history in econometrics. Pioneering work of Rubin, Klein, Wald, and Theil in the late 1940s and early 1950s had little practical impact and was ignored for some time. More comprehensive papers, oriented toward practical applications, were written in the late 1960s by Rao, Fisk, Hildreth and Houk, and Swamy. The pre-1970 literature is reviewed almost completely in the monograph by Swamy (1971). In the 1970s, a substantial body of theory was developed, and a number of useful review papers appeared. We mention Rosenberg (1973), Spjøtvoll(1977), and Mundlak (1978). Chapter 17 in Maddala (1977) and a recent chapter by Chow (1984) are also very useful. Annotated bibliographies have been published by Johnson (1977, 1980).

Most of these econometric papers discuss models that are less general than our model in the previous section. In the second-stage specification (Equation (3)), econometric models have $q_s = 1$, and $z_{js} = 1$. Thus $q = p$, and Equation (7) becomes simply $\underline{\beta}_j = \theta + \underline{\delta}_j$, because $Z_j = I$ for all j. There is effectively no second-stage model of independent interest, which makes these econometric models not very useful 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 that is very similar to our model. Unfortunately, Hanushek does not clearly distinguish between random and fixed variables, 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, which is discussed very briefly, is identical to our model (Equations 6 and 8), but the assumptions on the disturbances and the characteristics of the matrix Z are quite different. The difference arises, of course, because the models are designed for different types of applications. A two-stage model very similar to Amemiya's has been studied 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 random coefficient models in econometrics are either not specific enough, or are 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 will do in the sequel. Moreover, many results in statistics deal with general mixed-linear models. They can be used for our model, too. Finally, our two-stage models are closely related to Bayesian and empirical Bayes methods for the linear model. These results are discussed roost completely in Lindley and Smith (1972) and in the contributions of the discussants of that paper.

In discussing the history of our model we must also discuss some recent history that has come to our attention while preparing the final version of this paper. The current interest in school effectiveness research has focused attention on multilevel modeling and analysis. Basically, the same model as proposed in this paper is

also studied by Ecob (1985), Goldstein (in press), Aitkin and Longford (in press), and Longford (1985). The emphasis and the details are often somewhat different, because these authors use variance component models as their starting point and are sometimes only interested in special cases of our general model. Nevertheless, the similarities with our work are much more pronounced than the differences, and virtually everything in these papers is relevant for our discussion.

In the context of longitudinal studies, versions of our multilevel model (in which measurement waves define the second level) have been developed by Laird and Ware (1982) and Ware (1985). The paper that is closest to ours, both in its starting point and its proposed models, is Mason, Wong, and Entwisle's (1984). They use the hierarchical linear models of Lindley and Smith (1972) and Smith (1973) as their starting point, also with the explicit purpose of providing a completely specified model for contextual analysis. Their estimation methods are somewhat different from ours, and their basic example is from comparative fertility research, but otherwise both their approach and their results are very close to ours.

There are also two important developments in the random coefficient literature that we have not incorporated in our model, although these developments could very well be useful in multilevel research. The first one, already discussed in connection with Tate and Wongbundhit (1983) and 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," that is, correlations between coefficients and independent random variables. Our basic approach in this paper, without transmitted errors, is applied to random regressor models, and even to path analysis models in De Leeuw (1985). A second omission, somewhat less serious perhaps, is the modeling of the first-level error variances as random variables as well. This could be useful as a "residuals as outcomes7' approach. Models that allow random variances are discussed by Aragon (1984). Both extensions of our basic model lead to many complications and into largely uncharted territory.

4. Least-Squares Estimation

In this section we discuss various aspects of ordinary (unweighted) least-squares estimation in our model (Equation (9)). We first consider Equations (6) and (8) separately and estimate the $\underline{\beta}_j$ from Equation (6). We must realize, of course, that we estimate random variables here, and not fixed constants. Nevertheless, the notions of bias and variance apply to the 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), Swamy (1970, 1971), and Pfefferman (1984). The relevant result for our model is that the minimum variance unbiased linear estimate of $\underline{\beta}_j$, is $\hat{b}_j = (X_j'X_j)^{-1}X_j'\underline{y}_j$. Using matrix notation, we can also write $\hat{b} = (X'X)^{-1}X'\underline{y}$, but the important thing to observe from a practical point of view is that we compute regression coefficients separately for each group.

The expectation of $\hat{b}_j$ is $\mathbf{E}(\hat{b}_j) = Z_j\theta$, and its variance is $W_j = \Omega+\sigma_j^2(X_j'X_j)^{-1}$. Again, it is convenient to define $W = W_1\dot{+}\cdots\dot{+}W_m$. Thus, $\hat{b}$ has expectation $Z\theta$ and dispersion W. Also define the residual $\underline{r}_j = \underline{y}_j - X_j\hat{b}$. The residual $\underline{r}_j$ has expectation zero, and

$$\mathbf{E}(\underline{r}_j\underline{r}_j') = \sigma_j^2[I - X_j(X_j'X_j)^{-1}X_j']. \tag{10}$$

It follows that $\mathbf{E}(\underline{r}_j'\underline{r}_j) = \sigma_j^2(n_j - p)$, and thus $\hat{\sigma}_j^2 = \underline{r}_j'\underline{r}_j/(n_j - p)$ is unbiased for σ_j^2. Ordinary regressions within groups give us unbiased estimates of the $\underline{\beta}_j$ and the σ_j^2.

In the next step we compute an estimate of θ. This is simply $\hat{\theta} = (Z'Z)'Z\hat{b}$. Again, from a practical point of view, this is most easily understood by writing it as $\theta_s = (Z_s'Z_s)^{-1}Z_s'\hat{b}_s$, where $\hat{b}_s$ contains the m regression coefficients for variable s in the m groups. Our second step ends by computing an estimate for Ω. This requires some thinking, because the unbiased estimates developed in Rao (1965a) and Swamy (1970) will not work for our more complicated model. They are based on the econometric model in which $\mathbf{E}(\underline{\beta}_j) = \theta$ for all j. Their basic idea can be generalized quite easily, however. Define residuals $\underline{t}_s = \hat{b}_s - Z_s\theta_s$. We can also write $\underline{t}_s = Q_s\hat{b}_s$, where $Q_s s = I - Z_s(Z_s'Z_s)^{-1}Z_s$. The $\underline{t}_s$ have expectation zero, and

$$\mathbf{E}(\underline{t}_s\underline{t}_r') = Q_s(\omega_{sr}I + \Sigma\nabla_{sr})Q_r. \tag{11}$$

In Equation (11) we have used Σ for the diagonal matrix with the σ_j^2, and ∇_{sr} for the diagonal matrix with all (s,r)-elements of the m matrices $(X_j'X_j)^{-1}$ on the diagonal. From Equation (11) we find the unbiased estimate

$$\hat{\omega}_{sr} = (\underline{t}_s'\underline{t}_r - \mathbf{tr}\ Q_s\hat{\Sigma}\nabla_{sr}Q_r)/\mathbf{tr}\ Q_sQ_r, \tag{12}$$

where $\hat{\Sigma}$ contains the $\hat{\sigma}_j^2$ on the diagonal. If all Z_s are the same, as in Van den Eeden and Saris (1984), then Equation (12) simplifies to

$$\hat{\Omega} = \{\underline{T}'\underline{T} - \sum_{j=1}^{m} \xi_j \hat{\sigma}_j^2 (X_j'X_j)^{-1}\}/(m - \overline{q}), \tag{13}$$

where the $m \times p$ matrix $\underline{T}$ contains the $\underline{t}_s$, where ξ_j is the j^{th} diagonal element of Q, and where $\overline{q}$ is the number of columns of each of the Z_s. Thus, we can estimate the dispersion on both levels from the ordinary least-squares residuals. It is somewhat unfortunate that estimate in Equation (12) need not be positive semidefinite. Compare the discussion in Swamy (1971, pp. 107-111).

After two ordinary least-squares steps, we have unbiased estimates of all parameters. This is quite satisfactory, because most people in educational research use simple least squares, and it is quite likely that they will continue to do so. By using Equation (12), the additional parameters of the random coefficient model can be computed quite simply from the least-squares residuals. However, another ordinary least-squares procedure can be applied to our model. If the model is written in the two-equation form (Equations (6) and (8)), the two-step least-squares estimate is natural. If it is written in the single-equation form (Equation (9)), we immediately think of the single-step estimate $\hat{\theta} = ((U'U)^{-1}U'\underline{y}$ (Boyd & Iversen, 1979, pp. 53-55).

We can compare the two estimates by using generalized inverses. The matrix U is the product of X and Z, which are both assumed to be of full column rank. It follows that, using superscript + for the Moore-Penrose inverse, Z^+X^+ is a generalized inverse of U, in fact, a left inverse because $Z^+X^+XZ = I$. If we let $U^- = Z^+X^+$,

then $\hat{\theta} = U^{-}\underline{y}$ is nothing but the two-step estimate of θ computed with a single-step formula. Clearly, $\hat{\theta} = U^{+}\underline{y}$ is the single-step estimate. Although we use the same symbol for the two estimates, they are in general different. Boyd and Iversen (1979, appendix C) give sufficient conditions for their equality. Our development suggests a simple formulation. We have equality for all y if, and only if, U^{-} is a Moore-Penrose inverse of U. This is the case if, and only if, $UU^{-} = XZZ^{+}X^{+}$ is symmetrical.

Additional insight can be obtained by using Cline's (1964) formula. Cline proves that $(XZ)^{+} = Z^{+}(XZZ^{+})^{+}$ if X has full column rank. Other relevant generalizations of the "reverse order law" for Moore-Penrose inverses are given by Greville (1966) and by Barwick and Gilbert (1974). It is clear by now that the terminology single-step and two-step is quite misleading. We have the formula $\hat{\theta} = U^{-}\underline{y}$, which is a single-step formula for the separate-equations estimate. We also have the formulas $\hat{b} = (XZZ^{+})^{+}\underline{y}$ and $\hat{\theta} = Z^{+}b$, which are two-step formulas for the single-equation estimate.

Because the disturbances have zero expectation, both the single-equation and the separate-equations estimate are unbiased. Comparisons between them have been given by Van den Eeden and Saris (1984). The separate-equations estimate is easier to compute. We know that Z is of the form $Z = P(Z_1 \dotplus \cdots \dotplus Z_p)$, with P a permutation matrix. Thus, $Z^{+} = (Z_1^{+} \dotplus \cdots \dotplus Z_p^{+})P'$, and $ZZ^{+} = P(Z_1Z_1^{+} \dotplus \cdots \dotplus Z_pZ_p^{+})P'$. But XZZ^{+} is generally a full matrix, of order $n \times mp$, and thus computation of$(XZZ^{+})^{+}$ is not a trivial matter. It has been suggested by Van den Eeden and Saris that as a consequence, single-equation estimates may be bothered

more by multi-collinearity, and that the separate-equations estimates are easier to understand and to interpret. We agree with this evaluation. We do not agree with the other reasons suggested by Van den Eeden and Saris for preferring the separate-equations estimate. Both procedures lead to unbiased estimates and do not take into account the structure of the disturbances. Tate and Wongbundhit (1983) also reach the conclusion that the procedures they have compared (single equation, separate equations, and mixed) all produce unbiased estimates. For the model they consider, this can be proved directly; there is no need to use a Monte Carlo study to confirm this.

5. Weighted Least Squares

In the previous section we discussed both separate-equations and single-equation least-squares methods. In the first analysis, the single-equation method has little to recommend it, and the separate-equations method seems preferable from a computational and interpretational point of view. In this section and in the next one, we will develop procedures that are more satisfactory from a statistical point of view, and that maintain the interpretational advantages of the separate-equations method.

From Equation (9) we know, using the Gauss-Markov theorem, that the best linear unbiased estimate of θ is given by

$$\hat{\theta} = (U'V^{-1}U)^{-1}U'V^{-1}\underline{y}. \tag{14}$$

This result, as such, is quite useless, because $V = V_1 \dotplus \cdots \dotplus V_m$ with $V_j = X_j\Omega X_j' + \sigma_j^2 I$ is generally unknown. Swamy (1970, 1971)

suggested substituting the estimates of σ_j^2 and Ω computed in the previous section in the definition of V. This gives an estimate $\hat{V}$, also unbiased. We then estimate θ by substituting $\hat{V}$ for V in Equation (14). 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 discuss this, we first point out a remarkable simplification of the estimate.

In the monograph by Swamy (1971, p. 101), we find the formula

$$V_j^{-1} = \sigma_j^{-2}[I - X_j(X_j'X_j)^{-1}X_j'] + X_j(X_j'X_j)^{-1}W_j^{-1}(X_j'X_j)^{-1}X_j', \tag{15}$$

where $W_j = \Omega + \sigma_j^2(X_j'X_j)^{-1}$, as before. This implies that $X_j'V_j^{-1}X_j = W_j^{-1}$ and that $X_j'V_j^{-1}\underline{y}_j = W_j^{-1}\hat{b}_j$. Thus

$$\hat{\theta} = (U'V^{-1}U)^{-1}U'V^{-1}\underline{y} = (Z'W^{-1}Z)^{-1}Z'W^{-1}\hat{b}. \tag{16}$$

This formula is very convenient from the computational point of view, because we have replaced inversion of matrices V_j, of order n_j, by inversion of matrices W_j, of order p. It is also clear from Equation (16) that the Gauss-Markov estimate can be interpreted as a two-step estimate.

On the other hand, a comparison of Equation (16) with the formulas for the single-equation and the separate-equations estimate seem to indicate that the single-equation estimate will generally be closer to the Gauss-Markov estimate. The single-equation estimate is optimal if $\Omega = 0$ and if all σ_j^2 are equal, that is, if $W = \sigma^2(X'X)^{-1}$. The separate-equations estimate is optimal if in addition $X'X = I$, i.e. $X_j'X_j = I$ for all j. Thus, for small Ω and for approximately

equal σ_j^2 the single-equation method will give a good approximation to W, while it is difficult to think of situations in which the separate-equations approximation will be better. This advantage of the single-equation method may offset its disadvantages in computational and interpretational aspects.

Again, we emphasize that our development here generalizes that of Rao, Swamy, and others, who only study the simple case in which each Z_s is a single column of ones. In this case, Equation (16) simplifies to

$$\hat{\theta} = \left\{\sum_{j=1}^{m} W_j^{-1}\right\}^{-1} \sum_{j=1}^{m} W_j^{-1}\hat{b}_j, \tag{17}$$

which shows that in this case θ is a matrix-weighted average of the $\hat{b}_j$. Swamy (1970) has studied the asymptotics of weighted least squares for the restricted model. His results can be easily extended to our more general case. We have to assume that both m and n_j tend to infinity. The matrices $n_j^{-1}X_j'X_j$ and $m^{-1}Z_s'Z_r$ must also tend to limits. Let C_{sr} be the limit of $m^{-1}Z_s'Z_r$. Then Swamy's result, translated into our more general context, says that $m^{1/}(\hat{\theta} - \theta)$ is asymptotically normal. Its asymptotic dispersion matrix is the inverse of a matrix with submatrices $\omega^{sr}C_{sr}$, where ω^{sr} is $(\Omega^{-1})_{sr}$. If all C_{sr} are the same, which happens if all Z_s are the same, then the asymptotic dispersion is $\Omega \otimes C^{-1}$, with $\otimes$ the Kronecker product. Under Swamy's assumptions, the separate-equations least-squares estimate has an asymptotic dispersion matrix with submatrices $\omega_{sr}C_{sr}^{-1}$. Thus, if all Z_s are equal, the two estimates are asymptotically equivalent. It is far less simple to find

the asymptotic dispersion of the single-equation least-squares estimate.

Johansen (1982,1984) improves the conditions under which Swamy's (1970) result holds, but also points out that the result may not be satisfactory in some situations. The fact that the weighted least-squares estimate has the same asymptotic distribution as the unweighted separate equations estimate already indicates that (asymptotically at least) there was no reason to weight in the first place. This is also indicated by the fact that the limit distribution does not depend on n_j, X_j or σ_j^2. Johansen proves a much more complicated result, which allows for an asymptotic effect of the weights. The result depends critically, however, on assuming Gaussian disturbances, and is not 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 section, we think that the weighted estimate will generally improve upon the unweighted estimates, although this is by no means certain. 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 converge to their limits. Clearly, what we really need are expansions, not limit theorems, in order to make more definite statements.

6. Maximum Likelihood

Maximum likelihood methods for mixed analysis of variance models (ANOVA) were first discussed systematically by Hartley and Rao

(1967); recent state-of-the-art reviews include 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 Andrade and Helms (1984). Iterative weighted least-squares algorithms for computing maximum likelihood estimates were proposed by Goldstein (in press), and scoring methods by Longford (1985, in press). Both Goldstein and Longford have developed their methods in the context of nested hierarchical models, and both have applied them to school effectiveness research.

Alternative non-maximum likelihood estimates for the dispersion of the residuals, at both stages, could be based on Rao's MINQUE theory, which is reviewed by Rao (1979), Kleffe (1980), and Rao and Kleffe (1980). We merely note this; we do not apply MINQUE to our random coefficient model in this paper. For the possibilities, we refer to the dissertations of Streitberg (1977) and Infante (1978).

One of the most interesting results in our previous two sections is that the simplest unweighted least-squares method and the weighted least-squares method both worked in two computational steps. In the first step, within-class regression coefficients were computed by ordinary least squares, together with the within-class residuals. In the second step, the within-class regression coefficients were used as dependent variables for the between-class analysis. This is an important property, because it implies that in the second step

we did not work with the original y_j and X_j any more, but with a much smaller reduced set of variables. This makes computation in the second step relatively inexpensive. In this section, we show that a similar result applies in the case of maximum likelihood estimation. Although this method is computationally much more complicated than the least-squares methods, it does share this basic simplifying property with them.

As is well-known, the method of maximum likelihood has a somewhat peculiar position in statistics, especially in applied statistics. Maximum-likelihood estimations are introduced as if they are by definition good, or optimal, in all situations. Another peculiarity of the literature is that maximum-likelihood methods are introduced by assuming a specific probability model, which is often false in the applications one has in mind. In our context, this means that typically it is assumed that the disturbances, and thus the observed y, are jointly normally distributed. Of course, such an assumption is highly debatable in many educational research situations, and quite absurd in others.

We take a somewhat different position. Least-squares estimates are obtained by minimizing a given loss function, and this is how they are defined. Afterwards, we derive their properties and we discover that they behave nicely in some situations. We approach multinormal maximum likelihood in a similar way. The estimates are defined as those values of θ, Ω and Σ that minimize the loss function

$$\log|V| + (\underline{y} - U\theta)'V^{-1}(\underline{y} - U\theta). \tag{18}$$

Again, at a later stage, we will say something about their properties. The important fact is that Equation (18) is quite a natural loss function. It measures closeness of $\underline{y}$ to $U\theta$ by weighted least squares, and it measures at the same time closeness of $(\underline{y} - U\theta)(\underline{y} - U\theta)'$ to V. This last property may not be immediately apparent from the form of Equation (18). It follows from the inequality $\log|A| + \mathbf{tr}\ A^{-1}B \geq \log|B| + m$,which is true for all pairs of positive definite matrices of order m. We have equality if and only if $A = B$. Thus, in our context, $\log|V| + \mathbf{tr}\ V^{-1}R(\theta)$, with $R(\theta) = (\underline{y} - U\theta)(\underline{y} - U\theta)'$ measures the distance between V and the residuals $R(\theta)$. We want to make residuals small and we want the dispersion to be maximally similar to the dispersion of the residuals. Moreover, we want to combine these two objectives in a single loss function.

We now simplify Equation (18), again by using the basic formula Equation (15). This gives

$$(\underline{y}_j - X_jZ_j\theta)'V_j^{-1}(\underline{y}_j - X_jZ_j\theta) = (n_j - p)\hat{\sigma}_j^2/\sigma_j^2 + (\hat{b}_j - Z_j\theta)'W_j^{-1}(\hat{b}_j - Z_j\theta) \tag{19}$$

Remember that the $\hat{\sigma}_j^2$ are the first-stage estimates of the residual variances, that is, $\hat{\sigma}_j^2 = \underline{r}_j'\underline{r}_j/(n_j - p)$. Another useful identity is

$$\log|V_j| = \log|X_j'X_j| + (n_j - p)\log\sigma_j^2 + \log|W_j|. \tag{20}$$

Combining Equations (19) and (20) shows that minimizing Equation (18) is the same thing as minimizing

$$\log|W| + (\hat{b} - Z\theta)'W^{-1}(\hat{b} - Z\theta) + \sum_{j=1}^{m}(n_j - p)(\log\sigma_j^2 + \hat{\sigma}_j^2/\sigma_j^2). \tag{21}$$

To assess goodness-of-fit, it is useful to compare Equation (21) with a lower bound. If we can find θ such that $Z\theta = \hat{b}$, then such a θ is the maximum-likelihood estimate. In this case, the maximum-likelihood estimate of Ω is the zero matrix, and the maximum-likelihood estimate of σ_j^2 is $\hat{\sigma}_j^2 = \underline{r}_j'\underline{r}_j/n_j$. These values define a lower bound of Equation (21) equal to

$$n - \log|X'X| + \sum_{j=1}^{m} n_j \log \hat{\sigma}_j^2. \tag{22}$$

For interpretation, it is consequently convenient to define a loss function equal to the difference of Equations (21) and (22).

Actual minimization of this maximum likelihood loss function is not simple. In closely related situations, Goldstein (in press) applies iterative generalized least squares, and Longford (1985, in press) applies the method of scoring. The two are essentially equivalent in this context. We have derived the necessary formulae for our model, and we present them in Appendix A. The algorithm based on the formulae seems to perform well. It would be interesting to compare its performance with the EM-algorithm used by Mason, Wong, and Entwisle (1984). It must be emphasized, however, that Mason, Wong, and Entwisle compute restricted maximum-likelihood estimates (REML), whereas we use unrestricted maximum likelihood (ML). Comparisons between REML and ML are in Harville (1977). Comparisons with MINQUE are in Rao (1979) and Rao and Kleffe (1980).

The asymptotic properties of maximum-likelihood estimates in mixed analysis of variance models, which include our random coefficient model as a special case, have been investigated most thoroughly

by Miller (1977). Assuming normally distributed errors, he proves consistency, asymptotic normality, and efficiency of the maximum-likelihood estimates by using an increasing sequence of designs (both the number of schools and the number of pupils converge to infinity).

As we already mentioned in the discussion of weighted least-squares estimation, 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 to use Monte Carlo methods next to asymptotic results as long as satisfactory expansions are not available. If the conditions used by Swamy in the case of weighted least-squares estimation are true, then the maximum-likelihood estimates are asymptotically equivalent to the weighted least-squares estimates. In our special model, however, simplifications are possible because it follows from Equation (21) that the maximum-likelihood estimates are a function of the $\hat{b}_j$ and the $\hat{\sigma}_j^2$, which are asymptotically normal. We follow Aitkin, Longford, Goldstein, and others in using the information matrix as an estimate of the dispersion of the maximum-likelihood estimates. The necessary formulae are in the Appendix.

7. A School Effects 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. We only use the 1959 cohort, consisting of 1,270 pupils in 37 schools. For each pupil, the individual-level independent variables we used were sex, IQ, and occupational level of the father. The dependent variable was teachers' advice on the form of secondary education. Thus, in our example $p = 4$ (constant term, SEX, IQ, SES) and $m = 37$. IQ was coded as a continuous variable; it has values between 58 and 148. Fathers' occupation had six possible values, and teachers' advice had seven.

Optimal scaling techniques indicate that integer-scoring of the categories leads to regressions that do not deviate much from linearity (Meester & De Leeuw, 1983). Thus, we treat occupation and advice as numerical variables, although this remains debatable. As the independent variables on the school-level, we use a constant term and average school IQ, the aggregated individual-level IQ. Thus $q = 2$ and all Z_i are the same. We have chosen the same school variables as Van den Eeden and Saris, but we have more individual-level predictors because they only use IQ and the constant term on the individual level as well. We have standardized all four variables SEX, IQ, SES, and ADV in such a way that they have mean zero and variance one over the 1,270 pupils.

Our first analysis step is to perform the 37 within-school regressions. In Table I we have collected the most important information relating to this stage of the analysis. The columns contain number of pupils, average IQ, variance of IQ, regression coefficients for constant, SEX, IQ, SES, and estimate of the residual variance. It follows

from the results we have derived that these school-level statistics are all that are needed to perform the second-stage analysis (except for the single-equation ordinary least-squares estimation).

TABLE 1. Within-School Statistics: Number of Pupils, Mean IQ, Variance IQ, Regression Coefficients, and Residual Variance

no	np	avIQ	vrIQ	rgCON	rgSEX	rgINT	rgSES	resi
1	29	-0.42	0.48	-0.38	0.08	0.34	0.09	0.11
2	33	0.60	0.52	-0.12	-0.13	0.99	0.14	0.30
3	31	0.18	1.08	0.26	0.01	0.61	0.12	0.28
4	66	0.47	1.04	0.14	-0.09	0.69	0.12	0.41
5	39	-0.70	0.58	-0.20	-0.07	0.59	-0.01	0.21
6	45	-0.43	0.70	0.18	0.04	0.55	0.18	0.57
7	39	-0.03	1.24	0.10	-0.09	0.65	0.13	0.40
8	31	-0.33	0.98	0.08	0.15	0.81	-0.04	0.17
9	53	0.59	0.70	-0.22	-0.05	0.82	0.16	0.34
10	31	-0.45	0.56	-0.01	-0.01	0.85	0.13	0.18
11	30	-0.50	0.66	-0.19	0.24	0.51	0.21	0.09
12	36	-0.26	0.94	0.03	0.02	0.70	0.04	0.28
13	52	-0.02	1.15	-0.10	0.01	0.58	0.22	0.25
14	29	0.10	1.17	0.16	-0.08	0.72	0.22	0.33
15	33	-0.24	1.27	-0.01	0.08	0.58	0.34	0.23
16	65	0.40	0.66	0.42	-0.05	0.79	0.16	0.38
17	57	0.43	1.40	-0.10	-0.32	0.76	0.29	0.41
18	31	-0.26	1.11	-0.06	-0.20	0.73	0.04	0.44
19	26	-0.49	0.75	0.00	0.10	0.56	0.14	0.20
20	27	-0.27	0.56	-0.09	-0.12	0.81	-0.14	0.26
21	25	-0.25	0.42	-0.54	0.03	0.22	0.11	0.20
22	27	-0.02	0.70	-0.16	-0.10	1.02	0.13	0.22
23	26	-0.15	0.80	-0.03	-0.11	0.77	-0.04	0.31
24	36	-0.68	0.75	-0.37	0.15	0.53	0.07	0.29
25	11	-0.92	0.49	-0.90	1.08	0.70	0.l51	0.25
26	27	0.00	0.74	-0.16	0.11	0.63	0.23	0.25
27	15	0.56	0.67	-0.21	-0.02	0.63	-0.03	0.35
28	27	-0.41	0.90	-0.14	0.10	0.69	0.20	0.25
29	20	0.14	0.86	-0.15	0.31	0.26	0.22	0.25
30	32	-0.44	0.66	-0.05	-0.11	0.55	0.19	0.28
31	49	0.43	1.03	-0.02	0.01	0.86	0.10	0.30
32	57	0.63	0.68	-0.11	-0.14	1,07	-0.01	1.02
33	37	0.32	0.63	-0.12	-0.10	1.02	0.09	0.23
34	30	0.50	0.52	0.04	0.03	1.02	0.07	0.31
35	35	-0.20	0.65	-0.02	-0.03	0.83	0.29	0.31
36	28	-0.39	0.44	-0.10	0.07	0.65	0.02	0.12
37	16	-0.42	1.03	-0.49	-0.19	0.81	-0.11	0.13

It is clear from Table I that there is considerable variation both in the regression coefficients and in the residual variances, and the second step of the analysis seems necessary to model at least some of this variation.

Before we proceed, we must make one thing clear about our analysis of this example. We can use such an analysis for at least three purposes. First, we can try to draw conclusions that are of value in understanding the real-world situation. These can be of interest either for school effect research in general or for describing the situation in Groningen in 1959 in particular. This is obviously not our strategy in this paper. A second purpose of the analysis could be to show that models make a difference. This is illustrated beautifully in the paper by Aitkin and Longford (in press), and to some extent also in Burstein, Linn, and Capell (1978), Tate and Wongbundhit (1983), and Ecob (1985). But again this is not our purpose. We merely want to investigate if choice of estimation method makes a difference, and if this is the case how large these differences are. It is clear that the other two questions are far more interesting, but also far more difficult to answer. We hope to address them in subsequent publications.

The first second-stage technique is ordinary least squares. We estimate the $2 \times 4 = 8$ elements of θ. They are given in the first row of Table 2. The first four elements show the regression of the individual-level regression coefficients on the school-level constants. The next four elements show regression of the regression coefficients on school-level aggregated intelligence. The single-equation ordinary least-squares regression coefficients are given

in the second row of Table 2. It is clear that differences between the two sets of estimates are minor and mainly occur in the small regression coefficients.

By using the separate-equations least-squares residuals we can estimate Ω by Equation (13). The estimate is given in Table 4. Although it is not positive definite, we fortunately have that $\hat{\Omega} + \hat{\sigma}_j^2(X_j'X_j)^{-1}$ is positive definite for all j. Thus, we can use the Swamy estimate in Table 4 to compute weighted least-squares estimates of θ. They are given in row three of Table 2. Again, they do not differ substantially from the unweighted estimates. The Swamy estimate of Ω can also be used in estimating standard errors of the weighted least-squares estimate. These are given in row three of Table 3. Because of the negative elements on the diagonal of $\hat{\Omega}$ we cannot really be satisfied with the estimate.

TABLE 2. Regression Coefficients of Within-School Regression Coefficients on School Variables for Separate-Equations OLS, Single-Equation OLS, Weighted Least-Squares with Swamy Weights, and Weighted Least-Squares with Maximum Likelihood Weights

school	AVE	AVE	AVE	AVE	INT	INT	INT	INT
pupil	AVE	SEX	INT	SES	AVE	SEX	INT	SES
OLS (2S)	-.081	-.002	.718	.121	.221	-.235	.210	-.029
OLS (1S)	-.027	-.025	.712	.134	.157	-.121	.163	-.005
WLS (SW)	-.057	-.017	.717	.119	.170	-.119	.182	.071
WLS (ML)	-.050	-.014	.712	.127	.186	-.129	.184	.040

If we try to improve our estimates by maximum likelihood we run into various troubles. If we do not constrain Ω then the method tries to converge to an indefinite $\hat{\Omega}$, and this we cannot allow. The

TABLE 3. Standard Errors of Regression Coefficients of Within-School Regression Coefficients on School Variables for Separate-Equations OLS, Single-Equation OLS, Weighted Least Squares with Swamy Weights, and Weighted Least Squares with Maximum Likelihood Weights

school	AVE	AVE	AVE	AVE	INT	INT	INT	INT
pupil	AVE	SEX	INT	SES	AVE	SEX	INT	SES
OLS (2S)	.031	.024	.023	.022	.092	.131	.062	.094
OLS (1S)	.030	.017	.022	.018	.073	.043	.057	.046
WLS (SW)	.034	.024	.029	.010	.083	.057	.069	.025
WLS (ML)	.029	.016	.021	.016	.069	.038	.050	.040

TABLE 4. Swamy Estimate Second-Stage Error Dispersion

	AVE	SEX	INT	SES
AVE	.0309	-.0116	.0003	-.0003
SEX	-.0116	.0106	.0004	.0007
INT	.0003	.0004	.0172	-.0002
SES	-.0003	.0007	-.0002	-.0033

same thing happens if we constrain Ω to be diagonal. If we require diagonality, and in addition set elements (2,2) and (4,4) equal to zero, then the technique converges to the estimate given in Table **??**. This constrained model tells us that the regression coefficients for SEX on SES are fixed, while the intercept and the regression coefficient for intelligence are random with variances .0185 and .0035, respectively. Thus, Tables **??**,2, and 3 tell us that, using

$\underline{z}$ for a standard normal disturbance,

$$(\text{intercept})_j = -.05 + .19(\text{average intelligence})_j, +.14\underline{z}$$

$$(\text{regression coefficient IQ})_j = .71 + .18(\text{average intelligence})_j, +.05\underline{z},$$

$$(\text{regression coefficient SEX})_j = -.01 - .13(\text{average intelligence})_j,$$

$$(\text{regression coefficient SES})_j = .13 + .04(\text{average intelligence})_j.$$

If we take the standard errors in Table I11 into account, we see significant effects of average intelligence on the intercept and the regression coefficient for intelligence. This suggests that individuals with average IQ, SES, and SEX get higher advice in schools with high average intelligence levels. It also suggests that the individuals' intelligence is a better predictor of advice in high-intelligence schools. There is also some indication, though not very strong, that in schools with high average IQ, boys of average IQ and SES get a higher advice than corresponding girls, while in schools with low average IQ the situation is more the other way around.

TABLE 5. Maximum Likelihood Estimate Second-Stage (Restricted) Error Dispersion

	AVE	SEX	INT	SES
AVE	.0185	–	–	–
SEX	–	–	–	–
INT	–	–	.00344	–
SES	–	–	–	–

Our most important conclusion, however, is that choice of estimation method does not seem to have much influence on the size of the regression coefficients. Estimates of the second-stage between-school disturbances are quite different, however. These random

effects on the first-stage regression coefficients are quite small in this example. The standard errors in Table 3 show that the really important regression coefficients are estimated with roughly equal precision by all techniques, while the small regression coefficients are estimated more precisely by maximum likelihood and weighted least-squares methods.

The conclusions that are suggested by the analysis of this example, as far as the size of the regression coefficients is concerned, are similar to those of Van den Eeden and Saris (1984). Our methodological conclusions are a bit different. Although we agree that the two-stage (or separate-equations) unweighted least-squares method has definite advantages from the computational and interpretational point of view, we also find that the single-equation method gives more precise estimates.

8. Conclusions and Recommendations

Our first and foremost 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 many other possibilities. In fact, we believe that our model, although 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 developments that are needed from the modeling point of view.

In a statistical sense our models 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 hypothesis that $\Omega = 0$, for instance, or that the σ_j^2 are equal, that some elements of θ are zero, and so on. Because we have concentrated on estimation, and not on testing and interpretation, we have not 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 various estimation methods do not seem to lead to large differences in results. The Rao-Swamy-Johansen weighted least-squares method seems an excellent method to estimate the unknown parameters of the model, at least in cases in which the estimate of Ω is not too negative.

APPENDIX A. COMPUTING MAXIMUM LIKELIHOOD ESTIMATES BY THE SCORING METHOD

Define

$$f_j(\theta,\Omega,\Sigma) = (n_j - p)(\log\sigma_j^2 + \hat{\sigma}_j^2/\sigma_j^2) + \log|W_j| + \\ + (\hat{b}_j - Z_j\theta)'W_j^{-1}(\hat{b}_j - Z_j\theta). \tag{23}$$

Then we must minimize the sum of the f_j. If we want to apply the scoring method, we need expressions of the first-order and second-order partial derivatives. These expressions can also be used to compute the information matrix, and they can be used, in principle, to construct a Newton-Raphson algorithm. The partials are given by approximating $f_j(\theta+\zeta,\Omega+\Delta,\Sigma+T)$ by the first two terms of its Taylor expansion around (Θ,Ω,Σ). The Newton-Raphson method minimizes this quadratic approximation in each step. The scoring method first replaces the second-order terms by their expectations and then minimizes.

We use $\underline{s}_j$ for $\hat{b}_j - Z_j\theta$ and D_j for $(X_j'X_j)^{-1}$. The first order terms are

$$\mathbf{tr}\, W_j^{-1}\Delta - \underline{s}_j'W_j^{-1}\Delta W_j^{-1}\underline{s}_j + \tau_j[(n_j-p)(\sigma_j^{-2} - \hat{\sigma}_j^2\sigma_j^{-4}) + \\ + \mathbf{tr}\, W_j^{-1}D_j - \underline{s}_j'W_j^{-1}D_jW_j^{-1}\underline{s}_j] + 2\zeta'Z_j'W_j^{-1}\underline{s}_j. \tag{24}$$

The six different types of second order terms are given next.

$$\begin{aligned}(25)\quad & \underline{s}_j'W_j^{-1}\Delta W_j^{-1}\Delta W_j^{-1}\underline{s}_j-\frac{1}{2}\mathbf{tr}\,W_j^{-1}\Delta W_j^{-1}\Delta+\\ &+\tau_j^2[(n_j-p)(\hat\sigma_j^2\sigma_j^{-6}-\frac{1}{2}\sigma_j^{-4})+\underline{s}_j'W_j^{-1}D_jW_j^{-1}D_jW_j^{-1}\underline{s}_j-\frac{1}{2}\mathbf{tr}\,W_j^{-1}D_jW_j^{-1}D_j]+\\ &+\zeta'Z_j'W_j^{-1}Z_j\zeta+\tau_j(2\underline{s}_j'W_j^{-1}\Delta W_j^{-1}D_jW_j^{-1}\underline{s}_j-\mathbf{tr}\,\Delta W_j^{-1}D_jW_j^{-1})+\\ &+2\zeta'Z_j'W_j^{-1}\Delta W_j^{-1}\underline{s}_j+2\tau_j\zeta'Z_j'W_j^{-1}D_jW_j^{-1}\underline{s}_j.\end{aligned}$$

It is clear that the Newton-Raphson method will be difficult to apply. Matters simplify greatly if we take expectations of the six terms of Equation (25). Because $\mathbf{E}(\underline{s}_j)=0$ the last two terms disappear. Using $\mathbf{E}(\underline{s}_j\underline{s}_j')=W_j$ and $\mathbf{E}(\hat\sigma_j^2)=\sigma_j^2$ we obtain

$$\begin{aligned}(26)\quad & \frac{1}{2}\mathbf{tr}\,\Delta W_j^{-1}\Delta W_j^{-1}+\frac{1}{2}\tau_j^2[(n_j-p)\sigma_j^{-4}+\mathbf{tr}\,W_j^{-1}D_jW_j^{-1}D_j]+\\ &+\zeta'Z_j'W_j^{-1}Z_j\zeta+\tau_j\mathbf{tr}\,\Delta W_j^{-1}\Delta W_j^{-1}.\end{aligned}$$

Iterations of the scoring method can now be described in a simple way. We first update θ by $\theta=(Z'W^{-1}Z)^{-1}Z'W^{-1}\hat b$. This leaves us with a quadratic in Δ and the τ_j. The optimum value of τ_j, in terms of Δ, is computed next. This is

$$(27)\qquad \hat\tau_j=\frac{(n_j-p)(\sigma_j^{-2}-\hat\sigma_j^2\sigma_j^{-4})+\mathbf{tr}\,W_j^{-1}D_j+tr\,\Delta W_j^{-1}D_jW_j^{-1}-\underline{s}_j'W_j^{-1}D_jW_j^{-1}\underline{s}_j}{(n_j-p)\sigma_j^{-4}+\mathbf{tr}\,W_j^{-1}D_jW_j^{-1}D_j}$$

If we substitute this we still have a quadratic in Δ only, which is then minimized. Then substitute the resulting Δ in Equation (27). This process is fairly efficient and can easily be adapted to cases in which some parameters are constrained to be equal to zero or constrained to be equal to each other. From Equation (26) we also obtain the information matrix directly, and thus we can deduce

the dispersion matrix of the asymptotic normal distribution of the estimates.

References

Aitkin, M. A., & Longford, N. T. (in press). Statistical modelling issues in school effectiveness studies. Journal of the Royal Statistical Society.

Alker, H. R. (1969). A typology of ecological fallacies. In M. Dogan & S. Rokkan (Eds.), Quantitative ecological analysis in the social sciences. Cambridge, MA: MIT Press.

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

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

Aragon, Y. (1984). Random variance linear models: Estimation. Computational Statistics Quarterly, 1, 295-309.

Averch, H., Carroll, S. J., Donaldson, T., Kiesling, H. J., & Pincus, J. (1974). How effective is schooling? A critical review of research. Englewood Cliffs, NJ: Educational Technology Publications.

Barwick, D. T., & Gilbert, J. D. (1974). On generalizations of the reverse order law. SZAM Journal of Applied Mathematics, 27, 326-330.

Blalock, H. M. (1979). Measurement and conceptualization problems: The major obstacle to integrating theory and research. American Sociological Review, 44, 881-894.

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

Boyd, L. H., & Iversen, G. R. (1979). Contextual analysis: Concepts and statistical techniques. Belmont, CA: Wadsworth.

Brookover, B. C., Flood, B., Schweiser, I., & Wisenbaker, I. (1979). School social systems and student achievement: Schools can make a difference. New York: Bergin.

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. (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, I: Issues in micro-analysk. Cambridge, MA: Ballinger.

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

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

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 (Eds.), Re-analyzing program evaluations. San Francisco: Jossey-Bass.

Burstein, L., Miller, M. D., & Linn, R. L. (1979). The use of within-group slopes rn indices of group outcomes. Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation.

Chow, G. C. (1984). Random and changing coefficient models. In Z. Griliches & M. D. Intriligator (Eds.), Handbook of econometrics (Vol. 2). Amsterdam, The Netherlands: North Holland Publishing.

Cline, R. E. (1964). Note on the generalized inverse of the product of matrices. SIAM Review, 6, 57-58.

De Leeuw, J. (1985). Path models with random coefficients. Leiden, The Netherlands: Department of Data Theory FSWIRUL.

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.

Dreeben, R., & Thomas, J. A. (Eds.). (1980). The analysis of educational productivity. Volume I, Issues in microanalysis. Cambridge, MA: Ballinger.

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 the Cultural Basis of Education, Paris, France.

Ecob, R. (1985). Multilevel mixed linear models and their application to hierarchically Random Coefficient Models nested data. Paper presented at the European Meeting of the Psychometric Society, Cambridge, Great Britain.

Goldstein, H. (in press). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika.

Greville, T. N. E. (1966). Note on the generalized inverse of a matrix product. SIAM Review, 8, 518–521.

Hannan, M. T. (1971). Aggregation and disaggregation in sociology. Lexington, MA: Heath-Lexington.

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. Statistics Neerlandica, 20, 1-8.

Infante, A. (1978). Die MINQUE-Schatzung bei verlaufskurvenmodellen mit zufalligen regressionskoeffizienten. Unpublished doctoral dissertation, University of Dortmund, Federal Republic of Germany.

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 re-gression, with applications 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 in MINQUE-theory: Nonnegative estimation, consistency, asymptotic normality and explicit formulae. Mathematische Oper-ationsforschung und Statistik, 11, 563-588.

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

Langbein, L. I. (1977). Schools or students: Aggregation problems in the study of student achievement. Evaluation Studies Review Annual, 2, 27c298.

Lazarsfeld, P. F., & Menzel, H. (1961). On the relation between individual and collective properties. In A. Etzioni (Ed.), Complex Organization: A sociological reader. New York: Holt, Rinehart, and Winston.

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

Longford, N. T. (1985). Mixed linear models and an application to school effectiveness. Computational Statistics Quarterly, 2, 109-117.

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

MacDuffee, C. C. (1946). The theory of matrices. Berlin, FRG: Springer Verlag.

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

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

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

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, 3&31, 483-509.

Oosthoek, H., & Van den Eeden, P. (Eds.). (1984). Education from a multilevel perspective: Models, methodology, and empirical findings. New York: Gordon & Breach.

Peschar, J. (1975). Milieu, school, beroep. Groningen, The Netherlands: Tjeek 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.

Purkey, S. C.,& Smith, M. S. (1983). Effective schools: A review. Elementary School Journal, 4, 427452.

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, 447458.

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 NML 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 (Vol. 1). Amsterdam, The Netherlands: North Holland Publishing.

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

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 maximum likelihood estimates of patterned covariance matrices by the EM algorithm. Biometrika, 69, 657460.

Smith, A. F. M. (1973). A general Bayesian linear model. Journal of the Royal Statistical Society, B35, 67-75.

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

Streitberg, B. (1977). Schatzung von Kovarianzstrukturen in linearen Zwei-und Mehrebenenmodellen. Unpublished doctoral dissertation, Free University of Berlin, Federal Republic of Germany.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. Econometrics, 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. Math-ematische Operationsforschung und Statistik, 11, 545-561.

Van den Eeden, P. (1985a). The conditional type of multilevel theory in educational research. Paper presented at the International Seminar on Linking Micro- and Macro-Approaches, Tel Aviv, Israel.

Van den Eeden, P. (1985b). A two-steps procedure for analyzing multilevel structured Random Coefficient Models data. In W. E. Saris & I. N. Gallhofer (Eds.), Sociometric Research. London: MacMillan Press.

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

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. The American Statistician, 39, 95-101.

STATE UNIVERSITY OF LEIDEN

UNIVERSITY OF AMSTERDAM